Materials and methods

*Sanger sequencing*

DNA was isolated using the modified “HotSHOT” protocol (Meeker *et al.* 2007) and subsequently stored at -20°C. Approximately 800 bp of mitochondrial cytochrome *b* (Cyt*b*) were amplified using the primers AOLF1 (5’ RGG-GGA-GAA-GAG-RGC-TAG-GGA-GGA-3’) and AOLR2 (5’TGC-CGA-GAC-GTY-AAC-TAC-GGC-TG-3’). We also amplified the first intron of the S7 ribosomal protein (S7) gene using S7RPEX1F and S7RPEX2R (Chow & Hazama 1998) and intron 2 of the Glyceraldehyde-3-phosphate dehydrogenase (Gpd2) gene using Gpd2F and Gpd2R (Hassan *et al.* 2002). Polymerase chain reactions (PCRs) for all three markers were carried out in a 10µl volume containing 2-15ng of template DNA, 0.2-0.3µM of each primer, 5µl of the premixed PCR solution BioMix Red™ (Bioline), and deionized water to volume. PCR reactions utilized the following cycling parameters: initial denaturation at 95°C and final extension at 72°C (10 min each), with an intervening 35 cycles of 30s at 94°C, 30s at the annealing temperature (56°C, Cyt*b*; 58°C, S7; 54°C, Gpd2), and 45s at 72°C. Amplification products were purified using 0.75 units of Exonuclease I with 0.5 units of Shrimp Alkaline Phosphatase (ExoSAP; USB) per 7.5µl PCR products at 37°C for 60 min, followed by deactivation at 85°C for 15 min. DNA sequencing was performed with fluorescently-labeled dideoxy terminators on an ABI 3730XL Genetic Analyzer (Applied Biosystems) at the University of Hawai‘i Advanced Studies of Genomics, Proteomics and Bioinformatics sequencing facility.

Sequences for each locus were aligned, edited, and trimmed to a common length using the DNA sequence assembly and analysis software GENEIOUS PRO 5.0 (Biomatters). After
trimming, the alignment was unambiguous with no indels or frame-shift mutations. A detailed analysis of genetic structure across the Hawaiian Archipelago and Johnston Atoll (JO) revealed no genetic partitioning between locations (unpublished data). For analyses presented here Sanger sequences from Hawai‘i were grouped with samples from the nearby Johnston Atoll. Allelic states of nuclear sequences with more than one heterozygous site were estimated using the Bayesian program PHASE 2.1 (Stephens & Donnelly 2003) as implemented in the software DnaSP 5.0 (Librado & Rozas 2009). We conducted three runs in PHASE for each data set. Each run had a unique random seed number. Two runs were conducted for 10000 iterations with 1000 burn-in iterations. To ensure proper allele assignment, a third run of 100000 iterations was conducted. All runs returned consistent allele identities.

**RADSeq library preparation and sequencing**

For RADSeq, DNA was extracted from a subset of the original samples using DNeasy Tissue kits (Qiagen) following the manufacturer’s protocol for animal tissues, and stored at 20ºC. We prepared RADSeq libraries for up to 18 individuals from each of six populations (A. olivaceus: Hawai‘i, Kiritimati, Palau, Marshall Islands, Bali; A. reversus: Marquesas) using the double digest protocol of Peterson et al. (2012). We proceeded with library preparation using only samples with concentrations higher than 10 ng/µl, and followed the Peterson et al. (2012) protocol closely with few modifications, including a 3 hour digest at 37ºC using the SphI and MluCl restriction endonucleases [New England Biolabs; enzyme pair was chosen based on Table 1 of Peterson et al. (2012)]. Following cleaning using Dynabeads M-270 Streptavidin (Life Technologies), universal P2 adaptors were ligated to each DNA fragment. Additionally, a uniquely barcoded P1 adapter was ligated to each of sixteen individuals. These individuals were later pooled and a unique Illumina index was added to identify that pool of sixteen. Prior to
pooling, DNA was cleaned using magnetic beads (as above). Pooled DNA was size-selected using a Pippin Prep (Sage Science) with all fragments between 376-450 bp recovered. The unique Illumina indices were incorporated onto the P2 adaptor end of DNA fragments using a real-time library amplification kit (Kapa Biosystems). The concentration of each pool is standardized and combined for Illumina sequencing. Libraries were quantified using a High Sensitivity DNA Kit on a 2100 Bioanalyzer (Agilent Technologies) and sequenced at the UCLA Neuroscience Genomics Core facility on an Illumina HiSeq 2000 (100 bp single end reads). Sequencing runs resulted in 135 million reads passing initial quality control at the sequencing facility.

References


