Genomics of the globally distributed Echinoid genus *Tripneustes*

A dissertation submitted to the graduate division of the University of Hawai'i at Mānoa in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Zoology (Ecology, Evolution, & Conservation Biology)

May 2018

By
Áki Jarl Láruson

Dissertation Committee:

Floyd A. Reed, Chairperson
Robert C. Thomson
Robert J. Toonen
Daniel Rubinoff
David B. Carlon

Keywords: Marine Invertebrate, Phylogenomics, Transcriptomics, Population Genomics

DEDICATION

I dedicate this dissertation to my grandfather, Marteinn Jónsson (*née* Donald L. Martin).

Acknowledgements

Every step towards the completion of this dissertation has been made possible by more people than I could possibly recount. I am profoundly grateful to my teachers, in all their forms, and especially my undergraduate advisor, Dr. Sean Craig, of Humboldt State University, for all the opportunities he afforded me in experiencing biological research. My dissertation committee deserves special mention, for perpetually affording me pressing encouragement, but also providing an attitude of support and positivity that has been formative beyond measure. My mentor and committee chair, Dr. Floyd Reed, has provided me with perspectives, insights, and advise that I will carry with me for the rest of my life. My family, although far from the tropical shores of Hawai'i, have been with me in so many ways throughout this endeavor, and I am so profoundly grateful for their love and support.

Abstract

Understanding genomic divergence can be a key to understanding population dynamics. As global climate change continues it becomes especially important to understand how and why populations form and dissipate, and how they may be better protected. To this effect, the globally distributed sea urchin genus *Tripneustes* has been highlighted as an ideal group for studying patterns of genomic divergence as the global distribution is split into two physically separated species (*T. ventricosus* in the Atlantic and *T. gratilla* the Pacific), and cryptic divergence in the absence of hard physical barriers has been suspected within each ocean. Molecular signatures of population divergence can be affected and skewed by a number of different biological realities. In the case of lower fitness of a heterozygous individual (underdominance), the degree as well as network shape of the connectivity between populations can determine wether rare alleles persist between populations, muddying population divergence signals, or are driven to fixation at one end of the population range while going extinct in the other, giving a signal of parapatric speciation. In order to address questions regarding the more nuanced molecular differences and broader evolutionary trajectories within the genus *Tripneustes* a draft transcriptome for the species *T. gratilla* was generated. In addition to showing an expansion in tumor suppressor genes when compared to the genome enabled sea urchin *Strongylocentrotus*

purpuratus, and sex-specific gene expression differences in Sex determining Region Y-associated High Mobility Group box (SOX) genes, the transcriptome allowed for easy recovery of the full mitochondrial genome. Following isolation and sequence confirmation, the mitochondrial genome of *T. gratilla* was next compared to all previously published sea urchins mitochondrial genomes. A phylogenetic comparison validates the morphologically proposed superfamily Odontophora, with an estimated genesis of the group during the Eocene-Oligocene epoch transition. Estimates of selection via proportional non-synonymous to synonymous site substitution ratios suggest that purifying selection is the primary force acting on echinoid mitochondria.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Representatives of the sea urchin genus *Tripneustes* can be encountered throughout shallow tropical (and even some temperate) waters. Despite its broad distribution the accepted taxonomy of the genus currently includes just three species: *T. ventricosus* in the Atlantic, *T. gratilla* in the Indian Ocean & West Pacific, *T. depressus* in the East Pacific, as well as a single sub-species *T. gratilla elatensis* in the Red Sea (Dafni 1983, Fell 1974). All of these species and the sub-species have been defined by morphological characters. An especially useful aspect of these organisms' biology in studying patterns of gene flow is that mating is exclusively accomplished through synchronous broadcast spawning, where individuals regularly release a plume of gametes into the water column, so egg and sperm encounters are effectively random. This nicely circumvents the potential problems of sexual conflict impacting gene flow, which can dynamically effect patterns of divergence (Gavrilet & Hayashi 2005, Kaneshiro 2006). These biological characteristics make these sea urchins a near ideal approximation of a Wright-Fisher population, and therefore very useful in assessing population models of genetic divergence (Charlesworth 2009). In 1954 Ernst Mayr published a paper broadening his taxonomic range of evolutionary inference

by examining the geographic speciation of tropical sea urchins, including the genus *Tripneustes*. His conclusion from studying the distribution of these animals was that just as in the terrestrial domain geographic isolation was the primary, if not only, method of evolutionary divergence for sexually reproducing animals (Mayr 1954). In 2005 Palumbi & Lessios analyzed the same groups of sea urchins as Mayr using a phylogenetic approach based on mitochondrial COI and nuclear bindin gene sequences. Their conclusions were in line with Mayr's evolutionary model of speciation, referred to as the "evolutionary animation", which holds that more recently diverged species require strong vicariance (e.g. physical barriers) to exist, while only older divergences can persist in sympatry. Further analysis suggested that within the Pacific ocean there effectively exists only one continuously connected population complex, with no discernible genetic variation between *T. gratilla* and *T. depressus* (Lessios et al. 2003, Palumbi & Lessios 2005). It was concluded that greater population subdivision is evident between the East and West Atlantic, among *T. ventricosus*, than across the entire Pacific ocean.

If the enormous supposed meta-population that occupies the largest ocean body on the planet (165.25 million km2) maintains little to no genomic divergence, there must be some remarkably strong drivers of stabilizing/purging selection at work across the Pacific. However, if greater divergence across the genomes of sympatric Pacific populations is identified, as the preliminary analysis suggests be the case, the genome regions with the greatest level of divergence can provide insight into the mechanisms that are driving the split, e.g. if greater divergence is seen in genes implicit in fertilization success than elsewhere in the genome. As the sequencing of whole genomes is no longer prohibitively costly, it is now possible to analyze differential genomic level divergence within one of the groups Mayr originally used to define the dominance of allopatric speciation.

Chapter 2 of this dissertation focuses on patterns of population connectivity from a network perspective, defining a situation where a rare genetic variant under the condition of lower heterozygote fitness (underdominance) can persist in highly structured populations. The effect of patterns of connectivity in a structured population are highlighted and the implication for these patterns on various biological systems, such as broadcast spawning marine invertebrates, is discussed.

Chapter 3 of this dissertation describes the usefulness and necessity for increased genome level resources for marine invertebrate species of interest, culminating in the generation of a draft transcriptome from multiple tissues of two adult *Tripneustes gratilla*, as well as a cohort of larvae. Expression profiles of an adult male and female *T. gratilla* collected from O'ahu, Hawai'i are explicitly compared, with specific focus on select expression differences in purported sex determination genes. Attention is also given to several potentially toxic gene expression products.

Finally, Chapter 4 of this dissertation defines the evolutionary relationship between the mitochondrial genome of *T. gratilla* and those of fourteen other echinoids across the Order Camaradonta. Comparing rates of molecular evolution, through the use of calibration points, a timeframe for the genesis of the Superfamily Odontophora is proposed.

1.1 References

Charlesworth, B. (2009) Effective population size and patterns of molecular evolution and variation. Nature Rev Genet 10:195-205.

Dafni, J. (1983) A New Subspecies of *Tripneustes gratilla* from the Northern Red Sea. Isr J Zool 32:1-12.

Fell, F. J. 1974. The echinoids of Easter Island (Rapa Nui). Pac Sci. 28(2):147-158.

Gavrilet, S., Hayashi, T.I. (2005) Speciation and sexual conflict. Evol Ecol 19:167-198.

Kaneshiro, K.Y. (2006) Dynamics of sexual selection in the Hawaiian Drosophilidae: A paradigm of evolutionary change. Proc Hawaiian Entomol Soc 38:1-19.

Lessios, H.A., Kane, J., Roberton, D.R. (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: Contrasting patterns of population structure between oceans. Evolution 57(9):2026-2036.

Mayr, E. (1954) Geographic Speciation in Tropical Echinoids. Evolution 8(1):1-18.

Palumbi, S.R., Lessios, H.A. (2005) Evolutionary animation: How do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? PNAS 102:6566-6572.

Chapter 2

Stability of Underdominant Genetic Polymorphisms in Population Networks

2.1 Abstract

Heterozygote disadvantage is potentially a potent driver of population genetic divergence. Also referred to as underdominance, this phenomena describes a situation where a genetic heterozygote has a lower overall fitness than either homozygote. Attention so far has mostly been given to underdominance within a single population and the maintenance of genetic differences between two populations exchanging migrants. Here we explore the dynamics of an underdominant system in a network of multiple discrete, yet interconnected, populations. Stability of genetic differences in response to increases in migration in various topological networks is assessed. The network topology can have a dominant and occasionally non-intuitive influence on the genetic stability of the system.

2.2 Introduction

Variation in the fitness of genotypes resulting from combinations of two alleles (*e.g.*, A- and B-type alleles combined into AA-, AB-, or BB-genotypes resulting in w_{AA}, w_{AB}, and w_{BB} fitnesses respectively) result in different evolutionary dynamics. The case in which a heterozygote has a lower fitness than either homozygote, $w_{AB} < w_{AA}$ and $w_{AB} < w_{BB}$, is termed underdominance or heterozygote disadvantage. In this case there is an internal unstable equilibrium so that the fixation or loss of an allele depends on its starting frequency. In a single population, stable polymorphism is not possible. However, when certain conditions are met, populations that are coupled by migration (the exchange of some fraction of alleles each generation) can result in a stable selection-migration equilibrium. This selection-migration equilibrium is associated with a critical migration rate (m^*); above this point the mixing between populations is sufficiently high for the system to behave as a single population and all internal stability is lost (Altrock et al. 2010).

Underdominance can be thought of as an evolutionary bistable switch. From the perspective of game-theory dynamics it can be interpreted as a coordination game (Traulsen et al. 2012). The properties of underdominance in single and multiple populations have led to proposals of a role of underdominance in producing barriers to gene flow during speciation (Faria & Navarro 2010, Harewood et al. 2010) as well as proposals to utilize underdominace both to transform the properties of wild populations in genetic pest management applications (Curtis 1968, Davis et al. 2001, Sinkins & Gould 2006, Reeves et al. 2014) and to engineer barriers to gene flow (transgene mitigation) from genetically modified crops to unmodified relatives (Soboleva et al 2003, Reeves & Reed 2014).

The properties of underdominance in a single population are well understood (Fisher 1922, Haldane 1927, Wright 1931) and the two-population case has been studied in some detail (Karlin

et al 1972a, Karlin et al. 1972b, Lande 1985, Wilson & Turelli 1986, Spirito et al. 1991, Altrock et al. 2010, Altrock et al. 2011), with fewer analytic treatments of three or more populations (Karlin et al. 1972a, Karlin et al. 1972b). Simulation-based studies have been conducted for populations connected in a symmetrical one- and two-dimensional lattice (Schierup & Christiansen 1996, Payne et al. 2007, Eppstein et al. 2009, Landguth et al. 2015) (see also (Kondrashov 2003, Hoelzer et al. 2008, Barton & De Cara 2009) for extensions to multiple loci) and "wave" approximations have been used to study the flow of underdominant alleles under conditions of isolation by distance (Fisher 1937, May et al. 1975, Barton et al. 1991, Piálek & Barton 1997, Soboleva et al. 2003, Barton, et al. 2011). However, despite this body of literature, underdominance remains relatively neglected in population genetic research (Bengtsson & Bodmer 1976). Models in which allele frequencies are distributed in continuous populations are easier to analyse, and are appropriate approximations when selection is weak and the number of discrete demes are large (Barton 1979). However, here we are interested in the cases where the number of demes are small or selection is strong (as is the case in potential applications of underdominace) or demes are connected in more complex topologies (*e.g.*, assymetrical arrays). This is where the continuous approximation breaks down and many of the effects we see in the context of small network topology have been previously overlooked. Furthermore, a large focus of earlier theoretical work with underdominance was on how new rare mutations resulting in underdominance might become established in a population (Wright 1941, Bengtsson & Bodmer 1976, Hedrick 1981, Walsh 1982, Hedrick & Levin 1984, Lande1984, Lande1985, Barton & Rouhani 1991, Spirito 1992). However, here we are addressing the properties of how underdominant polymorphisms may *persist* once established within a set of populations rather than how they were *established* in the first place.

We explicitly focus on discrete populations that are connected by migration in a population network. We have found that the topology of the network has a profound influence on the

stability of underdominant polymorphisms aspects of which have been otherwise overlooked. This influence is not always intuitive *a priori*. These results have implications for the effects of network topology on a dynamic system (see for review Strogatz 2001), particularly for interactions related to the coordination game (such as the stag hunt Skyrms 2001), theories of speciation, the maintenance of biological diversity, and applications of underdominance to both protect wild populations from genetic modification or to genetically engineer the properties of wild populations—depending on the goals of the application.

2.3 Methods and Results

We are considering simple graphs in the sense of graph theory to represent the population network: each pair of nodes can be connected by at most a single undirected edge. A graph $\mathbb{G} = (\mathscr{N}, \mathscr{E})$, is constructed from a set of nodes, $\mathscr{N}$ (also referred to a vetexes), and a set of edges, $\mathscr{E}$, that connect pairs of nodes. For convenience $V = |\mathscr{N}|$ and $E = |\mathscr{E}|$, we chose V (for vertex) to represent the number of nodes to avoid future conflict with N symbolizing finite population size in population genetics. A node corresponds to a population made up of a large number of random-mating (well mixed) individuals (a Wright-Fisher population, (Fisher 1922, Wright 1931) with independent Hardy-Weinberg allelic associations, (Hardy 1908)) and the edges represent corridors of restricted migration between the populations. We are also only considering undirected graphs: in the present context this represents equal bidirectional migration between the population nodes. Furthermore, we are only considering connected graphs (each node can ultimately be visited from every other node) and unlabeled graphs so that isomorphic graphs are considered equivalent.

The network graph $\mathbb{G}$ is represented by a symmetric $V \times V$ adjacency matrix $\mathscr{A}$.

$$\mathscr{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,V} \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots & a_{2,V} \\ a_{3,1} & a_{3,2} & a_{3,3} & \dots & a_{2,V} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{V,1} & a_{V,2} & a_{V,3} & \dots & a_{V,V} \end{bmatrix}$$

The presence of an edge between two nodes is represented by a one and the absence of an edge is a zero. The connectivity of a node is

$$c_i = \sum_{j \in \mathscr{N}} a_{i,j}$$

Each generation, g, the allele frequency, p of each population node, i, is updated with the fraction of immigrants from n adjacent populations, j, at a migration rate of m.

$$p_{i,g} = (1 - c_i m) p_{i,g-1} + \sum_{j=1}^{N} m a_{i,j} p_{j,g-1}$$

Note that this equation will not be appropriate if the fraction of alleles introduced into a population exceeded unity. See the discussion of the star topologies illustrated in Figure 2.1 and the Supplemental Methods for discussion of an alternative approach.

The frequencies, adjusted for migration, are then paired into genotypes and undergo the effects of selection. The remaining allelic transmission sum is normalized by the total transmission of all alleles to the next generation to render an allele frequency from zero to one.

$$p'_{i,g} = \frac{p_{i,g}^2 + \omega p_{i,g}(1 - p_{i,g})}{p_{i,g}^2 + 2\omega p_{i,g}(1 - p_{i,g}) + (1 - p_{i,g})^2}$$

Note that here for simplicity we set the relative fitness of the homozygotes to one, $w_{AA} = w_{BB} = 1$ and the heterozygote fitness is represented by $w_{AB} = \omega$.

2.3.1 Analytic Results

Underdominance in a single population has one central unstable equilibrium and two trivial stable equilibria at $p = 0$ and $p = 1$. When one considers multiplying the three fixed points of a single population into multiple dimensions it can be seen that, if migration rates are sufficiently small, nine equilibria (3×3) exist in the two-population system. Again, there is a central unstable equilibrium, the two trivial stable equilibria, and two additional internal stable equilibria. The remaining four fixed points are unstable saddle points that separate the basins of attraction (see Figure 3 of Altrock et al. 2010). However, as the migration rates increase the internal stable points merge with the neighboring saddle points and become unstable themselves (Karlin & McGregor 1972, Altrock et al. 2010). In three populations there are a maximum of 27 ($3 \times 3 \times 3$) fixed points. The three types found in a single population, six internal stable equilibria, and 18 saddle points. The general pattern is that in V populations there are a maximum of 3^V equilibria if migration rates are sufficiently low. There will always be one central unstable equilibrium (a trajectory starting near this point can end up in any of the basins of attraction) and two trivial fixation or loss points. A V-dimensional hypercube represents the state space of joint allele frequencies of a V-population system. This hypercube has 2^V vertexes (corners). It can be seen that the internal stable points, if they exist, are near these corners when the corners represent a mix of zero and one allele frequencies. There are exactly two corners, the trivial global fixation and loss points, that do not contain a mix of unequal coordinate frequencies. Thus, there are $2^V - 2$ possible internal stable points that represent alternative configurations of a migration-selection equilibrium polymorphism. Finally there are up to $3^V - 2^V - 3$ saddle points that separate 2^V possible (but never less than two) domains of attraction.

We have a system of equations, $p'_{i,g}$, that describe the dynamics of allele frequency change within each population and these dynamics are coupled by migration. It is useful to think of the difference in frequency each generation, $\delta_{p,i} = p'_{i,g} - p'_{i,g-1}$. We can set $\delta_{p,\forall i} = 0$ to solve for fixed points in the state space. However, the system needs to be simplified in order to be tractable. For example if we only look along the $p_1 = p_2$ axis in the two-population case we get three solutions, $p = 0$, $p = 1$, and

$$\hat{p}_{\forall i \in \mathcal{N}} = \frac{w_{AA} - w_{AB}}{w_{AA} + w_{BB} - 2w_{AB}}$$

in the general case and

$$\hat{p}_{\forall i \in \mathcal{N}} = \frac{1 - \omega}{2 - 2\omega} = \frac{1}{2}$$

in the simplified (equal homozygote fitness) case. The first two points are the trivial loss of polymorphism. The third is identical to the internal unstable equilibrium point in a single population (Altrock et al. 2010).

In fact this unstable equilibrium point is always found along the $p_1 = p_2 = \cdots = p_V$ axis. Note that the migration rate, m, is not a part of the solution. The position of this point in the state space is independent of migration rates. Since it falls along the axis where the allele frequencies of all populations are equal, migration between populations, and in fact the population network topology itself, has no effect.

While the position of this internal point remains fixed regardless of the number of interacting populations in the network, there is a multiple population effect on the rate of change away from this point. Solving for the eigenvalues, λ_i, of a Jacobian matrix, $\mathbb{J}$, of partial derivatives of the system along the $p_1 = p_2 = \cdots = p_V$ axis shows that the rate of change follows the pattern

$$\lambda_1|_{p=\frac{1}{2}} = \frac{2V}{1 + \omega} - V.$$

Thus, as the number of interconnected populations increase the magnitude of flow away from

the internal point increases. This rate is a function of both the heterozyote fitness (ω) and the number of interlinked populations. At $\omega = 0$, or lethality of the heterozygous condition, the rate of change is equal to the number of connected populations.

We are also interested in the internal stable equilibria that allow differences in allele frequencies among populations to be maintained. In the two-population special symmetric case this can be solved from $\mathbb{J}$ along the $p_1 = 1 - p_2$ axis and yields

$$\lambda_2 = -\frac{1 - \omega + 2m^*(2m^* - 3)}{(1 - 2m^*)^2}$$

$$m^*|_{\lambda_2 = 0} = \frac{1}{4}\left(3 - \sqrt{5 + 4\omega}\right)$$

(see Appendix A of Altrock et al. 2010, for more detail). Unfortunately, with three or more populations, even with highly symmetrical configurations, we have not found a single axis or plane through the state space that captures these internal stable points. Therefore, we have used numerical methods to characterize the critical points.

2.3.2 Numerical Simulations

Sets of populations were initialized with approximately half (depending on if there were an even or odd number of populations) of the nodes in the network at allele frequencies near zero and half near a frequency of one. The allele frequencies were offset by a small random amount from zero or one to avoid being symmetrically balanced on unstable trajectories. Migration rates were slowly incremented at each step and the system was allowed to proceed to near equilibrium, when the difference in allele frequencies between generations was less than 10^{-10}, before the next step. The process was repeated until the point where a collapse in differences of allele

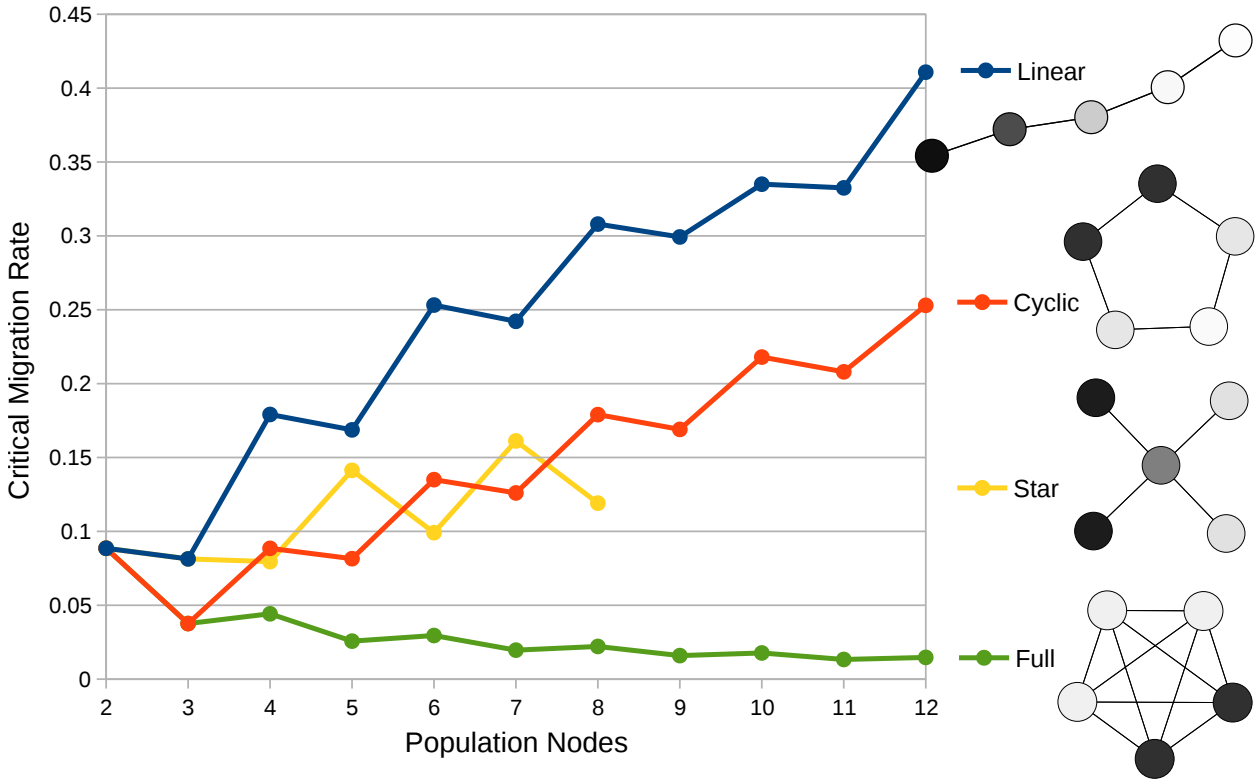

FIGURE 2.1: The stability of some simple network configurations. For each number of nodes and network topology (linear in blue, cyclic in red, star-like in yellow, and fully interconnected in green) the critical migration rate allowing polymorphic underdominant polymorphism to persist at $\omega = 1/2$ is plotted. Examples of each type of network at $V = 5$ are plotted as graphs in the legend to the right. The shading of the nodes represents the allele frequency between zero and one of each population near the critical migration rate value.

frequencies between all populations was detected. This point was then reported as the critical migration rate of the network.

Example Topologies

The stability of a range of basic network topologies were investigated, shown in Figure 2.1. In general, in these examples, the diameter of the network is predictive of migration-selection stability. Linear configurations had the highest stability, cyclic configurations approximately reduced the both the diameter and stability in half. Fully interconnected populations had both the smallest diameter and lowest stability.

Another pattern that became apparent is an even-odd alternation in stability. Except for starlike networks, an odd number of nodes results in a relatively lower stability than an even number of nodes.

Starlike networks showed an interesting pattern. They were approximately of the same stability as cyclic networks; however, the even-odd alternation was inverted—odd V graphs showed enhanced stability, showing that the even-odd pattern is not absolute. At this heterozygote fitness ($\omega = 1/2$) starlike networks with greater than eight nodes could not be evaluated. At $V \geq 9$ before the critical migration rate is reached the total amount of immigration into the central population exceeds 100%.

Fully interconnected networks had the lowest stability and by far the greatest number of edges. Unlike the other networks the fully-interconnected systems declined at higher V. However, the number of edges grew much faster than the number of nodes. Note that the even-odd alternation in relative stability is still apparent, even in these graphs.

The effects of a range of topologies for six nodes and five edges was also explored, shown in Figure 2.2. All of these networks have identical treeness, $\phi_{tree} = 1$ (*sensu* Xie & Levinson 2007). In

general the stability is correlated with the diameter of the network. However, the clear exception is the “double-Y” topology, which has the highest stability of all. This has inspired an alternative measure of the treeness of a network that we will refer to as “dendricity” to avoid conflicting with prior definitions of treeness in the literature.

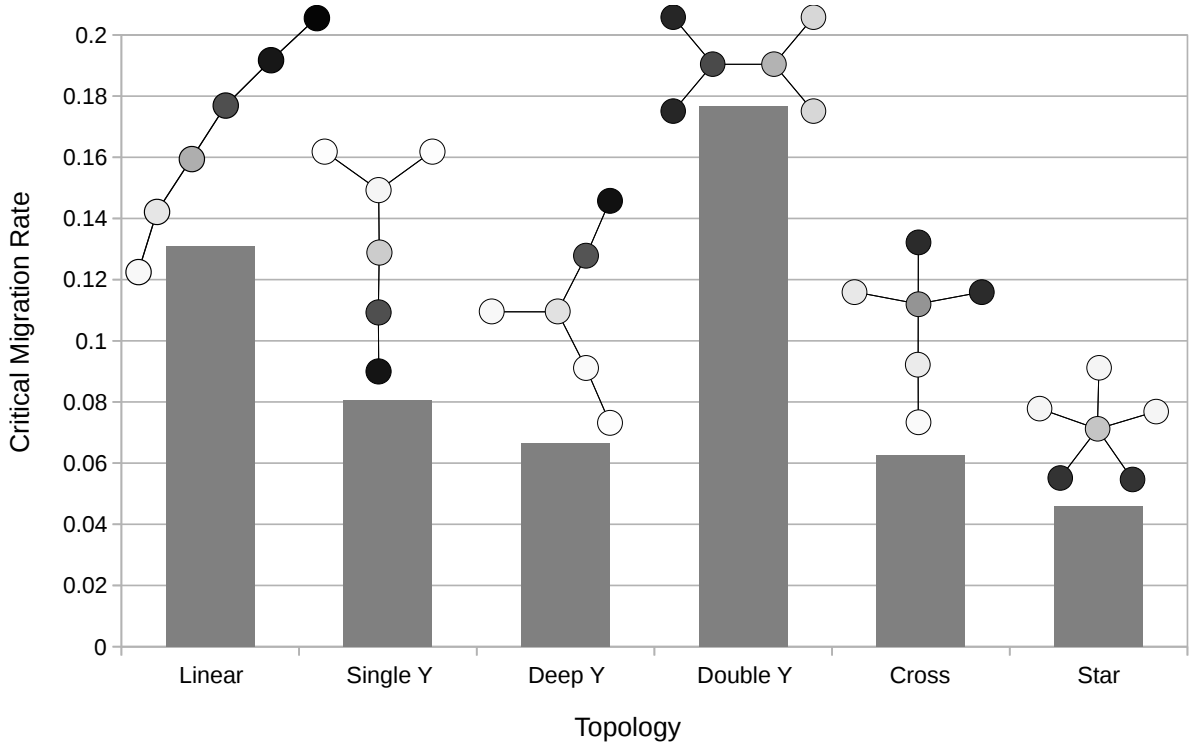

FIGURE 2.2: The stability of all possible simple connected networks made of six populations with five corridors of migration. Here the critical migration rate was evaluated at a relative heterozygote fitness of $\omega = 3/4$. The networks are arranged by diameter of the network declining from five on the left to two on the right. The shading of the nodes represents the allele frequency between zero and one of each population near the critical migration rate value. The shading of the bars is set to 50% gray to aid visualization of the allele frequencies of the nodes.

Random Graphs

In order to evaluate general correlations between migration-selection stability and network summary statistics we generated 100,000 random connected graphs of up to 20 nodes in size and evaluated their stability. The results are summarized in Supplementary Table 1. We found that the most stable network configurations contained nodes with at most three edges (see the next section below) so, in order to avoid the problem of the rate of immigration more than replacing a local population we set a maximum migration rate of $m = 1/3$ and reported this as m^* for the subset of highly stable network topologies—only 0.64% of the random networks reached this point of $m = 1/3$—these highly stable networks are explored in the next section.

The following parameters were estimated from these networks: *Variance*, which refers to the variance in connectivity (c_i the number of edges incident with node i) of all the nodes in the network. *Efficiency* refers to the shortest path lengths between nodes in the network according to

$$\phi_{efficiency} = \frac{1}{V(V-1)} \sum_{i<j \in V} \frac{1}{d(i,j)}$$

where $d(i,j)$ is the minimum path length between nodes i and j. The *diameter* of the graph is the maximum $d(i,j) \in \mathbb{G}$. *Dendricity* is the fraction of nodes incident with three edges where at least one edge is a "bridge" edge (removing bridge edges results in an unconnected graph) out of the total number of internal nodes. *Evenness* is simply a binary variable of zero or one to indicate if an odd or even number of nodes are present in the graph. Finally, *terminalness* indicates the fraction of nodes in the graph that are terminal (or leaf) nodes.

Each of the summary statistics we addressed were significant predictors of network stability; however, because of correlations between these measures caution must be used to interpret the results. For example, contrary to intuition the number of nodes was negatively correlated with stability. This is because the number of possible edges, which generally lower stability, increases

dramatically with the number of nodes. When the number of edges is controlled the number of nodes becomes strongly positively correlated. Simply the number of nodes per edge (V/E) is a powerful predictor of stability. In general *diameter* is a strong predictor of stability, particularly if the number of nodes are held constant (compare to Figure 2.2). *Dendricity* and *terminalness* also performed well as general predictors of stability. *Evenness* continues to be a predictor of stability, specifically even ordered networks have a greater stability than odd ordered ones, but the predictive power is weak compared to other measures. *Variance* and *efficiency* are a bit more difficult to understand. Increased variance in the number of edges per node is associated with lower stability, while intuitively one might expect the opposite. When this is measured as *variance* divided by the total number of edges the correlation almost disappears and in fact becomes slightly positive—as expected—unless the number of nodes are held constant. *Efficiency* also varies in a non-intuitive way. It is positively correlated unless either the number of nodes or the number of edges are held constant where it becomes strongly negatively correlated; however, if both are constant efficiency becomes slightly positively correlated again.

When exploring model selection to find the minimal adequate model to predict m^* via adjusted r^2, and Mallows' C_p as implemented in the R package "leaps" all ten summary statistics were retained using all four methods (R Development Core Team 2008, Lumley 2009). Using all of the predictors in the full linear model explains the majority, $r^2 = 0.72$, of the variation in m^*.

Evolving Networks

In order to more fully explore the upper edge of highly stable networks for a given V we wrote a program that would make random changes to the network and evolve higher stability configurations. Starting from a fully interconnected network with as close to half of the nodes near an allele frequency of zero or one as possible, edges were randomly selected to be removed or added with the constraint that the new network remain connected. Most frequently a single

edge was altered but with reducing frequency two or more edges could be changed simultaneously to allow larger jumps in topology and movement away from locally stable configurations. When a network is altered its m^*_{new} value is determined. If the new critical value is higher than the value of the current network $m^*_{current}$, the new network is adopted for the next step. If the new critical value is lower than the current network, the new network is adopted with a probability equal to the ratio of the new and current critical migration values ($m^*_{new}/m^*_{current}$). This also allows the network evolution to explore regions off of local maxima. Up to $V = 5$ the most stable network configuration was a linear topology. From $V \geq 6$ networks with greater stability than the linear configuration were found and are illustrated in Figure 2.3. Note that the number of possible connected networks increases dramatically with larger V. The most stable networks found in Figure 2.3 are not expected to result from an exhaustive search, particularly for $V \geq 10$. They are however meant to illustrate some general properties of highly stable networks.

2.3.3 Software Availability

All simulations for both random and evolving networks were written in Python 2.7.10. The code is freely available on GitHub: `https://github.com/akijarl/NetworkEvolve`

2.4 Discussion

One result that is beginning to emerge from the study of evolutionary dynamics on graphs is that the resulting properties can be sensitive to the network topology, but often in a non-intuitive way, (*e.g.* Hindersin & Traulsen 2014, Hindersin & Traulsen 2015). There are some general factors that influence, or are predictive, of the stability of underdominant polymorphisms in a population network. In smaller networks the influence of diameter and evenness are apparent. The larger

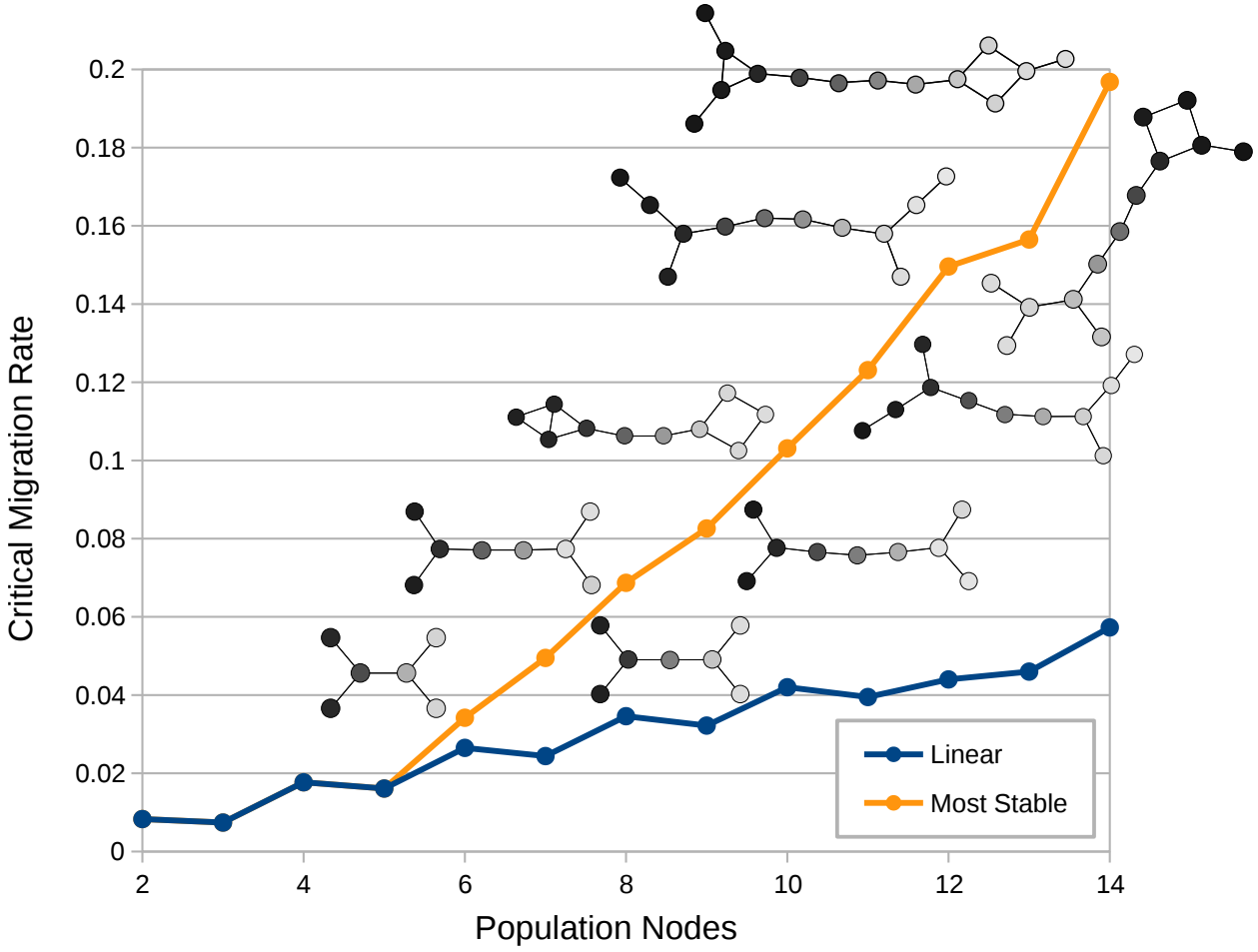

FIGURE 2.3: A plot of the stability of highly evolved networks ($\omega = 0.95$). For comparison the stability of corresponding linear structures is also plotted (linear, blue). The most stable network found for each V is given (at the end of 10,000 steps of random changes with a Metropolis-Hastings-like chain update, starting from a fully interconnected network, over 8 independent replicate runs), even to the left and odd to the right near each corresponding plotting point. With small change, networks can be substantially more stable than the linear configuration. This seems to derive from a balance of increasing diameter and forked anchoring structures at either end.

the diameter of the network the effectively lower the migration rate is between the edges of the system, because alleles have to be exchanged via intermediate nodes; it is well understood that a lower migration rate enhances migration-selection stability (Altrock et al. 2010). In contrast the more edges there are in a system the higher the effective migration rate across the network, which results in lower stability.

The role of the even-odd number of nodes is more subtle. In a system of coupled populations having an allele frequency near $p = 1/2$ is inherently unstable with underdominant fitness effects. An odd ordered network that is anchored at high and low frequencies near its edges pushes the central population near $p = 1/2$, which destabilizes the entire system. Of course there are exceptions to this rule such as the star network topology (star networks also have other unusual properties such as acting as amplifiers of selection Frean et al. 2013, Adlam et al. 2015, Hindersin & Traulsen 2015), and the importance of evenness declines with larger networks. This pattern may also change if there are three or more alleles that are underdominant with respect to each other. Interestingly, this effect seems to have been completely overlooked in previous work using either numerical techniques or wave approximations to study underdominant-like effects.

The networks that were evolved to higher stability illustrate the effects of having two "anchor" nodes at each end of a central linear network "trunk." The anchors are made up of variations on a theme of a node with three edges with at least one of these edges being a bridge edge. Intuitively, the flow of one allele along two paths into a population can overcome the flow of the alternative allele along a single path, thereby enhancing the stability of the anchored edges of the system (*cf.* the discussion of "stem" structures, reservoirs, and the movement of clusters of mutants within "superstar" networks in Jamieson-Lane & Hauert 2015). In contrast a strictly linear network allows adjacent populations to collapse one by one without any local restrictions in gene flow. Additionally the internal linear trunk has a large diameter further restricting gene flow according to the discussion of the effects of diameter above.

As the number of demes in one- or two-dimensions becomes large the natural limit of continuous population structure is approached (Barton 1979). Patterns found in small networks, such as the even-odd stability pattern, become minimized and the dynamics are consistent with continuous approximations. Studies of underdominance in continuous populations have found that polymorphisms can persist when "hybrid zones" are trapped by natural barriers such as regions of lower population density or restrictions in gene flow (Barton 1979, Barton & Turelli 2011). There are also parallels in ecological models, *e.g.* "invasion pinning" in Keitt et al 2001. This is consistent with the highly stable double-anchored tree networks. The central trunk acts as a local restriction to gene flow and pins the cline in allele frequency from invading across the structure (Barton 1979).

2.4.1 Implications

Natural Systems

This work was originally motivated by biological systems. It is interesting to ask, based on these results, where we might expect to see underdominant polymorphism being maintained in wild populations. Population network configurations exist in which even subtle levels of underdominance can maintain stable geographic differences between populations with substantial rates of migration. Chromosomal rearrangements that lead to strong underdominance can occur at relatively high rates and are rapidly established between closely related species (White 1978, Jacobs 1981). Subtle underdominant interactions may be more widespread than previously appreciated and may have played a large role in shaping gene regulatory networks (Stewart et al. 2013). Note also that weak effects among loci can essentially self organize to become coupled and that this effect may extend to a broader class of underdominant-like effects such as the well known Dobzhansky-Muller incompatabilites (Barton

& De Cara 2009, Landguth et al. 2015). We have found that bifurcating tree-shaped (dendritic) networks have very high stability. Key natural occurrences of dendritic habitats include freshwater drainage systems, as well as oceanic and terrestrial ridge systems. Indeed, a role of underdominance in shaping patterns of population divergence across connected habitats has been implicated, either directly or indirectly, in freshwater fish species (Fernandes-Matioli & Almeida-Toledo 2001, Alves et al. 2003, Nolte et al. 2009), salamanders (Fitzpatrick et al. 2009, Feist 2013), frogs (Bush et al. 1977) (see also Wilson et al. 1974), semiaquatic marsh rats (Nachman & Myers 1989), and *Telmatogeton* flies in Hawai'i, which rely on freshwater streams for breeding environments (Newman 1977) (in the case of Dipterans we are ignoring chromosomal inversions which do not result in underdominance in this group (Coyne et al. 1991, Coyne et al. 1993). In fact, alpine valleys around streams also follow a connected treelike branching pattern and there are examples of extensive underdominance in small mammals found in valleys in mountainous regions (Piálek et al. 2001, Basset et al. 2006, Britton-Davidian et al. 2000). To the extent that persisting underdominant and underdominant-like fitness effects may promote speciation (rates of karyotype evolution and speciation are correlated, see Levin & Wilson 1976, Bush et al. 1977) it should be noted that freshwater streams contain 40% of all fish species yet are only 1% of the available fish habitat and that a higher rate of speciation is indeed inferred for freshwater versus marine systems (Bloom et al. 2013).

However, there are also examples of the maintenance of underdominant polymorphisms that are not found in species associated with limnological structures. For example, the island of Sulawesi itself has an unusual branching shape and a large number of terrestrial mammal species with a 90% or greater rate of endemism excluding bats (Groves 2001). Other factors that are associated with underdominant stability are the diameter of the network and having an even order of nodes. The Hawaiian islands essentially form a linear network of four major island groups (Ni'ihau & Kaua'i—O'ahu—Maui Nui—Hawai'i) and are known for their high species diversity

and rates of speciation with examples in birds (Lerner et al 2011), spiders (Gillespie 2004), insects (Magnacca & Price 2015), and plants (Helenurm & Ganders 1985, Givnish et al. 2009). For example, the Hawaiian *Drosophila crassifemur* complex has maintained chromosomal rearrangements between the islands that are predicted to result in underdominance (Yoon et al 1975). Perhaps the network topology of the Hawaiian archipelago (in addition to the diversity of micro-climates, environments, and ongoing inter-island colonizations) has contributed to the high rates of diversification found on these islands.

In contrast, areas where we might expect to see less maintenance of genetic diversity that can contribute to boundaries to gene flow are in highly interconnected networks with low diameters such as, perhaps, marine broadcast spawners with long larval survival times that are associated with the shallow waters around islands (*i.e.*, the network nodes). Examples of a lack of speciation in such groups, distributed over areas as large as half of the Earth's circumference, exist (Palumbi 1992, Lessios et al 2003).

Another type of network is one that is distributed over time rather than space. Underdominant interactions have been inferred in the American bellflower *Campanula americana* (Galloway & Etterson 2005). This species is unusual in that individuals can either be annual or biennial depending on the time of seed germination. Given that the majority of seeds are expected to germinate within a single year (Galloway 2001), even-year biennials may form a somewhat distinct population from odd-year biennials with gene flow occurring by the subset of annual plants—forming an even ordered network. Finally, a tantalizing combination exists in the pink salmon, *Oncorhynchus gorbuscha*, of the North Pacific. This species is both biennial and returns to native freshwater streams to spawn. Indeed, artificial crosses between even and odd year individuals (of the same species) have revealed extensive genetic differences with hybrid disgenesis (Gharrett & Smoker 1991, Limborg et al. 2014).

Applications

Various "transgene mitigation" methods have been proposed to prevent the transfer of genetic modifications from genetically engineered crops to traditional varieties or wild relatives (Lee & Natesan 2006, Daniell 2002, Hills et al. 2007, Kwit et al. 2011) including the use of underdominant constructs (Reeves & Reed 2014, Soboleva et al. 2003). In a species with limited pollen dispersal, it may be tempting to plant a buffer crop area between a GMO crop with underdominant transgene mitigation and an adjacent unmodified population. However, these results suggest that, over multiple generations, this configuration may actually destabilize the system and promote the spread or loss of the genetic modification. In a simplistic scenario a single flanking buffer population results in an odd number of populations, breaking the evenness rule of stability (unless the populations are arranged in essentially a $V = 5$ star pattern, Figure 2.1). Depending on local conditions, it may be preferable to plant two distinct yet adjacent buffer crops or none at all. Of particular note are genetically modified and wild species that exist in freshwater systems, such as rice (Lu & Yang 2009) and fish (Devlin et al. 2006). Our predictions suggest that underdominant containment may in general have enhanced stability in these situations. However, this can work both ways. Genetic modifications may be amiable to underdominant mitigation strategies to prevent establishment in the wild; yet, difficult to remove from freshwater systems if established.

Using underdominance to stably yet reversibly genetically modify a wild population is one goal within the field of genetic pest management. The potential implications of these results depend on the amount of modified individuals that could be released into the wild. If the numbers are sufficiently high to transform an entire region then highly interconnected populations would be ideal to ensure full transformation. However, as is much more likely, if the number of individuals that can be released is much smaller than the total wild population, transformation might best be achieved in a stepwise strategy utilizing linear or treelike population configurations. In the case

of Hawai'i, limiting the effects of avian malaria by modifying non-native Culex mosquitoes has been proposed as a method to prevent further extinctions of native Hawaiian forest birds (Clarke 2002). Linear island archipelagos and their river valleys (Culex are more common at lower elevations, see van Ripper III et al. 1986) may be ideal cases to both transform local populations yet prevent genetic modifications from becoming established outside of the intended area.

2.4.2 Future Directions

It will come as no surprise to an evolutionary biologist that systems with greater genetic isolation (such as freshwater streams versus marine environments) will lead to increased genetic divergence and rates of speciation; however, the implication we are focusing on here is the influence of the population network topology. We are suggesting that, for the same degree of migration rate isolation, alternative network topologies might be compared to inferred rates of speciation and/or enhanced genetic diversity that leads to hybrid dysgenesis. A consideration of the geological history is also appropriate to incorporate effects such as stream capture and the merging of islands on biological diversity. A proper meta-analysis or experimental evolution of this network topology effect is beyond the scope of the current manuscript but would be useful projects to further explore these effects.

2.5 References

Adlam, B., Chatterjee, K., Nowak M.A. (2015) Amplifiers of selection. Proc R Soc A, 471:20150114

Altrock, P.M, Traulsen, A., Reeves, R.G., Reed, F.A. (2010) Using underdominance to bi-stably transform local populations. Journal of Theoretical Biology 267(1):62-75

Altrock, P.M., Traulsen, A., Reed, F.A. (2011) Stability properties of underdominance in finite subdivided populations. PLoS Computational Biology 7(11)

Alves, A.L., Oliveira, C., Foresti, F. (2003) Karyotype variability in eight species of the subfamilies Loricariinae and Ancistrinae (Teleostei, Siluriformes, Loricariidae). Caryologia 56(1):57-63

Barton, N.H., Rouhani, S. (1991) The Probability of Fixation of a New Karyotype in a Continuous Population. Evolution 45(3):499-517

Barton, N.H., Turelli M. (2011) Spatial waves of advance with bistable dynamics: cytoplasmic and genetic analogues of allee effects. The American Naturalist 178(3):E48-E75

Barton, N.H. (1979) The dynamics of hybrid zones. Heredity 43(3):341-359

Barton, N.H., De Cara, M.A.R. (2009) The evolution of strong reproductive isolation. Evolution 63:1171-1190

Basset, P., Yannic, G., Hausser J. (2006) Genetic and karyotypic structure in the shrews of the *Sorex araneus* group: are they independent? Molecular Ecology 15(6):1577-1587

Bengtsson, B.O., Bodmer, W.F. (1976) On the increase of chromosome mutations under random mating. Theoretical Population Biology, 9(2):260-281

Bloom, D.D., Weir, J.T., Piller, K.R., Lovejoy, N.R. (2013) Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (Atherinopsidae). Evolution 67:2040-2057

Britton-Davidian, J., Catalan, J., Ramalhinho, M.d.G, Ganem, G., Auffray, J.-C., Capela, R., Biscoito, M., Searle, J.B., Mathias, M.d.L. (2000) Rapid chromosomal evolution in island mice. Nature 403:158

Bush, G.L., Case, S.M., Wilson, A.C., Patton, J.L. (1977) Rapid speciation and chromosomal evolution in mammals. PNAS 74(9):3942-3946

Clarke, T. (2002) Mosquitoes minus malaria. Nature 419(6906):429-430

Coyne, J.A., Aulard, S., Berry, A. (1991) Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. Genetics, 129(3):791-802

Coyne, J.A., Meyers, W., Crittenden, A.P., Sniegowski, P. (1993) The fertility effects of pericentric inversions in *Drosophila melanogaster*. Genetics, 134(2):487-496

Curtis, C.F. (1968) Possible use of translocations to fix desirable genes in insect pest populations. Nature, 218:368-369

Daniell, H. (2002) Molecular strategies for gene containment in transgenic crops. Nature Biotechnology, 20(6):581-586

Davis, S., Bax, N., Grewe, P. (2001) Engineered underdominance allows efficient and economical introgression of traits into pest populations. Journal of Theoretical Biology, 212(1):83-98

Devlin, R.H.,Sundström, R.F., Muir, W.F. (2006) Interface of biotechnology and ecology for environmental risk assessments of transgenic fish. Trends in Biotechnology, 24(2):89-97

Eppstein, M.J., Payne, J.L., Goodnight, C.J. (2009) Underdominance, multiscale interactions, and self-organizing barriers to gene flow. Journal of Artificial Evolution and Applications, 2009:1-13

Faria, R., Navarro, A. (2010) Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends in Ecology & Evolution, 25(11): 660-669

Feist, S.M. (2013) Hellbender (*Cryptobranchus alleganiensis*) gene flow within rivers of the Missouri ozark highlands. Master's thesis, University of Missouri-Columbia, May 2013.

Fernandes-Matioli, F.M.C., Almeida-Toledo, L.F. (2001) A molecular phylogenetic analysis in *Gymnotus* species (pisces: Gymnotiformes) with inferences on chromosome evolution. Caryologia, 54(1):23-30

Fisher, R.A. (1922) On the dominance ratio. Proceedings of the Royal Society of Edinburgh, 42:321-341

Fisher, R.A. (1937) The wave of advance of advantageous genes. Annals of Eugenics, 7:355-369

Fitzpatrick, B.M., Johnson, J.R., Kump, D.K., Shaffer, H.B., Smith, J.J., Voss, S.R. (2009) Rapid fixation of non-native alleles revealed by genome-wide snp analysis of hybrid tiger salamanders. BMC Evolutionary Biology, 9(1):176

Frean, M., Rainey, P.B., Traulsen, A. (2013) The effect of population structure on the rate of evolution. Proceedings of the Royal Society of London B: Biological Sciences, 280(1762):20130211

Galloway, L.F., Etterson, J.R. (2005) Population differentiation and hybrid success in *Campanula americana*: geography and genome size. Journal of Evolutionary Biology, 18:81-89

Galloway, L.F. (2001) The effect of maternal and paternal environments on seed characters in the herbaceous plant *Campanula americana* (Campanulaceae). American Journal of Botany, 88(5):832-840

Gharrett, A.J., Smoker, W.W. (1991) Two generations of hybrids between even- and odd-year pink salmon (Oncorhynchus gorbuscha): a test for outbreeding depression? Canadian Journal of Fisheries and Aquatic Sciences, 48 (King 1955):1744-1749

Gillespie, R. (2004) Community assembly through adaptive radiation in Hawaiian spiders. Science, 303(5656):356-359

Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., Wood, K.R., Sytsma, K.J. (2009) Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). Proceedings of the Royal Society of London B: Biological Sciences, 276(1656): 407-416

Groves, C. (2001) Faunal and floral migrations and evolution in SE Asia-Australia, chapter Mammals in Sulawesi: where did they come from and when, and what happened to them when they got there, pages 333-342. CRC Press

Haldane, J.B.S. (1927) A mathematical theory of natural and artificial selection. part v. selection and mutation. Proc Camb Philol Soc, 23:838-844

Hardy, G.H. (1908) Mendelian proportions in a mixed population. Science, 28:49-50

Harewood, L., Schütz, F., Boyle, S., Perry, P., Delorenzi, M., Bickmore, W.A., Reymond, A. (2010) The effect of translocation-induced nuclear reorganization on gene expression. Genome Research, 20(5):554-564

Hedrick, P.W. (1981) The establishment of chromosomal variants. Evolution, 35(2):322-332

Hedrick, P.W., Levin, D.A. (1984) Kin-founding and the fixation of chromosomal variants. The American Naturalist, 124(6):789-797

Helenurm, K., Ganders, F.R. (1985) Adaptive radiation and genetic differentiation in Hawaiian bidens. Evolution, pages 753-765

Hills, M.J., Hall, L., Arnison, P.G., Good, A.G. (2007) Genetic use restriction technologies (GURTs): strategies to impede trans-gene movement. Trends in Plant Science, 12(4):177-183

Hindersin, L., Traulsen, A. (2014) Counterintuitive properties of the fixation time in network-structured populations. Journal of The Royal Society Interface, 11(99):20140606

Hindersin, L. Traulsen, A. (2015) Almost all random graphs are amplifiers of selection for birth-death dynamics, but suppressors of selection for death-birth dynamics. arXiv preprint arXiv:1504.03832

Hoelzer, G.A., Drewes, R., Meier, J., Doursat, R. (2008) Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. PLoS Comput Biol, 4(7):e1000126-e1000126

Jacobs, P.A. (1981) Mutation rates of structural chromosome rearrangements in man. American Journal of Human Genetics, 33:44-54

Jamieson-Lane, A., Hauert, C. (2015) Fixation probabilities on superstars, revisited and revised. Journal of Theoretical Biology, 382: 44-56

Karlin, S., McGregor, J. (1972a) Application of method of small parameters to multi-niche population genetic models. Theoretical Population Biology, 3:186-209

Karlin, S., McGregor, J. (1972b) Polymorphisms for Genetic and Ecological Systems with Weak Coupling. Theoretical Population Biology, 3: 210-238

Keitt, T.H., Lewis, M.A., Holt, R.D. (2001) Allee effects, invasion pinning, and species' borders. The American Naturalist, 157(2):203-216

Kondrashov, A.S. (2003) Accumulation of Dobzhansky-Muller incompatibilities within a spatially structured population. Evolution, 57(1):151-153

Kwit, C., Moon, H.S., Warwick, S.I., Stewart, C.N. (2011) Transgene introgression in crop relatives: molecular evidence and mitigation strategies. Trends in Biotechnology, 29(6):284-293

Lande, R. (1984) The expected fixation rate of chromosomal inversions. Evolution, 38(4):743-752

Lande, R. (1985) The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. Heredity, 54:323-332

Landguth, E.L., Johnson, N.A., Cushman, S.A. (2015) Clusters of incompatible genotypes evolve with limited dispersal. Frontiers in Genetics, 6, 2015.

Lee, D., Natesan, E. (2006) Evaluating genetic containment strategies for transgenic plants. Trends in Biotechnology, 24(3):109-114

Lerner, H.R.L., Meyer, M., James, H.F., Hofreiter, M., Fleischer, R.C. (2011) Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Current Biology, 21(21):1838-1844

Lessios, H.A., Kane, J., Robertson, D.R., Wallis, G. (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: contrasting patterns of population structure between oceans. Evolution, 57(9):2026-2036

Levin, D.A., Wilson, A.C. (1976) Rates of evolution in seed plants: Net increase in diversity of chromosome numbers and species numbers through time. Proceedings of the National Academy of sciences, 73(6):2086-2090

Limborg, M.T., Waples, R.K., Seeb, J.E., Seeb, L.W. (2014) Temporally isolated lineages of pink salmon reveal unique signatures of selection on distinct pools of standing genetic variation. Heredity, 105(6):1-11

Lu, B-R., Yang, C. (2009) Gene flow from genetically modified rice to its wild relatives: Assessing potential ecological consequences. Biotechnology Advances, 27(6):1083-1091

Magnacca, K.N., Price, D.K. (2015) Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing Drosophila (diptera: Drosophilidae). Molecular Phylogenetics and Evolution

May, R.M., Endler, J.A., McMurtie, R.E. (1975) Gene frequency clines in the presence of selection opposed by gene flow. American Naturalist, 109: 659-676

Nachman, M.W., Myers, P. (1989) Exceptional chromosomal mutations in a rodent population are not strongly underdominant. PNAS, 86(17): 6666-6670

Newman, L.J. (1977) Chromosomal evolution of the Hawaiian Telmatogeton (Chironomidae, Diptera). Chromosoma, 64(4):349-369

Nolte, A.W., Gompert, Z., Buerkle, C.A. (2009) Variable patterns of introgression in two sculpin hybrid zones suggests that genomic isolation differs among populations. Molecular Ecology, 18:2615-2627

Palumbi, S.R. (1992) Marine speciation on a small planet. Trends in Ecology & Evolution, 7(4):114-118

Payne, J.L., Eppstein, M.J., Goodnight, C.C. (2007) Sensitivity of self-organized speciation to long-distance dispersal. Proceedings of the 2007 IEEE Symposium on Artificial Life, CI-ALife 2007, pages 1-7, 2007. doi: 10.1109/ALIFE.2007.367651.

Piálek, J., Barton, N.H. (1997) The spread of an advantageous allele across a barrier: The effects of random drift and selection against heterozygotes. Genetics, 145:493-504

Piálek, J., Hauffe, H.C., Rodríguez-Clark, K.M., Searle, J.B. (2001) Raciation and speciation in house mice from the alps: the role of chromosomes. Molecular Ecology, 10(3):613-625

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL http://www.R-project.org

Reeves, R.G., Reed, F.A. (2014) Stable transformation of a population and a method of biocontainment using haploinsufficiency and underdominance principles. WO Patent App. PCT/EP2013/077856, WO2014096428 A1.

Reeves, R.G., Bryk, J., Altrock, P.M., Denton, J.A., Reed, F.A. (2014) First steps towards underdominant genetic transformation of insect populations. PLoS One, 9(5):e97557

Schierup, M.H., Christiansen, F.B. (1996) Inbreeding depression and outbreeding depression in plants. Heredity, 77(July 1995):461- 468

Sinkins, S.P., Gould, F. (2006) Gene drive systems for insect disease vectors. Nature Reviews Genetics, 7(6):427-435

Skyrms, B. (2001) The stag hunt. Proceedings and Addresses of the American Philosophical Association. 75. 10.2307/3218711.

Soboleva, T.K., Shorten, P.R., Pleasants, A.B., Rae, A.L. (2003) Qualitative theory of the spread of a new gene into a resident population. Ecological Modelling, 163:33-44

Spirito, F. (1993) The exact values of the probability of fixation of underdominant chromosomal rearrangements. Theoretical Population Biology, 41:111-120

Spirito, F., Rossi, C., Rizzoni, M. (1991) Populational interactions among underdominant chromosome rearrangements help them to persist in small demes. Journal of Evolutionary Biology, 3:501-512

Stewart, A.J., Seymour, R.M., Pomiankowski, A., Reuter, M. (2013) Under-Dominance constrains the evolution of negative autoregulation in diploids. PLoS Computational Biology, 9(3):e1002992

Strogatz, S.H. (2001) Exploring complex networks. Nature, 410(March):268-276

Lumley, T, using Fortran code by A. Miller. (2009) leaps: regression subset selection. URL https://cran.r-project.org/web/packages/leaps/. R package version 2.9.

Traulsen, A., Reed, F.A. (2012) From genes to games: Cooperation and cyclic dominance in meiotic drive. Journal of Theoretical Biology, 299: 120-125

van Riper III, C., van Riper, S.G., Goff, M.L., Laird, M. (1986) The epizootiology and ecological significance of malaria in Hawaiian land birds. Ecological Monographs, 56(4):327-344

Walsh, J.B. (1982) Rate of accumulation of reproductive isolation by chromosome rearrangements. The American Naturalist, 120(4):510-532

White, M.J.D. (1978) Modes of Speciation. W. H. Freeman, San Francisco.

Wilson, A.C., Sarich, V.M., Maxson, L.R. (1974) The importance of gene rearrangement in evolution: evidence from studies on rates of chromosomal, protein, and anatomical evolution. PNAS, 71(8):3028-3030

Wilson, D.S., Turelli, M. (1986) Stable underdominance and the evolutionary invasion of empty niches. The American Naturalist, 127 (6):835-850

Wright, S. (1931) Evolution in Mendelian populations. Genetics 16(2):97-159

Wright, S. (1941) On the probability of fixation of reciprocal translocations. Am Nat 75(761):513-522

Xie, F., Levinson, D. (2007) Measuring the structure of road networks. Geographical Analysis, 39:336-356

Yoon, J.S., Resch, K., Wheeler, M.R., Richardson, R.H. (1975) Evolution in Hawaiian Drosophilidae: Chromosomal phylogeny of the *Drosophila crassifemur* complex. Evolution 29:249-256

Chapter 3

Gene expression across tissues, sex, and life stages in the sea urchin *Tripneustes gratilla*

3.1 Abstract

The pan-tropical sea urchin *Tripneustes gratilla* is an ecologically and economically important shallow water algal grazer. The aquaculture of *T. gratilla* has spurred growing interest in the population biology of the species, and by extension the generation of more molecular resources. To this purpose, *de novo* transcriptomes of *T. gratilla* were generated for two adults, a male and a female, as well as for a cohort of approximately 1,000 plutei larvae. Gene expression profiles of three adult tissue samples were quantified and compared. These samples were of gonadal tissue, the neural ring, and pooled tube feet and pedicellariae. Levels of shared and different gene expression between sexes, as well as across functional categories of interest, including the immune system, toxins, genes involved in fertilization, and sensory genes are highlighted. Differences in expression of Sex determining Region Y-related High Mobility Group box groups and general isoform expression between the sexes is observed. Additionally an expansion of the tumor suppressor DMBT1 was observed in *T. gratilla* when compared to the annotated genome of

the sea urchin *Strongylocentrotus purpuratus*. The draft transcriptome of *T. gratilla* is presented here in order to facilitate more genomic level analysis of de-novo sea urchin systems.

3.2 Introduction

The Phylum Echinodermata occupies a unique place on the evolutionary tree of life. Together the Echinodermata and Hemichordates, or acorn worms, form the clade Ambulacraria, which are the most closely related invertebrate taxa to the Chordates (Metschnikoff 1881, Furlong & Holland 2002, Satoh et al. 2014). Sea urchins (Class Echinoidea) have served as a model organism in developmental biology for over 150 years, owing to their frequent and voluminous broadcast spawning behavior and the ease with which gametes are collected and observed for in vitro fertilization (McClay 2011). The importance of sea urchins to the field of developmental biology resulted in an official call by Davidson & Cameron (2002) to sequence a complete sea urchin genome. In 2006 the full genome of the East Pacific urchin *Strongylocentrotus purpuratus* (Stimpson 1857) was published, and the annotated genome made freely available online (Sea Urchin Genome Sequencing Consortium 2006, Cameron et al. 2009). This development inspired significant and ongoing molecular work on members of the Strongylocentrotidae family (Kober & Bernardi 2013, Oliver et al. 2010, Walters et al. 2008), and more recently, a handful of transcriptomes and representative genomes for Echinoderms beyond the Strongylocentrotidae have been published (Dilly et al. 2015, Israel et al. 2016). While there is routine use of sea urchins as laboratory organisms and a growing body of molecular data available, many aspects of sea urchin biology still remain unknown. For example, while there have been hypotheses generated about the sex determination of sea urchins, it has yet to be concluded whether sea urchins operate under genetic sex determination or even possess distinct sex chromosomes (Eno et al. 2009; Bachtrog et al. 2014). Many species of sea urchin are also highly commercially valued in the

food industry. Notably *Tripneustes gratilla* (Linnaeus 1758), commonly known as the collector urchin, and the Pacific congeneric of the Caribbean "sea egg" *T. ventricosus* (Lamark 1816). Tripneustes gratilla has one of the most expansive ranges of all shallow water echinoids, occurring everywhere from the Hawaiian islands in the central Pacific, to the shores of South Africa (Mayr 1954, but see Bronstein et al. 2017). As an ecosystem engineer, *T. gratilla* greatly affects the shallow water community composition during rapid population expansions and mass die-offs that frequently occur subsequent to rapid expansions (Valentine & Edgar 2010). As such, *T. gratilla* is currently the object of a large-scale aquaculture and outplanting effort by the state of Hawai?i as a biocontrol agent against invasive algae such as Acanthophoraspicifera, Gracilaria salicornia, Eucheuma denticulatum and Kappaphycus clade B (Westbrook et al. 2014). A member of the family Toxopneustidae (Troschel 1872), a sister family to the Strongylocentrotidae, (Láruson 2017), *T. gratilla* is a confamilial of the highly venomous sea urchin, *Toxopneustes pileolus* (Lamark 1816). The increased collection and cultivation of T. gratilla has spawned increased interest in understanding the population structure of this broadly distributed urchin (Cyrus et al. 2014, Westbrook et al. 2015). This paper presents comparative descriptions of annotated draft transcriptomes from a male and female adult as well as a cohort of larval *T. gratilla* to determine how gene expression differs in both somatic and gonadal tissues between the sexes, and how gene expression profiles vary at different life stages (planktotrophic larvae versus benthic adults). A focus is placed on key similarities in presence or absence of expressed genes between these broadly split life stages. As suggested by Tu et al. 2012, the combination of larval developmental stage and adult tissue transcriptome can be integral for a more accurate context of genome level sequences. We present this draft transcriptome in response to the need for more large scale molecular resources for this ecologically and economically important sea urchin.

3.3 Materials & Methods

3.3.1 Collection and Extraction

Whole RNA was extracted using a Qiagen RNeasy extraction kit from two adult *T. gratilla*, collected from near-shore waters of southwest O'ahu, Hawai'i (approximately 21°21'13"N, 158°7'54"W). Three distinct tissues were individually sampled and indexed from each adult: gonadal tissue was sampled from each of the five gonadal lobes; neural tissue was sampled from the neural ring; tube feet and pedicellariae were sampled from several places across the animals' external surface. Whole RNA was similarly extracted from approximately 1,000 plutei-stage T. gratilla larvae, acquired from the Ānuenue Fisheries Research Center on Sand Island, O'ahu, Hawai'i (Hawaii Department of Land and Natural Resources). The larval cohort consisted of mixed progeny from five adult females and five adult males. These parental *T. gratilla* were mature (>65mm in diameter) individuals wild caught off the coast of western O'ahu, Hawai'i.

3.3.2 Sequencing

Extracted RNA was sent to the Hawai'i Institute of Marine Biology Genetics Core for cDNA library construction and sequencing. Extracts were treated with Epicentre Rnase-free Dnase I; 1 uL of Dnase I was added per 20uL of sample, then incubated at 37°C for 30 min. Samples were then cleaned using the Qiagen RNeasy Minelute Cleanup Kit. Quality was assessed on an Agilent 2100 Bioanalyzer. Poly-A tail isolation and cDNA synthesis were performed with the Illumina TruSeq Stranded mRNA Sample Preparation Kit (Protocol Part # 15031047 Rev. E) with no fragmentation time to allow for larger RNA fragments. cDNAs were sequenced on two lanes of an Illumina MiSeq with single strand chemistry for 300 cycles, using a V2 reagent kit. Sequence files for each sample are available on the NCBI SRA database (Female Gonad:

SRR6844874, Female Neural: SRR6844873, Female Tubefeet: SRR6844872, Male Gonad: SRR6844871, Male Neural: SRR6844877, Male Tubefeet: SRR6844876, Larvae: SRR6844875).

3.3.3 Sequence Assembly, Annotation, and Analysis

All sequence reads were filtered and trimmed using Trimmomatic (Bolger et al. 2014): matches to TruSeq3 index adapters were purged, leading and trailing bases falling under a Phred33 quality score of 3 were trimmed, a sliding window trim was performed with a window size of four bases and an average window quality score threshold of 15, and a minimum read threshold of 36 was set. Cleaned reads from each sample were then assembled individually with Trinity (Haas et al. 2013). In order to avoid an excess of falsely identified isoform variants resultant assembly was clustered at a 99% identity threshold with CD-HIT (Fu et al. 2012, Li & Godzik 2006). Assembly quality was partly assessed via recovery of conserved metazoan single copy orthologs, as defined through the software BUSCO (Simão et al. 2015). Reciprocal best matches to the annotated genomes of *S. purpuratus* (Spur4.2) and *Lytechinus variegatus* (Lvar2.2), acquired from www.echinobase.org, were performed with a local tBLASTx alignment with an e-value cut off of 1e-10. Additional annotation was performed by mapping the assembly, as well as the longest open reading frames from Transdecoder (Haas et al. 2013), to the SwissProt (Bairoch & Apweiler 2000) and Pfam (Finn et al. 2016) databases. The resulting hits were assembled into a database with Trinnotate (https://trinotate.github.io/), with Gene Ontology (GO) categories assigned to said hits. Transcript quantification, in terms of Transcripts Per kilobase Million (TPM), was accomplished with RSEM (Li & Dewey 2011). Manhattan distance values, summary statistics, and visualization was generated in R (R Core Team 2017).

3.4 Results

In total, 14,903,815 raw sequence 300bp reads were generated. Trinity assembly of sequence data clustered at 99% identity from all seven samples resulted in 172,841 transcripts, with 157,610 flagged as unigenes. The final assembly had an average contig length of 721.09, an N50 of 917, and a GC content of 37.04%. Assembly quality was partly assessed via the recovery of 86.6% of highly conserved metazoan single copy orthologs. For comparison, the full *S. purpuratus* genome recovered 90.3% of the same dataset of metazoan orthologs. Only 10,027 unigenes had proposed Trinity identified splice variants, or isoforms. Some 11,689 unigenes were annotated via reciprocal best matches to *S. purpuratus*, while an additional 3,519 unigenes were annotated via matches to *L. variegatus* for a total of 15,208 annotated unigenes. Of the remaining 142,402 unannotated unigenes, 6,452 matches to the SwissProt and Pfam databases for a total of 21,660 annotated transcripts. This corresponds to just over 70% (21,660/29,948) of the total number of annotated genes of the *S. purpuratus* genome.

Between all annotated genes over 56% were observed exclusively in the adult tissues, and only 2% (422 genes) were uniquely observed in the larval transcriptome (Figure 3.1A). 86.8% of all annotated genes were observed in both individuals, with 6.9% of annotated genes being unique to the female, and 6.3% being unique to the male (Figure 3.1B). 58% of all annotated genes were found to be co-occurring across the three adult tissues (gonad, neural, and tube feet/pedicellariae). The largest number of tissue-specific expression was seen in the gonads (7.7%), while the fewest uniquely expressed genes were identified in the tube feet/pedicellariae (1.3%, Figure 3.1C).

Across broadly identified functional categories, after excluding ribosomal genes as well as genes with unknown annotations, by far the greatest number of annotated genes were identified as being involved in metabolism (1,038). The 20 most highly represented functional gene categories

are highlighted in Figure 3.1D.

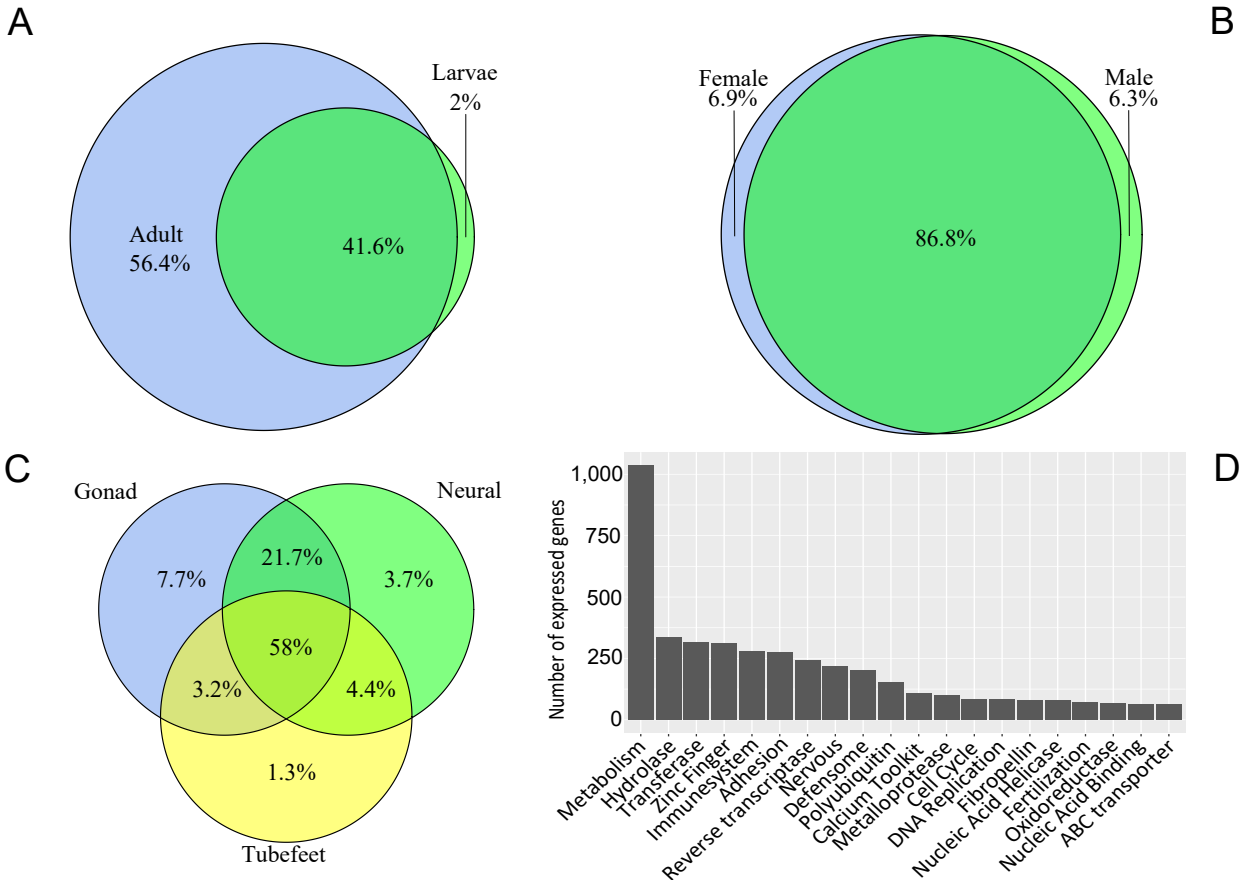

FIGURE 3.1: Annotated genes across samples. A. Genes identified across adult tissues versus larval sample. B. Genes identified across all tissues in a female versus in a male. C. Genes identified across different tissues. D. Number of identified genes expressed across functional categories identified via annotation to Strongylocentrotus purpuratus, Lytechinus variegatus, as well as Pfam and Swissprot datasets. Only the top 20 largest categories are displayed

3.4.1 Fertilization and Sex differences

A Manhattan distance calculation of expression profiles between adult tissues clusters each adult separately, with the neural and tube feet/pedicellariae tissues clustering more closely to each other than to the gonadal tissue. However when isoforms are clustered so that only unigene expression information is considered, the variation between individual gonad expression profiles is reduced such that the male and female gonad samples cluster together (3.2).

Ninety genes were identified as being involved in fertilization, with 42 of those expressed in both the adult samples as well as the larvae. Two genes, acid-sensing ion channel-like (AsicL) and sodium channel (NaC) were only observed in the larvae. The dual oxidase homolog Udx1, which has an important role in the fast-block to polyspermy (Wong et al. 2004), was one of the four most highly expressed fertilization genes across all samples, except for in the male gonad. Comparing male and female samples across all tissues recovered 1,335 genes that were exclusively found in the male and not the female, and 1,467 genes were only seen in the female and not the male. The only sex determination associated protein recovered in one sex was Wnt-4, which was found in the male and the larvae. Eleven Sex determining Region Y related High Mobility Group box (Sox) gene fragments were recovered. These represented the SoxB1, SoxB2, SoxC, SoxD, SoxE, SoxF, and SoxH groups; these groups were seen across all adult tissues, while SoxD was the only group not recovered from the larvae (Figure 3.3).

3.4.2 Immune system

Immune system genes were identified via known annotated functional category (Hibino et al. 2006) or GO categorization. Comparing only the most highly expressed immune genes, Ubiquitin-40S ribosomal protein S27a (RPS27A) fusion protein was the most highly expressed in all female tissue samples, the larvae sample, as well as the male neural and tube

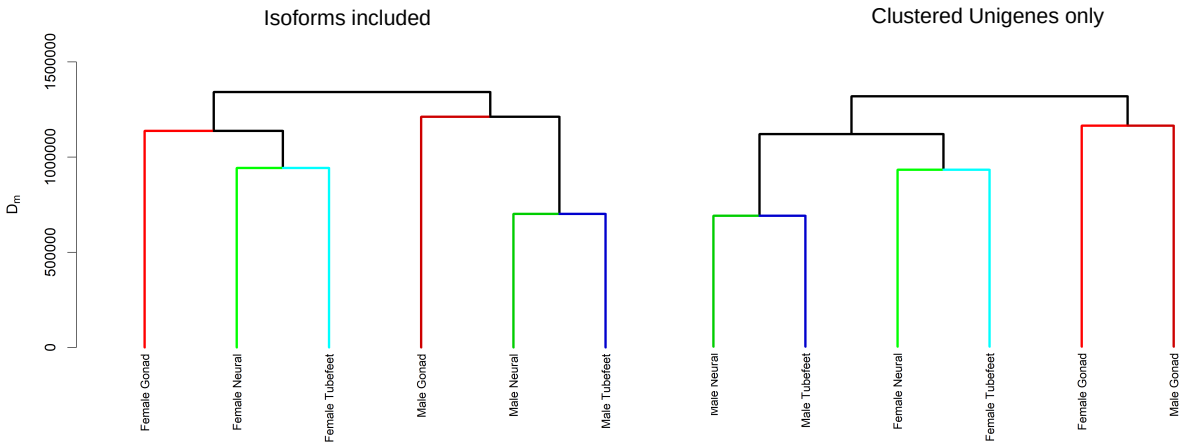

FIGURE 3.2: Dendrogram of Manhattan distances observed between expression profiles in adult tissues. On the left including all isoforms identified in each sample, and on the right considering only the expression of unigenes (clustered isoforms). The placement of Female Gonad and Male Gonad profiles as a distinct unit when considering only the unigene expression suggest that splice variation underlies the differences in male and female gonadal expression

feet/pedicellariae samples. In the male gonad sample, however, a Scavenger Receptor Cysteine Rich (SRCR) gene (Srcr197) was the most highly expressed immune gene observed. SRCR domains were seen expressed in all samples, with 33 unique SRCR genes identified in total. Two genes involved in the clotting response, Amassin (Hillier & Vacquier 2003), and a Kringle domain containing gene (PlgL2), were observed to be two of most highly expressed genes in the neural tissue of both the female and male samples. Additionally, PlgL2 was observed to be present at elevated levels in the tubefeet/pedicellariae from both adult samples. In total, nine sea urchin TLR genes were identified, along with six Tlr-like genes. Tlr003 was identified through the SwissProt database, but failed as a reciprocal best match to that of either the *S. purpuratus* or *L. variegatus* genomes.

Twenty-nine NACHT domain and leucine-rich repeat (NLR) proteins were identified. Overall 15

elements annotated to the immunoglobulin superfamily (IgSF) and 12 genes connected to V(D)J recombination were identified, as well as 18 tumor necrosis factor (TNF) functional category genes. Two RAS domains and two Fat-like cadherin-related tumor suppressor homologs were also identified. An unexpected finding was the identification of 69 Deleted in Malignant Brain Tumor 1 (DMBT1) unigenes. The majority of expressed DMBT1 genes were observed only in the adult tissues, but eight were also identified in the larvae. Results are summarized in Table 3.1.

TABLE 3.1: Identified unigenes from select functional gene classes implicated in immune system response

Gene class	Number of unigenes identified
Deleted in malignant brain tumors 1 (DMBT1)	69
Scavenger receptor cysteine-rich repeat (SRCR)	33
NACHT domain and leucine-rich repeat (NLR)	29
Tumor necrosis factor (TNF)	18
Toll-like receptor (TLR)	15
Immunoglobin superfamily (IgSF)	15

3.4.3 Toxins

Six gene fragments annotated as putative toxins were identified through the Swissprot/Pfam database. Venom Protein 302 was identified in all adult tissues, absent in the larvae, but expressed most highly in the tube feet and pedicellariae samples of both adults. Snake venom metalloprotease inhibitor 02A10 was identified in the male gonad and Venom serine protease Bi-VSP was found in all adult tissues except for the male gonad. C-type lectin lectoxin-Enh5 was found in all female tissues, the larvae, and the male neural tissue. Fourteen transcripts with Phospholipase A2 (PLA2) activity that were most highly represented in the tubefeet/pedicellariae were identifed. Fragments annotated as Stonustoxin alpha subunit and Verrucotoxin beta subunit, both toxic proteins characterized from Stonefish (genus *Synanceia*),

were expressed in the female tube feet/pedicellariae sample (Table 3.2). Alpha-latrotoxin-Lhe1a was also found in the female tube feet/pedicellariae. This fragment contained ankyrin repeats typically seen in Alpha-latrotoxin-Lhe1, but was missing the N-terminal domain associated with the neurotoxin from the western black widow spider *Latrodectus hesperus* Chamberlin & Ivie, 1934.

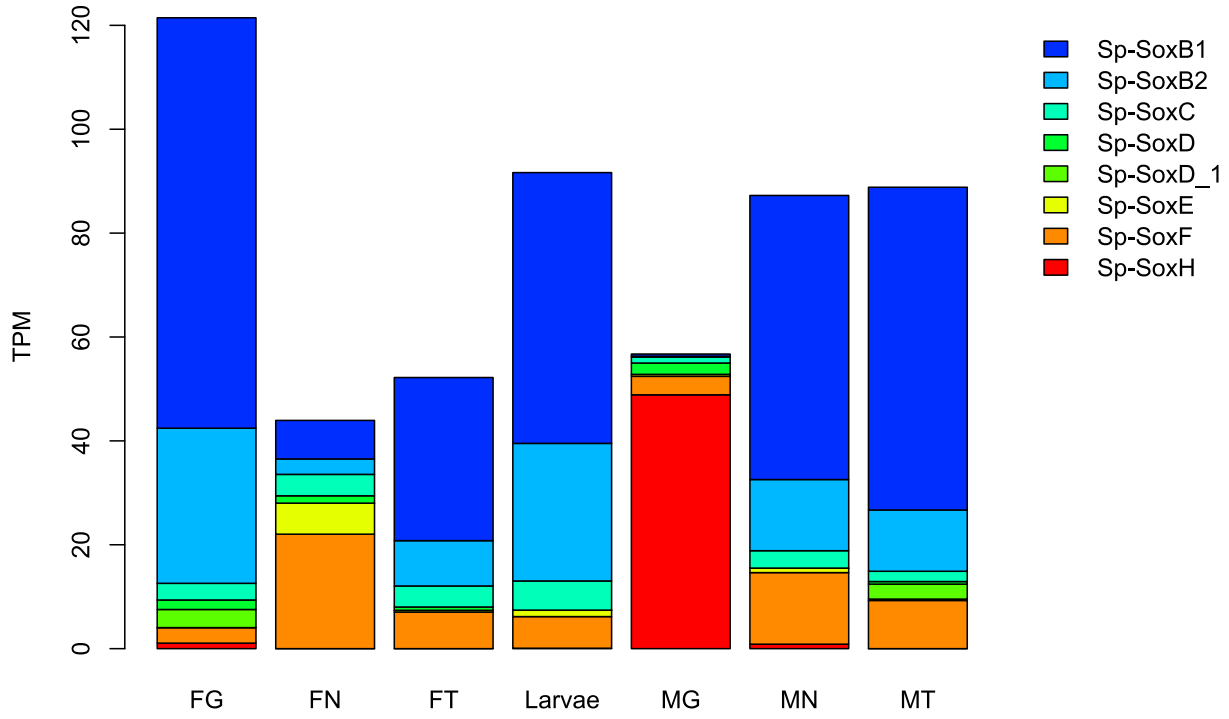

FIGURE 3.3: Barplot of Transcripts Per kilobase Million levels of Sex determining Region Y-related High Mobility Group box (SRY-HMG or SOX) group expression across the different samples. Elevated SoxH expression was observed in the Male Gonad compared to the remaining samples, while higher levels of SoxB1, SoxB2, and SoxC were observed in the female and the larvae.
F = Female, M=Male, G=Gonad, N=Neural, T=Tubefeet/Pedicellariae

3.4.4 Sensory

Only four of the eight prototypical sea urchin opsin genes (Raible et al. 2006, D'Aniello et al. 2015, Lowe et al. 2017) were recovered: opsin 2, opsin 5, opsin 3.2, and opsin 4-like-1. All opsins were only found expressed in adult tissues. The most highly expressed sensory protein identified in the larvae was mechanosensory protein 2, which was also the most highly expressed sensory protein in the male and female tube feet/pedicellariae as well as the male neural tissue. Six distinct Sensory GPCR Rhodopsins and a FAD binding blue light photoreceptor sequence were recovered from adult as well as larval sequences.

TABLE 3.2: Venom/Toxin associated gene fragments identified. Venom Protein 302 was the only transcript observed at its highest expression in the tube feet/pedicellaria of both individuals. All values reported in Transcripts Per kilobase Million

Transcripts	Annotation	F. Gonad	F. Neural	F. Tubefeet	Larvae	M. Gonad	M. Neural	M. Tubefeet
TR37298_c0_g1	Venom protein 302	0.67	5.2	15.27	0	0	7.54	10
TR47326_c0_g2	Stonustoxin subunit alpha	0	0	18.3	0	2.07	0	0
TR47326_c0_g3	Stonustoxin subunit alpha	0.52	0	6.61	0	0	0	0
TR47326_c1_g1	Verrucotoxin subunit beta	0	0	10.92	0	0	0	0
TR9549_c0_g1	Alpha-latrotoxin-Lhe1a	0	0	7.02	0	0	0	0
TR80588_c0_g1	Venom serine protease Bi-VSP	6.68	4.68	4.22	0	0	10.93	5.5
TR48480_c0_g1	C-type lectin lectoxin-Enh5	2.32	0.9	0.81	2.21	0	2.81	0
TR8624_c1_g2	Snake venom metalloprotease inhibitor 02A10	0	0	0	0	31.3	0	0

3.5 Discussion

Some 21,660 unigenes identified from a larval cohort, as well as three tissues of an adult male and female, are presented here from *Tripneustes gratilla*. This represents over 70% (21,660/29,948) of the total number of annotated genes of the *S. purpuratus* genome, and includes 86.6% of highly conserved metazoan single copy orthologs (Simão et al. 2015). This represents a significant portion of the *T. gratilla* transcriptome.

Since the sex determination mechanism of sea urchins remains unclear, expression profile differences between three individually extracted tissue samples from a male and female *Tripneustes gratilla* provides some insight into potentially contributing molecular factors. The distinct gene expression profiles of the male and female gonads suggested by the Manhattan distance on clustered unigenes, disappears when measuring differences in expression profiles considering isoforms (Figure 3.2). While differences were observed between isoform expression profiles in male and female tissues, no isoforms were found to exclusively occur in one sex. This suggests that similar genes are expressed in both the male and female gonads, and levels of expression of splice variants are primarily contributing to the functional distinction of testis and ovary. While the major sex determination protein SoxA (Sex determining Region Y) is absent in the presented *T. gratilla* dataset, as well as the *S. purpuratus* and *L. variegatus* transcriptomes (Cameron et al. 2009), seven of the remaining eight Sox gene groups are represented in this dataset (SoxB1, SoxB2, SoxC, SoxD, SoxE, SoxF, and SoxH). Somewhat curiously, the expression patterns of Sox gene groups in the larvae much more closely resembled that of the female than that of the male tissues (Figure 3.3). Out of all Sox gene expressions, SoxH was found to be most highly expressed in the male compared to the female, with 48.87 TPM observed in the male gonad, compared to 1.06 TPM in the female gonad. A homolog to the human Sox30, SoxH has been suggested to be involved in the differentiation of male germ cells (Osaki et al.1999). Wnt-4, generally understood in mammals to be involved in the suppression of masculinization in early development (Chassot et al. 2012), was recovered in the larvae as well as in adult male neural tissue. It was however not found to in the adult female. It is therefore possible the functional distinction of Wnt-4 may have the opposite effect, promotion of masculinization, in sea urchins.

The small relative number of uniquely larval expressed genes could be a result of the expression profile of several adult tissues being compared to a single developmental stage larvae; alternatively it could possible be an effect of sequencing effort: while there was effectively equal

sequencing effort in this study for the larval and the adult tissue sequencing (one illumina MiSeq lane each), a comparison to *S. purpuratus* datasets with seven fold greater sequencing effort for the larvae (2 lanes for adult, 14 for larvae, Pespeni et al. 2013a, Pespeni et al. 2013b) showed virtually no uniquely adult expressed genes, and over 32% of annotated genes being observed only in the larvae (supplementary figure 1).

This annotated draft transcriptome of *T. gratilla* captured some key immune system genes. Interestingly, only 15 sea urchin TLR and TLR-like genes were identified across all samples. This is a low number compared to the 222 TLR genes identified in *S. purpuratus* (Sea Urchin Genome Sequencing Consortium 2006), and more in line with the 10 genes observed in *Homo sapiens* Linnaeus, 1758 or the 28 *S. purpuratus* TLR transcripts recovered by Tu et al. 2012. Similarly, 33 unique SRCR domains identified in total, across all samples, which is a greater number of SRCR genes than the 16 genes in *H. sapiens*, but still unlikely to represent the full suite of SRCR diversity, as 218 SRCR genes have been identified in *S. purpuratus* (Sea Urchin Genome Sequencing Consortium 2006). However, an expansion was observed in a protein member of the SRCR superfamily: deleted in malignant brain tumor 1 (DMBT1), which was represented by 69 unigenes. Only 1 DMBT1 and 11 DMBT1-like genes have been identified in the *S. purpuratus* genome. DMBT1 are known to be involved in regulating both the immune system and tumor cells (via apoptosis or immune response to tumor cells). Tumor suppressing proteins such as Tumor Necrosis Factor (TNF) and Rassf5 were also observed in all adult tissues as well the larvae. This enhanced diversification of immune and tumor suppressing genes may be involved in the suppression of transmissible cancer cells in a marine environment (Metzger et al. 2016). Furthermore, this could suggest alternative immune response strategies between *S. pupuratus* and *T. gratilla*: amplification of TLR genes (but see also Tu et al. 2012) versus DMBT1 duplication. Additionally, the role of immune system genes in sea urchin tissue regeneration is worthy of future study (Ramírez-Gómez et al. 2008).

A number of potentially toxic proteins were identified in this dataset. Being a member of family Toxopneustidae, *T. gratilla* is closely related to the most venomous sea urchin *Toxopneustes pileolus*. Venom isolated from the globiferous pedicellaria of *T. gratilla* has been shown to be lethal to mice and rabbit subjects, with respiratory distress and terminal tonic convulsions reported prior to death (Alender 1963). *Tripneustes gratilla* is not considered hazardous to humans, however irritation of the skin after handling adult individuals has been observed by the authors. Also highly expressed in the tubefeet and pedicellariae were fourteen PLA2 activity genes. PLA2 share structural similarity with the *T. pileolus* Contractin A venom (Hatakeyama et al. 2015). Venom Protein 302, an insulin-like growth factor binding protein, originally identified in the venom gland of the Chinese Swimming Scorpion (*Lychas mucronatus* (Fabricus 1798) was observed in all tissue samples, aside from the male gonad and the larvae. It is the only annotated toxin to be present at its highest expression value in the tube feet/pedicellariae of both adult individuals.

Verrucotoxin is a dimeric lethal toxin identified in the Stonefish *Synanceia verrucosa* (Garnier et al. 1995). Verrucotoxin subunit B annotated fragments were observed exclusively expressed in the tube feet/pedicellaria; however only in the female adult. Two transcript variants of another Stonefish toxin, Stonustoxin subunit B (Poh et al. 1992), were also observed in the female adult's tube feet/pedicellariae. While Stonustoxin has previously been considered a novel marine toxin (Ghadessy et al. 1996), more recent analysis suggests that it may in fact have more broad functional ancestry, and as is true for many toxins need not necessarily have a toxic function across broader taxonomic identification (Ellisdon et al. 2015). The expression location of these supposed toxins in the tube feet/pedicellariae could suggest the underlying function of toxic defense against predation, the globiferous pedicellaria of *T. gratilla* are defensive structures (see review in Coppard et al. 2012) and are potentially pursuit-deterrents to predatory fish (Sheppard-Brennand, et al. 2017). Future efforts will be needed to fully isolate and confirm the

effects of these potential toxins.

Lastly it should be mentioned that a large number of reverse transcriptase genes (242) were identified in the dataset (Figure 3.1D). Sea Urchin Retroviral-Like (SURL, Mag family of Ty3/Gypsy retrotransposons) mRNA recovered suggesting that this TE is still quite active in *T. gratilla*. (Springer et al.1991, Gonzales and Lessios 1999)

The contribution of this *T. gratilla* transcriptome, with 21,660 identified unigene sequences, will allow for further research into comparative genetic characterization of echinoids, as well as molecular diversity and population structure of this broadly distributed, economically, and environmentally important species.

3.6 Data repository

All raw sequence reads are to be deposited to the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/SRP135810).

3.7 Acknowledgements

This research was funded and made possible by the Jessie D. Kay Memorial Fellowship; the Elizabeth A. Kay Endowed Award; the Charles H. & Margaret B. Edmondson Research Fund; the Watson T. Yoshimoto Fellowship, and the Hampton & Meredith Carson Fellowship, both administered by the Ecology, Evolution, and Conservation Biology specialization program at the University of Hawai'i at Mānoa. Thanks to Victoria Sindorf, for helpful comments and suggestions. Special thanks to David Cohen and staff of the Ānuenue Fisheries Research Center (DLNR), and to Amy Eggers, of the Hawai'i Institute of Marine Biology Genetics Core.

3.8 References

Alender, C.B. (1963) The venom from the heads of the globiferous pedicellariae of the sea urchin, *Tripneustes gratilla* (Linnaeus) [dissertation]. Honolulu, HI. University of Hawai'i at Mānoa. p.126.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., ... & Perrin, N. (2014). Sex determination: why so many ways of doing it? PLoS biology, 12(7), e1001899.

Bairoch, A., Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28(1):45-48.

Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30:15 2114-2120.

Bronstein, O., Kroh, A., Tautscher, B., Liggins, L., Haring, E. (2017) Cryptic speciation in the pan-tropical sea urchins: a case study of an edge-of-range population of *Tripneustes* from the Kermadec Islands. Sci Rep 7:5948.

Cameron R.A., Samanta M., Yuan A., He D., Davidson E. (2009) SpBase: the sea urchin genome database and web site. Nucleic Acids Research 37:D750-754.

Chamberlin, R.V. , Ivie, W. (1935) The black widow spider and its varieties in the United States. Bulletin of the University of Utah 25(8):1-29.

Coppard, S.E., Kroh, A., Smith, A.B. (2012) The evolution of pedicellariae in echinoids: an arms race against pests and parasites. Acta Zoologica, 93:125-148.

Cyrus, M.D., Bolton, J.J., De Wet, L., Macey, B.M. (2014) The development of a formulated feed containing Ulva (Chlorophyta) to promote rapid growth and enhanced production of high quality roe in the sea urchin Tripneustes gratilla. Aquaculture Research 45: 159-176.

D'Aniello, S., Delroisse, J., Valero-Gracia, A., Lowe, E.K., Byrne, M., Cannon, J.T., Halanych, K.M., Elphick, M.R., Mallefet, J., Kaul-Strehlow, S., Lowe, C.J., Flammang, P., Ullrich-Lüter, E., Wanninger, A., Arnone, M.I. (2015) Opsin evolution in the

ambulacraria. Marine Genomics 24:177-183.

Davidson, E.H., Cameron, A.R. (2002) Arguments for sequencing the genome of the sea urchin *Strongylocentrotus purpuratus*. Retrieved from https://www.genome.gov/pages/research/sequencing/seqproposals/seaurchin_genome.pdf.

Dilly, G.F., Gaitán-Espitia, J.D., Hofmann, G.E. (2015), Characterization of the Antarctic sea urchin (*Sterechinus neumayeri*) transcriptome and mitogenome: a molecular resource for phylogenetics, ecophysiology and global change biology. Mol Ecol Resour, 15:425-436.

Ellisdon, A.M., Reboul, C.F., Panjikar, S., Huynh, K., Oellig, C.A., Winter, K.L., Dunstone, M.A., Hodgson, W.C., Seymour, J., Dearden, P.K., Tweten, R.K., Whisstock, J.C., McGowan, S. (2015) Stonefish toxin defines an ancient branch of the perforin-like superfamily. PNAS 112(50):15360-15365.

Eno, C.C., Böttger, S.A., Walker, C.W. (2009) Methods for karyotyping and for localization of developmentally relevant genes on the chromosomes of the purple sea urchin, *Strongylocentrotus purpuratus*. Biol Bull 217(3):306-12.

Fabricius, J.C. (1798) Scorpio. Supplementum Systema Systematicae, Impensis C.G. Proft, Hafniae 294-295.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279-D285.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. (2012) CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics 28(23):3150-3152.

Furlong, R.F., Holland, P.W. (2002) Bayesian Phylogenetic Analysis Supports Monophyly of Ambulacraria and of Cyclostomes. Zool Sci 19:593-599.

Garnier, P., Goudey-Perriere, F., Breton, P., Dewulf, C., Petek, F., Perriere, C. (1995) Enzymatic properties of the stonefish (*Synanceia verrucosa* Bloch and Schneider, 1801) venom and

purification of a lethal, hypotensive and cytolytic factor. Toxicon 33(2):143-155.

Ghadessy F.J., et al. (1996) Stonustoxin is a novel lethal factor from stonefish (*Synanceja horrida*) venom. cDNA cloning and characterization. J Biol Chem 271(41):25575-25581.

Gonzalez, P., Lessios, H.A. (1999). Evolution of sea urchin retroviral-like (SURL) elements: evidence from 40 echinoid species. Mol Biol Evol 16(7), 938-952.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, E., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocols 8(8):10.1038/nprot.2013.084.

Hatakeyama, T., Higashi E., Nakagawa, H. (2015) cDNA cloning and expression of Contractin A, a phospholipase A2-like protein from the globiferous pedicellariae of the venomous sea urchin *Toxopneustes pileolus*. Toxicon 108:46-52.

Hibino, T., Loza-Coll, M., Messier, C., Majeska, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K., Fugmann, S.D., Anderson, M.K., Pancer, Z., Cameron, R.A., Smith, L.C., Rast, J.P. (2006) The immune gene repertoire encoded in the purple sea urchin genome. Dev Bio 300:349-365.

Israel, J.W., Martik, M.L., Byrne, M., Raff, E.C., Raff, R.A., McClay, D.R., Wray, G.A. (2016) Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus Heliocidaris. PLOS Biology 14(3):e1002391.

Kober K.M., Bernardi G. (2013) Phylogenomics of strongylocentrotid sea urchins. BMC Evolutionary Biology 13:88.

Lamarck, J.B.M. de. (1816). Histoire naturelle des animaux sans vertèbres. Tome troisième. Paris: Deterville/Verdière. 612.

Láruson, Á. J. (2017). Rates and relations of mitochondrial genome evolution across the Echinoidea, with special focus on the superfamily Odontophora. Ecology and Evolution 7(13): 4543-4551.

Li. B., Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li, W., Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Linnaeus, C. (1758) Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata. Laurentius Salvius: Holmiae. ii, 14-664.

Lowe, E.K., Garm, A., Ullrich-Lüter, E., Arnone M.I. (2017) The crowns have eyes: Multiple opsins found in the eyes of the Crown-of-Thorns Starfish *Acanthaster planci*. bioRxiv 173187.

Mayr. E. (1954) Geographic speciation in tropical echinoids. Evolution 8(1):1-18.

McClay, D.R. (2011) Evolutionary crossroads in developmental biology: sea urchins. Development 138:2639-2648

Metschnikoff, V.E. (1881) Uber die systematische Stellung von Balanoglossus. Zool Anzeiger 4:153-157.

Metzger, M.J., Villalba, A., Carballal, M.J., Iglesias, D., Sherry, J., Reinisch, C., Muttray, A.F., Baldwin, S.A., Goff, S.P. (2016). Widespread transmission of independent cancer lineages within multiple bivalve species. Nature 534(7609), 705-709.

Oliver, T.A., Garfield, D.A., Manier, M.K., Haygood, R., Wray, G.A., Palumbi S.R.(2010) Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. Genome Biology and Evolution 2(1):800-814.

Osaki, E., Nishina, Y. Inazawa, J., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Ohsugi, M., Tezuka, T., Yoshida, M., Semba, K. (1999). Identification of a novel Sry-related gene and

its germ cell-specific expression. Nucleic Acids Res 27(12):2503-2510.

Pespeni, M.H., Sanford, E., Gaylord, B., Hill, T.M., Hosfelt, J.D., Jaris, H.K., LaVigne, M., Lenz, E.A., Russell, A.D., Young, M.K., Palumbi, S.R. (2013a) Evolutionary change during experimental ocean acidification. PNAS 110(17):6937-6942.

Pespeni, M. H., Barney, B. T. and Palumbi, S. R. (2013b) Differences in the regulation of growth and biomineralization genes revealed through long-term common garden acclimation and experimental genomics in the purple sea urchin. Evolution 67:1901-1914.

Poh, C.H., Yuen, R., Chung, M.C.M., Khoo, H.E. (1992) Purification and partial characterization of hyaluronidase from stonefish (*Synanceia horrida*) venom. Comu Biochem Physiol 101B:159-163.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raible, F. et al. (2006) Opsins and clusters of sensory G-protein-coupled receptors in the sea urchin genome. Developmental Biology 300(1):461-475.

Ramírez-Gómez, F., Ortíz-Pineda, P.A., Rojas-Cartagena, C., Su'arez-Castillo, E.C., García-Ararr'as, J.E. (2008) Immune-related genes associated with intestinal tissue in the sea cucumber *Holothuria glaberrima*. Immunogenetics 60:57-71.

Satoh, N., Rokhsar, D., Nishikawa, T. (2014) Chordate evolution and the three-phylum system. Proc R Soc B 281(1794):20141729.

Sea Urchin Genome Sequencing Consortium (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. Science 314(5801):941-952.

Sheppard-Brennand, H., Poore, A.G.B., Dworjanyn, S.A. (2017) A waterborne pursuit-deterrent signal deployed by a sea urchin. Am Nat 189(6):700-708.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210-3212.

Springer, M.S., Davidson, E.H., Britten, R.J. (1991). Retroviral-like element in a marine invertebrate. PNAS 88(19), 8401-8404.

Stimpson, W. (1857). On the Crustacea and Echinodermata of the pacific shores of North America. Boston Journal of Natural History. 6:444-532.

Troschel, F.H. (1872) Die Familie der Echinocidariden. Archiv für Naturgeschichte 38:293-356.

Tu, Q., Cameron, R.A., Worley, K.C., Gibbs, R.A. (2012) Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. Genome Res 22:2079-2087.

Valentine, J.P., Edgar, G.J. (2010) Impacts of a population outbreak of the urchin *Tripneustes gratilla* amongst Lord Howe Island coral communities. Coral Reefs 29:399-410.

Westbrook, C.E., Ringang, R.R., Cantero, S.M.A., HDAR & TNC Urchin Team, Toonen, R.J. (2015) Survivorship and feeding preferences among size classes of outplanted sea urchins, *Tripneustes gratilla*, and possible use as biocontrol for invasive alien algae. PeerJ 3:e123.

Chapter 4

Rates and relations of mitochondrial genome evolution across the Echinoidea, with special focus on the superfamily Odontophora

4.1 Abstract

In order to better characterize the placement of genus *Tripneustes*, as a representative of the Toxopneustidae family within the broader sea urchin mitochondrial (MT) phylogeny, the complete MT genome of *Tripneustes gratilla* was generated and compared with all published echinoid MT genomes currently available on NCBI GenBank. The MT genome phylogeny supports the existence of the superfamily Odontophora (consisting of the families Strongylocentrotidae, Echinometridae, and Toxopneustidae). A relaxed molecular-clock time calibration suggests a split between the three key Odontophore MT lineages occurred during the late Eocene/Oligocene. Major global oceanographic changes have been inferred during this time frame, potentially driving species diversification through environmental selection pressures. To test for signatures of selection acting on the mitochondria, the historical rate of gene evolution of individual MT genes was assessed through a branch-site comparison of non-synonymous to

synonymous substitution ratios (ω). Models of positive selection and neutral evolution, as compared via a likelihood ratio test, show no evidence of strong historical positive selection on mitochondrial genes at the genesis of the Odontophora. However, while pairwise ω comparison revealed signatures of strong negative selection, relatively elevated ω values were observed within the *Strongylocentrotus* genus.

4.2 Introduction

There are presently over 1,000 described species of sea urchins that collectively make up the class Echinoidea, the sea urchins, within the Phylum Echinodermata (Appeltans et al. 2012). Arising early on in the fossil record (Ordovician) as a distinct class, and lending themselves well to fossilization, echinoids have long been a hallmark of systematic research (Agassiz and Clark 1907-1917, Clark 1912, Fell 1974, Kroh and Smith 2010, Littlewood and Smith 1995, Smith 1988, Smith and Savill 2001). Following the publication of Theodore Mortensen's detailed alpha taxonomy of echinoids between 1928 and 1951, sea urchins became a tractable system to investigate questions of evolution and speciation (Mayr 1954, Mortensen 1928-1954, Palumbi and Lessios 2005). Within the Echinoidea resides the ominously named Toxopneustidae family, whose members include the highly venomous "flower urchin," *Toxopneustes pileolus* (Nakagawa et al. 1996), the much-studied Caribbean "green sea urchin," *Lytechinus variegatus* (Watts et al. 2013), and the Indo-Pacific "collector urchin," *Tripneustes gratilla* (Lawrence and Agatsuma 2013), along with eight other recognized species (Appeltans et al. 2012). Though these species are well-recognized and studied, their familial phylogeny is not as well established. Previous attempts at placement of the Toxopneustidae family within the greater sea urchin phylogenetic tree using morphology has suggested that the Toxopneustidae, Echinometridae, and the Strongylocentrotidae together form the superfamily Odontophora (Kroh and Smith 2010). The

complete MT genome of *T. gratilla*, reported here, represents the last genome needed in order to compare all MT family lineages within the proposed Odontophora superfamily. Markers derived from MT genomes have long served as the molecular standard for species delineation, and population connectivity, as well as deeper evolutionary relationships (Avise et al. 1987, Ballard and Whitlock 2004). Part of the appeal of the MT marker has been not only its ease of recovery during DNA extraction owing to the high copy numbers of MT genetic material per cell, but also the lack of recombination and a general assumption of freedom from strong positive selection (Grey 1989). However, a growing number of studies have suggested natural selection on MT genes may not be as rare as previously assumed (Ballard and Whitlock 2004, Bazin et al. 2006, Doi et al. 1999, Stojković et al. 2016). Genetic variation of MT genes may be driven by thermal adaptation in ectothermic poikilotherms, such as sea urchins, as the thermal stability of transcribed proteins is critically important to function (Guderley and St-Pierre 2002, Hazel 1999). Indeed, periods of great oceanographic change, including significant global ocean cooling, have been linked to species divergence as a result of climate driven selection pressures (Prothero and Berggren 1992). Beyond thermal adaptation, cytonuclear incompatibility can also serve as a selection force on MT genes, which rely on nuclear molecular machinery. Cytonuclear incompatibility may arise through population divergence, as nuclear and MT genes evolve within a population, and can serve as a mechanism to restrict gene flow during secondary contact between populations (Burton and Barreto 2012). One approach to detect molecular signatures of selection across lineages is to quantify the proportion of nucleotide substitutions occurring at non-synonymous codon sites compared to the proportion of synonymous substitutions, accounting for the degree of degeneracy at each codon site. This ratio of non-synonymous to synonymous substitutions (dN/dS or ω) can serve as a highly conservative estimate of selection when comparing lineages. An assumption is made that the majority of synonymous substitutions are selectively neutral, while the occurrence of non-synonymous substitutions are presumably selection driven. Values of ω close to 1 suggest neutrality, values

much less than 1 suggest the action of purifying selection in removing amino acid substitutions, and values much greater than 1 pointing to positive selection on amino acid change (Nielsen and Yang 2003, Kryazhimskiy and Plotkin 2008). Extending this approach to inferred historical sequences of shared ancestors allows for an estimation of past instances of selection-driven divergence. This study does just that by estimating ω values for each branch of a phylogenetic tree generated from complete MT genomes of 14 taxa across the Echinoidea, including 10 from the Odontophora. Two branch-site tests of positive selection, a "strict" and "relaxed" variety, were performed on the branch giving rise to the Odontophora, in order to assess the influence of climate shifts driving lineage divergence through positive natural selection.

4.3 Materials & Methods

Whole RNA was extracted from two adult T. gratilla, as well as from approximately 1,000 plutei-stage T. gratilla larvae, using a Qiagen RNeasy extraction kit. All specimens were acquired from O'ahu, Hawai'i. Poly A-tail hybridization and cDNA generation was accomplished with an Illumina TruSeq kit and sequencing was performed on an Illumina MiSeq (300 cycle, single end) at the Hawaii Institute of Marine Biology (HIMB) Core Genetics facility, O'ahu, Hawai'i. Raw reads were filtered for adapter sequences using BBDuk in the BBMap package (Bushnell 2015) and assembled in Trinity (v.3.1.1, Grabherr et al. 2011). MT gene sequences were identified via annotation to the MT genome of the Strongylocentrotid *Hemicentrotus pulcherrimus* in Geneious (v.6). A draft MT genome for *T. gratilla* was generated by annotating the concatenated MT gene sequences at a 90% identity threshold to 12 echinoid MT genomes. All 13 coding DNA sequences (CDS), 22 tRNA, and 2 rRNA sequences were thus identified on the draft MT genome. Using the draft MT genome, 40 pairs of novel primers were designed using the Owczarzy et al. (2004) salt correction formula, and thermodynamic parameters from SantaLucia (1998), as implemented in

Primer3plus (http://www.primer3plus.com, Untergasser et al. 2012). A combination of regular and long-range primers (used with NEB Long Range Taq 2x) were designed to generate overlapping products across the whole *T. gratilla* MT genome. An annealing temperature of 58°C was used for all thermocycler runs. DNA from a third adult T. gratilla, collected from Western O'ahu, Hawai'i, was amplified using the novel primers in PCR, and resultant products were sequenced via Sanger sequencing on a 3730 ABI DNA Analyzer at the Advanced Studies in Genomics, Proteomics, and Bioinformatics (ASGPB), O'ahu, Hawai'i. Overlap of resultant amplicons allowed for the inclusion of intergenic sequences and clarification of a few low quality areas from the draft MT genome to give a complete and high confidence MT genome sequence. The *T. gratilla* MT genome is available on NCBI GenBank (Accession number: KY268294). In total, 12 complete MT genomes of echinoids were acquired from NCBI Genbank. This included *Arbacia lixula* (NC001770, De Giorgi 1996), *Paracentrotus lividus* (NC001572, Cantatore 1989), *Loxechinus albus* (NC024177, Jung and Lee 2015), *Sterechinus neumayeri* (NC020771, Dilly et al. 2015), *Heliocidaris crassispina* (NC023774, Jung et al. 2016), *Hemicentrotus pulcherrimus* (NC023771, Jung et al. 2014), *Mesocentrotus franciscanus* (NC024177, Gaitan-Espitia 2016), *Mesocentrotus nudus* (NC020771, Jung et al. 2013), *Pseudocentrotus depressus* (NC023773), *Strongylocentrotus purpuratus* (NC001453, Jacobs et al. 1988, Qureshi and Jacobs 1993, Valverde et al. 1994), *S. droebachiensis* (NC009940), *S. pallidus* (NC009941), and *S. intermedius* (NC023772). Initially the MT genome of *Temnopleurus hardwickii* (NC026200, Fu et al. 2014), from the family Temnopleuridae, was included in the analysis. However, due to ambiguous phylogenetic signals across the *T. hardwickii* MT genome and a polytomous assignment of the species, it was dropped from the final analysis. Full MT genome alignments were performed with MAUVE as implemented in Geneious v.6 (Darling et al. 2010). Sequence length comparisons of the 15 MT genes were also performed in Geneious v.6. The two non-coding 12s-rRNA and 16s-rRNA sequences were aligned in T-COFFEE using the ribosomal folding algorithm implemented in the RCoffee mode (Notredame et al. 2000, Wilm et al. 2008). Comparative alignments of the rRNA sequences were

executed with the Muscle (Edgar 2004) and ClustalW (Thompson et al. 1994) algorithms, as implemented in Geneious v.6, under default settings. The 13 coding DNA sequences (CDS) were aligned using the amino acid alignment algorithm in Geneious v.6, using the echinoderm MT codon table. Best fit nucleotide substitution models for the initial phylogenetic analysis for each of the 15 gene alignments were determined with JModeltest 2.1.7 (Darriba et al. 2012). From the 15 gene alignment dataset, a phylogenetic tree was generated with MrBayes 3.2.3, which was run for 5,000,000 steps; with a 1,250,000 step burn-in (Huelsenbeck et al. 2001, Ronquist and Huelsenbeck 2003), with *A. lixula* designated as an out group in order to root the tree. To confirm the robustness of the topology, a phylogenetic tree was also generated from the same data using RAxML v.8, which was run for 50,000 steps, with non-parametric bootstrapping enabled (Stamatakis 2014). For the time calibration and gene evolution analysis a second Bayesian majority consensus tree was generated, under identical parameters as the prior tree, but including only the 13 CDS, not the 12S and 16S sequences. This CDS tree was generated to prevent confounding of codon substitution rate assessments with non-coding gene sequences. Divergence calibration using the CDS majority consensus tree under a Birth-Death speciation process was performed in BEAST v1.8.7 (Drummond et al. 2012), with eight calibration points set. One log-normally distributed fossil calibration point was set at the split of the Odontopohora from the Parechinidae (represented by *Loxechinus albus* and *Paracentrotus lividus*) and Echinidae (*Sterechinus neumayeri*), with an initial value of 42.5mya (μ=5, offset=40, σ=1, Smith 1988, Lee 2003). All remaining node calibrations were normally distributed, and centered on splitting times derived from previous work on echinoid MT sequence evolution: *Sterechinus neumayeri* from the Parechinidae (μ=29.5, σ=4, Lee et al. 2004); *Loxechinus albus* from *Paracentrotus lividus* (μ=25.5, σ=3, Lee et al. 2004); root node of the Strongylocentrotidae (μ=15.5, σ=2, Lee 2003); *Hemicentrotus pulcherrimus* from Strongylocentrotus (μ=8.6, σ=1, Lee 2003); root node of *Mesocentrotus* (μ=6.9, σ=1.5, Lee 2003); root node of *Strongylocentrotus* (μ=5.6, σ=1, Lee 2003); and *S. droebachiensis* from *S. pallidus* (μ=2.6, σ=1, Lee 2003). Partitioning schemes for models of molecular evolution for the

fossil calibration run were determined with PartitionFinder v1.1.1 (Lanfear et al. 2012). Marginal likelihood estimations using path sampling and stepping stone sampling were used for molecular clock models ranging from a strict to relaxed model and resulted in the selection of the relaxed log-normal clock for the analysis (Baele et al. 2012). Calibrations were run for 100,000,000 generations and repeated 5 times, before being combined with logcombiner (Drummond et al. 2012). Proper Markov chain mixing was confirmed for both MrBayes and BEAST runs with Tracer 1.6 (Rambaut et al. 2014). The resulting tree was visualized using the R packages phytools (Revell 2012), PHYLOCH (Heibl 2008), strap (Bell and Lloyd 2014), and CODA (Plummer et al. 2006). A free rate model estimating the ratio of non-synonymous to synonymous substitutions (ω) independently across every branch of the CDS tree was estimated with the codeml package within PAML (Yang 2007, Zhang et al. 2005), and pairwise ω were estimated with the Nei and Gojobori (1986) method. Three branch-site models of ω variation along the Odontophore branch of the MT genome tree were then compared. The echinoderm specific MT codon table was specified for all calculations. Model MA fixes ω at 1 for every branch except for the specified branch leading to the Odontophora, where ω is assumed to be greater than 1. This model serves as the alternative to the two following null models. Model M1a similarly fixes ω at 1 for every branch except for the Odontophora branch, where ω is assumed to be ranging from 0 to 1; and MAnull fixes ω at 1 for every branch in the tree. Both MAnull and M1a were compared via a Likelihood Ratio Test (LRT) to model MA. Comparing MA to MAnull is the "strict" branch-site test for positive selection, as significance requires $\omega > 1$ at some subset of codon sites along the Odontophore branch. The LRT of MA to M1a is the "relaxed" test of positive selection, as significance requires only that a subset of codon sites along the branch of interest display an elevated ω when compared to the same sites along the background branches, and can thus identify instances of relaxation of selective constraints as well as positive selection. Significance of two times the difference in likelihood ($2\Delta\ell$) was estimated on a 50:50 mix of point mass 0 and distribution, with critical values of 2.71 for 5% and 5.41 for 1% significance levels (Self and Liang

1987, Zhang et al. 2005).

4.4 Results

The *Tripneustes gratilla* MT genome is 15,725 bp long, with a GC% of 40.3. This is just above the average MT genome length of 15,700.4 bp (s.d. 23.6) and mean GC% of 40.2. The largest MT genome so far reported is that of *P. depressus* (15,729 bp), while the smallest is that of *M. franciscanus* (15,649 bp). MT genome alignments in Mauve showed no gene re-arrangements in any taxa. Comparative gene length summaries across the MT genomes revealed roughly four categories of nucleotide length variation across the 14 species. In order to compare length variance across genes of different length, the coefficient of variation was found by normalizing the standard deviation for each gene by the gene length. Category A (genes COX2, COX3, ND3, and ND4L) had no deviation in length across all sampled taxa; Cat. B (COX1, ND1, ND2, and ND4) averaged a coefficient of variation (CV) of $7.5\text{x}10^{-4}$; Cat. C (16S, ATP6, ATP8, and CYTB) had an average CV of $4.8\text{x}10^{-3}$; while Cat. D (12S, ND5, and ND6) had a value of $9.7\text{x}10^{-3}$. There was no significant correlation between gene length and CV (Pearson's r = 0.12). The categories did not strictly reflect the number of codon inserts or deletions across the samples. Values are summarized in Table 4.1.

Individual gene trees for 16S, ATP8, COX1, COX3, ND2, and ND3 recovered a monophyletic Odontophora superfamily. However both ATP6 and CYTB failed to place *T. gratilla* in close association with Echinometridae & Strongylocentrodidae, while 12S, COX2, ND1, ND4, ND4L, ND5, and ND6 did not recover a grouping of Strongylocentrodidae with either Echinometridae or Toxopneustidae. The Bayesian consensus tree strongly supported the placement of *Tripneustes gratilla*, representative of the Toxopneustidae family, as the closest MT sister clade to the Strongylocentrotidae with a 100% posterior probability, as shown in Figure 4.1. Within the

TABLE 4.1: Average length and the coefficient of variation of the 15 MT genes across 14 echinoid taxa

Gene	Mean Length	CV	CV Category
ND4L	294	0	A
ND3	351	0	A
COX2	690	0	A
COX3	783	0	A
COX1	1,553.8	0.000516	B
ND2	1,058.8	0.000757	B
ND1	971.8	0.000825	B
ND4	1,389	0.000885	B
16S	1,539.3	0.004198	C
ATP8	165.2	0.004853	C
CYTB	1,144.5	0.004904	C
ATP6	691.1	0.005278	C
ND5	1,914	0.009077	D
12S	893.2	0.009276	D
ND6	493.1	0.010833	D

Strongylocentrotidae the *Mesocentrotus* and *Pseudocentrotus* genera formed a unique clade, with *P. depressus* nesting within the *Mesocentrotus*, sister to *M. franciscanus*. *H. pulcherrimus* appeared as a sister taxa to the *Strongylocentrotus genus*. The four *Strongylocentrotus* species formed a clade with *S. pallidus* and *S. droebachiensis* as the most recent split, followed by *S. purpuratus*, and finally *S. intermedius*. Outside the Strongylocentrotidae and Toxopneustidae clade, the representative of the Echinometridae *Heliocidaris crassispina* branched off, completing the three family cluster identified as the superfamily Odontophora by Kroh and Smith (2010). Splitting from the Odontophora within the Camarodonta order are the families Echinidae (represented by *S. neumayeri*) and Parechinidae (represented by *P. lividus* and *L. albus*). *A. lixula*, the outgroup, displays the split between Order Arbacioida and the Camarodonta (Figure 4.1). A maximum likelihood tree generated with RAxML yielded an identical topology, so only the Bayesian tree is shown. The Bayesian concensus tree generated using only CDS (excluding 12S and 16S) also

yielded an identical topology, and was used for ω estimation. Free rate estimation of ω for each coding gene across all branches suggested no strong signal at any one point in the CDS tree. Any elevated values of ω were found to be caused by synonymous substitution site proportions of zero, which inflated the effect of comparatively minute values of associated non-synonymous substitution proportions. Although elevated non-symonymous (dN) substitution rates were observed on the *S. purpuratus* branch, relative to all other branches within the Strongylocentrotidae. A one tailed t-test, comparing the proportion of non-synonymous substitutions averaged across all CDS (dN) of the S. purpuratus branch to the dN value averaged across the remaining seven branches of Strongylocentrodiae was significant at the 5% threshold with a Bonferroni correction (n=182, p=1.117x10^{-3}). Pairwise ω values were low across all genes, suggesting purifying selection, with one notable exception. ATP8 gave an ω value of 0.92 in the pairwise comparison between *P. lividus* and *A. lixula*, and values of 1.44, 0.94, and 1.06 in the pairwise comparisons between *S. purpuratus* and *S. droebachiensis*, *S. intermedius*, and *S. pallidus*, respectively. Averaging across all gene pairwise comparisons, the highest relative ω values were consistently observed between *S. purpuratus* and its three congenerics. Average pairwise ω are summarized in Figure 4.2. Testing specifically the Odontophore branch, no CDS had significant results of positive selection according to the "strict" branch-site test, with only three genes (ND1, ND5, and COX1) exhibiting differences greater than zero. Of those three, only ND1 showed a significant difference at the 5% level in the "relaxed" test for selection (Table 4.2).

4.5 Discussion

The placement of the Toxopneustidae family, represented here by the *Tripneustes gratilla* MT genome, as a closer sister group to the Strongylocentrotidae than the Echinometrid *Heliocidaris crassispina* contradicts the accepted species tree (Kroh and Smith 2010). It must be specified that

TABLE 4.2: Results of the strict and relaxed branch-site tests of positive selection

	ND1	ND5	COX1
Strict test			
$2\Delta\ell$	1.6998	1.0394	0.0157
p-value	0.0962	0.154	0.4501
Relaxed test			
$2\Delta\ell$	4.2609	1.0394	0.2097
p-value	0.0195	0.154	0.3235

this relationship is based solely on MT sequences and is thus an insight into the evolutionary relationships of the Echinoid mitochondria, not a definitive depiction of the full evolutionary history of these species. Incomplete lineage sorting between the mitochondrial genome and the nuclear genome, or possibly lineage capture, could account for the contradictory signals of the MT and species tree in the sequence of splitting events at the cladogenesis of the Odontophora. The placement of *Pseudocentrotus depressus* as nested within the genus *Mesocentrotus* is consistent with previous MT gene analysis of the Strongylocentrotidae (Kober and Bernardi 2013). While comparing the individual gene phylogenies of the taxa considered here, it became clear that the choice in alignment method of ribosomal RNA 12S and 16S sequences had a significant impact on the phylogentic signal of these two markers. Alignments using the built-in algorithm of the Geneious software, as well as ClustalW, and Muscle alignments of the rRNA markers yielded trees for each marker that differed in topology from the CDS tree consensus. Yet when the rRNA markers were aligned using the ribosomal alignment algorithm (rcoffee mode) in T-Coffee, which uses ribosomal folding patterns as a cue, the 16S alignment resulted in an identical tree topology with the CDS consensus tree. Regardless of alignment method, the shorter 12S rRNA alignment was unable to resolve most relationships, showing primarily polytomies. 12S and 16S have been commonly used in many phylogenetic studies, however the inability of either of these markers to recover even a resemblance of the consensus tree when aligned with commonly used nucleotide

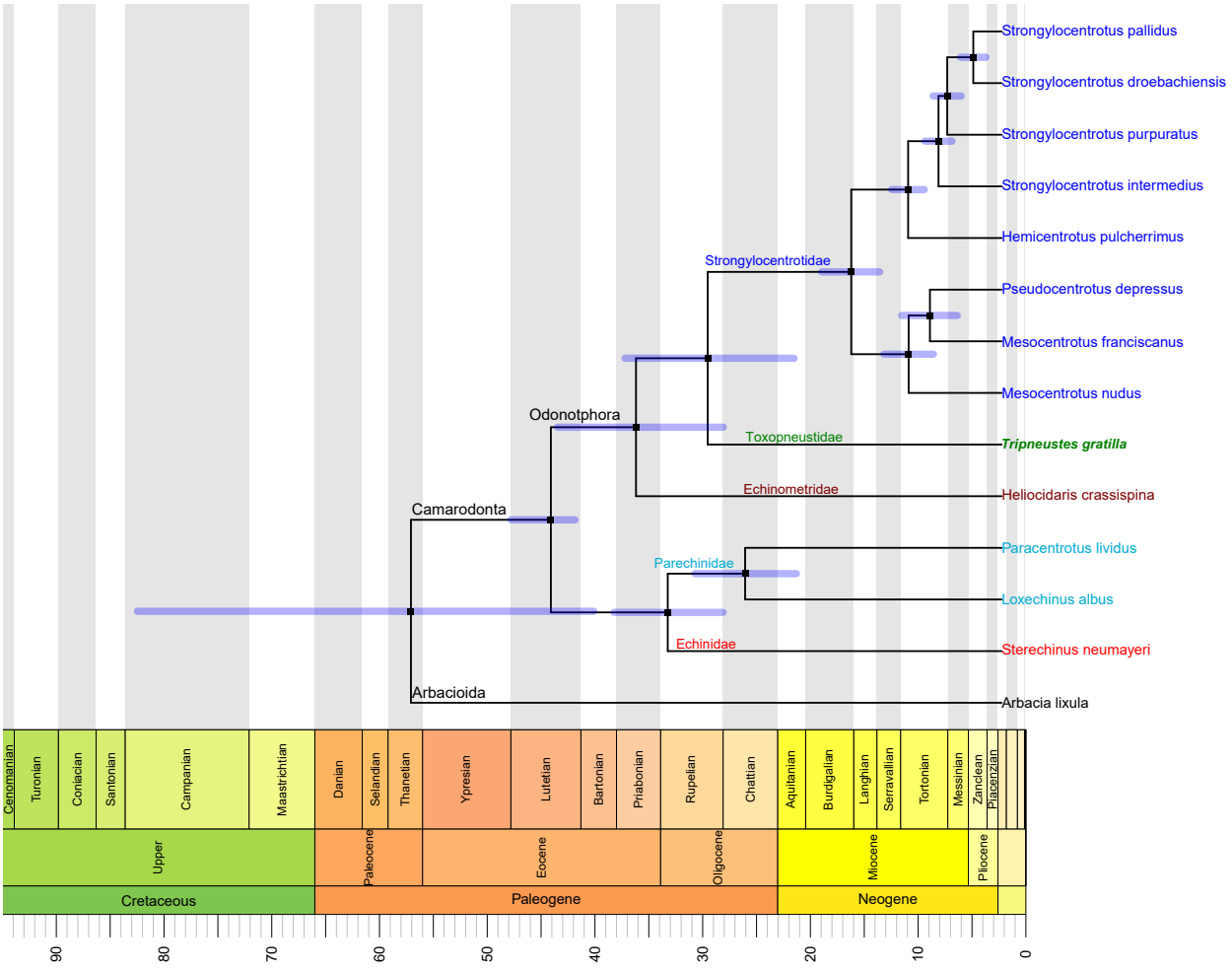

FIGURE 4.1: Bayesian tree showing mitochondrial genome relationships among the Echinoids. Posterior probabilities at all bifurcating nodes were 100%. Species are displayed to the right of the branch tips and color coded to their representative families. Blue error bars represent the 95% CI of the node height.

alignment algorithm could cast doubt on phylogenetic inferences derived solely from rRNA markers aligned via traditional means.

In contrast to the Smith et al. (2006) estimation of divergence times within the Class Echinoidea, which utilized fewer calibration points based on fossil dates of deep tree nodes, divergence

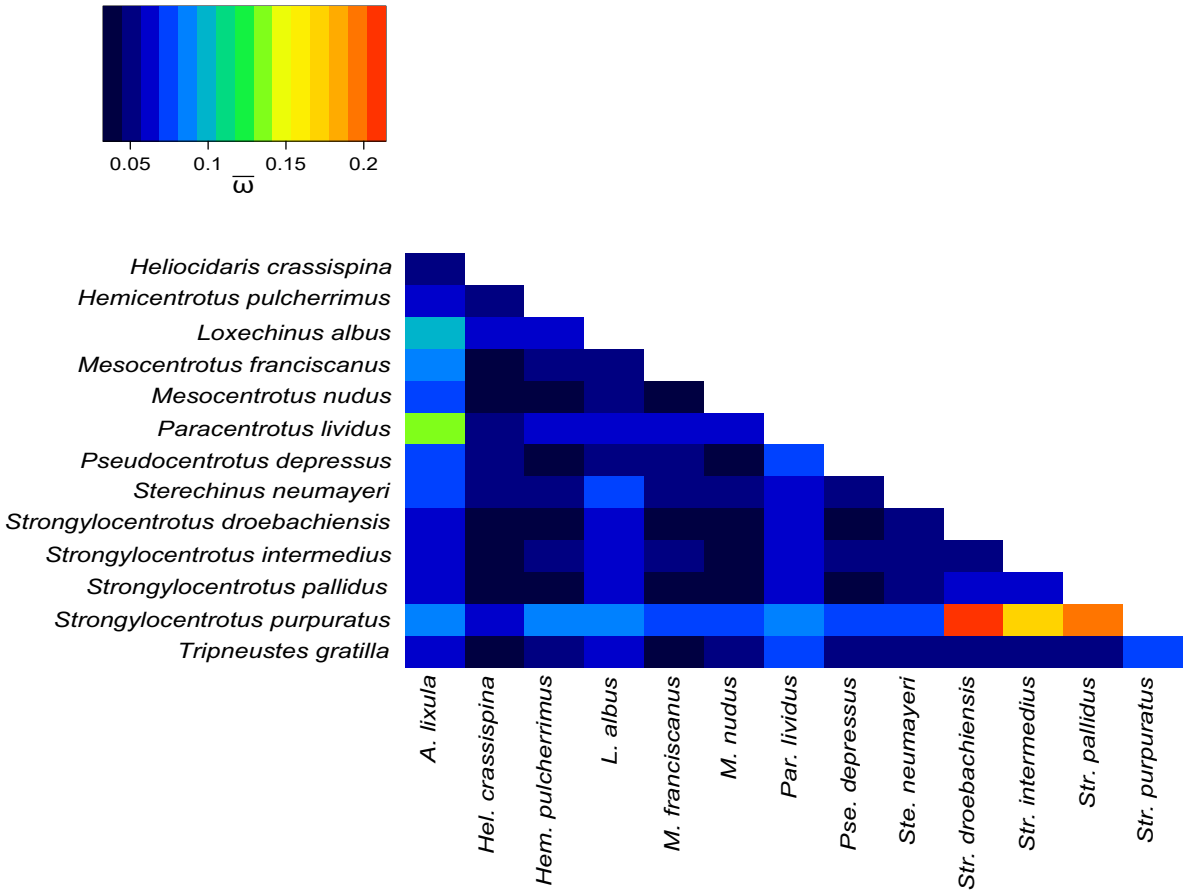

FIGURE 4.2: A heatmap depicting ω averaged across all 13 CDS, for each pairwise comparison. Lower values are depicted as cooler colors (blue), and higher values are presented as warmer colors (red). The largest values were consistently observed within genus *Strongylocentrotus*, specifically when *S. purpuratus* was compared to its congenerics *S. pallidus* and *S. intermedius*.

estimates in this study relied on multiple calibrations of mostly shallow nodes. The calibration of the fully bifurcating tree indicates that the split between the three major Odontophores (Toxopneustidae, Strongylocentrotidae, and Echinometridae) occurred within a chronological range of the Eocene/Oligocene boundary. This period saw significant oceanographic changes due to a number of factors, such as the opening of south sea passage ways, and increased ice

sheet coverage in the Arctic and Antarctic. A marked increase in atmospheric oxygen concentrations during this time contributed broadly to global cooling, with mean ocean surface temperatures estimated to have dropped by 4.4°C (Liu et al. 2009). This period gave rise not only to the Odontophora, but also to the splitting of the Parechinidae and Echinidae families; the Echinidae containing a prominent cold water sea urchin, the Antarctic *Sterechinus neumayeri*. While modern psychrospheric (deep cold ocean) fauna are traced back to this period of large scale climate change (Prothero and Berggren 1992), and the divergence of toothed and baileen whales appears to have been specifically driven by abiotic factors during this time (Steeman et al. 2009), no strong signature of selection was detected at the genesis of the Odontophora. Consistent with previous analysis, whole MT genome alignments show no gene re-arrangements across the Echinoidea (Lee et al. 2004). The absence of regular recombination, generally assumed with mitochondria, can significantly inhibit the rates of fixation of advantageous mutants (Hill and Robertson 1966). This could explain the lack of a strong positive selection driven signal. However, this does not discount the hypothesis that ecological and climatic shifts may have driven lineage diversification, but simply that MT genes were not predictably contributing. Nuclear genes, especially those involved in environmental stress response, could better indicate a pattern of abiotic selection pressure. Pespeni et al. (2013) showed that within *S. purpuratus* nuclear genes underlying some 40 functional classes of proteins undergo a shift in allele frequencies between populations reared at different levels of CO2 concentrations. As more nuclear markers and representative genomes become available for members of the echinoidea, so will a more definitive test of selection driven diversification. Mapping exclusively non-synonymous (dN) proportions onto the CDS tree showed no single lineage with especially increased non-synonymous substitution rates, save for one. The *S. purpuratus* branch across all MT genes consistently showed comparatively elevated non-synonymous ratios, and the largest pairwise values of ω . As a model organism further study of the Strongylocentrotidae is warranted in order to determine whether the MT of *S. purpuratus* is indeed uniquely divergent

from its congenerics due to selection (Sea Urchin Genome Sequencing Consortium 2006). However, consistent pairwise ω values much below one in most all taxa comparisons suggests that the major force affecting MT sequences is purifying, or negative, selection. The generation of a complete MT genome of *Tripneustes gratilla*, the first for a member of Toxopneustidae, has allowed for the confirmation of the existence of the Odontophora superfamily. This underscores the fact that MT genomes can serve as a powerful analytical unit when estimating species relationships, especially in systems that may otherwise lack publicly available nuclear sequences for comparisons. It is the hope of the author that researchers generating NGS data will view the isolation, proper annotation, and dissemination of captured MT genome sequences to be a worthwhile endeavor.

4.6 Acknowledgments

This research was funded and made possible by the Jessie D. Kay Memorial Fellowship, the Hampton & Meredith Carson Fellowship administered by the Ecology, Evolution, and Conservation Biology specialization program at the University of Hawai'i at Mānoa, a Sigma Xi Grant-in-Aid of Research, and the Linnean Society of London & Systematics Association Research Fund. The author would like to gratefully acknowledge the input and guidance of Dr. Floyd Reed and Dr. Robert Thomson, both at the University of Hawai'i at Mānoa, Department of Biology. Special thanks to Victoria Sindorf for comments and corrections. Additional thanks to associate editor Will Allen and two anonymous reviewers, for helpful comments and suggestions.

4.7 References

Appeltans, W., P. Bouchet, G. A. Boxshall, K. Fauchald, D. P. Gordon, B. W. Hoeksema, G. C. B. Poore, R. W. M. Van Soest, S. Stöhr, and T. C. Walter, et al. 2012. World register of marine species. Proceedings of the future of the 21st century ocean: Marine Sciences and European Research Infrastructures, Brest, France, 28:30.

Agassiz, A., and H. L. Clark. 1907-1917. Hawaiian and other Pacific Echini. UK:Cambridge.

Avise, J.C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Eco. Syst. 18:489-522.

Baele, G., P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard, and A. V. Alekseyenko. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29(9):2157-2167.

Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13(4):729-744.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312(5773):570-572.

Bell, M. A., and G. T. Lloyd. 2014. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. Palaentology 58(2):379-389.

Burton, R.S., and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol. Ecol. 21:4942-4957.

Bushnell, B. 2015. BBMap - sourceforge.net/projects/bbmap/

Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta, and C. Saccone. 1989. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of Paracentrotus lividus. J. Biol. Chem. 264(19):10965-10975.

Clark, H. L. 1912. Hawaiian and other Pacific Echini. UK: Cambridge.

Darling, A. E., B. Mau, N. T. Perna. 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. PLoS One. 5(6):e11147.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772

De Giorgi, C., A. Martiradonna, C. Lanave, and C. Saccone. 1996. Complete sequence of the mitochondrial DNA in the sea urchin Arbacia lixula: Conserved features of the echinoid mitochondrial genome. Mol. Phylogenet. Evol. 5(2):323-332.

Dilly, G. F., J. D. Gaitn-Espitia, and G. E. Hofmann. 2015. Characterization of the antarctic sea urchin (Sterechinus neumayeri) transcriptome and mitogenome: a molecular resource for phylogenetics, ecophysiology and global change biology. Mol. Ecol. Resour. 15(2):425-436.

Doi, A., H. Suzuki, and E. T. Matsuura. 1999. Genetic analysis of temperature dependent transmission of mitochondrial DNA in drosophila. Heredity 82(5):555-560.

Drummond, A.J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7 Mol. Biol. Evol. 29:1969-1973.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput Nucleic Acids Res. 32(5):1792-1797.

Fell, F. J. 1974. The echinoids of Easter Island (Rapa Nui). Pac Sci. 28(2):147-158.

Fu, W., C. Liu, X. Zeng. 2014. The complete mitochondrial genome of Temnopleurus hardwickii (Camaradonta: Temnopleuridae). Mitochondr. DNA 29:1-2.

Gaitan-Espitia, J. D., G. E. Hofmann. 2016. Mitochondrial genome architecture of the giant red sea urchin Mesocentrotus franciscanus (Strongylocentrotidae, Echinoida). Mitochondr. DNA A. 27(1):591-592.

Grabherr, M. G., B. J. Haas, M .Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 15;29(7):644-52.

Grey, M. W. 1989. Origin and evolution of mitochondrial DNA. Annu. Rev. Cell Bio. 5:25-50.

Guderley, H., and J. St-Pierre. 2002. Going with the flow or life in the fast lane: contrasting mitochondrial responses to thermal change. J. Exp. Biol. 205:2237-2249.

Heibl, C. 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. http://www.christophheibl.de/Rpackages.html. Version 1.5-3.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8: 269-294.

Huelsenbeck, J.P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.

Jacobs, H. T., D. J. Elliott, V. B. Math, and A. Farquharson. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. 202(2):185-217.

Jung, G., H. J. Choi, J. G. Myoung, and Y. H. Lee. 2014. Complete mitochondrial genome of sea urchin: Hemicentrotus pulcherrimus (Camarodonta, Strongylocentrotidae). Mitochondr DNA 25(6):439-440.

Jung, G., C. G. Kim, and Y. H. Lee. 2016. Complete mitochondrial genome sequence of Heliocidaris crassispina (Camarodonta, Echinometridae). Mitochondr DNA A. 27(4):2393-2394.

Jung, G., H. J. Choi, S. Pae, and Y. H. Lee. 2013. Complete mitochondrial genome of sea urchin: Mesocentrotus nudus (Strongylocentrotidae, Echinoida). Mitochondr. DNA 24(5):466-468.

Jung, G., and Y. H. Lee. 2015. Complete mitochondrial genome of chilean sea urchin: Loxechinus albus (Camarodonta, Parechinidae). Mitochondr. DNA 26(6):883-884.

Kober, K. M., and G. Bernardi. 2013. Phylogenomics of strongylocentrotid sea urchins. BMC Evol. Biol. 13(1):88.

Kroh, A., and A. B. Smith. 2010. The phylogeny and classification of post-Palaeozoic echinoids. J Syst Palaeontol. 8(2):147-212.

Kryazhimskiy, S., and J. B. Plotkin. 2008. The population genetics of dN/dS. PLoS Genet. 4(12):e1000304.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29(6):1695-1701.

Lawrence, J. M., and Y. Agatsuma. 2013. Tripneustes. In: Lawrence JM, editor. Sea Urchins: Biology and Ecology. Oxford, UK: Elsevier. p. 491-507.

Lee, Y. H. 2003. Molecular phylogenies and divergence times of sea urchin species of Strongylocentrotidae, Echinoida. Mol. Biol. Evol. 20(8):1211-1221.

Lee, Y. H., M. Song, S. Lee, R. Leon, S. O. Godoy, and I. Canete. 2004. Molecular phylogeny and divergence time of the Antarctic sea urchin (Sterechinus neumayeri) in relation to the South American sea urchins. Antarct. Sci. 16(1):29-36.

Littlewood, D. T., and A. B. Smith. 1995. A combined morphological and molecular phylogeny for sea urchins (Echinoidea: Echinodermata). Philos. Trans. R. Soc. Lond. B Biol. Sci. 347(1320):213-234.

Liu, Z., M. Pagani, D. Zinniker, R. DeConto, M. Huber, H. Brinkhuis, S. R. Shah, R. M. Leckie, and A. Pearson. 2009. Global cooling during the Eocene-Oligocene climate transition. Science 323:1187-1190.

Mayr, E. 1954. Geographic speciation in tropical echinoids. Evolution 8(1):1-18.

Mortensen, T. 1928-1951. A monograph of the Echinoidea. A Monograph of the Echinoidea. Copenhagen, DK: Reitzel CA.

Nakagawa, H., H. Hashimoto, H. Hayashi, M. Shinohara, K. Ohura, E. Tachikawa, and T. Kashimoto. 1996. Isolation of a novel lectin from the globiferous pedicellariae of the sea urchin Toxopneustes pileolus. In: Singh BR, Tu AT, editors. Natural Toxins 2: Structure, Mechanism of Action, and Detection. Boston, MA: Springer US. p. 213-223.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3(5):418-426.

Nielsen, R., and Z. Yang. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. Mol. Biol. Evol. 20:1231-1239.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302:205-217.

Owczarzy, R., Y. You, B. G. Moreira, J. A. Manthey, L. Huang, M. A. Behlke, and J. A. Walder. (2004). Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. Biochemistry 43(12):3537-3554.

Palumbi, S. R., and H. A. Lessios. 2005. Evolutionary animation: how do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? PNAS 102:6566-6572.

Pespeni, M.H., E. Sanford, B. Gaylord, T.M. Hill, J.D. Hosfelt, H.K. Jaris, M. LaVigne, E.A. Lenz, A.D. Russell, M.K. Young, S.R. Palumbi. 2013. Evolutionary change during experimental ocean acidification. PNAS 110(17):6937-6942.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6(1):7-11.

Prothero, D. R., and W. A. Berggren, eds. 1992. Eocene-Oligocene climatic and biotic evolution. Princeton Univ. Press, New Jersey, USA.

Qureshi, S.A., and H. T. Jacobs. 1993. Two distinct, sequence-specific DNA-binding proteins interact independently with the major replication pause region of sea urchin mtDNA. Nucleic Acids Res. 21(12):2801-2808.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer

Revell, L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217-223.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

SantaLucia, J., Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. PNAS 95(4):1460-1465.

Stamatakis, A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics 30(9): 1312-1313.

Steeman, M.E., M.B. Hebsgaard, R.E. Fordyce, S.Y.W. Ho, D.L. Rabosky, R. Nielsen, C. Rahbek, H. Glenner, M.V. Sørensen, E.Willerslev. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. Syst Biol 58 (6):573-585. doi: 10.1093/sysbio/syp060

Sea Urchin Genome Sequencing Consortium. 2006. The genome of the sea urchin Strongylocentrotus purpuratus. Science 314(5801):941-952.

Self, S.G., and K. Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J. Am. Stat. Assoc. 82:605-610.

Smith, A. B. 1988. Phylogenetic relationship, divergence times, and rates of molecular evolution for Camarodont sea urchins. Mol. Biol. Evol. 5(4):345-365.

Smith, A. B., and J. J. Savill. 2001. Bromidechinus, a new Ordovician echinozoan (Echinodermata), and its bearing on the early history of echinoids. Trans. R. Soc. Edinburgh Earth Sciences 92(02):137.

Smith, A. B., D. Pisani, J. A. Mackenzie-Dodds, B. Stockley, B. L. Webster, and D. T. J. Littlewood. 2006. Testing the Molecular Clock: Molecular and Paleontological Estimates of Divergence Times in the Echinoidea (Echinodermata). Mol. Biol. Evol. 23(10):1832-1851.

Stojković, B., A. Sayadi, M. Đorđević, J. Jović, U. Savković, and G. Arnqvist. 2016. Divergent evolution of life span associated with mitochondrial DNA evolution. Evolution. doi:10.1111/evo.13102

Thompson, J.D., D.G. Higgins, T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22):4673-4680.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3 - new capabilities and interfaces. Nucleic Acids Res. 40(15):e115.

Valverde, J. R., R. Marco, and R. Garesse. 1994. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. PNAS. 91(12):5368-5371.

Watts, S. A., J. B. McClintock, and J. M. Lawrence. 2013. Lytechinus. In: Lawrence JM, editor. Sea Urchins: Biology and Ecology. Oxford, UK: Elsevier. p. 475-490.

Wilm, A., D. C. Higgins, and C. Notredame. 2008. R-Coffee: a method of multiple alignment of non-coding RNA. Nucleic Acids Res. 36(9):e52.

Yang, Z. 2007. PAML 4: a program for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24(8):1586-1591.

Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22(12):2472-2479.

Chapter 5

Conclusion

The growing access to genomic resources from broad taxonomic sources will continue to define much of biological research to come. While it is clear that the interplay of theoretical expectations, practical implementation, and complex biological signals will continue to be tackled by the field, the value of rationalizing biological conclusions from focused datasets, and addressing focused questions with specific data, seems a worthy practice to which to endeavor. *Tripneustes* gonads are harvested across their distribution, including historically and contemporarily in Hawai'i, for consumption as well as for the decorative aquarium trade (Martins 2012). As a result of this harvesting steady stock declines have been reported for a number of years for both *T. gratilla* and *T. ventricosus* and there have been recent attempts at stock restoration efforts (Hughes et al. 1987, Scheibling & Mladenov 1987, Smith & Berkes 1991, Junio-Meñez et al. 2008). Furthermore the apparent recent success at transplanting *T. gratilla* from across the Hawai'ian islands to Kāne'ohe bay in removing invasive algal cover has led to a possible interest in the broader use of *Tripneustes* in invasive algae removal of threatened coral reefs (Vasserot 1992, Birkeland 2006, Stimson & Conklin 2008). Effective conservation efforts are uniquely dependent on correct phylogenetic assessment (Faith & Baker 2006). If stocks of *Tripneustes* are to be properly managed and even transplanted for conservation work then an accurate and up to date molecular resource is necessary to avoid unintentional dilution of

biodiversity as well as potentially unbalancing heretofore unrealized population structures.

5.1 References

Birkeland, C., Jayewardene, D. (2006) How many fish does it take to keep the alien algae out? NOAA Final Report 2004-2005

Faith, D.P., Baker, A.M. (2006) Phylogenetic Diversity (PD) and Biodiversity Conservation: Some Bioinformatics Challenges. Evol Bioinform Online 2:121-128

Hughes, T.P., Reed, D.C. , Boyle, M.-J. (1987) Herbivory on Coral Reefs: Community Structure Following Mass Mortalities of Sea Urchins. J Exp Mar Biol Ecol 113:39-59

Juinio-Meñez, M.A., Pastor, D., Bangi, H.G. (2008) Indications of Recruitment Enhancement in the Sea Urchin Tripneustes gratilla Due to Stock Restoration Efforts. ICRS 2:1024-1028

Martins, L., Souto, C., Magalhães, W.F., Alves, O.F.de S., Rosa, I.L., Sampaio, C.L.S. (2012) Echinoderm Harvesting in Todos-os-Santos Bay, Bahia State, Brazil: The Aquarium Trade. SCB 12(1)53-59

Scheibling, R.E., Mladenov, P.V. (1987) The decline of the sea urchin, *Tripneustes ventricosus*, fishery of Barbados: A survey of fishermen and consumers. Mar Fish Rev 49(3):62-69

Smith, A.H., Berkes, F. (1991) Solutions to the “Tragedy of the Commons": Sea-urchin management in St Lucia, West Indies. Environ Conserv 18(2):131-136

Stimson, J., Conklin, E. (2008) Potential Reversal of a Phase Shift: The Rapid Decrease in the Cover of the Invasive Green Macroalga Dictyosphaeria cavernosa Forsskøal on Coral Reefs in Kāne‘ohe Bay, Oahu, Hawai‘i. Coral Reefs 27:717-726

Vasserot, J. (1992) Prospects of Biological Control of Harmful Proliferations of Multicellular Algae: 2. Possibilities for Control and Utilization by Grazing Animals of Sargassum muticum. J Rec Oceanogr 16:77