# Support Information of the Infor Prada and Hellberg 10.1073/pnas.1208931110

### SI Materials and Methods

Survivorship. Colonies were surveyed annually from 2003 to 2007 at eight reefs in Parguera (Puerto Rico), including Media Luna, the location used for the transplant experiment and genetic analysis. We monitored three depths per reef. At each depth, a replicated 20-m transect was placed and four  $1-m^2$  quadrats were randomly selected (total of  $8 \text{ m}^2$  per depth). In total, we analyzed 161 quadrats (we excluded 31 quadrats with no or ambiguous Eunicea flexuosa data). A full description of the site and the experimental design are described elsewhere (1).

Reciprocal Transplant. We selected 40 colonies from each of two depths (<3 m shallow and >20 m deep). Each colony was divided and the resulting fragments then transplanted to both shallow and deep areas. Survivorship was recorded yearly for 2 y for native and foreign colonies. We also estimated survivorship by pooling these current data with previous (2) estimates. Both transplant experiments were carried out in Parguera (Media Luna reef), Puerto Rico, for the same colonies used in the genetic analysis (see below). We estimated survivorship for one generation by multiplying the annual survivorship rate by 38.88 (the estimated time to reach 70 cm, assuming a 1.8-cm/y growth rate). We then estimated the selection advantage as the survivorship of one lineage over the total survivorship of both lineages. For example for the shallow area, the advantage of the Shallow lineage is as follows: Shallow/(Shallow + Deep). We then assume a 1:1 settlement and estimate Hardy–Weinberg proportions.

Population Genetic Sampling. Our sampling scheme spans regions where geographical patterns of genetic differentiation have been seen previously, including Exuma Sound and Mona Passage (3, 4). We collected adult ( $>50$  cm) colonies at least 5 m apart to decrease the sampling of clone mates and preserved them in 95% (vol/vol) ethanol at −20 °C. We extracted genomic DNA from these samples using the QIAGEN DNeasy Kit following the manufacturer's protocols.

Generation of Sequence Data. As a mitochondrial marker, we sequenced the MSH region, which encodes an ORF unique among animals to some anthozoans, using published primers (5). We developed three new nuclear sequence markers (Table S7) using sequences generated from a partial 454 run. Briefly, we extracted total RNA from a shallow individual from Puerto Rico using TRIzol (6). We then produced cDNA by reverse transcription of the RNA following (7), using a combination of Clontech SMART RT and Invitrogen SuperScriptII. The resulting PCR was verified in 1% agarose gel with fragments ranging between 500 and 4,000 bp. This PCR cDNA was fragmented by sonication and pyrosequenced at the University of California, Los Angeles, Genomic Center.

We generated a total of 27,197 reads with an average read length of 305 bp. Reads were assembled into 3,685 contigs and 12,457 singletons using CAP3 (8). Annotation with the Swiss-Prot database identified 3,344 homologous sequences. We selected 20 different genes to make primer pairs. First, we aligned the selected E. flexuosa sequence against the closet homolog in the Nematostella (anemone) genome to identify the possible placement of introns. We then designed intron-spanning primer pairs anchored in coding regions using Primer  $3(9)$ . From the initial 20 candidates, three were selected because they amplified across samples with a desirable sequence size of 350–700 bp. Initial screening showed these markers were single copy  $(\leq 2$  alleles per individual) and variable within populations.

PCR amplifications were performed in a Bio-Rad T100 with the same cycling conditions (except for the annealing temperature) for all genes: an initial denaturation cycle of 3 min at 95 °C, 2 min annealing at 50–56 °C, and 2 min extension at 72 °C; followed by 38 cycles of 30 s at 95 °C, 45 s at 50–56 °C, and 45 s at 72 °C, and a final extension cycle at 72 °C for 10 min. Amplicons were directly sequenced in both directions in an ABI 3100 using BigDye chemistry, version 3.1, and the amplification primers.

We resolved indel heterozygotes containing a single indel using CHAMPURU (10). We resolved haplotypes using PHASE, version 2.1 (11). We used 90% probability as our cutoff, because detailed cloning studies have shown that haplotypes inferred by PHASE with a probability >70% accurately reflect haplotypes (12). We cloned individuals using the Invitrogen TOPO TA Cloning Kit and sequencing at least five clones per reaction.

**Analysis of Genetic Data.** We aligned sequences with Clustal  $X(13)$ and inferred substitution models for each marker using jModelTest (14). Using the suggested model of evolution, we tested for intralocus recombination using both GARD and SBP as implemented in Hy-Phy (15, 16). We also inspected for recombination using the difference of sums of squares (DSS) method, with a sliding window of 100- and 10-bp step size as implemented in TOPALi, version 2 (17).

A parsimony haplotype network was constructed for each marker using the algorithm of Templeton et al. (18) implemented in TCS 1.21 (19). Each network was constructed with confidence level set at 95% and excluding gaps.

Hierarchical genetic subdivision was analyzed using the analysis of molecular variance (AMOVA) framework as implemented in GENODIVE (20). AMOVAs were performed with 10,000 permutations using conventional F statistics. We first defined populations by depth and location. To test for geographical differentiation within lineages, we then used clustering information from the STRUCTURE analysis (see below) to classify individuals by lineage (i.e., Shallow or Deep). We calculated both AMOVA and pairwise  $F_{ST}$  between locations within each lineage was calculated in GENODIVE (20).  $F_{ST}$  values were plotted against pairwise geographical distances among populations calculated in Google Earth 6.2 using the shortest nautical distance among populations.

Sequences were recoded as frequency data (codominant) in DNAsp 4.0 (21) and used to infer population subdivision using Bayesian clustering in STRUCTURE (22). We used the admixture model with a burnin of 100,000 steps followed by 20 million iterations and 10 replicates per run. We ran STRUCTURE without information of the origin of each individual, thereby reducing potential biases. We inferred population structure using a range of K values (number of inferred populations) from 1 to 8 (maximum number of populations). We then used STRUC-TURE HARVESTER (23), which implements the Evanno method (24) to estimate the most likely K based on the difference of likelihood scores. Because STRUCTURE uncovers only the deepest genetic subdivision (24), we hierarchically ran the program within each subsequent cluster.

To test the extent to which depth, geography, and their interaction have played roles in lineage splitting and to recover the evolutionary history of all populations, we used a combination of gene trees/species tree approaches and model-based inference approaches  $(25-27)$ . In STEM, we iterated the process five times, so that in each round gene genealogies were reestimated, and

species tree likelihood scores recalculated. We then used the average likelihood scores across replicates to infer information theory statistics and rank all possible topologies following Anderson (27) and Carstens and Dewey (28). In STEM, we used clustering information from the STRUCTURE analysis to classify individuals and included only individuals with assignment probabilities >90%; individuals with lower probability may be the result of interspecific gene flow, which may introduce bias.

We ran IMa first using broad priors (q1, q2, qA = 20, m1, m2 = 10,  $t = 10$ ) and generated more specific priors (q1, q2, qA = 3,  $m1 = 1$ ,  $m2 = 3$ ,  $t = 5$ ) that encompassed the entire distribution of parameters in subsequent runs. We ran IMa two times starting from different seeds, each run with 20 chains, a heating scheme of  $-g1$  0.8  $-g2$  0.9, and 2–4 million steps (2 million burnin). Once runs converged, we used the L mode (load trees) to estimate likelihood scores for all nested IM models. We repeated this process at least two times for a dataset including all samples divided by lineage as suggested by STRUCTURE, habitat and morphology (we used 2 mm spindle length as a cutoff). We also run IMa for each location for each grouping (lineage, habitat, and morphology). To convert IMa effective size estimates into demographic units, we used the size at first reproduction (30 cm). Assuming a conservative annual growth rate of 2 cm, the generation time for E. flexuosa is 15 y. Estimates based using other reported (2, 29, 30) growth rates fall within the reported 90% high posterior density range (Table S6).

We present the results of such analysis on Table S6 and Fig. S2. Migration rates should be interpreted with caution. We present data from Bahamas, Curaçao, and Panama to show that our results hold for most comparisons but recognize that our sampling design may be inappropriate to test for asymmetrical migration at these locations. In Bahamas and Curaçao, we suspect the deep habitat may be deeper, whereas the uniqueness of Panama and the blur in the distinctness of morphologies may require a larger sampling at this locale.

We suspect that only in Puerto Rico we sampled the opposite ends of the depth gradient, and thus the asymmetry in gene flow

- 1. Yoshioka P (2009) Sediment transport and the distribution of shallow-water gorgonians. Caribb J Sci 45(2):254–259.
- 2. Prada C, Schizas NV, Yoshioka PM (2008) Phenotypic plasticity or speciation? A case from a clonal marine organism. BMC Evol Biol 8:47.
- 3. Taylor MS, Hellberg ME (2006) Comparative phylogeography in a genus of coral reef fishes: Biogeographic and genetic concordance in the Caribbean. Mol Ecol 15(3):695–707.
- 4. Baums IB, Miller MW, Hellberg ME (2005) Regionally isolated populations of an imperiled Caribbean coral, Acropora palmata. Mol Ecol 14(5):1377–1390.
- 5. France SC, Hoover LL (2001) Analysis of variation in mitochondrial DNA sequences (ND3, ND4L, MSH) among Octocorallia (=Alcyonaria) (Cnidaria: Anthozoa). Bull Biol Soc Wash 10:110–118.
- 6. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162(1): 156–159.
- 7. Hale M, McCormick C, Jackson J, DeWoody J (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (Acipenser fulvescens): The relative merits of normalization and rarefaction in gene discovery. BMC Genomics 10:11.
- 8. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9(9):868–877.
- 9. Rozen S, Skaletsky J (2000) Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols: Methods in Molecular Biology, eds Krawetz S, Misener S (Humana, Totowa, NJ), pp 365–386.
- 10. Flot J-F (2007) Champuru 1.0: A computer software for unraveling mixtures of two DNA sequences of unequal lengths. Mol Ecol Notes 7(6):974–977.
- 11. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68(4):978–989.
- 12. Harrigan RJ, Mazza ME, Sorenson MD (2008) Computation vs. cloning: Evaluation of two methods for haplotype determination. Mol Ecol Resour 8(6):1239–1248.
- 13. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21): 2947–2948.
- 14. Posada D (2008) jModelTest: Phylogenetic model averaging. Mol Biol Evol 25(7): 1253–1256.
- 15. Pond SL, Frost SD (2005) Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21(10):2531–2533.

can be more rigorously tested. Puerto Rico is the most depthsegregated location, with almost all (except for two Shallow lineage individuals in deep), and thus estimates derived from it eliminate any sampling artifact derived from the clinal distribution of both lineages. We then used model-based selection to calculate evidence ratios and rank all possible models (27). Because we reject a strict isolation model, we used a full model to estimate demographic parameters.

Morphological Measurements. We used 5% bleach to clear the organic surface from colonies, washed the colonies twice, and dried them in ethanol. We took digital images of each colony through a Diagnostic Instrument SPOT RT Slider CCD camera attached to a Leica MZ7 stereomicroscope.

To compare populations and understand the relative importance of different factors, we fit spindle length to a generalized linear model. We used a Gaussian distribution and geography, depth, and lineage as factors. We constructed all 17 possible models from these factors and used model-based approaches based on Akaike information criterion scores (similar to the STEM analysis) to rank each model (27). We also used an honestly significant difference (HSD) Tukey test to identify which comparisons among locations within lineages were significant. Analysis was performed with raw measurements and log-transformed data with similar results.

SI ACKNOWLEDGMENTS. We thank D. Beltrán, D. Ruiz, M. Vermeij, H. Lasker, and R. Collin for help during collections. We thank Caribbean Research and Management of Biodiversity (CARMABI) (Curaçao), Smithsonian Tropical Research Institute at Bocas del Toro, Panama, and the Department of Marine Sciences of the University of Puerto Rico for allowing the use of their facilities during sampling. We thank N. Schizas and R. Appeldorn, who provided laboratory space and materials during field collections in Puerto Rico. We thank Matt Brown at the Louisiana State University Socolofsky Microscopy Center and D. Beltrán for help with morphological measurements. We thank the governments of Puerto Rico, Panama, Curaçao, and Bahamas for granting collecting permits. Discussions with participants of the 2010 Summer School in Speciation Genetics (Eawag, Kastanienbaum, Switzerland) improved this study.

- 16. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23(10):1891–1901.
- 17. Milne I, et al. (2004) TOPALi: Software for automatic identification of recombinant sequences within DNA multiple alignments. Bioinformatics 20(11):1806–1807.
- 18. Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132(2):619–633.
- 19. Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. Mol Ecol 9(10):1657–1659.
- 20. Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4(4): 792–794.
- 21. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19(18):2496–2497.
- 22. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959.
- 23. Earl D, vonHoldt E (2011) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Cons Gen Res 4(2):359–361.
- 24. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol 14(8):2611–2620.
- 25. Pigliucci M, Kaplan J (2006) Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology (Chicago Univ Press, Chicago).
- 26. Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19(2):101–108.
- 27. Anderson DR (2008) Model Based Inference in the Life Sciences (Springer, New York).
- 28. Carstens BC, Dewey TA (2010) Species delimitation using a combined coalescent and information-theoretic approach: An example from North American Myotis bats. Syst Biol 59(4):400–414.
- 29. Beiring EA, Lasker HR (2000) Egg production by colonies of a gorgonian coral. Mar Ecol Prog Ser 196:169–177.
- 30. Yoshioka PM, Yoshioka BB (1991) A comparison of the survivorship and growth of shallow-water gorgonian species of Puerto Rico. Mar Ecol Prog Ser 69:253–260.



Fig. S1. Parsimony haplotype networks for each locus. The size of the circles is proportional to haplotypes frequency. Nuclear markers have twice as many haplotypes.



Fig. S2. Migration rate estimates between Shallow and Deep lineages per location obtained by fitting the IMa model to all four loci. Samples were divided by habitat (Top), lineage (Middle), and morphology (Bottom). The solid lines represent migration from Shallow to Deep, and the dotted lines from Deep to Shallow. Note that migration estimates by lineage are lower than for habitat or morphology, but that the probabilities of the estimates are higher.

Table S1. Annual survivorship per size class for E. flexuosa across reefs and depths over 4 y



Table S2. Annual survivorship for reciprocally transplanted colonies for two independent experiments



\*Extracted from ref. 1.

PNAS PNAS

1. Prada C, Schizas NV, Yoshioka PM (2008) Phenotypic plasticity or speciation? A case from a clonal marine organism. BMC Evol Biol 8:47.



#### Table S3. Localities of the collections and number of colonies sampled per depth (N)

# Table S4. Within-lineage pairwise  $F_{st}$  values



Shallow and Deep lineage comparisons are below and above the diagonal, respectively. Significant values after Bonferroni correction are in bold.





The first three letters of the models represent the populations sizes for the three lineages (Deep, Shallow, and ancestral). The next two letters (or 0) indicate the migration from Deep to Shallow and from Shallow to Deep. Lineages with the same letter indicate same population size and zero indicates no migration. AIC, Akaike information criterion inferred from the generalized linear model fitting; Δi, difference in AIC score with respect to the best model; model likelihoods, relative likelihood of the model given the data; wi, model probabilities; evidence ratio, difference in model probabilities between the proposed model and the best model (ABC0D).

SVNG PNS





The geometric mean of the mutation rate per year was 4.96 x 10<sup>-7</sup>. Generation time was 17 y. The posterior distribution for the time parameter was flat. The 90% HPD was estimated as the widest range across the three runs. Estimates from Bahamas, Curaçao, and Panama should be interpreted with caution as the shallow and deep habitats may represent intermediate areas and not the strict opposite sides of the depth gradient.  $\theta_1 = 4N1u$ , the population mutation rate Shallow;  $\theta_2 = 4N2u$ , the population mutation rate for Deep; m<sub>1</sub>/μ, the migration rate, per mutation, from Deep to Shallow; m<sub>2</sub>/μ, the migration rate, per mutation, from Shallow to Deep; t/μ, divergence time scaled by mutation rate; estimates should be interpreted carefully as the distribution was flat; N<sub>1</sub>, effective size of Shallow; N<sub>2</sub>, effective size of Deep; N<sub>A</sub>, ancestral effective population size; N<sub>1</sub>m<sub>1</sub>, migrants per generation from Deep to Shallow; N<sub>2</sub>m<sub>2</sub>, migrants per generation from Shallow to Deep.

\*Estimates inferred from Puerto Rico only.

PNAG PNAS

## Table S7. Genetic markers used in this study with closest blastx match



\*Primers are from France and Hoover (1).

PNAS PNAS

1. France SC, Hoover LL (2001) Analysis of variation in mitochondrial DNA sequences (ND3, ND4L, MSH) among Octocorallia (=Alcyonaria) (Cnidaria: Anthozoa). Bull Biol Soc Wash 10:110–118.