PATTERNS OF POPULATION STRUCTURE AND GENE FLOW OF GASTROPODS AT HYDROTHERMAL VENTS AT THE MATA VOLCANOES

by

Jack Towson

Submitted in partial fulfillment of the requirements for graduation as an Honors Scholar with a

Bachelor of Science

at

POINT LOMA NAZARENE UNIVERSITY

April 2024

Supervised by thesis advisor

Walter Cho (Ph.D.)

and thesis committee members

Andrew Nosal (Ph.D.) and Carlson Triebold (Ph.D.)

Abstract

Gastropod *Ifremeria nautilei* was collected at deep-sea hydrothermal vents at the underwater Mata Volcanoes near Tonga to investigate gene flow and population structure between and within vent communities. Nuclear DNA was extracted from gastropod samples from four volcanoes, and fragments of the nuclear genome were amplified using six different microsatellite loci. Results from various population analyses showed no significant genetic structure of the gastropod populations in regard to depth or geography, indicating high levels of gene flow between hydrothermal vent communities, and supported the results of previous studies performed on these populations using mitochondrial markers. Results from this study may provide insight into managing these unique habitats as human activity, such as deep-sea mining, increases in the deep sea.

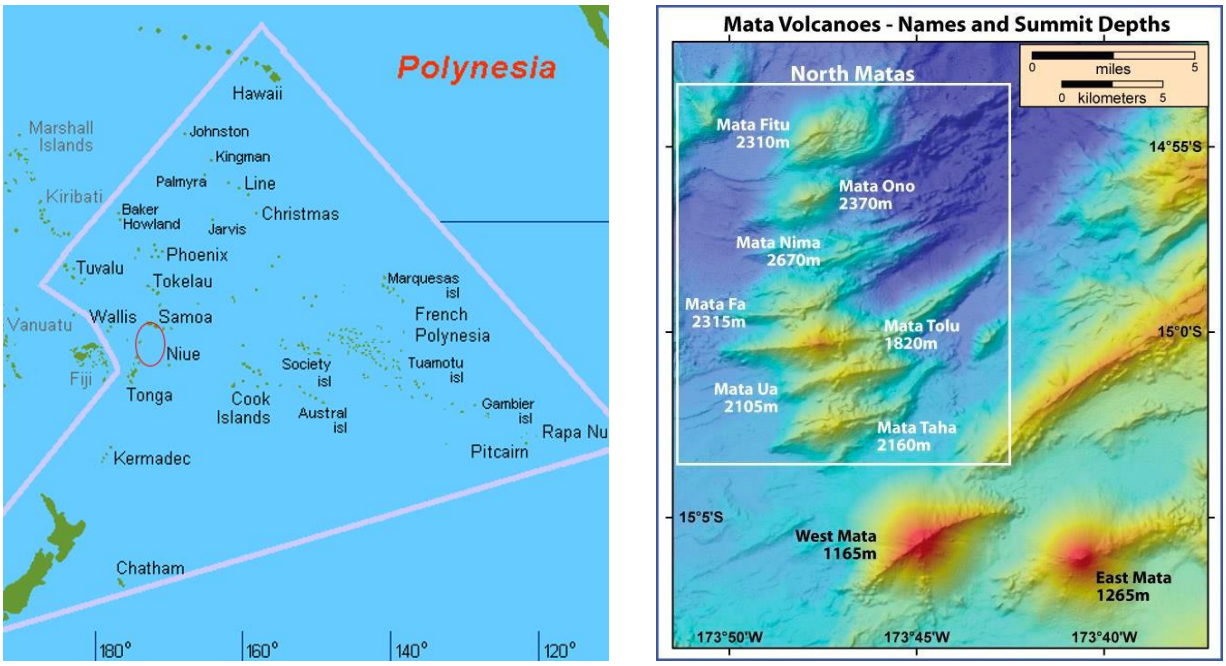
Introduction

Hydrothermal vents are one of the most unique environments to viably support an ecosystem, as they are filled with diverse species and ecosystems despite lying thousands of meters below the surface of the ocean with little to no light penetration. These vents expel heated, chemical-rich, and nutrient-rich seawater through chimney-like openings on the ocean crust. Chemosynthetic bacteria convert nutrient-rich fluid into energy by chemosynthesis, and they serve as a source of energy and nutrients for other species within the vent community. *Ifremeria nautilei*, a local gastropod species, has a mutualistic symbiosis with chemosynthetic bacteria. The snails provide the bacteria with a place to live in an extremely stressful environment, and, in return, the chemosynthetic bacteria provide the snails with sufficient nutrients required to survive (Thaler et al., 2010). Considering the widespread distribution of hydrothermal vents across the ocean floor and the high abundance of chemosynthetic bacteria at these vents, comparable mutualistic symbiotic relationships are observed in other active vent communities (Amend, 2004).

Hydrothermal vents are ephemeral, meaning they appear and disappear rapidly. As a result, vent species must have a wide larval dispersal to prevent local extinction. *I. nautilei* are brooders, so adult gastropods incubate their eggs prior to dispersal. Upon dispersion, *I. nautilei* larvae, known as Waren's larvae, exit their egg from their mother's internal brood pouch as free-swimming, fully ciliated, lecithotrophic larvae (Reynolds et al., 2010). In other words, these free-swimming larvae are non-feeding and are covered in sensory hairs that assist in propulsion through the ocean. After 15 days, the Waren's larvae metamorphose into the typical larval form of gastropods, or the Veliger form (Reynolds et al., 2010). The ability of *I. nautilei* larvae to efficiently travel long distances to other vents is critical to the survival of this species, and it also allows different populations of gastropods to interact with each other and increase biodiversity.

The Mata Volcanoes are a group of underwater volcanoes that promote the emergence of hydrothermal vents. These nine distinct volcanoes are located in the South Pacific at varying depths, ranging from 1800 meters to 2700 meters, in the Northeast region of the Lau Basin, off the northeast coast of Tonga (Clague et al., 2011). The Mata Volcanoes are also all active and relatively young. Furthermore, the individual volcanos are within notable proximity of each other, located no farther than 1.5 to 1.75 kilometers apart (Figure 1). The diversity and

abundance of life present at the Mata Volcanoes is surprising, ranging from chemosynthetic bacteria to various marine invertebrates. Gastropods, in particular, were discovered in abundance around the hydrothermal vents scattered throughout the volcanoes. *I. nautilei* was identified as one of the two most prominent gastropod species found (the other being *Alviniconcha boucheti*). Most species discovered at these hydrothermal vents are endemic to vent ecosystems, meaning that vent sites are the only places where these species are found in the world.



(A) (B)

Figure 1: Volcanoes in the North Mata range and their associated approximate depths. (B) is a zoomed-in map of the red circle depicted in (A).

I. nautilei is a part of the phylum Mollusca, class Gastropoda, family Provannidae. Organisms within the Provannidae family are classified as having a conical shell that is 6 to 15 millimeters in length (Figure 2). Furthermore, these gastropods possess reduced eyes near the outside of the cephalic tentacles, a paired jaw, a stomach with a developed style sac and gastric shield with a simple chamber, and digestive glands that open to the stomach. The females have gonopericardial ducts, whereas the males have an aphallate with dimorphic sperm (Warén and

Ponders, 1991). Interestingly, the *Ifremeria* genus only has one identified species: *I. nautilei* (Bouchet and Warén, 1991). Like most vent species, *I. nautilei* are endemic to hydrothermal vents and thus cannot be located elsewhere.

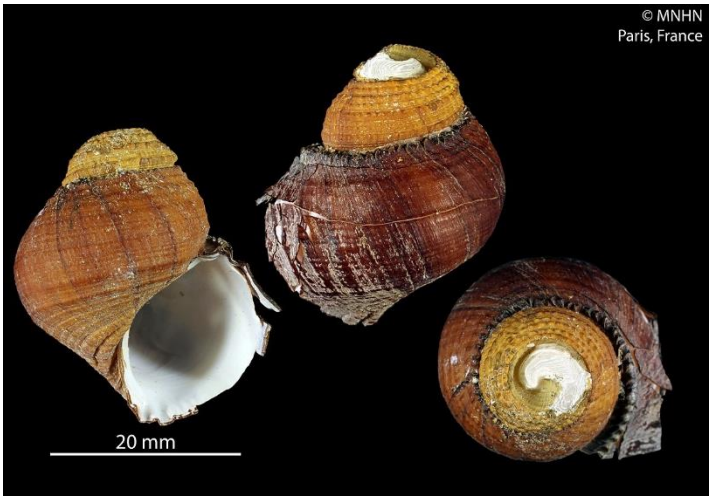


Figure 2: *Ifremeria nautilei* external morphology (Waren & Bouchet, 1993)

Previous studies have been done on the population structure of specialized hydrothermal vent communities in the western Pacific. Thaler et al. (2011), in particular, studied the scale at which genetic differentiation occurred among populations of *I. nautilei* by using DNA microsatellite markers to determine patterns of gene flow and isolation for a series of basins off of the coasts of Australia and Fiji. More specifically, Thaler compared population structure and diversity within and between *I. nautilei* populations at the Manus, North Fiji, and Lau basins. While prominent levels of gene flow were found to occur between the North Fiji and Lau basin, indicating that gastropods from these two locations were interacting with each other, this was not found to be true of the Manus Basin. Thus, a contemporary restriction of gene flow isolated the Manus Basin population of *I. nautilei* from the *I. nautilei* populations native to the North Fiji and Lau Basins (Thaler et al., 2011). The Thaler et al. (2011) study not only showed that an understanding of hydrothermal vent populations at various scales can help define natural conservation units and minimize genetic diversity losses in regions where human interference is

proposed, but it also illustrated that microsatellite DNA markers are a viable option for analyzing population structure and genetics of deep-sea communities and the species that comprise them.

Research on *I. nautilei* populations at the Mata Volcanoes has also been completed by Davis et al. (2022), in which genetic structure across varying depths, vent locations, and populations was characterized by analyzing the cytochrome oxidase I (COI) mitochondrial gene. They identified all 57 samples collected from the Mata Volcanoes as *I. nautilei* and discovered that the genetic sequences of these individuals were extremely similar, signifying close relationships between the samples. The haplotype network of the analyzed gastropod samples provided evidence for high levels of gene flow between different communities along the Mata Volcanoes and refuted the potential idea these communities were physically and genetically isolated from each other. Because closely related alleles were found across different volcanoes at different coordinates, depths, and vents, it was determined that genetic exchange was occurring. Additionally, a nested clade analysis and analysis of molecular variance (AMOVA) found no population structure and supported high gene flow for all locations. Thus, it is expected that microsatellite analysis of the genetic structure of the Mata Volcano *I. nautilei* populations will corroborate the results from mitochondrial markers. However, this verification of high gene flow between *I. nautilei* populations at the Mata Volcanoes remains yet to be completed.

Microsatellites are tandem repeats of approximately 1 to 6 nucleotides found in the nuclear genomes of most taxa (Selkoe and Toonen, 2006). The flanking region, or the DNA encompassing the microsatellite locus within the genome, is usually conserved across individuals of the same species. As a result, primers can be easily made to bind to the specific flanking region in a polymerase chain reaction (PCR). Microsatellites have relatively high mutation rates, thus enabling the study of allelic diversity on relatively short ecological timescales (Selkoe and Toonen, 2006). While mitochondrial markers can only distinguish between populations of the same species, microsatellites possess a much greater resolution: they distinguish differences between individuals of the same species within the same population (Selkoe and Toonen, 2006). Because of their high mutation rates, species-specificity, and high resolution for a DNA marker, microsatellites offer another dependable tool for population genetics analysis and the study of community structure across unique habitats such as hydrothermal vent communities.

Current microsatellite analysis is commonly computed using STRand Analysis Software used at the University of California, Davis' Veterinary Genetics Lab. It speeds up the analysis of DNA fragment length polymorphism samples run on fluorescence-based gels (Toonen, 2003). In this study, STRand was utilized to analyze highly multiplexed, short tandem repeat (STR) markers for *I. nautilei*. However, the potential replication of a more automated, user-friendly version of the STRand program was explored in the R programming language, and a script using the *Fragman* and *MsatAllele* packages was drafted to offer the same functions as STRand in a more efficient manner.

Methods and Materials

Sample Collection

The *I. nautilei* samples were collected on the vessel *R/V Falkor* during the 2017 expedition to Samoa and the Mata Volcanoes, FK171110 Underwater Fire. More specifically, the remotely operated vehicle (ROV) SuBastian collected the specimens from the hydrothermal vents scattered throughout the Mata Volcanoes. Samples were directly collected using either a multi-chamber suction sampler or scoop nets. The samples were preserved on the vessel by either directly freezing them at -80ºC or storing them in 95% EtOH or RNALater before freezing. After arriving at the laboratories, the samples were transferred to a -80ºC freezer for long-term storage and usage.

Microsatellite DNA Processing

In the laboratory, DNA extractions were performed on the gastropod samples using the DNEasy Blood & Tissue Kit from Qiagen (Hilden, Germany), according to the manufacturer's protocols with the following modifications: for DNA elution, 100 microliters of buffer AE was used and immediately followed by an additional 100 microliters of buffer AE. During the final collection step, the samples were centrifuged at 13,200 rpm, rather than the suggested 14,000 rpm. DNA extractions were stored in 1.5 mL microcentrifuge tubes and stored at -20ºC for short-term storage when used frequently or at -80ºC for long-term storage when used infrequently. After completing the DNA extractions successfully, the samples were inspected for the presence of DNA using gel electrophoresis. The gel was made with 1.5% agarose gel, 1X TBE buffer, and

ethidium bromide. Furthermore, the DNA concentrations of the samples were verified with a NanoDrop 2000c spectrophotometer (ThermoFisher Scientific, Waltham, MA).

Six different primer pairs were obtained from ThermoFisher Scientific (Waltham, MA) for microsatellite markers identified by Thaler et al. (2010). Of the 12 viable microsatellite loci for *I. nautilei* suggested by Thaler, three primers (Ifr78, Ifr94, and Ifr103) were selected based on their ability to cross-amplify with *A. boucheti*. The additional three microsatellite markers (Ifr40, Ifr43, and Ifr93) were selected due to their allelic size ranges (Thaler et al., 2010). The primers were mapped using Multiplex Manager 1.2 (Holleley and Geerts, 2018) for the identification of the appropriate primer pair combinations. Once obtained, the labeled primers were completely wrapped in tin foil to preserve their viability and they were stored alongside the unlabeled primers at -20ºC.

Primer stocks containing the unlabeled markers were made up to determine if the DNA sequences would amplify and to identify the ideal conditions for amplification in a multiplex PCR. A gradient PCR was performed to determine the best annealing temperature for the *I. nautilei* samples (61ºC). After confirming that the unlabeled primers worked for *I. nautilei*, the labeled primers were tested. The molecular markers were allocated into two primer stocks according to their allelic size ranges: the first primer stock (PS1) contained four primers (Ifr40, Ifr43, Ifr93, and Ifr94), whereas the second primer stock (PS2) contained the remaining two primers (Ifr78 and Ifr103). Molecular markers were initially amplified in multiple PCRs using 12 µL of 2X Type-It Multiplex PCR Master Mix, 2.5 µL of 10X primer mix (2 µM of each primer), 2 µL of template DNA, and 8 µL of RNase-free water for a final concentration of 25 µL. Specific PCR parameters ran with the Type-It samples are listed in Table 1.

However, little success was observed in amplifying the PS2 markers using the Type-It kit. Alternatively, the molecular markers were tested and amplified in multiple PCRs using 5 µL of 4X Bio-Rad Reliance One-Step Multiple Supermix, 2 µL of 10X primer mix, 2 µL of template DNA, and 11 µL of MilliQ water was added to bring the final volume of each reaction to 20 µL. Clearer, more reliable results were obtained using the 4X Bio-Rad Reliance One-Step Multiple Supermix. The PS1 markers and PS2 markers produced ideal results under different PCR parameters, specifically the annealing time for each cycle, so two slightly different PCR

programs were used to produce the greatest amplification of each primer stock. These different PCR programs for Bio-Rad samples are shown in Table 1.

Following a PCR, the products were run through 1.5% agarose gels via gel electrophoresis to check for the presence of DNA within the expected base pair range of each molecular marker. If bands were present, they were classified on a scale of bright, medium, or faint. Samples with bright and medium bands were stored at -20 ºC until they were sent out for analysis, and samples with dull bands were rerun on a separate PCR. Samples that continuously produced faint bands were also stored at -20 ºC for analysis if they had relatively low DNA concentrations. The viable microsatellite PCR products that contained clear bands were diluted to appropriate concentrations to be sent out to Eton Bioscience for fragment analysis.

Table 1: Multiplex PCR conditions for microsatellite marker amplification for *I. nautilei*

Samples/Kit	1) Intial denaturation	2) Denaturation	3) Annealing	4) Extension	Number of cycles	5) Final extension	Completion
Type-It	95ºC for 5:00	95ºC for 5:00	61ºC for 1:30	72ºC for 0:30	Repeat 2-4 30X	60ºC for 30:00	End (4º for ∞)
Bio-Rad PS1	95ºC for 10:00	95ºC for 0:30	58ºC for 1:30	60ºC for 0:30	Repeat 2-4 35X	N/A	End (4º for ∞)
Bio-Rad PS2	95ºC for 10:00	95ºC for 0:30	58ºC for 1:00	60ºC for 0:30	Repeat 2-4 35X	N/A	End (4º for ∞)

Microsatellite Genetic Analysis

The fragment analysis results were received from Eton Bioscience as chromatogram files and imported into STRand. Expected allelic size ranges for each microsatellite marker were defined prior to the analysis of the samples. The GeneScan 500 LIZ size standard (35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, 500) was used to score the alleles of each sample, and raw allelic sizes for each locus of each sample were determined. Typically, more than 16 fragment sizes were called automatically by STRand, resulting in incorrectly defined sizes for the microsatellite loci of the samples. Thus, the manual definition of the actual 16 size standard peaks in the electropherogram was often necessary. After correcting the size standard, the correct allelic sizes for each microsatellite loci were defined. After all the viable samples were analyzed and the messy or overdiluted samples were removed, the final database,

containing correctly called fragment sizes, was exported into GENEPOP ON THE WEB to perform population analyses.

Due to the lack of automated success in STRand, alternative options were explored to obtain identical results in a more adept approach. Thus, the fragment analysis results received from Eton Bioscience as chromatogram files were additionally imported into R. For each sample, allelic size ranges were defined and the raw allelic sizes of all six molecular markers, if present, were binned into whole number allele lengths (number of base pairs) using the *Fragman* package (Covarrubias-Pazaran et al., 2016). Initially, the size standards and allelic sizes were called automatically by *Fragman*. However, if the program identified strange peaks that resulted from stuttering or extra noise, manual classification of a marker's peak was also available and offered to the user (Covarrubias-Pazaran et al., 2016). Additionally, messy or over-diluted samples were identified and removed from the imported database. Individual databases for each molecular marker, as well as one collective database for all six markers, were created for future analysis. The databases containing the identified peaks for each sample were then imported into the *MsatAllele* package and checked for any erroneous sizes that were incorrectly called by *Fragman* and unnoticed by the user during manual identification. These erroneous peaks were located through graphical figures, specifically histograms and cumulative allelic distributions, which illustrated the distribution of allelic sizes for a given loci across all samples possessing that particular marker (Alberto, 2009). Peaks that fell notably outside the expected size range for a marker were re-imported into the *Fragman* package, verified, and removed from the dataset. The final database, containing the samples with correctly called peaks, was then available to export for further analyses completed outside of R.

Input files for some external analyses were created from the molecular marker database in the 6-digit GENEPOP ON THE WEB version 4.7 format (Raymond & Rousset, 1995). Data from this file type were then reformatted by GENEPOP ON THE WEB to create input files for FSTAT version 2.9.3.2 (Goudet, 2001) and ARLEQUIN version 3.5.2.2 (Excoffier et al., 2007). Additionally, an MS Excel spreadsheet containing identical data to the molecular marker database, sorted by vent location collection site, was used to create an input file in the 3-digit MICRO-CHECKER version 2.2.3 format (van Oosterhout et al., 2004). The FSTAT file format was used in PCA-GEN version 1.2 (Goudet, 1999). The 3-digit allele length data utilized in

MICRO-CHECKER were also manually formatted for subsequent analyses in STRUCTURE version 2.3.4 (Pritchard et al., 2000; Falush et al., 2003, 2007).

MICRO-CHECKER was used similarly to the MsatAllele package in R, as it screened each molecular marker for any potential scoring errors—null alleles, large allele dropout, and stuttering—that would cause the dataset to unnaturally deviate from Hardy-Weinberg equilibrium (HWE) expectations (run at default 95% CI, 1000 permutations). The probabilities for observed homozygote-size class frequencies were calculated by a cumulative binomial distribution (Weir, 1996) and by ranking the observed frequency in a distribution of randomized genotypes (van Oosterhout et al., 2004). The obtained *P* values were then combined to reveal deviations from Hardy-Weinberg proportions (van Oosterhout et al., 2004). Once alleles were ready to be reliably scored, the number of alleles, allelic richness, allelic size range, and size and frequency of each allele were calculated within and across all subpopulations in FSTAT.

In ARLEQUIN, nonbiased expected and observed heterozygosity values were determined following Fusaro (2008). Additionally, Fisher-type exact tests of HWE per locus were conducted in ARLEQEUIN with a recommended Markov chain of 2,000,000 steps and 200,000 dememorization steps for reproducibility (L. Excoffier, Genetic Software Forum pers. comm.). To assess locus conformance to HWE expectations of random mating, F_{IS}-based estimates were calculated in FSTAT (Weir & Cockerham, 1984) with the following parameters: 2000 permutations (recommended value for < 10 loci = 1000 permutations). Furthermore, the dataset was tested for linkage disequilibrium in GENEPOP ON THE WEB using a probability test on all locus-specific contingency tables under the null hypothesis of independence (Cockerham & Weir, 1979). The parameters for the linkage disequilibrium test were comprised of 5000 dememorization steps, 500 batches, and 2000 iterations per batch—at which variation in *P* value magnitude among multiple runs did not alter the level of pairwise significance. All significance tests were run at a nominal alpha level of 0.05 with sequential Bonferroni correction (Rice, 1989).

To detect heterogeneity among and between pooled populations and cohorts, exact tests of differentiation in allele (genic) and genotype (genotypic) distributions between sample pairs were conducted in GENEPOP ON THE WEB. The same parameters used to test for linkage disequilibrium were used to test for heterogeneity among and between populations: 5000

dememorization steps, 500 steps, and 2000 iterations per batch. A principal components analysis (PCA) based on F_{ST} comparisons was used to identify clusters of subpopulations from genotypic data using PCA-GEN.

Overall population structure was estimated using hierarchical analysis of molecular variance (AMOVA) in ARLQEUIN with 20,000 permutations. AMOVA quantified the magnitude of genetic variation among geographic regions, relative to other sources of genetic variation among populations, within populations, and within individuals. Significant differences from zero in Wright's fixation indices at these four levels were tested for deviation from HWE by comparing observed values to a null distribution by randomly permutating populations defined by volcano region (north volcanoes vs. south volcanoes), depth, and individual volcano.

Furthermore, the samples were examined for additional genetic structure among one another and between volcanoes using an admixture model of ancestry and correlated frequencies in STRUCTURE (burnin of 20,000, followed by 20,000 MCMC replicates; tested for K=1 through K=5). This program uses an algorithm that assigns individuals to groups assuming that loci are at HWE within each group and estimates the origin population for each individual according to their observed genotypes (Pritchard et al., 2000). The posterior probability of the data (PPD) was plotted as in Garnier et al. (2004) to determine the best number of population groups (K) given the results.

Estimates of Weir & Cockerham's allele frequency-based θ_{WC} (estimates Wright's 1951 F_{ST}), assuming the infinite alleles model of mutation, and Slatkin's (1995) allele repeat-based R_{ST}, assuming the stepwise model of microsatellite mutation, were calculated between population and cohort pairs in ARLEQUIN. 20,000 permutations were used to test for significance alongside sequential Bonferroni correction for all tests. F_{ST} was also calculated and tested for significance per locus across all populations in FSTAT. These comparisons considered genetic variability within and between populations, where the genic and genotypic differentiation tests ran in GENEPOP ON THE WEB were based on the distribution of alleles or genotypes among populations.

Genetic isolation by geographical distance was evaluated for significance using the Mantel test implemented in GENEPOP ON THE WEB (500,000 permutations). Because the distance between populations was greater than the habitat width of *I. nautilei*, Slatkin's (1995)

linearized F_{ST} was plotted against geographical distance to determine the significance of the relationship (Rousset, 1997). The linear distance between sites was calculated as the shortest route between given coordinates of latitude and longitude using a great circle calculator (available at https://edwilliams.org/gccalc.htm).

The effective number of migrants among populations or cohorts (Nm) was estimated using the rare alleles method offered in GENEPOP ON THE WEB (Slatkin, 1985; Barton & Slatkin, 1986). This overall estimate of gene flow is calculated using the average frequency of alleles found in only one population.

Results

Genetic variability

A total of 57 *I. nautilei* individuals were collected and analyzed from four populations at six polymorphic microsatellite loci. The volcanoes from which these individual gastropods were collected ranged in depths from 1823 to 2623 meters (Table 2; Figure 3)**.**

Table 2: Metadata for *I. nautilei* samples collected and analyzed in this study.

Volcano	Dives	Collection Date	Latitude (South)	Longitude (West)	Depth (m)	*I. nautilei* Samples
Mata Ua	S89, S100	11/30/17, 12/11/17-12/12/17	-15.017736	-173.788517	2359	26
Mata Tolu	S91, S94	12/03/17, 12/05/17-12/06/17	-15.004639	-173.792739	1823	16
Mata Fitu	S97	12/08/17-12/09/17	-14.913476	-173.792733	2623	2
Mata Ono	S101, S102	12/13/17, 12/14/17	-14.940578	-173.799375	2360	13

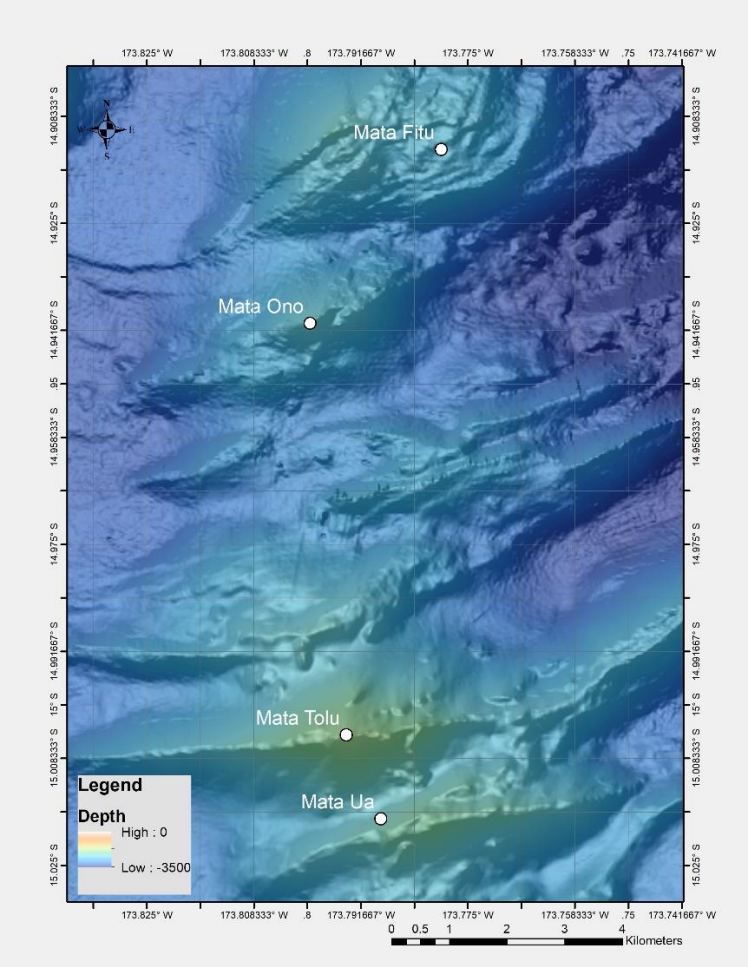


Figure 3: Map of Mata Volcanoes where *I. nautilei* samples used in this study were collected.

All 57 individuals were genotyped at all loci and replicated per plate and between runs to promote consistent allele scoring. Across all populations, the total number of alleles per locus showed fair variability, ranging from two (Ifr43) to fourteen (Ifr93), with a mean total allele count per locus of 5.5. Allelic richness normalized to a minimum of 10 individuals per sample ranged from 1.750 (Ifr43, Ifr94) to 6.750 (Ifr93) per locus and population (mean = 1.833). Allele lengths ranged from 170 bp (Ifr94) to 346 bp (Ifr103). The most common alleles and their frequencies within populations are provided in Table 3.

Hardy-Weinberg equilibrium

Non-biased expected heterozygosities (gene diversity) ranged vastly from 0.03846 (Ifr94 at Mata Ua) to 0.83333 (Ifr40 at Mata Fitu) (mean = 0.28617) (Table 3). Observed heterozygosity ranged from 0.03846 (Ifr94 at Mata Ua) to 1.000 (Ifr43, Ifr94, Ifr78, and Ifr103 at Mata Fitu; Ifr78 at Mata Ua) (Table 3). MICRO-CHECKER revealed that no issues of scoring error due to stuttering, large allele dropout, or evidence for null alleles in any of our populations (the different volcanoes). Thus, there was no evidence of significant departure from HWE. In FSTAT, all single locus F_{IS}-based exact tests were consistent with Hardy-Weinberg equilibrium expectations in the absence of inbreeding (Table 4). F_{IS} values approaching 1 suggest heterozygote deficiency and promote inbreeding, whereas F_{IS} values approaching -1 suggest heterozygote excess and promote inbreeding avoidance. Values close to or equal to 0 indicate a fairly even distribution of homozygotes and heterozygotes in the population. While loci in certain populations did approach deviation from HWE (Ifr103 at Mata Ono), none of the loci truly departed from HWE, suggesting complete locus conformance to HWE expectations of random mating.

Table 3: Allelic variability at six microsatellite loci in *Ifremeria nautilei* populations. Allelic richness based on a minimum sample size of two individuals, absolute allele count, allele size ranges, size and frequency of most common allele (FSTAT), and H_E (non-biased) and H_O with test for departure from Hardy-Weinberg expectations (ARLEQUIN). N/A values for H_E and H_O signify that the locus is monomorphic for that population (H_O = 1.0000). Significant departures from Hardy-Weinberg expectations are highlighted in bold. *n*, sample size; *k*, number of alleles; R_S, allelic richness; A_R, allele size range; *S*, size in base pairs; *F*, frequency of the most common alleles; H_E, expected Nei's non-biased heterozygosity; H_O, observed heterozygosity. Per locus across all populations: k_{TOT}, total number of alleles; k_{mean}, mean number of alleles; $R_{S\ mean}$, mean allelic richness; $A_{R\ TOT}$, total allele size range; $H_{E\ mean}$, mean non-biased expected heterozygosity.

		Locus					
Population	**Statistic**	Ifr93	Ifr40	Ifr43	Ifr94	Ifr78	Ifr103
Mata Fitu	*n*	2	2	2	2	2	2
	k	3	2	1	1	1	1
	R_S	3.000	2.000	1.000	1.000	1.000	1.000
	A_R	238-246	198-212	172	250	228	274
	S	238/244/246	198/212	172	250	228	274
	F	0.500	0.750	1.000	1.000	1.000	1.000
	H_E	0.50000	0.83333	N/A	N/A	N/A	N/A
	H_O	0.50000	0.50000	N/A	N/A	N/A	N/A
Mata Ono	*n*	13	13	13	13	13	13
	k	5	3	2	2	4	3
	R_S	2.250	1.443	1.408	1.154	1.817	1.443
	A_R	238-246	190-202	170-172	250-262	226-230	226-274
	S	238/240/242/ 244/246	190/198/ 202	170/172	250/262	226/228/ 230	226/270/ 274
	F	0.654	0.885	0.885	0.769	0.769	0.885
	H_E	0.56000	0.21846	0.21231	0.07692	0.39692	**0.21846**
	H_O	0.69231	0.23077	0.23077	0.07692	0.38462	0.07692
Mata Tolu	*n*	16	16	16	16	16	16
	k	9	3	2	2	2	3
	R_S	2.697	1.363	1.340	1.238	1.238	1.363
	A_R	230-254	190-202	170-172	250-258	228-230	246-274

	S	230/232/234/236/238/242/244/246/254	190/198/202	170/172	250/258	228/230	246/270/274
	F	0.531	0.906	0.906	0.938	0.938	0.906
	H_E	0.70161	0.17944	0.17540	0.12097	0.12097	0.17944
	H_O	0.75000	0.18750	0.06250	0.12500	0.12500	0.18750
Mata Ua	*n*	26	26	26	26	26	26
	k	10	3	2	2	1	4
	R_S	2.381	1.299	1.498	1.077	1.000	1.580
	A_R	232-260	190-202	170-172	250-258	228	270-346
	S	232/238/240/242/244/246/248/250/256/260	190/198/202	170/172	250/258	228	270/274/278/346
	F	0.615	0.923	0.846	0.981	1.000	0.846
	H_E	**0.60106**	0.14781	0.26546	0.03846	N/A	0.28054
	H_O	0.61538	0.15385	0.15385	0.03846	N/A	0.23077
Total	k_{TOT}	14	4	2	3	4	6
	k_{mean}	6.750	2.750	1.750	1.750	2.000	2.750
	$R_{s\ TOT}$	2.439	1.367	1.412	1.138	1.267	1.469
	$R_{s\ mean}$	2.582	1.526	1.312	1.117	1.264	1.347
	$A_{R\ TOT}$	230-260	190-212	170-172	250-262	226-230	226-346
	$H_{E\ mean}$	0.59067	0.34476	0.21772	0.07878	0.25895	0.22615

Table 4: F_{IS} per locus and population (FSTAT). P-values in parentheses (1440 randomizations). Values of N/A indicate complete homozygosity for a specific loci in a particular population.

	Population			
Locus	Mata Fitu	Mata Ono	Mata Tolu	Mata Ua
Ifr93	0.500 (0.3208)	-0.249 (0.1396)	-0.071 (0.4729)	-0.024 (0.5646)
Ifr40	0.000 (1.0000)	-0.059 (0.8896)	-0.047 (0.9146)	-0.042 (0.8750)
Ifr43	N/A	-0.091 (0.8813)	0.651 (0.1000)	0.425 (0.0792)
Ifr94	N/A	0.000 (1.0000)	-0.034 (0.9667)	0.000 (1.0000)
Ifr78	N/A	0.032 (0.7479)	-0.034 (0.9646)	N/A
Ifr103	N/A	0.657 (0.0667)	-0.047 (0.8958)	0.180 (0.1917)
All loci	0.333 (0.3208)	-0.006 (0.6208)	0.028 (0.4146)	0.108 (0.1167)

Allelic and genotypic variation

The probability test on all locus-specific contingency tables under the null hypothesis of independence for linkage disequilibrium revealed that none of the loci were linked (Table 5). In other words, no combination of two loci will be inherited together. Exact tests of genic and genotypic variation in pairwise population comparison detected mostly high genetic exchange between the Mata Volcanoes, with only two of 16 tests of genic and genotypic differentiation being significant. Locus Ifr78 contributed with significance to both types of differentiation across all populations when all the samples were pooled by location (Table 6). Despite the significant population differentiation at both the genic and genotypic level for Ifr78, the overall genic and genotypic differentiation across all loci is not significant. Thus, Mata Volcano

populations do not significantly differ from each other in genic and genotypic comparisons (despite locus-specific differentiation in one locus).

Table 5: Probability test *P* values for loci linkage disequilibrium.

Locus	Ifr93	Ifr40	Ifr43	Ifr94	Ifr78	Ifr103
Ifr93		0.99576	0.96891	0.44996	0.25849	0.44886
Ifr40			0.15076	1.00000	0.91111	0.64456
Ifr43				0.43884	0.95193	0.39970
Ifr94					1.00000	1.00000
Ifr78						0.92480
Ifr103						

Table 6: Genic and genotypic differentiation over all populations (GENEPOP ON THE WEB: 5000 dememorization steps, 500 batches, 2000 iterations per batch). Bold *P* values are signification with sequential correction for 8 tests at a level of 5%.

	Genic (allelic)		Genotypic	
Locus	*P* value	S.E.	*P* value	S.E.
Ifr93	0.76934	0.00498	0.82714	0.00386
Ifr40	0.53622	0.00273	0.50562	0.00267
Ifr43	0.70644	0.00128	0.80867	0.00120
Ifr94	0.40834	0.00219	0.40317	0.00215
Ifr78	**0.00878**	0.00050	**0.01297**	0.00057
Ifr103	0.30760	0.00360	0.45115	0.00373
All loci	0.18737		0.28396	

Overall population structure

Principal component analysis of pairwise F_{ST} was used to determine how overall populations and individuals within those populations clustered with one another (Figure 4). The PCA plot illustrates the different populations and the different individuals within those populations in a nondimensional space of principal components, so that the samples that are closely related will cluster together. Visually apparent was the distinct separation of populations at each volcano, as the four numbers, each representing one of the four Mata Volcano populations, did not overlap at all (Figure 4A). If the individuals from each population (or individuals from the same volcano) were all similar to each other, clear clustering of same-colored numbers should result. However, because the 57 individuals included in the PCA were fairly widespread throughout the diagram, regardless of color, there was no apparent pattern in clustering and thus no population structure (Figure 4B).

F_{ST}-based AMOVA tests in ARLEQUIN were run between the *I. nautilei* samples for the following covariates: different volcanoes, different depth ranges, and northern versus southern volcanoes (Table 7). The AMOVA detected no significant differences between the samples among groups, among populations within regions, and within populations for any of the covariates in the F_{ST}-based tests at a significance threshold of 0.05. However, significant genetic structure at the within populations level was approached in all three covariate tests. The majority of the microsatellite allele frequency variation was found within individuals (>94%).

Additionally, when plotting the posterior probabilities of the population data for K = 1 through K = 5 (where K is the number of expected populations), there was no clear indication of how many distinct populations made up the sample set. When analyzing STRUCTURE results, it is important to note that vertical bars are representative of low gene flow, whereas horizontal bars correlate to high gene flow among the clusters (O'Donnell, 2014). Across all five bar graphs illustrated in Figure 5, each representative of a different expected number of populations, the resulting horizontal bars shared an even proportion of the total area, depending on how many assumed populations there were. Thus, the lack of population structure supports high gene flow between the different colored clusters in the STRUCTURE output.

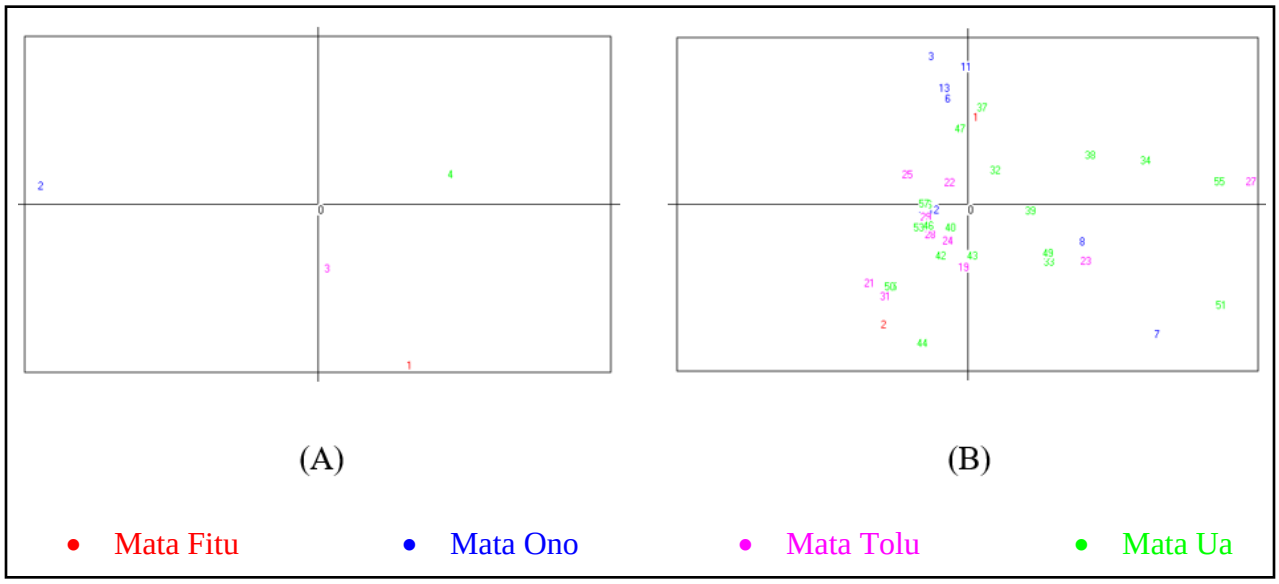


Figure 4: Principal component analysis of genetic differentiation among (A) four *I. nautilei* populations and (B) 57 *I. nautilei* individuals, showing no isolation between or within the Mata Volcanoes. The numbers are color-coded by volcano geography, representing (A) different populations or (B) separate individuals belonging to one of the four populations.

Table 7: Analysis of molecular variance testing between northern and southern Mata Volcanoes, depth, and individual Mata Volcanoes. Bold values are significant at the 5% nominal alpha level (ARLEQUIN: 20022 permutations)

Covariate	Source of Variation	*df*	Variance	%	*F*
Volcano Region (North vs. South)	Among regions	1	0.00401	0.55	F_{CT} = 0.42908
	Among populations within regions	5	-0.00847	-1.17	F_{SC} = 0.76221
	Within populations	50	0.04526	6.24	F_{IS} = 0.09145
	Within individuals	57	0.68421	94.27	F_{IT} = 0.11797
	Total	113	0.72500		
Depth	Among regions	2	-0.01305	-1.82	F_{CT} = 0.95200
	Among populations within regions	4	0.00141	0.2	F_{SC} = 0.38922
	Within populations	50	0.04526	6.3	F_{IS} = 0.08860
	Within individuals	57	0.68421	95.32	F_{IT} = 0.12097
	Total	113	0.71782		
Volcano	Among regions	3	-0.00366	-0.51	F_{CT} = 0.72690
	Among populations within regions	3	-0.00321	-0.44	F_{SC} = 0.53786
	Within populations	50	0.04526	6.26	F_{IS} = 0.09764
	Within individuals	57	0.68421	94.69	F_{IT} = 0.12276
	Total	113	0.72260		

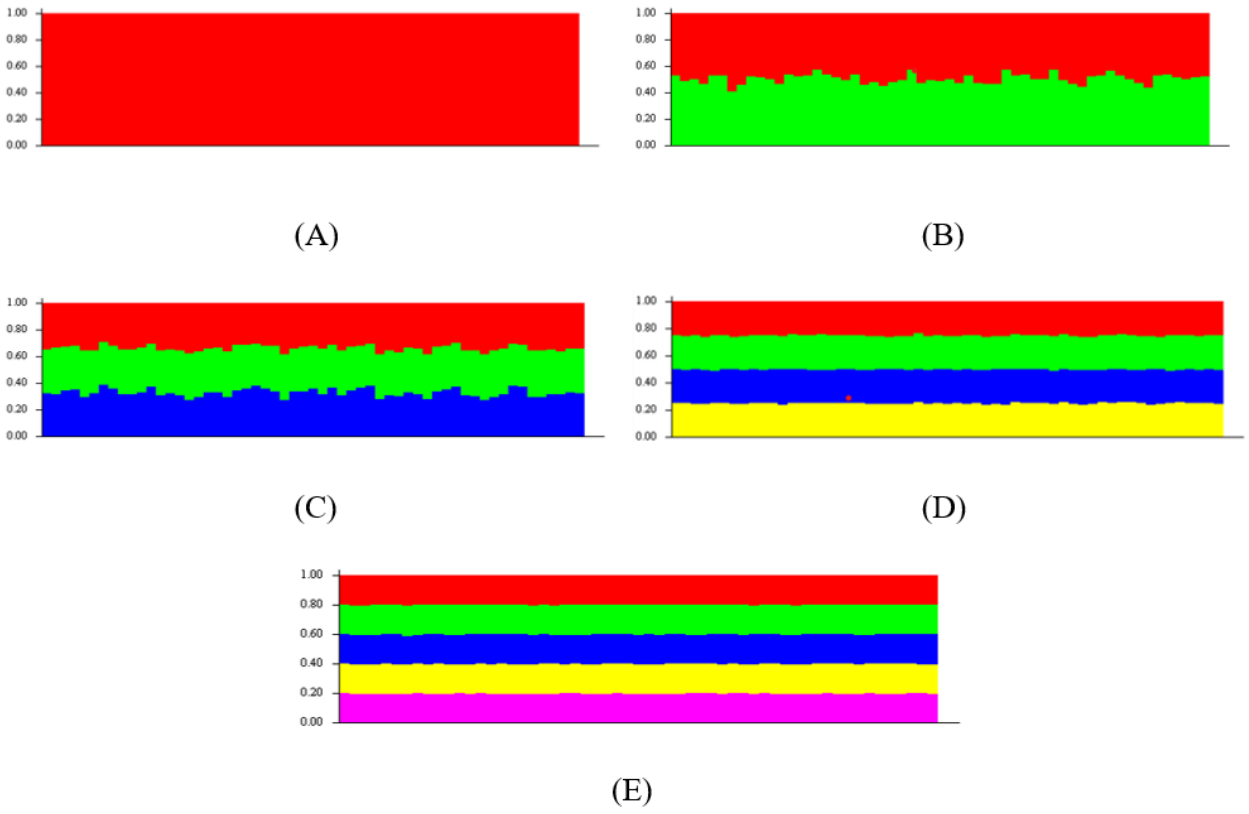


Figure 5: Results from STRUCTURE with K = 1 to K = 5 populations (20,000 burnin, 20,000 MCMC reps, admixture model, correlated allele frequencies, no probability calculation for K). Each individual, scored at six microsatellite loci, is represented by a vertical line broken into two segments representing the estimated proportion of the individual's genome originating from each inferred cluster. No clusters were evidenced among the Mata Volcano pooled populations.

Pairwise genetic differentiation

F_{ST}-based comparisons consider genetic variability within and between populations. F_{ST} values close to or less than 0 indicate complete sharing of genetic material, thus revealing prominent levels of gene flow. In comparison, F_{ST} values that are close to 1 correlate to no sharing of genetic material, suggesting genetic isolation. Per locus F_{ST} values over all populations were greatest at the Ifr103 locus, with an overall loci and all populations F_{ST} of ~3% (no significance test available; Table 8). Tests of multi-locus estimates of F_{ST} ($\theta\hat{}$) found no

significant differences between any pooled combination of individual Mata Volcanoes (Table 9). Therefore, all of the Mata Volcano populations share their genetic material, and elevated levels of gene flow occur.

Table 8: F_{ST} per locus over all populations (FSTAT).

Ifr93	Ifr40	Ifr43	Ifr94	Ifr78	Ifr103	All Loci
-0.041	0.014	-0.055	-0.045	0.029	-0.056	-0.028

Table 9: Genetic differentiation between populations (F_{ST}, above diagonal) and respective *P* values (below diagonal). No pairwise significance at the 5% nominal level was detected between any of the individual Mata Volcanoes.

	Mata Fitu	Mata Ono	Mata Tolu	Mata Ua
Mata Fitu		-0.0595	-0.0900	-0.0650
Mata Ono	0.80000		-0.0053	-0.0080
Mata Tolu	0.82500	0.62500		-0.0069
Mata Ua	0.59167	0.30000	0.33330	

Isolation and migration

The Mantel test in GENEPOP ON THE WEB revealed extremely no isolation-by-distance between the separate populations present at the Mata Volcanoes (P = 0.91654 with 500,000 permutations).

Estimation of the number of migrants per generation (Nm) using the rare allele method implemented by GENEPOP ON THE WEB and corrected for sample size suggested moderate gene flow of approximately 3.6 migrants per generation when populations consisted of individuals pooled by location. Only one migrant is required per generation to keep two

populations mixed. Thus, because the number of migrants obtained between the Mata Volcano populations is greater than one, it is acceptable to assume that gene flow is occurring between the four populations.

Discussion

This study provided greater clarity regarding the population structure and dynamics of hydrothermal vent communities scattered throughout the Mata Volcanoes, particularly around the *I. nautilei* gastropod species. DNA from 57 *I. nautilei* samples was extracted and amplified using microsatellite markers, and fragment analyses were performed to reveal allelic size differences within and across multiple vent populations. All population genetic analyses revealed that there is no significant population structure between any of the *I. nautilei* populations, including comparisons of populations by depth and volcanoes. Thus, these results produced using microsatellite genetic markers verify and strengthen the results obtained from mitochondrial genetic analyses in Davis et al. (2022). The lack of *I. nautilei* population structure is not surprising due to the ephemeral nature of hydrothermal vents; the rapid appearance and disappearance of vent communities make it extremely difficult for any population to foundationally establish itself at a particular location. Thus, larvae of vent-associated fauna are often capable of wide dispersal to survive.

The genetic variation observed within and among populations aligns with the results of previous population genetic studies using microsatellites for *I. nautilei* (Thaler et al., 2011). The high genetic diversity exhibited by this species can be explained by the high fitness of this marine gastropod species in highly heterogeneous environments, such as hydrothermal vents (Holland, 2001). Due to the ephemeral nature of hydrothermal vents, *I. nautilei*, as well as other vent species, frequently experience bottlenecks, forcing them to migrate, relocate, and reestablish themselves in a new environment. The successful metapopulation properties (recolonization after habitat extinction and habitat creation) have been suggested to strongly influence the high genetic diversity of vent species (Vrijenhoek et al., 1998; Born et al. 2008). Furthermore, the swift and efficient migration of *I. nautilei* also allows for a wide geographical distribution of alleles in a preexisting population, thus promoting genetic diversity by introducing opportunities

for multiple gene pools to expand to incorporate new alleles. This pattern of genetic diversity is consistent with the results obtained in this study.

MICRO-CHECKER verified that there were no null alleles present in this data set. Of all the individuals analyzed, all six of the locus-specific genotypes tested were obtained for all fifty-seven samples. However, if null alleles were present, it is expected that there would be a substantial proportion of null homozygotes at a given locus. However, this is not the case, so population admixture must be considered. If deviations from HWE resulted from inbreeding, heterozygote deficits should be present across most, if not all, loci. However, the overall F_{IS} value across all loci was essentially zero, indicating a fairly even distribution of homozygotes and heterozygotes across all Mata Volcano populations. Elevated F_{IS} values for individual loci at particular volcanoes may result from underlying initial effects resulting from rapid colonization of a new hydrothermal vent region that was then eliminated after one generation of random mating (Vrijenhoek, 1997). Regardless, the lack of genetic disequilibrium suggested that these loci are inherited independently. Furthermore, these populations did not depart from HWE, suggesting that random mating is occurring and there are no selective pressures on the populations.

The results from the principal component analysis (PCA), analysis of molecular variance (AMOVA), and STRUCTURE program all indicated no population structure between any of the Mata Volcano populations and supported high gene flow. The PCA illustrated that while the four volcano populations are clearly distinct from one another, the individuals that make up the four populations overlap noticeably, signifying that no geographical isolation is preventing individuals from separate populations from mixing and interacting with each other (Figure 4). The F_{ST}-based AMOVA tests revealed that there was no population structure when populations were compared based on volcano region (whether the volcano was in the northern or southern half of the Mata Volcanoes), depth ranges, or individual volcano. This implies that *I. nautilei* populations scattered throughout hydrothermal vents at the Mata Volcanoes are not genetically distinct across all depth ranges and volcanoes in the approximate range of the Mata Volcanoes. Regardless of the number of inferred populations, STRUCTURE revealed that there is high gene flow occurring between the populations, as each individual had an equal chance of originating from any of the inferred populations and was not biased towards certain populations (Figure 5).

Genetic variability was also noticeably low within and across the four *I. nautilei* populations at the Mata Volcanoes. Individuals within and across populations did not have clear differences between each other at the genetic level; rather, there was heavy overlap in the specific allele sizes found for the microsatellite loci across all populations. The F_{ST}-based comparisons also correlated to the complete sharing of genetic material within and between populations. No significant genetic differentiation was observed for loci across all populations. Furthermore, no genetic differentiation resulted for pairwise comparisons of the Mata Volcano populations. These analyses indicate that there is no genetic isolation between the *I. nautilei* populations and high levels of gene flow are present. Additionally, the proximity between each of the individual Mata Volcanoes allows for migration and prevents geographical isolation, and there are enough migrants per generation to assume that substantial levels of gene flow are occurring between populations.

The presence of high gene flow and sparse population structure not only confirms the results in Davis et al. (2022) using mitochondrial markers but also aligns with the results of previous studies that implemented microsatellite fragment analysis to study *I. nautilei* populations. Thaler et al. (2011) observed similar patterns of high gene flow in *I. nautilei* populations within the Manus Basin of the West Pacific. Additionally, no genetic differentiation was detected between gastropod individuals from different populations in the North Fiji and Lau Basins. However, a clear restriction of gene flow was observed between the Manus Basin and the North Fiji and Lau Basins. These genetic differences were likely a result of geographical isolation, as the substantial distance of 3500 kilometers that separated the Manus Basin from the other two basins prevented *I. nautilei* larval dispersal from genetically connecting all three basins (Thaler et al., 2011). The results obtained in this study did not illustrate such contemporary restriction of gene flow because the distance between each hydrothermal vent population at the Mata Volcanoes is smaller, ranging no farther than 1.75 kilometers, and *I. nautilei* larvae easily disperse to other vent communities near their origin of dispersal.

Additional studies of vent communities, such as the study performed by Beedessee et al. (2013) at the Central Indian ridge, revealed no genetic differentiation between vent endemic organisms. High connectivity between slow and intermediate (<60 nm/year) spreading ridge systems was revealed, yet these systems had higher variability in chemical and physical vent

activities (Beedessee et al., 2013). As a result, the fluctuations in vent activity along the Central Indian ridge created highly specific niches that promoted niche partitioning among endemic species and maintained high levels of biodiversity. Therefore, if geographical distance is not significantly hindering larval dispersal between two populations, it seems possible that most deep-sea vent populations may not be genetically isolated from each other.

This study also illustrates the success of microsatellites as a viable and useful genetic marker for measuring genetic structure that cannot be initially detected using mitochondrial analyses used in previous studies. The species-specificity of microsatellites makes them a valuable resource in determining genetic differentiation between not just separate populations of a particular species, but also differentiation between individuals within the same population. While the initial success of microsatellites in studying *I. nautilei* populations is encouraging, future studies using the high-resolution microsatellite markers to study the population structure of other endemic vent species, such as the other key gastropod species *A. boucheti* or shrimp species *Opaepele loihi* and *Rimicaris variabilis*, is recommended. Additional analysis of *I. nautilei* populations at the Mata Volcanoes using a greater sample size will also provide a more accurate representation of the gene flow of *I. nautilei* populations within and between these hydrothermal vent communities.

Additionally, the success of the R script in replicating the microsatellite results from the STRand program signifies optimistic implications for efficient analysis in the future. The R script constructed primarily from the *MsatAllele* and *Fragman* packages performed all of the tasks typically performed in STRand, yet it did so in a much more efficient and just as effective manner. While the raw allelic sizes called in the *Fragman* package in R occasionally deviated from the sizes called originally in STRand, these slight fluctuations did not impact the overall results and implications of the population analyses performed. While the accuracy of STRand and the R script is comparable, the efficiency at which it is performed differs immensely: the programming script completes all of the tasks that STRand does, and it also provides additional services to help identify null alleles or clear errors in the peak scoring. Furthermore, provided a viable DNA standard, it offers much more accurate automatic peak calling compared to the automatic peak calling of STRand. It is not necessarily difficult to manually call loci sizes for each sample, but it is certainly time-consuming and tedious. Moving forward, the R script

created and tested in this study should continue to be used to test its viability. If continuous success with other microsatellite markers is observed, it should be implemented as another efficient tool for utilizing microsatellite markers in research.

The research discussed in this study is particularly important in understanding the complex relationships that characterize the communities present in hydrothermal vent ecosystems. In recent years, there has been heightened interest in mining the area around vents for precious metals and abundant minerals that have significant economic value. By acquiring genetic data from vent organisms, it becomes possible to determine if human interference (i.e., mining) or even natural disasters, such as eruptions, have the opportunity to completely destroy these unique species. If populations of certain species are genetically isolated from other populations, various forms of stochasticity, particularly natural catastrophes, could lead to local extinction. Although this study and complementary past studies concluded gene flow between hydrothermal vent systems, it is critical to acquire more data from different sites around the Mata Volcanoes to provide guidance for any future potential mining operations, ensuring that precautionary measures can be implemented to preserve the rich biodiversity that characterizes this unique ecosystem.

Acknowledgments

I would like to deeply thank my mentor and advisor, Dr. Walter Cho, for working alongside me and consistently supporting and challenging me throughout this project. I would also like to express my gratitude to my additional committee members Dr. Andrew Nosal and Dr. Carlson Triebold for their advice, guidance, and instruction in maneuvering through my research. I owe much of my success to the Point Loma Nazarene University Biology Summer Research, as well as the Research Associates, for giving me an opportunity and making my research possible by providing the necessary funding to complete this project. Additionally, I am grateful to the 2022 and 2023 Cho Summer Research Teams for their hard work and support that put me in the position I am in today. Finally, I would like to extend my thanks to Dr. Mark Mann for allowing me to participate in the PLNU Honors Program to be able to share my research.

References

Alberto F. (2009). MsatAllele_1.0: An R package to visualize the binning of microsatellite alleles. *The Journal of heredity*, *100*(3), 394–397. https://doi.org/10.1093/jhered/esn110.

Amend, J. P., Rogers, K. L., & Meyer-Dombard, D. R. (2004). Microbially mediated sulfur-redox: Energetics in Marine Hydrothermal Vent Systems. *Sulfur Biogeochemistry - Past and Present*. https://doi.org/10.1130/0-8137-2379-5.17.

Barton, N. H., & Slatkin, M. (1986). A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity*, *56 (Pt 3)*, 409–415. https://doi.org/10.1038/hdy.1986.63.

Beedessee G, Watanabe H, Ogura T, Nemoto S, Yahagi T, Nakagawa S, et al. (2013) High Connectivity of Animal Populations in Deep-Sea Hydrothermal Vent Fields in the Central Indian Ridge Relevant to Its Geological Setting. PLoS ONE 8(12): e81570. https://doi.org/10.1371/journal.pone.0081570.

Born, C., Kjellberg, F., Chevallier, M.-H., Vignes, H., Dikangadissi, J.-T., Sanguié, J., Wickings, E. J., & Hossaert-McKey, M. (2008). Colonization processes and the maintenance of genetic diversity: Insights from a pioneer rainforest tree,*Aucoumea Klaineana*. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1647), 2171–2179. https://doi.org/10.1098/rspb.2008.0446.

Clague, D.A., Paduan, J.B., Caress, D.W., Thomas, H. Chadwick, W. W. Jr., and Merle, S. G. (2011). Volcanic morphology of west mata volcano, NE lau basin, based on high-resolution bathymetry and depth changes. *Geochem. Geophys. Geosyst.* 12:QOAF03. doi: 10.1029/2011GC003791.

Covarrubias-Pazaran, G., Diaz-Garcia, L., Schlautman, B., Salazar, W., & Zalapa, J. (2016). Fragman: an R package for fragment analysis. *BMC genetics*, *17*, 62. https://doi.org/10.1186/s12863-016-0365-6.

Davis, K., Mauch, H., and Cho, W. (2022) Population Genetics of Gastropods at Hydrothermal Vents at the Mata Volcanoes. 45th Annual West Coast Biological Sciences Undergraduate Research Conference. San Diego, CA.

Excoffier, L., Laval, G., & Schneider, S. (2007). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, *1*, 47–50.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587. https://doi.org/10.1093/genetics/164.4.1567.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes*, *7*(4), 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x.

Fusaro, A. J. (2008). Spatial and Temporal Population Genetics at Deep-Sea Hydrothermal Vents along the East Pacific Rise and Galapagos Rift. https://doi.org/10.1575/1912/2564.

Garnier, S., Alibert, P., Audiot, P., Prieur, B., & Rasplus, J. Y. (2004). Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, Carabus solieri. *Molecular ecology*, *13*(7), 1883–1897. https://doi.org/10.1111/j.1365-294X.2004.02212.x.

Goudet, J. (1999) PCA-GEN for Windows, V. 1.2. Distributed by the author, Institute of Ecology, Univ. of Lausanne.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Updated from Goudet (1995).

Holland B. S. (2001). Invasion without a bottleneck: Microsatellite variation in natural and invasive populations of the brown mussel Perna perna (L). *Marine biotechnology (New York, N.Y.)*, *3*(5), 407–415. https://doi.org/10.1007/s1012601-0060-z.

Holleley, C. E., & Geerts, P. G. (2009). Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques*, *46*(7), 511–517. https://doi.org/10.2144/000113156.

O'Donnell, J. (2014). *Interpreting structure plots*. Jimmy O'Donnell. https://jimmyodonnell.wordpress.com/2014/05/25/interpreting-structure-plots/.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICROCHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945.

Reynolds, K. C., Watanabe, H., Strong, E. E., Sasaki, T., Uematsu, K., Miyake, H., Kojima, S., Suzuki, Y., Fujikura, K., Kim, S., & Young, C. M. (2010). New molluscan larval form: brooding and development in a hydrothermal vent gastropod, Ifremeria nautilei (Provannidae). *The Biological bulletin*, *219*(1), 7–11. https://doi.org/10.1086/BBLv219n1p7.

Rice W. R. (1989). ANALYZING TABLES OF STATISTICAL TESTS. *Evolution; international journal of organic evolution*, *43*(1), 223–225. https://doi.org/10.1111/j.1558-5646.1989.tb04220.x .

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*(3), 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573.

Rousset F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1093/genetics/145.4.1219.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, *9*(5), 615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x.

Slatkin M. (1985). RARE ALLELES AS INDICATORS OF GENE FLOW. *Evolution; international journal of organic evolution*, *39*(1), 53–65. https://doi.org/10.1111/j.1558-5646.1985.tb04079.x.

Slatkin M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457–462. https://doi.org/10.1093/genetics/139.1.457.

Thaler, A. D., Zelnio, K., Jones, R., Carlsson, J., Van Dover, C. L., & Schultz, T. F. (2010). Characterization of 12 polymorphic microsatellite loci in Ifremeria nautilei, a chemoautotrophic gastropod from deep-sea hydrothermal vents. *Conservation Genetics Resources*, *2*(1), 101–103. https://doi.org/10.1007/s12686-010-9174-9.

Thaler, Andrew D., Zelnio, K., Saleu, W., Schultz, T. F., Carlsson, J., Cunningham, C., Vrijenhoek, R. C., & Van Dover, C. L. (2011). The spatial scale of genetic subdivision in populations of Ifremeria nautilei, a hydrothermal-vent gastropod from the Southwest Pacific. *BMC Evolutionary Biology*, *11*(1), 372. https://doi.org/10.1186/1471-2148-11-372.

Toonen, R. J., & Hughes, S. (2001). Increased throughput for fragment analysis on an ABI PRISM 377 automated sequencer using a membrane comb and STRand software. *BioTechniques*, *31*(6), 1320–1324.

Vrijenhoek R. C. (1997). Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *The Journal of heredity*, *88*(4), 285–293. https://doi.org/10.1093/oxfordjournals.jhered.a023106.

Vrijenhoek, R., Shank, T., & Lutz, R.A. (1998). Gene flow and dispersal in deep-sea hydrothermal vent animals. *Cahiers de Biologie Marine*. 39. 363-366.

Waren, A., & Bouchet, P. (1993). New Records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps*. *Zoologica Scripta*, *22*(1), 1–90. https://doi.org/10.1111/j.1463-6409.1993.tb00342.x.

Waren, A., & Ponder, W. F. (1991). New Species, anatomy, and systematic position of the hydrothermal vent and hydrocarbon seep gastropod family Provannidae fam.n. (Caenogastropoda). *Zoologica Scripta*, *20*(1), 27–56. https://doi.org/10.1111/j.1463-6409.1991.tb00273.x.

Weir, B. S. (1997). Genetic Data Analysis II. *Biometrics*, *53*(1), 392. https://doi.org/10.2307/2533134.

Weir, B. S., & Cockerham, C. C. (1979). Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, *42*(1), 105–111. https://doi.org/10.1038/hdy.1979.10.

Weir, B. S., & Cockerham, C. C. (1984). ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. *Evolution; international journal of organic evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.

Wright S. (1951). The genetical structure of populations. *Annals of eugenics*, *15*(4), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x.