DR. CYNTHIA RIGINOS (Orcid ID : 0000-0002-5485-4197)
JOSHUA A. THIA (Orcid ID : 0000-0001-9084-0959)

Article type : Research Article

Global connections with some genomic differentiation between Indo-Pacific and Atlantic Ocean wahoo, a circumtropical large pelagic fish

Isabel Haro-Bilbao[1,2], Cynthia Riginos[1], John D. Baldwin[3], Mitchell Zischke[4,5], Ian R. Tibbetts[1], Joshua A. Thia[1,6,*]

[1] School of Biological Sciences, The University of Queensland, QLD, Australia

[2] Charles Darwin Research Station, Santa Cruz, Galapagos Islands, Ecuador

[3] Department of Biological Sciences, Florida Atlantic University, Davie Florida, USA

[4] Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, USA

[5] Illinois-Indiana Sea Grant, West Lafayette, IN, USA

[6] School of BioSciences, Bio21 Institute, The University of Melbourne, VIC, Australia

* Corresponding author: josh.thia@live.com, joshua.thia@unimelb.edu.au

Running title: Genetic structure in a pelagic fish

Key words: *Acanthocybium solandri,* population genomics, demographic inference, connectivity, population structure, fisheries management, wahoo

SIGNIFICANCE STATEMENT

Our study is the most comprehensive genetic investigation to date for the wahoo, *Acanthocybium solandri*, a pelagic fish with increasing importance to marine fisheries. Using population genomics approaches, we identify regional differentiation at the world's largest biogeographic scales, namely between the Indo-Pacific and Atlantic Oceans. Demographic analyses revealed there has been considerable gene flow within these ocean basins over evolutionary timescales. Our findings highlight how genomics can uncover subtle geographic differentiation in highly dispersive marine animals and provide new insights on appropriate wahoo stock definitions for fisheries management.

ABSTRACT

Aim

Globally distributed pelagic fishes are typified by very low to negligible genetic differentiation at oceanic scales arising from high gene flow and (or) large population sizes. Genomic approaches employing thousands of loci to characterise genetic variation can, however, illuminate subtle patterns of genetic structure and facilitate demographic inference, such that effects arising from gene flow and population size can be partially decoupled. We used a population genomics approach to identify putative stocks in a circumtropical pelagic fish, wahoo, and to assess global connectivity in this species.

Location

Indo-Pacific and Atlantic Oceans.

Taxon

Wahoo, *Acanthocybium solandri* (Cuvier, 1832)

Methods

Globally distributed wahoo samples from 11 locations (representing a total of 296 individuals) were sequenced using a pool-seq ezRAD approach to obtain 1,289–9,825 genome-wide SNP loci per population pair for analyses of genetic structure at MAF >0.05. Demographic inference using a diffusion approximation method (∂A∂I) was performed using 11,495–12,812 SNPs per population pair at a MAF >0.02.

Results

Genetic structure, measured as F_{ST}, was overall low indicating very little heterogeneity among sample pairs (pairwise F_{ST} ≤0.021). However, there was a clear signal of regional genetic structuring between ocean basins. A principal coordinate analysis separated samples from the Indo-Pacific with those from the Atlantic, and an analysis of molecular variance suggested that ~77% of variation in genetic structure was among regions. Our demographic analyses found greater support for models including migration over simple models of isolation.

Main conclusions

Our study provides the most thorough genetic investigation of wahoo to date. We provide evidence for global connectivity of wahoo populations over their evolutionary history, but we also provide the first indication of subtle regional structure between the Indo-Pacific and Atlantic Oceans, which occurs against a background of high gene flow. The identification of regional stocks will inform new management strategies and guide future investigations in wahoo, an increasingly important species in global fisheries.

INTRODUCTION

Large pelagic fishes are highly mobile and inhabit an ecosystem with few barriers to movement. Their ranges are extensive and, in the extreme, include worldwide pelagic waters. Historically, evidence of minimal genetic differentiation across global spatial scales (for example, between ocean basins) appears consistent with extensive dispersal, at least when using traditional genetic markers (mtDNA, microsatellites, and allozymes: reviewed by Hauser & Ward, 1998; Gaither, Bowen, Rocha, & Briggs, 2016). Genetic homogeneity of large pelagic fishes contrasts with frequent observations of geographic differentiation for coastal marine species (Lessios & Robertson, 2006; Rocha et al., 2007; Gaither & Rocha, 2013; Ludt & Rocha, 2014; Crandall et al., 2019), especially at the scale of ocean basins where landmasses, constrictions, and currents limit dispersal (such as the Isthmus of Panama, Sunda Shelf, Straits of Gibraltar, and Benguela Current), as do large expanses of deep waters (constituting the Eastern Pacific and Mid-Atlantic barriers). For high dispersal species such as pelagic fishes, the ability to query many loci using contemporary genomic methods, however, provides greater sensitivity to detect subtle population structure. Indeed, recent genomic surveys of circumtropical pelagic fishes have reported slight, but significant, genetic differentiation *within* ocean basins, for example, in yellowfin tuna (Grewe et al., 2015; Barth, Damerau, Matschiner, Jentoft, & Hanel, 2017; Mullins, McKeown, Sauer, & Shaw, 2018; Pecoraro et al., 2018), for albacore tuna (Laconcha et al., 2015; Vaux, Bohn, Hyde, O'Malley, 2021), for marlin (Mamoozadeh, Graves, & McDowell, 2019), and for dolphinfish (Maroso, Franch, Dalla Rovere, Arculeo, & Bargelloni, 2016). Genomic results comparing *between* ocean basins reinforce emergent results from earlier genetic studies: for circumtropical pelagic fishes, Mediterreanean populations can be very distinctive (for example, in albacore tuna: Laconcha et al., 2015), reflecting well noted biogeographic and phylogeographic transitions between the Atlantic and Mediterranean (Patarnello, Volckaert, & Castilho, 2007). Differentiation between Atlantic and Pacific Ocean populations can also be significant (albacore tuna: Laconcha et al., 2015; yellowfin tuna: Barth et al., 2017; Pecoraro et al., 2018), as can that between Indian and Pacific Ocean populations (albacore tuna: Laconcha et al., 2015; yellowfin tuna: Pecoraro et al., 2018; marlin: Mamoozadeh et al., 2019).

Although pelagic fishes inhabiting Atlantic and Indian Oceans can manifest patterns of divergence consistent with long-standing isolation, intriguingly, genomic investigations of yellowfin tuna have detected ongoing connections with substantial migration into Atlantic Ocean populations over evolutionary time frames (Barth et al., 2017) and among present day populations (Mullins et al., 2018). Occasional Atlantic–Indian Ocean connections likely arise from the complex oceanography of southern Africa. The Benguela Current on the Atlantic coast of Africa transports cold, upwelled waters northward along the southeastern Atlantic, and its low temperatures are lethal for most tropical species. However, the warm waters from the Angola (tropical east Atlantic) and Agulhas (western Indian Ocean) currents mix and generate the Agulhas Rings, a series of eddies that can be advected northward in the Benguela Current (Peeters, Acheson, Brummer, Wilhelmus, & et al., 2004; Hutchings et al., 2009). These warm-water filaments or edges breaking into the Benguela Current may allow the sporadic dispersal of some species between the Indo-Pacific and the Atlantic Oceans, which is reflected in the genetic similarity between Atlantic and Indian Ocean populations of *pelagic* marine species that can move as adults (reviewed by Gaither et al., 2016). Colonization of the Atlantic by Indo-Pacific *benthic* marine species, however, has also been inferred, for example in a coral goby (Rocha et al., 2005).

In addition to uncovering subtle genetic differentiation, population genomic data can also be used for demographic inference to infer relative contributions of restricted gene flow and divergence times to genetic differentiation (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009), as exemplified by Barth et al. (2017) with yellowfin tuna. Though the high dispersal capacity of large pelagic fishes suggests that migration might maintain genetic homogeneity between ocean basins, such a genetic pattern can also be generated from a large effective population size. This is because large effective population sizes resist the effects of genetic drift and such a process is common in marine species (Waples, 1998). Therefore, genetic homogeneity does not necessarily implicate high dispersal. Moreover, simple patterns of genetic differentiation preclude assessment of possible asymmetries in migration: populations contributing more to migration could be considered more important in maintaining global connectivity. Therefore, demographic inference not only provides opportunity to assess whether populations are connected by migration, but also the nature of these connections.

In this study, we focus on the wahoo, *Acanthocybium solandri* (Cuvier, 1832), as an exemplar of circumtropical pelagic fishes. Wahoo are members of the Scombridae, a family of large predatory fishes that are highly valued for commercial and sport fishing. Wahoo are strong swimmers and globally distributed in epipelagic tropical and temperate waters. These fish are primarily caught as a byproduct in commercial tuna, swordfish, and dolphinfish fisheries and are also targeted in artisanal, subsistence and recreational fisheries worldwide (Collette & Nauen, 1983; Luckhurst & Trott, 2000; Zischke, 2012). Over the past ten years, the average annual global landings for wahoo in commercial fisheries have been ~4,500 tonnes, which represents an increase of 40% compared to the catches reported in the preceding decade (FAO, 2019). The increasing importance of wahoo as a fishery, but scant information on its population genetic structure and biogeography, warrants increased investigation into patterns of population connectivity at a global scale (Oxenford, Murray, & Luckhurst, 2003; Luckhurst, 2007; Theisen, Bowen, Lanier, & Baldwin, 2008; Zischke, 2012).

The natural history of wahoo sets an *a priori* expectation for low genetic differentiation at broad spatial scales. Wahoo are hypothesized to spawn in proximity to major oceanic surface currents that facilitate the dispersal of their buoyant eggs and pelagic larvae (Brown-Peterson, Franks, & Burke, 2000; Wollam, 1969), which potentially enhances their dispersal capacity (Jenkins & McBride, 2009; Zischke, Farley, Griffiths, & Tibbetts, 2013). Spawning probably occurs during the warmer summer months when individuals are diffusely distributed, and adult wahoo are not known to form large spawning aggregations (Jenkins & McBride, 2009; Zischke et al., 2013). Tagging studies have revealed that adults can swim at high speeds (>77 km h^{-1}) (Walters & Fierstine, 1964) and can rapidly traverse long distances. For instance, Theisen & Baldwin (2012) described an individual in the Atlantic Ocean that traveled 1,960 km in 30 days. In the Pacific Ocean, an adult tagged in the vicinity of Hawaii was recaptured after ~200 days near Kiribati, more than 2,500 km away (NMFS, 1999). Consequently, wahoo have high dispersal capacity at all life stages, which should favour the homogenisation of genetic variation worldwide.

We use a population genomics approach to re-examine the patterns of genetic differentiation in wahoo and test competing demographic hypotheses that might explain the distribution of present-day genetic variation. Previous genetic studies support the concept that wahoo exist as a single globally homogeneous population. Comparisons between the Atlantic and Pacific using the mtDNA control region (Garber, Tringali, & Franks, 2005), and broad global assessments using the mtDNA *cytb* and nuclear *ldhA6* sequences (Theisen et al., 2008), have failed to detect any genetic differentiation. Using thousands of SNP loci, we observed a signal of weak genetic differentiation and regional structuring between the Indo-Pacific and the Atlantic Oceans. Despite this regional structuring, demographic models suggest that ongoing gene flow occurs between ocean basins. Collectively, our study identifies potential wahoo management units at the level of ocean basins, with the considerations that these management units are likely connected by migration.

METHODS

Collection of samples

Wahoo were sampled at 11 localities throughout their global distribution between 1998 and 2015 from recreational and artisanal commercial fisheries: American Samoa, Bimini, Christmas Island, Eastern Australia, Galapagos, Grand Cayman, Hawaii, North Carolina, Palau, Thailand, and Trinidad & Tobago (Figure 1a).

DNA extraction and pooled RAD-seq libraries preparation

Total genomic DNA was extracted from individual tissue samples (muscle, gill, or fin) using the E-Z 96 Tissue DNA Kit (Omega Bio-tek). The DNA extractions were visualized on 1% agarose gels to assess quality and quantity. Only samples that showed no signs of degradation were used for reduced representation library construction, as evidenced by high molecular weight bands and an absence of sample smearing on the gel. The final set consisted of 296 samples: American Samoa (AmSam, *n* = 30); Bimini (*n* = 30); Christmas Island (ChrIsl, *n* = 24); Eastern Australia (EAus, *n* = 30); Galapagos (Gal, *n* = 30); Grand Cayman (GrandCay, *n* = 30); Hawaii (*n* = 14); North Carolina (NCar, *n* = 30); Palau (*n*= 30); Thailand (Thai, *n*= 18); and Trinidad & Tobago (TrinTab, *n*= 30) (Figure 1a; Table S1).

We used the ezRAD method (Toonen et al., 2013) and a pool-seq approach to sample genomic variation. When DNA pool sizes are large (≥ 30 diploids) allele frequency estimates show strong correlations to true population values (Futschik & Schlötterer, 2010; Gautier et al., 2013; Schlötterer, Tobler, Kofler, & Nolte, 2014; Hivert, Leblois, Petit, Gautier, & Vitalis, 2018). Hence, pool-seq was an attractive and preferred method for us to obtain genetic information (allele frequencies) from many individuals and sampling locations. We improved accuracy in allele frequency estimation by creating two replicate pooled ezRAD libraries per sample location. Replicate library preparation affords estimation of error from genetic sampling and technical artifacts (Gautier et al., 2013). For each library, 500 ng DNA was pooled per sampling location, with equimolar amounts per contributing fish. The same individuals went into each replicate library. Double digestion was performed with the MboI and Sau3AI enzymes (GATC). Library construction and sequencing were outsourced to Texas A&M University-Corpus Christi Genomics Core Lab (TAMUCC; http://genomics.tamucc.edu/). A total of 22 libraries (11 localities with 2 replicates) were sequenced on a single lane of Illumina HiSeq 4000 (150 bp paired-ends).

De novo assembly and data processing

Raw demultiplexed sequencing reads (paired-end) were screened using the program FastQ Screen (Andrews, 2011) with the aligner software BWA (Li, 2013) to assess the presence of potential contaminants. A reference repository was generated containing genomic sequences downloaded from NCBI: yeast, *Saccharomyces cerevisiae* (GCF_000146045.2); fruit fly, *Drosophila serrata* (GCF_002093755.1); *E. coli*, *Escherichia coli* (GCF_000005845.2); human, *Homo sapiens* (GCF_000001405.37); mouse, *Mus musculus* (GCF_000001635.25). Illumina adapters were also screened. Reads were only kept for downstream analyses if they did not align to any of these reference genomes. Since FastQ Screen can produce a desynchronization of paired read files, forward and reverse read files were resynchronized using PAIRFQ_LITE.PL script (Staton, 2013). Only the reads that were correctly paired were used for further analysis. The restriction enzyme cut site was removed from the reads using the program SEQTK (Li, 2016) by trimming the first four bases in every read.

The main *de novo* assembly of RAD contigs was performed using the DDOCENT V2.2 pipelines (Puritz, Hollenbeck, & Gold, 2014). Briefly, PEAR (Zhang, Kobert, Flouri, & Stamatakis, 2014) removes reads that overlap. TRIMMOMATIC (Bolger, Lohse, & Usadel, 2014) is used to trim low-quality bases (Phred score < 20) from the extremes of each read, followed by a 5 bp sliding window approach to remove bases when the average Phred score < 10. Trimmed and untrimmed reads are used for different steps of the assembly. Contigs are assembled from untrimmed reads using a combination of the RAINBOW algorithm (Chong, Ruan, & Wu, 2012) and CD-HIT (Li & Godzik, 2006). Trimmed reads are then aligned to the contigs using the BWA MEM algorithm (Heng, Ruan, & Durbin, 2008) and SNPs are called with FREEBAYES (Garrison & Marth, 2012).

In our assembly, we implemented the DDOCENT in a series of discrete stages. First, we trimmed reads. We then utilised the DDOCENT script, REFMAPOPT.SH, to assemble contigs outside the main pipeline. Since two replicates from each ezRAD library were available, the replicate with the largest file size (from each pooled locality) was selected to generate the reference assembly, with a clustering threshold of 98%. This reference was then incorporated back into the DDOCENT pipeline for read mapping and variant calling. We used BWA MEM parameters: –A 1 (match score), –B 3 (mismatch penalty), –O 20 (gap penalty), –E 10 (gap extension penalty) –U 20 (unpaired penalty). The gap open penalty was set high to avoid alignments with many gaps. The DDOCENT script was edited to include the gap extension penalty (to further prevent large gaps in assembly) and the unpaired penalty (to reduce the splitting of reads across different contigs). The raw FREEBAYES SNPs were then treated with an initial filtering using VCFTOOLS (Danecek et al., 2011), which removed indels and SNPs with more than two alleles, required a minimum mean depth of one read, a mapping quality of 30, and no missing data.

SNPs derived from mitochondrial contigs (non-nuclear) were removed. The wahoo mitogenome (Accession AP012945) was downloaded from NCBI. RAD contigs were split into their forward and reverse sequences if they were scaffolded (joined by a series of “NNNNNNNNNN”).

Separated forward and reverse RAD contig ends, and contiguous RAD contigs, were mapped against the wahoo mitogenome using BOWTIE2 (Langmead & Salzberg, 2012) extracted using SAMTOOLS. Any SNPs derived from these identified mitochondrial RAD contigs were removed prior to analyses of geographic differentiation or demographic inference.

Analytical approaches: F_{ST} and allele frequencies

Using our pool-seq allele counts, we estimated geographic differentiation and conducted demographic inference analyses to determine the possible demographic scenarios that might have led to the observed distribution of genetic variation in wahoo. Our pool-seq allele counts were treated in slightly different ways for each of these respective analyses. To estimate population genetic structure, we combined reads across replicate libraries and estimated genetic structure (F_{ST}) using R's 'poolfstat' package (Hivert et al., 2018). For demographic inference analyses, we imputed allele frequencies (p) using POOLNE_ESTIM (Gautier et al., 2013), which leverages information across replicate libraries. Hence, although the total read counts were the same, the analyses differed regarding whether the replicate library information was considered in the estimation of summary statistics (F_{ST} or p). We elaborate on these differences below.

Geographic differentiation

We estimated geographic differentiation using R's 'poolfstat' package. This package does not accommodate replicate library information but has been optimised to provide highly robust estimates of F_{ST} from pool-seq data using an ANOVA framework (Hivert et al., 2018). We calculated F_{ST} for all geographic sample pairs and estimated genetic structure using two sets of SNPs. Firstly, we identified a *sample pair specific* SNP set, whereby loci were filtered for each sample pair independently, with the assumption that these loci are randomly sampled from across the genome and are all drawn from the same F_{ST} distribution. Secondly, we identified a *shared* SNP set, whereby loci were filtered with respect to all sampling locations. The *sample pair specific* SNP set allowed us to obtain a greater number of loci per sample pair, 1,289 ≤ n ≤ 9,825 loci, which is useful for identifying subtle patterns of genetic differentiation (Table S2).

The *shared* SNP set allowed us to evaluate how non-overlapping loci in the *sample pair specific* SNP set might have influenced our interpretation of geographic differentiation, using *n* = 945 loci. We summed reads across replicate libraries and randomly sampled one SNP locus per RAD contig. For both SNP sets, loci were kept if the read depth (between sample pairs or across all samples) was 50 ≤ total reads ≤ 1,000 and if they had minor allele frequency (MAF) of 0.05. Note, the 99% percentile for read depth at a locus was 806, with a maximum 14,432 reads. Our minimum and maximum read depth requirements were chosen to balance good coverage with exclusion of loci with unusually high read depth.

F_{ST} was calculated in 'poolfstat' using the "Anova" method using POOLFSTAT_DT() function from R's 'genomalicious' (Thia & Riginos, 2019), which is a wrapper function for POOLFSTAT::COMPUTEFST(). For the *sample pair specific* SNP set, we calculated the mean multilocus F_{ST} in two ways: (i) the empirically observed value using all loci available per sample pair; and (ii) the bootstrapped mean F_{ST} using 1,000 bootstrap replicates of *n* = 1,000 randomly drawn loci. The bootstrapping procedure allowed us to compare sample pairs for equal number of loci in the sample specific SNP sets; we found that the correlation between the empirical (all loci) and the mean bootstrapped F_{ST} was high (*r* > 0.99) for the sample specific SNP set, so we proceeded with further analyses using the mean bootstrapped F_{ST}. This bootstrapping procedure also facilitated comparison to the *shared* SNP set, where *n* = 945 loci across all geographic locations, by virtue of a similar number of loci analysed. In the *shared* SNP set, F_{ST} was calculated as the empirical value (using all loci, no bootstrapping). We did not remove outlier loci prior to F_{ST} calculation because we were interested in genome-wide patterns of geographic differentiation.

We further interrogated patterns of geographic differentiation using a principal coordinate analysis (PCoA) for visualisation of spatial genetic relationships, and an analysis of molecular variance (AMOVA) (Excoffier, Smouse, & Quattro, 1992) to partition the variance at different spatial scales. For both SNP sets, we generated an F_{ST} distance matrix between sample pairs and used the PCOA() function from R's 'ape' package (Paradis & Schliep, 2018) to conduct PCoAs.

We used the 'pegas' package (Paradis, 2010) to conduct AMOVAs in R, fitting a model where F_{ST} was predicted by basins (Indian, Pacific, or Atlantic) nested within regions (Indo-Pacific, or Atlantic). Because negative F_{ST}-values are not interpretable and are effectively "zero", we converted all negative F_{ST} estimates to zero prior to PCoA and AMOVA.

Demographic inference analyses

Marine organisms are characterised by their large population sizes and very high dispersal potential (Palumbi, 1994). Because of these characteristics, low genetic structure could be the product of two different processes: (1) large effective population sizes that resist drift and maintain allelic diversity and low genetic differentiation in the absence of high gene flow; or (2) high gene flow that facilitates genetic homogenisation among populations. To understand which processes might affect the global distribution of genetic variation in wahoo, we attempted to disentangle the relative likelihood of three demographic scenarios: isolation, symmetric migration, or asymmetric migration.

First, we imputed sample location allele frequencies (*p*) using the program POOLNE_ESTIM (Gautier et al., 2013). The algorithm in POOLNE_ESTIM leverages information across replicate libraries to estimate the potential error associated with allele frequency imputation resulting from pooling (unequal contributions due to technical and sequencing biases): this allows estimation of the effective diploid number (n_e) for a pool-seq experiment (Table S1). Imputed allele frequencies were used to estimate the site frequency spectrum (SFS) for demographic analyses. Loci were again filtered for read depth of 50 ≤ total reads ≤ 1,000. However, we chose a MAF of 0.02 to prevent dropout of rare alleles, which would bias the SFS estimation (Matz, 2018; Linck & Battey, 2019). Our demographic analyses focused on sample locations where 30 diploid individuals were collected. Therefore, in a sample of 60 haploid chromosomes, a MAF of 0.02 equates to ~1 minor allele.

Sampling locations were also chosen for demographic analyses based on read depth. Read depth was consistently right-skewed across samples (low frequency of high coverage sites), but

Indo-Pacific pooled samples received less coverage than those from the Atlantic. For the Indo-Pacific locations, American Samoa and the Galapagos had the highest coverage (Figure S1). Among the Atlantic sampling sites, North Carolina and Trinidad & Tobago received less coverage than Bimini and Grand Caiman (Figure S2), exhibiting a read depth distribution more similar to those from the Indo-Pacific. We therefore selected American Samoa and the Galapagos to represent the Indo-Pacific, and North Carolina and Trinidad & Tobago to represent the Atlantic. For each of these geographic location samples, we used the DADI_INPUTS_POOLS() function in R's 'genomalicious' to create allele counts in the ∂A∂I input format based on the imputed population allele frequencies: for a given number of pooled diploids, the SFS was calculated as the expected number of reference and alternate alleles in the haploid sample (60 unique chromosomes for 30 diploids). Demographic scenarios were assessed for sampled locations within each ocean basin (American Samoa/Galapagos and North Carolina/Trinidad & Tobago), and for all sample pairs between ocean basins (American Samoa/North Carolina, American Samoa/Trinidad & Tobago, Galapagos/North Carolina, and Galapagos/Trinidad & Tobago). The number of SNPs available for each sample pair and their median read depth statistics are tabulated in Table S3. On average, 11,979 loci were available for pairwise analyses of demography, with a range of 11,495–12,812 loci, and an average median depth of 139.5 reads per locus, with a range of 110–167 median reads per locus. The SFS for each sample pair was then projected down to 10-by-10 alleles and was folded prior to performing demographic simulations using the ∂A∂I python function, DADI.SPECTRUM.FROM_DATA_DICT().

Support for our three demographic scenarios (isolation, symmetric migration, and asymmetric migration) was assessed using the demographic simulator, ∂A∂I (Gutenkunst et al., 2009) (Figure 2). In the isolation scenario, it was assumed that populations have diverged from some ancestral population in the past and have not exchanged genes since divergence. In the symmetric scenario, it was assumed that gene flow has occurred since divergence and the rate of gene exchange is equal between populations. In contrast, the asymmetric migration scenario assumed that gene exchange is biased in one direction. If N_A is the ancestral population size, N_i

the contemporary size for population *i*, the key parameters being estimated by our models are as follows: *T* ("T"), the effective number of generations since divergence, with *t* number of generations (= t / $2N_{\mathrm{A}}$); v_i ("nu"), the relative contemporary population size parameter for population *i* (=N_i/ N_{A}); and M_{ij}, the scaled migration rate from population *j* into population *i*, relative to m_{ij}, the proportion of chromosomes that move between populations per generation (= $2N_{\mathrm{A}}m_{\mathrm{ij}}$). Hence, M_{ij} describes gene flow in numerical terms, that is, the total number of chromosomes that migrate. When symmetric migration was modelled, it was assumed that the scale parameter was identical for both directions of gene flow ("M"), whereas in the asymmetric scenario we modelled the movement of genes from population 1 into 2 ("M21") and from population 2 into 1 ("M12").

Estimating demographic parameters in ∂A∂I requires running multiple replicate simulations to explore parameter space and check for model convergence (Gutenkunst et al., 2009; Rougemont et al., 2017; Rougeux, Gagnaire, & Bernatchez, 2019). We ran 100 simulations per scenario, per sample pair, each with 100 optimisation iterations. Based on preliminary tests, we set the respective initial parameter values in the isolation scenario as T = nu1 = nu2 = 0.05 with upper bounds of 5 and lower bounds of 1e−10. For the symmetric and asymmetric scenarios, initial parameter values were T = nu1 = nu2 = m or m12 = m21 = 5, with upper bounds of 100 and lower bounds of 1e−3. For each simulation, we perturbed the initial parameters by a magnitude of three and used the linear extrapolation function in ∂A∂I. Once the simulations had finished running, we summarised the parameter estimates and log-likelihoods for the top 10 models.

For each sample pair, the AIC for the best model in each scenario was calculated as AIC = $2k - 2\ln(L)$, where *k* was the number of estimated parameters, and ln(*L*) was the log-likelihood. To test for differential support among scenarios for each sample pair, we calculated ΔAIC scores as $\Delta\mathrm{AIC} = \mathrm{AIC}_S - \mathrm{AIC}_B$, where subscript *S* indicates the focal scenario and subscript *B* the scenario with the best AIC. A focal scenario was considered substantially different from the best scenario when ΔAIC > 10 (Burnham & Anderson, 2002).

RESULTS

RAD-seq assembly

After bioinformatic processing, the number of reads for any single library ranged from 196,568 to 20,280,396. The total number of reads for any sampled location ranged from 5,350,856 to 32,670,537, with sample means ranging from 2,680,428 to 16,335,269 reads across both replicate libraries.

The percentage similarity between the effective number of pooled diploids (n_e) versus the true number of pooled diploids (n) ranged from 2.9% to 100%. The mean percentage similarity between n_e and n was 71.32% for any one library preparation. Of the 24 libraries, 13 had an n_e similar to n (> 85%). Values are tabulated in Table S1.

For each sample pair, after applying MAF and depth filters, the number of loci used for analyses of geographic differentiation ranged from 1,298–9,825 loci in the sample *pair specific* SNP set (average of 5,905 loci) (Tables S2). The *shared* SNP set contained 945 loci that met our MAF (>0.05) and depth filters in all samples. For demographic analyses, 11,495–12,812 SNP loci were available following depth and MAF (>0.02) filtering (Table S3).

Geographic differentiation

Using our *population pair specific* SNP set (1,289 ≤ n ≤ 9,825 loci), mean bootstrap F_{ST} estimates showed non-zero genetic differentiation ($F_{ST} > 0$) between almost all sample pairs according to bootstrap 2.5% and 97.5% percentiles (Tables 1 & S2). The exceptions being the Christmas Island/Galapagos pair ($F_{ST} = 0.002$), East Australia/Thailand ($F_{ST} = 0.001$) and Palau/Thailand ($F_{ST} = 0.003$), where estimates of genetic differentiation were not statistically different from zero. Overall, however, values of mean bootstrap F_{ST} were low, with a range of 0.001–0.021, and a mean of 0.010, suggesting that despite the presence of genetic structure, there was very little variation among geographic locations.

Despite apparent low genetic structure across large geographic extents, a pattern of regional geographic differentiation emerged. Comparisons of F_{ST} distributions between ocean basins indicated that genetic structure was lower within the Indo-Pacific and Atlantic than between these oceanic regions (Figure 1b). This regional structuring was further evidenced in a PCoA (Figure 1c), where the first axis (capturing 53% of the variance in pairwise F_{ST}-values) separated wahoo samples from the Atlantic from those in the Indo-Pacific. The second PCo axis (capturing 14.8% of the variance in pairwise F_{ST}-values) was associated with variance within regions. Finally, AMOVA (Table 2) suggested that 77.41% of genetic variation was among regions, which was statistically significant ($p < 0.001$), whereas only 0.82% of genetic variation was among basins nested within regions and this was statistically non-significant ($p = 0.695$).

Measures of genetic differentiation from our *shared* SNP set ($n = 945$ loci) provided a different picture of wahoo geographic differentiation. Using this set of loci, many more sample pairs exhibited F_{ST}-values that were non-significantly different from zero (26/55 sample pairs) with values ranging from −0.009–0.014 (Figure S3a). In contrast to the sample pair specific SNP set, there was no evidence of regional structuring when considering the smaller *shared* SNP set. There was no clear separation of the Atlantic and Indo-Pacific in a PCoA of pairwise F_{ST}-values (Figure S3b). An AMOVA found a non-significant effect of region, despite 57.43% of variation being attributed to region ($p = 0.335$), and there was also a non-significant effect of basin nested in region ($p = 0.278$), which explained 16.53% of the variation (Table 2).

Our analyses of genetic structure using the two different SNP sets highlight the subtlety in regional differences among wahoo populations. All loci in the *shared* SNP set were in the *sample pair specific* SNP sets, yet they did not capture the regional signal between the Atlantic and the Indo-Pacific. Hence, despite being drawn from the same F_{ST} distribution, it appears there were too few loci in the *shared* SNP set to adequately sample variation in F_{ST} across loci, and that this set of 945 loci on average sat in the lower range of F_{ST}-values. We note that the *sample pair specific* SNP sets used an equivalent number of loci to estimate the mean bootstrapped F_{ST} (n = 1,000 randomly subsampled loci per bootstrap replicate, without replacement) to the *shared* SNP set, indicating the number of genetic markers used to perform F_{ST} calculations cannot explain differences between these SNP sets. Instead, more loci available in the *sample pair specific* SNP sets likely better characterised the F_{ST} distribution. The correlation between the number of loci and the mean bootstrapped F_{ST} was r = 0.32, indicating a moderately positive effect of a larger SNP set in capturing greater signals of genomic divergence within a sample pair. Additionally, the correlation between the number of loci and the bootstrapped interval width (2.5% and 97.5% percentiles) was r = 0.60, indicating there was greater variation in mean bootstrapped F_{ST}-values with larger SNP sets when subsampling for 1,000 loci per bootstrap. Collectively, the larger number of loci obtained when SNPs were filtered by sample pair facilitated identification of regional structuring among sampled locations through greater genomic sampling.

Demographic inference analyses

In all sample pairs considered, the allele frequency correlations were very high, r > 0.88 (Figure S4), irrespective of whether pairs were between or within ocean basins. Using the demographic approximation method, ∂A∂I, we attempted to assess whether such highly correlated allele frequencies were better explained by large population sizes and (or) high migration rates, through estimation of effective divergence time (T), relative contemporary population sizes (nu1 and nu2), and migration rates (M, or M12 and M21). Log-likelihoods for the best models and their associated parameter sets are summarised in Figures S5 and Figures S6–S8, respectively.

When considering the best model for each scenario, we observed high concordance between simulated SFSs and the observed SFS estimated from pool-seq ezRAD data (e.g., Figure 3; see also Figures S9–S14). This indicated that the parameter combinations estimated in the best models provided reasonable reconstruction of the observed data. We do note that our models tended to overestimate the number of joint allele counts at the lower ends of the SFSs, for example, 1:1 or 1:2 or 1:3 for sample 1:sample 2, or vice versa. In other words, rare alleles were less frequent in our observed data, which might be partially attributable to our SFS estimation from imputed allele frequencies. However, most joint counts between simulated and observed SFS for any scenarios, in any sample pair, were similar, resulting in residual distributions centered on zero (with a few large outliers due to the overestimation of rare alleles in the simulated SFSs).

Overall, our demographic analyses suggest that a scenario including migration among wahoo populations describes patterns of genetic variation better than a scenario of isolation without migration. Across the top 10 simulations per scenario, convergence on similar log-likelihoods was greater for the isolation scenario, whereas greater variance in log-likelihoods was exhibited for the symmetric and asymmetric migration scenarios (Figure S5). Nonetheless, the best isolation scenario models were less likely and had worse AIC scores relative to those that included migration (Table 3; Figure S5). Indeed, to explain contemporary patterns of genetic variation under the isolation scenario, wahoo populations would have had to diverge very recently and be of a much smaller size, relative to their shared ancestral population—as indicated by very small values of T, nu1, and nu2 (approaching zero) (Table 3; Figure S6).

Symmetric migration was the most likely scenario for the American Samoa/North Carolina and Galapagos/North Carolina pairs (between Indo-Pacific and Atlantic) (Table 3). Note, however, that despite symmetric migration being the best scenario for the Galapagos/North Carolina pair, this was equivalent to the asymmetric scenario, based on ΔAIC < 10 (Table 3). Similarly, although not the best scenario for the North Carolina/Trinidad & Tobago pair, symmetric migration was nearly equivalent to the asymmetric scenario, AIC = 11 (see below). The best symmetric migration scenarios were characterised by more recent divergences between contemporary populations (T < 1), smaller but similar contemporary population sizes relative to the ancestral population (nu1 and nu2 < 1, and nu1 ≈ nu2), and considerable movement of genes between populations (M > 50) (Table 3). Across the top 10 models, although the exact value of M was variable among simulations, there was a general trend of large M, small T, and small nu1 and nu2 (Figure S7).

The sample location pairs where asymmetric migration was the best scenario were American Samoa/Galapagos (within the Indo-Pacific), American Samoa/Trinidad & Tobago and Galapagos/Trinidad & Tobago (between the Indo-Pacific and Atlantic), and North Carolina/Trinidad & Tobago (within the Atlantic). However, based on ΔAIC, the symmetric migration scenario had an equivalent (or nearly equivalent) likelihood to asymmetric migration scenario for the Galapagos/North Carolina and North Carolina/Trinidad & Tobago pairs (Table 3). Asymmetric migration models exhibited greater variability in parameter combinations, relative to the other scenarios (Figure S8). Qualitatively, there were two sets of parameters that emerged across the top 10 models in the asymmetric migration scenario: (1) more ancient divergence times with smaller migration rates, and (2) more recent divergence times with larger migration rates. For all sample pairs, the best asymmetric model was one where scaled divergence time was small ($T < 1$), contemporary effective population sizes were small (nu1 and nu2 < 1), and scaled migration rates were large (M12 and M21 > 30).

Based on our demographic inference analyses, we can be relatively confident that large population size alone, in the absence of gene flow, is not a major mechanism for low geographic structure in wahoo. However, there were no clear or consistent patterns with respect to the directionality of gene flow among oceanic regions. Comparisons between samples from the Indo-Pacific versus the Atlantic yielded a mix of symmetric models being the most likely, asymmetric models being the most likely, or both migration models being indistinguishable (Table 3). We can therefore only conclude that migration has played an important role in maintaining shared genetic variation between the Indo-Pacific and Atlantic.

DISCUSSION

Subtle regional genetic differentiation and challenges to demographic inference for a high gene flow pelagic fish

Large circumtropical fishes are typified by minimal genetic differentiation over large geographic distances, including between ocean basins. Our study is the first to use genome-wide SNP data to assess global genetic patterns in wahoo. Prior investigations using single loci have been unable to discern putative boundaries (mtDNA and nuclear LDH: Garber et al., 2005; Theisen et al., 2008). Yet here, using 1000s of genome-wide SNPs, we recovered a regional signal that separated wahoo from the Indo-Pacific with those from the Atlantic Ocean (Figure 1b & 1c; Table 2). This regional structuring was, however, weak, as evidenced by very low F_{ST} (≤ 0.021, Table 1) and highly correlated allele frequencies (Figure S4).

Differentiation between the Atlantic and Indo-Pacific Ocean wahoo populations in our study conforms to phylogeographic observations from other pelagic species exhibiting inter-oceanic genetic structure (reviewed by Hauser & Ward, 1998; Theisen et al., 2008; Gaither et al., 2016), especially yellowfin tuna (Barth et al., 2017; Mullins et al., 2018). Weak-but-significant genetic structuring can result either from recent divergence, high migration rates, or a combination of the two processes, and both processes are commonly observed in marine fisheries species (Waples, 1998). When considering results of our ∂A∂I demographic analyses across all scenarios and sample pairs (Table 3), isolation models performed worse than those including migration. The large scaled migration parameters observed (M >50, M12 and M21 > 50) indicates that substantial gene flow has occurred in the evolution history of wahoo.

Based on oceanography, it is expected that migration would occur from the Indo-Pacific into the Atlantic via advection of the warm water off the southern African coast, the Agulhas Rings, by the Benguela Current (Peeters et al., 2004; Hutchings et al., 2009). Contrastingly, migration around the southern tip of South America is a potentially unlikely route of connection due to the consistently cold sea surface temperatures (5–10°C). With respect to cross-Pacific movement through the East Pacific Barrier, asymmetric migration scenario was the most likely model for the American Samoa/Galapagos. Based on final parameters, greater dispersal was inferred from American Samoa into the Galapagos (M21 = 98.91) versus the reverse direction (M12 = 73.94). However, the high parameter values of M12 and M21 for the American

Samoa/Galapagos sample pair indicate that migration between the east and west Pacific has been extensive through time in both directions (Table 3). Regardless of the direction of movement, the deep waters of the East Pacific clearly do not constitute a barrier to movement in wahoo.

One caveat in our analyses of geographic differentiation is that our F_{ST} genetic distance matrix used for PCoA and AMOVA is derived from a non-overlapping set of loci in the *sample pair specific* SNP set. The AMOVA framework partitions variance in a genetic distance matrix with respect to hierarchical population structure and was originally formulated with respect to haplotypes sampled in all samples and populations (Excoffier et al., 1992). By virtue that the response variable in an AMOVA is a distance metric, we were able to overcome missing data limits and obtain many more genetic markers for estimating FST in our *sample pair specific* SNP set (1,289–9,825 loci) relative to the *shared* SNP set (945 loci). Hence, we acknowledge that our measures of genetic distance using the *sample pair specific* SNP set are not directly comparable among sample pairs. Yet as evident from our parallel analyses of both datasets, the greater number of loci in the *sample pair specific* SNP set allowed us to better sample the F_{ST} distribution across geographically distributed wahoo, whereas the *shared* SNP set was underpowered to capture regional structuring (Figures 1c versus Figure S3c; Table 2a versus Table 2b). Therefore, we believe our approach is justified and has allowed us to observe novel population genetic patterns in wahoo that are in line with general biogeographic expectations.

An additional caveat pertains interpretation of gene flow directionality in our demographic analyses. Demographic inference methods, such as ∂A∂I (Gutenkunst et al., 2009), can sometimes resolve the influence of divergence timing and migration. Yet in practice, extremely recent divergence and high migration rates are nearly impossible to resolve with confidence (Robinson, Coffman, Hickerson, & Gutenkunst, 2014). The ability to obtain well supported demographic inferences is dependent on various factors, such as the number of loci and individuals sampled, the complexity of the actual demographic history, and how well the true demography is reflected in models (which are undoubtedly over-simplified). Although we ran many simulations per geographic sample pairs (100 simulations, each with 100 optimising iterations), three features of our results suggest that discerning the role of asymmetric versus symmetric migration is challenging with our present data. Firstly, both migration scenarios exhibited considerable variation among the top 10 models with respect to their migration parameters (Figure S7 & S8). Secondly, no clear pattern of asymmetric migration being more likely than symmetric migration (or vice versa) was recovered by our simulations (Table 2). Finally, all the best asymmetric models inferred high migration in both directions, and in some cases, the M12 and M21 parameters were of comparable magnitude (Table 3), which is numerically equivalent to symmetric migration. Hence, further investigations are required to fully characterise patterns of dispersal of wahoo (discussed below).

In summary, our data provides good support for two major conclusions: (1) wahoo likely exist as two weakly differentiated stocks between the Indo-Pacific and the Atlantic Oceans; and (2) this weak differentiation occurs against a backdrop of considerable gene flow in the evolutionary history of wahoo. We do, however, caution against direct interpretation of our results with respect to asymmetry and directionality of migration, for reasons mentioned above. Indeed, the biology of wahoo implies a parameter space where inference is notoriously difficult (Robinson et al., 2014; Rougemont et al., 2017) and there is great scope for future work to more thoroughly examine the eco-evolutionary processes that shape connectivity and the distribution of genetic and phenotypic variation in this species.

Future prospects and implications

In this study, wahoo samples were collected over a 17 year window (1998–2015) to obtain globally distributed samples. In highly migratory, globally distributed marine species, such temporal separation of samples is the norm because broad biogeographic distributions are recalcitrant to collections within a narrow timeframe (Vaux et al., 2021). For demographic inference, which are on evolutionary timescales, the temporal separation in our study is unlikely to affect conclusions, unless major changes in the SFSs within and between locations have occurred over the last two decades. Targeted sampling could provide new insights into temporal stability of genetic patterns, the distribution of life history traits, and measures of individual dispersal trajectories in wahoo. Augmented insights have been obtained, for example, in sampling young-of-the-year from the circumtropical bluefin tuna (Carlsson, McDowell, Carlsson, & Graves, 2007; Boustany, Reeb, & Block, 2008), and using temporal replication when sampling white marlin (Mamoozadeh, McDowell, Rooker, & Graves, 2018). For wahoo, reproductive areas, sex-specific movement patterns, habitat preferences, and philopatric behaviors are largely unknown, hindering the development of biologically informed sampling design (Zischke, 2012; Lascelles et al., 2014). Future work employing individual-based genotyping and increased representation of the Indian Ocean, Central and Eastern Atlantic Ocean, Mediterranean, and African localities would be important steps in resolving putative dispersal patterns in wahoo; for example, a recent genomic investigation of Albacore tuna was able to delineate North versus South Pacific populations using individual focused analyses (Vaux et al. 2021).

The implications of our findings for fisheries are somewhat ambiguous, as wahoo clearly inhabit the "Waples Zone" of weak genetic differentiation arising from a combination of high migration and large effective population sizes (*sensu* Kelley et al., 2010, referring to Waples 1998). If indeed wahoo do travel large distances and mix readily (especially within ocean basins), then these linkages among geographically distant populations would imply that overharvesting in particular locations could affect population numbers elsewhere, particularly if there are seasonal aggregations in regions with limited resources to effectively regulate acute local harvest. On the other hand, recreational and artisanal harvesting is unlikely to affect local stocks at the oceanic stock scale. However, whether genetically similar yet geographically distant populations are ecologically cohesive remains an open question (i.e., Waples & Gaggiotti, 2006; Lowe & Allendorf, 2010). Multidisciplinary approaches that incorporate information based on morphometrics, parasite sharing, and tagging (Sepulveda et al., 2011; Zischke et al., 2012) alongside individual-based genotyping (e.g. Vaux et al., 2021) may uncover connectivity dynamics in a timeframe better matched to wahoo fisheries management than the evolutionary timescales reflected in allele-frequency based genetic data.

Conclusions

Genetic tools are useful for developing practical population delimitations for management purposes. However, characterising discrete stocks of cosmopolitan pelagic fishes is challenging because their large effective population sizes and (or) high dispersal can obscure signals of spatial genetic differentiation. We provide evidence that wahoo populations can be characterised as two weakly differentiated stocks based on genome-wide SNP loci: an Indo-Pacific and an Atlantic stock. Despite this regional structuring, our demographic analyses indicated that these populations are likely globally connected by high gene flow. These findings are in line with genetic-based biogeographic investigations of other large pelagic fishes that highlight substantial evolutionary connections over vast geographic distances.

ACKNOWLEDGEMENTS

We thank the many fishers (commercial, charter, and recreational) and fellow researchers who helped us obtain samples for our study and compile a truly global wahoo dataset. Samples from the Galapagos Islands were collected by I.H., whereas the rest of our samples were collected as part of work by Zischke and colleagues (2012) (provided by M.Z.) and Theisen and colleagues (2008) (provided by J.D.B.) with acknowledgements contained within their respective works.

Collection of samples and their exportation from the Galapagos was via the *Galapagos National Park Directorate Permits* PC-16-15 and 059-2015 DPNG (granted to Dr. Pelayo Salinas from the Charles Darwin Foundation). Importation of wahoo samples into Australia from overseas was via the *Permit to Import Quarantine Material*, IP14019478 (granted to C.R.). Permission to catch wahoo in Queensland was via the *Queensland Department of Primary Industries and Fisheries General Fisheries Permit*, 56095. Permission to catch wahoo in New South Wales was via the *New South Wales Department of Primary Industries Scientific Collection Permit*, P09/0061-1.0.

Finally, financial support was provided by the Ecuadorian Government through a National Secretary of Higher Education, Science and Technology (SENESCYT) scholarship (granted to I.H.) and from discretionary funds to C.R. from the School of Biological Sciences, UQ.

Additionally, we thank Editor Giacomo Bernadi and three anonymous referees who provided very supportive and constructive feedback on our study.

DATA AVAILABILITY

Our scripts and analyses have been uploaded to Dryad: Thia (2021), https://doi.org/10.5061/dryad.dncjsxkz4. Raw ezRAD pool-seq reads have been uploaded to NCBI's SRA as has the relevant BioSample information: BioProject PRJNA683059. Additional metadata for these pool-seq reads are available through GeOME: https://n2t.net/ark:/21547/DhI2.

BIOSKETCH

The team behind this work constitutes a diverse group of marine biologists, biogeographers, and population geneticists. We share a general interest in understanding the processes that have shaped the distribution and evolution of marine life. Additionally, we seek to use our insights of eco-evolutionary processes to inform management decisions that benefit the sustainability of our oceans.

REFERENCES

Abaunza, P., Murta, A. G., Campbell, N., Cimmaruta, R., Comesana, A. S., Dahle, G., … Zimmermann, C. (2008). Stock identity of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean Sea: Integrating the results from different stock identification approaches. *Fisheries Research*, 89(2), 196–209. (WOS:000254162300013). https://doi.org/10.1016/j.fishres.2007.09.022

Andrews, S. (2011). FastQ Screen (http://www.bioinformatics.bbsrc.ac.uk/projects/fastq_screen/): NGS reads quality control. Retrieved from http://www.bioinformatics.bbsrc.ac.uk/projects/fastq_screen/

Barth, J. M. I., Damerau, M., Matschiner, M., Jentoft, S., & Hanel, R. (2017). Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna

(*Thunnus albacares*) populations. *Genome Biology and Evolution*, *9*(4), 1084–1098. (28419285). https://doi.org/10.1093/gbe/evx067

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boustany, A. M., Reeb, C. A., & Block, B. A. (2008). Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Marine Biology*, 156(1), 13–24. https://doi.org/10.1007/s00227-008-1058-0

Brown-Peterson, N. J., Franks, J. S., & Burke, A. M. (2000). Preliminary observations on the reproductive biology of wahoo, *Acanthocybium solandri*, from the northern Gulf of Mexico and Bimini, Bahamas. *Proceedings of the Gulf and Caribbean Fisheries Institute*, 51, 414–427.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference A Practical Information-Theoretic Approach* (2nd ed..). New York, NY: Springer New York.

Carlsson, J., McDowell, J. R., Carlsson, J. E. L., & Graves, J. E. (2007). Genetic identity of YOY bluefin tuna from the Eastern and Western Atlantic spawning areas. *Journal of Heredity*, *98*(1), 23–28. https://doi.org/10.1093/jhered/esl046

Chong, Z., Ruan, J., & Wu, C.-I. (2012). Rainbow: An integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, 28(21), 2732–2737. https://doi.org/10.1093/bioinformatics/bts482

Collette, B., & Nauen, C. (1983). FAO species volume 2. Scombrids of the world. An annotated and illustrated catalogue of tunas, mackerels, bonitos and related species known to date. *FAO Fisheries Synopsis*, 125.

Crandall E.D., Riginos C., Bird C.E., Liggins L., Treml E.A., Beger M., Barber P.H., Connolly S.R., Cowman P.F., … Gaither, M.R. (2019) The molecular biogeography of the Indo-Pacific: Testing hypotheses with multispecies genetic patterns. *Global Ecology and Biogeography*, 58(5), 403-418.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group, (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Excoffier, L., Smouse, P.E., & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* 131(2), 479–491.

FAO. (2019). *FISHSTATJ Plus: Universal software for fishery statistical time series*.

Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, 186(1), 207. https://doi.org/10.1534/genetics.110.114397

Gaither M.R., & Rocha L.A. (2016) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. *J Biogeography*, 40(9),1638-1648.

Gaither, M. R., Bowen, B. W., Rocha, L. A., & Briggs, J. C. (2016). Fishes that rule the world: Circumtropical distributions revisited. *Fish and Fisheries*, 17(3), 664–679. https://doi.org/10.1111/faf.12136

Garber, A. F., Tringali, M. D., & Franks, J. S. (2005). Population genetic and phylogeographic structure of wahoo, *Acanthocybium solandri*, from the Western Central Atlantic and Central Pacific Oceans. *Marine Biology*, 147(1), 205–214.

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

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., … Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Molecular Ecology*, 22(14), 3766–3779. https://doi.org/10.1111/mec.12360

Grewe, P. M., Feutry, P., Hill, P. L., Gunasekera, R. M., Schaefer, K. M., Itano, D. G., … Davies, C. R. (2015). Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports*, *5*, 1–9. https://doi.org/10.1038/srep16916

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Hauser, L., & Ward, R. D. (1998). *Population identification in pelagic fish: The limits of molecular markers. In: Advances in Molecular Ecology* (G. R. Carvalho, Ed.). Amsterdam: IOS Press.

Hawkins, S. J., Bohn, K., Sims, D. W., Ribeiro, P., Faria, J., Presa, P., … Genner, M. J. (2016). Fisheries stocks from an ecological perspective: Disentangling ecological connectivity from genetic interchange. *Fisheries Research*, 179, 333–341. https://doi.org/10.1016/j.fishres.2016.01.015

Heng, L., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, 18(11), 1851–1858. https://doi.org/10.1101/gr.078212.108

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from pool-seq data. *Genetics*, *210*(1), 315. https://doi.org/10.1534/genetics.118.300900

Hutchings, L., van der Lingen, C. D., Shannon, L. J., Crawford, R. J. M., Verheye, H. M. S., Bartholomae, C. H., … Monteiro, P. M. S. (2009). The Benguela Current: An ecosystem of four components. *Progress in Oceanography*, 83(1), 15–32. https://doi.org/10.1016/j.pocean.2009.07.046

Jenkins, K. L. M., & McBride, R. S. (2009). Reproductive biology of wahoo, *Acanthocybium solandri*, from the Atlantic coast of Florida and the Bahamas. *Marine and Freshwater Research*, 60(9), 893–897.

Laconcha, U., Iriondo, M., Arrizabalaga, H., Manzano, C., Markaide, P., Montes, I., … Estonba, A. (2015). New nuclear snp markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). *PLoS One*, 10(6). (26090851). https://doi.org/10.1371/journal.pone.0128247

Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4), 357–360.

Lascelles, B., Notarbartolo Di Sciara, G., Agardy, T., Cuttelod, A., Eckert, S., Glowka, L., … Ridoux, V. (2014). Migratory marine species: Their status, threats and conservation management needs. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 24(S2), 111–127.

Lessios H.A., & Robertson D.R. (2006). Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society B*, 273(1598), 2201-2208.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv:1303.3997*.

Li, H. (2016). Seqtk: A fast and lightweight tool for processing FASTA or FASTQ sequences. Retrieved from https://github.com/lh3/seqtk/.

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Linck, E., & Battey, C. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3), 639–647.

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, 19(15), 3038–3051. https://doi.org/10.1111/j.1365-294x.2010.04688.x

Luckhurst, B. E. (2007). Large pelagic fishes in the wider Caribbean and northwest Atlantic Ocean: Movement patterns determined from conventional and electronic tagging. *Gulf and Caribbean Research*, 19(2), 5–14.

Luckhurst, B. E., & Trott, T. (2000). Bermuda’s commercial line fishery for wahoo and dolphinfish: Landings, seasonality and catch per unit effort trends. *Proceedings of the Gulf and Caribbean Fisheries Institute*, 51, 404–413.

Ludt W.B., Rocha L.A. (2014). Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, 42(1), 25-38.

Mamoozadeh, N. R., Graves, J. E., & McDowell, J. R. (2019). Genome-wide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (*Kajikia audax*), a broadly distributed and highly migratory pelagic species. *Evolutionary Applications*, *13*(4), 677–698. https://doi.org/10.1111/eva.12892

Mamoozadeh, N. R., McDowell, J. R., Rooker, J. R., & Graves, J. E. (2018). Genetic evaluation of population structure in white marlin (*Kajikia albida*): The importance of statistical power. *ICES Journal of Marine Science*, 75(2), 892–902. https://doi.org/10.1093/icesjms/fsx047

Maroso, F., Franch, R., Dalla Rovere, G., Arculeo, M., & Bargelloni, L. (2016). RAD SNP markers as a tool for conservation of dolphinfish *Coryphaena hippurus* in the Mediterranean Sea: Identification of subtle genetic structure and assessment of populations sex-ratios. *Marine Genomics*, 28, 57–62. https://doi.org/10.1016/j.margen.2016.07.003

Matz, M. V. (2018). Fantastic beasts and how to sequence them: Ecological genomics for obscure model organisms. *Trends in Genetics*, 32(2), 121–132.

Mullins, R. B., McKeown, N. J., Sauer, W. H. H., & Shaw, P. W. (2018). Genomic analysis reveals multiple mismatches between biological and management units in yellowfin tuna (*Thunnus albacares*). *ICES Journal of Marine Science*, 75(6), 2145–2152. https://doi.org/10.1093/icesjms/fsy102

NMFS. (1999). *Billfish Newsletter. The Southwest Fisheries Science Center's. Prepared by David Holtz and Douglas Prescott. Southwest Fisheries Science Center, PO Box 271, La Jolla, CA 92038-0271*.

Kelly R.P., Oliver T.A., Sivasundar A., & Palumbi S.R. (2010) A method for detecting population genetic structure in diverse, high gene-flow species. *Journal of Heredity*, 101(4), 423-436.

Oxenford, H. A., Murray, P. A., & Luckhurst, B. E. (2003). The biology of wahoo (*Acanthocybium solandri*) in the western central Atlantic. *Gulf and Caribbean Research*, 15(1), 33–49.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology, Evolution, and Systematics*, 25(1), 547–572. https://doi.org/10.1146/annurev.es.25.110194.002555

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26, 419–420.

Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Patarnello, T., Volckaert, F. A. M. J., & Castilho, R. (2007). Pillars of Hercules: Is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, 16(21), 4426–4444. https://doi.org/10.1111/j.1365-294x.2007.03477.x

Pecoraro, C., Babbucci, M., Franch, R., Rico, C., Papetti, C., Chassot, E., … Tinti, F. (2018). The population genomics of yellowfin tuna (*Thunnus albacares*) at global geographic scale challenges current stock delineation. *Scientific Reports*, 8(1), 13890. https://doi.org/10.1038/s41598-018-32331-3

Peeters, F. J. C., Acheson, R., Brummer, G.-J. A., Wilhelmus, P. M. de R., Schneider, R.R., Ganssen, G.M., ... Kroon, D. (2004). Vigorous exchange between the Indian and Atlantic oceans at the end of the past five glacial periods. *Nature*, 430(7000), 661–665. https://doi.org/10.1038/nature02785

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2014*(1). https://doi.org/10.7717/peerj.431

Robinson, J. D., Coffman, A. J., Hickerson, M. J., & Gutenkunst, R. N. (2014). Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evolutionary Biology*, 14(1). https://doi.org/10.1186/s12862-014-0254-4

Rocha L.A., Craig M.T., & Bowen B.W. (2007) Phylogeography and the conservation of coral reef fishes. *Coral Reefs* 26(3), 501-512.

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., & Evanno, G. (2017). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, 26(1), 142–162. https://doi.org/10.1111/mec.13664

Rougeux, C., Gagnaire, P., & Bernatchez, L. (2019). Model-based demographic inference of introgression history in European whitefish species pairs. *Journal of Evolutionary Biology*, 32(8), 806–817. https://doi.org/10.1111/jeb.13482

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749–763. https://doi.org/10.1038/nrg3803

Sepulveda, C. A., Aalbers, S. A., Ortega-Garcia, S., Wegner, N. C., & Bernal, D. (2011). Depth distribution and temperature preferences of wahoo (*Acanthocybium solandri*) off Baja California Sur, Mexico. *Marine Biology*, 158(4), 917–926.

Staton, E. (2013). Pairfq: Sync paired-end FASTA/Q files and keep singleton reads. Retrieved from https://github.com/sestaton/Pairfq

Theisen, T. C., & Baldwin, J. D. (2012). Movements and depth/temperature distribution of the ectothermic Scombrid, *Acanthocybium solandri* (wahoo), in the western North Atlantic. *Marine Biology*, 159(10), 2249–2258.

Theisen, T. C., Bowen, B. W., Lanier, W., & Baldwin, J. D. (2008). High connectivity on a global scale in the pelagic wahoo, *Acanthocybium solandri* (tuna family Scombridae). *Molecular Ecology*, 17(19), 4233–4247. https://doi.org/10.1111/j.1365-294X.2008.03913.x

Thia, J. A., & Riginos, C. (2019). genomalicious: Serving up a smorgasbord of R functions for population genomic analyses. *BioRxiv*, 667337. https://doi.org/10.1101/667337

Thia, Joshua (2020), Population structure and demographic analyses of *Acanthocybium solandri* from the Indo-Pacific and Atlantic oceans, Dryad, Dataset, https://doi.org/10.5061/dryad.dncjsxkz4

Thia, Joshua (2021), Population structure and demographic analyses of Acanthocybium solandri from the Indo-Pacific and Atlantic oceans, Dryad, Dataset, https://doi.org/10.5061/dryad.dncjsxkz4

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). EzRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ*, e203(1). https://doi.org/10.7717/peerj.203

Vaux F, Bohn S, Hyde JR, O'Malley KG. (2021). Adaptive markers distinguish North and South Pacific Albacore amid low population differentiation. *Evolutionary Applications*, 1-22. https://doi.org/10.1111/eva.13202

Walters, V., & Fierstine, H. L. (1964). Measurements of swimming speeds of yellowfin tuna and wahoo. *Nature*, 202, 208–209.

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89(5), 438–450.

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6), 1419–1439. https://doi.org/10.1111/j.1365-294x.2006.02890.x

Wollam, M. B. (1969). Larval wahoo, *Acanthocybium solanderi* (Cuvier) from the straits of Yucatan and Florida. *Florida Department of Natural Resources, Division of Marine Resources Marine Research Laboratory Leaflet Series*, *4*, 1–7.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zischke, M. T. (2012). A review of the biology, stock structure, fisheries and status of wahoo (*Acanthocybium solandri*), with reference to the Pacific Ocean. *Fisheries Research*, 119, 13–22.

Zischke, M. T., Farley, J. H., Griffiths, S. P., & Tibbetts, I. R. (2013). Reproductive biology of wahoo, *Acanthocybium solandri*, off eastern Australia. *Reviews in Fish Biology and Fisheries*, 23(4), 491–506.

Zischke, M. T., Griffiths, S. P., Tibbetts, I. R., & Lester, R. J. (2012). Stock identification of wahoo (*Acanthocybium solandri*) in the Pacific and Indian Oceans using morphometrics and parasites. *ICES Journal of Marine Science*, 70(1), 164–172.

TABLE AND FIGURE LEGENDS

Figure 1. Sampling spatial distribution and population structure of wahoo. **(a)** Global distribution of our focal wahoo locations with number of sampled fish in parentheses. Inset is an illustration of a wahoo. **(b)** Distribution of pairwise mean bootstrap F_{ST} estimates from genome-wide SNPs using sample pair specific SNP sets (1,289 ≤ n ≤ 9,825 loci). Estimates are grouped with respect to comparisons within versus between ocean basins. Boxplot outliers were only present in the Indian/Pacific comparisons: AmSam/ChrIsl (F_{ST} = 0.012) and AmSam/Thai (F_{ST} = 0.008) were upper outliers, whereas EAus/Thai (F_{ST} = 0.001) was a lower outlier. **(c)** Principal coordinate analysis (PCoA) of pairwise mean bootstrap F_{ST} estimates derived from sample pair specific SNP sets. Numbers in parentheses for axes labels indicate the proportion of variance captured by each PCo axis. Colours represent ocean basins (see legend). Location abbreviations: AmSam = American Samoa; Bimini = Bimini; ChrIsl = Christmas Island; EAus = East Australia; Gal = Galapagos; GrandCay = Grand Cayman; Hawaii = Hawaii; NCar = North Carolina; Palau = Palau; Thailand = Thailand; TrinTab = Trinidad & Tobago.

Figure 2. Demographic scenarios for gene flow between two populations (1 and 2): isolation, symmetric migration, and asymmetric migration. T, effective number of generations since divergence, nu, relative contemporary population size parameter for each population, M, symmetrical migration, M12, scaled migration rate from population 2 into population 1, M21, scaled migration rate from population 1 into population 2.

Figure 3. Folded site frequency spectrums (SFS) for the American Samoa/North Carolina wahoo population pair for the observed data and our three simulated scenarios (Isolation, Symmetric, Asymmetric). Similarity between the simulated scenarios and the observed SFS indicate that reasonable parameter combinations were identified in our demographic analyses. The x-axes are the allele counts for North Carolina, the y-axes are the allele counts of American Samoa, and coloured cells illustrate the frequency of joint allele counts between the populations (see coloured scale bar). The SFSs presented were filtered to a minimum joint allele count of 1 for visualisation.

Table 1. Pairwise mean bootstrapped F_{ST} estimates between wahoo sample pairs (ordered by ocean basin), across 1,000 bootstrap replicates of 1,000 subsampled loci (without replacement), using sample pair specific SNP sets ($1{,}289 \leq n \leq 9{,}825$ loci).

Note: Pairs with $F_{ST} > 0$ based on bootstrap confidence intervals are in bold (see also Table S2). Population abbreviations are as follows: AmSam = American Samoa; ChrIsl = Christmas Island; EAus = East Australia; Gal = Galapagos; GraCay = Grand Cayman; NCar = North Carolina; Thai = Thailand; TrinTab = Trinidad & Tobago.

Table 2. Analysis of molecular variance (AMOVA) of wahoo populations among basins nested within regions, using SNP sets specific to sample pairs ($1{,}289 \leq n \leq 9{,}825$ loci) or a common shared SNP set among all sample pairs (n = 945 loci).

Table 3. Estimated demographic parameters from ∂A∂I analyses between wahoo sample pairs, within or between ocean basins, and for different demographic scenarios.

[a] Location abbreviations: AmSam = American Samoa; Gal = Galapagos; NCar = North Carolina; TrinTab = Trinidad & Tobago.

[b] Underlined scenarios were deemed the most likely for their sample pair, based on lowest AIC, and ΔAIC (< 10 considered equivalent).

[c] Migration parameters (M, M12, and M21) are forward in time estimates, the number of chromosomes moving between populations per generation.

Table 1. Pairwise mean bootstrapped F_{ST} estimates between wahoo sample pairs (ordered by ocean basin), across 1,000 bootstrap replicates of 1,000 subsampled loci (without replacement), using sample pair specific SNP sets (1,289 ≤ *n* ≤ 9,825 loci).

		Atlantic				**Indian**		**Pacific**				
		Bimini	**GraCay**	**NCar**	**TrinTab**	**ChrIsl**	**Thai**	**AmSam**	**EAus**	**Gal**	**Hawaii**	**Palau**
Atlantic	**Bimini**	0										
	GraCay	**0.004**	0									
	NCar	**0.004**	**0.004**	0								
	TrinTab	**0.004**	**0.007**	**0.008**	0							
Indian	**ChrIsl**	**0.015**	**0.017**	**0.015**	**0.011**	0						
	Thai	**0.016**	**0.012**	**0.017**	**0.012**	**0.005**	0					
Pacific	**AmSam**	**0.02**	**0.018**	**0.021**	**0.017**	**0.012**	**0.008**	0				
	EAus	**0.013**	**0.011**	**0.013**	**0.012**	**0.003**	0.001	**0.007**	0			
	Gal	**0.017**	**0.016**	**0.019**	**0.014**	0.002	**0.004**	**0.008**	**0.003**	0		
	Hawaii	**0.017**	**0.018**	**0.021**	**0.014**	**0.004**	**0.004**	**0.013**	**0.005**	**0.006**	0	
	Palau	**0.014**	**0.012**	**0.015**	**0.009**	**0.004**	0.003	**0.009**	**0.004**	**0.004**	**0.004**	0

Note: Pairs with $F_{ST} > 0$ based on bootstrap confidence intervals are in bold (see also Table S2). Population abbreviations are as follows: AmSam = American Samoa; ChrIsl = Christmas Island; EAus = East Australia; Gal = Galapagos; GraCay = Grand Cayman; NCar = North Carolina; Thai = Thailand; TrinTab = Trinidad & Tobago.

Table 2. Analysis of molecular variance (AMOVA) of wahoo populations among basins nested within regions, using SNP sets specific to sample pairs (1,289 ≤ *n* ≤ 9,825 loci) or a common shared SNP set among all sample pairs (*n* = 945 loci).

SNP set	Term	SSD	MSD	DF	% var	*p*-value
Specific	Regions	5.44e−4	5.44e−4	1	77.41	< 0.001
	Basins	5.73e−6	5.73e−6	1	0.82	0.695
	Error	1.53e−4	1.91e−5	8	21.78	
	Total	7.02e−4		10		
Shared	Regions	4.43e−5	4.43e−5	1	57.43	0.335
	Basins	1.28e−5	1.28e−5	1	16.53	0.278
	Error	2.01e−5	2.51e−6	8	26.04	
	Total	7.72e−5		10		

Table 3. Estimated demographic parameters from ∂A∂I analyses between wahoo sample pairs, within or between ocean basins, and for different demographic scenarios.

Sample pair[a]	Basin	Scenario[b]	Log-likelihood	T	nu1	nu2	M[c]	M12[c]	M21[c]	AIC	ΔAIC
AmSam/Gal	Within	Isolation	-583	1.69e−10	7.98e−09	6.35e−09				1172	665
		Symmetric	−280	0.82	0.33	0.37	55.60			567	60
		Asymmetric	−249	0.55	0.26	0.20		73.94	98.91	508	0
NCar/TrinTab	Within	Isolation	−505	2.56e−07	1.19e−05	1.50e−05				1016	497
		Symmetric	−261	0.88	0.34	0.40	52.15			530	11
		Asymmetric	−254	0.77	0.24	0.51		96.80	46.18	519	0
AmSam/NCar	Between	Isolation	−551	1.41e−08	5.76e−07	3.61e−07				1108	482
		Symmetric	−309	0.69	0.42	0.23	63.71			626	0
		Asymmetric	−344	0.94	0.67	0.31		37.96	60.74	699	73
AmSam/TrinTab	Between	Isolation	−556	4.20e−09	1.75e−07	1.25e−07				1118	539
		Symmetric	−297	0.74	0.31	0.30	63.07			601	23
		Asymmetric	−284	0.46	0.28	0.15		98.42	94.36	579	0
Gal/NCar	Between	Isolation	−580	6.15e−08	2.36e−06	1.73e−06				1167	563

		Symmetric	−298	0.73	0.35	0.27	62.87			604	0
		Asymmetric	−301	0.64	0.22	0.36		94.60	59.13	611	7
Gal/TrinTab	Between	Isolation	−572	7.46e−08	2.90e−06	2.47e−06				1150	612
		Symmetric	−290	0.78	0.36	0.30	58.54			589	51
		Asymmetric	−264	0.49	0.21	0.19		99.78	89.43	538	0

[a] Location abbreviations: AmSam = American Samoa; Gal = Galapagos; NCar = North Carolina; TrinTab = Trinidad & Tobago.

[b] Underlined scenarios were deemed the most likely for their sample pair, based on lowest AIC, and ΔAIC (< 10 considered equivalent).

[c] Migration parameters (M, M12, and M21) are forward in time estimates, the number of chromosomes moving between populations per generation.

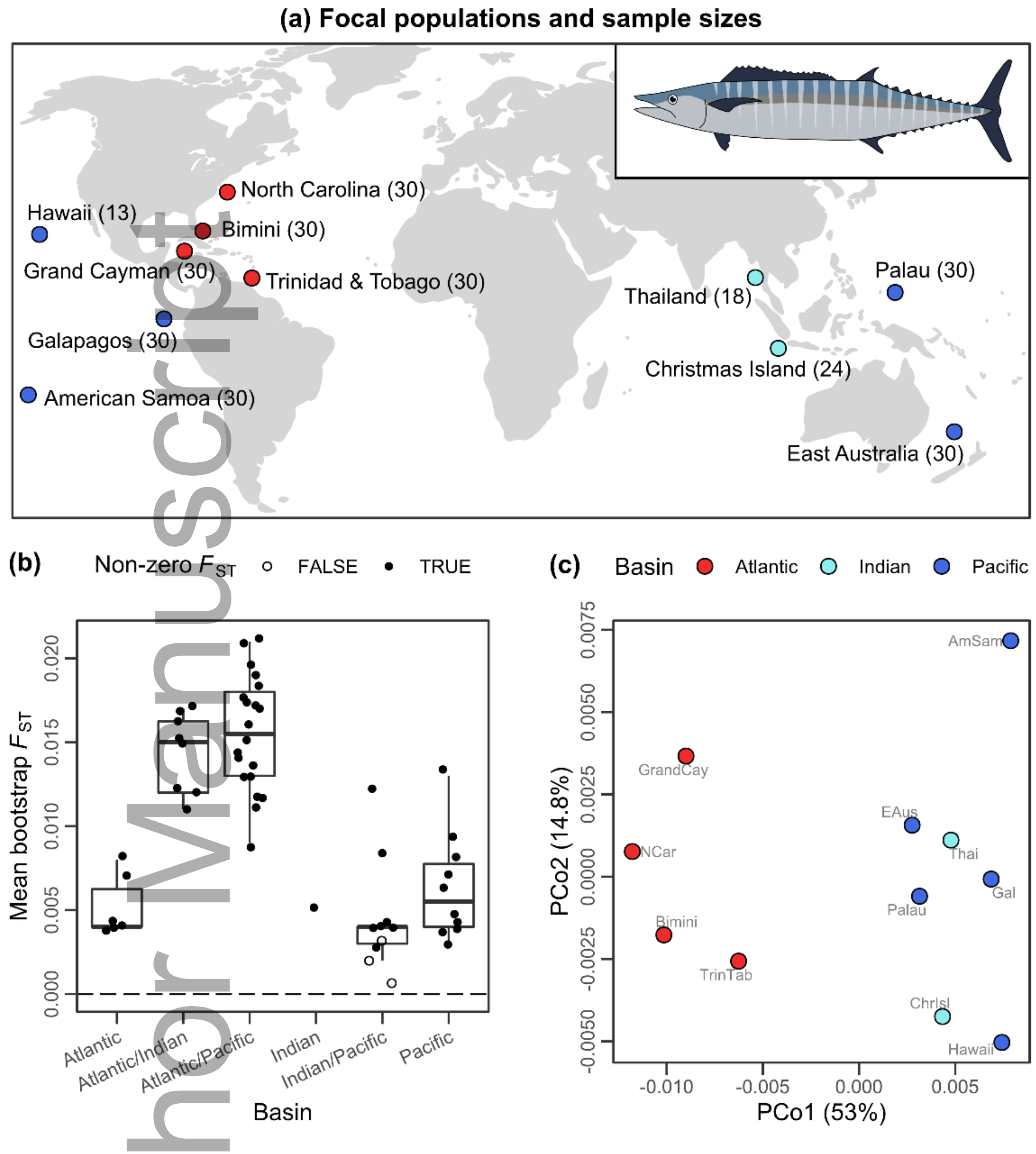


Figure 1. Sampling spatial distribution and population structure of wahoo. **(a)** Global distribution of our focal wahoo locations with number of sampled fish in parentheses. Inset is an illustration of a wahoo. **(b)** Distribution of pairwise mean bootstrap F_{ST} estimates from genome-wide SNPs using sample pair specific SNP sets (1,289 ≤ n ≤ 9,825 loci). Estimates are grouped with respect to comparisons within versus between ocean basins. Boxplot outliers were only present in the Indian/Pacific comparisons: AmSam/ChrIsl (F_{ST} = 0.012) and AmSam/Thai (F_{ST} = 0.008) were upper outliers, whereas EAus/Thai (F_{ST} = 0.001) was a lower outlier. **(c)** Principal coordinate analysis (PCoA) of pairwise mean bootstrap F_{ST} estimates derived from sample pair specific SNP sets. Numbers in parentheses for axes labels indicate the proportion of variance captured by each PCo

axis. Colours represent ocean basins (see legend). Location abbreviations: AmSam = American Samoa; Bimini = Bimini; ChrIsl = Christmas Island; EAus = East Australia; Gal = Galapagos; GrandCay = Grand Cayman; Hawaii = Hawaii; NCar = North Carolina; Palau = Palau; Thailand = Thailand; TrinTab = Trinidad & Tobago.

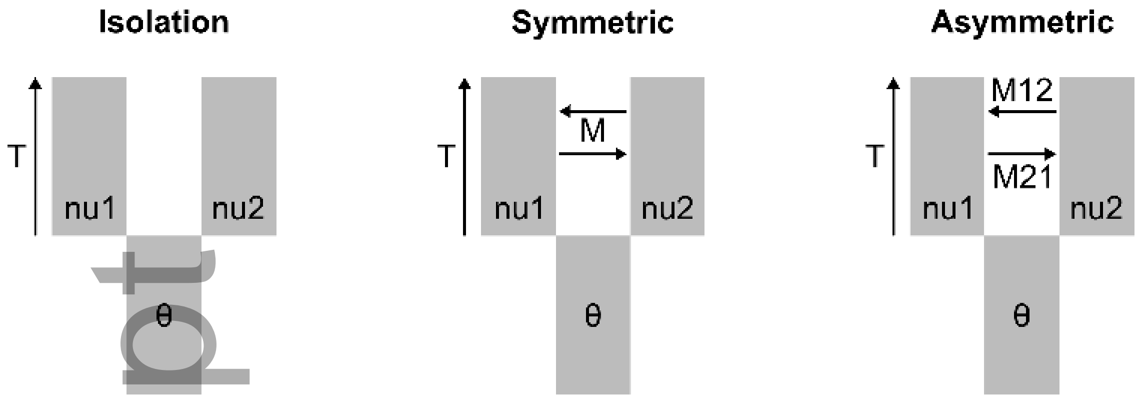


Figure 2. Demographic scenarios for gene flow between two populations (1 and 2): isolation, symmetric migration, and asymmetric migration. T, effective number of generations since divergence, nu, relative contemporary population size parameter for each population, M, symmetrical migration, M12, scaled migration rate from population 2 into population 1, M21, scaled migration rate from population 1 into population 2.

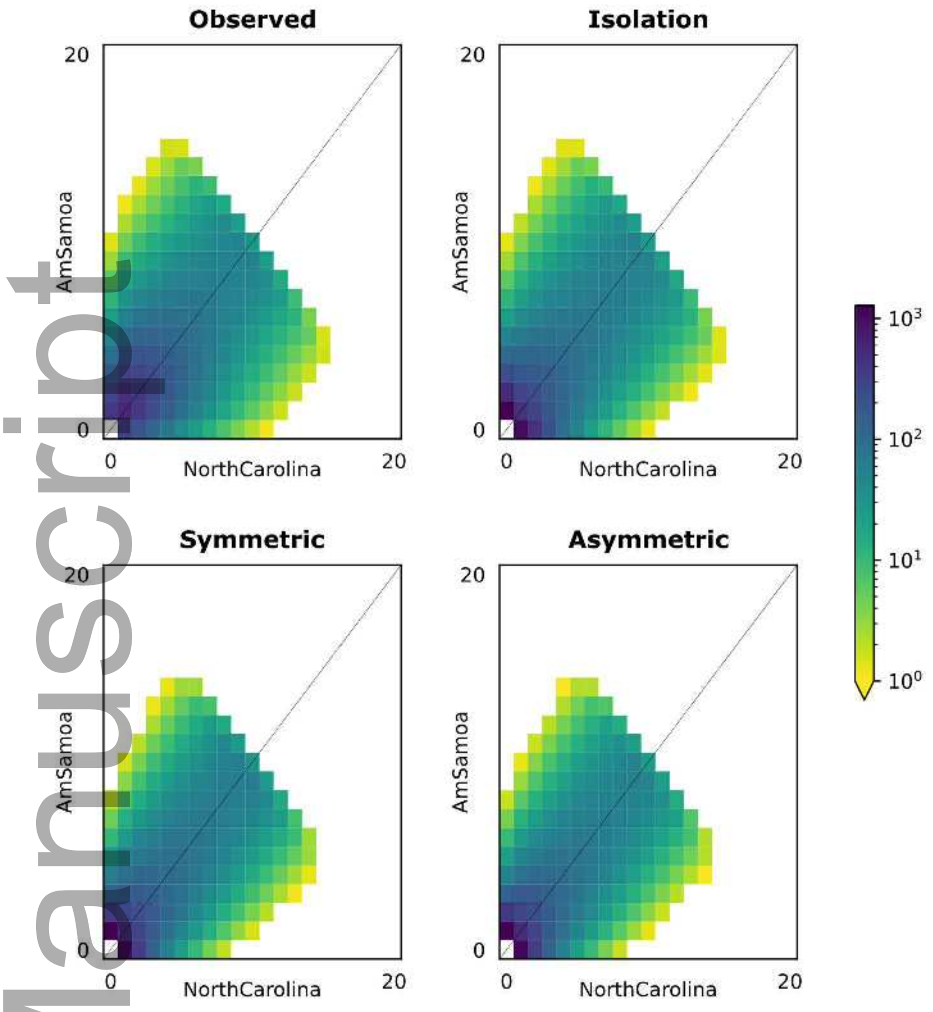
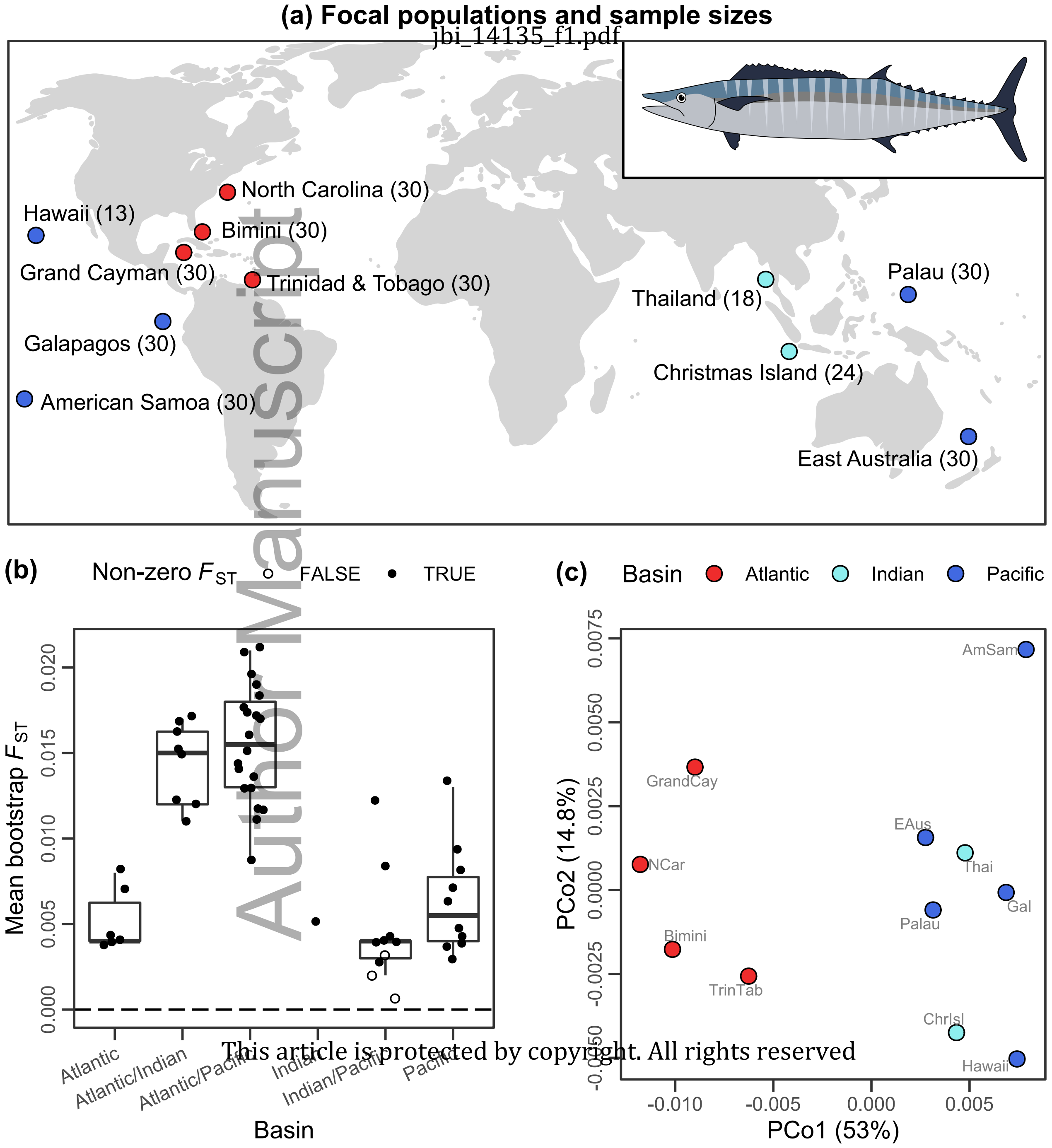
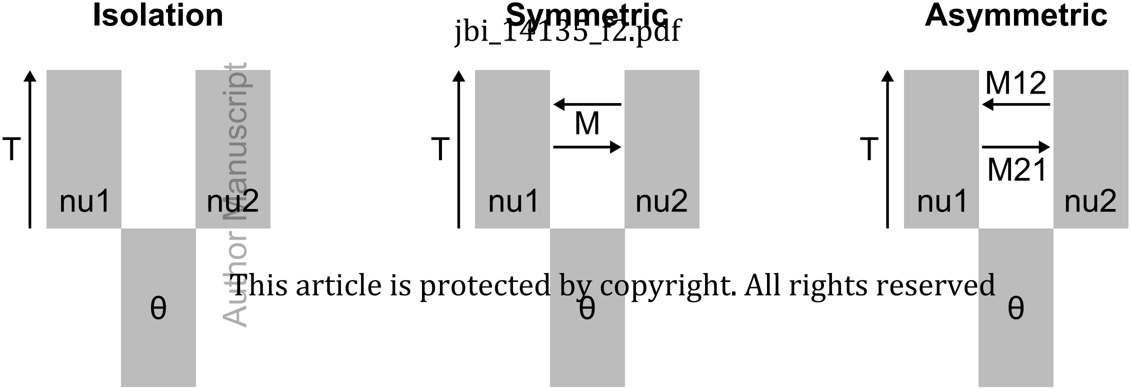


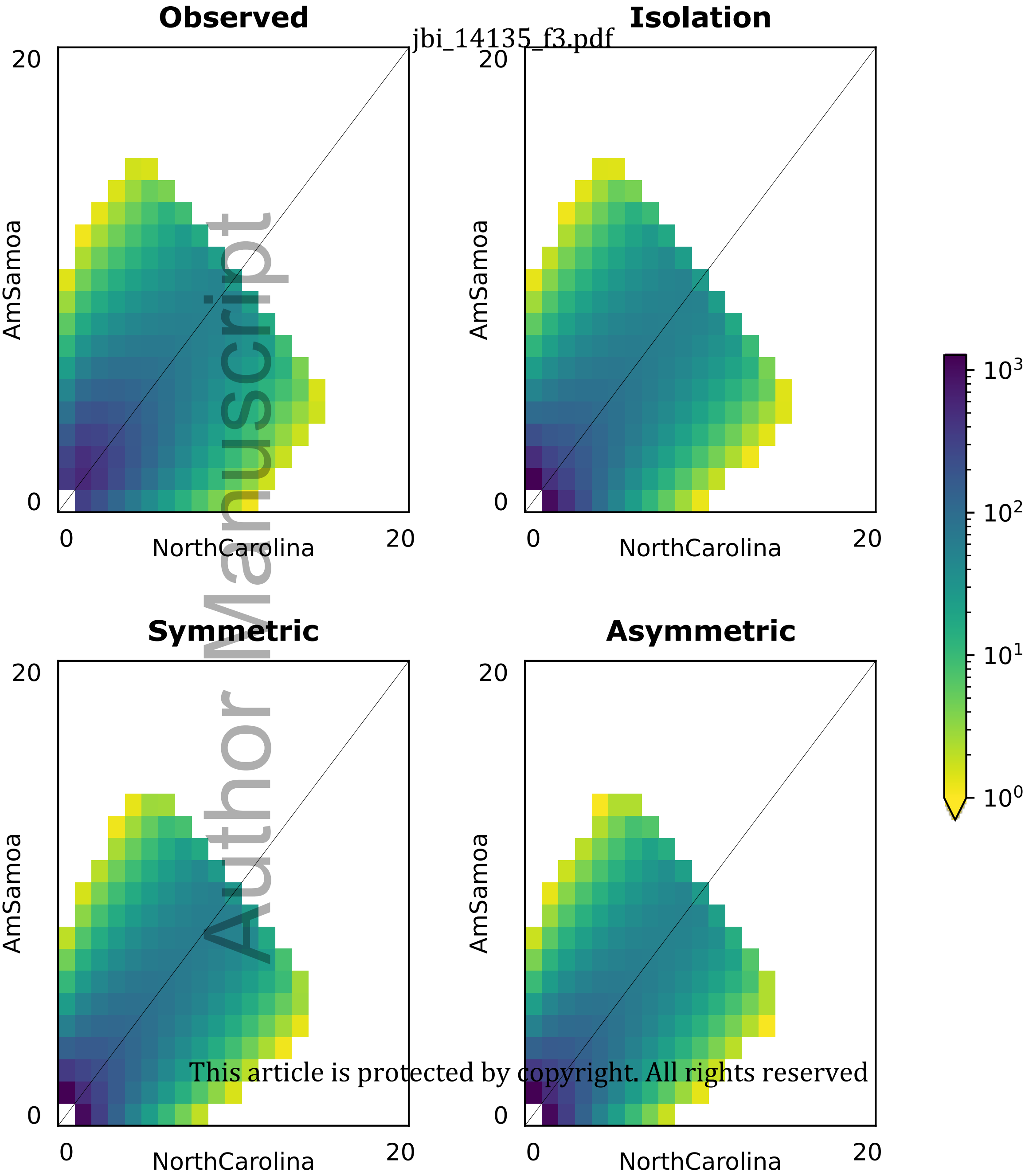
Figure 3. Folded site frequency spectrums (SFS) for the American Samoa/North Carolina wahoo population pair for the observed data and our three simulated scenarios (Isolation, Symmetric, Asymmetric). Similarity between the simulated scenarios and the observed SFS indicate that reasonable parameter combinations were identified in our demographic analyses. The x-axes are the allele counts for North Carolina, the y-axes are the allele counts of American Samoa, and coloured cells illustrate the frequency of joint allele counts between the populations (see coloured scale bar). The SFSs presented were filtered to a minimum joint allele count of 1 for visualisation.

jbi_14135_f1.pdf



jbi_14135_f2.pdf




Observed
Isolation
Symmetric
Asymmetric
20
0
AmSamoa
NorthCarolina
0
20
10^3
10^2
10^1
10^0