Population Structure and Phylogeography of *Macrobrachium* Species on the Hawaiian Archipelago

A Major Qualifying Project Report

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

By

Kathleen Duffy

Tracy Rhode

Submitted on: April 28, 2022
Approved by: Prof. Lauren Mathews,
Prof. Michael Buckholt

Table of Contents

Abstract

Macrobrachium is a genus of freshwater prawns with species found around the world. Several *Macrobrachium* species are present on the Hawaiian archipelago, including the native *M. grandimanus* and the recently introduced *M. lar*. The mitochondrial COI gene was sequenced and analyzed from samples of the Hawaiian populations of these two species. Data analysis suggested significant differences between species' haplotype networks. Data for *M. lar* supported expectations for a recent founder population. Data for *M. grandimanus* suggested gene flow between islands, as well as a possible recent bottleneck event. Both species exhibited an overall lack of genetic diversity. However, due to the small sample size of this study, further research is required to confirm these conclusions.

Introduction

Investigations of population genetic structure in marine or aquatic organisms are frequently used to further understand ecological and evolutionary processes affecting these species. This research project focuses on the population genetics of two prawn species found in the Hawaiian archipelago within the contexts of their unique marine and brackish habitats on these islands. These focal species are *Macrobrachium grandimanus* and *Macrobrachium lar*. *M. grandimanus*, also known as the Hawaiian prawn or ʻŌpae ʻoehaʻa, is a naturally occurring Hawaiian species (Division of Aquatic Resources, 2019). *M. lar*, also known as the Tahitian prawn, is an invasive species that was artificially introduced for aquaculture purposes (Brasher et al., 2006). Like other *Macrobrachium* species, both *M. lar* and *M. grandimanus* exhibit amphidromous life cycles, spending their larval stage at sea before migrating back to brackish or freshwater as juveniles (McDowall, 2004).

Macrobrachium is a genus of freshwater prawns or shrimp (Bate, 1868). *Macrobrachium* species are native to all continents, excluding Europe (Holthuis & NG, 2009). Members of the genus can be categorized by the number of larval stages required for metamorphosis (Jalihal et al., 1993). Most species require 8 or more molts and thus can be categorized as having type I: (prolonged) larval development. Type I can be subdivided into two groups. Group A species have larvae that require 10 or more molts for metamorphosis. Both *M. grandimanus* and *M. lar* are members of this group. For species in this group, larval development is dependent on salinity, so gravid females will migrate downstream toward estuaries. Species belonging to this group exhibit sexual dimorphism; adult males display an enlarged second pair of pereiopods (Jalihal et al., 1993).

M. grandimanus is considered a native species in Hawaii, though its exact origins on the archipelago are unknown. Adults grow to approximately 8cm in length and feed on scavenged plant and animal material (Division of Aquatic Resources, 2019). Previous studies have noted at least two separate major populations of *M. grandimanus*: one occurring on the Ryukyu island chain on the southern end of Japan, and the other occurring in the Hawaiian Islands. These populations are genetically distinct (Jose & Harikrishnan, 2018). Despite being considered native, *M. grandimanus* was only found to be present in 35% of surveyed streams on Oahu, Kauai, and Hawaii, as of a study in 2001 (Brasher et al., 2006). *M. lar* was introduced to Hawaii from the Marquesas Islands in the 1950s (Benson, 2015). Previous phylogenetic data lists native populations of *M. lar* appearing in Ryukyu, in common with *M. grandimanus*, as well as Taiwan and the Philippines (Jose & Harikrishnan, 2018). Adults grow to approximately 18cm in length. *M. lar* was present in 71% of the surveyed streams on Oahu, Kauai, and Hawaii, as of 2001 (Brasher et al., 2006). Thus, despite being recently introduced, *M. lar* had dispersed to a majority of these streams throughout the islands. As of 2021, *M. lar* has been recorded on the islands of Hawaii, Maui, Kauai, and Oahu (Benson, 2015).

Amphidromy plays an important role in the brackish and freshwater habitats of tropical or subtropical islands like Hawaii. It also has significant implications for population genetic relationships. Species that exhibit this adaptive life cycle appear predominantly in tropical or subtropical islands, where they can make up most or all of freshwater fauna, in contrast to their sparsity in continental habitats. The presence of their marine larval stage may explain how this fauna that can live in freshwater came to inhabit such remote islands (McDowall, 2004). In Hawaii, all native fish species in inland streams consist of amphidromous gobies, often studied for their unique adaptations. Some species of these fish, such as *Lentipes concolor*, have been observed climbing rock faces behind waterfalls to reach ideal habitat in maturity (McDowall, 2007). Population genetics studies conducted on these gobies have noted a significant lack of phylogeographic structure within the species, possibly due to gene flow occurring in the marine larval stage (Chubb et al., 1998). Similar results have been observed in *M. grandimanus* as well, which in one small study shared COI sequence haplotypes between Kauai and Hawaii, again implying gene flow of amphidromous species via the ocean between islands (Bebler & Foltz, 2004). However, *M. grandimanus* remains a species with very little prior research available on its phylogeographic history. This project aims to fill in some of the wide gap in knowledge on this species.

This project involves the gathering and analyzing of genetic data from *M. grandimanus* and *M. lar* to produce phylogenetic information. The data analysis of this project aims to establish degrees of genetic relationship between groups of mature specimens belonging to each species located in different bodies of freshwater across Hawaii. This research also aims to investigate the impact of gene flow between these groups, which may occur during the marine larval stage of their amphidromous life cycles. These genetic relationships are inferred by the comparison of mitochondrial DNA sequences for COI, or cytochrome oxidase 1. COI plays a critical role in the electron transport phase of respiration, and consequently, it is one of the most conserved mitochondrial protein-coding genes (Strüder-Kypke & Lynn, 2010).

Mitochondrial DNA is useful in phylogeographic studies for various reasons. Many copies of mitochondrial DNA are present in cells, thus it is easily amplified. Mitochondrial DNA can also be amplified without purification, which allows for nondestructive sampling when needed. Additionally, mitochondrial DNA mutates at a higher rate than nuclear DNA, allowing it to better distinguish individuals geographically. Mitochondrial DNA is maternally inherited and as such does not recombine with paternal DNA. This lack of significant recombination allows genetic relationships to be inferred among haplotypes. However, because mtDNA is maternally inherited, it is not useful in studies on species that disperse differently based on sex (Rowe et al., 2017). Furthermore, mtDNA tends to have higher sensitivity to diversity loss in events such as bottlenecks, a significant potential drawback to consider in the context of this study (Freeland, 2005). Through sequencing and bioinformatic analysis of the mitochondrial COI gene in our target species, this research can divulge new information about the phylogeographic patterns and history of *M. grandimanus* and its recently introduced relative, *M. lar,* on the Hawaiian archipelago.

Methods

Collection of Samples:

M. grandimanus and *M. lar* specimens were collected at several numbered sites on the island of Hawaii, in the winter of 2020, by L. Mathews. Kauai and Oahu samples were collected by L. Mathews in the winter of 2022. These samples were shipped to Worcester Polytechnic Institute's facilities, preserved in containers with 90% isopropanol (Hawaii Island specimens) or DNA/RNA Shield (Zymo Research; Oahu and Kauai Island specimens). Samples were kept at -20ºC until DNA extraction. Specimens included in our analysis are listed in Table 1.

Table 1. Specimens included in the study by location

Island	Site Code	Site Name	Species Collected	Number of Specimens
Hawaii	H5	Aleamai Stream	*M. lar*	8
Hawaii	H9	Hawaii Ocean Science & Tech anchialine ponds	*M. grandimanus*	5
Hawaii	H11	Kiholo Bay anchialine ponds	*M. lar*	9
Hawaii	H17	Wailoa River	*M. grandimanus*	9
Hawaii	H19	Aamakao Gulch	*M. lar*	2
Kauai	KI2	Wailua River	*M. grandimanus*	19
Oahu	OI1	Paukauila Watershed	*M. grandimanus*	5
Oahu	OI2	Waihe'e Stream	*M. grandimanus*	7

Total of 19 *M. lar* samples and 45 *M. grandimanus* samples.

DNA Extraction:

For each sample in a set, a master mix of 120µL of 0.5M EDTA solution (pH 8.0) and 500µL of Nuclei Lysis Solution (Promega Wizard Genomic DNA Purification Kit) were added to a centrifuge tube and chilled on ice. A piece of muscle tissue from each specimen in the set was diced with a razor and placed in separate labeled Eppendorf tubes. A volume of 600µL master mix and 17.5µL of 20mg/ml proteinase K were added to each of these tubes, which were then incubated at 55 degrees Celsius overnight. After incubation, 200µL of Protein Precipitation Solution was added to each tube, which was then vortexed briefly and chilled for 5 minutes on ice. The tubes were centrifuged for 4 minutes at 15,000 × g to form a protein pellet.

The supernatant was transferred to new labeled tubes each containing 600µL of 100% isopropanol, then gently mixed. These tubes were centrifuged for 1 minute at 15,000 × g. The supernatant was then decanted. 600µL of 70% ethanol was added to each tube. The tubes were gently inverted before careful removal of this ethanol. A volume of 100µL DNA Rehydration Solution was added to tubes after drying to rehydrate the DNA in incubation at 65°C for 1 hour. The tubes were then stored at -20ºC. Agarose gels (1%) for each set of prepared DNA extractions were prepared with SYBR green added for staining. After setting, each well was filled with 1.5µL purple gel loading dye mixed with 1µL of assigned DNA extraction. Gels were run for 45 minutes at 117V, then photographed with a gel documentation system to assess quality and concentration.

PCR:

Sets of up to 9 PCR tubes were placed in a tray on ice for each set of reactions. Separate master mixes were assembled when PCRs included specimens of both *M. grandimanus* and *M. lar* (Table 2). Master mixes were assembled for 21µL reactions including 10µL of OneTaq 2X Master Mix with Standard Buffer, 40 ng each of forward and reverse primers (10µmol/L, 0.4µL each), and 8.2µL deionized water per PCR reaction. A volume of 20µL master mix and 1µL of a DNA extraction solution were added to each PCR tube. PCR tubes were placed in a Bio-Rad MyCycler™ thermal cycler, and the cycler conditions consisted of an initial denaturation step of 94ºC for 1 min, followed by 45 cycles of 94ºC for 30 seconds, 56ºC for 30 seconds, and 68ºC for 1 minute, followed by a final 10-minute stage at 68ºC. After cycling was completed, PCR products were stored at 4ºC until gel analysis. Agarose gels (1%) for each set of PCR products were prepared with SYBR green added for staining. After setting, each well was filled with 2µL purple gel loading dye mixed with 2µL of an assigned PCR product tube. Gels were run for 45 minutes at 117V, then photographed with a UV lightbox machine to assess quality and concentration. After gel analysis, PCR products were stored at -20ºC.

Table 2. *M. grandimanus* and *M. lar* COI Primers

Name	Sequence	Species
MgrandCOIF	CTATTCTGYTTCTTCGGCCACCC	*M. grandimanus*
MgrandCOIR2	AGTCATTTKGGGTTTAGTGA	*M. grandimanus*
MlarCOIF2	CTACAGTAATCAACATACGATC	*M. lar*
MlarCOIR	TGCCGGCGAAGATTCCGAATAC	*M. lar*

Primers were custom designed by L. Mathews based on alignments of existing *Macrobrachium* sequences from GenBank.

Preparation of PCR products for automated sequencing:

For reactions that yielded a single, clear band of about the expected size (~550 bp), the PCR product was cleaned by mixing 15μL of product with 6μL of exoSAP, which comprised 2μL of exonuclease 1 and 4μL of shrimp alkaline phosphatase (New England Biolabs). The product was incubated at 37°C for 15 minutes, then heated to 80°C for 15 minutes to inactivate enzymes. The purified PCR products were sent to Eton Bioscience for direct sequencing of both strands.

Preparation of sequence alignments:

Nucleotide sequences for the forward strands were aligned with the reverse complement of their reverse strand using BioEdit version 7.2.5 with the ClustalW multiple alignment function (Hall, 1999). Consensus sequences were confirmed with chromatograms viewed in FinchTV software version 1.4.0 (Geospiza, Inc.). Consensus sequences within each species were aligned with each other.

Data Analysis:

Alignment files saved from BioEdit were converted into nexus-type files for analysis (Hall, 1999). From alignment files of the consensus sequences of *M. grandimanus* and *M. lar* samples, haplotype networks were created using the PopART version 1.7 freeware software. Each alignment was imported, and the program was then run to create a minimum spanning network (Leigh & Bryant, 2015).

Haplotypes groupings were confirmed by manual analysis of sequence alignments in BioEdit, in which single nucleotide polymorphisms (SNPs) could be identified and categorized. The SNPs identified are reflected in the haplotype networks created in PopART in which each hatch between the haplotype groups, represented by colored circles, represents a single SNP setting two groups apart. The alignments created for the obtained *M. grandimanus* and *M. lar* sequences were further analyzed in DnaSP v.5 software to obtain a variety of relevant statistics (Librado & Rosaz, 2009). PopART v.1.7 software was then also used to confirm statistical values for both species data sets, including nucleotide diversity and Tajima's D value.

A separate alignment was created in BioEdit with each unique COI haplotype in our collection of *M. grandimanus* and *M. lar* sequences, in addition to COI sequences of these and other *Macrobrachium* species sourced from the GenBank database. GenBank sequences were chosen by conducting BLAST searches from consensus nucleotide sequences of *M. grandimanus* and *M. lar* from alignments and selecting sequences from the results. A variety of different *Macrobrachium* species COI sequences were chosen from different studies, varying in their relatedness to *M. grandimanus* and *M. lar*, and from a variety of different locations. A GenBank sequence of *Alpheus angulosus*, a marine shrimp species from a genus related to *Macrobrachium*, was included as well to serve as an outgroup on the final phylogenetic tree. Collection locations for sequences were recorded as given on GenBank files.

The alignment of *M. grandimanus* and *M. lar* haplotypes with GenBank sequences was exported from BioEdit and then converted to nexus file format before formatting in BEAUti version 2.6.6 to an XML file for use in BEAST 2 (Bouckaert et al., 2014). BEAUti configuration included setting of the HKY nucleotide substitution model, a strict clock, and a chain length of 100,000,000 and burn-in of 100000. Additionally, a monophyletic “macrobrachium” prior was set containing all of the *Macrobrachium* sequences used. The output of the BEAST 2 run of this alignment was used to create a final phylogenetic tree displayed in the FigTree program version 1.4.4. Posterior values for each node were calculated by FigTree 1.4.4 and recorded.

Results

Multiple haplotypes were identified for the *M. grandimanus* and *M. lar* sequences based on identified SNPs in the sequence alignments. In a 400 bp length alignment of all *M. grandimanus* sequences, 6 SNPs were identified. The PopART v 1.7 program identified these as 6 segregating sites, of which 2 were parsimony-informative, where multiple sequences shared an SNP. This data identified a total of 7 unique haplotypes for the *M. grandimanus* alignment. In a 435 bp length alignment of all *M. lar* sequences used, 2 SNPs were identified. The PopART program identified these as 2 segregating sites, of which both were parsimony-informative. The haplotypes identified were visualized by minimum spanning haplotype network analysis created in the PopART software.

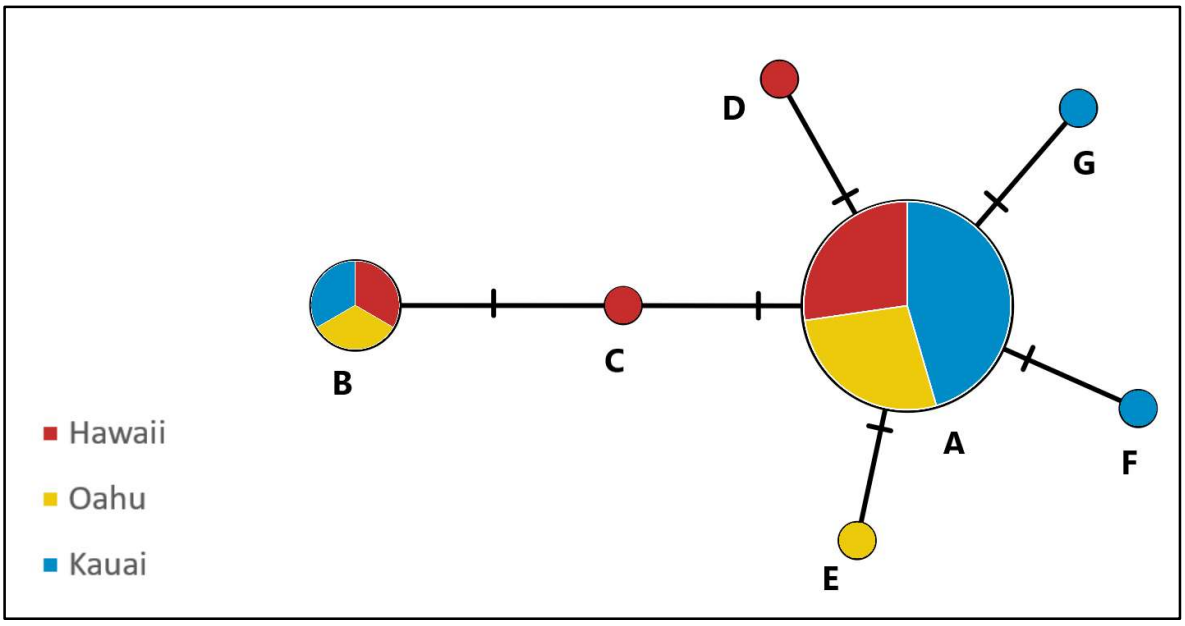

Figure 1. Minimum spanning haplotype network for *M. grandimanus*, created in PopART v. 1.7. Haplotype groups labeled A-G with colors indicating island source. Hatch marks represent SNPs between groups.

The majority of *M. grandimanus* sequences belonged to haplotype A (Figure 1). haplotypes C, D, E, F, and G differed from haplotype A by one SNP. Haplotype B differed from haplotype A by two SNPs yet only differed from haplotype C by one SNP. Haplotypes C, D, E, F, and G only consisted of one individual per haplotype.

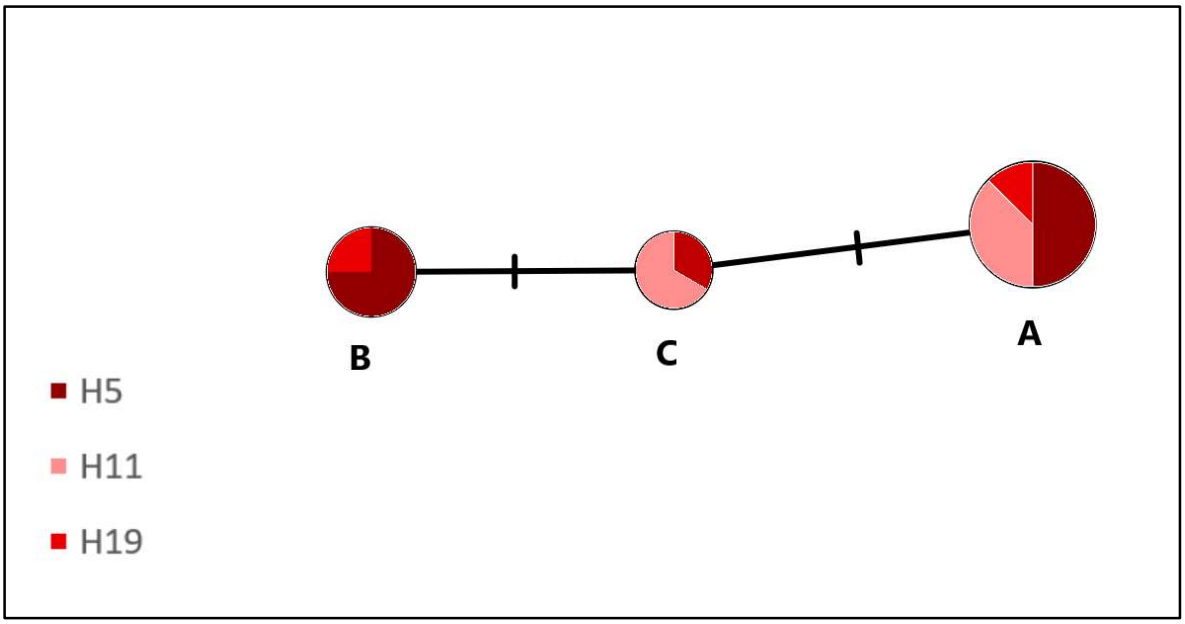

Figure 2. Minimum spanning haplotype network for *M. lar*, created in PopART v. 1.7. Haplotype groups are labeled A-C with colors indicating the sample site on Hawaii island. Hatch marks represent SNPs between groups.

The majority of *M. lar* sequences belonged to haplotype A (Figure 2). Haplotype B differed from haplotype A by two SNPs and differed from haplotype C by one SNP. Haplotype C differed from haplotype A by only one SNP.

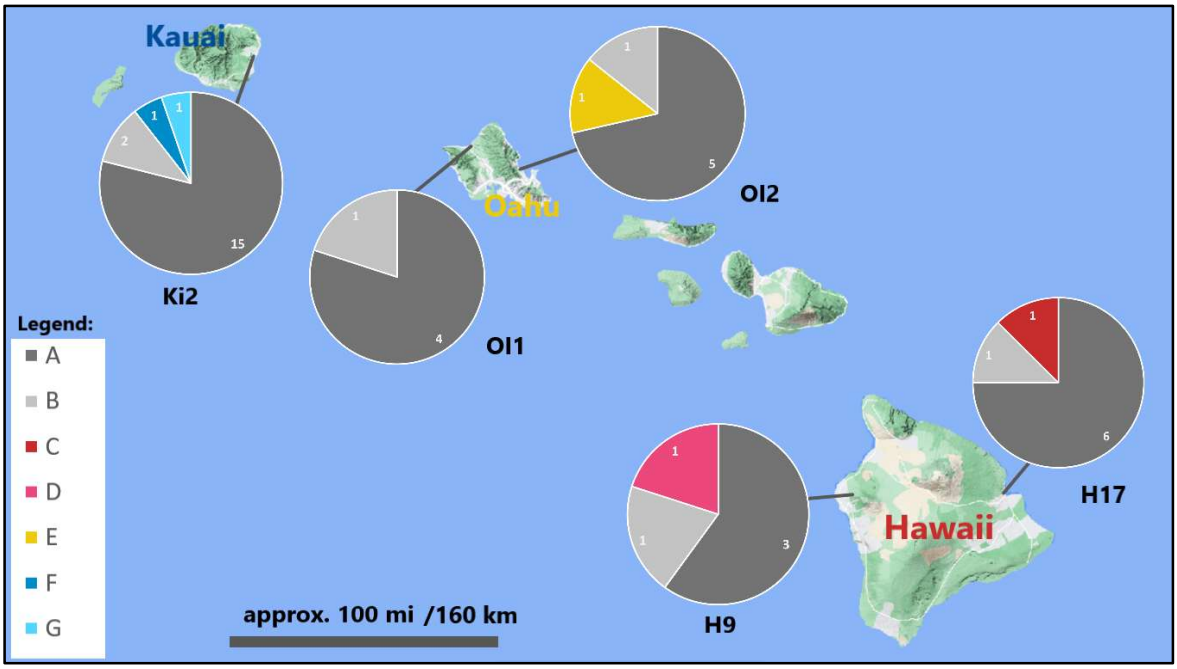

Figure 3. Haplotype geographical distribution map for *M. grandimanus*, over terrain map of Hawaii from Google Maps. Haplotype numbers correspond with those designated in Figure 1.

For *M. grandimanus*, haplotypes A and B were present in all sampled sites (Figure 3). The majority of samples within each site were haplotype A. All haplotypes unique to sites(C-G) consist of only one sampled individual. Kauai was the only site to have two unique haplotypes.

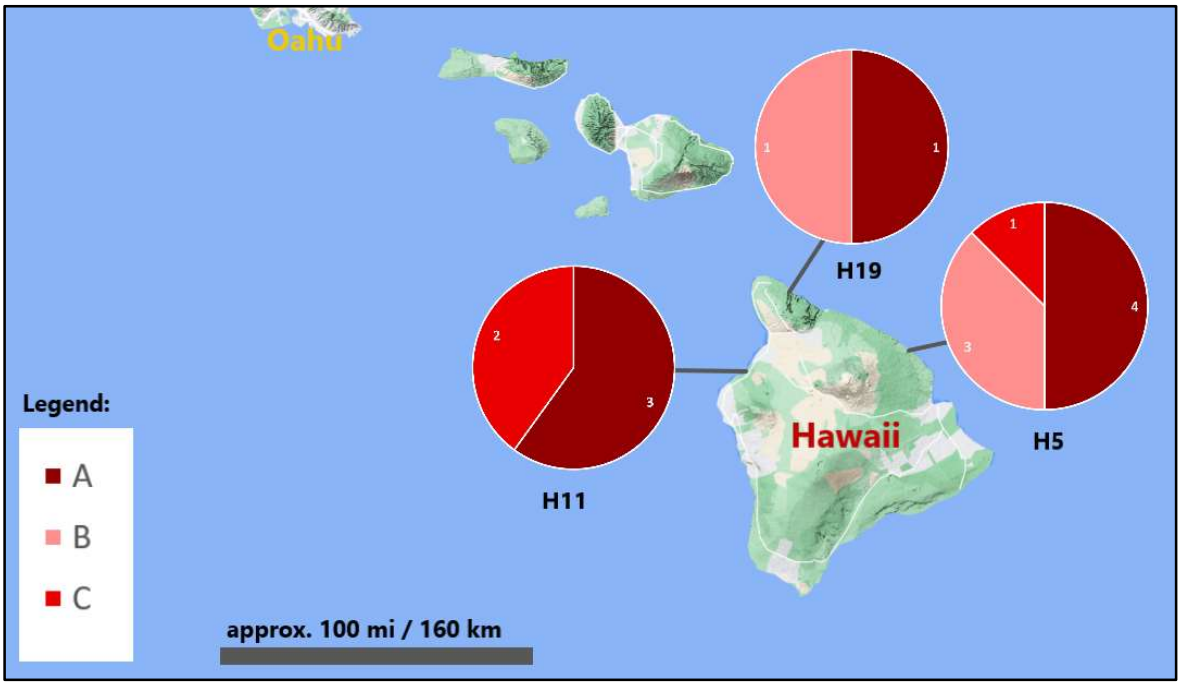

Figure 4. Haplotype geographical distribution map for *M. lar*, over terrain map of Hawaii from Google Maps. Haplotype letters correspond with those designated in Figure 2.

For *M. lar*, haplotype A was present in all sampled sites (Figure 4). The majority of samples within each site were haplotype A.

The same sequence alignments used to create the haplotype networks seen in Figure 1 and Figure 2 were opened in the DnaSP v.5 software for statistical analysis, thus maintaining values such as sequence lengths and the number of sequences in each alignment (Librado & Rosaz, 2009). The analyses “Tajima’s Test” and “Fu and Li’s (and other) Tests” were run within the DnaSP program for each alignment. Relevant statistical values obtained from these analyses were recorded in Table 3. Nucleotide diversity and Tajima’s D values were confirmed by statistical analysis in PopART v.1.7, which returned the same values for these tests as DnaSP v.5 (Leigh & Bryant, 2015).

Table 3. Statistical values for each species alignment as calculated by DnaSP v.5. The statistics in this table represent alignments of 44 sequences of 400 bp length for *M. grandimanus* and 15 sequences of 435 bp length for *M. lar.*

Species	**Nucleotide diversity (π)**	**Tajima's D**	**Fu's Fs**
M. grandimanus	0.00174	-1.29388 (P > 0.10)	-3.501
M. lar	0.00219	1.44289 (P > 0.10)	0.823

Table 4. Statistical values for each species alignment as calculated by DnaSP v.5, continued. The statistics in this table represent alignments of 44 sequences of 400 bp length for *M. grandimanus* and 15 sequences of 435 bp length for *M. lar.*

Species	**Number of segregating sites**	**Average nucleotide differences (k)**	**Haplotype diversity (Hd)**
M. grandimanus	6	0.69662	0.426
M. lar	2	0.95238	0.648

The values of nucleotide diversity, Tajima's D, average nucleotide differences, and the number of segregating sites were obtained from the "Tajima's Test" analysis in the DnaSP program. Nucleotide diversity (pi) gives the average number of nucleotide differences per site between two sequences. Tajima's D statistic tests the selective neutrality hypothesis of mutations. DnaSP provides the D value along with a significance test of this value (any P > 0.05 is not statistically significant). Therefore, the Tajima's D statistical values obtained for this data are not significant, though the differences in the values between the two species may still hold relevance. The D value calculation depends on other values collected from DnaSP, including the average number of nucleotide differences between sequence pairs, and the number of segregating sites in the sequences, also representing the number of mutations (Tajima, 1989).

Haplotype diversity values and Fu's Fs statistic were obtained from the "Fu and Li's (and others) Tests" analysis in the DnaSP program. Fu's Fs value provides another test of mutation neutrality, based on haplotype frequencies. In addition to Tajima's D, this statistic can be used to detect population expansions or bottlenecks (Fu, 1997).The haplotype diversity values are given by Hd, alongside these values' sampling variance (Rozas et al., 2016).

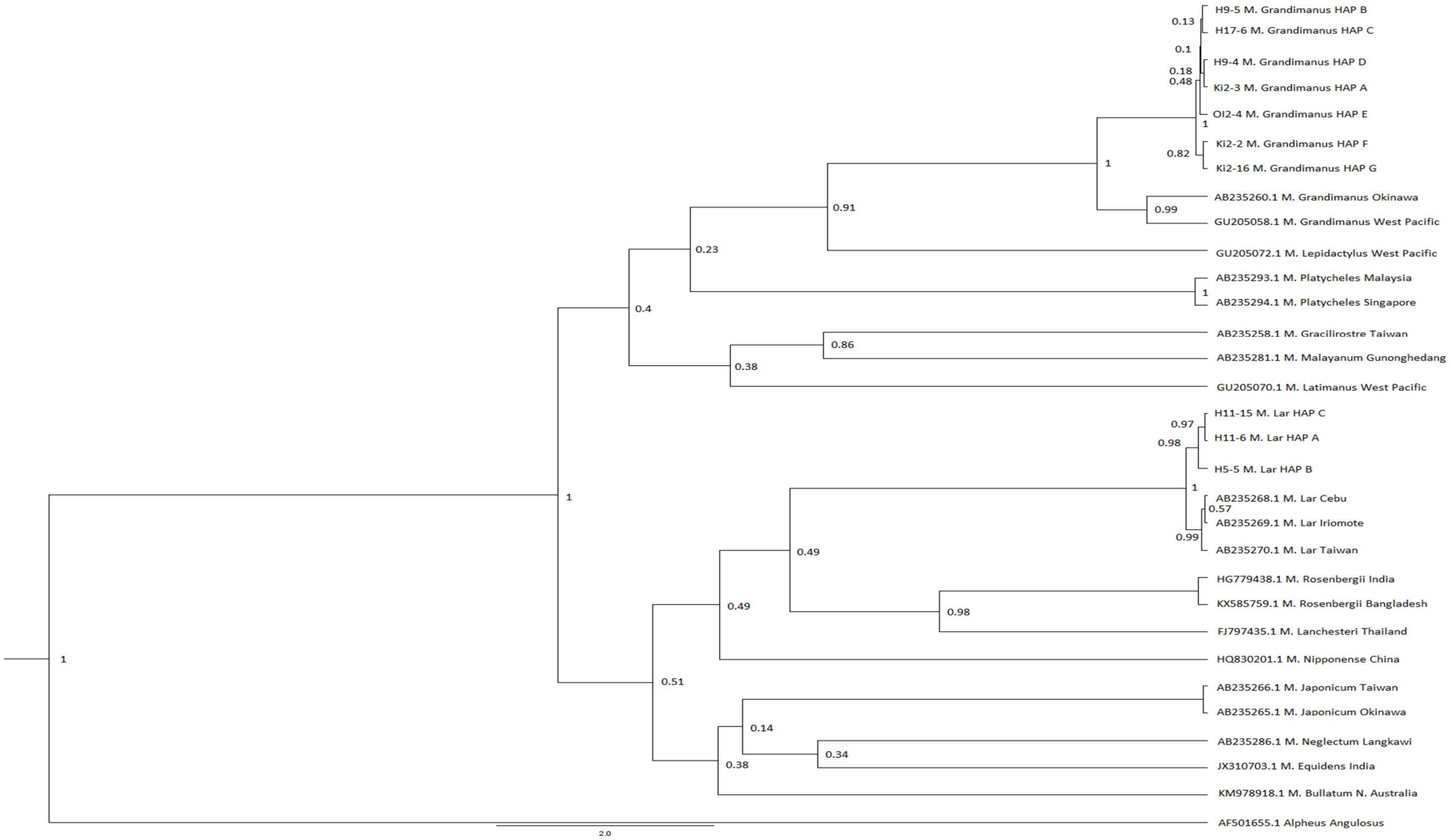

Figure 5. *Macrobrachium* phylogenetic tree created with BEAST 2 software package. Node labels represent posterior probability values. Tips are labeled by haplotype and sample number from alignment data or with GenBank accession number and species' sample location for GenBank-sourced sequences

The *Macrobrachium* genus phylogenetic tree (Figure 5) separated *Macrobrachium* species into two distinct main clades, separated at a node of posterior value 1. All *M. grandimanu*s sequences, both the Hawaii samples and those from GenBank, were grouped into one clade and all the *M. lar* sequences were grouped into another clade.

M. grandimanus sequences from Hawaii were clustered with *M. grandimanus* Genbank sequences from Okinawa (Ryukyu islands, Japan) and another sequence only labeled in the GenBank entry as from the West Pacific region. *M. grandimanus* sequences were separated into two distinct clades, separated at a node of posterior value 1. The Hawaiian population formed one clade, and another clade included the Okinawa sequence, confirmed to originate from the Japanese population of *M. grandimanus* (Liu et al., 2006). *M. lar* sequences from Hawaii were clustered with *M. lar* Genbank sequences from the islands of Cebu island (Philippines), Iriomote island (South Japan), and Taiwan. *M. lar* sequences were separated into two distinct clades as well, separated at a node of posterior value 1. The Hawaiian population formed one clade, while all other *M. lar* sequences formed another. The non-Hawaiian sequences of *M. lar* sourced from GenBank originated from one study (Liu et al., 2006), of which the sequence originating from Taiwan was most similar to the Hawaiian population. The other sequences, from the Philippines and Japan, were notably identical. Within each of the two main clades of the phylogenetic tree, the *M. grandimanus* sequences were clustered with *M. lepidactylus* and *M. platycheles* sequences, while the *M. lar* sequences were clustered with *M. rosenbergii* and *M. lanchesteri* sequences.

GenBank accession numbers and sampling locations from GenBank files were recorded to display on the phylogenetic tree for each sequence used from the database. Specificity of sampling locations for each sequence varied by source, however. Some sequences originated from studies that listed specific islands or regions as sampling sites, while others originated from studies that only listed sampling sites as a country or a large geographic region, such as the “West Pacific” region referenced for some sequences.

GenBank data was severely limited for the target species. No usable sequences from Hawaii could be found for either species for comparison. Very few usable COI sequences for *M. grandimanus* were available on the GenBank database, limiting analysis of this species. Results for the relationship of the Hawaiian population of *M. lar* with populations from other locations was limited as well. Usable sequences from GenBank of *M. lar* COI were only found from one study (Liu et al., 2006). Furthermore, no COI sequences for *M. lar* were available on GenBank originating from French Polynesia / the French Marquesas islands, where sources claim that the Hawaiian *M. lar* population was artificially introduced from (Benson, 2015). Thus, this history could not be investigated by COI sequence analysis.

Discussion

M. grandimanus and *M. lar* exhibited notably different patterns of haplotype networks from the analysis of their respective sequence alignments. *M. lar* exhibited three haplotype groups, each with multiple members (Figure 2). *M. grandimanus* exhibited a 'star-shaped' haplotype network, with single-member haplotypes branching off around the main haplotype group, haplotype A, consisting of a majority of this species' samples (Figure 1). The star-shape of the *M. grandimanus* haplotype network may indicate a population bottleneck event, potentially followed by sudden expansion (Hwang et al., 2018)(Johnson et al., 2007). Haplotype A, at the center of the star-shaped pattern, contains a significant majority of the *M. grandimanus* samples analyzed, including 33 of the 44 total or 75% (Figure 1). Such a high-frequency haplotype in the data set, especially in an interior position in the network, is more likely to represent an older haplotype relative to the other haplotypes present (Freeland, 2005). The presence of haplotype A on the archipelago therefore most likely predates any recent bottleneck events that may have occurred. In contrast, the peripheral haplotypes D-G, with only one representative each, may represent newer haplotypes, potentially diverging from haplotype A during expansion after a bottleneck event (Freeland, 2005).

The presence of *M. grandimanus* haplotypes A and B in all sampled sites, on the islands of Hawaii, Oahu, and Kauai, certainly indicates gene flow between islands (Figure 3). Considering the amphidromous nature of this species, it is likely that *M. grandimanus* larvae of haplotype A and B spread between islands by sea during this marine stage of the life cycle. *M. grandimanus* haplotype B differs from haplotype A by two SNPs, as demonstrated in the species haplotype network (Figure 1), in contrast to the remaining haplotypes, which each differ from haplotype A by one unique SNP. Haplotypes C-G form the star-shape network, radiating around haplotype A on the network map (Figure 1), from which each differs by one unique SNP. These haplotypes may represent an expansion from the main population, haplotype A.

Comparisons of the samples of both populations on the island of Hawaii may supplement understanding of the differences between them, as *M. lar* was only sampled on the island of Hawaii, and sample sizes of both populations from this island are similar. On this island, 69% of *M. grandimanus* samples belonged to haplotype A (Figure 3), while 53% of *M. lar* samples belonged to haplotype A (Figure 4). Additionally, 4 haplotypes of *M. grandimanus* were present on the island of Hawaii (Figure 3), while 3 haplotypes of *M. lar* were present (Figure 4).

Common trends appear between the analyses of the island of Hawaii populations and the entire sample size, which includes *M. grandimanus* samples from other islands. *M. grandimanus* shows a greater number of haplotypes, yet a larger portion belonging to a main haplotype, than *M. lar* in both cases. Statistical analysis of the *M. grandimanus* sequence alignment produced lower values for nucleotide diversity and haplotype diversity (Table 3, Table 4) than *M. lar* as well, although *M. grandimanus* had a higher number of segregating sites and distinct haplotypes. This difference may reflect the significantly higher portion of *M. grandimanus* that belonged to the majority haplotype, haplotype A, compared to the haplotype A of *M. lar*, indicating lower diversity overall (Freeland, 2005)(Rozas et al., 2016).

The statistical analyses completed for each species' alignments can provide more insight into the potential phylogeographic history of these species on Hawaii. In the data set analyzed, the *M. grandimanus* network's Tajima's D value was a negative value, while the *M. lar* network's D value was a positive value of somewhat higher magnitude. This significant difference in D values between the species could correspond to differences in their phylogeographic histories, despite the P-value of each individual species' D values implying a lack of significance ($P > 0.10$). Generally, a positive Tajima's D value indicates a lack of rare alleles, while a negative value indicates a greater abundance of rare alleles. In the data set, this could correspond to the greater number of unique haplotypes, many with only one representative sample, in the *M. grandimanus* alignment as compared to that of *M. lar*, which lacked individuals with unique haplotypes. The stronger positive value of the *M. lar* network's D value could also indicate a contraction of some kind of the population (Berwick, 2005)(Tajima, 1989). In consideration of the species' recent history being introduced artificially to the archipelago, this could indicate a founder's effect.

These inferences are supported by additional insight gained from the Fu's Fs statistic. The *M. grandimanus* alignment produced a strong negative Fs value, while the *M. lar* alignment produced a weaker, positive value. The differences in sign and magnitude once again allude to significant differences between the two species' populations. Positive Fs values, as seen in *M. lar*, generally imply a deficiency of allele diversity, possibly due to a bottleneck event such as a founder's effect. Negative Fs values, as seen in *M. grandimanus*, generally imply an excess of allele diversity, and possible recent population expansion. Though Fs value analysis is similar to that of Tajima's D statistic, simulations have shown that Fs can be more sensitive to population expansion (Rosaz et al., 2016).

The combination of the results of both statistical tests for the two species studied provides support for plausible recent trends in their phylogeographic histories.
The overall statistical results indicate that *M. lar* exhibits a lack of allele diversity due to a recent bottleneck event, most likely a founder's effect from their recent artificial introduction to Hawaii. The results also suggest that *M. grandimanus* may have recently undergone population expansion after a bottleneck event, demonstrated by a small number of rare haplotypes alongside more frequent, potentially older haplotypes. Statistical analysis returned relatively low values for nucleotide and haplotype diversity for both species, which may indicate bottleneck events in their recent histories resulting in this overall lack of genetic diversity observed. The type of bottleneck event that *M. grandimanus* may have undergone is unclear, however.

Bayesian statistical analyses completed through the construction of a phylogenetic tree with other *Macrobrachium* species were done to gather more information on the history of the target species in Hawaii (Figure 5). *M. grandimanus* sequences formed two distinct clades, separating the Hawaiian sequences from sequences including a sample from Okinawa, thus supporting a genetic divide between the Hawaiian and Japanese populations of the species. *M. lar* sequences similarly formed two clades separating the Hawaiian population from other *M. lar* sequences. The *M. lar* sequences from Hawaii were most similar to the GenBank *M. lar* sequence from Taiwan, potentially indicating a more recent common ancestor between the Taiwanese and Hawaiian populations of the species. More data is needed however, to determine the relationships of the Hawaiian populations of these species to other populations, due to the small number of available sequences on GenBank.

The Bayesian phylogenetic tree analyses may also indicate the relatedness of identified haplotypes within the Hawaiian populations of either species. *M. grandimanus* haplotypes are divided into two distinct clades. One clade contains haplotypes A, B, C, D and E, while the other contains haplotype F and G, both notably single-member haplotypes from the island of Kauai. *M. lar* haplotypes are similarly divided into two distinct clades. One clade contains haplotypes A and C, while the other contains haplotype B. Haplotypes on the same clades may have more recent common ancestors to each other than to the other haplotypes. Additional trials of Bayesian analysis may be needed to confirm these inferences, however.

The cluster of *M. grandimanus* sequences with *M. lepidactylus* and *M. platycheles* sequences suggests that *M. lepidactylus* and *M. platycheles* are most likely close relatives of *M. grandimanus* within the genus. *M. lepidactylus* is native to Reunion Island, which is a few hundred miles east of Madagascar (Hilgendorf, 1879). The vast distance between the ranges of these species notably contrasts with their genetic similarity. The cluster of *M. lar* sequences with *M. rosenbergii* and *M. lanchesteri* sequences suggest that *M. rosenbergii* and *M. lanchesteri* are likely close relatives to *M. lar* within the genus. These species' genetic similarities are consistent with their proximal ranges (Benson, 2015; Smithsonian Environmental Research Center, n.d.; Chong & Khoo, 1988). Prior research, in addition to the BEAST phylogeny created in this study, supports that *M. lar* and *M. grandimanus* most likely do not share common history on the Hawaiian archipelago, but rather that these populations derive from separate ancestral sources in the Indo-Pacific region.

The sample sizes used for analysis in this report are relatively small, as previously noted, compared to some other phylogeographic studies referenced. Additionally, only one gene was analyzed, COI, from the mitochondrial genome. The inferences made in the discussion, therefore, primarily serve as probable, though not definitive, conclusions supported by the limited evidence provided here. Additional sampling of *M. grandimanus* across the archipelago would further clarify the haplotype structure of the species, and the full extent of the rare haplotypes present in this species. A sampling of *M. lar* on additional islands could enable a comparison of this species' inter-island dispersal to that of *M. grandimanus*. Additional sampling of *M. lar* in Hawaii and other locations across the Pacific region, including the Marquesas islands, could confirm the origins of the Hawaiian population of this species, a task unable to be

completed in this study due to lack of available data. More data may also be acquired by sequencing of other genes from the samples. This would allow for more genetic diversity analysis, as well as comparison to a greater dataset from GenBank, which may have *Macrobrachium* sequence data for other genes.

Understanding the relationships between native species, often threatened on the Hawaiian archipelago, and their invasive counterparts can be paramount to proper conservation efforts to preserve native biodiversity. Nearly 40% of native bird and plant species have been lost on these islands in the past 200 years, along with many others (Freeland, 2005). This study aims to uncover information that could contribute to the building of datasets for future conservation efforts. Research to understand the dispersal and diversity of native and invasive species informs the management of these species to promote the conservation of biodiversity. Thus, the applied science of molecular ecology and phylogenetics plays a crucial role in conservation biology (Freeland, 2005). The history of *M. grandimanus* on the Hawaiian archipelago is yet to be fully understood. The results of this study may be informative, however, the gap of knowledge on this species is still too large to conclude its origins. The data gathered in this project contributes to crucial knowledge, knowledge that is currently severely lacking, for the understanding of *Macrobrachium* species and all other amphidromous organisms in Hawaii, and how they disperse and interact with their environment. If these species can be understood, conservation efforts may be better equipped to preserve the role of amphidromous species in these island habitats and others like it around the world.

Acknowledgments

The authors would like to acknowledge that in addition to sequences obtained solely by the project team, raw sequence data from previous work from the laboratory of Professor Lauren Mathews was made available for the analysis done in this project

The authors would also like to thank Worcester Polytechnic Institute faculty and professors: Mihail Bocka, Michael Buckholt, Louis Roberts, and JoAnn Whitefleet-Smith for their laboratory assistance.

Authorship

All parts of this report were completed in cooperation of both student authors, Kathleen Duffy and Tracy Rhode.

References

Bate, C. S.. (1868). The Proceedings of the Scientific Meetings of the Zoological Society of London. Zoological Society of London. https://books.google.com/books?id=1vAKAAAAIAAJ

Bebler, M. H., & Foltz, D. W. (2004). Genetic Diversity in Hawaiian Stream Macroinvertebrates. *Micronesia*, *37*, 119-128.

Benson, A. J. (2015). Tahitian prawn (*Macrobrachium lar*) - species profile. Retrieved from https://nas.er.usgs.gov/queries/factsheet.aspx?SpeciesID=1201.

Berwick, R. (2005). Calculating Tajima's D. HST.508: Quantitative Genomics. Harvard-MIT Division of Health Sciences and Technology.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M., Rambaut, A., & Drummond, A. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1003537

Brasher, A. M., Luton, C. D., Goodbred, S. L., & Wolff, R. H. (2006). Invasion patterns along elevation and urbanization gradients in Hawaiian streams. *Transactions of the American Fisheries Society*, *135*(4), 1109–1129. https://doi.org/10.1577/t05-083.1

Chubb, A. L., Zink, R. M., & Fitzsimons, J. M. (1998). Patterns of mtDNA variation in Hawaiian freshwater fishes: The phylogeographic consequences of amphidromy. *Journal of Heredity*, *89*(1), 8-16. doi:10.1093/jhered/89.1.8

Chong, S. S. C., & Khoo, H. W. (1988). The Identity of *Macrobrachium lanchesteri* (De Man, 1911) (Decapoda, Palaemonidae) from Peninsular Malaysia and Singapore, and a Description of Its First Zoea. *Crustaceana*, *54*(2), 196–206. http://www.jstor.org/stable/20104352

Freeland, J. R. (2005). *Molecular Ecology*. Wiley.

Fu, Y. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*.

Hall, T. A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*.

Hilgendorf, F. (1879). Die von Hrn. W. Peters in Moçambique gesammelten Crustaceen. Monatsbericht der königlich preussischen Akademie der Wissenschaften zu Berlin, Sitzung der Physikalisch-mathematische Klasse. 1878: 782-851, pls. 1-4 [imprint 1878]., available online at https://www.biodiversitylibrary.org/page/36097499

Holthuis, L. B., & Ng, P. K. (2009). Nomenclature and taxonomy. Freshwater Prawns, 12–17. https://doi.org/10.1002/9781444314649.ch2

Hwang, J. Y., & Cho, G. J. (2018). Identification of novel haplotypes and interpretation of gene flow of mitochondrial DNA control region of Eurasian otter (*Lutra lutra*) for the effective conservation. *The Journal of veterinary medical science*, *80*(11), 1791–1800. https://doi.org/10.1292/jvms.17-0678

Jalihal, D., Sankolli, K., & Shenoy, S. (1993). Evolution of Larval Developmental Patterns and the Process of Freshwaterization in the Prawn Genus *Macrobrachium* Bate, 1868 (Decapoda, Palaemonidae). *Crustaceana*, *65*(3), 365-376. doi:10.1163/156854093x00793

Jose, D., & Harikrishnan, M. (2018). Evolutionary history of genus *Macrobrachium* inferred from mitochondrial markers: A molecular clock approach. *Mitochondrial DNA Part A*, *30*(1),92-100. doi:10.1080/24701394.2018.1462347

Johnson, J. A., Dunn, P. O., & Bouzat, J. L. (2007). Effects of recent population bottlenecks on reconstructing the demographic history of prairie-chickens. *Molecular Ecology*, *16*(11), 2203–2222. https://doi.org/10.1111/j.1365-294x.2007.03285.x

Leigh, J. W., & Bryant, D. (2015). popart : Full-feature software for Haplotype Network Construction. *Methods in Ecology and Evolution*, *6*(9). https://doi.org/10.1111/2041-210x.12410

Librado, P., & Rozas, J. (2009). DnaSP V5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11). https://doi.org/10.1093/bioinformatics/btp187

Liu, M.-Y., Cai, Y., & Tzeng, C.-S. (2006). Molecular systematics of the freshwater prawn genus Macrobrachium Bate, 1868 (Crustacea: Decapoda: Palaemonidae) from mtDNA sequences, with emphasis on East Asian species. *Zoological Studies*, *46*.

Macrobrachium Rosenbergii. Marine Invasions Research at the Smithsonian Environmental Research Center. (n.d.). Retrieved from https://invasions.si.edu/nemesis/species_summary/96343#:~:text=Macrobrachium%20rosenbergii%20

McDowall, R.M., (2004). Ancestry and amphidromy in island freshwater fish faunas. *Fish and Fisheries*, *5*(1), pp.75-85.

McDowall, R. (2007). Hawaiian Stream Fishes: The Role of Amphidromy in History, Ecology, and Conservation Biology. Biology of Hawaiian Streams and Estuaries.

Native animals. Division of Aquatic Resources. (2019, August 12). Retrieved from https://dlnr.hawaii.gov/dar/habitat/about-streams/native-animals/

Rowe, G., Sweet, M., & Beebee, T. J. (2017). *An introduction to molecular ecology* (3rd ed.). Oxford, United Kingdom: Oxford University Press.

Rozas, J., et al. (2016). DnaSP v5. Tutorial. Universitat de Barcelona.

Strüder-Kypke, M. C., & Lynn, D. H. (2010). Comparative analysis of the mitochondrial cytochrome oxidase subunit I (COI) gene in ciliates (Alveolata, Ciliophora) and evaluation of its suitability as a biodiversity marker. *Systematics and Biodiversity*, *8*(1), 131–148. https://doi.org/10.1080/14772000903507744

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3). https://doi.org/10.1093/genetics/123.3.585

Appendices

Figure 6. Photos of *Marcrobrachium grandimanus* (right) and *M. lar* (left) from Hawaii island

Table 4. Members of Haplotype groups by sample ID number for *M. grandimanus*

Haplotype Group	Sample Numbers
A	H9-1, H9-2, H9-3, H17-1, H17-2, H17-5, H17-9, H17-10, H17-12, OI1-5, OI1-6, OI1-7, OI1-8, OI2-2, OI2-3, OI2-5, OI2-12, OI2-13, KI2-3, KI2-6, KI2-8, KI2-9, KI2-11, KI2-12, KI2-13, KI2-14, KI2-15, KI2-17, KI2-18, KI2-19, Ki2-20, KI2-21, KI2-22
B	H9-5, H17-4, OI1-17, OI2-7, KI2-4, KI2-10
C	H17-6
D	H9-4
E	OI2-4
F	KI2-2
G	KI2-16

Table 5. Members of Haplotype groups by sample ID number for *M. lar*

Haplotype Group	Sample Numbers
A	H5-2, H5-3, H5-6, H5-8, H11-5, H11-6, H11-12, H19-2
B	H5-5, H5-10, H5-12, H19-3
C	H5-1, H11-8, H11-15