Molecular Phylogenetics and Evolution

Surgeons and suture zones: hybridization among four surgeonfish species in the Indo-Pacific with variable evolutionary outcomes

Joseph D. DiBattista[1,2,*], Jonathan Whitney[3], Matthew T. Craig[4], Jean-Paul A. Hobbs[1], Luiz A. Rocha[5], Kevin A. Feldheim[6], Michael L. Berumen[2], Brian W. Bowen[3]

[1]Department of Environment and Agriculture, Curtin University, PO Box U1987, Perth, WA 6845, Australia

[2]Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

[3]Hawai'i Institute of Marine Biology, Kāne'ohe, HI 96744, USA

[4]Fisheries Resource Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA

[5]Section of Ichthyology, California Academy of Sciences, San Francisco, CA 94118, USA

Kevin A. Feldheim

[6]Field Museum, Pritzker Laboratory for Molecular Systematics and Evolution, Chicago, IL 60605, USA ,

*Corresponding author

E-mail: josephdibattista@gmail.com

Abstract

Closely related species can provide valuable insights into evolutionary processes through comparison of their ecology, geographic distribution and the history recorded in their genomes. In the Indo-Pacific, many reef fishes are divided into sister species that come into secondary contact at biogeographic borders, most prominently where Indian Ocean and Pacific Ocean faunas meet. It is unclear whether hybridization in this contact zone represents incomplete speciation, secondary contact, an evolutionary dead-end (for hybrids) or some combination of the above. To address these issues, we conducted comprehensive surveys of two widely-distributed surgeonfish species, *Acanthurus leucosternon* ($N = 141$) and *A. nigricans* ($N = 412$), with mtDNA cytochrome *b* sequences and ten microsatellite loci. These surgeonfishes are found primarily in the Indian and Pacific Oceans, respectively, but overlap at the Christmas and Cocos-Keeling Islands hybrid zone in the eastern Indian Ocean. We also sampled the two other Pacific members of this species complex, *A. achilles* ($N = 54$) and *A. japonicus* ($N = 49$), which are known to hybridize with *A. nigricans* where their ranges overlap. Our results indicate separation between the four species that range from the recent Pleistocene to late Pliocene (235,000 to 2.25 million years ago). The Pacific *A. achilles* is the most divergent (and possibly ancestral) species with mtDNA $d_{corr} \approx 0.04$, whereas the other two Pacific species (*A. japonicus* and *A. nigricans*) are distinguishable only at a population or subspecies level ($\Phi_{ST} = 0.6533$, $P < 0.001$). Little population structure was observed within species, with evidence of recent population expansion across all four geographic ranges. We detected sharing of mtDNA haplotypes between species and extensive hybridization based on microsatellites, consistent with later generation hybrids but also the effects of allele homoplasy. Despite extensive introgression, 98% of specimens had concordance between mtDNA lineage and species identification based on external morphology, indicating that species integrity may not be eroding. The *A. nigricans* complex demonstrates a range of outcomes from incomplete speciation to secondary contact to decreasing hybridization with increasing evolutionary depth.

Keywords: Acanthuridae; coral reef fish; introgression; reverse speciation; species complex

1. Introduction

Much of the debate on the formation of species has focused on allopatry (Dobzansky, 1937; Mayr, 1942) versus sympatry (Bolnick and Fitzpatrick, 2007; Bird et al., 2012). The allopatric model has been favored for decades and stresses geographic isolation with negligible gene flow. However, the importance of allopatric speciation in the marine environment is uncertain given the paucity of physical barriers and the high dispersal ability of most marine organisms (Bowen et al., 2013). Closely related marine species often have overlapping distributions, which can obscure the role of geographic isolation in evolutionary partitioning (Bellwood and Wainwright, 2002; Rocha and Bowen, 2008). Ecological factors, on the other hand, may be a driving force behind species formation under sympatric or parapatric scenarios (Choat, 2006; Bird et al., 2011). Indeed, there are several recent examples of ecological speciation (Nosil, 2012) or speciation with gene flow (Feder et al., 2012; Abbott et al., 2013).

The prevalence of closely-related cohabiting taxa on coral reefs provides extensive opportunity for hybridization. Secondary contact among recently diverged taxa, along with a host of other factors (e.g. external fertilization, weak behavioral isolation, niche overlap and unequal abundance of parental species) facilitate potential interspecific mating (Gardner, 1997; Volmer and Palumbi, 2002; Hobbs et al., 2013). There are several possible outcomes when two species interbreed. Low levels of mixing can lead to an influx of novel and potentially beneficial genes, which may allow hybrids to exploit new niches (i.e. adaptive introgression; Rieseberg, 2011; Pardo-Diaz et al., 2012). These hybrids can form new species that may become adaptive radiations if they assortatively mate or occupy different niches relative to parental species (Seehausen, 2004). In contrast, high levels of mixing between species may swamp the gene pool, removing adaptive gene variants (Rhymer and Simberloff, 1996), and may also lead to two species blending into one (reverse speciation; Seehausen, 2006; Coleman et al., 2014). Hybrids may have greater fitness than their parents (hybrid vigor), or selection against hybrids can reduce the fitness of the F1 generation but more so F2 generations (Barton, 2001). Indeed, a range of behavioral, genetic and physiological traits may evolve and prevent hybrids from breeding with the parent species, thus limiting any further introgression via a process known as reinforcement (Kirkpatrick, 2001). Hybridization can therefore enhance, inhibit or have no consequence to the adaptive evolution of parental species**.**

Coral reef fishes represent the most diverse assemblage of vertebrates (Jones et al., 2002), and yet hybridization was traditionally considered unimportant to the evolution of this group (Hubbs, 1955). It is now evident that hybridization is prevalent in coral reef fishes and make up the majority of the 173 marine fishes reported to hybridize (Montanari et al., 2016). Hybridization is also geographically and taxonomically widespread in this group (e.g. Hobbs and Allen, 2014; DiBattista et al., 2015) and includes surgeonfishes (Randall, 2002; Marie et al., 2007), butterflyfishes (McMillan et al., 1999; Hobbs et al., 2013), angelfishes (Pyle and Randall, 1994), wrasses (Yaakub et al., 2006, 2007), damselfishes (van Herwerden and Doherty, 2006) and groupers (van Herwerden et al., 2006). Although morphological characterization of hybrids is still common, molecular techniques can reveal hybrids that may otherwise go undetected (Kuriiwa et al., 2007; Montanari et al., 2012, 2014) and the geographic extent of introgression (McMillan et al., 1999). Such advances are especially relevant for identifying hybrids between closely related species, which may be challenging to classify based solely on morphological characters. Bayesian analytical approaches have improved our ability to not only genetically detect hybrids in nature, but also discriminate between F1 and later generation hybrids (Schwarz and Beheregaray, 2008).

Suture zones are areas of secondary contact where multiple pairs of sister species interbreed (Remington, 1968). A suture zone for reef fishes occurs in the eastern Indian Ocean (Christmas Island and Cocos-Keeling Islands), where Indian and Pacific Ocean faunas come into contact. Hobbs and Allen (2014) documented 15 pairs of hybridizing species from eight families in this suture zone. Climatic cycles over the past 2 million years have allowed allopatric divergence of Indian and Pacific lineages, which have since come back into contact in this area (Briggs and Bowen, 2012). Another prominent suture zone for reef fishes was recently discovered in the Socotra Archipelago, at the intersection of four biogeographic provinces in the north-western Indian Ocean (DiBattista et al., 2015). More suture zones are suspected in southern Japan and the Marshall Islands (Hobbs et al., 2013). The presence of multiple suture zones across the Indo-Pacific provides the opportunity to study hybridization and evolutionary processes (e.g. reproductive isolation and reinforcement) operating in these species complexes.

In this study we examine hybridization in four species of the *Acanthurus nigricans* complex (also known as the *Acanthurus achilles* complex; Randall and Frische, 2000): the powder blue surgeonfish (*Acanthurus leucosternon* Bennett 1833) in the Indian Ocean; the goldrim surgeonfish (*Acanthurus nigricans* Linnaeus, 1758) distributed throughout the Pacific and eastern Indian Oceans; the Achilles surgeonfish (*Acanthurus achilles* Shaw 1803) restricted to the Hawaiian Islands and other oceanic islands of Oceania and the North Pacific Ocean; and the white-nose surgeonfish (*Acanthurus japonicus* Schmidt, 1831) restricted to Japan, Taiwan, the Philippines and northern Indonesia (Fig 1). All species are ecologically and morphologically similar (Randall, 1956; Randall, 2002; Marie et al., 2007), with the exception of distinct color patterns (Fig 2; also see Robertson et al., 1979; Kuiter and Debelius, 2001). *Acanthurus nigricans* and *A. leucosternon* hybridize at the eastern Indian Ocean suture zone (Marie et al., 2007). The presence of individuals with intermediate coloration in the Marshall Islands indicates hybridization between *A. nigricans* and *A. achilles* (Randall, 1956), and intermediates between *A. nigricans* and *A. japonicus* have been documented in southern Japan and Taiwan (Randall and Frisch, 2000).

This study expands on the initial genetic documentation of hybridization between *A. leucosternon* and *A. nigricans* in the eastern Indian Ocean (Marie et al., 2007) with near range-wide sampling using mitochondrial and microsatellite markers, along with Bayesian approaches, to infer the extent and direction of introgression. Using the same approach, we examine hybridization involving *A. nigricans* and the other two species in the complex (*A. achilles* and *A. japonicus*). We hypothesize that each species will show genetic evidence of hybridization in areas of range overlap, but that they will also maintain unique lineages where the parental species do not co-occur.

2. Materials and Methods

2.1 Sample collection

A total of 54, 49, 141 and 412 tissue samples of *A. achilles*, *A. japonicus*, *A. leucosternon* and *A. nigricans,* respectively, were collected at 21 locations across the Indo-Pacific (Fig 1). Although *A. nigricans* collections include multiple locations across the Main Hawaiian Islands (separated by ~500 km), these were combined for analysis owing to the majority of sampling occurring at a single site (Kauai, *N* = 18; Oahu, *N* = 3; Big Island, *N* = 2) with no detectable genetic differentiation among sites (mtDNA: Φ_{ST} = -0.15, *P* = 0.71; microsatellites: F_{ST} = -0.0041, *P* = 0.57). Putative hybrids between species, identified by intermediate coloration (Fig. 2), were also collected in areas of overlap for *A. achilles* and *A. nigricans* (Kosrae, Caroline Islands, *N* = 6; Marshall Islands, *N* = 1), *A. japonicus* and *A. nigricans* (Japan, *N* = 1; Guam, Mariana Islands, *N* = 1) and *A. leucosternon* and *A. nigricans* (Christmas Island, *N* = 25; Cocos-Keeling Islands, *N* = 14). Tissue was preserved in a saturated salt-DMSO solution, total genomic DNA was extracted using a "HotSHOT" protocol (Meeker et al., 2007) and samples were subsequently stored at -20 ˚C.

2.2 Mitochondrial DNA analysis

A 491 base pair (bp) segment of the mitochondrial DNA (mtDNA) cytochrome *b* (cyt *b*) gene was resolved using heavy-strand (5' - GTGACTTGAAAAACCACCGTTG - 3', Song et al., 1998) and light-strand primers (5' - AATAGGAAGTATCATTCGGGTTTGATG - 3', Taberlet et al., 1992). Polymerase chain reaction (PCR) amplification was carried out in 15 µl volumes containing BioMix Red (Bioline Ltd., London, UK), 0.26 µM of each primer and 5 to 50 ng template DNA. PCRs used an initial denaturing step at 95 °C for 3 minutes, then 35 cycles of amplification (30 seconds of denaturing at 94 °C, 45 seconds of annealing at 63 °C and 45 seconds of extension at 72 °C), followed by a final extension at 72 °C for 10 min.

PCR products were purified by incubating with exonuclease I and FastAP™ thermosensitive alkaline phosphatase (ExoFAP; Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C for 60 min, followed by 85 °C for 15 min. All samples were sequenced with fluorescently labeled dye terminators (BigDye 3.1, Applied Biosystems Inc., Foster City, CA, USA) and analyzed using an ABI 3130XL Genetic Analyzer (Applied Biosystems). The sequences were aligned, edited and trimmed to a common length using Geneious Pro 4.8.4 (Drummond et al., 2009); cyt *b* sequences were deposited in GenBank (accession numbers: KM456297to KM456989). jModelTest 1.0.1 (Posada, 2008, but also see Guindon and Gascuel, 2003) was used with an Akaike information criterion (*AIC*) test to determine the best nucleotide substitution model in each species; the HKY (Hasegawa et al., 1985), TPM1uf (Kimura, 1981), TPM3 + alpha parameter (gamma distribution) of 0.015 (Posada, 2008) and TIM1 + alpha parameter (gamma distribution) of 0.635 models were selected for *A. achilles*, *A. japonicus*, *A. leucosternon* and *A. nigricans,* respectively.

ARLEQUIN 3.5 (Excoffier et al., 2005) was used to calculate haplotype (*h*) and nucleotide diversity (π) for each species and collection site (if $N \geq 6$). Deviations from neutral sequence evolution were assessed with Fu's *Fs* (Fu, 1997) using ARLEQUIN; significance was tested with 99,999 permutations. Each site (and species) was also fitted with the population parameter τ in order to estimate the time since the most recent population expansion ($\tau = 2\mu t$; Rogers and Harpending, 1992), where t is the age of the population in generations and μ is the mutation rate per generation for the sequence (μ = number of bp ⋅ divergence rate within a lineage ⋅ generation time in years). We used a range of approximate cyt *b* mutation rates available from previous fish studies (1% per million years [MY] to 1.55% per MY within lineages; Bowen et al., 2001; Lessios, 2008; Reece et al., 2010) and a generation (replacement) time of 12 years for all species based on existing life-history information (i.e. age-at-maturity ~ two years, longevity ~ 27 to 36 years; J.H. Choat pers. comm.). Although absolute values generated by these calculations are approximations, relative comparisons among species and sites are more robust. The time of divergence between species was estimated in ARLEQUIN as the pairwise sequence distance between species minus the pairwise sequence distance within species, the corrected sequence divergence (d_{corr}).

We tested for range-wide patterns of population structure within *A. leucosternon* and *A. nigricans* (excluding morphological hybrids in this case) but not *A. achilles* and *A. japonicus*, given our small sample size for the latter two species. Genetic differentiation among sampling sites was first estimated with an analysis of molecular variance (AMOVA, Excoffier et al., 1992) in ARLEQUIN; only sites with $N \geq 6$ were considered. Deviations from null distributions were tested with non-parametric permutation procedures (N = 99,999) and P-values for pairwise Φ_{ST} were adjusted according to the modified false discovery rate (as per Narum, 2006).

Evolutionary relationships within and among all four species were assessed with an unrooted network constructed with NETWORK 4.5.1.0 (www.fluxus-engineering.com/network_terms.htm) using a median joining algorithm and default settings (as per Bandelt et al., 1999). Each haplotype was divided into sample site (and species) contributions as reflected by the pie diagrams. Putative hybrids were included to identify the maternal contribution of each hybrid individual and the degree of haplotype mixing between species. The network was simplified by removing all haplotypes occurring in single specimens (N = 103), except for the hybrids; this reduced multiple connections but did not influence the overall pattern.

2.3 Microsatellite genetic analysis

Each species was genotyped at 10 microsatellite loci (GenBank Accession numbers: HQ130123 to HQ130132) using PCR reaction mixes and cycling parameters described in DiBattista et al. (2011). PCR products labeled with different dye colors were pooled for genotyping and resolved using an ABI 3130XL Genetic Analyzer (Applied Biosystems) along with a fluorescently labeled internal size standard (LIZ-500; Applied Biosystem); allele sizes were assigned manually with GENEMAPPER 3.7 (Applied Biosystems). For each locus, the mean number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities, departure from Hardy-Weinberg proportions (HWE) and linkage disequilibrium (LD) were assessed with ARLEQUIN. The possible presence of null alleles was tested with MICRO-CHECKER 2.2.3 (van Oosterhout et al., 2004). Genotyping error rate was also estimated by independently re-genotyping randomly selected specimens (12% of all specimens per locus, for each species; see Hoffman and Amos, 2005); PCR amplification or scoring errors were negligible in this case (< 0.86%, < 0.64%, < 1.05% and < 1.79% for *A. achilles*, *A. japonicus*, *A. leucosternon* and *A. nigricans*, respectively, at each individual locus). Most specimens amplified for at least 9 (109 out of 654 samples) or 10 (509 out of 654 samples) loci, and only two specimens (both *A. nigricans* from Palmyra Atoll) were excluded from further analysis because they amplified at fewer than five microsatellite loci. Also note that only *A. japonicus* from Japan (N = 6), an *A. japonicus* x *A. nigricans* from Guam and an *A. japonicus* x *A. nigricans* from Japan were genotyped given that the remainder (N = 43) were opportunistically collected at a later stage in the study.

Nuclear population structure was assessed for *A. leucosternon* and *A. nigricans* only (excluding morphological hybrids) as described above: 1) AMOVA and 2) pairwise population F_{ST} estimation. To define genetic clusters within each species without *a priori* information on the geographical origin of specimens, we used STRUCTURE 2.3.3 (Pritchard et al., 2000). The most likely number of clusters in the dataset was identified based on the probability of $K = 1$ to $K = 4$ or $K = 1$ to $K = 15$ for *A. leucosternon* and *A. nigricans*, respectively, averaged over five replicate runs with 1,000,000 Markov Chain Monte Carlo (MCMC) repetitions and a burn-in of 100,000 iterations. Correlated allele frequencies and admixed populations were assumed. The most likely value of K was determined by plotting the mean Ln probability of the data (Ln P[D]) over multiple runs versus K for each species.

2.4 Genetic characterization of hybrids

To determine the power of our microsatellite markers to identify morphologically cryptic hybrid individuals in our sample, we used the program HYBRIDLAB 1.0 (Nielsen et al., 2006), which simulates multi-locus genotypes for offspring of purebred populations. We selected samples collected from areas where morphological hybrids were not detected, farthest from the hybrid zone and with the largest sample size to represent the microsatellite allele frequency variation for parental species; Seychelles for *A. leucosternon* ($N = 31$) and Kiritimati for *A. nigricans* ($N = 36$). For each dataset ($N = 5$ replicates), we simulated genotypes for 450 parental *A. leucosternon* and 450 parental *A. nigricans*, which were then used to simulate F1 hybrids, F2 hybrids (F1 x F1) and B2 backcrosses to each species ($N = 25$ for each hybrid class)(see Fig. S2)

The five simulated datasets were analyzed with two different Bayesian methods to establish a threshold (q) for assigning parental or hybrid status, and a range of q-values to discriminate among each hybrid class. Bayesian methods are favored because they do not require diagnostic alleles to assign individuals to the genotype classes or that “pure” members of each species are sampled. First, in STRUCTURE, we ran the admixture model for 1,000,000 iterations and a burn-in of 500,000, with independent allele frequencies between the two species, all fish classified as "unknown" and $K = 2$. These parameters resulted in a q-value for each individual ranging from 0 to 1, representing the proportion of a fish's genotype that had *A. nigricans* or *A. leucosternon* ancestry. We considered a range of q-values averaged over multiple runs to define each hybrid class and a threshold for the parental categories. All individuals, from all five independent datasets, were used to calculate a conservative 99% confidence interval for each category. We redid these analyses using the correlated allele frequencies option and the results were not different; the former approach with independent allele frequencies is therefore presented here.

To calculate the probability of an individual belonging to each parental category and hybrid class, we implemented a model with uniform priors for allele frequency and admixed distributions in NEWHYBRIDS 1.0 (Anderson and Thompson 2002). We ran each of the five simulated datasets for 1,000,000 iterations and a burn-in of 500,000, five separate times. If the assignment to a hybrid class was of high probability (i.e. > 90%) we accepted the assignment. If the probability was divided among the hybrid classes, we summed the probability over all classes, and if the sum was ≥ 0.9 we considered this a hybrid with an undefined class.

To determine the proportion of previously unknown hybrids in our collections we analyzed the 478 genotyped specimens in STRUCTURE and NEWHYBRIDS using the same parameters as outlined for the simulations. A lack of within-species population structuring (see Results) mitigated bias associated with considering the entire dataset versus a subset of the data. Putative hybrids were included in this dataset as a form of quality control for hybrid assignment. A final species (or hybrid) class was assigned to each individual based on a combination of the two methods and thresholds defined by our simulations. The level of error within each program, based on the standard deviation (SD) among replicate runs, was negligible (STRUCTURE SD = 0.0003 to 0.0030; NEWHYBRIDS SD = 0 to 0.0116), although this represents only one source of potential error.

Given that the above analyses are based on the assumption that we sampled pure populations of each parental species, we reran analyses in STRUCTURE without this assumption using a broader dataset, including specimens of all four surgeonfish species and their putative hybrids as outlined above. In brief, we tested $K = 1$ to $K = 8$ for the combined dataset averaged over five replicate runs with 1,000,000 Markov Chain Monte Carlo (MCMC) repetitions and a burn-in of 100,000 iterations, which allowed us to assess fine-scale genetic structure both *within* and *between* putative species. Correlated allele frequencies and admixed populations were assumed. The most likely value of K was determined by plotting the mean Ln probability of the data (Ln P[D]) over multiple runs versus K for each species (plot not shown). Average assignment of individuals across runs to each cluster was calculated using CLUMPAK 1.1 (Kopelman et al., 2015). We additionally reran analyses in STRUCTURE using only *A. leucosternon*, *A. nigricans* and their hybrids; the result was the same (data not shown) and so only the former results are presented here. Moreover, to ensure that our results were not driven by the small sample size of *A. japonicus* (i.e. $N = 6$), we repeated these same analyses by randomly selecting six individuals per species.

A discriminant analysis of principal components (DAPC; Jombart et al., 2010) was also run on all loci to investigate the relationship between genotype and geographical location using the "*find_clusters*" option with the 'adegenet' package (Jombart, 2008) in R (R Development Core Team, 2014). For each species, the number of principal components that explained 80% of the variation in the data were retained to estimate the most likely number of genetic clusters (*K*). To identify the most likely *K*, the program was run sequentially with increasing values of *K*. The Bayesian Information Criterion (BIC) was used to compare the different clustering solutions, with the optimal clustering solution corresponding to the lowest BIC value. The number of principal components retained for genotypic variability was equal to the number of individuals divided by three; the number of DA eigenvectors corresponded to the number of populations minus one.

3. Results

3.1 Molecular characteristics

Cytochrome *b* sequence data from *A. achilles, A. japonicus, A. leucosternon* and *A. nigricans* revealed 23, 17, 43 and 84 haplotypes, respectively (range: 2 to 25 per site), with haplotype diversity ranging from $h = 0.34$ to 0.94 and nucleotide diversity ranging from $\pi = 0.0009$ and 0.0066 across all sites and species (Table 1). The most common haplotype for each species was detected at almost every sampling site and shared by 9, 19, 39 and 214 individuals, respectively.

Table 1. Sample size and molecular diversity indices for *Acanthurus achilles*, *A. japonicus*, *A. leucosternon* and *A. nigricans* based on mitochondrial DNA (cytochrome *b*) sequence data. Time since the most recent population expansion was calculated using a range of mutation rates (1% to 1.55% per million years within lineages; Bowen et al., 2001, Lessios, 2008, Reece et al., 2010) and a generation time of 12 years for all species (see Materials and Methods).

Collection locality	N	H_N[b]	Time since expansion (Yrs)	Haplotype diversity ($h \pm SD$)	Nucleotide diversity ($\pi \pm SD$)	Fu's F_S
Acanthurus achilles						
Marshall Islands (MAR)	10	8	231,259-358,452	0.93 ± 0.08	0.0066 ± 0.0042	**-3.34**[a]
American Samoa (AMS)	26	14	173,970-269,654	0.94 ± 0.03	0.0054 ± 0.0033	**-7.31**
Cook Islands (CKI)	11	8	218,119-338,085	0.95 ± 0.05	0.0060 ± 0.0038	-3.10
Nuku Hiva, French Polynesia (NUK)	2	2	n/a[c]	n/a	n/a	n/a
Johnston Atoll (JON)	3	2	n/a	n/a	n/a	n/a
Hawaiian Islands (HAW)	2	2	n/a	n/a	n/a	n/a
All samples	54	23	180,146-279,226	0.93 ± 0.02	0.0054 ± 0.0033	**-15.65**
Acanthurus japonicus						
Philippines (PHI)	7	4	105,841-164,053	0.86 ± 0.10	0.0025 ± 0.0021	-1.06
Paracel Islands (PAR)	36	16	57,224-88,697	0.80 ± 0.07	0.0037 ± 0.0024	**-11.50**
Japan (JAP)	6	5	234,544-363,544	0.93 ± 0.12	0.0056 ± 0.0040	-1.47
All samples	48	17	131,792-204,277	0.83 ± 0.05	0.0040 ± 0.0025	**-10.48**
Acanthurus leucosternon						
Socotra, Yemen (SOC)	5	n/a	n/a	n/a	n/a	n/a
Republic of Seychelles (SEY)	31	17	137,967-213,849	0.92 ± 0.04	0.0055 ± 0.0033	**-10.57**
Chagos Archipelago (DIG)	63	25	147,559-228,717	0.91 ± 0.02	0.0050 ± 0.0031	**-18.75**
Cocos-Keeling Islands (COC)	34	18	57,486-89,104	0.87 ± 0.05	0.0062 ± 0.0037	**-10.19**
Christmas Island (XMA)	8	4	51,836-80,346	0.75 ± 0.14	0.0059 ± 0.0039	0.90
All samples	141	43	104,724-162,322	0.89 ± 0.02	0.0054 ± 0.0032	**-26.70**

Acanthurus nigricans						
Chagos Archipelago (DIG)	14	9	132,449-205,295	0.88 ± 0.08	0.0050 ± 0.0032	**-4.05**
Cocos-Keeling Islands (COC)	36	14	119,309-184,929	0.75 ± 0.08	0.0030 ± 0.0021	**-10.06**
Christmas Island (XMA)	43	16	121,805-188,798	0.75 ± 0.07	0.0033 ± 0.0022	**-11.51**
Republic of Palau (PAU)	49	18	75,685-117,312	0.72 ± 0.07	0.0025 ± 0.0018	**-17.82**
Guam, Mariana Islands (MAI)	11	7	129,098-200,102	0.82 ± 0.12	0.0036 ± 0.0026	**-3.32**
Marshall Islands (MAR)	15	9	199,921-309,878	0.85 ± 0.09	0.0052 ± 0.0033	**-3.57**
Kosrae, Caroline Islands (CAR)	27	13	135,471-209,980	0.80 ± 0.08	0.0036 ± 0.0024	**-8.56**
American Samoa (AMS)	20	10	108,797-168,635	0.76 ± 0.10	0.0029 ± 0.0021	**-6.55**
Tokelau Islands (TOK)	35	13	72,728-112,729	0.71 ± 0.09	0.0028 ± 0.0019	**-9.24**
Moorea, French Polynesia (MOR)	30	12	6,898-10,692	0.61 ± 0.11	0.0027 ± 0.0019	**-8.51**
Nuku Hiva, French Polynesia (NUK)	41	16	72,400-112,220	0.69 ± 0.08	0.0023 ± 0.0017	**-15.70**
Kiritimati, Republic of Kiribati (KIR)	36	14	169,766-263,137	0.75 ± 0.08	0.0037 ± 0.0024	**-8.39**
Palmyra Atoll (PAL)	22	5	197,096-305,499	0.34 ± 0.13	0.0009 ± 0.0009	**-3.20**
Hawaiian Islands (HAW)	22	9	63,925-99,084	0.61 ± 0.12	0.0021 ± 0.0016	**-6.40**
All samples	401	84	108,534-168,228	0.71 ± 0.03	0.0030 ± 0.0020	**-27.90**

[a] Numbers in bold are significant, $P < 0.05$ (or $P < 0.02$ for Fu's F_S estimates; Fu, 1997)
[b] Abbreviations is as follows: H_N, number of haplotypes.
[c] Metrics were not calculated for individual sites with $N < 6$, these samples were instead included in the species-level estimates.

Tests for cyt *b* neutrality revealed negative and significant Fu's F_S values (with $P < 0.02$) at most sites for all species, with only a few exceptions, and all with sample size less than 12 (*A. japonicus*, Japan, Fu's *Fs* = -1.06, *P* = 0.082; *A. japonicus*, Philippines, Fu's *Fs* = -1.47, *P* = 0.094; *A. leucosternon*, Christmas Island, Fu's *Fs* = 0.688, *P* = 0.213; Table 1). The range of mutation rates and τ values yielded somewhat overlapping estimates of time since the last population expansion for *A. achilles* (180,146 - 279,226 years), *A. japonicus* (131,792 - 204,277 years), *A. leucosternon* (104,724 - 162,322 years) and *A. nigricans* (108,534 - 168,228 years; Table 1). These findings indicate recent population expansion in all species throughout their ranges.

Among the individuals scored for the ten microsatellite loci (Table S1), the number of alleles per locus at each sampling site ranged from 2 (Ahy49) to 39 (Ahy182) and observed heterozygosity ranged from 0.10 (Ahy49) to 1.00 (at multiple loci). Based on within-site comparisons, few loci deviated from Hardy-Weinberg equilibrium (*A. achilles*: six out of 27, $P \leq 0.013$; *A. japonicus*: one out of 10, $P \leq 0.017$; *A. leucosternon*: six out of 40, $P \leq 0.012$; *A. nigricans*: 12 out of 140, $P \leq 0.009$) and linkage disequilibrium between loci was rare (*A. achilles*: three out of 108, $P \leq 0.009$; *A. japonicus*: zero out of 45, $P \leq 0.011$; *A. leucosternon*: nine out of 180, $P \leq 0.0083$; *A. nigricans*: six out of 630, $P \leq 0.007$) after correcting for multiple tests (Narum, 2006). Evidence of null alleles was detected in only three out of 27 (*A. achilles*), one out of 10 (*A. japonicus*), six out of 40 (*A. japonicus*) and 14 out of 140 (*A. nigricans*) within-site comparisons, although two of the loci were disproportionately represented; Ahy54 and Ahy203 tested positive for null alleles in 50% and 18% of the sampled populations, respectively. We therefore ran all subsequent analyses with and without these two loci to mitigate bias (data not shown); our findings were no different between datasets, and so we present results including all 10 microsatellite loci.

3.2 Population structure analysis

Despite a wide geographic distribution, we found modest to no population structure at mtDNA for either surgeonfish species with near range-wide sampling (*A. leucosternon*, $\Phi_{ST} = 0.038$, $P = 0.12$; *A. nigricans*, $\Phi_{ST} = 0.0065$, $P = 0.05$) and at microsatellite markers (*A. leucosternon*, $F_{ST} = 0.0063$, $P = 0.005$; *A. nigricans*, $F_{ST} = 0.0005$, $P = 0.49$) based on AMOVA. Indeed, most of the mtDNA (*A. leucosternon*: 100%; *A. nigricans*: 99.35%) and microsatellite variation (*A. leucosternon*: 99.38%; *A. nigricans*: 99.95%) was explained by within-site variation. To ensure that the observed outcome for microsatellite markers was not being driven by a single locus, we repeated AMOVA analyses for each species by removing one locus at a time, which gave consistent results for F_{ST} but variable significance estimates (*A. leucosternon*: $F_{ST} = 0.0020$ to 0.0078, $P = 0.003$ to 0.64; *A. nigricans*: $F_{ST} = -0.0019$ to 0.00062, $P = 0.02$ to 0.91). Sequence divergence between all four species ranged from $d_{corr} = 0.0047$ (*A. japonicus* and *A. nigricans*) to $d_{corr} = 0.0450$ (*A. achilles* and *A. leucosternon*), which is consistent with a period of separation between the lineages of *ca.* 235,000 to 2.25 million years based on an accepted benchmark for cyt *b* calibration in teleost fishes (see Materials and Methods). The time-frame for divergence is also consistent with estimates from Sorenson et al. (2013), a more robust phylogeny based on two mitochondrial and six nuclear loci. Due to sharing of haplotypes between *A. japonicus* and *A. nigricans*, we compared the largest sample of *A. japonicus* (at the Paracel Islands) to the nearest sample of *A. nigricans* (at Palau) with ARLEQUIN and observed strong population structure (Φ_{ST} = 0.6533, $P < 0.001$).

Population pairwise tests (Φ_{ST} or F_{ST}) revealed that mtDNA haplotype frequencies were not significantly different in all comparisons for *A. leucosternon* and 86 out of 91 comparisons for *A. nigricans* after correcting for multiple tests (Table 2; Table S2). Microsatellite allele frequencies (F_{ST}) were also not significantly different in five out of six comparisons for *A. leucosternon* and 88 out of 91 comparisons for *A. nigricans* after correcting for multiple tests (Table 2; Table S2). Notably, Hawai'i had no significant pairwise comparisons, an unusual finding for this isolated archipelago (DiBattista et al., 2011; Gaither et al., 2011; Eble et al., 2015).

Table 2. Matrix of population pairwise Φ_{ST} values with associated *P*-values in parentheses, based on mitochondrial DNA (mtDNA) cytochrome *b* sequence data from (a) *Acanthurus leucosternon* (*N* = 141) and (b) *A. nigricans* (*N* = 401) sampled at sites across the Indo-Pacific region. Morphological hybrids between the two species were omitted from this analysis and all negative Φ_{ST} values were adjusted to 0.

(a)

Location[a]	SEY	DIG	COC	XMA
SEY	–			
DIG	0 (0.53)	–		
COC	0.021 (0.23)	0.010 (0.03)	–	
XMA	0 (0.47)	0.081 (0.16)	0 (0.79)	–

[a] Site abbreviations are described in Table 1.

(b)

Location[a,b]	COC	XMA	DIG	PAU	MAI	MAR	CAR	AMS	TOK	MOR	NUK	KIR	PAL	HAW
COC	–													
XMA	0.002 (0.37)	–												
DIG	0.012 (0.22)	0.009 (0.26)	–											
PAU	0.012 (0.08)	0.011 (0.07)	**0.047** (0.02)	–										
MAI	0 (0.58)	0 (0.74)	0 (0.73)	0.011 (0.22)	–									
MAR	0.014 (0.18)	0.007 (0.27)	0 (0.55)	**0.058**[†] (<0.01)	5.2 x10^{-4} (0.44)	–								
CAR	0.004 (0.30)	0 (0.66)	0.028 (0.66)	0 (0.92)	0 (0.73)	0.026 (0.06)	–							
AMS	0.010 (0.22)	0.006 (0.27)	0 (0.54)	**0.059**[†] (<0.01)	0.016 (0.23)	0 (0.53)	**0.033** (0.04)	–						
TOK	0 (0.43)	0 (0.66)	0.015 (0.18)	0 (0.51)	0 (0.41)	0.018 (0.16)	0 (0.57)	0.021 (0.10)	–					
MOR	0 (0.56)	0 (0.47)	0.015 (0.17)	0.004 (0.25)	0 (0.55)	0.023 (0.11)	0 (0.43)	0.019 (0.12)	0 (0.71)	–				
NUK	0.008 (0.19)	0 (0.96)	0.033 (0.06)	0 (0.51)	0 (0.68)	**0.050** (0.02)	0 (0.72)	**0.038** (0.03)	0 (0.77)	0.001 (0.39)	–			
KIR	0.004 (0.28)	0 (0.96)	0.011 (0.22)	**0.015** (0.03)	0 (0.70)	0.009 (0.23)	0 (0.50)	0.003 (0.31)	5.30 x10^{-4} (0.41)	0 (0.71)	0 (0.52)	–		

PAL	**0.039** (0.01)	0.006 (0.26)	**0.083**† (<0.01)	0.001 (0.39)	**0.059** (0.03)	**0.086**† (<0.01)	0.009 (0.18)	**0.110**† (<0.01)	0.011 (0.18)	0.018 (0.08)	0.004 (0.31)	0.014 (0.11)	–	
HAW	3.10 x10^{-4} (0.41)	0 (0.80)	0.020 (0.13)	0 (0.90)	0 (0.49)	0.033 (0.07)	0 (0.79)	0.037 (0.06)	0 (0.79)	0 (0.76)	0 (0.79)	0 (0.59)	0 (0.63)	–

---

[a] Site abbreviations are described in Table 1.

[b] Significant values are indicated in bold ($P \leq 0.05$). "†" indicates $P \leq 0.01$ (corrected as per Narum, 2006).

STRUCTURE identified $K = 1$ as the most likely for *A. leucosternon* or *A. nigricans* (Fig S1). Similar results were obtained using the Delta *K* method (the second order rate of change of Ln P[D] in relation to *K*; Evanno et al., 2005) and so only the results of the former are presented here. Genetic differentiation was therefore consistently weak or absent within each species.

Haplotype networks were consistent with minimal genetic differentiation among sampling sites within-species (Fig 3; also see Fig S3). Each surgeonfish species shared a common haplotype among all sites (with few exceptions), and there was extensive geographic representation of haplotypes that were peripheral in the network. Despite discrete clusters of haplotypes unique to each species, haplotypes were shared among all species except *A. achilles*. These included the following: 1) all seven putative hybrid *A. achilles* x *A. nigricans* sampled at Kosrae and the Marshall Islands had *A. nigricans* mtDNA, 2) both putative hybrids of *A. japonicus* x *A. nigricans* sampled at Guam and Japan had *A. nigricans* mtDNA and 3) 51% of the putative hybrid *A. leucosternon* x *A. nigricans* sampled at Christmas and Cocos-Keeling Islands had *A. leucosternon* mtDNA and 49% had *A. nigricans* mtDNA.

Interestingly, no specimen with morphology matching *A. achilles* had an mtDNA haplotype affiliated with any of the other species, whereas four “pure” looking *A. japonicus* (Japan, $N = 3$; Paracel Islands, $N = 1$) had *A. nigricans* mtDNA, seven *A. nigricans* (Christmas Island, $N = 1$; Cocos-Keeling Islands, $N = 1$; Kiritmati, $N = 1$; Marshall Islands, $N = 2$; Tokelau Islands, $N = 2$) had *A. japonicus* mtDNA and six *A. leucosternon* (Christmas Island, $N = 1$; Cocos-Keeling Islands, $N = 3$; Diego Garcia, $N = 1$; Seychelles, $N = 1$) had *A. nigricans* mtDNA. Based on the assumption that F1 individuals have intermediate coloration, these findings indicate later generation hybrids at locations far outside the suture zone.

3.3 Genetic characterization of hybrids

Bayesian analyses of simulated datasets demonstrated that our microsatellite markers had sufficient power to assign parental and broad hybrid categories to individual fish (Fig S2). Therefore based on the 99% confidence limits for each parental category and hybrid class defined by simulations run in STRUCTURE, we found that 326 (out of 448) of our surgeonfish specimens were identified as F1 or F2, 76 were strictly F2, 22 were B2xAleu and 4 were B2xAni (Fig 4, where Aleu = *A. leucosternon* and Ani = *A. nigricans*); 20 individuals fell *between* the confidence limits of the F2 and B2xAleu hybrid classes. Of the 29 morphological hybrids we included for analysis, we found that 24 individuals were grouped as F1 or F2, an additional three individuals were classified as strictly F2, and one individual was grouped as a backcross (B2xAleu); only one individual fell *between* the confidence limits of the F2 and B2xAleu hybrid classes.

For NEWHYBRIDS, 250 of the total 448 individuals were identified as F2 with 90% probability, and of the remainder, 173 were categorized as “undefined hybrid” based on a 90% cumulative probability across all hybrid classes. Of the 29 morphological hybrids we included for analysis, all were identified as “undefined hybrid”, possibly indicating less power in assigning hybrid individuals to a particular hybrid class. No specimens were identified as parental by either program, despite well-defined confidence limits based on simulations in STRUCTURE, and a high probability of detection reported in NEWHYBRIDS.

Even though our simulations suggest high power for identifying parental species and their hybrids, this was based on highly variable loci (range of alleles per species: 20 to 74) and a small number of starting "pure" individuals for simulating hybrids (N = 31 and N = 36 for the two species). This deficiency limits the number of alleles represented in the simulated offspring and may even underestimate homoplasy, a common problem encountered with genetic stock identification of monitored fisheries (e.g. Anderson et al., 2008).

By running STRUCTURE on the entire data set without any assumptions regarding the purity of parental species, we found that K = 3 or 4 was supported (K = 1, meanLnP(K) = -36883.42 ± 0.33; K = 2, meanLnP(K) = -36470.46 ± 3.47; K = 3, meanLnP(K) = -35711.08 ± 3.04; K = 4, meanLnP(K) = -35669.82 ± 10.52; K = 5, meanLnP(K) = -36084.96 ± 142.33; K = 6, meanLnP(K) = -36989.40 ± 709.16; K = 7, meanLnP(K) = -37526.34 ± 1280.05; K = 8, meanLnP(K) = -37792.52 ± 947.76), and on average, *A. achilles* individuals almost always assigned to *A. achilles*, *A. japonicus* individuals were assigned to a mixture of all species except *A. achilles*, *A. leucosternon* individuals almost always assigned to *A. leucosternon* and *A. nigricans* individuals were equally assigned to *A. japonicus* and *A. nigricans* (Fig 5). In most cases, individuals identified as hybrids in the field were assigned to both of the suspected parent species (Fig 5). The STRUCTURE results were not qualitatively different when locus Ahy75 was omitted from the analysis given that every *A. achilles* was missing data from this locus (Fig S4; K = 1, meanLnP(K) = -33125.4 ± 0.50; K = 2, meanLnP(K) = -32763.5 ± 1.91; K = 3, meanLnP(K) = -32022.14 ± 4.07; K = 4, meanLnP(K) = -32033.12 ± 19.15; K = 5, meanLnP(K) = -32409.82 ± 129.92; K = 6, meanLnP(K) = -32806.18 ± 100.06; K = 7, meanLnP(K) = -33332.70 ± 954.83; K = 8, meanLnP(K) = -33526.72 ± 427.94). The results with STRUCTURE were similar when hybrids were placed in their own "populations" and when only species with the largest sample size were run together (*A. leucosternon*, *A. nigricans* and their hybrids), although when standardized to the lowest sample size (*N* = 6), the four “species” could not be separated (*K* = 1, meanLnP(*K*) = -1251.24 ± 0.32; *K* = 2, meanLnP(*K*) = -1251.64 ± 0.42; *K* = 3, meanLnP(*K*) = -1252.80 ± 1.06; *K* = 4, meanLnP(*K*) = -1258.84 ± 4.28). A DAPC analysis confirmed these findings (Fig 6 and Fig S5).

4. Discussion

The four surgeonfish species in our study diverged during the Pleistocene and have undergone more recent population expansion (100,000 to 280,000 years ago) based on mtDNA data. Contemporary secondary contact at range edges appears to have affected each species differently. Despite hybridizing with *A. nigricans*, *A. achilles* remains genetically distinct. *Acanthurus leucosternon*, on the other hand, exhibits introgression where its range overlaps with *A. nigricans* in the eastern Indian Ocean, but maintains distinct mtDNA haplotypes outside of the hybrid zone. *Acanthurus japonicus* co-occurs and hybridizes with *A. nigricans* throughout its small geographic range, leading to widespread introgression that makes it difficult to distinguish *A. japonicus* based on our genetic assays.

We observed a range of mtDNA divergence in the *A. nigricans* complex. The greatest mtDNA divergence distinguished *A. achilles* (d_{corr} with *A. leucosternon* = 0.044; d_{corr} with *A. nigricans* = 0.038; d_{corr} with *A. japonicus* = 0.040), consistent with previous phylogenetic analyses of the group (Sorenson et al., 2013). The Indian Ocean *A. leucosternon* and Indo-Pacific *A. nigricans* had lower sequence divergence (d_{corr} = 0.012), consistent with a separation approximately 600,000 years ago under a conventional molecular clock (see Materials and Methods). In contrast, *A. japonicus* (from the Northwest Pacific) contained a distinct cluster of haplotypes separated by only a few mutations from *A. nigricans*, but also shared haplotypes with the latter. Based on the mtDNA data used in this study, and the gradient of color patterns observed in the field, it is dubious to regard *A. japonicus* as a species distinct from *A. nigricans*, and further studies may indicate that the former is a regional color morph or subspecies (as originally suggested by references in Randall, 1956). Coloration differences do not always equate to species distinctions in reef fishes (DiBattista et al., 2012a,b). Our data also support the conclusion of Montanari et al. (2014) that closely-related reef fishes (*A. japonicus* and *A. nigricans* in this case) hybridize and introgress more readily than their distantly-related congeners (*A. achilles* and *A. nigricans*).

All available evidence indicates that *A. leucosternon* in the Indian Ocean and *A. nigricans* in the Pacific Ocean (and eastern Indian Ocean) are young species with similar morphology and ecology. Our estimate of divergence in these two species (approximately 600,000 years ago) indicates that Pleistocene glacial cycles played a role in their differentiation as sea level fluctuations episodically restricted connections between the Indian and Pacific Oceans (reviewed in Gaither and Rocha, 2013). Given the primarily Indian and Pacific distributions of *A. leucosternon* and *A. nigricans,* respectively, it is likely that the vicariant separation of ocean basins was the starting point for this evolutionary chronicle. The Indo-Pacific Barrier (i.e. Sunda Shelf) is ephemeral however, with separate Indian and Pacific Ocean faunas at glacial maxima, followed by a biogeographic gradient during interglacial periods as the two faunas expand into the alternate ocean basin (Briggs and Bowen, 2012; Gaither and Rocha, 2013; Hodge et al., 2014; Eble et al., 2015). Hence, evolutionary divergence between sister taxa may begin at the glacial-induced barrier, but species cohesion depends on maintaining separate gene pools during the interglacial periods of secondary contact (Chenoweth et al., 1998; van Herwerden and Doherty, 2006; Sorenson et al., 2014).

Even though *A. leucosternon*, *A. japonicus* and *A. achilles* maintain distinct geographic ranges, the range edges of all three species overlap with the Pacific *A. nigricans*. In all cases where the ranges overlap, the parent species are observed in mixed social groups, intermediate color morphs are present and the occurrence of these hybrids is verified with our DNA data. In the regions around Christmas Island, the Cocos-Keeling Islands and Bali in Indonesia, areas where the Pacific and Indian Ocean fauna mix, *A. nigricans* hybridizes with *A. leucosternon* to produce an intermediate (Fig 2c). Eastwards in the Hawaiian Islands and Micronesia, *A. nigricans* hybridizes with *A. achilles* (Fig 2f). In southern Japan, Taiwan and the northern Philippines, *A. nigricans* hybridizes with the genetically similar *A. japonicus* (Fig 2i). Our field observations and genetic analyses confirm previous reports of hybridization at regions of range overlap in this complex (Randall, 1956; Randall, 2002; Allen et al., 2007; Marie et al., 2007; Hobbs et al., 2009). Here our discussion will focus on the widespread Indian and Pacific Ocean species (*A. leucosternon* and *A. nigricans*) for which we have range-wide sampling, therefore allowing us to address broader questions.

4.1 Hybridization or incomplete lineage sorting?

In young species like *A. leucosternon* and *A. nigricans*, it can be difficult to separate contemporary hybridization from incomplete lineage sorting. This problem has attracted much interest (Sang and Zhong, 2000; Buckley et al., 2006; Holland et al., 2008), and yet few effective approaches exist for distinguishing these processes. Fortunately the range-wide sampling with mtDNA and microsatellite markers facilitates some level of resolution.

Lineage sorting eliminates ancestral polymorphism over time such that sister species eventually become reciprocally monophyletic (Avise, 2000), although monophyly is expected to be incomplete early in the divergence process (Sullivan et al., 2002). Under a scenario of complete isolation, mtDNA lineages are expected to attain reciprocal monophyly at approximately $2N_e$ generations on average, where N_e is effective population size (Avise, 2000). This process is longer for diploid nuclear loci, averaging $4N_e$ generations (Avise, 2000). Given that each species is a single population spread across a vast geographic range, historical N_e values must be in the millions. Hence a timeframe of 600,000 years for divergence of *A. leucosternon* and *A. nigricans* is certainly within the realm of incomplete lineage sorting for nuclear loci. Furthermore there is evidence for inter-specific hybridization, most robustly demonstrated by intermediate coloration and mtDNA lineages transferred between species. We conclude that the finding of no pure parents in STRUCTURE and NEWHYBRIDS is a product of both incomplete lineage sorting and hybridization, as might be expected in young species with incomplete reproductive barriers.

Introgression in reef fishes (and other taxa) appears to be characteristic of more recently separated species (e.g. van Herwerden et al., 2006; Yaakub et al., 2006; Marie et al., 2007), with 2% mtDNA divergence suggested as a benchmark for the initiation of this process (Montanari et al., 2014). Among the seven specimens that had discordance between mtDNA lineage and species ID (via coloration), four are inside the suture zone, three are from distant locations in the Indian Ocean and none are observed in the Pacific range of *A. nigricans* (Fig. 3). The alternative of extreme sex-biased mating seems unlikely given that these reef fish are dioecious, do not guard their eggs and fertilize externally in open water (Thresher, 1984; but see Marie et al., 2007).

A prominent caveat to microsatellite-based conclusions is the homoplasy of alleles, those that are identical in size but are not identical by descent. Some of the alleles “shared” between species arrived at their identical size by independent evolutionary pathways, and this phenomenon has yielded ambiguous claims of hybridization in other marine fishes (see Henrique et al., 2016). This caveat will certainly apply to species separated by 600,000 years. Hence the finding of a hybrid origin for all specimens is probably based on a combination of incomplete lineage sorting, introgression and allele homoplasy; fixed alleles with more diagnostic markers are likely required to tease apart their relative contribution.

4.2 Is there gene flow from the suture zone via hybrid offspring?

In the first genetic documentation of hybrids between *A. leucosternon* and *A. nigricans*, Marie et al. (2007) noted that a broader phylogeographic study would be necessary to understand the evolutionary scope of these events. Here we show that the genetic consequences of hybridization in the suture zone encompass the range of both species and are not restricted to the areas of sympatry. This effect is likely facilitated by larval dispersal of hybrid and backcrossed individuals out of the hybrid zone. Indeed, individuals of hybrid coloration have been observed outside the hybrid zone (Craig et al., 2008; J.P.A. Hobbs pers. obs.). Previous reef fish studies have also detected introgression in allopatric populations that are well beyond the hybrid zone (McMillan et al., 1999; DiBattista et al., 2012). Despite a high probability of detecting pure members of each species based on simulations (Fig S2), our analysis detected no pure parental individuals in all of our surgeonfish samples (Fig 4). Indeed, 90% to 100% were identified as F1 or F2 hybrids based on microsatellites, with the remainder identified as some form of later generation backcross. The use of species diagnostic microsatellite markers would improve our resolution of hybrid categories (*e.g.*, Roberts et al., 2009), but the identification of hybrids based on both coloration and genetics validates our field classification of "hybrid" fishes.

Even though individuals with intermediate coloration were confirmed to be hybrids in this study (and in other reef fishes: Hobbs and Allen, 2014); backcrossed individuals of many reef fish species often have the same coloration as the parent species and go undetected (McMillan et al., 1999; Yaakub et al., 2006; Montanari et al., 2012; 2014). As hybrid larvae disperse out of the hybrid zone, settle onto reefs and backcross with populations containing individuals that look like the parent species, the hybrid coloration is rapidly lost (e.g. McMillan et al., 1999). The signal of introgression, however, remains, and continues to spread throughout the range of a species as larvae disperse between populations and interbreed. Consequently, the zone of introgression is usually much larger than the hybrid zone in reef fish (McMillan et al., 1999). Our use of allopatric populations of *A. leucosternon* and *A. nigricans* to represent "purebred" individuals may blur the distinction among hybrid classes, but we suspect that hybrid ancestry is widespread in this species complex. A corollary of this conclusion is that F1 hybrids must have some degree of fertility to produce backcrossed offspring. We conclude that hybrid offspring are not an evolutionary dead end in this system, rather they lead to widespread introgression.

The Indian Ocean *A. leucosternon* and Pacific Ocean *A. nigricans* have negligible population structure across their ranges, and both have a lengthy post-larval stage that may facilitate dispersal between populations (Randall, 2002). While a link between dispersal ability and duration of the pelagic stage has considerable intuitive appeal, the relationship is not a simple one (Selkoe and Toonen, 2011; Selkoe et al., 2014). Nonetheless, fishes with a long pelagic larval stage tend to have extensive dispersal, broad ranges and dominate communities on isolated islands (Horne et al., 2008; Reece et al., 2011; Hobbs et al., 2013). Some of the most dispersive reef fishes are those that can stay in the water column after transforming from larva to juvenile (or pre-juvenile), including trumpetfishes (Bowen et al., 2001), squirrelfishes (Craig et al., 2007) and (most relevant) other surgeonfishes (Thresher, 1984). The sister species surveyed here provide additional examples of highly dispersive fishes, a trait that may be a general feature of the genus *Acanthurus* (Eble et al., 2011). In these circumstances, perhaps it is inevitable that the hybrid larvae can also disperse widely and backcross far from the point of origin. A 2008 expedition to the Chagos Archipelago revealed a previously unknown population of *A. nigricans* in the central Indian Ocean, and new *A. leucosternon* x *A. nigricans* hybrid records at Chagos and other archipelagos in the Indian Ocean support a broad geographic distribution of introgressed individuals in these two species (Craig, 2008). We conclude that hybridization coupled with high dispersal is the key to understanding the pattern of hybrid ancestry across the *A. nigricans* complex.

4.3 Evolutionary consequences

The suture zone at Christmas and Cocos-Keeling Islands is where many allopatric Indian and Pacific Ocean sister species come into secondary contact, including *A. leucosternon* and *A. nigricans* (Hobbs and Allen, 2014). At this suture zone some sister species exhibit subtle differences in diet and habitat use (Hobbs et al., 2010; Hobbs and Allen, 2014), however, other sister species have seemingly identical patterns of habitat use, diet, behavior and are equally likely to form homo- or heterospecific mating pairs (Montanari et al., 2012; 2014). Within their allopatric distributions, *A. leucosternon* occurs at inshore reefs in large aggregations, whereas *A. nigricans* tends to be more common on the outer reef below the surge zone (Kuiter and Debelius 2001; Randall, 2002). However, at Christmas and Cocos-Keeling Islands the two species occupy the same habitat, form mixed schools and feed together, which may break down pre-zygotic barriers to hybridization (Marie et al., 2007; Hobbs and Allen, 2014; Montanari et al., 2016). Furthermore, the hybrid individuals also cohabitate and feed with the parents indicating ecological and behavioral post-zygotic barriers are also overcome. The breakdown of these pre- and post-zygotic barriers aids hybridization and, along with dispersal out of the hybrid zone and lack of selection against introgressed individuals, would explain the pattern of widespread introgression in *A. leucosternon* and *A. nigricans*.

Our field observations also confirm that the other species in the complex (*A. achilles* and *A. japonicus*) cohabitate and feed with *A. nigricans* in the narrow areas of range overlap. Furthermore, the hybrids are seen to mix and feed with both parent species in these narrow contact zones (J.P.A. Hobbs, unpubl. data). However, overcoming ecological and behavioral pre- and post-zygotic barriers to breeding has had contrasting evolutionary outcomes. *Acanthurus japonicus* is genetically indistinct from *A. nigricans* throughout its small range, however *A. achilles* maintains its genetic identity. This indicates that some post-zygotic mechanisms (e.g. failed production of F2 individuals, selection against introgressed individuals) may be maintaining species integrity.

Hybridization is increasingly being detected in reef fishes and is not surprising given that most reef fishes have traits that would diminish barriers to reproductive isolation: broad distributions, closely related species in sympatry, external fertilization and a lack of parental care (Palumbi 1994; Montanari et al., 2016). It is now evident that hybridization in reef fishes occurs under different scenarios, and across a range of taxa and geographic locations (Hobbs et al., 2013; Hobbs and Allen, 2014; DiBattista et al., 2015). Hybridization in reef fishes has a range of evolutionary consequences (Richards and Hobbs, 2014), even within a species complex (this study). These hybridization events provide exciting opportunities to investigate speciation in reef fishes and the role of introgression in structuring phylogenetic relationships (Kuriiwa et al., 2007). Although the use of genetic approaches is critical to confirm hybridization, it is also important to distinguish between incomplete lineage sorting, introgression and allele homoplasy (Henrique et al., 2016). All of these processes may occur in young hybridizing species and overlap to produce similar genetic signatures. The presence of suture zones allows for comparative studies across multiple taxa that are hybridizing in one location (Hobbs and Allen, 2014; DiBattista et al., 2015). Furthermore, species complexes (such as the *A. nigricans* complex) that span multiple suture zones provide an ideal scenario to determine the role that hybridization plays in the evolutionary history of reef fishes.

Acknowledgements

This research was supported by the National Science Foundation grants OCE-0929031 to BWB, NOAA National Marine Sanctuaries Program MOA No. 2005-008/66882 to R.J. Toonen, Seaver Institute, KAUST Office of Competitive Research Funds under Award No. CRG-1-2012-BER-

002 to MLB, baseline research funds to MLB, National Geographic Society Grant 9024-11 to JDD and by a Natural Sciences and Engineering Research Council of Canada postgraduate fellowship to JDD. For specimen collections we thank Kim Andersen, Paul Barber, J. Howard Choat, Richard Coleman, Joshua Copus, Toby Daly-Engel, Joshua Drew, Jeff Eble, Iria Fernandez-Silva, Kevin Flanagan, Michelle Gaither, Brian Greene, Song He, Matthew Iacchei, Stephen Karl, Randall Kosaki, Carl Meyer, Yannis Papastamatiou, David Pence, Mark Priest, Richard Pyle, Joshua Reece, D. Ross Robertson, Jennifer Schultz, Tane Sinclair-Taylor, Derek Smith, Zoltan Szabo, Kim Tenggardjaja, Bill Walsh, Ivor Willliams, Zeng Xiaoqi, Jill Zamzow and the crew of the R.V. *Hi'ialakai*. For logistic support we thank Robert Toonen, Randall Kosaki, Serge Planes, Jo-Ann Leong, Charles Sheppard, Salah Saeed Ahmed, Fouad Naseeb, Thabet Abdullah Khamis, Ahmed Issa Ali Affrar (Socotra Specialist Tours), Hawaii Department of Land and Natural Resources, Coral Reef Research Foundation, Papahānaumokuākea Marine National Monument, the Ocean University of China - College of Fisheries, Ministry of Water and Environment of Yemen, the Environment Protection Authority (EPA) Socotra, Administration of the British Indian Ocean Territory, Western Australia Department of Fisheries, Parks Australia and U.S. Fish and Wildlife Service. We thank Tane Sinclair-Taylor, Keoki Stender, Kenji Sorita and Hiroshi Senou (Kanagawa Prefectural Museum of Natural History) for providing images. Thanks to J. Howard Choat and Ben Victor for providing life-history information, Catherine Cullingham for providing a copy of HYBRIDLAB, Stefano Montanari for the map figure, members of the ToBo lab for logistic support and the Center for Genomics, Proteomics and Bioinformatics at the University of Hawai'i at Mānoa for their assistance with genotyping. We also thank Giacomo Bernardi and two anonymous reviewers for their helpful

comments on an earlier version of this manuscript. This is contribution no. XX from the Hawai'i Institute of Marine Biology and no. XX from the School of Ocean and Earth Science and Technology.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., 2013. Hybridization and speciation. Journal of Evolutionary Biology, 26, 229-246.

Anderson, E.C., Thompson, E.A., 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics, 160, 1217-1229.

Anderson, E.C., Waples, R.S., Kalinowski, S.T., 2008. An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences, 65, 1475-1486.

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge MA.

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16, 37-48.

Barton, N.H., 2001. The role of hybridization in evolution. Molecular Ecology, 10, 551-568.

Bellwood, D.R., Wainwright, P.C., 2002. The history and biogeography of fishes on coral reefs. In: Coral Reef Fishes. Dynamics and Diversity on a Complex Ecosystem (ed Sale PF), pp. 5-32. Academic Press, New York NY.

Bird, C.E., Holland, B.S., Bowen, B.W., Toonen, R.J., 2011. Diversification of endemic sympatric limpets (*Cellana* spp.) in the Hawaiian Archipelago. Molecular Ecology. 20:2128-2141.

Bird, C.E., Fernandez-Silva, I., Skillings, D.J., Toonen, R.J., 2012. Sympatric speciation in the post "Modern Synthesis" era of evolutionary biology. Evolutionary Biology, 39, 158-180.

Bolnick, D.B., Fitzpatrick, B.M., 2007. Sympatric Speciation: Models and Empirical Evidence. Annual Review of Ecology, Evolution, and Systematics, 38, 459-87.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S., Robertson, D.R., 2001. Phylogeography of the trumpetfish (*Aulostomus* spp.): ring species complex on a global scale. Evolution, 55, 1029-1039.

Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., Craig, M.T., DiBattista, J.D., Eble, J.A., Gaither, M.R., Skillings, D., Bird, C.E., 2013. Origins of tropical marine biodiversity. Trend in Ecology and Evolution, 28, 359-366.

Briggs, J.C., Bowen, B.W., 2012. A realignment of marine biogeographic provinces with particular reference to fish distributions. Journal of Biogeography, 39, 12-30.

Buckley, T.R., Cordeiro, M., Marshall, D.C., Simon, C., 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). Systematic Biology, 55, 411-425.

Chenoweth, S.F., Hughes, J.M., Keenan, C.P., Lavery, S., 1998. When oceans meet: a teleost shows secondary intergradation at an Indian-Pacific interface. Proceedings of the Royal Society of London Series B-Biological Sciences, 265, 415-420.

Choat, J.H., 2006. Phylogeography and reef fishes: bringing ecology back into the argument. Journal of Biogeography, 33, 967-968.

Coleman, R.R., Gaither, M.R., Kimokeo, B., Stanton, F., Bowen, B.W., Toonen, R.J., 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.

Craig, M.T., 2008. The goldrim surgeonfish (*Acanthurus nigricans*; Acanthuridae) from Diego Garcia, Chagos Archipelago: first record for the central Indian Ocean. Zootaxa, 1850, 65-68.

Craig, M.T., Eble, J.A., Robertson, D.R., Bowen, B.W., 2007. High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). Marine Ecology Progress Series, 334, 245-254.

DiBattista, J.D., Wilcox, C., Craig, M.T., Rocha, L.A., Bowen, B.W., 2011. Phylogeography of the Pacific Blueline Surgeonfish *Acanthurus nigroris* reveals a cryptic species in the Hawaiian Archipelago. Journal of Marine Biology, Article ID 839134.

DiBattista, J.D., Waldrop, E., Bowen, B.W., Schultz, J.K., Gaither, M.R., Pyle, R.L., Rocha, L.A., 2012a. Twisted sister species of Pygmy Angelfishes: Discordance between taxonomy, coloration, and phylogenetics. Coral Reefs, 31, 839-851.

DiBattista, J.D., Craig, M.T., Rocha, L.A., Feldheim, K.A., Bowen, B.W., 2012b. Phylogeography of the Indo-Pacific butterflyfishes, *Chaetodon meyeri* and *Chaetodon ornatissimus*: Sister species reveal divergent evolutionary histories and discordant results from mtDNA and microsatellites. Journal of Heredity, 103, 617-629.

DiBattista, J.D., Rocha, L.A., Hobbs, J-P.A., He, S., Priest, M.A., Sinclair-Taylor, T.H., Bowen, B.W., Berumen, M.L., 2015. When biogeographic provinces collide: Hybridization at the crossroads of marine biogeographic provinces in the Arabian Sea. Journal of Biogeography, 42, 1601-1614

Dobzhansky, T., 1937. Genetic and the Origin of Species. Columbia University Press, New York.

Drew, J.A., Allen, G.R., Erdmann, M.V., 2010. Congruence between mitochondrial genes and color morphs in a coral reef fish: population variability in the Indo-Pacific damselfish *Chrysiptera rex* (Snyder, 1909). Coral Reefs, 29, 439-444.

Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., et al., 2009. Geneious v4.8. Available from http://www.geneious.com.

Eble, J.A., Rocha, L.A., Craig, M.A., Bowen, B.W., 2011. Not all larvae stay close to home: Insights into marine population connectivity with a focus on the Brown Surgeonfish (*Acanthurus nigrofuscus*). Journal of Marine Biology, Article ID 518516.

Eble, J.A., Bowen, B.W., Bernardi, G., 2015. Phylogeography of coral reef fishes. Pp. 64-75 *In* C. Mora (ed.) Ecology of Fishes on Coral Reefs. University of Hawaii Press, Honolulu.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online, 1, 47-50.

Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends in Genetics, 28, 342-350.

Fu, Y.-X, 1997. Statistical tests of neutrality against population growth, hitchhiking and background selection. Genetics, 147, 915-925.

Gaither, M.R., Rocha, L.A., 2013. Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. Journal of Biogeography, 40, 1638-1648.

Gaither, M.R., Jones, S.A., Kelley, C., Newman, S.J., Sorenson, L., Bowen, B.W., 2011. High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian Archipelago. PLoS One, 6, e28913.

Gardner, J.P.A., 1997. Hybridization in the sea. In: Advances in Marine Biology (eds Blaxter J, Southward A), pp. 1-78. Academic Press, New York.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology, 52, 696-704.

Henriques, R., von der Heyden, S., Matthee, C.A., 2016. When homoplasy mimics hybridization: a case study of Cape hakes (*Merluccius capensis* and *M. paradoxus*). Peer J, 4, e1827.

Hobbs, J.P., Jones, G.P., Munday, P.L., 2010. Rarity and extinction risk in coral reef angelfishes on isolated islands: interrelationships among abundance, geographic range size and specialization. Coral Reefs, 29, 1-11.

Hobbs, J.P.A., Jones, G.P., Munday, P.L., Connolly, S.R., Srinivasan, M., 2012. Biogeography and the structure of coral reef fish communities on isolated islands. Journal of Biogeography, 39, 130-139.

Hobbs, J.P.A., van Herwerden, L., Pratchett, M.S., Allen, G.R., 2013. Hybridisation among butterflyfishes. Pp 48-69. In: Pratchett, M.S., Berumen, M.L., Kapoor, B. (Eds) Biology of Butterflyfishes. CRC Press, Boca Raton, FL.

Hobbs, J.P., van Herwerden, L., Jerry, D.R., Jones, G.P., Munday, P.L., 2013. High genetic diversity in geographically remote populations of endemic and widespread coral reef angelfishes (Genus: Centropyge). Diversity, 5, 39-50.

Hodge, J.R., van Herwerden, L., Bellwood, D.R., 2014. Temporal evolution of coral reef fishes: global patterns and disparity in isolated locations. Journal of Biogeography, 41, 2115-2127.

Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V., Huber, K.T., 2008. Using supernetworks to distinguish hybridization from lineage-sorting. BMC Evolutionary Biology, 8, 202.

Horne, J.B., van Herwerden, L., Choat, J.H., Robertson, D.R., 2008. High population connectivity across the Indo-Pacific: Congruent lack of phylogeographic structure in three reef fish congeners. Molecular Phylogenetics and Evolution, 49, 629-638.

Hubbs, C.L., 1955. Hybridization between fish species in nature. Systematic Zoology, 4, 1-20

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403-1405.

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11, 94.

Jones, G.P., Caley, M.J., Munday, P.L., 2002. Rarity in coral reef fish communities. In: Sale PF (ed) Coral reef fishes: dynamics and diversity in a complex ecosystem. Academic Press, San Diego, pp 81-101.

Kimura, M., 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proceedings of the National Academy of Sciences USA, 78, 454-458.

Kirkpatrick, M., 2001. Reinforcement during ecological speciation. Proceedings of the Royal Society of London B: Biological Sciences, 268, 1259-1263.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources.

Kuiter, R.H., Debelius, H., 2001. Surgeonfishes, Rabbitfishes and their Relatives. A Comprehensive Guide to Acanthuroidei (Marine Fish Families). TMC Publishing, Chorleywood, NJ.

Kuriiwa, K., Hanzawa, N., Yoshino, T., Kimura, S., Nishida, M., 2007. Phylogenetic relationships and natural hybridization in rabbitfishes (Teleostei: Siganidae) inferred from mitochondrial and nuclear DNA analyses. Molecular Phylogenetics and Evolution, 45, 69-80.

Lessios, H.A., 2008. The great American schism: Divergence of marine organisms after the rise of the Central American isthmus. Annual Reviews of Ecology Evolution and Systematics, 39, 63-91.

Marie, A.D., van Herwerden, L., Choat, J.H., Hobbs, J.-P.A., 2007. Hybridization of reef fishes at the Indo-Pacific biogeographic barrier: a case study. Coral Reefs, 26, 841-850.

Mayr, E., 1942. Systematics and the Origin of Species from the Viewpoint of a Zoologist. Columbia University Press, New York.

McMillan, W.O., Weigt, L.A., Palumbi, S.R., 1999. Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). Evolution, 53, 247-260.

Meeker, N.D., Hutchinson, S.A., Ho, L., Trede, N.S., 2007. Method for isolation of PCR-ready genomic DNA from zebrafish tissues. BioTechniques, 43, 610-614.

Montanari, S., van Herwerden, L., Pratchett, M., Hobbs, J.-P., Fugedi, A., 2012. Reef fish hybridization: lessons learnt from butterflyfishes (genus *Chaetodon*). Ecology and Evolution, 2, 310-328.

Montanari, S.R., Hobbs, J.-P.A., Pratchett, M.S., Bay, L.K., van Herwerden, L., 2014. Does genetic distance between parental species influence outcomes of hybridization among coral reef butterflyfishes? Molecular Ecology, 23, 2757-2770.

Montanari, S.R., Hobbs, J-P.A., Pratchett, M.S., van Herwerden, L., 2016. The importance of ecological and behavioural data in studies of hybridisation among marine fishes. Reviews in Fish Biology and Fisheries. DOI 10.1007/s11160-016-9420-7

Narum, S.R., 2006. Beyond Bonferroni: less conservative analyses for conservation genetics. Conservation Genetics, 7, 783-787.

Nielsen, E.E., Bach, L.A., Kotlicki, P., 2006. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. Molecular Ecology Notes, 6, 971-973.

Nosil, P., 2012. Ecological Speciation. Oxford University Press, London.

Palumbi, S.R., 1994. Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics, 25, 547-572.

Pardo-Diaz, C., Salazar, C., Baxter, S.W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W.O., Jiggins, C.D., 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. PLoS Genetics, 8, e002752.

Posada, D., 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution, 25, 1253-1256.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Pyle, R.L., Randall, J.E., 1994. A review of hybridization in marine angelfishes (Perciformes: Pomacanthidae). Environmental Biology of Fishes, 41, 127-145.

R Development Core Team, 2014. Tools for general maximum likelihood estimation. *R package version* 1.5.2

Randall, J.E., 1956. A revision of the surgeonfish genus *Acanthurus*. Pacific Science, 10, 159-235.

Randall, J.E., 2002. Surgeonfishes of Hawaii and the World. Bishop Museum Bulletin in Zoology 4. Mutual Publishing, Honolulu HI.

Randall, J.E., Frische, J., 2000. Hybrid surgeonfishes of the *Acanthurus achilles* complex. *Aqua*, 4, 51-56.

Reece, J.S., Bowen, B.W., Smith, D.G., Larson, A.F., 2010. Molecular phylogenetics of moray eels (Murenidae) demonstrates multiple origins of a shell-crushing jaw (*Gymnomuraena, Echidna*) and multiple colonizations of the Atlantic Ocean. Molecular Phylogenetics and Evolution, 57, 829-835.

Reece, J.S., Bowen, B.W., Larson, A., 2011. Long larval duration in moray eels (Muraenidae) ensures ocean-wide connectivity despite differences in adult niche breadth. Marine Ecology Progress Series, 437, 269-277.

Remington, C.L., 1968. Suture-zones of hybrid interaction between recently joined biotas. In *Evolutionary Biology* (ed Dobzansky T), 321-428. Plenum Press, New York.

Rhymer, J.M., Simberloff, D., 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics, 27, 83-109.

Rieseberg, L., 2011. Adaptive introgression: the seeds of resistance. Current Biology, 21, R581-R583.

Roberts, D.G., Gray, C.A., West, R.J., Ayre, D.J., 2009. Evolutionary impacts of hybridization and interspecific gene flow on an obligately estuarine fish. Journal of Evolutionary Biology, 22, 27-35.

Robertson, D.R., Polunin, N.V.C., Leighton, K., 1979. The behavioral ecology of three Indian Ocean surgeonfishes (*Acanthurus lineatus, A. leucosternon* and *Zebrasoma scopas*): their feeding strategies, and social and mating systems. Environmental Biology of Fishes, 4, 125-170.

Rocha, L.A., Bowen, B.W., 2008. Speciation in coral reef fishes. Journal of Fish Biology, 72, 1101-1121.

Rogers, A.R., Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution, 9, 552-569.

Sang, T., Zhong, Y., 2000. Testing hybridization hypotheses based on incongruent gene trees. Systematic Biology, 49, 422-434.

Schwartz, T.S., Beheregaray, L.B., 2008. Using genotype simulations and Bayesian analyses to identify individuals of hybrid origin in Australian bass: lessons for fisheries management. Journal of Fish Biology, 72, 435-450.

Seehausen, O., 2004. Hybridization and adaptive radiation. Trends in Ecology & Evolution, 19, 198-207.

Seehausen, O., 2006. African cichlid fish: a model system in adaptive radiation research. Proceedings of the Royal Society of London B: Biological Sciences, 273, 1987-1998.

Selkoe, K.A., Toonen, R.J., 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Marine Ecology Progress Series, 436, 291-305.

Selkoe, K.A., Gaggiotti, O., Andrews, K.R., Bernal, M.A., Bird, C., Bolick, H., et al., 2014. Emergent patterns of population genetic structure for a coral reef community. Molecular Ecology, 23, 3064-3079.

Song, C.B., Near, T.J., Page, L.M., 1998. Phylogenetic relations among percid fishes as inferred from mitochondrial cytochrome *b* DNA sequence data. Molecular Phylogenetics and Evolution, 10, 343-353.

Sorenson, L., Santini, F., Carnevale, G., Alfaro, M.E., 2013. A multi-locus timetree of surgeonfishes (Acanthuridae, Percomorpha) with revised family taxonomy. Molecular Phylogenetics and Evolution, 68, 150-160.

Sorenson, L., Allen, G.R., Erdmann, M.V., Dai, C.-F., Liu, S.-Y.V., 2014. Pleistocene diversification of the *Pomacentrus coelestis* species complex (Pisces: Pomacentridae): historical biogeography and species boundaries. Marine Biology, 161, 2495-2507.

Sullivan, J.P., Lavoué, S., Hopkins, C.D., 2002. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). Evolution, 56, 597-616.

Taberlet, P., Meyer, A., Bouvet, J., 1992. Unusual mitochondrial DNA polymorphism in two local populations of blue tit *Parus caeruleus*. Molecular Ecology, 1, 27-36.

Thresher, R.E., 1984. Reproduction in Reef Fishes. TFH publications, Neptune City, NJ.

van Herwerden, L., Doherty, P.J., 2006. Contrasting genetic structures across two hybrid zones of a tropical reef fish, *Acanthochromis polyacanthus* (Bleeker 1855). Journal of Evolutionary Biology, 19, 239-252.

van Herwerden, L., Choat, J.H., Dudgeon, C.L., Carlos, G., Newman, S.J., Frisch, A., van Oppen, M., 2006. Contrasting patterns of genetic structure in two species of the coral trout *Plectropomus* (Serranidae) from east and west Australia: Introgressive hybridization or ancestral polymorphism. Molecular Phylogenetics and Evolution, 41, 420-435.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, 4, 535-538.

Volmer, S.V., Palumbi, S.R., 2002. Hybridization and evolution of reef coral diversity. Science, 296, 2023-2025.

Yaakub, S.M., Bellwood, D.R., van Herwerden, L., Walsh, F.M., 2006. Hybridization in coral reef fishes: introgression and bi-directional gene exchange in *Thalassoma* (family Labridae). Molecular Phylogenetics and Evolution, 40, 84-100.

Yaakub, S.M., Bellwood, D.R., van Herwerden, L., 2007. A rare hybridization event in two common Caribbean wrasses (genus *Halichoeres;* family Labridae). Coral Reefs, 26, 597-602.

Data Accessibility

mtDNA sequences: Genbank accession numbers KM456297to KM456989.

Microsatellite primer sequences: Genbank accession numbers HQ130123 to HQ130132.

Sample locations and microsatellite genotypes: DRYAD entry doi: 10.5061/dryad.2m973.

Figure Captions

Figure 1. Scaled map indicating collection sites for *Acanthurus achilles* (light pink circles), *A. japonicus* (light green circles), *A. leucosternon* (light blue circles) and *A. nigricans* (light orange circles), or morphological hybrids between the species (grey circles) in the Indo-Pacific region. Dashed lines outline the known distributions for each species based on Randall (2002). Site abbreviations are defined in Table 1.

Figure 2. *Acanthurus achilles* (e, Johnston Atoll), *A. japonicus* (h, Anilao, Philippines), *A. leucosternon* (b, Christmas Island, Australia), *A. nigricans* (a, d, g, Christmas Island, Australia) and the hybrids between them (c, *A. leucosternon* × *A. nigricans*, Christmas Island, Australia; f, *A. achilles* × *A. nigricans*, Big Island, Hawaiian Archipelago; i, *A. japonicus* × *A. nigricans*,

Hachijo-jima, Izu Islands). Photo credit: a,b,c,d,g = Tane Sinclair-Taylor, e,h = Luiz Rocha, f = Keoki Stender, i = Kenji Sorita.

Figure 3. Median-joining statistical parsimony networks based on 491 bp of mitochondrial cytochrome *b* from *Acanthurus achilles* (*N* = 54), *A. japonicus* (*N* = 49), *A. leucosternon* (*N* = 110), *A. nigricans* (*N* = 344) and morphological hybrids between the species (*N* = 48) sampled across the Indo-Pacific region. Each circle represents a haplotype and its size is proportional to total frequency; for scale, the largest circle represents 230 individuals. Branches and black crossbars represent a single nucleotide change unless otherwise noted, whereas small, open circles indicate unsampled haplotypes; colors denote collection location or species as indicated by the embedded key. The network was simplified by removing all haplotypes occurring in single specimens (*N* = 103), except for the hybrids; this did not influence the overall pattern.

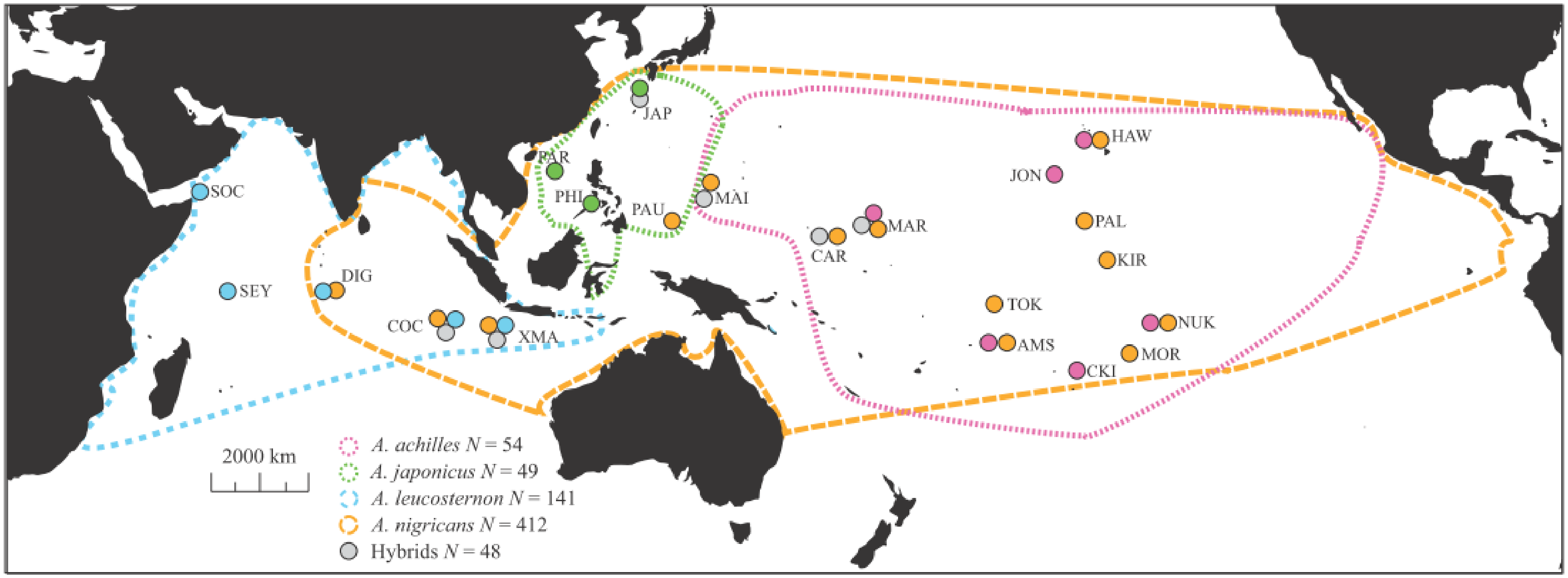
Figure 4. Distribution of *q*-values (i.e. proportion of genotype with *Acanthurus leucosternon* or *A. nigricans* ancestry) from STRUCTURE for all surgeonfish collections (*N* = 478) partitioned by sample site. A *q*-value of 0.92 is the *A. leucosternon* threshold of assignment and 0.08 is the *A. nigricans* threshold of assignment. Each individual is given a symbol based on their final classification in NEWHYBRIDS (F2 = circle; undefined hybrid = square; no classification = an X). Morphological hybrids were included in this analysis as a form of quality control for class assignment and therefore highlighted in green. *A. leucosternon* and *A. nigricans* samples were also highlighted, in red and blue, respectively. Dashed black lines represent the threshold for

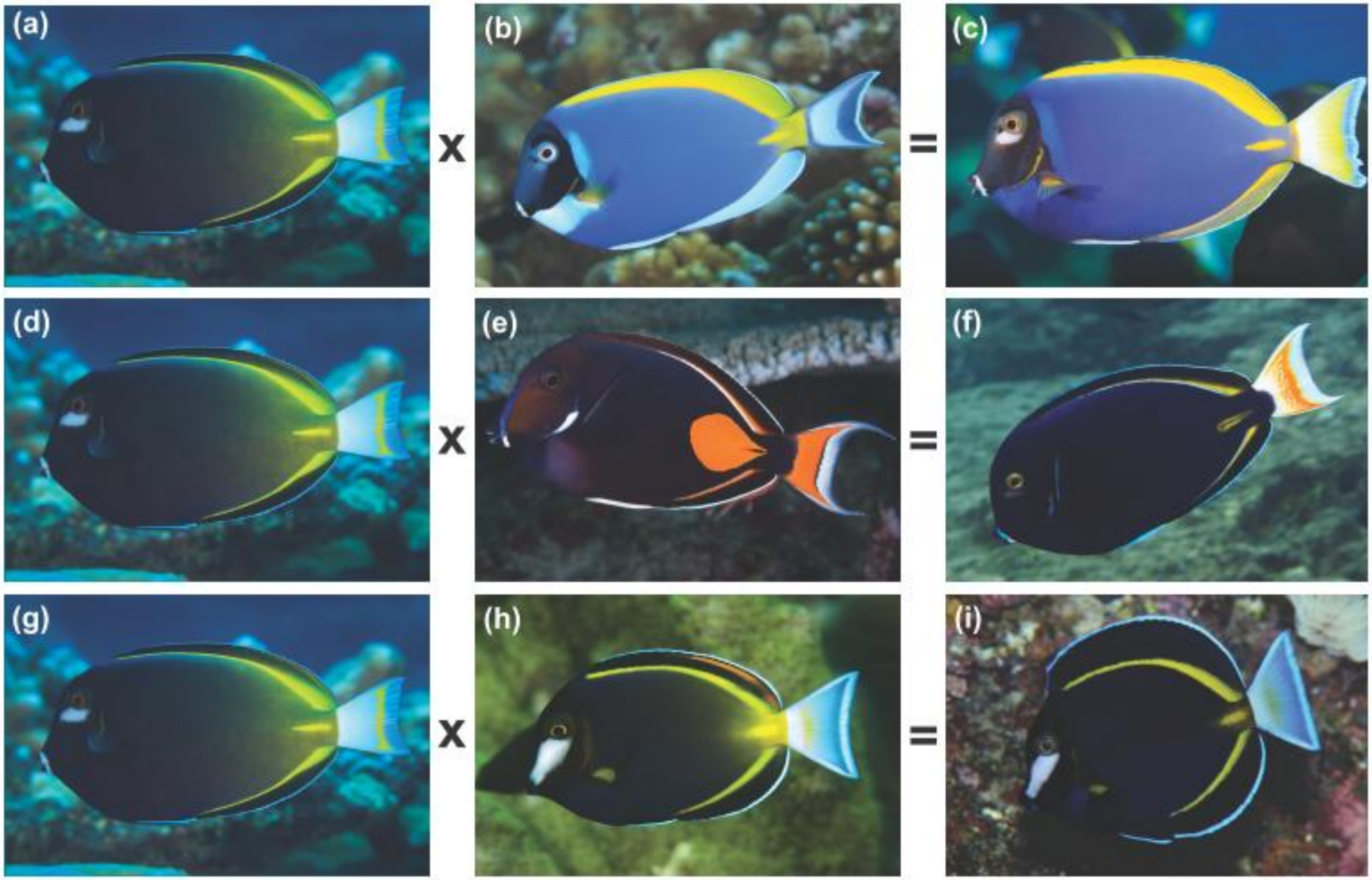
assignment to parental *A. leucosternon* or parental *A. nigricans* categories. Site abbreviations are described in Table 1.

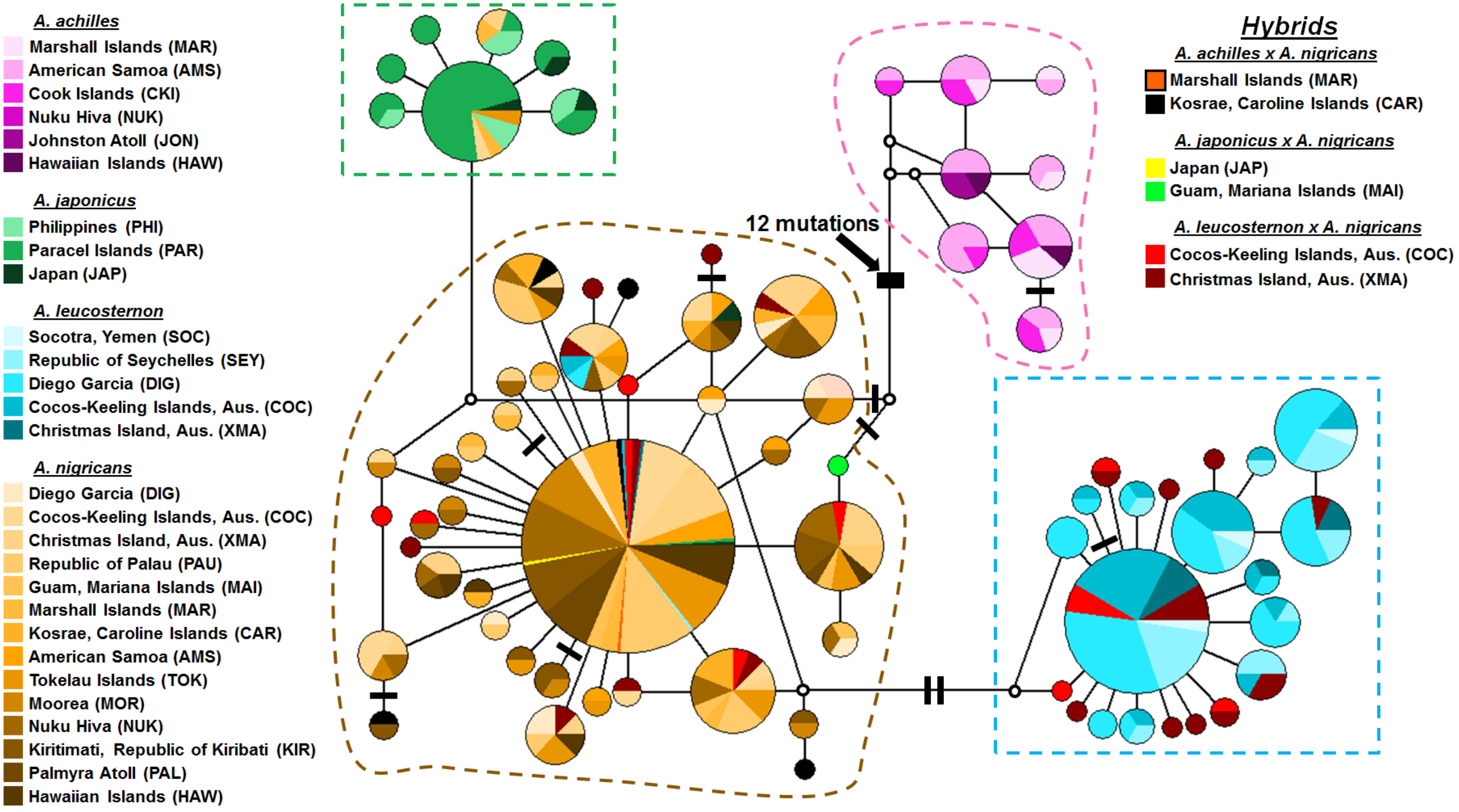
Figure 5. Results of STRUCTURE analysis performed on 10 microsatellite loci with *K* = 1 to *K* = 4 categories for the four species of surgeonfish (*Acanthurus achilles, A. japonicus, A. leucosternon* and *A. nigricans*) and their morphological hybrids. *K* = 3 (meanLnP(*K*) = -35711.08 ± 3.04) or *K* = 4 (meanLnP(*K*) = -35669.82 ± 10.52) was identified as the most likely in this case. For *K* = 4, light pink represents *A. achilles*, light green represents *A. japonicus*, light blue represents *A. leucosternon* and light orange represents *A. nigricans*. Populations are listed as abbreviations outlined in Table 1. Morphological hybrids are included within the following *A. nigricans* populations: Japan (*A. japonicus* x *A. nigricans, N* = 1); Cocos-Keeling, Australia (*A. leucosternon* x *A. nigricans, N* = 14); Christmas Island, Australia (*A. leucosternon* x *A. nigricans, N* = 25); Guam, Mariana Islands (*A. japonicus* x *A. nigricans, N* = 1); Kosrae, Caroline Islands (*A. achilles* x *A. nigricans*, *N* = 6); Marshall Islands (*A. achilles* x *A. nigricans*, *N* = 1). In addition to the main STRUCTURE plots, morphological hybrids are highlighted in the inset.

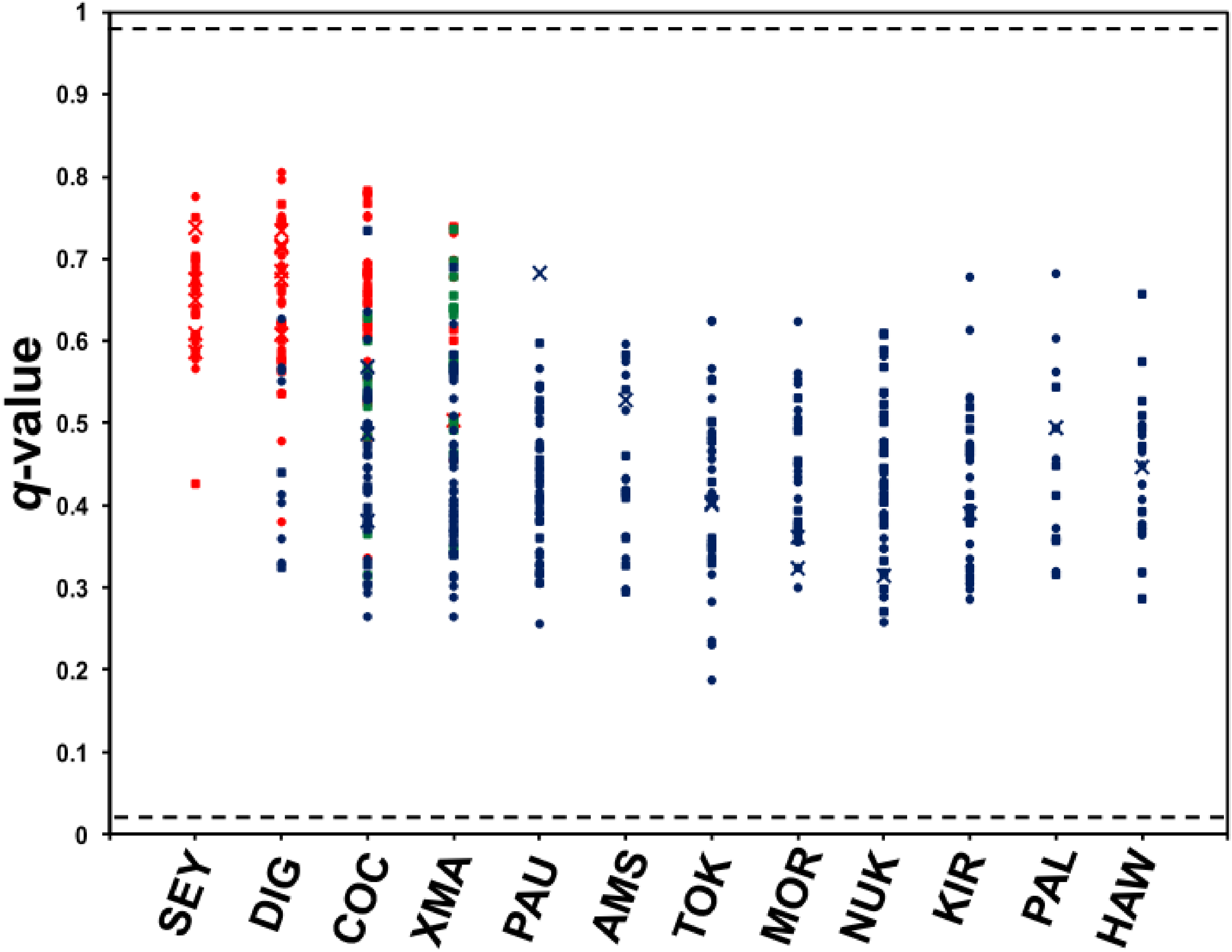
Figure 6. Scatterplot of DAPC performed on 10 microsatellite loci for the four species of surgeonfish (*Acanthurus achilles, A. japonicus, A. leucosternon* and *A. nigricans*) and their morphological hybrids. Species and their hybrids are shown by a color scheme consistent with Fig. 3 (*A. achilles*: pink; *A. japonicus*: dark green; *A. leucosternon*: light blue; *A. nigricans*:

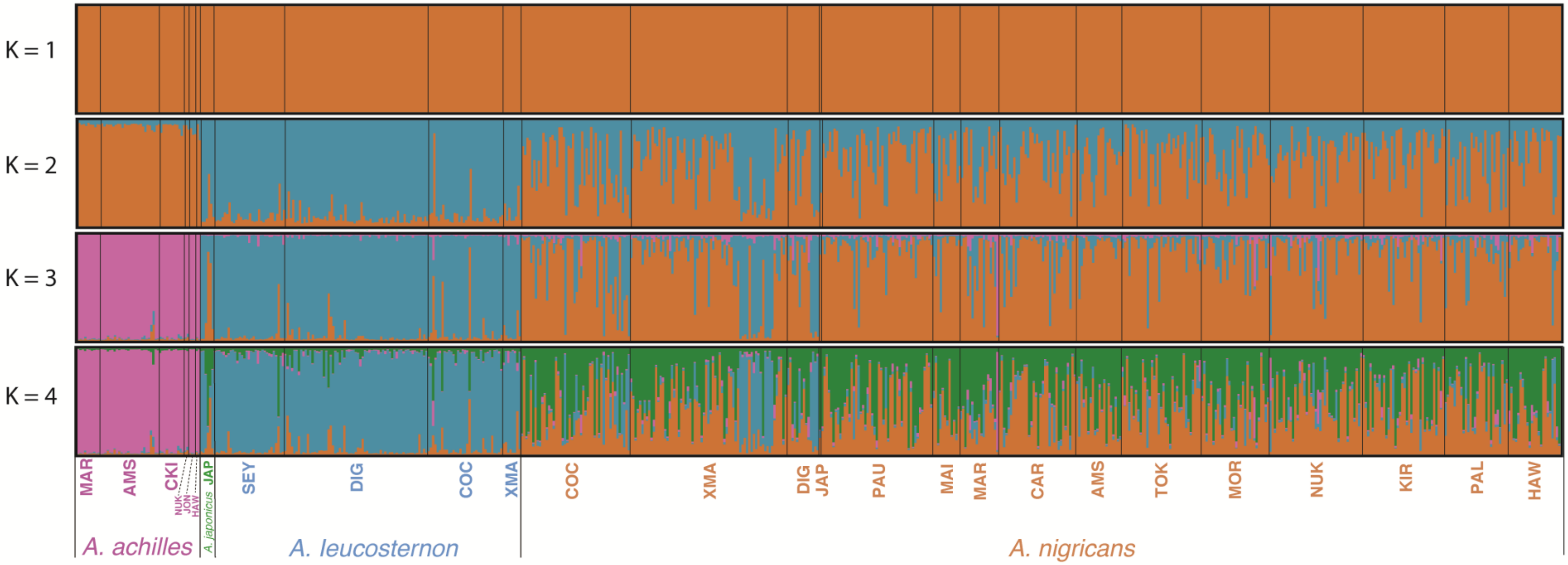
brown; *A. achilles* × *A. nigricans*: black; *A. japonicus* × *A. nigricans*: yellow; *A. leucosternon* × *A. nigricans*: red), 95% inertia ellipses and an optimal number of seven clusters based on BIC. Symbols represent individual genotypes indicated by the embedded key; axes show the first two discriminant functions.

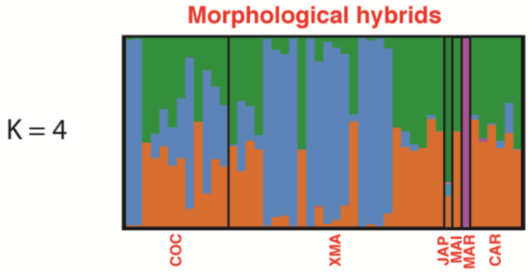

JAP
PAR
PHI
PAU
MAI
SOC
SEY
DIG
COC
XMA
CAR
MAR
HAW
JON
PAL
KIR
TOK
NUK
AMS
MOR
CKI
2000 km
A. achilles N = 54
A. japonicus N = 49
A. leucosternon N = 141
A. nigricans N = 412
Hybrids N = 48

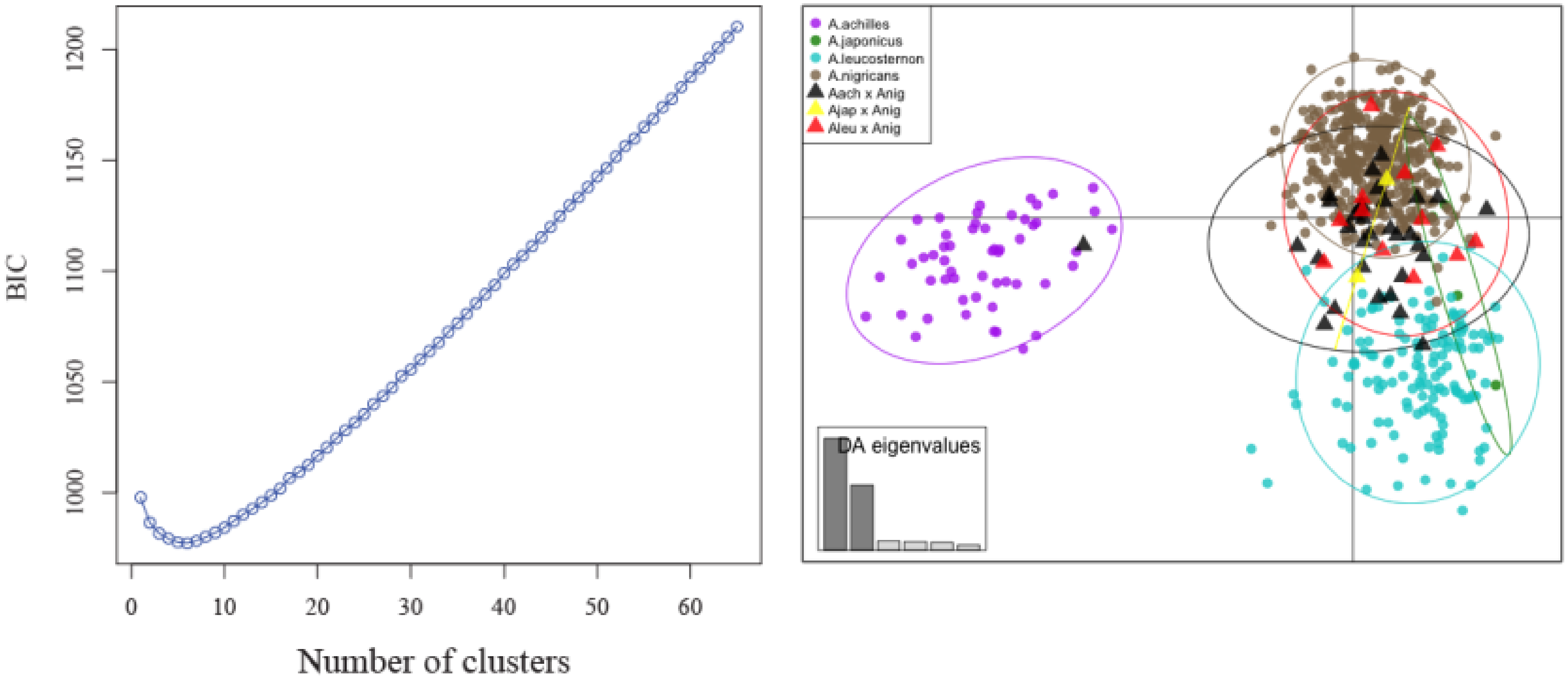

(a)
x
(b)
=
(c)
(d)
x
(e)
=
(f)
(g)
x
(h)
=
(i)


A. achilles
Marshall Islands (MAR)
American Samoa (AMS)
Cook Islands (CKI)
Nuku Hiva (NUK)
Johnston Atoll (JON)
Hawaiian Islands (HAW)
A. japonicus
Philippines (PHI)
Paracel Islands (PAR)
Japan (JAP)
A. leucosternon
Socotra, Yemen (SOC)
Republic of Seychelles (SEY)
Diego Garcia (DIG)
Cocos-Keeling Islands, Aus. (COC)
Christmas Island, Aus. (XMA)
A. nigricans
Diego Garcia (DIG)
Cocos-Keeling Islands, Aus. (COC)
Christmas Island, Aus. (XMA)
Republic of Palau (PAU)
Guam, Mariana Islands (MAI)
Marshall Islands (MAR)
Kosrae, Caroline Islands (CAR)
American Samoa (AMS)
Tokelau Islands (TOK)
Moorea (MOR)
Nuku Hiva (NUK)
Kiritimati, Republic of Kiribati (KIR)
Palmyra Atoll (PAL)
Hawaiian Islands (HAW)
12 mutations
Hybrids
A. achilles x A. nigricans
Marshall Islands (MAR)
Kosrae, Caroline Islands (CAR)
A. japonicus x A. nigricans
Japan (JAP)
Guam, Mariana Islands (MAI)
A. leucosternon x A. nigricans
Cocos-Keeling Islands, Aus. (COC)
Christmas Island, Aus. (XMA)

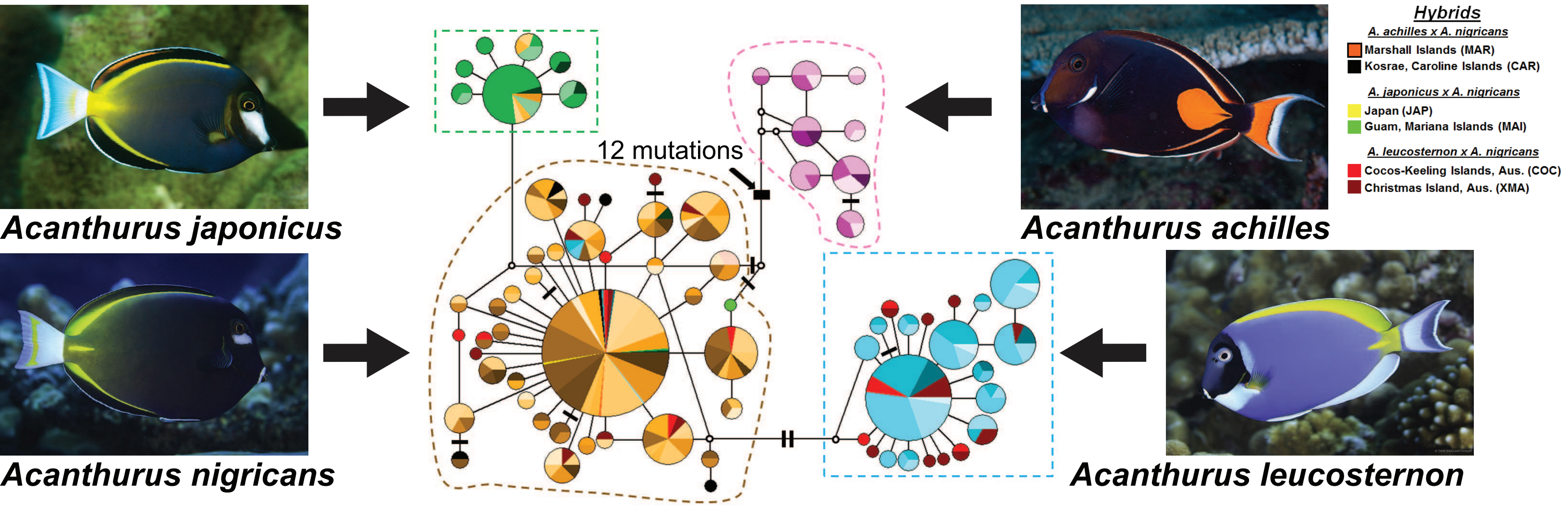

q-value
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SEY
DIG
COC
XMA
PAU
AMS
TOK
MOR
NUK
KIR
PAL
HAW


K = 1
K = 2
K = 3
K = 4
MAR
AMS
CKI
NUK
JON
HAW
JAP
SEY
DIG
COC
XMA
COC
XMA
DIG
JAP
PAU
MAI
MAR
CAR
AMS
TOK
MOR
NUK
KIR
PAL
HAW
A. achilles
A. japonicus
A. leucosternon
A. nigricans


Morphological hybrids
K = 4
COC
XMA
JAP
MAI
MAR
CAR


BIC
1000
1050
1100
1150
1200
0
10
20
30
40
50
60
Number of clusters
A.achilles
A.japonicus
A.leucosternon
A.nigricans
Aach x Anig
Ajap x Anig
Aleu x Anig
DA eigenvalues

Graphical Abstract



Highlights

- Evolutionary consequences of hybridization in a coral reef fish complex
- Genetic analyses of 645 individuals from four species of surgeonfish sampled across the Indo-Pacific
- Recent separation between all species of 235,000 to 2.25 million years
- Hybridization between all species at suture zones due to secondary contact
- Introgression levels vary among species producing different evolutionary outcomes