# POPULATION COMPOSITION OF AN EXPLOITED HAWAIIAN FISHERY

A Thesis

by

PATRICIA MALAMALAMA COCKETT

Submitted to the Office of Graduate and Professional Studies of Texas A&M University
and the Graduate Faculty of The Texas A&M University – Corpus Christi
in partial fulfillment of the requirements for the joint degree of

MASTER OF SCIENCE

| | |
|---|---|
| Chair of Committee, | Christopher E. Bird |
| Committee Members, | Derek J. Hogan |
| | Jonathan B. Puritz |
| Head of Department, | Anna Armitage |

May 2015

Major Subject: Marine Biology

# POPULATION COMPOSITION OF AN EXPLOITED HAWAIIAN FISHERY

A Thesis

by

PATRICIA MALAMALAMA COCKETT

Submitted to the Office of Graduate Studies of Texas A&M University
and Texas A&M University - Corpus Christi
in partial fulfillment of the requirements for the joint degree of

MASTER OF SCIENCE

Approved as to style and content by:

5/6/15

Dr. Christopher E. Bird
(Chair of Committee)

Dr. Derek J. Hogan
(Member)

Dr. Jonathan B. Puritz
(Member)

Dr. Joe Fox
(Department Chair)

Dr. Frank Pezold
(Dean, College of Science & Engineering)

May, 2015

Major Subject: Marine Biology

## ABSTRACT

POPULATION COMPOSITION OF AN EXPLOITED HAWAIIAN FISHERY

Patricia Malamalama Cockett, B.A., University of Hawai‘i at Mānoa

Chair of Advisory Committee: Dr. Christopher E. Bird

Overharvesting has been implicated in altering the population structure of marine organisms, reducing genetic diversity and adaptive capacity. Overharvested fisheries can be particularly vulnerable to environmental and anthropogenic stressors due to the loss of advantageous mutations. The Hawaiian broadcast-spawning limpet, *Cellana exarata*, is subject to varying levels of harvesting pressure on different islands, ranging from no harvest on the uninhabited island of Nihoa in the Papahānaumokuākea Marine National Monument, to a reduction in population density on Maui and Kaua‘i, to near extirpation on O‘ahu, the most populous of the Hawaiian Islands. In this study, we use genome-wide surveys of genetic variation (ezRAD, >21,000 loci) on *C. exarata* from the islands of Nihoa, Kaua‘i, Maui, O‘ahu, and Hawai‘i to test for relationships between genetic diversity, population size, island age, and harvest pressure. Global estimates of genetic differentiation among islands are greater than those estimated with mtDNA. Pairwise comparisons among islands indicate a substantial difference in genetic composition between the inhabited Main Hawaiian Islands and the uninhabited island, Nihoa. Estimates of nucleotide diversity ($\pi$) were greatest on Nihoa ($\pi = 2.05 \times 10^{-3}$), despite

having the smallest estimated population size (without harvesting); estimates of nucleotide diversity on the Big Island of Hawai'i are the lowest of all the islands in this study ($\pi = 1.71 \times 10^{-3}$), despite having the largest estimated population size. This difference in genetic diversity, while initially counter intuitive, is correlated with island age and indicates that *C. exarata* populations within the MHI experienced a recent bottleneck. Overall, these results suggest that the PMNM harbors a stockpile of genetic diversity for *C. exarata,* despite relatively small population sizes when compared to the MHI.

**Table of Contents**

**List of Tables and Figures**

**Acknowledgements**

I thank the Pacific Missile Range Facility, Texas A&M University – Corpus Christi, the TOBO lab, Papahānaumokuākea National Marine Monument, The Nature Conservancy, and ʻOpihi Partnership. My warmest aloha and mahalo to the Kapiʻolani Community College STEM program for being my foundation. Thank you Anty Keolani Noa for helping me find the courage to pursue my dream of studying ʻOpihi. Thank you Dr. Hadfield and Wendy Kuntz for your guidance and support. I thank the HOBI lab in all of their glory. Thank you Sharon Furiness for all of your manaʻo. A special thanks to all the rock star undergraduate students in the HOBI lab. Mahalo nui loa to Alin Gonzalez, Lauren Gurski, and Luza. Thanks Mom and Dad, you have been great.

## Introduction

Overharvesting of a fishery removes alleles from populations causing permanent changes in genetic composition (Kenchington *et al.* 2003; Kenchington 2003; Ryman *et al.* 1995). These genetic changes exacerbate the risk of extinction, reduce resiliency, and may select for earlier maturation rates (Hutchings 2000; Law 2007; Olsen *et al.* 2004; Walsh *et al.* 2006). Changes in heritable traits in fish (e.g. weight at age, length at age, spawning season etc.) following a collapse have been well documented in Atlantic cod, haddock and Pollock stocks (Hutchings 2000; Swain *et al.* 2007; Trippel *et al.* 1997; Wright 2005; Wright & Tobin 2013). Swain et al. (2007) reported a decrease in the average size and average age of Pacific Salmon, attributing the changes to harvesting practices and pressures. However, there are few studies that directly demonstrate a reduction in genetic diversity in a fishery due to overharvesting and even fewer that demonstrate a loss of genetic variation in invertebrate populations. The few examples include New Zealand snapper (Hauser *et al.* 2002) and the Atlantic Cod (Hutchinson *et al.* 2003). Each exhibited slightly lower levels of microsatellite allelic variation following collapses in these fisheries.

Gene flow can be an important determinant of genetic diversity in addition to effective population size, selection and genetic mutation rate (Kenchington 2003). For example, the exploited population of black-lip abalone from Tasmania, Australia displayed higher levels of microsatellite allelic variation in contrast to populations that have not been overharvested (Miller *et al.* 2009). One possible explanation for the increased genetic variation of the abalone in the collapsed population is that gene flow from other populations introduced new alleles as habitat became available for new

recruits (Miller *et al*. 2009). This underscores the importance of identifying subpopulations within a fishery that can be sources of gene flow for declining populations.

If an overexploited population is isolated, meaning there is no introduction of new alleles from other populations (gene flow), then there will be reductions in genetic diversity, therefore lowering that population's ability to adapt to environmental changes (Hutchings 2000; Jackson *et al.* 2001; Smith *et al.* 1991). An exploited population's capacity to recover after a decline can be related to changes in allele frequencies or even the loss of alleles (Frankham *et al.* 2002). Populations that have collapsed can yield low productivity because of reduced fitness or decreased allelic diversity (Allendorf *et al.* 2008; Guinand *et al.* 2003; Hauser & Carvalho 2008; Ryman *et al.* 1995; Wang *et al.* 2012) thereby relying on other source populations for production and genetic diversity.

The collapse of the North Sea cod exemplifies the need for the conservation of genetic variation in the face of a changing climate (O'Brien *et al.* 2000). The number of cod was reduced to near extirpation, causing the individuals within the population to be much smaller and less fecund. Extreme environmental changes may be the *coup de grace* for the North Sea cod. The ability of species to adapt to new environmental conditions is critical in the next few centuries as the world's human population continues to rise. This will inevitably lead to an increase in intensity of human impacts on exploited fisheries such as, extreme exploitation of natural resources, anthropogenic pollution, climate change and habitat loss (Halpern *et al.* 2009; Halpern *et al.* 2008; Selkoe *et al.* 2009). The world's fisheries are in decline and will continue to decline without improved management practices (Thorpe *et al.* 2000).

Advances in population genetic techniques, such as restriction site associated DNA sequencing (RADseq, Miller *et al.* 2007), allow the precise identification of subpopulations, inference of migration rates, and quantification of genetic diversity by measuring allele frequencies among population samples and heterozygosity within population samples using tens of thousands of genetic markers. RAD markers are a repeatable, yet semi-randomly sampled loci across the genome. By identifying spatial patterns of gene flow and genetic diversity, population genetic analysis can be used to determine the spatial scale at which management plans should be executed and can be used to develop detailed management plans, such as a network of harvesting refuges, or pu‘uhonua.

Hawaiian endemic limpets (*Cellana* spp.), locally referred to as ‘opihi, are a culturally and economically important fishery that collapsed in the early 1900’s (Kay & Magruder 1977). The fishery has not recovered since, and populations continue to decline (~150,000 lbs./year in the 1900’s to ~10,000 lbs./year in 1944). *Cellana exarata* are one of three ‘opihi species that occupy the Hawaiian rocky intertidal shores. They are broadcast-spawning limpets that release their gametes into the water column followed by a ~2-18 day pelagic larval duration (Bird *et al.* 2007; Corpuz 1983) which is indicative of a population with the potential for high levels of gene flow among islands.

In population genetic analyses of a single mtDNA marker (Bird *et al.* 2007) and six nDNA microsatellites (Pennoyer et al., in prep), the *C. exarata* population exhibits a strong gene-flow restriction between the inhabited main Hawaiian Islands (MHI) and the uninhabited Northwestern Hawaiian Islands (NWHI), also known as Papahānaumokuākea Marine National Monument (PMNM). mtDNA population genetic

structure was much weaker within the MHI and NWHI, with some islands exhibiting no differentiation. With the six microsatellites, gene flow restrictions were identified among the primary MHI, but little genetic structure was detected among the NWHI. Despite the amount of population genetic data for *C. exarata*, there is no information regarding the genetic composition of *C. exarata* on the island of Oʻahu, where populations have been harvested to near extirpation. There are varying levels of harvesting pressure among Hawaiian islands because each island has a different human populous and a different amount of viable ʻopihi habitat.

The relationship between humans and the amount of ʻopihi on each island can be described by the meters of coastline divided by the human census population, meters of habitat per capita. The meters of *C. exarata* habitat per capita are related to the density of *C. exarata* as measured in transect surveys for four of the main Hawaiian islands (MHI) (Fig. 2, Bird et al, in prep), thereby demonstrating a potential link between human population, harvest pressure, and ʻopihi populations. It should be noted that the density of ʻopihi in the NWHI, where harvest is prohibited, is approximately $30/m^2$.

Identification of *C. exarata* subpopulations and their genetic diversity will provide Hawaii's managers with the required information to manage the limpet fishery on a population level as well as provide insight for the biology of broadcast spawning limpets. There are currently minimal regulations for the harvesting of Hawaiian *Cellana* spp.; legal size of collection is 3.1 cm, with no other restrictions. Despite eight years of attempts by the Hawai'i State Senate, new regulations for the ʻopihi fishery have not been passed. One reason for the failure of legislation is the lack of scientific evidence demonstrating that humans are impacting ʻopihi populations.

In this study, we test for a relationship between genetic diversity and: (1) island age (2) amount of habitable *C. exarata* coastline per capita (harvesting pressure) (3) the amount of habitable coastline. A decline in genetic diversity of *C. exarata* with increasing human population size (decreasing habitable coastline per capita) would be a strong indication that humans are affecting the ‘opihi and their ability to recover. Additionally, we test for gene flow restrictions among Nihoa, Kaua‘i, O‘ahu, Maui and the Big Island of Hawai‘i using ~22,000 loci and compare our results to that of Bird et al. (2007) which assessed population structure among islands using one mtDNA locus.

## Methods

### *Collection sites*

*Cellana exarata* reside exclusively on rocky intertidal shoreline in the Hawaiian archipelago. Forty-eight individuals were collected in 2013 from one site on each of four main Hawaiian Islands (MHI) and one Northwestern Hawaiian Island (NWHI) from Southeast to Northwest: Hilo break wall, Big Island of Hawai‘i (BI); Waihe‘e, Maui (MA); Ka‘ena, O‘ahu (OA); Ke‘e, Kaua‘i (KA); and Adam’s Bay, Nihoa (NI, Fig. 1).

### *Island characteristics*

The Hawaiian Islands decrease in size and increase in age from the Southeast to the Northwest, from the Big Island of Hawai‘i to Nihoa (Clague 1996). Island ages were obtained from Clague (1996). The amount of *C. exarata* habitat on each island was estimated by multiplying the circumference of each island by the proportion of rocky shoreline on each island (Hawai’i DLNR Report, Papahānaumokuākea Marine National Monument). The human population on each island was obtained from the 2014 United States Census (www.census.gov). The peak human population size on Nihoa (175) was obtained from www.hawaiianatolls.org. The length of habitable coastline per capita was estimated by dividing the length of coastline that is habitable by *C. exarata* by the number of people residing on each island (175 for Nihoa).

### *Sample processing*

*Cellana exarata* were dissected immediately after collection, and DNA was isolated from ~150 mg of mantle tissue with the Omega E-Z 96 ® Tissue DNA Kit. A portable extraction kit including a vacuum pump, vacuum manifold, Eppendorf ThermoMixer,

and Rainin p-200 and p-1000 multichannel pipette were used to perform extractions on live tissue to maximize the yield of high molecular weight DNA, which is critical to the success of ezRAD (Toonen *et al.* 2013). DNA samples were transported to the Hawai'i Institute of Marine Biology, and lyophilized before final transport to the laboratory at Texas A&M University – Corpus Christi.

*Quantification and RADseq library preparation*

All DNA samples were subjected to gel electrophoresis on a 1% agarose gel in 1x Tris-acetate-EDTA (TAE) to assess the length distribution of isolated DNA fragments. SPRIselect paramagnetic beads were used to isolate high molecular weight DNA using a 0.4x bead: sample reaction ratio. Size-selected samples were re-inspected via gel electrophoresis and used for library preparation if low molecular weight DNA was successfully removed. DNA concentration was quantified with the AccuBlue High-Sensitivity fluorescence assay on a SprectraMax M3 plate reader according to the manufacturer's protocol, and DNA concentrations were equalized. Normalized DNA samples were pooled (6 - 12 individuals per pool, 3 - 4 four pools per site, Table 1) for a total of 80 ng of DNA in each pooled sample prepared with the Illumina TruSeq Nano DNA LT kit and 500 ng of DNA in each pooled sample prepared with the Illumina TruSeq DNA - PCR free LT kit. RAD libraries were prepared according to the ezRAD protocol (Toonen *et al.* 2013) with modifications to reduce cost and DNA loss. Pooled DNA samples (16.67 µl) were purified with 33.3 µl of AmpureXP beads, eluted with 16.9 µl of nanopure water and digested with the isoschizomers MboI and Sau3AI in 25 µl reactions. Following Fisher et al. (2011), the AmpureXP paramagnetic beads were not removed from the samples, and were reactivated with 3M NaCl in 20% polyethylene

glycol at the end of each reaction step. Digested samples were inserted into the Illumina TruSeq DNA Sample Prep Kit protocol (immediately after the DNA shearing step). Sites Nihoa, Kauaʻi, Oʻahu, and Maui were PCR amplified but the Big Island was not subjected to PCR amplification. Libraries were quantified using the Kapa qPCR Library Quantification Kit; equal concentrations were pooled, and submitted for paired-end 125 bp sequencing to the Genomic Sequencing and Analysis Facility (University of Texas at Austin) on an Illumina HiSeq 2500.

*Data Processing*

Sequence quality and characteristics were assessed using FASTQC (Babraham Bioinformatics). The dDocent bioinformatics pipeline (Puritz *et al.* 2014) was used to quality control RADseq data, generate *de novo* reference RAD contigs, map sequence reads to reference contigs, and call single nucleotide polymorphisms (SNPs), with some modifications to accommodate pooled ezRAD sequences. One sample from each island was used to construct a single *de novo* reference "genome" with *Rainbow* (Chong *et al.* 2012). In order for sequence reads to be used in the construction of the reference, Trim Galore! (Babraham Bioinformatics) was used to ensure they contained no adapter, had a per base Phred quality score above 15, length of 118 bp, and were present in at least 3 samples with at least 3 supporting reads. For reads mapped to the reference, adapter and restriction site sequence were removed, only sequence with a per base Phred quality score above 20 and a length greater than 50 bp were retained for mapping with *BWA* (Li & Durbin 2010). SNPs were then called using *freebayes* with the "--pooled-continuous" argument, which tallies read frequency rather than assuming a particular ploidy (Garrison & Marth 2012). Loci with a mean total read depth less than 20x among samples were

removed. Loci were retained for further analysis if they met the following criteria: present in all samples, maximum read frequency > 0.1 in at least one sample, biallelic, had two or more polymorphic samples. The remaining loci were filtered in each sample, independently, with respect to depth of read coverage using R (R Core Team, 2014, Fig. 3). Specifically, all loci with read depths greater than double the sample-specific mode read depth were removed. This effectively removed the long right tails of the read depth distributions. Then, 2.5 % of the remaining loci with the highest and the lowest read depth were removed from each sample (5% total per sample). Finally, any loci not represented in all samples were removed. The remaining loci were used for subsequent genetic statistical analyses. All loci were allelotyped, where sequence read frequencies were used to estimate the most likely allele frequencies within each of the pooled samples.

*Analysis of Genetic Differentiation*

An analysis of molecular variance (AMOVA), coded in the R statistical package, was used to estimate two of Wright's (1951) *F*-statistics: $F_{CT}$ and $F_{SC}$ (Excoffier *et al.* 1992). Due to the sampling design employed here, $F_{CT}$ is an overall measure of genetic fixation and differentiation among islands (Bird *et al.* 2011), and $F_{SC}$ is a measure of differentiation among samples nested within islands, where multiple samples were collected from a single site on each island (Table 1). Following Weirand Cockerham (1984), *F*-statistics were calculated across several loci by independently summing the variance components associated with island, sample, and within sample.

In addition to global estimates of $F_{CT}$ and $F_{SC}$, pairwise estimates of $F_{CT}$ and $F_{SC}$ were calculated among islands in order to identify the island populations driving the global

genetic structure. Pairwise $F_{CT}$ values were regressed against cumulative stepping-stone distance between islands (Bird *et al*. 2007). A mantel test was performed to assess whether genetic differentiation conforms to an isolation by distance model using the Isolation by distance web service (Wright 1943; Jenson *et al.* 2005).

In an effort to reduce computational expense, we estimated the number of loci required to consistently estimate $F_{CT}$ and $F_{SC}$ with minimal variation. To accomplish this, 1, 10, 100, 1000, 10,000 and 22,474 loci (1 SNP per locus) were randomly drawn from the data set and the F-statistics were calculated. This procedure was repeated 1000 times to obtain estimates of the mean and variance associated with the *F*-statistics given the number of loci interrogated. The mean *F*-statistics and variances were plotted against the number of loci to determine an acceptable level of precision, as indicated by an asymptote (Fig. 5).

To test for a non-random pattern of genetic differentiation in pooled samples, it is critical to model the error associated with both the sampling of individuals from the population and the sampling of sequence reads from pooled samples. For the test of $F_{CT}$ and $F_{SC}$, a standard null resampling procedure for a nested AMOVA is employed that is similar to that described in Excoffier *et al.* (1992), with the exception that bootstrapping is used rather than permutation (Fig. 4) and we subsampled loci (see previous paragraph). (1) A sample of k loci (1 SNP per locus, k is a constant) are drawn from the samples to be compared and the observed F-statistics are calculated; (2) the sample allele frequencies of each locus from each sampling site are calculated from the data and used as estimates of the true frequencies; (3) two alleles per individual are randomly sampled from each island to obtain simulated allele frequencies for each pooled sample; (4) the observed

number of sequence reads are randomly sampled from each pooled sample; (5) each locus in each sample is allelotyped; (6) the null $F_{SC}$ is calculated and recorded; (7) the samples are randomly assigned to islands and null $F_{CT}$ is calculated; and (8) this procedure is repeated from Step 1 in order to generate the observed and null distributions of the *F*-statistics. The statistical significance of the observed $F_{CT}$ and $F_{SC}$ values is determined by comparison of the mean observed values to their null distributions (Excoffier *et al.* 1992).

*Analysis of Genetic Diversity*

Nucleotide diversity (π, Nei and Li 1979) was calculated for each island at each sequence locus according to the following equation:

$$\pi = \frac{2n \times \sum pq}{(n-1) \times sequence\ length}$$

where n is the number of alleles sampled, sum pq refers to the summation of the products of reference and the variant allele frequencies (after allelotyping) for all polymorphic positions in the sequence locus, and j is the sequence length, in base pairs. There was a significant relationship between number of individuals in a pool and π, therefore the alleleotypes from pools within islands were combined such that approximately 30 individuals per island were used in the calculation of π (Table 1). After performing this procedure, there was no relationship between the number of individuals and nucleotide diversity.

In addition to nucleotide diversity, the proportions of sequence loci that were polymorphic at each island were estimated by dividing the number of polymorphic loci by the total number of loci. Both the nucleotide diversity and the polymorphic proportion of loci were used to test for differences in genetic diversity among islands using the

statistical software packages R and *JMP* Pro (Version 11.2.0). Specifically, the polymorphic proportion of loci was used to assess overall levels of genetic diversity across all loci, while $\boldsymbol{\pi}$ was used to assess the degree of diversity exhibited in polymorphic loci. The estimates of genetic diversity (polymorphic proportion of loci and mean $\pi$ in polymorphic loci) were regressed against three predictor factors that are thought to affect genetic diversity: island age, amount of habitat, and harvest pressure. The mean estimate of $\boldsymbol{\pi}$ and the values for each factor were log transformed to meet the assumptions of least squares linear regression and the F-distribution was used to test if each slope was significantly different from zero (Zar 1984). To gain better insight of how each locus responds to changes in nucleotide diversity with respect to island age, we regressed $\boldsymbol{\pi}$ against island age for each locus and tabulated the number of loci with each specific bin of $R^2$ values. Additionally, we calculated the proportion of loci with a positive and negative slope within each bin of $R^2$ values.

A post-hoc Tukey HSD test was employed to identify differences in $\boldsymbol{\pi}$ among islands. Finally, each predictor factor was log transformed and Pearson product-moment correlation coefficients were computed to investigate collinearity between each predictor factor and each other predictor factor using the statistical software package *JMP*.

## Results

### *SNP Discovery*

Raw sequence reads input to the dDocent variant calling pipeline, produced 159,677 SNPs within 48,764 loci. Removing loci with <20x mean coverage, multiallelic loci, and any loci that were not present in all samples yielded 156,042 SNPs within 48,764 loci. After removing SNP loci with the sample-specific filter for high and low depth of coverage, there were 70,014 SNPs within 23,453 loci (Fig. 3). After removing SNPs with a variant frequency <0.1 in at least one sample and SNPs that were present in at least two samples, there were 59,730 SNPs remaining within 22,474 loci. Following adjustments made to include ~30 individuals per site by randomly selecting samples within each site for genetic diversity analyses, 21,529 loci remained.

### *Analysis of Molecular Variance*

In order to evaluate the precision and accuracy of *F*-statistic estimates generated with SNPs, the mean global $F_{CT}$ and $F_{SC}$ were plotted against number of randomly drawn loci (one randomly drawn SNP per locus, Fig. 5). The mean $F_{CT}$ and $F_{SC}$ both increased with an increasing number of loci, from 1 to 100, before reaching an asymptote of $F_{CT} = 0.07$ and $F_{SC} = 0.002$ for 1,000 through 22,474 loci. The standard deviation of the mean *F*-statistic decreased with increasing numbers of loci for both $F_{CT}$ and $F_{SC}$, as expected. One-thousand randomly sampled loci was deemed to be sufficient to calculate $F_{CT}$ and $F_{SC}$ values with an acceptable amount of error for this study ($stdev_{FCT}$=0.012, $stdev_{FSC}$=0.004). From here forward, all reported *F*-statistics are means obtained from 1000 randomly drawn sets of 1000 loci.

The overall population genetic structuring detected among islands was significantly greater than zero ($F_{CT}$ = 0.071, P < 0.001, Table 2), and there was no detectable genetic differentiation among replicate sample pools ($F_{SC}$ = 0.002, P = 0.811, Table 2). Pairwise estimates of genetic differentiation ($F_{CT}$) between Nihoa and each MHI (Kaua‘i, O’ahu, Maui, Big Island of Hawai‘i) are consistently and substantially greater than estimates among the MHI (mean $F_{CTpairwiseNIvsMHI}$ = 0.164, stdev = 0.01, versus mean $F_{CTpairwiseMHIvsMHI}$ = 0.015, stdev = 0.01; Fig. 6). Within the MHI, the island of Maui exhibited a small amount of genetic differentiation from the other MHI, with a mean pairwise $F_{CTpairwiseMAvsMHI}$ = 0.024 (stdev = 0.01). For all other pairwise comparisons among the MHI, however, $F_{CT}$ < 0.01, thereby indicating very little genetic structuring (Fig. 6). There was no pattern of isolation by distance when regressing pairwise $F_{CT}$ against cumulative stepping-stone distance between islands. There is a positive linear relationship between $F_{CT}$ and stepping-stone distance, however, the slope is not significantly greater than zero ($r^2$ = 0.25, P = 0.125).

In contrast, pairwise estimates of genetic differentiation among samples collected within islands ($F_{SC}$) were lower than the $F_{CT}$ values (maximum $F_{SC}$ = 0.0164, minimum $F_{SC}$ = - 0.0084), with one exception on the island of Kaua‘i (Fig. 6). The $F_{SC}$ values in pairwise comparisons between Kaua‘i and the other MHI were generally similar, but greater than the $F_{CT}$ values ($F_{SCpairwiseKAvsMHI}$ = 0.012, StDev = 0.005 versus $F_{CTpairwiseKAvsMHI}$ = 0.006, StDev = 0.003).

*Genetic Diversity Among Islands*

Despite filtering out several thousands of monomorphic loci and combining pooled samples within islands (Table 1), an average of 53% of the all the remaining loci within

each island were invariant, where $\pi = 0$. Consequently, two summary indices of genetic diversity are presented: the proportion of loci within each island that are polymorphic and the mean nucleotide diversity of the polymorphic loci within each island.

The polymorphic proportion of loci was greatest on the island of Nihoa (0.630), and decreased progressively on each island from Northwest to Southeast: Kaua‘i (0.533), O‘ahu (0.504), Maui (0.495), and the Big Island of Hawai‘i (0.491, Fig. 7). Almost all of the variation in the proportion of loci that are polymorphic can be explained by the age of the island ($r^2$=0.998, $P < 0.0018$, Fig. 8), where older islands have more polymorphic loci. Large portions of the variation in the proportion of loci that are polymorphic can also be explained by the amount of habitat, in linear meters of habitable coastline ($r^2$=0.98, $P = 0.0066$, negative relationship, Fig. 9), and the amount of habitable coastline per person ($r^2$=0.92, $P < 0.011$, positive relationship, Fig. 10). There is a negative linear relationship between the amount of habitable coastline and polymorphic proportion of loci, and there is a positive linear relationship between the amount of habitable coastline per capita and the polymorphic proportion of loci.

It is clear that there are correlations among the predictor variables employed here (Fig. 11). There was a negative correlation between the amount of habitable coastline per capita and the amount of habitable coastline, where the amount of habitat generally decreases as the amount of habitat per capita increases (r = -0.86, Fig. 11.A). There was a positive correlation between the age of the islands and the amount of habitable coastline per capita, in which the amount of habitat per capita increases as island age increases (r = 0.47, Fig. 11.B). There was a negative correlation between the amount of habitable

coastline and the age of the islands, where the amount of habitable coastline decreases as island age increases (r = -0.81, Fig. 11.C).

Mean nucleotide diversity ($\pi$) of the polymorphic loci was highest on Nihoa ($\pi = 2.05 \times 10^{-3}$), followed by Maui ($\pi = 1.87 \times 10^{-3}$ ), O‘ahu ($\pi = 1.78 \times 10^{-3}$ ), Kaua‘i ($\pi = 1.75 \times 10^{-3}$ ), and finally Hawai‘i ($\pi = 1.71 \times 10^{-3}$, Table 1, Fig. 12). Unlike with the polymorphic proportion of loci, *post hoc* tests using the Tukey HSD could be calculated on $\pi$ because there were many observations of $\pi$ for each island. *Post hoc* tests indicate that the mean nucleotide diversity on Nihoa is significantly greater than that on Maui ($P < 0.0001$), $\pi$ on Maui is significantly greater than that on Kaua‘i ($P < 0.0001$) and O‘ahu ($P < 0.0001$), and $\pi$ on Hawai‘i is significantly less than that on the other islands ($P < 0.0001$). A regression of the mean nucleotide diversity of polymorphic loci plotted against island age reveal a positive relationship, where mean nucleotide diversity increases as island age increases ($r^2 = 0.0037$, $P < 0.0001$, Figure 13). Nucleotide diversity of polymorphic loci regressed against the amount of habitable coastline exhibit a negative relationship, in which nucleotide diversity decreases as the amount of habitable coastline increases ($r^2 = 0.0073$, $P < 0.0001$, Fig. 14). There was a positive relationship between nucleotide diversity and the amount of habitable coastline per person, where nucleotide diversity increases as the amount of habitable coastline per person increases ($r^2 = 0.0058$, $P < 0.0001$, Fig. 15). All three predictor factors had a significant linear relationship with nucleotide diversity. Mean nucleotide diversity for each island are as follows: Nihoa, 95%CI: $2.0310^{-3}$ to $2.0710^{-3}$; Kaua‘i, 95%CI: $1.7310^{-3}$ to $1.7810^{-3}$; O‘ahu, 95%CI: $1.7610^{-3}$ to $1.8010^{-3}$; Maui, 95%CI: $1.8610^{-3}$ to $1.9010^{-3}$; Big island of Hawai‘i, 95%CI: $1.6910^{-3}$ to $1.7410^{-3}$.

It is important to recognize that while the diversity results presented thus far are averages, there was a lot of variation in the pattern of genetic diversity among islands for each locus. For nucleotide diversity regressed against island age for each locus (including both polymorphic and monomorphic loci), ~9000 loci have a low coefficient of determination ($r^2$ = 0-0.2) compared to the number of loci (~1000) with a high coefficient of determination ($r^2$ = 0.8-1, Fig. 16). There was a fair amount of loci (~5000) with a coefficient of determination between 0.6 and 0.8 with remaining loci (~7250) exhibiting a coefficient of determination between 0.2 and 0.6. For loci where the relationship between diversity and island age is not significantly different than zero, there is an expectation that there will be equal numbers of loci with positive and negative slopes. Indeed, loci with a coefficient of determination between 0 and 0.4 show even proportions of positive (~50%) and negative slopes (~%50). The proportion of loci with positive slopes increases with the coefficient of determination: ~65% of loci have a positive slope when $r^2$ = 0.4-0.6 and ~80% of loci have a positive slope when $r^2$ = 0.6-0.8. However, only ~75% of the loci with the highest coefficient of determination values ($r^2$ = 0.8-1) have a positive slope. In other words, the nucleotide diversity for a substantial number of loci (>1500) are strongly negatively related to island age, and are thus positively related to the amount of habitat.

## Discussion

### *Genetic Differentiation Among Islands*

Analysis of thousands of loci in *C. exarata* indicates that there is population structure between the MHI and the NWHI as previously documented with mtDNA (Bird *et al.* 2007). These results support the conclusions of Bird *et al.* (2007) and Toonen *et al.* (2011): the 255 km channel separating the MHI and NWHI restricts gene flow for intertidal and coastal marine species. In addition to being isolated, Nihoa is a small island with 4.73 km of coastline that is habitable by *C. exarata*, relative to ~109 km on Kaua‘i. Without the influence of harvesting, the population on Nihoa should be smaller than that of Kaua‘i. Genetic drift is more powerful in smaller populations (Lacy 1987), and given sufficient time, could accentuate the observed level of genetic differentiation. On the other hand, it is plausible that some of the genetic differentiation between Nihoa and the MHI is driven by the difference in genetic diversity which is expected to decrease if effective population sizes remain at present levels (see Genetic Diversity below).

Estimates of genetic differentiation using 22,474 loci (1 SNP per locus) provide somewhat greater resolution to detect population structure within the MHI, compared to a single mtDNA locus, but not as much as anticipated. Differences in genetic structure between mtDNA and a genome-wide survey of SNPs can be caused by a variety of factors, such as: increased mutation rates, separate replication events, modes of inheritance, and the absence of recombination in mtDNA (mtDNA are matrilineal; Reigh & Luck 1966; Zink & Barrowclough 2008). Bird et al. (2007) sampled *C. exarata* from Kaua‘i, Moloka‘i, and the Big Island of Hawai‘i in the MHI, and detected genetic partitioning solely between the central island (Moloka‘i) and both Kaua‘i and the Big

Island of Hawai‘i. While this was suggestive of Moloka‘i being genetically partitioned from the other MHI, Moloka‘i was not partitioned from a second site on Hawai‘i. Here, a central island, Maui, is clearly partitioned from the other MHI sampled (Kaua‘i, O‘ahu, and the Big Island of Hawai‘i) where $F_{CT}$ between Maui and the other MHI is 4x greater than that among the other MHI (Fig. 6). No other genetic structure is evident within the MHI.

When considering the connection between Moloka‘i and Maui, the SNP and mtDNA data for *C. exarata* are consistent with each other, but not with the general pattern of genetic structure exhibited by several species (Toonen et al. 2011). Populations on Moloka’i and Maui are expected to be genetically similar when considering that Moloka‘i and Maui are presently separated by a 13.5 km channel, and fuse into a single island, Maui Nui, during ice ages when sea level is can be ~120 meters lower than present day. Indeed, Toonen *et al.* (2011) surveyed the population structure of 23 marine species and found little genetic structure among the islands of Maui Nui. Rather, it is the Ka‘ie‘ie Waho (the channel separating Kaua‘i and O‘ahu, ~116km) and the ‘Alenuihaha (the channel separating Maui and the Big Island of Hawai‘i, ~45km) that are associated with genetic discordance in the MHI (Toonen et al. 2011). In *C. exarata*, however, the observed genetic discordance is specifically between Maui Nui and the other MHI. For example, the Big Island of Hawai’i is not differentiated from O’ahu or Kaua’i.

It is not immediately clear why Maui is genetically differentiated from the other MHI, but it is likely a consequence of historical demographics. Given the relative estimated population sizes and lack of genetic structure among Kaua’i, O’ahu, and Hawai’i, the genetic differentiation of Maui Nui from the other MHI is indicative of a

non-equilibrium condition. Effective population size ($N_e$), migration rate ($m$), and mutation rate ($\mu$) interact to affect genetic structure. In an Island Model (Wright 1943), large populations, where $2N_em>>1$ or $4N\mu>>1$, converge upon the same allele frequencies at equilibrium and are unlikely to exhibit genetic structure (Crow & Kimura 1970), but the populations here do exhibit structure. Smaller populations where $2N_em<1$ or $4N\mu<1$ would be expected to be differentiated at equilibrium. In a Stepping-Stone Island Model (Kimura & Weiss 1964), genetic differentiation would be expected to increase among islands that are farther apart. The observed pattern of genetic structure in the MHI, however, is unrelated to geographic distance or the amount of habitable coastline (proxy for population size without harvest), which is roughly similar among Kaua‘i, O‘ahu, and Maui, and 3-4x greater on the Big Island of Hawai‘i. During the last low sea level stand, the amount of habitat on Maui Nui was roughly similar to Hawai‘i. While identifying the demographic scenarios consistent with the observed pattern are beyond the scope of this paper, Maui Nui is the only instance where *C. exarata* populations are appreciably split and joined with changes in sea level, and this process may be involved.

*Genetic Differentiation Among Pooled Samples Collected Within Sites*

The advantage of preparing multiple pooled samples per site, rather than the more common practice of preparing a single pooled sample, is that the amount of error introduced by sampling, sample preparation, and sequencing is estimated and accounted for in comparisons of sampling locations. Errors may stem from pipetting, DNA quantification, or sample collection among other possibilities that could result in the misrepresentation of a population. Here, the original population sample was divided into

3-4 sub samples of 6-12 ‘opihi per site, but an alternative would be to make one large pooled sample and replicate it in triplicate, which would incur the same cost as the strategy employed here. The global $F_{SC}$ was not significant, indicating that a small amount of error was introduced by sampling and sample prep (Table 2). Two sub-samples, however, did not conform to expectations, one from O’ahu and one from Kaua’i. The sample from O’ahu was flagged very early on in the quality control process for having a low read depth and skewed frequency distributions relative to other samples taken from the same location and was removed prior to data analysis. The sample from Kaua’i was identified with $F_{SC}$, an index of the level of fixation among samples (Fig. 6). If a sub-sampling or replication procedure were not employed, Kaua’i may have been erroneously identified as genetically differentiated from other locations, and the O’ahu sample may have been lost. The sample from Kaua‘i could be removed from the analysis, however, it is most likely unnecessary as $F_{CT}$ estimates should be unaffected by the differences among samples from Kaua‘i. In order to better understand the effects of pooling procedures, further tests that directly compare results from multiple small pools, one large pool, replicates of large pools made with the same individuals would be instructive.

*Mean Genetic Diversity*

Both genome-level metrics that were used to quantify genetic diversity (the proportion of loci that are polymorphic and the mean nucleotide diversity among loci) yielded qualitatively similar results. Both diversity metrics are positively related to island age and the length of habitable coastline per person (for *C. exarata*) and negatively related to the length of habitable coastline, a proxy for population size without harvesting. This

was unexpected because, in *C. exarata*, mitochondrial COI (Bird et al 2007) and a nDNA intron, ATPSß (Bird et al 2011), both exhibit the opposite qualitative relationship where the larger populations in the MHI have greater nucleotide diversity than the smaller populations in the NWHI. There are other species in the Hawaiian Archipelago with higher levels of mitochondrial nucleotide diversity ($\pi$) in the NWHI compared to the MHI, including *C. sandwicensis* (Bird et al. 2007), *Holothuria atra* (Skillings et al. 2011), *Acanthaster planci* (Timmers et al. 2011), *Acanthurus nigroris* (DiBattista et al. 2010), and *Epinephelus quernus* (Rivera et al. 2004). Of these, only *C. sandwicensis*, a sibling species of *C. exarata*, is known to exhibit large decreases in population size in the NWHI due to a decrease in habitat.

Population genetic theory, such as the infinite sites model, indicates that at equilibrium (or a steady state) and all else being equal, larger populations will exhibit greater genetic diversity (Crow & Kimura 1970; Hartl *et al.* 1997). Empirical studies have found positive relationships between genetic diversity and population size (Soule 1976; Frankham 1996) but Bazin *et al.* (2006) did not find the expected relationship with mtDNA. In contrast, the nucleotide diversity of mtDNA in *C. exarata* conforms to expectations and the overall genetic diversity of the RAD loci (mostly nuclear) does not (Figs. 9 & 14). The negative relationship between genetic diversity and population size in *C. exarata* indicates that most loci have not achieved equilibrium with respect to effective population size, mutation, and migration. Supporting this inference, both the COI and ATPSß loci in *C. exarata* exhibit a signature starburst haplotype network Bird *et al.* (2007, 2011) that is consistent with a population bottleneck followed by population

expansion. The timing of the bottleneck in *C. exarata* is estimated to be greater than 2000 years old, prior to human harvesting of *C. exarata* (unpublished, but see Bird *et al.* 2011)

The positive relationship between the polymorphic proportion of loci and island age is striking ($r^2$=0.998, Fig. 8). Nihoa, by far the oldest and smallest island sampled, has the greatest polymorphic proportion of loci and nucleotide diversity. There are two primary hypotheses that would explain the observed pattern. (1) The ancestor of *C. exarata* colonized the Hawaiian archipelago via Nihoa (or an island farther to the Northwest, ~5.1 Mya according to Bird *et al.* 2011) then subsequently colonized each island to the Southeast as they emerged. (2) There was a more recent population bottleneck that affected the MHI more severely and/or more recently than Nihoa. The latter hypothesis seems more plausible given the shallow divergences among haplotypes of COI and ATPSβ (Bird *et al.* 2007, 2011), and the proportionately low nucleotide diversities for the RAD loci at all locations (Fig. 12). Further, island age is likely to be correlated with the amount of time since population expansion began given the progressively lower levels of genetic diversity on each island to the Southeast, which each have greater amounts of habitat. We expect that the genetic diversity of *C. exarata* will increase in the MHI if effective population sizes are maintained.

Population bottlenecks caused by overharvesting can also reduce the genetic diversity, but hard evidence demonstrating this phenomenon in marine fish and invertebrates is uncommon (Allendorf *et al.* 2008; Hauser *et al.* 2002). *Cellana exarata* exhibit lower genetic diversity on islands with less habitat per person (MHI) than Nihoa, an island presently unoccupied by humans and estimated to be able to support 175 humans prior to European contact (Cleghorn 1988). O'ahu is the island with the greatest

harvesting pressure (~0.12 meters of habitat per person) that has resulted in low population densities relative to other islands (pers. obs.). However, genetic diversity on O‘ahu is greater than on the Big Island of Hawai‘i (~2.2 meters of habitat per person). Detecting a decline in genetic diversity due to overharvesting may be difficult for *C. exarata*, given that genetic diversity is likely to be low in the MHI due to a relatively recent population bottleneck that predates human occupation. The COI nucleotide diversity of both *C. sandwicensis* and *C. talcosa* are much greater than in their sibling species, *C. exarata,* and if at equilibrium, should be more sensitive to population bottlenecks caused by harvesting.

Perhaps the most intriguing pattern in genetic diversity observed in this study is the relatively high nucleotide diversity for polymorphic loci on Maui relative to the other MHI (Fig. 12) despite a low proportion of loci that are polymorphic (Fig. 7). In Figures 13-15, Maui is the sample that deviates from the overall trend. Pairwise $F_{SC}$ estimates of pooled samples within Maui were not unusual as seen with pooled samples within Kaua‘i; hence, sequencing and laboratory errors are unlikely to explain the observed nucleotide diversity. As mentioned earlier, the population on Maui fuses with the populations presently isolated on Moloka‘i, Lana‘i, and Kaho‘olawe during low sea level stands. The Wahlund Principle predicts that the alternating fusion and isolation of populations should lead to a reduction in genetic diversity relative to a population that is not subdivided (Hartl & Clark 2007). However, when a population is not at equilibrium and the historical demographics are unknown, seemingly odd patterns of diversity may develop. To understand gene flow among the islands of Maui Nui, additional genetic information will need to be collected and analyzed. Population genetic simulations

mimicking the effects of sea level rise and fall on populations of *C. exarata* on Maui Nui would provide insight as to whether additional historical demographic events are required to explain the observed levels of nucleotide diversity.

*Locus by Locus Analysis of Nucleotide Diversity (π)*

While the average diversity estimated using RAD loci suggests that island age is its strongest correlate, not all RAD loci follow this trend (Fig. 16). Thus far, inferences about genetic diversity have been made from averages of 21,529 loci, but the true power of RAD sequencing is the ability to analyze each locus. For ~10% of all loci investigated, nucleotide diversity exhibits a negative relationship with island age ($r^2$>0.4), and thus a positive relationship with the length of habitable coastline (a proxy for population size without harvesting). With this knowledge, the positive relationships between nucleotide diversity and habitable coastline for COI (Bird *et al.* 2007) and ATPSβ (Bird et al. 2011) are consistent with the RAD loci. Specifically, COI and ATPSβ are popular genetic markers because of their high mutation rates, low $N_e\mu$, and subsequent utility in detecting genetic structure. Given a bottleneck, followed by population expansion, nucleotide diversity is expected to rise the fastest in the loci with high mutation rates and low $N_e\mu$ (see Zink *et al.* 2008 for review). Therefore, we propose that the ~10% of loci characterized by a negative relationship between nucleotide diversity and island age are likely to have mutation rates that are higher than average. Frankham (1996) notes that the relationship between population size and genetic diversity is likely to differ among loci, but selection is advanced as the primary reason, which cannot be ruled out here.

It is instructive to consider that with a handful of sequenced DNA loci or a sample of 100 RAD loci (~200bp per locus), this pattern (30% of loci positively correlated with

island age and 10% negatively correlated) may have gone undetected or overlooked as an anomaly. We suggest that at least 1000 loci should be surveyed when investigating patterns of genetic diversity, which coincidently, was the same number of loci required to obtain a reasonable estimate of $F_{CT}$. Further, a few loci, selected haphazardly are not expected to be representative of the whole genome.

*Conclusions*

Restriction site associated DNA sequencing (RAD) has promised to bring population genomic to non-model species at a reasonable cost (Davey & Blaxter 2010). Here, we employed pooled sample sequencing with ezRAD to further reduce the costs. Compared to the mtDNA study of *C.exarata* (Bird *et al.* 2007), RAD facilitated larger population samples, marginally greater resolution for the detection of genetic structure with *F*-statistics, much greater confidence in conclusions of no genetic structure, and a much more detailed inferences about the processes driving observed patterns of genetic diversity. Both the genetic structure and diversity of *C. exarata* are strongly influenced by historical demographics. It would be worthwhile to compare multiple locations on each island to confirm the results presented here. Finally, the results here do not preclude the utility of RAD markers in detecting patterns of gene flow at very small scales, within islands; and sequencing individuals would facilitate additional analysis options.

For fishery management, a more detailed analysis of migration rate and population size is required than what is presented or warranted here. *F*-statistics are notoriously insensitive to intermediate levels of $N_em$ and while it would superficially seem that there is connectivity among most of the MHI, realize that there are differences in genetic diversity and genetic structure among islands that would be erased by high

migration rates at a scale where stock replenishment among islands could occur on a reasonable time scale. Counter intuitively, the majority of genetic diversity is found in the smallest population investigated. Falling within the Papahānaumokuākea Marine National Monument, Nihoa houses a genetic stockpile for the *C. exarata* in the MHI. The other three NWHI harboring *C. exarata* should be investigated to determine if they have equivalent levels, or even greater levels of genetic diversity than Nihoa.

**References**

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution* **23**, 327-337.

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

Bird CE, Holland BS, Bowen BW, Toonen RJ (2007) Contrasting phylogeography in three endemic Hawaiian limpets (Cellana spp.) with similar life histories. *Molecular ecology* **16**, 3173-3186.

Bird CE, Karl SA, Smouse PE, Toonen RJ (2011) Detecting and measuring genetic differentiation. *Phylogeography and population genetics in Crustacea* **19**, 31-55.

Chong Z, Ruan J, Wu C-I (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics* **28**, 2732-2737.

Clague DA (1996) The growth and subsidence of the Hawaiian-Emperor volcanic chain. *The origin and evolution of Pacific Island Biotas, New Guinea to Eastern Polynesia: pattern and processes*.

Cleghorn PL (1988) The settlement and abandonment of two Hawaiian outposts: Nihoa and Necker Islands. *Bishop Museum Occasional Papers* **28**, 35-49.

Corpuz G (1983) *Life History Traits Influencing Vertical Zonation in the Hawaiian Intertidal Species of Cellana*, University of Hawai'i.

Crow JF, Kimura M (1970) An introduction to population genetics theory. *An introduction to population genetics theory.*

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings*

*in Functional Genomics* **9**, 416-423.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Frankham R, Ballou JD, Briscone DA (2002) *Introduction to Conservation Genetics* Cambridge Univerity Press, Cambridge.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Guinand B, Scribner K, Page K (2003) Genetic variation over space and time: analyses of extinct and remnant lake trout populations in the Upper Great Lakes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 425-433.

Halpern BS, Kappel CV, Selkoe KA*, et al.* (2009) Mapping cumulative human impacts to California Current marine ecosystems. *Conservation Letters* **2**, 138-148.

Halpern BS, Walbridge S, Selkoe KA*, et al.* (2008) A global map of human impact on marine ecosystems. *Science* **319**, 948-952.

Hartl DL, Clark A (2007) Principles of Population Genetics (Sinauer, Sunderland, MA). *the text*.

Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). *Proceedings of the National Academy of Sciences* **99**, 11742-11747.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries* **9**, 333-362.

Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature* **406**, 882-885.

Hutchinson WF, van Oosterhout C, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (Gadus morhua). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 2125-2132.

Jackson JB, Kirby MX, Berger WH*, et al.* (2001) Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629-637.

Jensen JL, Bohonak AJ, and Kelley ST (2005) Isolation by distance, web service. BMC Genetics 6: 13. v.3.23 http://ibdws.sdsu.edu/

JMP®, Version 11.2.0. SAS Institute Inc., Cary, NC, 1989-2007.

Kay A, Magruder W (1977) The Biology of Opihi. *Department of Planning and Economic Development.*

Kenchington E, Heino M, Nielsen EE (2003) Managing marine genetic diversity: time for action? *ICES Journal of Marine Science: Journal du Conseil* **60**, 1172-1176.

Kenchington EL (2003) 14 The Effects of Fishing on Species and Genetic Diversity. *Responsible fisheries in the marine ecosystem*, 235.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561.

Lacy RC (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology* **1**, 143-158.

Law R (2007) Fisheries-induced evolution: present status and future directions. *Marine Ecology Progress Series* **335**, 271-277.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589-595.

Miller K, Maynard B, Mundy C (2009) Genetic diversity and gene flow in collapsed and healthy abalone fisheries. *Molecular ecology* **18**, 200-211.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research* **17**, 240-248.

O'Brien CM, Fox CJ, Planque B, Casey J (2000) Fisheries: climate variability and North Sea cod. *Nature* **404**, 142-142.

Olsen EM, Heino M, Lilly GR*, et al.* (2004) Maturation trends indicative of rapid evolution preceded the collapse of northern cod. *Nature* **428**, 932-935.

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431.

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reich E, Luck D (1966) Replication and inheritance of mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **55**, 1600.

Ryman N, Utter F, Laikre L (1995) Protection of intraspecific biodiversity of exploited fishes. *Reviews in Fish Biology and Fisheries* **5**, 417-446.

Selkoe K, Halpern B, Ebert C*, et al.* (2009) A map of human impacts to a “pristine” coral reef ecosystem, the Papahānaumokuākea Marine National Monument. *Coral Reefs* **28**, 635-650.

Smith P, Francis R, McVeagh M (1991) Loss of genetic diversity due to fishing pressure. *Fisheries Research* **10**, 309-316.

Soulé M (1976) Allozyme variation: its determinants in space and time. *Molecular evolution*, 60-77.

Swain DP, Sinclair AF, Hanson JM (2007) Evolutionary response to size-selective mortality in an exploited fish population. *Proceedings of the Royal Society B: Biological Sciences* **274**, 1015-1022.

Thorpe J, Solé-Cava A, Watts P (2000) Exploited marine invertebrates: genetics and fisheries. *Hydrobiologia* **420**, 165-184.

Toonen RJ, Andrews KR, Baums IB*, et al.* (2011) Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian Archipelago. *Journal of Marine Biology* **2011**.

Toonen RJ, Puritz JB, Forsman ZH*, et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**, e203.

Trippel EA, Morgan M, Fréchet A*, et al.* (1997) *Changes in age and length at sexual maturity of northwest Atlantic cod, haddock and pollock stocks, 1972-1995* Department of Fisheries & Oceans, Biological Station.

Walsh MR, Munch SB, Chiba S, Conover DO (2006) Maladaptive changes in multiple traits caused by fishing: impediments to population recovery. *Ecology letters* **9**, 142-148.

Wang L, Shi X, Su Y, Meng Z, Lin H (2012) Loss of genetic diversity in the cultured stocks of the large yellow croaker, Larimichthys crocea, revealed by microsatellites. *International journal of molecular sciences* **13**, 5584-5597.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370.

Wright S (1921) Systems of mating, I–V. Genetics, 6, 111–178.

Wright S (1931) Evolution in Mendelian populations. Genetics, 16, 97–159.

Wright S (1943) Isolation by distance. Genetics, 28, 114–138.

Wright S (1951) The genetical structure of populations. Annals of Eugenics, 15, 323–354.

Wright S (1965) The interpretation of population structure by Fstatistics with special regard to system of mating. Evolution, 19, 395–420.

Wright S (1977) Evolution and the Genetics of Population, Experimental Results and Evolutionary Deductions. The University of Chicago Press, Chicago.

Wright S (1978) Evolution and the Genetics of Population, Variability Within and Among Natural Populations. The University of Chicago Press, Chicago.

Wright PJ (2005) Temporal and spatial variation in reproductive investment of haddock in the North Sea. *ICES CM* **2005**.

Wright PJ, Tobin D (2013) Maturation differences between sub-stocks of haddock, Melanogrammus aeglefinus. *Marine biology* **160**, 231-239.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology* **17**, 2107-2121.

Zar, J. H. (1984). Simple linear regression. Biostatistical analysis, 4, 324-359.

## Tables

**Table 1** Region, island, number of pools, number of individuals per pool, mean nucleotide diversity ($\pi$) for polymorphic loci, and percent polymorphic loci for *Cellana exarata*. Samples used to calculate nucleotide diversity and polymorphic proportions of loci are denoted with an asterisk (*).

| Region | Island | # of pools | Individuals/pool | Mean $\pi$ | Polymorphic loci (%) |
|---|---|---|---|---|---|
| NWHI | Nihoa | 3 | 7*,7*,8* | $2.05 \times 10^{-3}$ | 62.988 |
| MHI | Kaua‘i | 4 | 6,7*,11*,12* | $1.75 \times 10^{-3}$ | 53.346 |
| | O‘ahu | 4 | 10*,8*,10* | $1.78 \times 10^{-3}$ | 50.427 |
| | Maui | 3 | 9*,10*,7* | $1.87 \times 10^{-3}$ | 49.546 |
| | Hawai‘i | 4 | 10*,10*,10*,10 | $1.71 \times 10^{-3}$ | 49.159 |

**Table 2** Analysis of Molecular Variance. The islands are a comparison of each island to each other island: Nihoa, Kaua‘i, O‘ahu, Maui, and the Big Island of Hawai‘i. The island variance component with respect to the total variance is $F_{CT}$. The among sample within island variance component is $F_{SC}$. Bold font indicates statistical significance $\alpha = 0.05$).

| Source | d.f | $F_{CT}$ / $F_{SC}$ | Sig |
|---|---|---|---|
| Island | 4 | 0.071 | < **0.001** |
| Sample(Island) | 16 | 0.002 | 0.811 |
| Total | 20 | | |

**Figures**

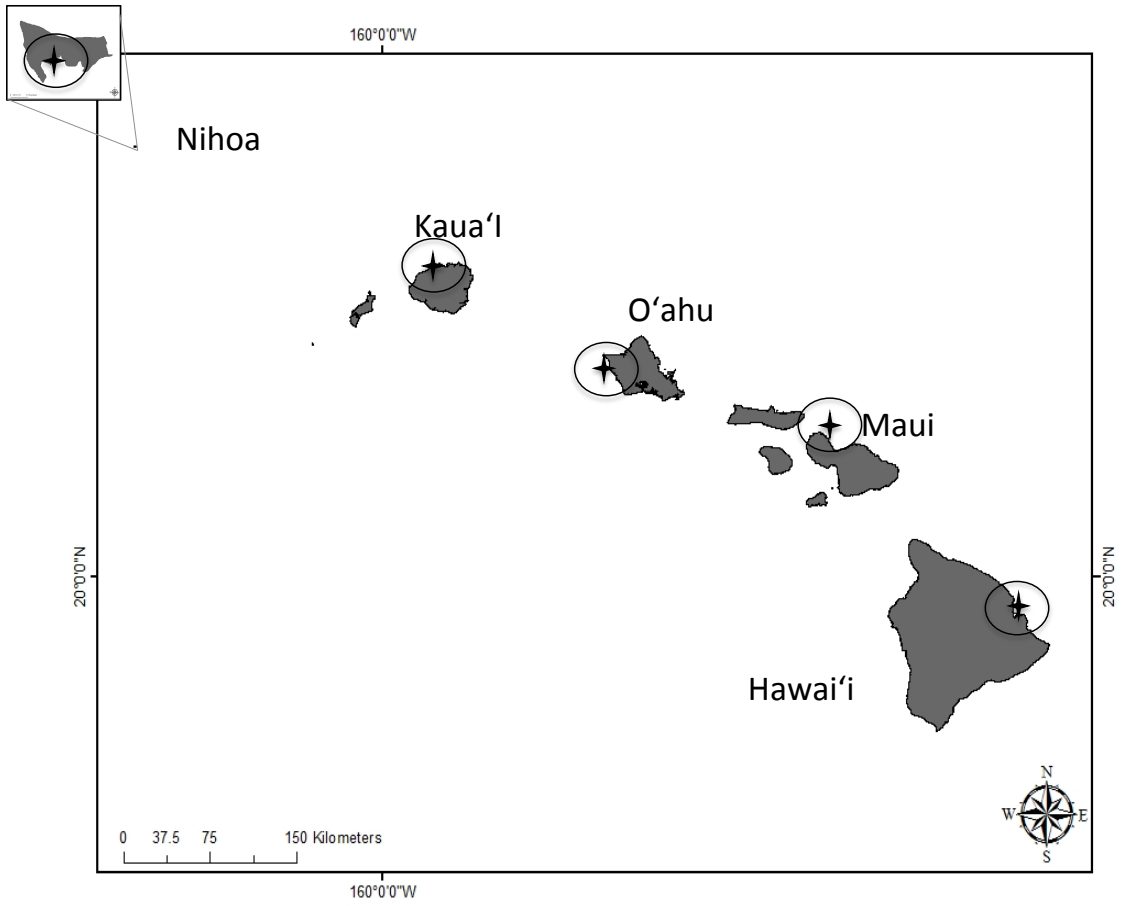


**Figure 1** Map of sample locations within the Hawaiian archipelago. There is one sample collection site within the Northwestern Hawaiian Islands (Adam's Bay, Nihoa) and four sample collection sites within the Main Hawaiian Islands (Ke'e, Kaua'i; Ka'ena, O'ahu; Waihe'e, Maui; Hilo break wall, the Big Island of Hawai'i).

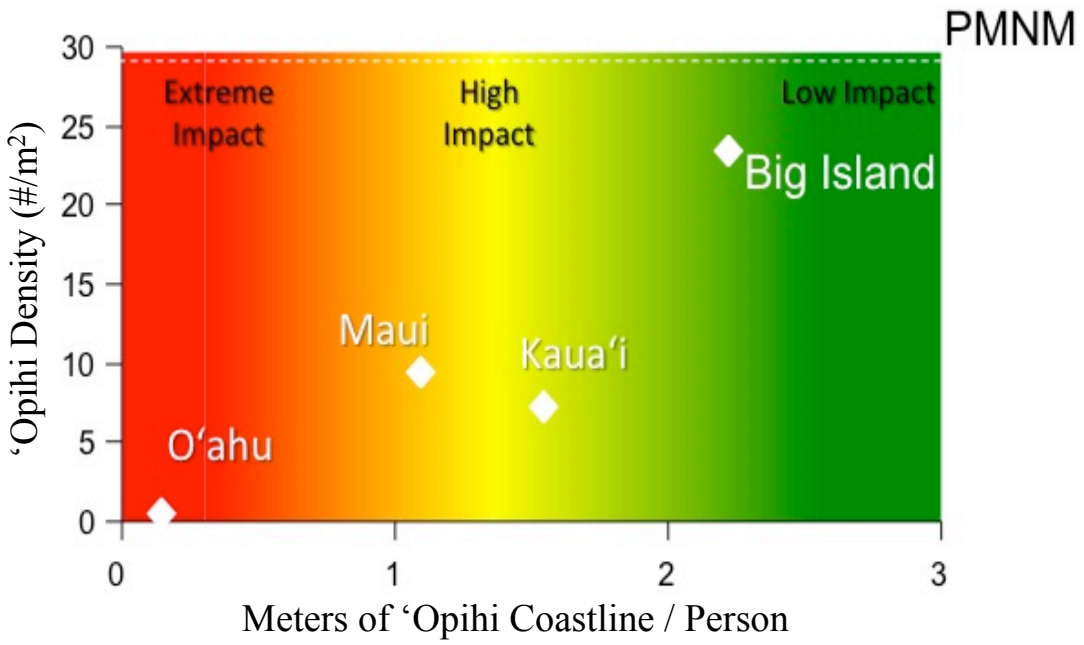


**Figure 2** The linear meters of *C. exarata* coastal habitat/ per person with respect to *C. exarata* densities (number of *C. exarata* /meter$^2$) on each of four of the largest MHI: O'ahu, Maui, Kaua'i, and Big Island. Densities of *C. exarata* populations were estimated via intertidal surveys (C.E.B unpublished data).

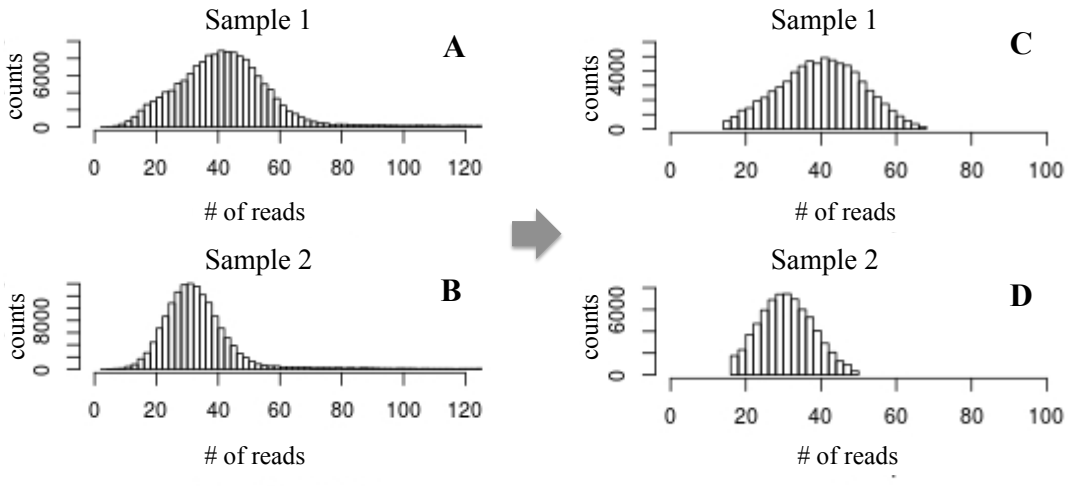


**Figure 3** Total depth of coverage distributions for loci within two samples prior to removing SNP loci with the sample-specific filter for high and low depth of coverage (A, B) and total depth of coverage distributions for SNP loci within the same two samples after removing sample specific loci for high and low depth of coverage (C, D).

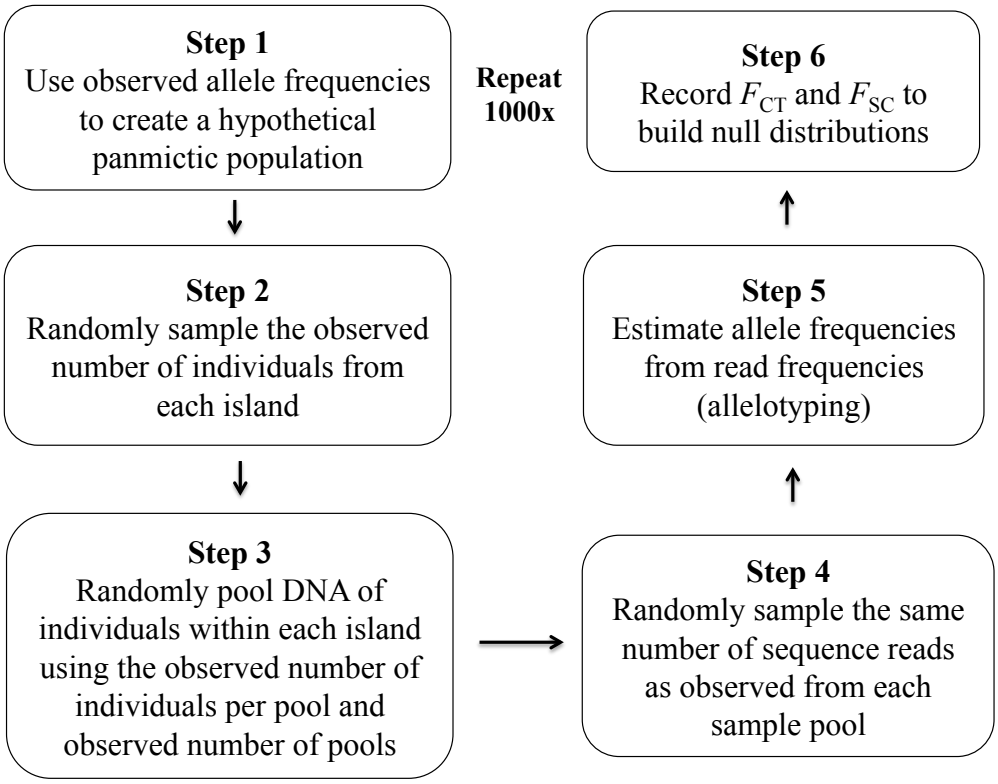


**Figure 4** Representation of the simulation coded in R that was used to create null $F_{CT}$ and $F_{SC}$ distributions. Steps 1 through 5 were simulated 1000 times. Randomly sampled individuals in step 2 were sampled from the same hypothetical panmictic population without resampling.

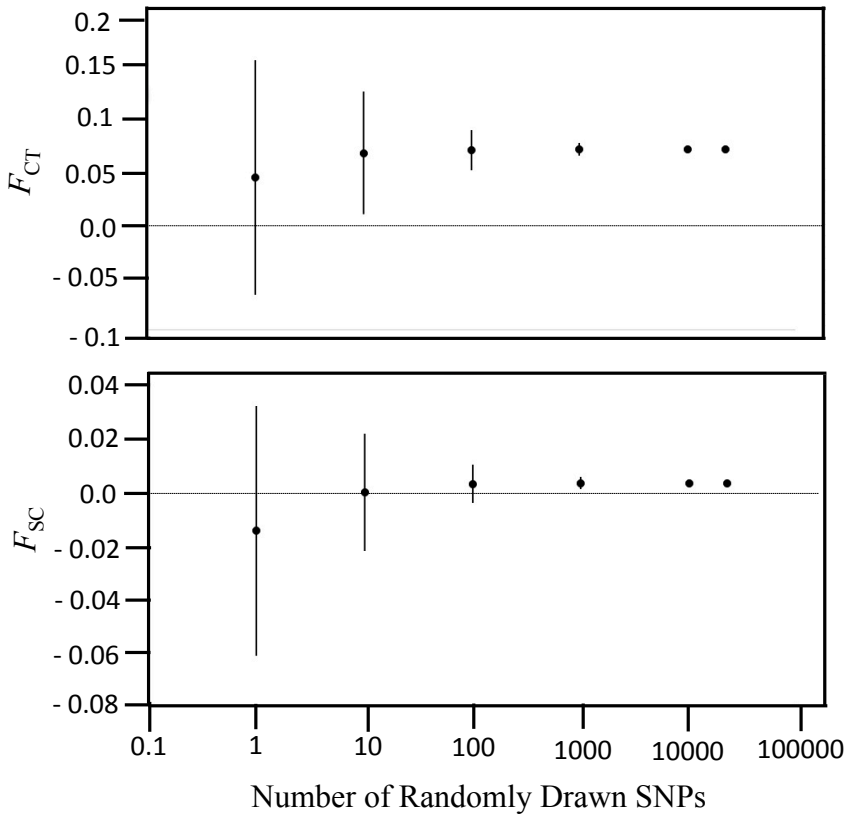


**Figure 5** Plots of mean $F_{CT}$ and $F_{SC}$ vs. 1, 10, 100, 1000, 10,000, and 22,474 randomly drawn SNPs from 22,474 loci. SNPs were drawn 1000x. Bars represent ±1 standard deviation.

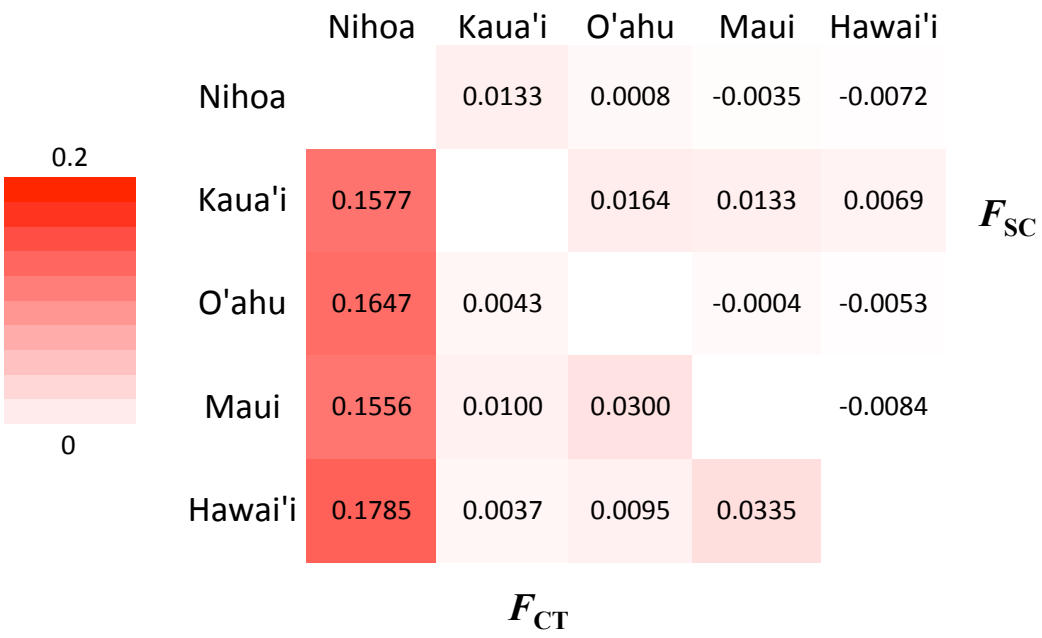


**Figure 6** Heat map of pairwise $F_{CT}$ and $F_{SC}$ estimates. An $F_{CT}$ or $F_{SC}$ value relatively close to zero (represented by white) indicates little to no genetic differentiation. $F_{CT}$ and $F_{SC}$ values greater than zero (represented by pink/red) indicates increased levels of genetic differentiation. $F_{CT}$ is a measure of variance among two islands and $F_{SC}$ is a measure of variance among samples within islands.

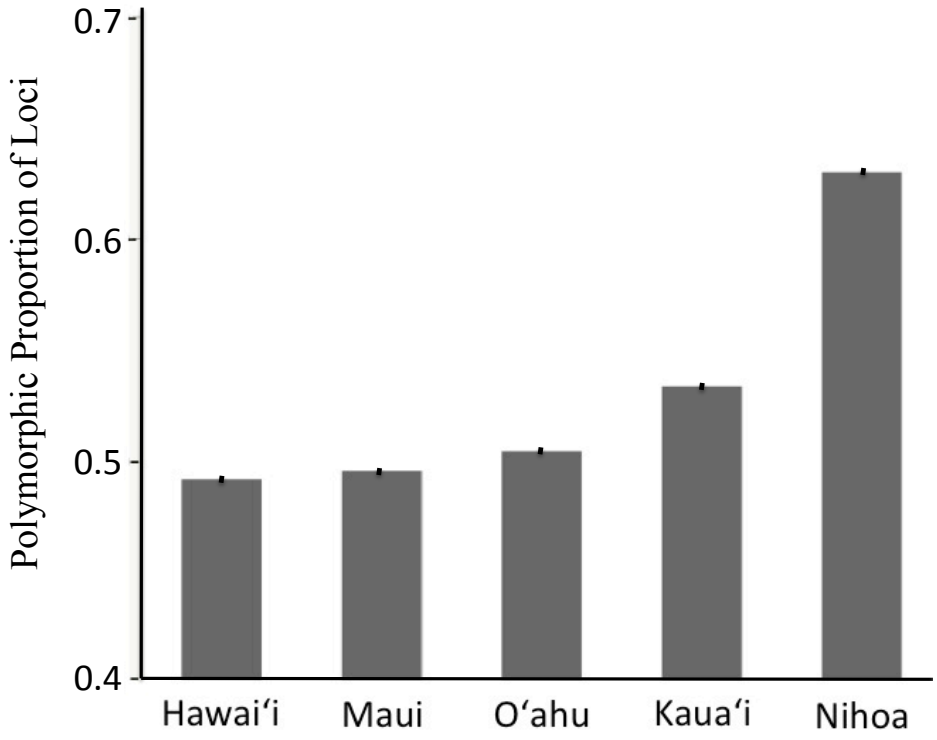


**Figure 7** Comparison of polymorphic proportions of loci by island. Bars represent binomial confidence intervals. Islands are in order from Southeast to Northwest.

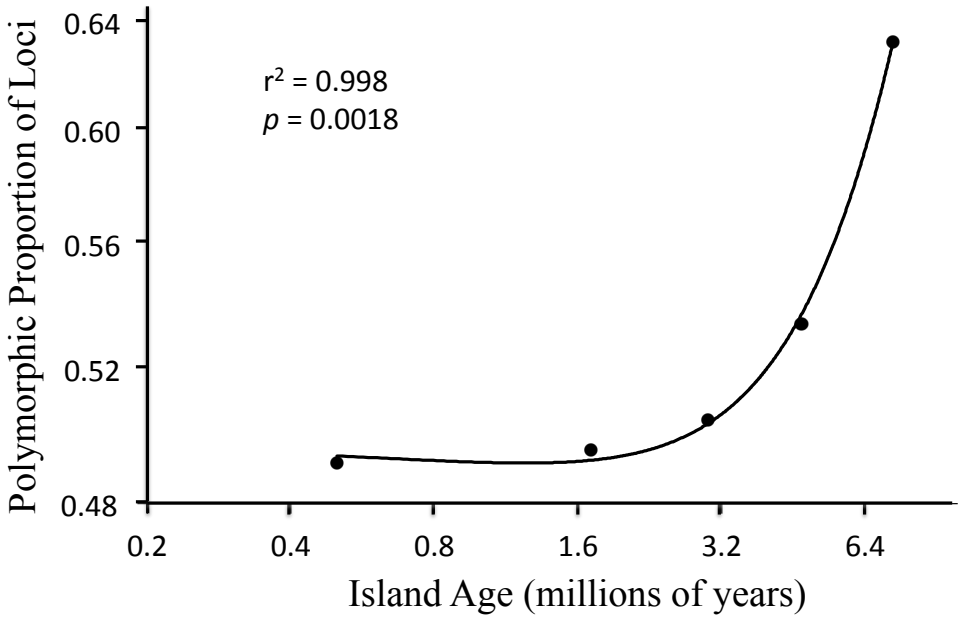


**Figure 8** Response of polymorphic proportions of loci (21,529 total loci) to island age. Island age is in millions of years ago. Island age is represented on a logarithmic scale ($log_2$).

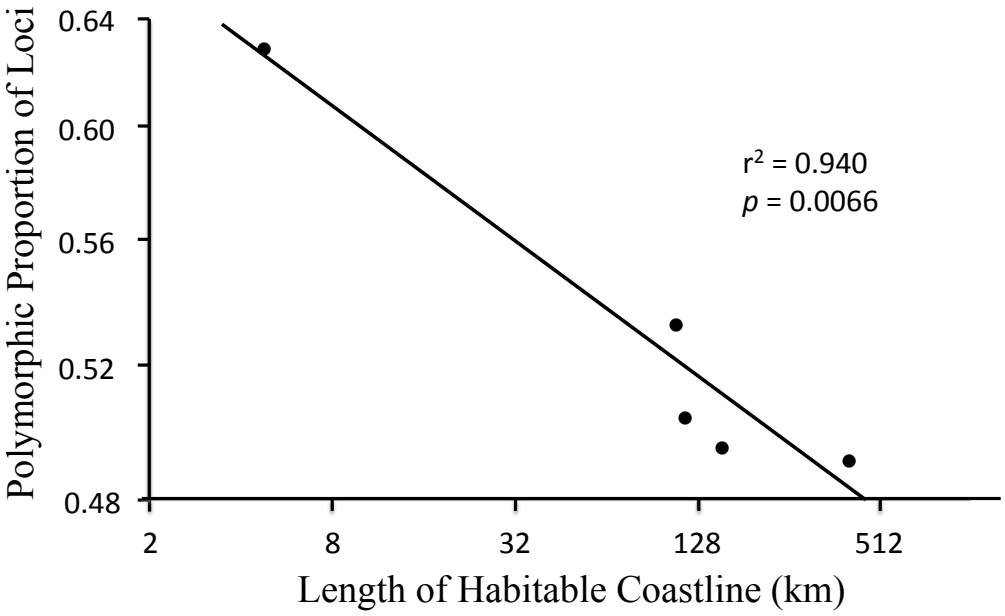


**Figure 9** Response of polymorphic proportions of loci (21,529 total loci) to the length of habitable coastline (km). The length of habitable coastline is represented on a logarithmic scale (Ln) and represents the linear available habitat for *C. exarata*.

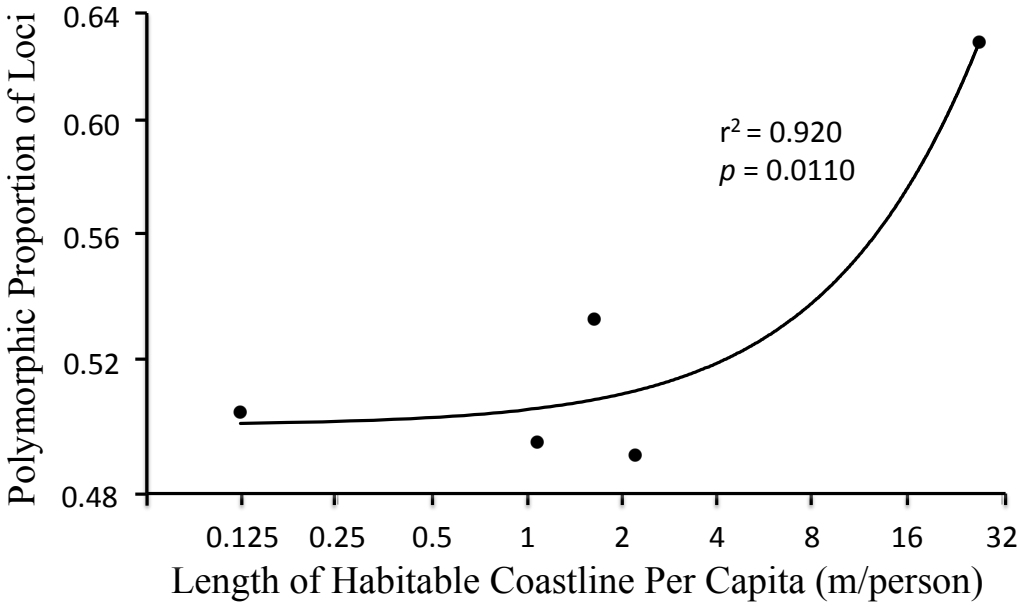


**Figure 10** Plot of polymorphic proportions of loci vs. the length of habitable coastline per capita. The length of habitable coastline per capita was calculated by dividing the length of habitable coastline by the human population size for each island.

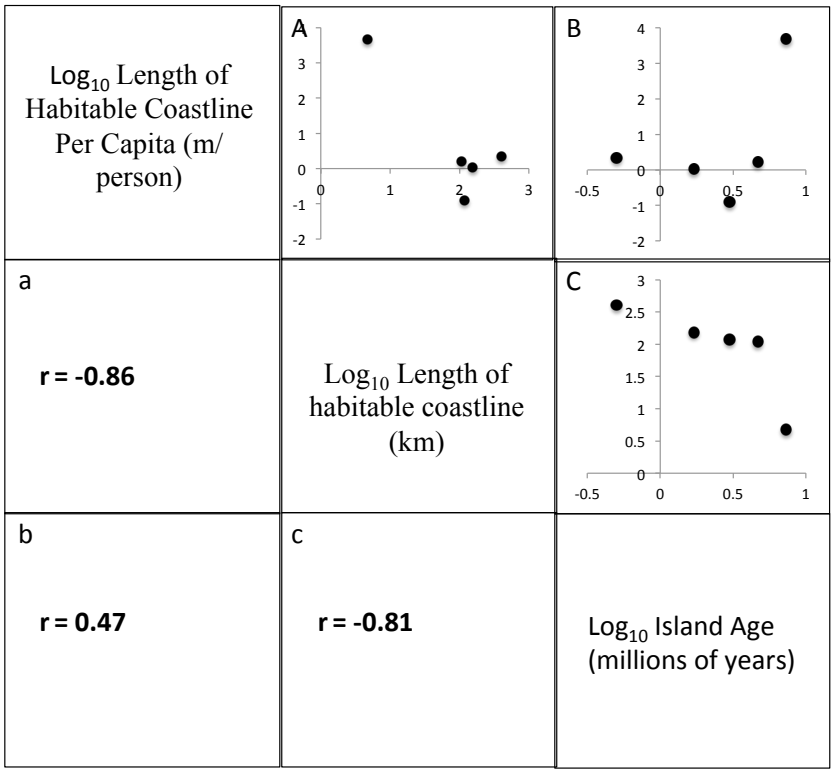


**Figure 11** Plots of length of habitable coastline (km) vs. length of habitable coastline per person (A), island age vs. length of habitable coastline (km, B), and island age vs. the length of habitable coastline (km, C). Each predictor was log transformed ($log_{10}$). Pearson product-moment correlation coefficients are included to match each plot of the predictors (a,b,c).

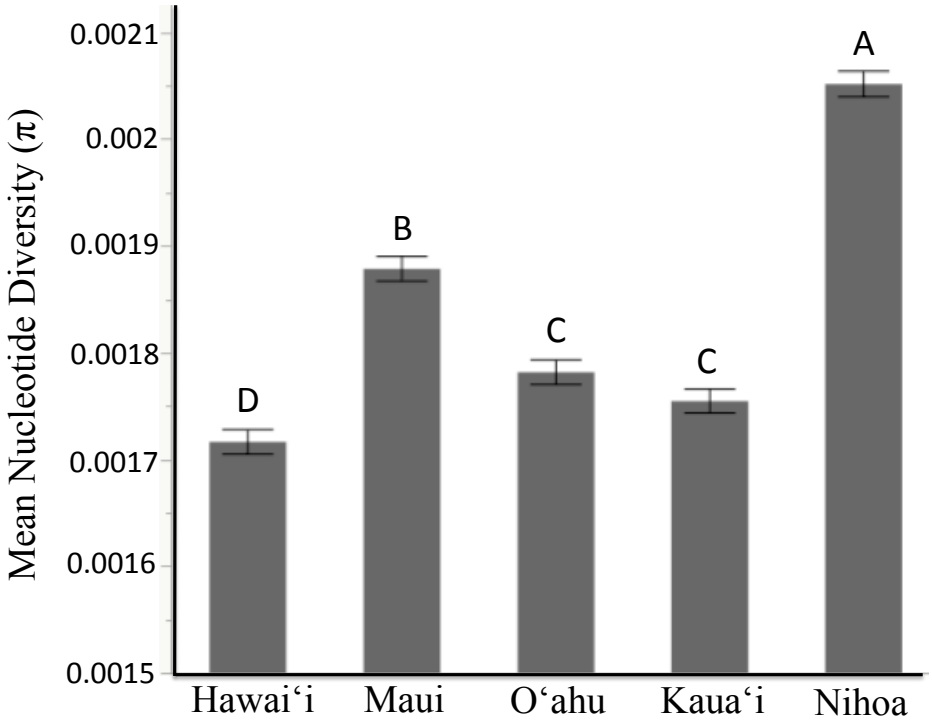


**Figure 12** Comparison of the mean nucleotide diversity of polymorphic loci by island. Bars are 95% confidence intervals and each island with a different letter is significantly different (Tukey's HSD, $p < 0.05$)

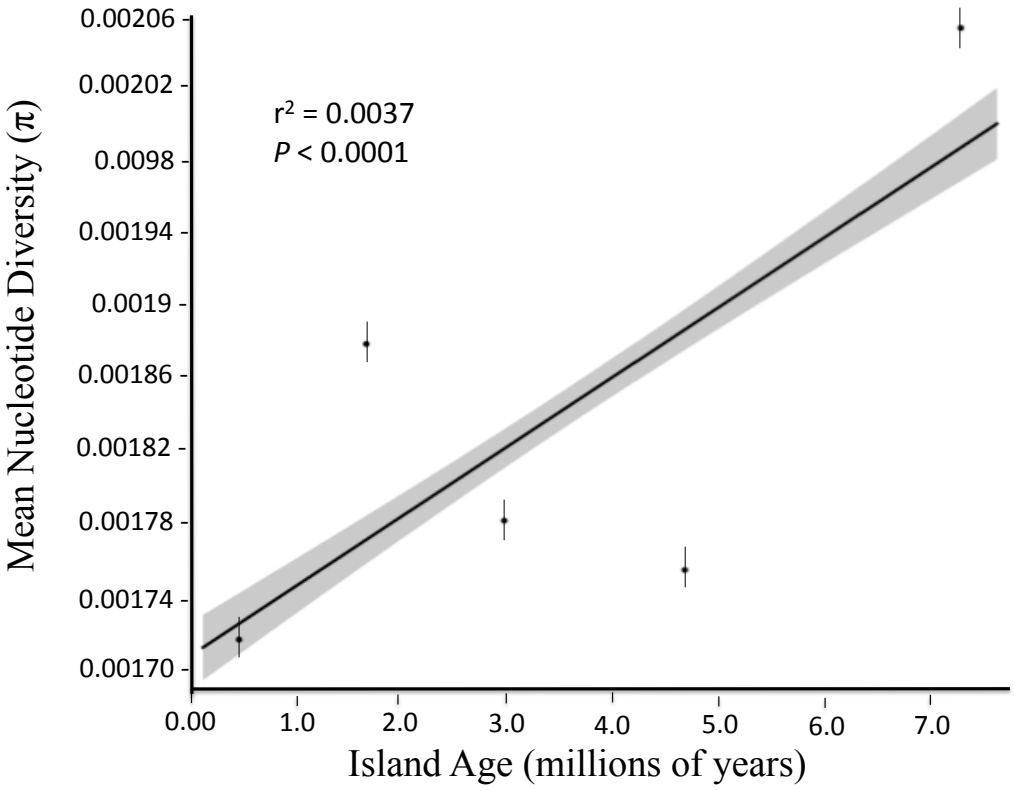


**Figure 13** Plot of mean nucleotide diversity of polymorphic loci vs. island age. Island age is represented as millions of years ago. Bars are 95% confidence intervals.

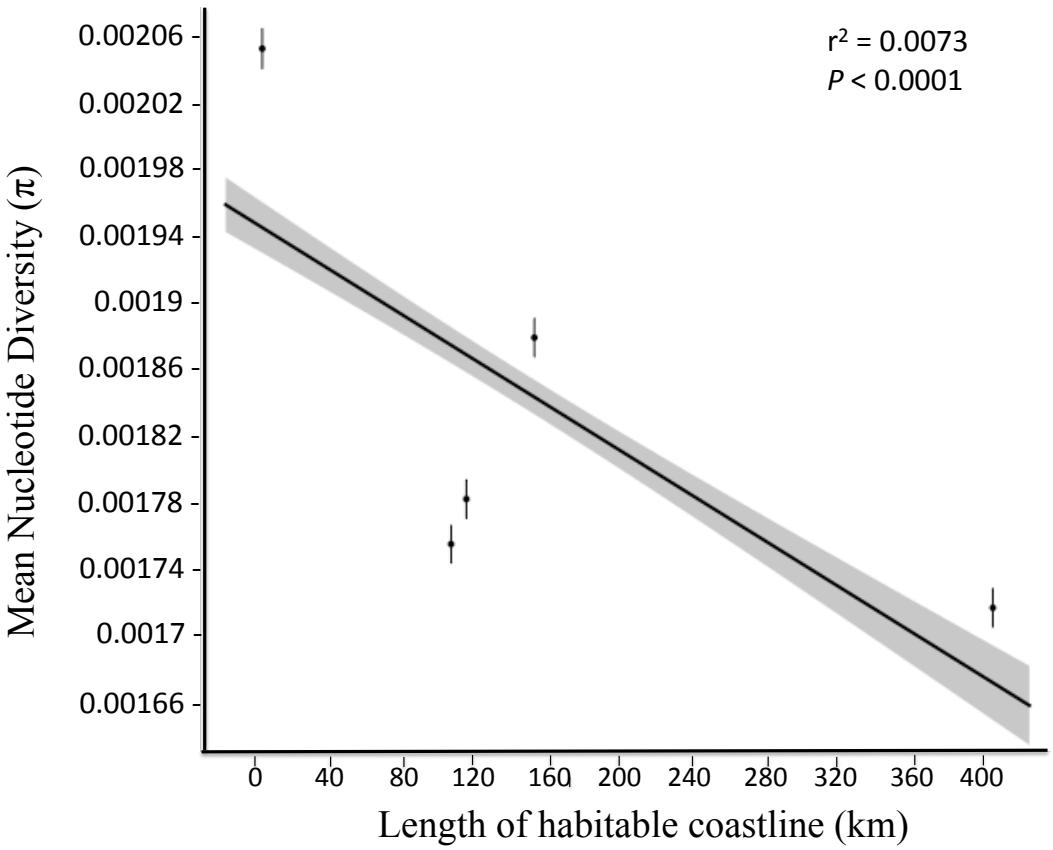


**Figure 14** Plot of mean nucleotide diversity of polymorphic loci vs. length of habitable coastline. Length of habitable coastline is represented in km. Bars are 95% confidence intervals.

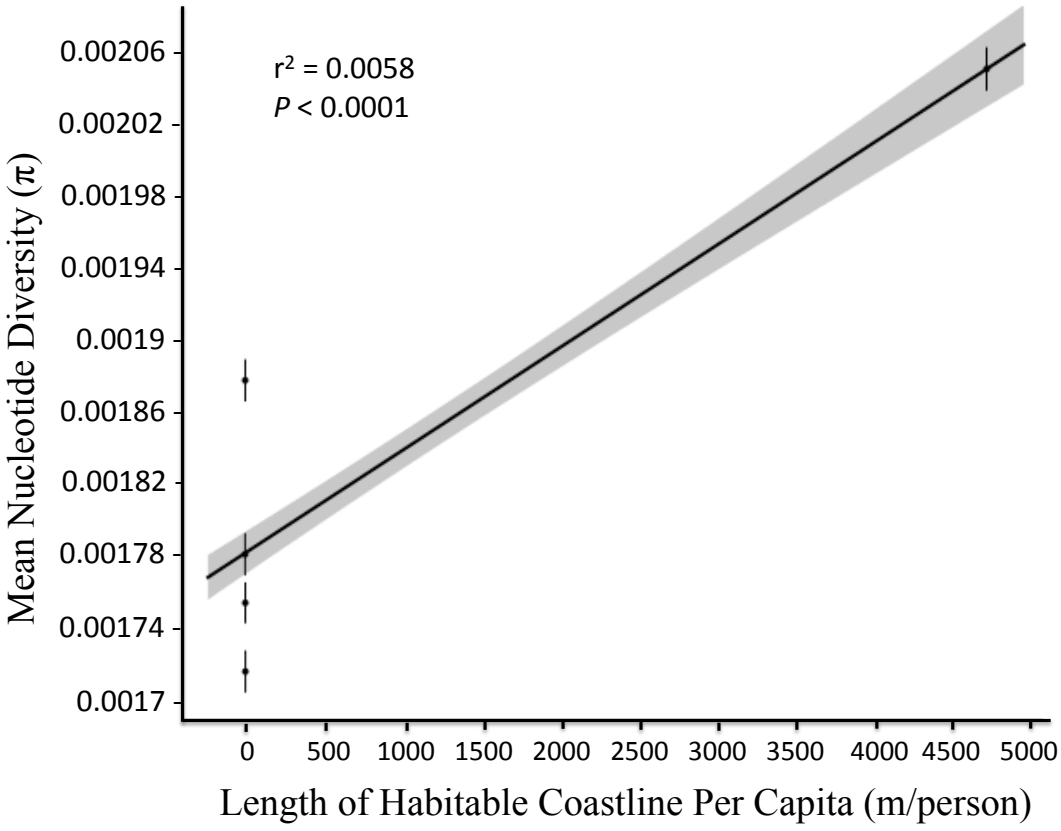


**Figure 15** Plot of mean nucleotide diversity of polymorphic loci vs. length of habitable coastline per capita. The length of habitable coastline per capita is measured by the linear meters of *C. exarata* habitat per person. Bars are 95% confidence intervals.

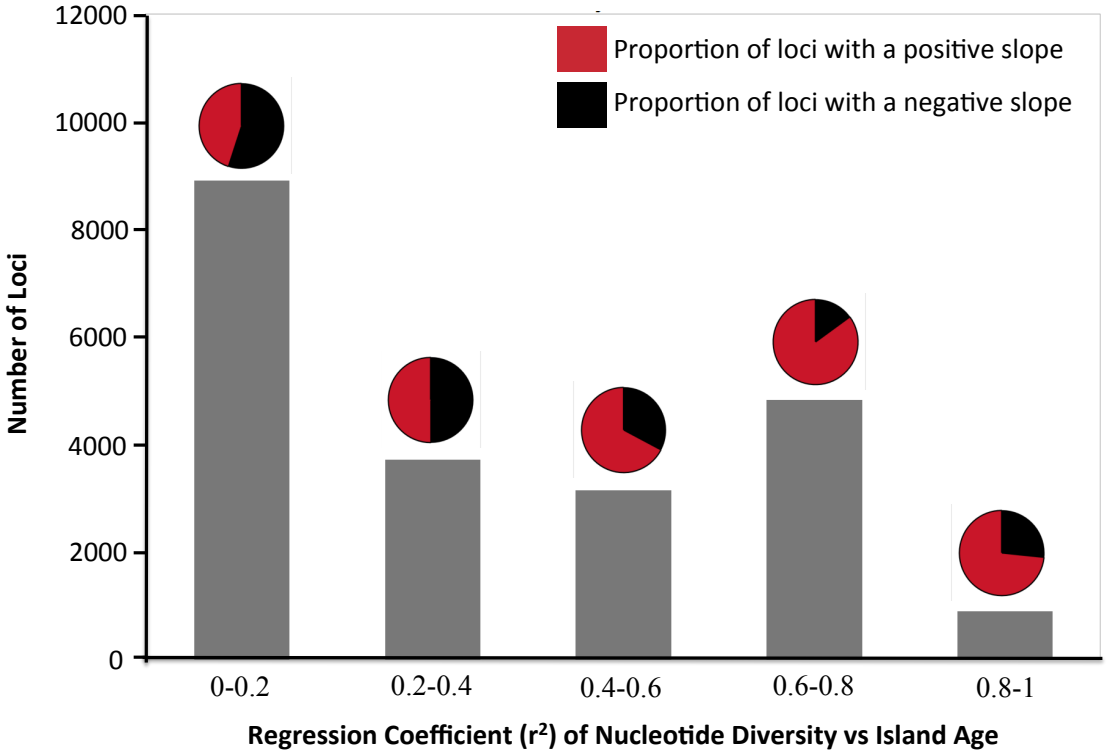


**Figure 16** Comparison of the number of loci with respect to the coefficient of determination calculated from regressions of 21,529 loci. Nucleotide diversity was regressed against island age. Pie charts represent the proportion of loci that share a positive (represented by red) or negative (represented by black) relationship.