PROCEEDINGS B

Research

Author for correspondence:
Kimberly A. Selkoe
e-mail: selkoe@nceas.ucsb.edu

THE ROYAL SOCIETY
PUBLISHING

The DNA of coral reef biodiversity: predicting and protecting genetic diversity of reef assemblages

Kimberly A. Selkoe[1,2], Oscar E. Gaggiotti[3], Eric A. Treml[4], Johanna L. K. Wren[5], Mary K. Donovan[6], Hawai'i Reef Connectivity Consortium† and Robert J. Toonen[1]

[1]Hawai'i Institute of Marine Biology, University of Hawai'i, Kāne'ohe, HI 97644, USA
[2]National Center for Ecological Analysis and Synthesis, 735 State Street, Santa Barbara, CA 93101, USA
[3]School of Biology, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife KY16 8LB, UK
[4]School of BioSciences, University of Melbourne, Parkville, Victoria 3010, Australia
[5]Department of Oceanography, and [6]Department of Biology, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA

iD KAS, 0000-0003-1790-5431

Conservation of ecological communities requires deepening our understanding of genetic diversity patterns and drivers at community-wide scales. Here, we use seascape genetic analysis of a diversity metric, allelic richness (AR), for 47 reef species sampled across 13 Hawaiian Islands to empirically demonstrate that large reefs high in coral cover harbour the greatest genetic diversity on average. We found that a species's life history (e.g. depth range and herbivory) mediates response of genetic diversity to seascape drivers in logical ways. Furthermore, a metric of combined multi-species AR showed strong coupling to species richness and habitat area, quality and stability that few species showed individually. We hypothesize that macro-ecological forces and species interactions, by mediating species turnover and occupancy (and thus a site's mean effective population size), influence the aggregate genetic diversity of a site, potentially allowing it to behave as an apparent emergent trait that is shaped by the dominant seascape drivers. The results highlight inherent feedbacks between ecology and genetics, raise concern that genetic resilience of entire reef communities is compromised by factors that reduce coral cover or available habitat, including thermal stress, and provide a foundation for new strategies for monitoring and preserving biodiversity of entire reef ecosystems.

1. Introduction

Known for their stunning arrays of colours, shapes and life forms, coral reefs are captivating examples of extreme biodiversity. Hidden within the taxonomic and life-history diversity found on reefs, but no less important, is the genetic diversity carried within individuals and populations. Genetic diversity is the seed of ecological and evolutionary processes like niche partitioning and species diversification that lead to the complex community structure typical of coral reefs and other highly biodiverse ecosystems. In turn, community-level processes no doubt have consequences for genetic diversity within populations, although mechanisms are not well studied. Because of the many possible ways in which genetic diversity may be linked to ecological functioning, adaptive capacity and extinction risk, conservation strategies often call for preserving areas of high genetic diversity [1,2]. When conservation strategies focus on habitats, communities or ecosystems, as is common for coral reefs, it is important to understand how genetic diversity patterns vary across co-distributed species, because including genetic data can shift conservation priorities dramatically [3].
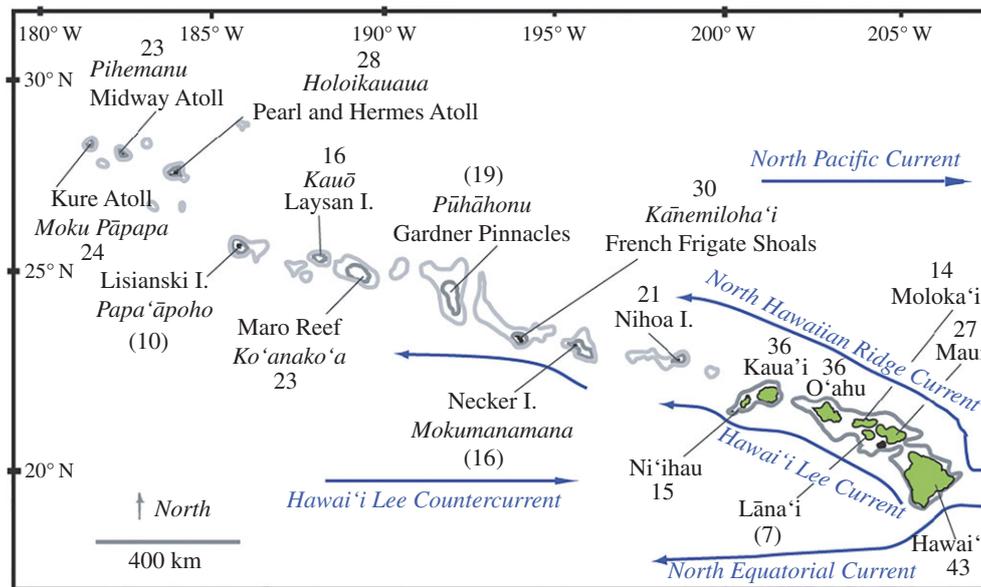
Figure 1. Map of the Hawaiian archipelago. Hawaiian (italics) and English (regular font) names of sampled islands and number of species genetically sampled per island are indicated; numbers in parentheses indicate islands excluded from community-level analyses of aggregate AR. Major currents are represented by arrows; 1000 and 2000 m isobaths are delineated. Islands east of 200° are the Main Hawaii Islands (MHI); islands west of 200° are the Northwest Hawaiian Islands (NWHI) and part of Papahānaumokuākea Marine National Monument.

Investigating patterns of genetic diversity at the community level can be framed by tests of foundational theory on drivers of biodiversity. Theory predicts that physical area constrains diversity by limiting carrying capacity and the genetic 'effective' population size, while immigration boosts diversity by bringing in new variants [4]. Species with similar constraints on habitat and movement may thus be expected to have similar patterns of genetic diversity [5], despite important trait differences across species [6,7]. Furthermore, historical events, such as major disturbances, may act on whole communities to produce a common signature of genetic bottleneck that depresses observed genetic diversity [7,8]. Accumulating evidence indicates that ecological interactions can shape genetic diversity, such as when the genetic diversity of habitat-forming species influences the diversity of associated fauna [9]. Patterns of species-level diversity and genetic diversity can be correlated, perhaps due to parallel responses to dominant environmental gradients [10] or causal relationships [11], suggesting the possibility of emergent genetic patterns at the community level. Despite these homogenizing shared forces, contrasts in life-history traits across species create differences in migration rates, habitat use and density, leading to variation in patterns of genetic diversity across co-distributed species. Life-history differences may also cause species to vary in which landscape features most strongly drive their spatial diversity patterns. There is much recent interest in investigating the major environmental correlates of population genetic patterns, partly due to the utility of protecting key landscape features or locations in conservation planning. However, prior landscape genetic studies have focused on, at most, a handful of species at a time. To date, few studies have explored the range of genetic diversity patterns and their drivers across a large sample of co-distributed species, and the role of life history in mediating these patterns and drivers [6–8,12]. None to date have focused on a sample of the 'meta-community' that includes representatives from multiple taxa, trophic levels and functional groups.

Here, we investigate how patterns of genetic diversity vary across the Hawaiian Archipelago for a sample of 47 reef-associated animal species and ask if observed variation can be ascribed to potential seascape drivers representing benthic cover, ocean currents, habitat loss caused by sea-level change, temperature stress and other site characteristics. At first glance, the nearly linear array of discrete islands in the Hawaiian Archipelago might be expected to produce high congruence of genetic patterns as they impose uniform physical constraints on population size and dispersal distance for most reef species (figure 1). However, there are few clear cases of isolation by distance, the null expectation for genetic structure across a linear island array. Instead, patterns of population genetic structuring are highly variable, and much of this variation probably stems from the great diversity of life history and demography across reef-dwelling species [13,14]. We leveraged the large number of species genetically sampled across Hawai'i to investigate whether species' genetic diversity patterns tend to covary with each other and with species diversity, and whether species traits account for observed variance in spatial patterns of genetic diversity and association to seascape drivers.

A priori hypotheses about drivers of diversity across the islands were derived from past theoretical and empirical studies. They were:

H1: Marginal populations show reduced genetic diversity due to greater isolation [15].
H2: Species richness and genetic diversity covary [4,10].
H3: Habitat size and immigration influence genetic and species diversity [16].
H4: Habitat loss during the last glacial maximum produced genetic bottlenecks that dampen diversity [7,8].
H5: The genetic diversity of habitat-forming species influences the genetic diversity of associated fauna [9,10].
H6: Ecological factors structuring reefs, namely thermal stress, coral cover and crustose coralline algae (CCA) cover, may influence genetic diversity.

We assessed these hypotheses based on both the responses of each species to the various drivers and the response of the aggregated data. The latter 'community-

Table 1. Summary of seascape effects on a reef community. Hypothesis column indicates which numbered hypothesis was tested with the dataset; symbols + and − indicate whether the predicted relationship to genetic diversity is positive or negative. Support column: Y indicates support at $p < 0.05$ for the hypothesis based on regression analyses of composite genetic diversity of mtDNA data (left), nucDNA data (middle) and linear mixed modelling (right). Italics indicate factors not included in multiple regressions due to colinearity or data gaps. LGM is last glacial maximum; CCA is crustose coralline algae.

seascape driver	hypothesis	support	brief description
fish species richness	H2 +	Y Y N	Bootstrapped 'Chao' estimates of species counts collected on underwater visual transects conducted by NOAA's Coral Reef Ecosystem Division from 2011 to 2012. Number of sites surveyed per island was roughly proportional to reef area. Values were natural log transformed.
coral species richness	H2 +	N N N	Same as for fish species richness except that survey data spanned years 2006–2010.
habitat area	H3 +	Y Y Y	Log_{10} $(x + 1)$ transformed estimates of total shallow-water area within the 10-fathom depth curve of each island.
potential larval immigration	H3 +	N N N	In-coming centrality metric calculated from modelled larval connectivity estimated from an oceanographic biophysical model parametrized for the species and habitat array.
nearest-neighbour distance	H3 −	N N N	Path distances in km between approximate centroids of islands estimated using Google Earth. The shortest distance to a neighbouring island was selected.
LGM habitat loss	H4 −	Y N Y	Estimate of the relative severity of population bottlenecks due to habitat loss 18 000 years ago during the LGM when sea level was 120 m lower; one minus the ratio of LGM to present-day habitat area.
M. capitata genetic diversity	H5 +	N N N	Rarefied AR averaged across five microsatellite loci of the coral *Montipora capitata* that each showed Hardy–Weinberg Equilibrium (non-significant tests that F_{IS} differs from zero).
coral cover	H6 +	Y Y Y	Number of pixels with more than 10% cover in IKONOS satellite imagery covering 0–30 m depth range. Expressed as a percentage by dividing by total number of pixels analysed.
CCA cover	H6 +	Y N Y	Same as for coral cover.
thermal stress	H6 −	Y Y N	Frequency of hotspot events, as defined by NOAA's Coral Reef Watch, when SST exceeded the maximum monthly mean temperature, over years 1985–2000, and measured at 4 km.
wave disturbance	H6 −	N N N	Yearly average of maximum monthly mean wave energy over years 1997–2010, measured at 1°.

level' analysis approach treats genetic diversity of all species sampled as an emergent trait of the community, and builds on previous reports of European forest plants and Mediterranean darkling beetles which found that mean genetic diversity of a community sample shows strong correlations to landscape features and species diversity [7,12].

2. Material and methods

(a) Seascape factors

Environmental and ecological data used to test our set of six hypotheses were assembled from both existing sources and new efforts. Table 1 summarizes each dataset included in analyses; see the electronic supplementary material for full descriptions. Given the small number of islands with adequate data to test our hypotheses, we sought to limit the number of seascape drivers compared in alternative model testing. Below we note our rationale for excluding additional available datasets from analysis. Estimates of benthic cover included macroalgae, sand, uncolonized, coral and CCA. We limited analyses of benthic cover type to coral and CCA cover based on established hypotheses about their roles structuring reefs. Although macroalgae also structure reefs, remote sensing methodology was unable to make the ecologically critical distinction between native and

invasive types. Multi-year environmental data were available for sea surface temperature (SST), wave energy, irradiance and chlorophyll-a [17]. We limited analysis to just two metrics that are often named as drivers of coral reef communities, SST and wave energy, both available as island level averages of all grid cells within the 30 m isobath. Lastly, two habitat area measures were available—total hard-bottom area classified from satellite imagery, and bathymetric area within the 10-fathom contour. The two estimates covaried and produced identical correlation to rarefied mean AR (electronic supplementary material, table S1); we present main results using the bathymetric estimate. A principal components analysis of variation in seascape factors across islands shows that approximately 60% of the variation occurs along two axes, with the primary axis most closely associated with fish species richness, coral cover and thermal stress (electronic supplementary material, figure S1).

(b) Genetic diversity metrics

Genetic data assembly is detailed elsewhere [14]. Existing genetic data from samples of reef species collected in Hawaiian waters were included here for species with a minimum of six specimens each taken from of two or more islands (electronic supplementary material, dataset S1). Mean sample size per species per island was 32; 10% of the species-site samples contained fewer than 10

4

individuals. Excluding samples based on 6–10 specimens had insubstantial effects on results (electronic supplementary material, table S2). Analyses primarily focused on mitochondrial sequence data because only a few species have nuclear datasets; however, we also used the 11 nuclear multi-locus datasets available to assess sensitivity of some results to the marker type (electronic supplementary material, dataset S2). In addition, microsatellite data for the coral *Montipora capitata* [18] was used as a predictor variable to test hypothesis H5. For all genetic datasets, rarefied allelic richness (AR) was calculated for each island using HP-RARE [19]. As a count of alleles, AR is unaffected by other locations but possibly influenced via gene flow [20].

Multi-species 'composite' AR means were calculated for each island based on AR values of all species sampled. To reduce sampling bias, composite AR was rarefied to a uniform size of 12 species per island by sampling species without replacement 500 times (electronic supplementary material, dataset S3). Bias in marker composition across islands was checked and found to be minimal (electronic supplementary material, 'Methods'). Lisianski, Kahoʻolawe and Lanaʻi were excluded from analyses of composite AR due to low genetic sample sizes ($n < 12$ species). See the electronic supplementary material for extensive sensitivity analyses of this metric to sampling parameters.

(c) Statistical analyses

Moran's I calculated with R packages *ape* and *ncf* [21] revealed no large-scale or small-scale positive spatial autocorrelation of AR across all sites, or NWHI ($p = 0.31$) and MHI ($p = 0.99$) subsets of sites. The same was true for diversity metrics of fishes ($p = 0.16$) and corals ($p = 0.91$). After assembly of all seascape data (electronic supplementary material, dataset S4), two islands were excluded from all analyses (Necker and Gardner), because they were missing data for several key seascape factors (electronic supplementary material, 'Methods'). When using composite AR, Lisianski, Kahoʻolawe and Lanaʻi were also excluded due to low sample size ($n < 12$ species), leaving 13 islands in the analysis.

Redundancy analysis (RDA) with the R package *vegan* [22] was used to visualize variation in species' correlations of AR to the seascape predictors and assess the influence of species traits on this variation. A set of species traits, summarized in table 2, was published previously [14] (electronic supplementary material, dataset S5). Pearson's r values describing each species's correlation of AR to each seascape factor were the dependent variables (electronic supplementary material, dataset S6). A partial RDA was also performed, using the sample size (i.e. number of islands), marker type and total marker diversity for each species as covariates, but these covariates lacked influence (electronic supplementary material, dataset S7). RDA was complemented by AIC_c model selection of linear models built with species and sampling traits to determine which traits most parsimoniously explained which species showed high or low correlations of AR to seascape factors.

Congruence of spatial patterns of AR across species was gauged by Pearson's correlation coefficient for each species's AR spatial pattern regressed against the composite AR pattern of all species, for species sampled at more than five islands ($n = 34$). A one-sided t-test indicated whether species tended to positively correlate with composite AR, to assess a community-level trend towards congruence as in [7]. Sensitivity of congruence to sampling was assessed (electronic supplementary material, 'Methods'). To assess the roles of life-history, sampling and genetic traits on congruence, these Pearson's r values were regressed against species traits in table 2 in linear AIC_c-based model selection using with JMP PRO v. 11.

The same model selection procedure was also used to assess relationships of composite AR to physical and ecological seascape factors in table 1. For comparison, a similar model selection procedure was implemented for the individual AR values for all species-by-marker-by-island combinations available ($n = 421$)

Table 2. Summary of species trait data used in analyses.

species traits	description
fish	species is fish (1) or invertebrate (0)
endemic	species is endemic to Hawaiʻi (1) or widespread in the Pacific (0)
PLD	pelagic larval duration in days, log transformed
length	maximum body length in cm, log transformed
min depth	minimum reported depth occurrence in meters, log transformed
depth range	maximum depth minus the minimum depth, log transformed
habitat specialist	species is tied to particular reef features (1) or found on most reef types (0)
θ_{ST}	strength of inter-island genetic differentiation
F_{CT}	strength of regional genetic differentiation (i.e. groups of adjacent islands)

using a linear mixed model which designated the species-by-marker label as a random effect (electronic supplementary material, dataset S8). The latter tests the aggregated response of individual species, which does not have to be the same as the response of the aggregated data (composite AR). Variation was high in AR values across species, and in which species were sampled across islands; the two modelling approaches address these issues differently but otherwise test the same hypothesis with the same response variable. The composite mean uses rarefaction to help standardize sampling variance across islands. The mixed modelling approach uses the species-level data to incorporate the variance across species and also the possibility that each species is drawn from its own distribution. Because of the large increase of parameters that need to be estimated, power is lower, but assumptions are fewer.

For both model selection procedures, data gaps for coral species richness and *M. capitata* genetic diversity estimates required omitting these predictors from model selection. Fish and coral species richness and wave disturbance were omitted due to colinearity with other factors (electronic supplementary material, 'Methods'; electronic supplementary material, table S1). Models were limited to one to three terms for model comparison to reduce model number given the small sample size of composite AR ($n = 13$ islands). Top models were defined as $\Delta AIC_c < 2.0$, where ΔAIC_c is the difference in AIC_c score from the model with the minimum observed AIC_c score. Model selection was repeated for regional subsets (i.e. seven islands in the NWHI and six in the MHI), motivated by the many differences between these regions that might produce distinct population genetic and ecological dynamics.

3. Results

(a) Spatial diversity patterns

Spatial patterns of genetic diversity varied considerably across the 34 well-sampled species (sampled at more than four islands). For eight species sampled with both nuclear and mitochondrial markers at 6–12 islands, spatial patterns positively correlated between the marker sets, suggesting single-marker spatial trends are interpretable (electronic supplementary

Table 3. Which species respond to which seascape drivers? Top models from multiple regression model selection built with species' traits (independent variables) and correlation coefficients for each species' AR regressed against individual seascape factors (dependent variable). Depth indicates depth range. (−) indicates negative relationships; all others are positive. $*p < 0.05$; $**p < 0.01$.

seascape factor	species trait predictors	Adj. r^2	K	AIC$_c$	w_i	coeff. 1	coeff. 2
habitat area	ϕ_{ST}	0.21**	2	27.9	0.48	6.15**	
coral cover	depth, ϕ_{ST}	0.24**	3	18.8	0.54	0.40**	4.86*
potential larval immigration	depth, ϕ_{ST}	0.29**	3	26.2	0.76	0.29*	5.77*
wave disturbance	depth (−), ϕ_{ST}	0.28**	3	38.2	0.35	−0.45**	−6.25*
CCA cover	depth, count	0.30**	3	2.00	0.24	0.24*	−0.09**
LGM habitat loss	PLD (−), taxonomy[a]	0.46**	5	23.3	0.30	−0.47**	−0.79**
	marker[b], depth (−)			34.2		0.53**	−0.002*
nearest-neighbour	F_{CT}	0.13*	2	39.8	0.45	5.68*	
fish spp. richness	herbivore (−)	0.15*	2	34.6	0.23	−0.27*	
coral spp. richness	herbivore, F_{CT}	0.29**	3	22.0	0.31	5.63**	−0.28*
thermal stress	no significant model						

[a]Invertebrates tend to show more negative correlation to LGM habitat loss than fishes.
[b]Cyt B dataset tend to show more positive correlation to LGM habitat loss than other marker types.

material, table S3). Furthermore, individual species' mtDNA AR patterns showed significant tendency for positive correlations to the composite mean AR pattern calculated using all 47 species (t-test 2.4, d.f. = 33, $p = 0.01$; electronic supplementary material, figure S2). Congruence retained significance when highly influential species were omitted, either because they were particularly well sampled or most positively correlated to the mean pattern (electronic supplementary material, 'Methods'). Species with shorter PLD tended to have higher congruence (adj. $r^2 = 0.13$, $p = 0.04$). Composite AR of all species ranged 3.0–3.8 haplotypes per six individuals across islands, and tended to show higher values at both margins of the island chain (quadratic $r^2 = 0.45$, $p = 0.05$; electronic supplementary material, figure S3), contrary to our first hypothesis that a stepping stone habitat array produces lower diversity at the margins due to reduced immigration. Importantly, habitat area ranges three orders of magnitude across the sampled islands (6.6–470 km^2), with large areas at the margins. High genetic diversity at the margins was also seen for the nucDNA version of composite AR of 10 species, because mtDNA and nucDNA composite AR values significantly correlated ($r^2 = 0.51$, $p = 0.01$; electronic supplementary material, table S1). A population genetic simulation reproduced the patterns of high diversity at the margins when effective population sizes were made to vary across demes with the same relative magnitude seen in habitat area across the Hawaiian Archipelago (i.e. approx. threefold; electronic supplementary material, 'Methods' and figure S4).

(b) Species-genetic diversity correlation

Estimated number of fish species (standardized for sampling effort) ranged 98–152 across islands and showed a negative quadratic fit with latitude, such that composite AR and fish species diversity significantly covaried ($r^2 = 0.51$, $p < 0.01$; electronic supplementary material, figure S3), supporting our second hypothesis. Coral species richness ranged from 10 to 25 per island and positively correlated with fish species richness ($r^2 = 0.30$, $p = 0.04$; electronic supplementary material,

table S1), but was not significantly correlated to composite AR ($r^2 = 0.27$, $p = 0.08$; electronic supplementary material, figure S3). Functional group specificity of species-versus-genetic diversity correlations may contribute to the mixed result [11]. Composite nucDNA AR showed almost identical relationships to species richness as mtDNA (electronic supplementary material, table S1). Considering species' individual responses, model selection showed that herbivores tended to respond positively to coral species richness and negatively to fish species richness (table 3). An obligate corallivore, *Chaetodon lunulatus*, was the species whose AR pattern most strongly correlated to coral species richness ($r^2 = 0.99$; electronic supplementary material, dataset S6).

(c) Effects of species traits on seascape drivers of genetic diversity

RDA showed that the 12 species traits together explained 49% of the variation in how species related to the seascape drivers ($p = 0.04$; electronic supplementary material, figure S5). Overall, species with stronger genetic structuring (i.e. pairwise differentiation between adjacent islands or groups of islands) showed stronger relationships to most drivers, and especially to habitat-related factors. The strongest trends in seascape associations were tied to the combined effects of taxonomy, habitat area and depth range. For example, species that showed the strongest response to habitat loss during the last glacial maximum tended to be invertebrates with shallow depth ranges, short PLD and low marker diversity, which typically indicates a past bottleneck. Shallowest species responded negatively to wave disturbance, and larger depth range was also associated with influence of coral cover and CCA cover on genetic diversity (table 3).

Interestingly, the few species that showed strong positive correlation of genetic diversity to simulations of potential larval immigration were deep-water, more genetically structured species, a rare combination in our dataset. The larval immigration model, which scaled larval production to habitat area, produced a strong peak in potential immigration at the

Table 4. Top seascape mixed models of species AR ($n = 421$). All models for which ΔAIC_c ranged $0-2$ are listed (excluding models for which coefficient signs opposed our hypothesized relationships). Arch. is archipelago-wide, K is the number of parameters, w_i is the Akaike weight. Coefficients are ordered by predictor order. ***$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$.

region	seascape predictors	K	AIC_c	ΔAIC_c	w_i	coeff. 1	coeff. 2	coeff. 3
Arch.	habitat area, LGM habitat loss	4	1167.7	0	0.16	0.20**	−0.29*	
Arch.	habitat area, coral cover	4	1168.9	1.2	0.09	0.14***	0.43*	
Arch.	habitat area, coral cover, LGM habitat loss	5	1169.2	1.5	0.07	0.18*	0.21	−0.22
Arch.	coral cover	3	1169.3	1.6	0.07	0.49*		
Arch.	habitat area, CCA cover, LGM habitat loss	5	1169.6	1.9	0.06	0.19*	3.54	−0.28*
NWHI	habitat area	3	601.5	0	0.08	0.10		
MHI	coral cover, CCA	4	634.2	1.1	0.08	0.61*	23.6*	

Table 5. Top seascape models of composite AR ($n = 13$). All models for which ΔAIC_c ranged $0-2$ are listed. Adj. is adjusted, other abbreviations and notation as in table 4.

region	seascape predictors	Adj. r^2	K	AIC_c	ΔAIC_c	w_i	coeff. 1	coeff. 2
Arch.	habitat area, coral cover	0.59**	3	−1.7	0	0.13	0.24*	0.56***
Arch.	habitat area, LGM habitat loss	0.59**	3	−1.5	0.2	0.12	0.30*	−0.29***
Arch.	habitat area, thermal stress	0.58*	3	−1.3	0.4	0.11	0.21	−0.07***
Arch.	habitat area	0.44*	2	−1.2	0.5	0.10	0.30*	
Arch.	CCA cover, thermal stress	0.56*	3	−1.0	0.8	0.09	20.6***	−0.09*
Arch.	thermal stress	0.39*	2	−0.9	0.8	0.09	−0.1*	
NWHI	habitat area	0.67*	2	3.9	0	0.83	0.27*	
MHI	coral cover	0.61***	2	11.1	0	0.52	1.02***	

centre of the chain (electronic supplementary material, figure S6). This pattern is expected by the stepping stone model and by the pattern of habitat area, but in conflict with the hypothesized positive effect of immigration on observed diversity, which showed high values at the margins. Larval inputs from Hawai'i's nearest neighbour, Johnston Atoll, an alternative potential cause of the uptick in diversity at the margins, is unlikely according to the biophysical transport model (electronic supplementary material, figure S6), as are larval inputs from farther away, which were not included. Estimates of potential larval immigration pattern were also relatively insensitive to PLD and spawning seasonality.

(d) Community-level seascape genetic analysis

A mixed modelling approach to assess the aggregated response of individual species to the seascape identified various combinations of habitat area, LGM habitat loss and coral cover as the most influential seascape correlates (table 4). Pairwise colinearity of habitat area, coral cover, thermal stress and LGM habitat loss at the 13 islands was not high ($r < 0.5$; electronic supplementary material, table S1). Model selection of composite AR provided nearly identical assessments of top drivers to the mixed modelling approach, although p-values differed. Pairing habitat area with coral cover, thermal stress or LGM habitat loss created competing top bi-variate models that explained 59% of the variation in composite AR across the archipelago (table 5). When limited to Northwestern Hawaiian Islands (NWHI), habitat area explained 67% of the variation in composite mtDNA diversity, whereas in the Main Hawaiian

Islands (MHI), coral cover was a more parsimonious model and explained 61% of variation in mtDNA AR (table 5). Overall, positive correlation between genetic diversity and habitat area appears to be the strongest community-level seascape genetics relationship. Interestingly, equal numbers of species showed positive and negative correlations with habitat area individually. When those species with the highest positive individual correlations to habitat area are removed from the composite mean, the strong positive effect of habitat area on composite diversity remains and exceeds any single species's correlation to habitat (electronic supplementary material, figure S7).

4. Discussion

This study breaks new ground in exploring how relationships between genetic diversity and seascape variables change when examined at the species and community levels. Variation in seascape relationships across species is high, as expected of a marine meta-community with diverse life histories and tendency for high gene flow. Nevertheless, there are logical ways in which species traits predict which seascape factors most influence a given species. Species with higher spatial genetic structure showed stronger links to seascape metrics, because they are demographically more sensitive to the local environment compared with species with genetic stocks spanning many islands. Herbivores responded positively to species richness of corals, which provide shelter and algal substrate, and negatively to fish species richness, perhaps due to

competition and predation. Shallower species showed stronger effects of both wave energy, where effects are concentrated, and past bottlenecks because shallow areas fluctuate considerably with sea level compared with deeper areas [23]. The lingering signal of historical habitat loss corroborates evidence for genetic bottlenecks tied to the last glacial maximum elsewhere in Polynesia, and may be most pronounced at islands with large shallow lagoons [24]. Only deep-water species with high genetic structuring, a rare combination in our dataset, showed strong influence of potential larval immigration (inferred from circulation models). The small population sizes at the many small islands in the middle of the island chain probably lead to such low diversity levels for shallow species there that even a relatively high rate of immigration cannot compensate. However, deeper species have larger and more stable habitat area even at small islands, and thus we speculate that larval immigration emerges as a detectable impact at depth, as long as genetic structuring is sufficient to create strong localized genetic responses.

(a) Congruence and composite AR

Low-dispersal species had a tendency for higher congruence of AR, echoing previous findings for European plants where dispersal ability predicted level of congruence in genetic diversity [7]. Finding significant congruence lends support to the interpretability of composite mean genetic diversity of all species, a metric which served two purposes. First, it improved upon the traditional use of one or a few exemplar species which are often used to represent genetic diversity in reef conservation planning. The conventional exemplar species strategy would fail to uncover most of the significant driver relationships found here and, due to the variation in genetic patterns among species, would not represent much of the reef-associated community. Second, it leveraged the many single-marker datasets as replicates to overcome sampling error in assessing the role of seascape factors in shaping patterns of genetic diversity at the community level. Despite finding that species traits created some differences in how species related to seascape drivers, using mean genetic diversity did not obscure relationships to seascape drivers. Instead, the mean showed patterns that fit with most of our hypotheses. Finding the same strong correlations to habitat area and fish species richness when composite genetic diversity was calculated from 11 multi-locus nuclear datasets indicates these relationships are not marker-specific. The overall message of the results is that the composite mean seascape models show the same qualitative patterns as the mixed modelling of species-level data, and the same patterns whether it is calculated with mtDNA or nucDNA samples. Although combining data from non-homologous mtDNA loci into a single mean is not ideal, the trends across locus types are invariant, and thus the non-homologous marker dataset is probably an adequate proxy for a homologous marker dataset in this situation (i.e. with a large sample of species and an island chain). These findings support the value of continued careful synthesis of the thousands of existing 'last-gen' datasets in the literature, and help establish the robustness and utility of the composite mean as a tool for applied population genetics. In sum, although using a composite mean comes with accompanying errors and assumptions, it has a potentially important role to play in distilling complex data and aiding the uptake of genetic data by conservation and management.

(b) Roles of seascape drivers

Habitat area appears to be a dominant influence on genetic diversity, with high predictive power for both genetic and species-level diversity across islands. The dominant effect of habitat area on diversity is widely understood [4]. Another logical finding was the strong influence of coral cover on reef genetic diversity. Coral cover acts as a modifier of habitat quality and quantity, by providing shelter, food and rugosity. Confirming basic relationships between diversity and habitat factors bolsters the validity of the unexpected influences of additional seascape factors at the island scale.

Finding little influence of our estimate of potential immigration on mean genetic diversity fits with the rarity of isolation by distance among these species shown previously across the Hawaiian Archipelago [14]. Interestingly, dispersal metrics had little spatial variation relative to the variation in habitat size across islands. This suggests that rate of drift may be more spatially variable than migration, and thus more influential to genetic diversity of shallow-water reef communities [25,26]. The increase in diversity at marginal locations is due to the coincidence that the largest habitat areas within the Hawaiian Archipelago are found at the margins, and this was confirmed by population genetic simulation (electronic supplementary material, figure S4). Additionally, realized immigration might differ from modelled biophysical transport perhaps due to effects of chemical cues from coral and CCA, post-settlement mortality or density-dependence affecting rates of settlement and recruitment [27,28].

Although the effect of recent thermal stress on genetic diversity was not one of the strongest seascape factors, it showed significant negative correlation to composite AR and roughly half of the species showed a moderate or strong correlation individually (electronic supplementary material, dataset S6). Thermal stress is known to directly affect health of coral and algal populations with cascading effects on reef habitat complexity, quality or productivity [29]. Low-latitude Pacific reefs have experienced thermal stress up to 240% higher than Hawai'i and documented ecological effects have also been more severe [30], suggesting that recent declines in genetic diversity might also be more severe at lower latitudes. Indirect or synergistic effects from other predictors or unanalysed 'latent' factors may influence which seascape drivers show strong influence on genetics. For example, thermal stress shows moderate correlation with fish and coral species richness (electronic supplementary material, figure S1), which may partially account for its correlation to genetic diversity.

(c) Conservation implications

The seascape relationships identified here lend crucial and novel empirical support to difficult, urgent and controversial decisions unfolding about how best to conserve genetic diversity at both community- and species-level scales. Specifically, these results lend a new form of support to the idea that conserving large, intact reefs with high coral cover protects the diversity of entire reef communities, and thus supports the emerging strategy of creating large-scale marine protected areas pursued by the Big Ocean Initiative, among others [31]. Active conservation of reef biodiversity will promote future resilience to mounting stress posed by increasing coastal development and adaptive potential to climate change [31–33]. In particular, the likely negative effect of thermal stress on genetic diversity may signal the potential for global warming to compromise the adaptive capacity and

genetic integrity of not just corals but the entire coral reef community. Papahānaumokuākea Marine National Monument safeguards the majority of Hawaiian reef biodiversity, but populations in the MHI tend to be genetically distinct [13] and warrant additional protections. Our study shows that genetic diversity varies across the MHI, extending the argument for targeted protection of genetic resources of reefs throughout the MHI. Notably, Hawai'i Island has the greatest amount of coral reef area, harbours maximal genetic diversity, on average, and probably serves a unique and particularly influential role due to its large size at the margin of the chain. Nihoa and Ni'ihau, which sit at the transition between NWHI and MHI, also warrant particular focus for future research and protection due to intriguing combinations of high genetic divergence and unusual fish composition [34].

(d) Eco-genetic feedbacks

It might seem puzzling that the multi-species mean showed strong positive correlation to habitat area even though equal numbers of individual species showed positive and negative correlations to habitat area (electronic supplementary material, figure S7). There are many ways that macro-ecological forces and species interactions could generate this seemingly contradictory pair of results. Note that which species are included in which islands' multi-species means is driven largely by ecology—most species do not occur commonly at all islands, and sampling efforts were in most cases exhaustive, such that the lack of DNA samples from some islands reflects the species's rarity or the absence there. The cause of such absence or rarity can not only be due to environmental mismatch (e.g. colder temperatures at the northern end of the chain), but also due to lottery/precedence effects on recruitment success or high mortality due to competitive exclusion, or intense predation.

A specific example that demonstrates the possible linkage of community ecology to community-level genetic patterns in Hawai'i is that endemic and widespread reef fishes exhibit opposite linear gradients in density across the archipelago, both in aggregate and for pairs of endemic and non-endemic congeners [34]. Whereas the density ratio of endemics to non-endemics is 2 : 3 at the southeastern end of the chain, it flips to 3 : 2 at the northwestern end, with an inflection point of 1 : 1 around French Frigate Shoals at the midpoint. The environmental climate of the northwestern end of the chain is unusual for Pacific coral reefs, and endemics show a competitive advantage over non-endemics under these conditions. Our genetic samples also showed an increase in proportion of endemics with latitude that was driven by the sampled fishes. Assuming genetic diversity scales with numerical abundance and only abundant species are genetically sampled, this swapping of competitive dominance across islands would lead to the same calculations of mean AR despite individual species showing very different spatial patterns of AR. In support of this possibility, we found that for the 10 endemic fishes in our dataset the mean correlation of genetic diversity to habitat is negative ($r = -0.06$), while for the 13 non-endemic fishes it is positive ($r = 0.11$). Nevertheless, other processes aside from endemism probably also contribute to the emergent trends in composite diversity.

As a whole, community-level genetic diversity reflects both the bottom-up result of each species's population genetic history, as is well known and understood, but also top-down influences of community filtering of species composition and interspecific constraints on the composite effective population size of the species assemblage. In other words, although species-level processes filter and constrain community-level genetic patterns, community-level processes also filter and constrain community-level genetic patterns. These latter effects are as yet not well documented and studied. However, recent demonstration that haplotype turnover closely tracks species turnover supports the idea that mean genetic diversity is constrained by macro-ecological forces in addition to well-understood species-level processes (e.g. genetic drift) [35].

(e) Composite genetic diversity as an emergent property?

It is interesting to consider how and why composite diversity can show higher correlation to habitat than any species shows individually. If the boost in correlation is entirely due to the large sample size (i.e. treating species as locus replicates), this suggests that all species would show the same high correlation to habitat area with better genomic sampling. However, it is likely that some fraction of species in a community will always show negative or no correlation with area due to particular species traits, species interactions and historical effects. Indeed, a recent synthesis of species–genetic diversity correlations finds both positive and negative relationships are prevalent at the species level [36]. We hypothesize that composite genetic diversity is likely to be an emergent property at the community level that shows responses to the seascape distinct from individual species' responses, due to the aforementioned linkage to other community-level traits such as species composition and diversity. The distinction between drivers of individual species' genetic diversity and drivers of multi-species genetic diversity is analogous to the distinction between drivers of abundance patterns within a species and drivers of total abundance of individuals within a community, which is well understood to be an emergent ecological property of the community. Mathematically, when species composition, species richness and/or species abundance distributions vary strongly across space, composite genetic diversity should diverge more strongly from representing the mean genetic diversity of individual species, but continue to mirror species diversity patterns. A mechanistic macro-ecological link between species and genetic diversity patterns exists if the same large-scale assembly rules that dictate species composition also influence composite genetic diversity [37]. Specifically, a species's rarity and small range size lead to low total gene diversity, and also lead to greater chance of local extinction, thereby impacting both composite genetic diversity and species composition. Further work is needed to better understand these sorts of feedbacks that link composite genetic diversity to species abundance distribution, community assembly and its drivers. Detecting macro-ecological regularities requires large sample sizes; using the multi-species mean overcomes stochasticity associated with any single dataset so that the most dominant pattern and its drivers are more powerfully assessed [12]. The composite mean may be a convenient and appropriate tool that can reveal underlying macro-ecological processes influencing diversity at all hierarchical levels, and reveal these effects more strongly than a comparison of the patterns of individual species. The findings as a whole demonstrate that continued expansion of integrative studies of community-level genetic diversity holds promise to elucidate the complex

interdependencies across biodiversity levels and provide critical information to stem the loss of global biodiversity.

9

rspb.royalsocietypublishing.org   Proc. R. Soc. B 283: 20160354

References

1. Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. 2008 Ecological consequences of genetic diversity. Ecol. Lett. 11, 609–623. (doi:10.1111/j.1461-0248.2008.01179.x)

2. Laikre L et al. 2010 Neglect of genetic diversity in implementation of the convention on biological diversity. Conserv. Biol. 24, 86–88. (doi:10.1111/j.1523-1739.2009.01425.x)

3. Beger M, Selkoe KA, Treml E, Barber PH, von der Heyden S, Crandall ED, Toonen RJ, Riginos C. 2014 Evolving coral reef conservation with genetic information. Bull. Mar. Sci. 90, 159–185. (doi:10.5343/bms.2012.1106)

4. Antonovics J. 1976 The input from population genetics: 'the new ecological genetics'. Syst. Bot. 1, 223–245. (doi:10.2307/2395303)

5. Laroche F, Jarne P, Lamy T, David P, Massol F. 2015 A neutral theory for interpreting correlations between species and genetic diversity in communities. Am. Nat. 185, 59. (doi:10.1086/678990)

6. Alvarez N et al. 2009 History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants. Ecol. Lett. 12, 632–640. (doi:10.1111/j.1461-0248.2009.01312.x)

7. Petit RJ et al. 2003 Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300, 1563–1565. (doi:10.1126/science.1083264)

8. Taberlet P et al. 2012 Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. Ecol. Lett. 15, 1439–1448. (doi:10.1111/ele.12004)

9. Whitham TG et al. 2006 A framework for community and ecosystem genetics: from genes to ecosystems. Nat. Rev. Genet. 7, 510–523. (doi:10.1038/Nrg1877)

10. Vellend M, Geber MA. 2005 Connections between species diversity and genetic diversity. Ecol. Lett. 8, 767–781. (doi:10.1111/j.1461-0248.2005.00775.x)

11. He T, Lamont BB, Krauss SL, Enright NJ, Miller BP. 2008 Covariation between intraspecific genetic diversity and species diversity within a plant functional group. J. Ecol. 96, 956–961. (doi:10.1111/j.1365-2745.2008.01402.x)

12. Papadopoulou A, Anastasiou I, Spagopoulou F, Stalimerou M, Terzopoulou S, Legakis A, Vogler AP. 2011 Testing the species–genetic diversity correlation in the Aegean Archipelago: toward a haplotype-based macroecology? Am. Nat. 178, 241–255. (doi:10.1086/660828)

13. Toonen RJ et al. 2011 Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian archipelago. J. Mar. Biol. 2011, 1–13. (doi:10.1155/2011/460173)

14. Selkoe KA, Gaggiotti OE, Lab T, Bowen BW, Toonen RJ. 2014 Emergent patterns of population genetic structure for a coral reef community. Mol. Ecol. 23, 3064–3079. (doi:10.1111/mec.12804)

15. Tajima F. 1990 Relationship between migration and DNA polymorphism in a local population. Genetics 126, 231–234.

16. MacArthur RH, Wilson EO. 1967 The theory of island biogeography. Princeton, NJ: Princeton University Press.

17. Gove JM, Williams GJ, McManus MA, Heron SF, Sandin SA, Vetter OJ, Foley DG. 2013 Quantifying climatological ranges and anomalies for Pacific coral reef ecosystems. PLoS ONE 8, e61974. (doi:10.1371/journal.pone.0061974)

18. Concepcion GT, Baums IB, Toonen RJ. 2014 Regional population structure of Montipora capitata across the Hawaiian Archipelago. Bull. Mar. Sci. 90, 257–275. (doi:10.5343/bms.2012.1109)

19. Kalinowski ST. 2005 HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol. Ecol. Notes 5, 187–189. (doi:10.1111/j.1471-8286.2004.00845.x)

20. Nei M. 1987 Molecular evolutionary genetics. New York, NY: Columbia University Press.

21. Paradis E, Claude J, Strimmer K. 2004 APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290. (doi:10.1093/bioinformatics/btg412)

22. Oksanen J et al. 2013 Vegan: community ecology package. R package version 2.0-10. See http://cran.r-project.org/web/packages/vegan/index.html.

23. Baums IB, Godwin LS, Franklin EC, Carlon DB, Toonen RJ. 2014 Discordant population expansions in four species of coral-associated Pacific hermit crabs (Anomura: Diogenidae) linked to habitat availability resulting from sea-level change. J. Biogeogr. 41, 339–352. (doi:10.1111/Jbi.12181)

24. Fauvelot C, Bernardi G, Planes S. 2003 Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. Evolution 57, 1571–1583.

25. Selkoe KA, Watson JR, White C, Horin TB, Iacchei M, Mitarai S, Siegel DA, Gaines SD, Toonen RJ. 2010 Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Mol. Ecol. 19, 3708–3726. (doi:10.1111/j.1365-294X.2010.04658.x)

26. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013 Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol. Ecol. 22, 5983–5999. (doi:10.1111/mec.12561)

27. Dixson DL, Abrego D, Hay ME. 2014 Chemically mediated behavior of recruiting corals and fishes: a tipping point that may limit reef recovery. Science 345, 892–897. (doi:10.1126/science.1255057)

28. Hixon MA, Pacala SW, Sandin SA. 2002 Population regulation: historical context and contemporary challenges of open vs. closed systems. Ecology 83, 1490–1508. (doi:10.1890/0012-9658(2002)083[1490:PRHCAC]2.0.CO;2)

29. Pratchett MS, Munday PL, Wilson SK, Graham NAJ, Cinner JE, Bellwood DR, Jones GP, Polunin NVC, McClanahan TR. 2008 Effects of climate-induced coral bleaching on coral-reef fishes—ecological and economic consequences. Oceanogr. Mar. Biol. 46, 251–296.

30. Jokiel PL, Brown EK. 2004 Global warming, regional trends and inshore environmental conditions influence coral bleaching in Hawaii. Glob. Change Biol. 10, 1627–1641. (doi:10.1111/j.1365-2486.2004.00836.x)

31. Toonen RJ et al. 2013 One size does not fit all: the emerging frontier in large-scale marine conservation. Mar. Pollut. Bull. 77, 7–10. (doi:10.1016/j.marpolbul.2013.10.039)

32. Nystrom M. 2006 Redundancy and response diversity of functional groups: implications for the resilience of coral reefs. *Ambio* **35**, 30 – 35. (doi:10.1579/0044-7447-35.1.30)

33. Singleton RL, Roberts CM. 2014 The contribution of very large marine protected areas to marine conservation: giant leaps or smoke and mirrors? *Mar. Pollut. Bull.* **87**, 7 – 10. (doi:10.1016/j.marpolbul.2014.07.067)

34. Williams ID, Baum JK, Heenan A, Hanson KM, Nadon MO, Brainard RE. 2015 Human, oceanographic and habitat drivers of Central and Western Pacific coral reef fish assemblages. *PLoS ONE*. **10**, e0120516. (doi:10.1371/journal.pone.0120516)

35. Baselga A, Gómez-Rodríguez C, Vogler AP. 2015 Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Glob. Ecol. Biogeogr*. **24**, 873 – 882. (doi:10.1111/geb.12322)

36. Vellend M, Lajoie G, Bourret A, Murria C, Kembel SW, Garant D. 2014 Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Mol. Ecol.* **23**, 2890 – 2901. (doi:10.1111/Mec.12756)

37. Bellwood DR, Hughes AR. 2001 Regional-scale assembly rules and biodiversity of coral reefs. *Science* **292**, 1532 – 1534. (doi:10.1126/science.1058635)

Contents

Appendix 1

Names and addresses of 31 contributors to the Hawaiʻi Reef Connectivity Consortium group authorship. Consortium members contributed existing data to the study, and verified proper data usage and interpretation of results.

Kimberly Andrews, Dept Fish and Wildlife Sciences, University of Idaho, 875 Perimeter Drive MS 1136, Moscow ID 83844-1136

Iliana Baums, Department of Biology, The Pennsylvania State University, 208 Mueller Laboratory University Park, PA, 16802, USA

Moisés A. Bernal, California Academy of Sciences, 55 Music Concourse Drive, Golden Gate Park, San Francisco, CA 94118, USA

Giacomo Bernardi, Department of Ecology and Evolutionary Biology, University of California Santa Cruz100 Shaffer Road, Santa Cruz, California, 95060

Christopher Bird, Marine Biology Program, Department of Life Sciences, Texas A & M University–Corpus Christi, 6300 Ocean Drive, Corpus Christi, Texas 78412, USA

Holly Bolick, Bishop Museum, 1525 Bernice St, Honolulu, HI, 96817, USA

Brian Bowen, Hawaiʻi Institute of Marine Biology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Richard Coleman, Hawaiʻi Institute of Marine Biology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Gregory T. Concepcion, Pacific Biosciences, 1380 Willow Rd, Menlo Park, CA, 94025, USA

Matthew T. Craig, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 8901 La Jolla Shores Drive, La Jolla, CA, 92037, USA

Toby S. Daly-Engel, Department of Biology, University of West Florida, 11000 University Parkway, Pensacola, FL, 32514, USA

Joseph D. DiBattista, Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia

Jeffrey Eble, Center for Environmental Bioremediation and Diagnostics, University of West Florida, 11000 University Parkway, Pensacola, FL, 32514, USA

Iria Fernandez-Silva, California Academy of Sciences, 55 Music Concourse Drive, Golden Gate Park, San Francisco, CA 94118, USA

Erik C. Franklin, Hawaiʻi Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Alan Friedlander, Fisheries Ecology Research Lab, Department of Biology, University of Hawaiʻi, Honolulu, Hawaiʻi 96822, USA and Pristine Seas-National Geographic, Washington DC 20036, USA.

Michelle R. Gaither, School of Biological and Biomedical Sciences, Durham University, South Road, Durham DH1 3LE, UK *and* California Academy of Sciences, Ichthyology, 55 Music Concourse Drive, San Francisco, CA 94118, USA

Jamison Gove, Ecosystems and Oceanography Division, Pacific Islands Fisheries Center, National Oceanographic and Atmospheric Administration Inouye Regional Center, Honolulu, HI, 96818

Mathew Iacchei, Department of Oceanography, University of Hawaiʻi at Mānoa, 1000 Pope Rd., Honolulu, HI, 96822, USA

Yanli Jia, International Pacific Research Center, University of Hawaiʻi at Mānoa, Honolulu, HI, 96822, USA

Donald Kobayashi, Ecosystems and Oceanography Division, Pacific Islands Fisheries Center, National Oceanographic and Atmospheric Administration Inouye Regional Center, Honolulu, HI, 96818

Nicholas R. Polato, Department of Ecology & Evolutionary Biology, 215 Tower Rd., Cornell University, Ithaca, NY 14853, USA

Malia Ana J. Rivera, Hawaiʻi Institute of Marine Biology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Luiz A. Rocha, California Academy of Sciences, 55 Music Concourse Drive, Golden Gate Park, San Francisco, CA 94118, USA

Joshua Reece, Valdosta State University, 1500 North Patterson Street, Valdosta, GA 31698, USA

Derek Skillings, Hawaiʻi Institute of Marine Biology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Scott R. Santos, Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

Zoltan Szabo, Hawaiʻi Institute of Marine Biology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Molly Timmers, Hawaiʻi Institute of Marine Biology, University of Hawaiʻi, Kāneʻohe, HI 97644, USA

Lisa Wedding, Center for Ocean Solutions, Stanford University, 99 Pacific Street, Suite 555E, Monterey, CA 93940

Gareth J. Williams, Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, La Jolla, CA 92083, USA

Nicholas M. Whitney, Behavioral Ecology and Physiology Program, Mote Marine Laboratory, Sarasota, FL 34236

Supplemental Methods concerning input data

Assembly of existing seascape datasets. 'Habitat area' used Log_{10} (x+1) transformed estimates of total shallow-water area within the 10-fathom depth curve of each island [1]. An alternative estimate of habitat area, the total hard-bottom area classified from satellite imagery (see below), strongly correlated with this metric and produced identical correlation to rarefied mean AR as the bathymetric estimate (Table S1).

Percentages of 'Coral cover' and 'Crustose coralline algal (CCA) cover' used published estimates produced from IKONOS satellite imagery covering 0-30m depth range in shallow-water coral reef ecosystems [2, 3]. Grid cells were classified as one of six benthic cover types when at least 10% of the grid cell contained a single cover type (i.e., hard coral, CCA, 'uncolonized hard bottom,' macroalgae, sand and 'hard bottom indeterminate cover'). In the MHI, finer subcategories of these six types were combined to match NWHI classes. Necker was excluded from analyses of benthic cover because of large classification gaps. Total area metrics were converted to percentage cover by dividing number of typed grid cells by the total grid cells classified at each island.

Multi-year environmental data were available for sea surface temperature (SST), wave energy, irradiance and chlorophyll-a as island level averages of all grid cells within the 30m isobath [4]. 'Thermal stress' represented the frequency of Hotspot events, as defined by NOAA's Coral Reef Watch, when SST exceeded the maximum monthly mean temperature over years 1985-2000, measured at 4 km. 'Wave disturbance' represented yearly average of maximum monthly mean wave energy over years 1997-2010, measured at one degree.

During the last glacial maximum (LGM) 18,000 years ago, sea level was as much as 120 m lower than present [5], severely restricting coral reef habitat in some parts of the archipelago. 'LGM habitat loss' compared current bathymetric area within 0-45 m depth and area within 120-165 m depth, which is the analogous depth range when sea level was 120m lower during the LGM [6]. Necker and Gardner Pinnacles were excluded because unusual bathymetry at these pinnacles makes this approach uncertain. To estimate the likely severity of population bottlenecks due to habitat loss during the LGM, we subtracted from one the ratio of LGM to present day area estimates.

Distance to nearest island was used as a simple metric of probable connectivity. 'Nearest neighbor distance' was estimated on GoogleEarth using path distances between approximate centroids of islands and selecting the shortest distance to a neighboring island. A second, bio-physical metric of connectivity is described below.

Larval Dispersal Model: A particle tracking transport model based simulated larval dispersal using a regional implementation of the Massachusetts Institute of Technology general circulation model [7] (MITgcm) for years 2009-2014 configured for the Hawaiian Archipelago (175°E-210°E and 15°N-35°N), with hourly output at 4 km horizontal resolution across 50 vertical layers. The model was forced by surface wind stress at a resolution of 0.25 degree from satellite scatterometer measurements for the period of May 2009 to May 2014 at daily intervals, using monthly climatological surface heat and fresh water fluxes, and ocean conditions from a global ocean estimation system at 8 km resolution at the lateral open boundaries. Reef habitat was defined for 687 pixels, representing all shallow reefs and submerged banks present in the Archipelago, using a combination of IKONOS-derived data [35, 36] and modeled coral distributions in the MHI [8]. The relative amount of larvae released per reef cell was determined by the percentage of reef habitat in each pixel. Larvae were released every 2 weeks for a maximum pelagic larval duration (PLD) of 45 days, an approximate median duration from the range of

PLD known for our genetically sampled species and the majority of common reef taxa [9]. Relative dispersal rates in Hawaii were insensitive to PLD beyond 40 days.

Larval dispersal estimates were compared from two alternative biophysical modeling approaches, an advection transport algorithm and an individual-based model (IBM), that both used the MITgcm data, the same habitat designations, and a 45-day maximum PLD bounded to 3 – 33 m depth. For the IBM, 50 particles were released each day (totaling 1673 days) from each habitat pixel, creating >57 million particles total/run. Particles were kept at 50 meters depth, allowed to passively disperse for 45 days without directional behavior. Particles that reached <5 km distance from defined reef pixels after 45 days were deemed settled and tallied by island. Results were averaged at monthly resolution for all release dates. For the second larval dispersal simulation, an advection transport algorithm [10, 11], dispersal kernels were resolved to densities of < 1/60 million larvae per km^2, allowing rare and/or infrequent events to be quantified. For both models, outputs were compared when averaged across all release dates, and when limited to release dates April through July, a window when the majority of the species in the genetic dataset spawn and settle. The results were highly correlated for release dates and for the two alternative modeling approaches. Thus only the data produced by the advection transport model for all months was used further.

A single metric from this output was used in analysis to best represent 'potential larval immigration': an estimate of in-closeness centrality, which measures the relative distances each node (island) is from all other nodes based on dispersal to that node (i.e., immigration) [12]. High closeness means that a site has strong, direct connections to many other sites instead of indirect or no connections. Exploratory analysis of other metrics from the model output, such as number of settlers, self-recruitment rate, betweeness centrality and out-closeness centrality revealed no further insights about mean genetic diversity.

Species richness datasets. 'Fish Species Richness' and 'Coral Species Richness' estimates came from coral and fish species counts collected on underwater visual transects by the Pacific RAMP, conducted by the NOAA Pacific Islands Fisheries Science Center's Coral Reef Ecosystem Division, from 2011-2012 for fishes and from 2006-2010 for corals. Estimates of mobile invertebrate species richness were unavailable. Fish surveys were based on a stratified random design across reef zone and depth zones of <30m depth on hard bottom habitats with a stationary point count [13]. Coral surveys were conducted at permanent sites distributed haphazardly within reef zones at <30m depth on hard bottom habitats using a linear-point intercept method [14]. Number of sites surveyed per island was roughly proportional to reef area. Both methods identified taxa to the finest taxonomic level possible and only observations at the species level were used in diversity calculations. Sites were pooled by island [15] and incidence-based bootstrapped 'chao' estimates were calculated with the *specpool* function in the R package *vegan* [16]; this procedure corrects for varying numbers of transects per island.

Colinearity of seascape predictors: Before model selection, colinearity between pairs of predictor variables was assessed with significance testing of linear and quadratic regression, and principal components analysis (Fig. S1). Regression tests showed covariance of latitude, coral cover and wave disturbance to be the most prominent concern. Wave disturbance increases linearly with latitude with near-perfect correlation. Coral cover declines linearly with latitude (and waves) until an asymptote north of 24 ˚N, creating a strong curvilinear relationship. Therefore of the three predictors, we include only

coral cover in all analyses because its positive correlations with both mean genetic diversity and fish species richness are more interpretable than the apparent bowl-shaped relationships of latitude and wave disturbance to diversity metrics (Fig. S5). However, latent effects of latitude and/or wave disturbance may confound interpretation of the relationship between coral cover and mean genetic diversity. Also, the co-variation of fish species richness, coral cover and thermal stress make the detected effects ascribed to any one of these factors potentially augmented effects from the other two (Fig. S1).

Other latitudinal gradients in the predictor data are noted below. LGM habitat loss showed a significant negative quadratic relationship with latitude, caused by four extreme low values at the archipelago's margins (r^2=0.64, p=0.004). We repeated analyses involving LGM habitat loss without these sites (Hawaiʻi, Maui, Molokaʻi and Kure) and found that significance of its relationship to genetic diversity were identical, so a latent effect of latitude is unlikely. Thermal stress showed a moderate positive quadratic fit to latitude and highest values at Lisianski, the epicenter of observed coral bleaching in the NWHI in 2002 [17]. Because the oceanographic mechanism producing this gradient is unknown [17] and thermal stress showed no linear or quadratic relationship to any other predictors, we retained this predictor. The remaining predictors, Habitat area, CCA cover, and *M. capitata* genetic diversity showed no significant linear or quadratic colinearity (i.e., Pearson's r>0.7; Table S1).

Linear modeling of allelic richness: Although Poisson-distribution with log link function and an over-dispersion parameter are often used in GLMs of high-variance count data (ie., richness), Poisson model fits and P values were identical to normal-distributed GLMs with identity link functions (i.e., ordinary linear models), so AIC-based model selection used the latter. Indeed, because the species richness and mean allelic richness data used here have no zeros and are means of count data, their distributions are expected to be symmetric and bell shaped, albeit limited to positive values. Species Richness data was natural log transformed. For all GLM models, all pairwise relationships, interactions, and residuals of multiple regressions were inspected to determine appropriate treatment, and residuals were tested for spatial autocorrelation.

Simulation of observed spatial pattern of composite genetic diversity. We simulated a linear stepping-stone model at equilibrium consisting of 12 populations and compared two scenarios: (1) all populations had equal effective sizes, and (2) the two marginal populations were three times larger than all other populations. Simulations were run with *Fastsimcoal2.5* [18]. Results are based on 200 replicates of each scenario and sample size of 10 haploid individuals. We simulated mtDNA sequences 1kb in length and mutation rate = $2x10^{-6}$. In the low-migration scenario *m* = 0.001 between neighbouring populations (i.e. total proportion of migrants in non-marginal demes was 0.002). In the high-migration scenario *m* = 0.01 (i.e. total proportion of migrants in non-marginal demes = 0.02). The results show that when migration rates are low (Fig. S3a), the effect of the reduced migration at the margins is more than compensated by the larger effective size of the marginal populations. The same overall pattern is observed for higher migration rates (Fig. S3b) but in this case, genetic diversity is more evenly distributed among demes so the pattern is less pronounced.

Sensitivity Analyses Methods and Results

We assessed sensitivity of results to various aspects of the study design, detailed below.

1. Marker Type - Due to existing concern about reliability of single-locus genetic data and interpretability of mitochondrial data, we sought to confirm that spatial patterns of AR based on mitochondrial data are consistent with other marker types within a species. We were able to compare rarefied AR from nuclear and mitochondrial datasets individually for 8 species with both types of markers available. Comparisons used Pearson's r values and confirmed positive covariation (Table S2).

Eleven species total were sampled with multiple nuclear markers across the Hawaiian Archipelago (Dataset S2). This nucDNA dataset includes two coral species not included in the mtDNA dataset; all other species are included in the mtDNA dataset. We used HPRARE to calculate rarefied allelic richness for these 11 datasets, using a rarefaction size of 20 alleles (minimum sample size of 10 individuals). Twelve islands were sampled with at least 6 species, including Necker and Gardener. The number of species sampled per island for the nuclear dataset was correlated with that of the mtDNA dataset (r^2=0.47, p=0.02). Because variation in sampled species per island was low (it ranged 6 to 11 species), rarefying the island means was not necessary. We compared the nucDNA values of mean allelic richness to the values from mtDNA. We again excluded Necker and Gardner from analyses as one or the other behaved as a strong outlier in all tests. For the remaining ten islands, the two estimates of composite genetic diversity are well-correlated (r^2=0.50, p = 0.02) and showed similar responses to most seascape factors (Table S1). Notably, fish species richness strongly predicted composite nucDNA diversity (r^2=0.73, p=0.002), with a stronger fit than for composite mtDNA diversity at this subset of ten islands (r^2=0.44, p=0.02). Habitat area was also a significant predictor of composite nucDNA diversity (r^2=0.41, p=0.05), and this was the top model in alternative model selection of univariate and bi-variate models of composite nuclear diversity. The most parsimonious bi-variate regression models were a competing set of habitat area paired with coral cover, CCA cover, thermal stress or LGM habitat loss all within ΔAIC_c of 2.0, a similar list to the top models for the mtDNA dataset. However, none of the secondary coefficients showed statistical significance. The smaller sample size of species and islands in the nucDNA dataset likely reduces power to pick up these secondary effects.

2. Minimum sample size - To assess sensitivity of the study's results to the rarefaction size of 6 haplotypes per sample, we repeated all analyses for a rarefaction size of ten haplotypes, excluding samples smaller than this minimum. The inclusion of small samples doesn't change any of the results (Table S3). The notion that polymorphic loci require large sample sizes is in fact a common misconception in the field [19-22].

3. Diversity metric - As an alternative genetic diversity metric to allelic richness, we calculated gene diversity (GD), the mean rarity of alleles and resampled island means to get a composite genetic metric. Mean GD correlated highly with mean AR (Table S1), so only AR was used in main statistical analyses.

4. MtDNA locus – In concept, pooling allelic richness data for non-homologous mitochondrial loci may seem problematic because the marker-species diversity estimates are not replicates from a single "population" but instead taken from a highly similar set of populations each representing a genome region. However, our results indicate that error associated with combining samples taken from several regions of mtDNA does not overwhelm the signal and bias is very low. Using a large sample (e.g., adding species) is a valid way to overcome noise, and the fact that we find a significant result indicates that the

signal is strong using our sample. Thus, the estimates made from non-homologous nuclear and mitochondrial loci can be viewed as reliable and useful proxies for estimates made from homologous loci. We detail our analyses supporting this statement below.

Species sampled with COI showed higher marker diversities on average than the other marker types. However, this trend was driven by three COI datasets with unusually high genetic diversity- two hermit crabs and a limpet (Calcinus haigae, C. seurati and Cellana talcosa). All three species had very poor sampling, and as such these species were excluded from species-level analyses and have little impact on composite AR. Marker type had no effect on species level results in the redundancy analysis. We also confirmed that correlation coefficients for each species' AR to the top seascape factors (habitat area, coral area, fish species richness, and thermal stress) are uninfluenced by marker type using ANOVA. Neither marker type nor gene diversity had any effect on congruence. Specifically, a t-test showed that species sampled with COI or CB were no more likely to show high or low congruence to the composite AR values than species sampled with other loci (t=1.2, p=0.22).

Bias could nevertheless arise in the composite AR results if certain islands were predominantly sampled with mtDNA regions of low diversity. Cytochrome B shows lower diversity than other mtDNA regions in our dataset. Average of total gene diversity from the sampled species is 0.66 for CytB and ~0.73 for all other markers – not a large difference. In our case, the most influential islands to the trends in patterns and drivers of composite AR are Hawaii (highest AR) and Nihoa (lowest AR), so bias in marker composition is most crucial for these two islands. In fact, 49% of species at Hawaii were sampled with Cyt B, whereas 65% of species at Nihoa were sampled with Cyt B, potentially amplifying their difference in AR. To figure out how much bias this difference in marker composition creates, we calculated the mean total genetic diversity at each location as a weighted average using the proportion of CytB markers in the island's sample. We find the expected difference in diversity due to the different marker ratio at the two islands is 1% (mean total gene diversity of 0.696 at Hawai'i vs. 0.685 at Nihoa). In comparison, the actual difference in rarefied AR for the two locations is 22%, and the relative difference in raw mean gene diversity (more comparable to total gene diversity) is 16%. A 1% bias cannot produce a 16% effect. Thus, bias is not contributing measurably to the results.

5. Species sampled - We explored the sensitivity of the congruence results to the species included in the composite mean. Because there are 37 species in the calculation of the rarified island mean, excluding one doesn't change the mean enough to make a difference in congruence results. We chose to test the stability of the composite mean and congruence results by jackknifing in ways to purposefully disrupt them. First, we recalculated the resampled island means after omitting, as a group, the three species with the highest congruence to the resampled mean (celexa, halorn and panpen, all with congruence r>0.7). Overall distribution of congruence values to the new resampled mean was nearly identical, and the positive skew of the distribution was nearly identical (with identical statistical significance p=0.03). Although the congruence of the three omitted species to the new mean was lower (r=0.45, 0.61 and 0.69), they remained some of the highest observed congruence values. We repeated this type of test a second time, omitting four other species with high influence on the results because they were collected at many islands, including islands with small total number of species sampled (acapla, ophpic,parmul,abuabd). These species showed a mix of low and high congruence (r = 0.05, 0.13, 0.44 and 0.55). Again overall distribution and skew of congruence to the new multi-species mean were unchanged when calculated without these four species, yet the four species showed no appreciable

8

change in their individual congruence to the mean from which they were omitted (r = 0.07, 0.08, 0.39 and 0.56).

6. Taxon - We re-calculated the rarefied multi-species mean allelic richness using only fishes, to test the possibility that the fishes were driving all of the seascape relationships. Rarefaction used a sample of 8 species per island instead of 12 to accommodate lower samples at some islands. Correlation of the fish-only mean to fish species richness, habitat area and other seascape factors were very noisy due to certain islands becoming large outliers. Even with these islands excluded, the mean of all 47 species showed higher correlation to fish species richness than the mean of the 37 fishes (all: R2= 0.67, p = 0.001; fishes: R2 = 0.46, p = 0.01; Fig S8). Most likely, running analyses with fish-only means of genetic diversity, although more comparable to fish species richness estimates, reduces sample sizes at some islands to low levels, compromising statistical power of seascape analyses.

Literature Cited

1.   Rohmann S.O., Hayes J.J., Newhall R.C., Monaco M.E., Grigg R.W. 2005 The area of potential shallow-water tropical and subtropical coral ecosystems in the United States. *Coral Reefs* **24**(3), 370-383. (doi:10.1007/s00338-005-0014-4).

2.   Battista J., Costa B.M., Anderson S.M. 2007 Shallow-water benthic habitats of the main eight Hawaiian Islands (DVD). In *NOAA Technical Memorandum NOS NCCOS* (Silver Spring, MD.

3.   Friedlander A., Keller K., Wedding L.M., Clarke A., Monaco M.E. 2009 A marine biogeographic assessment of the Northwestern Hawaiian Islands. In *NOAA Technical Memo NOS NCCOS 84* (p. 363. Silver Spring, MD.

4.   Gove J.M., Williams G.J., McManus M.A., Heron S.F., Sandin S.A., Vetter O.J., Foley D.G. 2013 Quantifying climatological ranges and anomalies for Pacific coral reef ecosystems. *Plos One* **8**(4), e61974. (doi:10.1371/journal.pone.0061974).

5.   Bard E., Hamelin B., Arnold M., Montaggioni L., Cabioch G., Faure G., Rougerie F. 1996 Deglacial sea-level record from Tahiti corals and the timing of global meltwater discharge. *Nature* **382**(6588), 241-244. (doi:Doi 10.1038/382241a0).

6.   Baums I.B., Godwin L.S., Franklin E.C., Carlon D.B., Toonen R.J. 2014 Discordant population expansions in four species of coral-associated Pacific hermit crabs (Anomura: Diogenidae) linked to habitat availability resulting from sea-level change. *J Biogeogr* **41**(2), 339-352. (doi:10.1111/Jbi.12181).

7.   Marshall J., Adcroft A., Hill C., Perelman L., Heisey C. 1997 A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research* **102**, 5753-5766.

8.   Franklin E.C., Jokiel P.L., Donahue M.J. 2013 Predictive modeling of coral distribution and abundance in the Hawaiian Islands. *Mar Ecol Prog Ser* **481**, 121-132. (doi:10.3354/meps10252).

9.   Shanks A.L. 2009 Pelagic larval duration and dispersal distance revisited. *Biological Bulletin* **216**(3), 373-385.

10.   Treml E.A., Halpin P.N., Urban D.L., Pratson L.F. 2008 Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecol* **23**, 19-36.

11.   Treml E.A., Roberts J.J., Chao Y., Halpin P.N., Possingham H.P., Riginos C. 2012 Reproductive Output and Duration of the Pelagic Larval Stage Determine Seascape-Wide Connectivity of Marine Populations. *Integr Comp Biol* **52**(4), 525-537. (doi:10.1093/icb/ics101).

12.   Borgatti S.P. 2005 Centrality and network flow. *Social Networks* **27**(1), 55-71. (doi:http://dx.doi.org/10.1016/j.socnet.2004.11.008).

13.   Williams I.D., Richards B.L., Sandin S.A., Baum J.K., Schroeder R.E., Nadon M.O., Zgliczynski B., Craig P., McIlwain J.L., Brainard R.E. 2011 Differences in reef fish assemblages between populated and remote reefs spanning multiple archipelagos across the central and western Pacific. *Journal of Marine Biology* **2011**, 14. (doi:10.1155/2011/826234).

14.   Brainard R., Asher J., Blyth-Skyrme V., Coccagna E.F., Dennis K., Donovan M.K., Gove J.M., Kenyon J., Looney E.E., Miller J.E., et al. 2012 Coral reef ecosystem monitoring report of the Mariana Archipelago: 2003-2007. In *PIFSC Special Publication* (p. 1019, Pacific Islands Fisheries Science Center,.

15.   Palmer M.W. 1990 The estimation of species richness by extrapolation. *Ecology* **71**, 1195-1198.

16.   Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2013 *vegan: Community Ecology Package. R package version 2.0-10.*

17.   Hoeke R., Brainard R., Moffitt R.B., Kenyon J. 2006 Oceanographic conditions implicated in the 2002 Northwestern Hawaiian Islands bleaching event. *Proc 10th Internat Coral Reef Symp*, 718-723.

18.   Excoffier L., Foll. M. 2011 Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332-1334.

19.   Yan L.N. and Zhang D.X. 2004. Effects of sample size on various genetic diversity measures in population genetic study with microsatellite DNA markers. *Acta Zool Sinica* **2**.

20.   Kalinowski S.T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* **34**, 33-36.

21.   Hale M.L., Burg T.M., Steeves T.E. 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PloS one*, **7**, e45170.

22.   Karl S.A., Toonen R.J., Grant W.S., Bowen, B. W. 2012. Common misconceptions in molecular ecology: echoes of the modern synthesis. *Mol Ecol* **21**, 4171-4189.

Table S1. Colinearity matrix based on Pearson's r values for all genetic diversity and seascape variables at 13 islands sampled across the archipelago. Column headers correspond to row labels. AR is allelic richness, mtDNA genetic metrics use 47 datasets, nuclear AR uses 11 multi-locus nuclear datasets sampled at 10 islands, GD is gene diversity, F_{ST} is genetic isolation.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Mean mtDNA AR	1.00																
2. Mean NucDNA AR	0.70	1.00															
3. Mean mtDNA GD	0.94	0.54	1.00														
4. Mean mtDNA F_{ST}	-0.67	-0.39	-0.60	1.00													
5. Habitat Area†	0.68	0.64	0.61	-0.51	1.00												
6. Habitat Area††	0.72	0.76	0.56	-0.36	0.80	1.00											
7. Coral Cover	0.56	0.62	0.48	-0.12	0.32	0.59	1.00										
8. CCA Cover	0.52	0.09	0.66	-0.02	0.41	0.31	0.39	1.00									
9. LGM Habitat Loss	-0.36	-0.26	-0.37	0.11	0.03	-0.20	-0.52	-0.10	1.00								
10. Nearest Neighbor	-0.16	-0.51	-0.10	0.17	-0.35	-0.39	-0.12	0.08	0.34	1.00							
11. Larval Immigration	-0.30	-0.26	-0.34	0.26	0.14	0.09	0.08	0.24	0.52	0.18	1.00						
12. Fish Spp Richness	0.72	0.85	0.53	-0.44	0.59	0.83	0.61	0.02	-0.47	-0.33	-0.25	1.00					
13. Coral Spp Richness	0.53	0.23	0.33	-0.52	0.57	0.59	0.52	0.11	-0.18	0.39	0.31	0.63	1.00				
14. Thermal Stress	-0.62	-0.59	-0.54	0.17	-0.44	-0.66	-0.42	-0.20	0.48	0.24	0.48	-0.69	-0.16	1.00			
15. *M. capitata* AR	-0.47	-0.29	-0.43	0.05	-0.33	-0.53	-0.47	-0.31	0.41	0.20	-0.08	-0.59	-0.39	0.09	1.00		
16. Wave Disturbance	-0.10	-0.26	-0.01	-0.35	-0.18	-0.49	-0.73	-0.28	0.33	0.21	-0.40	-0.35	-0.27	0.37	0.25	1.00	
17. Latitude	-0.01	-0.21	0.08	-0.40	-0.13	-0.43	-0.71	-0.26	0.23	0.15	-0.52	-0.26	-0.28	0.22	0.26	0.98	1.00

† Based on bathymetry
†† Based on classified satellite imagery

7

Table S2. Differences in main results based on rarefying allelic richness to 6 haplotypes (RAR 6) versus 10 haplotypes (RAR 10). A. Redundancy analysis; 'cum.' indicates cumulative. B. Congruence analysis. C. Seascape model selection analysis. * is $p \leq 0.05$, ** is $p \leq 0.01$, *** is $p \leq 0.001$.

A. Redundancy Analysis

Dataset	cum. r^2	axis 1 r^2	axis 2 r^2
RAR 6	0.47*	0.48***	0.37***
RAR 10	0.49*	0.54***	0.28***

B. Congruence

Dataset	Mean r	t-test	N
RAR 6	0.13	2.21*	36
RAR 10	0.16	2.38**	34

C. Seascape model selection of composite AR

Predictors	RAR 6 Adj. r^2	Δ AIC$_c$	RAR 10 Adj. r^2	Δ AIC$_c$
Habitat area, Coral cover	0.59***	0.00	0.63*	0.00
Habitat area, LGM habitat loss	0.59***	0.23	0.57*	0.68
Habitat area, Thermal stress	0.58**	0.40	0.40	6.29
Habitat area	0.44**	0.50	0.35*	2.57
CCA cover, Thermal stress	0.56**	0.76	0.53	3.37
Thermal stress	0.39*	0.82	0.23	4.57
% Coral cover	0.34*	2.04	0.48*	2.24
% Coral cover, LGM habitat loss	0.34	6.26	0.50*	4.26

Table S3. Correlation of spatial patterns of allelic richness for 8 fishes with paired nuclear and mitochondrial datasets. All samples rarefied to 10 individuals. Sources indicated by superscripts.

Species	Second Marker type	Loci	Pearson's r value	N islands
Abudefduf abdominalis[1]	Nuclear sequence	3	0.83	8
Abudefduf vaigiensis[1]	Nuclear sequence	3	0.61	6
Epinephelus/Hyporthodus quernus[2]	Microsatellite	6	0.32	9
Etelis coruscans[3]	Microsatellite	8	0.13	7
Etelis marshi[3]	Microsatellite	9	0.06	9
Lutjanus kasmira[4]	Nuclear sequence	3	0.26	8
Pristipomoides filamentosus[5]	Microsatellite	13	0.11	8
Zebrasoma flavescens[6]	Microsatellite	13	0.57	12

Source Data

1. Coleman RR, Gaither MR, Kimokeo B, Stanton FG, Bowen B, Toonen RJ (2014) Large-scale introduction of the Indo-Pacific damselfish Abudefduf vaigiensis into Hawai'i promotes genetic swamping of the endemic congener A. abdominalis. *Molecular Ecology*, 23, 5552-5565.

2. Rivera MAJ, Andrews KR, Kobayashi DR, Wren JLK, Kelley C, Roderick GK, Toonen RJ (2011) Genetic analyses and simulations of larval dispersal reveal distinct populations and directional connectivity across the range of the Hawaiian Grouper (*Epinephelus quernus*). Journal of Marine Biology, 2011, 1–11.

3. Andrews KR, Moriwake VN, Wilcox C, Grau EG, Kelley C, Pyle RL, Bowen BW (2014) Phylogeographic analyses of submesophotic snappers *Etelis coruscans* and *Etelis "marshi"* (Family Lutjanidae) reveal concordant genetic structure across the Hawaiian Archipelago. Plos One 9, e91665-e91665.

4. Gaither MR, Bowen BW, Toonen RJ, Planes S, Messmer V, Earle J, Robertson DR (2010) Genetic consequences of introducing allopatric lineages of Bluestriped Snapper (*Lutjanus kasmira*) to Hawaii. Molecular Ecology, 19, 1107-1121.

5. Gaither MR, Jones S a, Kelley C, Newman SJ, Sorenson L, Bowen BW (2011) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian archipelago. PloS one, 6, e28913.
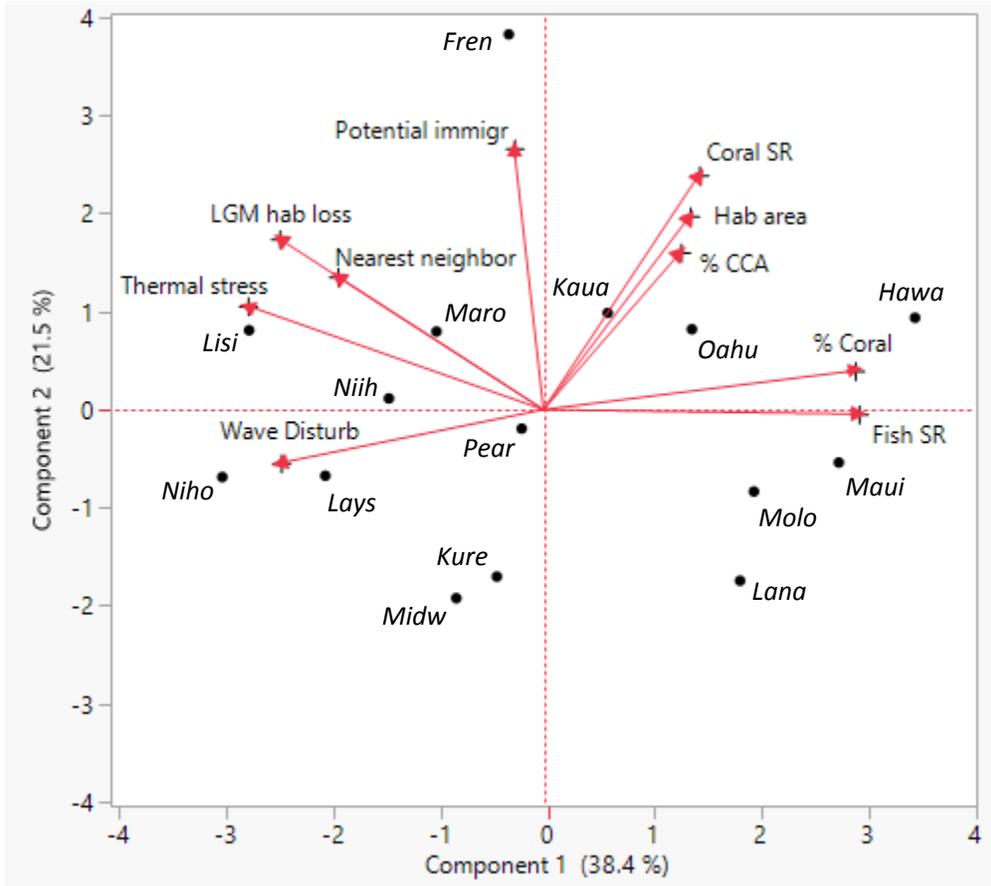
6. Eble J, Toonen R, Sorenson L, Basch L, Papastamatiou Y, Bowen B (2011) Escaping paradise: larval export from Hawaii in an Indo-Pacific reef fish, the yellow tang *Zebrasoma flavescens*. Marine Ecology Progress Series, 428, 245–258.

**Figure S1. Principal Components Analysis of seascape factors at all islands.** See Table 2 in main text for explanation of factors.

59 **Figure S2. Congruence of spatial patterns of allelic richness across species.** Distribution of
60 correlations between individual species' mtDNA allelic richness and the composite mtDNA
61 mean for 34 species sampled at ≥ 5 islands. Box plot horizontal lines indicate maxima and
62 quartiles, diamond is centered on mean with tips at 95% confidence interval; red bracket shows
63 range of 50% observation density.
64
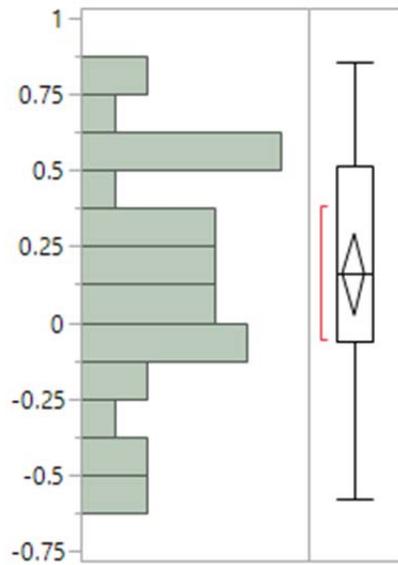
Figure S3. **Plots of composite mtDNA allelic richness against 12 factors.** Numbers indicate islands in order from southeast to northwest (1. Hawaiʻi, 2. Maui, 3. Molokaʻi, 4. Oahu, 5. Kauaʻi, 6. Niʻihau, 7. Nihoa, 8. French Frigate Shoals, 9. Maro, 10. Laysan, 11. Pearl and Hermes, 12. Midway, 13. Kure).
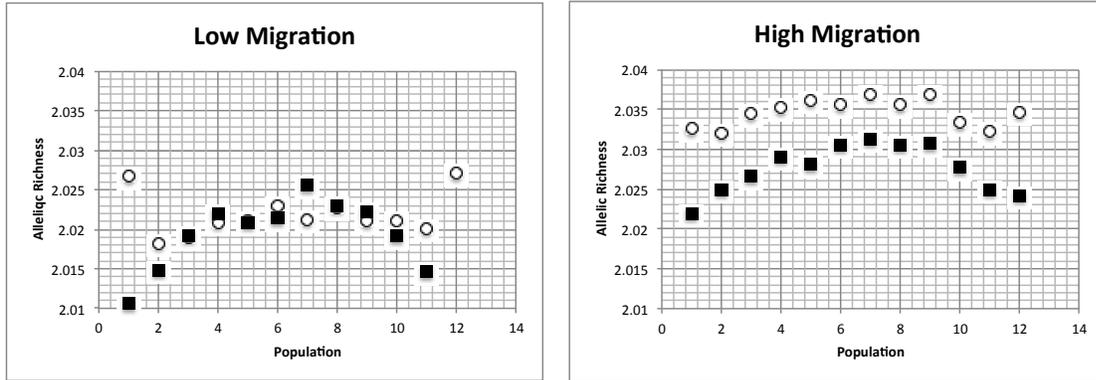

17
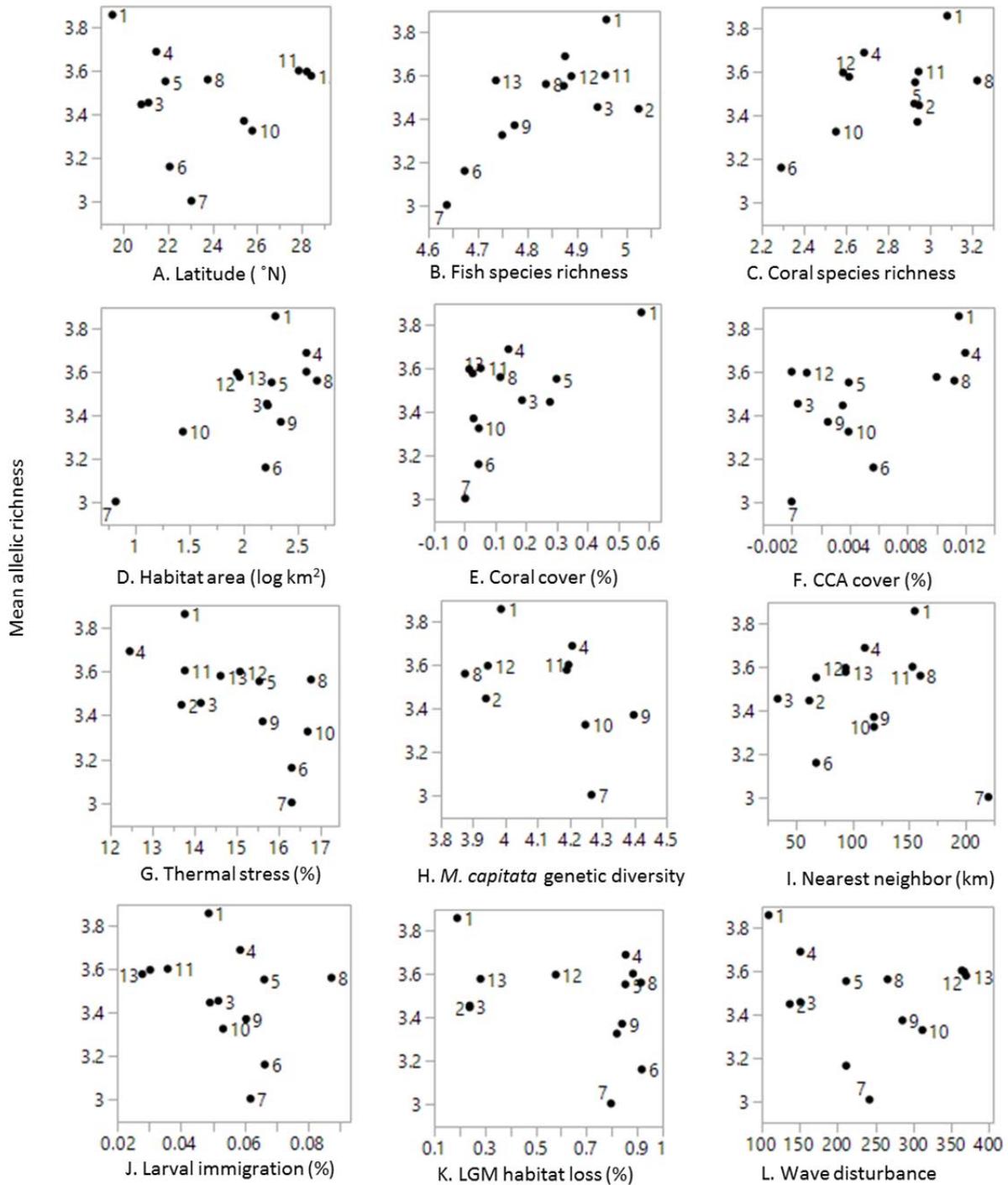
Fig. S4. Output of genetic simulations. Plots show allelic richness along a linear habitat array with migration among neighboring populations only. Solid squares correspond to the scenario where all populations have equal sizes ($N_e = 1000$); circles correspond to a scenario were the two marginal populations are three times larger than the others ($N_e = 3000$ vs $N_e = 1000$).

**Figure S5. Redundancy analysis relating genetic diversity, seascape variables and species'
traits.** Predictive power of species traits (blue font) for the strength of spatial correlation of each
species' AR values with seascape factors (red font). Seascape factors adjacent to trait vectors are
under positive influence, factors opposing are under negative influence; these associations are
driven by the species names clustered adjacent or opposing to the vectors (black font, see Dataset
S1 for full names).

87   **Figure S6. Modeled larval immigration rates.** A. Map of Hawaii overlaying estimates of
88   'larval immigration' (i.e., in-closeness centrality; green circles) and pairwise rates of larval
89   exchange (greyscale lines), with southeastward movement above the archipelago and
90   northwestern movement below the archipelago. Island to the south is Johnston Atoll. B. Plot of
91   Larval Immigration against latitude for the 13 Hawaiian Islands analyzed.

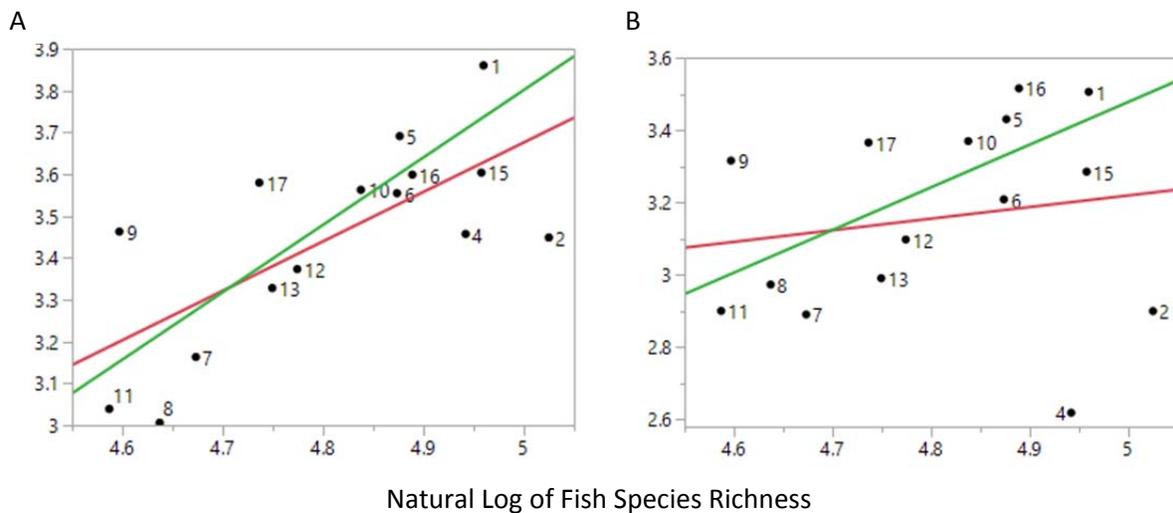

20

96
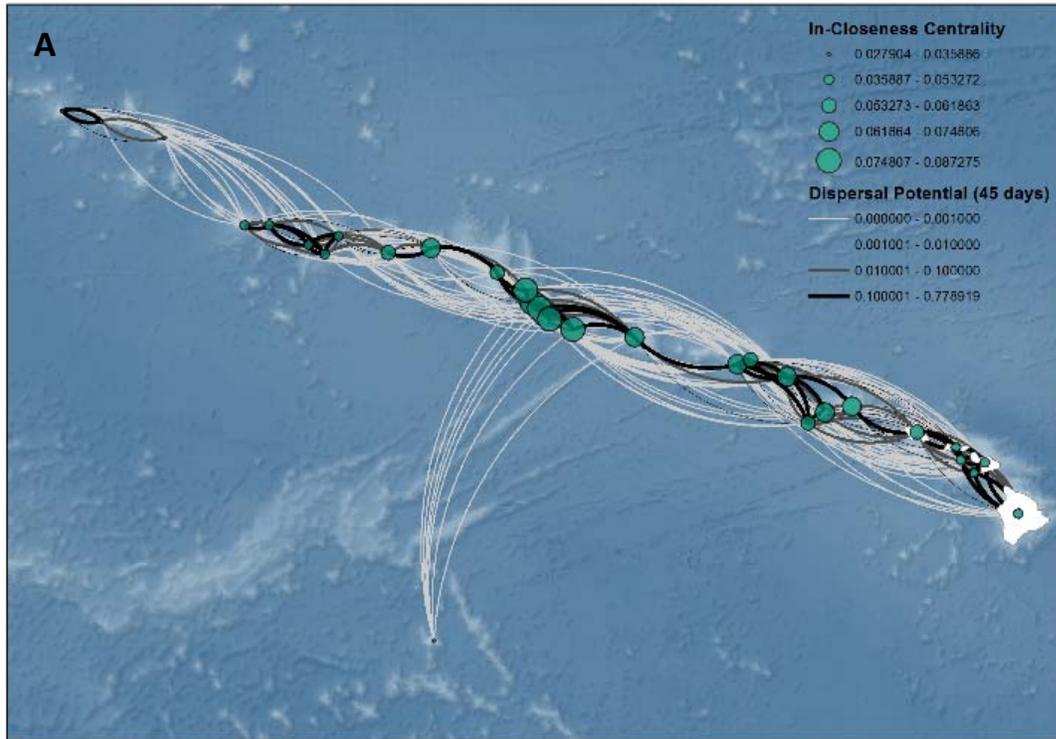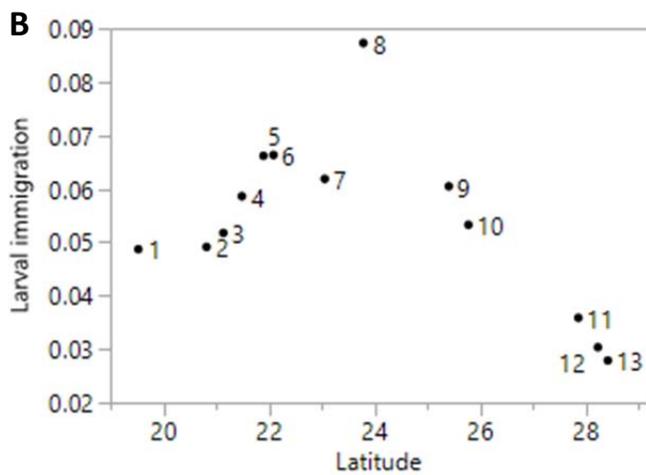
Fig. S7. The distribution of species' correlations of mtDNA AR to habitat area. Solid arrow indicates the composite AR correlation to habitat area inclusive of all species (r = 0.66). Shaded units indicate 8 species with highest individual correlations. Dashed arrow indicates the composite AR correlation to habitat area with these 8 species omitted from the mean (r = 0.61). Box plot horizontal lines indicate maxima and quartiles, diamond is centered on mean with tips at 95% confidence interval; red bracket shows range of 50% observation density; A: for all species (mean r = 0.05, N = 39), B: when 8 species are omitted (mean r = -0.10, N = 31). Identities of omitted species are: *C. exarata, C. lunulatus, C. miliaris, H. ornatissimus, L. kasmira, M. vanicolensis, O. erinaceus and P. pencillatus*; three are invertebrates and three are endemics.

107
108
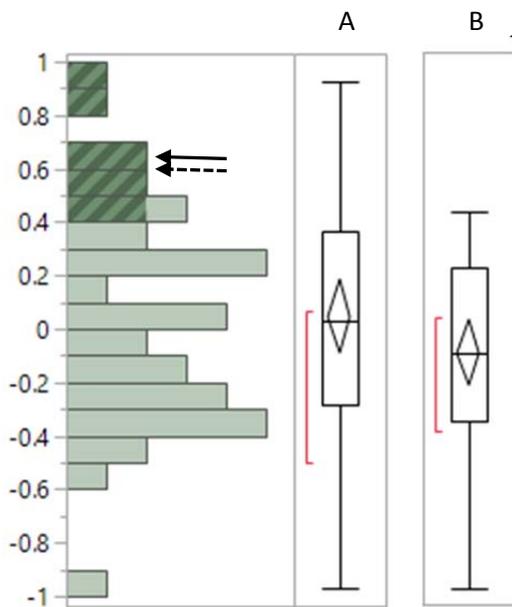
Fig S8. Comparison of correlations of mean allelic richness to fish species richness. A. Mean
mtDNA AR based on all species; B. Mean mtDNA AR on all 37 fishes. Red lines include all
islands; green lines exclude islands numbered 2 (Maui) and 4 (Molokai) which are outliers in
plot B.

Natural Log of Fish Species Richness