## ORIGINAL ARTICLE

# A multi-locus assessment of connectivity and historical demography in the bluehead wrasse (*Thalassoma bifasciatum*)

RA Haney[1], BR Silliman[2] and DM Rand[1]

[1]*Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA and* [2]*Department of Zoology, University of Florida, Gainesville, FL, USA*

The pelagic larval stage of most coral reef fishes might allow extensive dispersal or, alternatively, some level of local recruitment might be important. Molecular markers can be used to obtain indirect estimates of dispersal to evaluate these alternatives, yet the extent of gene flow among populations is known for only a small number of species. The use of such markers must take into account the properties of the markers and the demographic history of the population when making inferences about current gene flow. In the Caribbean bluehead wrasse, *Thalassoma bifasciatum*, previous studies have found both substantial levels of local recruitment, in studies interpreting otolith microchemistry and, conversely, a lack of genetic differentia- tion inferred from mitochondrial DNA (mtDNA) restriction- fragment length polymorphism (RFLP) data and allozymes. However, if subtle differentiation exists, larger sample sizes and highly variable markers may be required to discern it. Here we present results from mitochondrial control region sequence and microsatellite data that indicate a lack of genetic differentiation at both small and large spatial scales. However, historical processes, such as changes in popula- tion size, may have affected the current distribution of genetic variation.

## Introduction

The tremendous diversity of coral reef fishes draws biologists and nonbiologists alike to the shallow tropical seas of the world, and in the past few decades important aspects of the biology of many reef fish species have become amenable to study in their underwater habitats (Ehrlich, 1975). However, the degree of connectivity among reef fish populations is unknown for the vast majority of species, and is of more than intellectual interest. As resource overexploitation is a global concern in marine, including reef, ecosystems (Hutchings, 1996; Birkeland, 1997; Myers *et al.*, 1997; Schiermeier, 2002), the concept of marine reserves has arisen in response, both to ensure fishery sustainability and to protect biodiver- sity and critical habitats (Sale, 2002). The proper design of marine reserves, whether single protected areas or coordinated networks, requires knowledge of the dis- persal capabilities of target species, and the degree to which populations are demographically connected by dispersal (Stobutzki, 2001; Botsford *et al.*, 2003; Gaines *et al.*, 2003; Palumbi, 2003). Studies of nontarget species may still be informative as to the oceanographic and life

history factors that contribute to a given magnitude of realized dispersal.

For species of fish on coral reefs, the most widely accepted connectivity model indicates an open-popula- tion structure. In such a model, local recruitment is independent of local reproduction, instead being depen- dent on processes operating at larger spatial scales (Caley *et al.*, 1996). Although the benthic adults of most species are relatively sedentary, most of the approxi- mately 100 families of fish on coral reefs have pelagic larvae, and a majority have pelagic eggs as well (Ehrlich, 1975). These eggs and larvae were believed to behave as passive particles, be transported at the whim of ocean currents, and thus to be widely dispersed.

This open-population paradigm has recently been challenged by data stressing that local recruitment, possibly owing to small-scale oceanographic phenomena or larval homing behavior, may also be an important process (Jones *et al.*, 1999; Swearer *et al.*, 1999; Atema *et al.*, 2002; Leis, 2002; Warner and Cowen, 2002). Swearer *et al.* (1999) exploited trace element signatures in otoliths to determine the developmental provenance, and hence the source, of larvae of *Thalassoma bifasciatum* recruited to three bays on the island of St Croix in the Caribbean. Results indicated that summer recruitment to two leeward bays was strongly dependent upon local reproduction, which was responsible for up to 89% of sampled larvae at those sites. A companion study performed on the Great Barrier Reef with the Damselfish *Pomacentrus amboinensis* also indicated higher than

Correspondence: *RA Haney, Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, RI 02912, USA.*
E-mail: Robert_Haney@brown.edu

expected levels of self-recruitment (Jones *et al.*, 1999). Larvae were marked by exposure to tetracycline and direct measurement of the number of adult individuals bearing a fluorescent signature in their otoliths yielded an estimate of up to 60% of juveniles recruiting from the natal population. Biophysical modeling in Caribbean reef fish species suggests that larval dispersal is limited, typically on a scale of 10 s to 100 s of kilometers (Cowen *et al.*, 2006).

Depending on the magnitude and persistence of local recruitment, a significant reduction in gene flow among populations may be apparent. Results from studies employing molecular markers indicate that local recruitment is sufficient for isolation of populations over evolutionary timescales in some species. Significant levels of differentiation among populations of several Indo–Pacific reef fish species have been found on both small and large spatial scales, using allozyme or mitochondrial sequence data (Planes *et al.*, 1998; Bernardi *et al.*, 2001; Fauvelot and Planes, 2002). In the Caribbean, Taylor and Hellberg (2003) found evidence for strong genetic structuring among populations of the goby *Elacatinus evelynae* using cytochrome *b* sequence data. Shulman and Bermingham (1995) found evidence for significant genetic differentiation in three of eight species studied. The genetic differentiation within one of these species (*Halichoeres bivittatus*) was later found to be because of a clade level divergence and argued to be owing to ecological speciation while genetic differentiation within clades was lacking over large areas of the Caribbean Basin (Rocha *et al.*, 2005). A similar result was observed in an earlier study of surgeonfish species, with a lack of genetic differentiation within the confines of the Caribbean Basin (Rocha *et al.*, 2002).

Molecular marker-based studies of population connectivity in reef fishes are of considerable interest, as direct estimates of dispersal are difficult to obtain and otolith techniques remain a labor-intensive enterprise. However, issues of scale may intrude (Levin, 1992). First, the spatial scale at which recruitment is 'local' has the potential to be quite different in molecular versus otolith studies. Second, short-term dispersal processes may differ from those important in creating patterns that reflect averages over longer timescales. In marine environments, multiple dispersal mechanisms operate, on varying spatio–temporal scales. In particular, the role of rare long-distance dispersal events may be of great importance over the longer timescales on which population connectivity is assessed by molecular markers, yet have little impact on demography measured over the short term (Kinlan *et al.*, 2005).

Although substantial self-recruitment on a small spatial scale was documented in *T. bifasciatum* (Swearer *et al.*, 1999), analysis of mitochondrial DNA (mtDNA) restriction-fragment length polymorphism (RFLP) data showed no evidence of genetic differentiation among six populations from across the Caribbean (Shulman and Bermingham, 1995). This work, however, was not completed in areas of known self-recruitment and differentiation in marine environments may be subtle, requiring larger sample sizes to discern. Further, it has been suggested that a greater reliance on rapidly evolving markers, such as microsatellites or examination of patterns of isolation by distance, may allow insight into ecological timescale processes in high-gene flow

species (Mora and Sale, 2002; Palumbi, 2003). In the current study, we test the hypothesis that short-term self-recruitment in the bluehead wrasse is coupled across evolutionary timescales to processes affecting population genetic structure using rapidly evolving markers from both the mitochondrial and nuclear genomes. We perform this test on the same spatial scale as documented self-recruitment on St Croix, using samples drawn from the same localities included in the otolith study of Swearer *et al.* (1999), as well as on larger scales across the species range, to more fully investigate limits on gene flow in this species.

When using molecular markers in an assessment of population connectivity, it is also important to consider how factors other than migration can impact the distribution of genetic variation (Rand, 1995; Skibinski, 2000; Ford, 2002). Selection and demographic fluctuations may act on markers assumed *a priori* to behave in a neutral manner and should be tested for and considered when inferring levels of migration from population structure revealed by molecular data. Therefore, we test whether the markers sampled for this study are evolving in a manner consistent with neutrality and/or have been subject to past demographic perturbations from mutation-drift equilibrium, possibly affecting estimates of current levels of gene flow.

An investigation of the population genetics of *T. bifasciatum* thus provides an excellent opportunity to examine the interface between independent sources of data, as well as testing for genetic differentiation in a reef fish with both a high potential for dispersal through a long pelagic larval stage and a propensity for local recruitment. Knowledge of the partitioning of genetic variation owing to processes operating on variable timescales may provide a link between ecological and evolutionary studies and provide insight into the processes that act to segregate genetic variation in the sea.

## Materials and methods

### Sampling and laboratory data collection
Individual fish were sampled from nearshore reefs with baited nets. Whole individuals or tissue samples were preserved in 60% ethanol or frozen for transport to the laboratory. Sampling was designed to include comparisons at a small spatial scale, with two sites on the leeward and two sites on the windward sides of the island of St Croix. This is the same scale as that of self-recruitment detected by otolith data (Swearer *et al.*, 1999). As the magnitude of self-recruitment documented at this scale, although substantial, allowed for sufficient exchange of migrants to homogenize gene frequencies among populations, we would expect no genetic differentiation to occur. Hence, two additional areas (four sites) were sampled at progressively larger distances to investigate gene flow on scales similar to the total range of the species, which spans the area from Belize and Mexico's Yucatan peninsula in the West to Barbados in the east, and from northern South America to Bermuda (Table 1, Figure 1). For the largest scale of comparison, that between western and eastern Caribbean sites, we pooled the western populations as sample sizes for these sites were lower than those in the east and

**Table 1** Sample sizes and variability for mitochondrial DNA sequence data

| Site | Map #-Fig. 1 | n-CR | h-CR | π-CR | k |
|---|---|---|---|---|---|
| Butler Bay, St Croix | 1 | 49 | 0.993 | 0.020 | 7.33 |
| Grapetree Bay, St Croix | 2 | 51 | 0.987 | 0.019 | 6.36 |
| Jack Bay, St Croix | 3 | 47 | 0.989 | 0.021 | 7.36 |
| Northstar Bay, St Croix | 4 | 51 | 0.981 | 0.018 | 6.96 |
| San Salvador, Bahamas | 5 | 39 | 0.982 | 0.019 | 7.23 |
| South Water Caye, Belize | 6 | 27[a] | 0.994[a] | 0.020[a] | 7.51 |
| Ambergris Caye, Belize | 7 | | | | |
| Cancun, Mexico | 8 | | | | |

[a]Sample size, gene and nucleotide diversity is shown for the pooled Western sample.
Columns indicate the number of haplotypes sampled for each location, gene diversity (h), nucleotide diversity (π) and average number of pairwise differences (k).
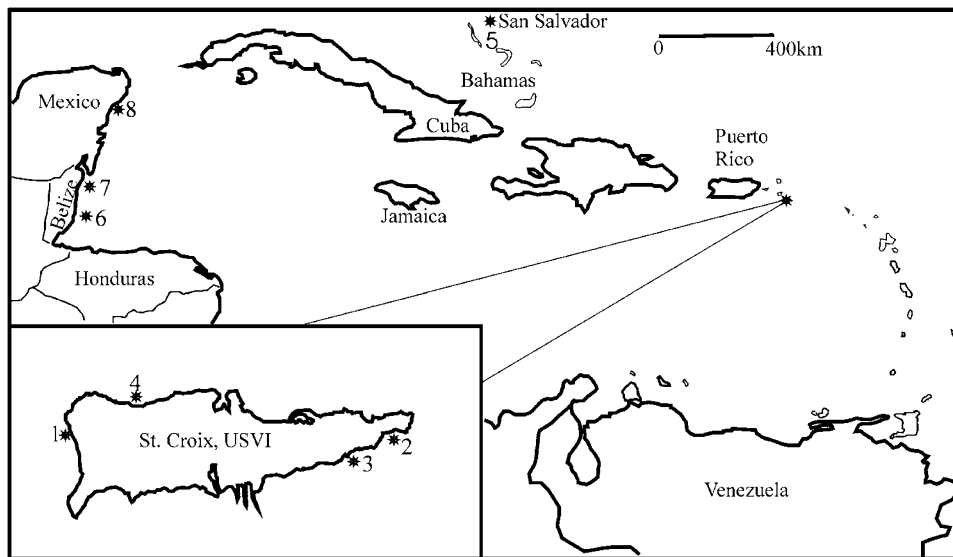


**Figure 1** Map of the Caribbean with sampling sites indicated by *. Inset displays sites sampled around the island of St Croix.

a preliminary analysis indicated no evidence of significant differentiation between them.

Genomic DNA was extracted from gill or muscle tissue by digestion with proteinase-K, followed by aqueous extraction with phenol/chloroform isoamyl alcohol and precipitation with 95% ethanol overnight. DNA from a small subset of samples was extracted using a DNeasy tissue kit (Qiagen, Rockville, MD, USA).

PCR amplification of a portion of the mitochondrial control region, including the hypervariable d-loop, was performed with a combination of three primers (A,K,E) described in Lee et al. (1995) in 50 μl reaction volumes consisting of 10× PCR buffer (1.5 mM MgCl$_2$), 10 mM dNTPs, 10 μM primers, 1 U Taq DNA polymerase and 4 μl of template. The thermal profile for the amplification was as follows: initial denaturation at 94°C for 2–4 min, 30–40 cycles at 94°C for 30 s, 50–57°C for 30 s and 72°C for 45 s, followed by a final extension at 72°C for 4–6 min. Amplification of microsatellite loci described in Wooninck et al. (1998) and Williams et al. (2004) was achieved with species-specific primers, the reaction ingredients listed above and the following amplification profiles: (1) 94°C for 2 min, 35–45 cycles at 94°C for 10–30 s, 50–58°C for 10–30 s and 72°C for 20–45 s, followed by a final extension at 72°C for 15 min (2) 94°C for 2 min, 94°C for

10 s, 50–53°C for 10 s, 72°C for 20 s for 10 cycles followed by 30–35 cycles at 94°C for 10 s, 53–55°C for 10 s, 72°C for 20 s, with a final extension of 15 min.

PCR products were electrophoresed on ABI377 or ABI3730 automated sequencers. For mtDNA, forward and reverse sequences were assembled in Sequencher v. 4.0 (Gene Codes Corporation, Ann Arbor, MI, USA) for the first 143 individuals. Owing to high sequence quality, short read length and the lack of ambiguous bases, the remaining 121 PCR products were sequenced in the forward direction only.

### Data analysis

Consensus sequences were imported into Bioedit (Hall, 1999) and saved into NBRF/PIR format. An initial sequence alignment was then obtained in Clustal X (Thompson et al., 1997), after which imported sequences were aligned by eye. A representative sequence for each different haplotype was deposited in Genbank under accession numbers DQ913896–DQ914071. Microsatellite genotypes were compiled in Microsoft excel and converted into appropriate formats for further analysis.

Variability in the mitochondrial sequence data was assessed by haplotype and nucleotide diversity and the

average number of nucleotide differences among sequences. Values were calculated in DNAsp v. 3.99 (Rozas and Rozas, 1999). For the microsatellite data, expected heterozygosity ($H_E$), and allelic richness (AR), a standardized measure of number of alleles sampled per population, were calculated in Fstat v. 2.9.3 (Goudet, 2001).

For the mitochondrial sequence data, estimation of the degree of genetic differentiation among subpopulations was based on $F_{st}$ incorporating Tamura-Nei distances among sequences. Both pairwise values and the overall average across all populations were calculated in Arlequin v. 2.000 (Schneider et al., 2000). Nonparametric tests of the null hypothesis of no differentiation were performed for each pairwise comparison, with significance of $F_{st}$ values determined by comparing observed values to a distribution obtained by permuting haplotypes among populations 10 000 times. A standard Bonferroni correction for multiple tests was applied. As the exact mutation model for the microsatellite data was unknown, both $F_{st}$ and $R_{st}$ values were calculated for comparison. $F_{st}$ values were estimated by $\theta$ (Weir and Cockerham, 1984) in Fstat v. 2.9.3 (Goudet, 2001). $R_{st}$ was estimated using RST CALC v. 2.2 (Goodman, 1997). Exact tests of differentiation were performed for all pairwise comparisons and over all populations simultaneously using the likelihood G-statistic with significance determined by 15 000 permutations of genotypes among populations (Goudet et al., 1996) in Fstat v. 2.9.3 (Goudet, 2001).

We further tested for the presence of population structure in the microsatellite data using the Bayesian clustering method implemented in Structure version 2.1 (Pritchard et al., 2000). Three independent runs of a Markov chain of 10 00 000 generations after a burn-in of 1 00 000 generations were performed for putative population numbers from 1 to 10 under a correlated allele frequencies model. The posterior probability of a given number of populations was then calculated via Baye's rule.

A test of the possible presence of isolation-by-distance in both data sets was performed using IBD v. 1.5 (Bohonak, 2002). This analysis assumes a stepping-stone model of gene flow in which genetic exchange is higher for spatially proximate populations. In marine species for which genetic differentiation is low relative to sampling variance, a signal of isolation-by-distance may be a more senstive and robust indicator of limits to dispersal (Palumbi, 2003). For the mitochondrial sequence data, average pairwise Tamura-Nei $\gamma$-corrected distances among populations were calculated in MEGA v. 2.0 (Kumar et al., 2001) and regressed on straight-line distance between the populations using both untransformed and log-transformed data. Microsatellite data was analyzed for all loci combined, with $(\delta\mu)^2$ distances (Goldstein et al., 1995) calculated in RSTCALC (Goodman, 1997) used as input and regressed on straight-line distances among sites.

DNAsp v. 3.99 (Rozas and Rozas, 1999) was used to test for deviation from a neutral model of molecular evolution in the mitochondrial sequence data using Tajima's $D$ and Fu and Li's $F^*$ and $D^*$. One thousand replicates of a coalescent simulation were used to determine the empirical distributions of the test statistics and to calculate confidence intervals and $P$-values for observed values of each. Population demography was investigated by visual inspection of the mismatch distribution and by testing a null model of constant population size by calculating the raggedness value (Harpending, 1994), Fu's $F_s$ (Fu, 1997) and $R_2$ (Ramos-Onsins and Rozas, 2002) and comparing each statistic to its empirical distribution generated by 1000 replicates of a coalescent simulation in DNAsp v. 3.99.

Fstat version 2.9.3 was used to test for Hardy–Weinberg equilibrium in the microsatellite data within population samples using $f$ as an estimator of $F_{is}$ (Weir and Cockerham, 1984). Significance was determined by comparison to a distribution of the test statistic generated by 5000 randomizations of alleles among individuals for each locus. A standard Bonferroni multiple test correction was applied to determine the significance level for each test. Linkage equilibrium among all pairs of loci was tested by calculating the likelihood-ratio G-statistic and comparing it to a distribution estimated by randomly associating genotypes at the two loci being tested.

The microsatellite data was used to perform a test for changes in population size owing to a bottleneck or expansion with the program Bottleneck (Cornuet and Luikart, 1996). The method implemented in this program takes advantage of the theoretical expectation that the number of alleles will be reduced to a greater extent than the gene diversity during a population size reduction owing to the preferential loss of rare alleles. Gene diversity was compared to a distribution of $H_{eq}$ obtained by coalescent simulation of a stable population at mutation-drift equilibrium under both the infinite-allele (IAM) and stepwise mutation (SMM) models conditioned on the number of alleles over 1000 replications for each locus. A Wilcoxon sign–rank test was performed to determine if a significant excess of loci showed an excess of gene diversity indicating a population expansion or bottleneck.

Lastly, as rapidly evolving microsatellite loci may be prone to null alleles (O'Connell and Wright, 1997; Dakin and Avise, 2004), we test for their presence in the microsatellite loci sampled for this study using Micro-checker v. 2.2.1 (van Oosterhout et al., 2004). Genotype frequencies were adjusted for the presence of segregating null alleles for loci in each population following Brookfield (1996). These adjusted genotypes were then used to calculate test statistics, which were then compared to the original values.

## Results

### Molecular marker data

A total of 264 sequences of the 5' end of the mitochondrial control region were obtained from individuals sampled at eight locations (Table 1). A total of 176 different haplotypes were recovered. The alignment consisted of 380 sites of which 94 were variable. There were zero gaps in the alignment. Haplotype and nucleotide diversity were 0.986 and 0.021, respectively, for the whole data set, values for individual populations were similar to the overall averages (Table 1).

Five microsatellite loci, T3235 (Wooninck et al., 1998), TbAAC50, TbAAT4, TbAAT49 and TbAAT42 (Williams et al., 2004) were scored for individuals from each

**Table 2** Sample sizes, variability measures, inbreeding coefficients and estimated null allele frequencies per locus and population for five microsatellite loci

| Site | T3235 | | | | | Tb50 | | | | | Tb49 | | | | | Tb42 | | | | | Tb4 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | AR | $H_E$ | f | nf | N | AR | $H_E$ | f | nf | N | AR | $H_E$ | f | nf | N | AR | $H_E$ | f | nf | N | AR | $H_E$ | f | Nf |
| Butler | 48 | 34 | 0.98 | 0.130* | 0.126 | 49 | 31 | 0.98 | 0.235* | 0.176 | 49 | 9 | 0.77 | 0.211 | 0.173 | 43 | 19 | 0.94 | 0.187* | 0.287 | 51 | 19 | 0.94 | 0.152* | 0.068 |
| Grapetree | 52 | 32 | 0.98 | 0.174* | 0.115 | 52 | 33 | 0.98 | 0.179* | 0.146 | 51 | 7 | 0.77 | 0.192 | 0.194 | 47 | 17 | 0.92 | 0.170* | 0.274 | 53 | 17 | 0.94 | 0.116 | 0.090 |
| Jack | 46 | 35 | 0.98 | 0.046 | | 48 | 34 | 0.98 | 0.110* | 0.086 | 48 | 7 | 0.74 | 0.188 | 0.128 | 45 | 16 | 0.93 | 0.191* | 0.216 | 48 | 17 | 0.93 | 0.153* | 0.109 |
| Northstar | 50 | 33 | 0.98 | 0.185* | 0.150 | 50 | 31 | 0.98 | 0.185* | 0.178 | 50 | 9 | 0.78 | 0.216 | 0.203 | 43 | 16 | 0.93 | 0.125 | 0.301 | 499 | 20 | 0.94 | 0.074 | |
| Sansal | 37 | 33 | 0.98 | 0.236* | 0.226 | 39 | 31 | 0.98 | 0.264* | 0.124 | 38 | 9 | 0.82 | 0.195 | 0.139 | 36 | 16 | 0.93 | 0.281* | 0.248 | 38 | 19 | 0.94 | 0.136 | 0.108 |
| West | 27 | 33 | 0.98 | 0.133* | 0.056 | 27 | 34 | 0.97 | 0.011 | | 27 | 7 | 0.75 | 0.210 | | 27 | 17 | 0.94 | 0.095 | | 27 | 16 | 0.94 | 0.251* | 0.112 |

Abbreviations: AR, allelic richness; $H_E$, expected heterozygosity; nf, null allele frequency.
Shown are number of diploid individuals analyzed (N), $F_{is}$ as estimated by f (Weir and Cockerham, 1984). Blank entries indicate that evidence for null alleles was not found for that particular locus/population combination. Significance levels of population-by-locus tests of Hardy–Weinberg equilibrium (HWE) were determined by 5000 randomizations of alleles among individuals for each locus. Values significant after standard Bonferroni correction are marked with an asterisk ($P < 0.0017$).

**Table 3** $F_{st}$ values for mitochondrial sequence data are below diagonal, $F_{st}$ averaged across five microsatellite loci above

| | Butler | Grapetree | Jack | Northstar | San Sal | West |
|---|---|---|---|---|---|---|
| Butler | 0.0000 | 0.0002 | −0.0047 | 0.0002 | −0.0007 | −0.0017 |
| Grapetree | −0.0028 | 0.0000 | −0.0015 | 0.0006 | 0.0022 | −0.0008 |
| Jack | −0.0080 | 0.0010 | 0.0000 | 0.0010 | −0.0001 | −0.0006 |
| Northstar | −0.0035 | 0.0065 | −0.0040 | 0.0000 | 0.0018 | −0.0013 |
| San Sal | 0.0009 | −0.0059 | −0.0043 | 0.0028 | 0.0000 | −0.0033 |
| West | −0.0057 | −0.0142 | −0.0085 | −0.0015 | −0.0054 | 0.0000 |

Values for mitochondrial sequence data were computed using a Tamura–Nei distance measure with $\gamma$ correction ($\alpha = 0.5$) based on 380 bp of mitochondrial control region sequence data.

population. Locus T3237 (Wooninck et al., 1998) did not amplify for a majority of individuals and the data are not analyzed herein. Locus T3231 (Wooninck et al., 1998) did not produce any products in the first 50 individuals attempted. Further amplifications were not attempted for this locus. Loci scored were highly variable, with the number of alleles ranging from 14 for locus TbAAT49 to 78 for locus T3235 and gene diversity ranging from 0.75 for TbAAT49 to 0.98 for T3235 and TbAAC50, across all populations. Sample sizes for individual loci and levels of variability were similar for each population (Table 2). Approximately 50% of population-by-locus tests for Hardy–Weinberg equilibrium were significant (Table 3). Five out of six populations exhibited significant deviations towards excess homozygosity after Bonferroni correction for loci T3235 and TbAAC50, three and four out of six were significant in this direction for loci TbAAT4 and TbAAT42, respectively, whereas no tests were significant for locus TbAAT49, although f was positive, suggesting a deficiency of zygotes at this locus. No comparisons between loci for linkage disequilibrium were significant.

### Genetic differentiation
For the mitochondrial control region data, no evidence of population structuring was found. Averaged across all population comparisons, $F_{st}$ was −0.003, whereas values ranged from negative values in most pairwise comparisons to a maximum of 0.006 between Northstar and Grapetree Bays, St Croix (Table 3). No comparisons were significant based on 10 000 permutations of haplotypes among populations. Regression of average pairwise genetic distances between populations on geographic

distance between sites was not significant, either with untransformed ($r^2 = 0.045$, $P = 0.264$) or log-transformed ($r^2 = 0.036$, $P = 0.242$) data.

The combined microsatellite data set revealed no evidence of differentiation among populations. Averaging over loci, the value of $F_{st}$ was 0.000, whereas $R_{st}$ over all population samples was −0.003. Pairwise values of $F_{st}$ ranged from −0.005 to 0.002 (Table 3). Pairwise values of $R_{st}$ and $F_{st}$ between given subpopulations were similar, as $R_{st}$ ranged from −0.007 to 0.009. A test of genetic differentiation over all subpopulations and loci using the likelihood G-statistic and not assuming HWE within samples was also not significant ($P = 0.291$), nor were any pairwise comparisons (P-values ranging from 0.17 to 0.96). Posterior probabilities for a given number of hypothesized populations between one and 10 were calculated from likelihood values obtained from three independent runs of 10 00 000 generations for hypothesized numbers of populations from 1–10 using the Bayesian clustering algorithm in Structure 2.1. A model with a single population has by far the highest likelihood and a posterior probability of 1. Regression of $(\delta\mu)^2$ values on geographic distance was not significant for untransformed values ($r^2 = 0.033$, $P = 0.361$) or for log-transformed data ($r^2 = 0.0009$, $P = 0.458$).

### Mutation-drift equilibrium
As there was no genetic differentiation among populations, test statistics for mutation-drift equilibrium were computed for the complete control region sequence data set. Tajima's D was significant ($P = 0.009$) at −1.79. Fu and Li's F* (−3.67, $P = 0.003$) and D* (−4.24, $P = 0.001$) were both significant. Inspection of the distribution of

pairwise differences between sequences revealed a strong, sharp peak with a smaller secondary peak (data not shown). The observed raggedness value ($r = 0.0093$) was not significant ($P = 0.298$), but both $R_2$ and $F_s$ were significant ($R_2 = 0.037$, $P = 0.02$; $F_s = -286.88$, $P = 0.000$).

For the microsatellite data, populations were pooled into one for analysis as no genetic differentiation was apparent. Under the IAM, the Wilcoxson sign–rank test shows an excess of loci with $H_{obs}$ greater than the expected value ($H_{eq}$) under stable population size simulations ($P = 0.016$; 0 loci with heterozygosity deficiency, 5 with heterozygosity excess). Under the SMM, the two most variable loci had to be removed from the analysis, as the program was unable to compute the likelihood distribution of $\theta$ given the number of alleles and sample size. The Wilcoxon test shows no excess of loci with heterozygosity excess under this model ($P = 1.000$), but a test for heterozygosity deficiency is close to 95% significance ($P = 0.063$). The results are the same if the dataset adjusted for null alleles is used in the analysis.

### Null alleles

All loci included in this study appear to have null alleles segregating in at least five of the six populations at frequencies up to 30% (Table 2). We adjusted for putative null allele homozygotes on a locus-by-locus basis by coding them as missing data based on the adjusted genotypes output by Microchecker v. 2.2.1 and used this dataset for analysis of genetic differentiation. Pairwise values of $F_{st}$ and $R_{st}$ were very similar to those for the original data. A test of genetic differentiation over all subpopulations and loci using the likelihood G-statistic was not significant ($P = 0.935$), nor were any pairwise values ($P$-values ranging from 0.23 to 0.99). Regression of $(\delta\mu)^2$ values on geographic distance was not significant for untransformed values ($r^2 = 0.0023$, $P = 0.306$) or for log-transformed data ($r^2 = 0.000013$, $P = 0.376$).

## Discussion

As in previous studies using whole mtDNA RFLP data (Shulman and Bermingham, 1995) or allozymes (Lacson, 1992), highly variable mitochondrial sequence data and microsatellites analyzed in the current study showed no evidence of differentiation between *T. bifasciatum* populations on either small or large scales across the Caribbean.

These results also indicate that *T. bifasciatum*, a species with substantial local recruitment on ecological time-scales, as inferred from otolith microchemistry (Swearer et al., 1999), may not have structured populations when viewed on longer evolutionary timescales. Yet these results are complementary as opposed to discordant and highlight the need for studies incorporating a pluralistic approach to data collection to capture information reflecting both ecological and evolutionary timescales in an assessment of dispersal.

A finding of substantial local recruitment but lack of genetic differentiation could arise under two distinct scenarios. First, even if gene flow were measured by neutral molecular markers in a system at mutation-drift equilibrium, it may still be the case that different types of data (otolith versus molecular) allow inference into processes operating on specific temporal scales. Pro-

cesses such as ecological-scale local recruitment may simply not be temporally persistent enough to affect population genetic structure, which represents a longer term average pattern and may still indicate high-gene flow if levels of local recruitment are both high and variable. This is likely the case if patterns of recruitment are driven by variability in small-scale physical oceanographic processes, as was suggested for St Croix (Swearer et al., 1999). In addition, levels of self-recruitment on small spatial scales, although substantial by ecological standards, may simply not be large enough in magnitude to affect genetic structure, as even small numbers of migrants between populations are sufficient to homogenize haplotype or allele frequencies (Wright, 1931). Even comparatively rare dispersal events could easily lead to panmixis as assayed by molecular markers (Palumbi, 2003; Kinlan et al., 2005).

Second, it is also possible that molecular markers may not be neutral or that the species gene pool may represent a nonequilibrium system owing to historical demographic events such as population expansion. Genetic homogeneity among populations in this case may not reflect patterns of current migration, but rather historical processes. Under such a scenario, even if self-recruitment is sufficient to lead to genetic differentiation, such a pattern may not yet be visible in the molecular data. The influence of historical population growth has been invoked to explain patterns of mitochondrial sequence variation for several species of reef fish in the Indo–Pacific (Fauvelot et al., 2003; Chen et al., 2004; Thacker, 2004) and also recently for Gulf of Mexico red snapper (*Lutjanus campechanus*) populations (Pruett et al., 2005). Such events may be common, even in tropical reef organisms inhabiting presumably stable environments (Benzie, 1999).

There is some evidence that *T. bifasciatum* has experienced an historical population expansion. The mtDNA of *T. bifasciatum* deviates from a neutral model at the population level. Tajima's $D$ and Fu and Li's $F^*$ and $D^*$ were all significantly negative. This result could be owing to the action of background selection against mildly deleterious mutations in mtDNA, a predominant mode of evolution for this marker and an equilibrium process (Rand, 2001), but could also be the result of a selective sweep or population expansion. The mismatch distribution is not smoothly unimodal nor is the raggedness value significant, as would be expected during population expansion. However, the $R_2$ and $F_s$ statistics were significant, and are far more powerful tests under a range of values for expansion parameters (Ramos-Onsins and Rozas, 2002). The statistics used could also be significant owing to recovery from a selective sweep of the mtDNA, but historical fluctuations in population size should affect all genomic regions and evidence from microsatellite data indicates a possible demographic signature in the nuclear genome. The gene diversity excess exhibited under an infinite–alleles model is consistent with a weak bottleneck and subsequent expansion in the evolutionary history of *T. bifasciatum*. In contrast, a possible heterozygosity deficiency occurs under a stepwise-mutation model. Such a deficiency under this model is consistent with a population bottleneck based on simulations (Cornuet and Luikart, 1996). However, this effect appears to occur only at allele numbers well below those observed for the loci in this

study and our result under this model is based on only three loci. Yet, as multiple loci deviate in a manner consistent with a population expansion and other Caribbean and Indo–Pacific reef fish species show similar patterns of mtDNA variation, a population expansion related to broad-scale climatic fluctuations seems a more likely explanation than multiple selective events. Additional data from nuclear genes, especially DNA sequence data, might help resolve more clearly whether past fluctuations in population size have affected the gene pool of *T. bifasciatum*.

Lastly we must consider that highly variable microsatellites, while providing power to discriminate subtle variation, are prone to null alleles, possibly owing to mutations occurring in primer regions (Brookfield, 1996; O'Connell and Wright, 1997). The presence of nonamplifying alleles may affect the results of population genetic analyses (Dakin and Avise, 2004), although there is no reason to suspect that their presence would bias towards finding high-gene flow. For example, other studies that made use of microsatellite data to examine population structure in marine and freshwater organisms have uncovered significant genetic structure in the presence of heterozygosity deficiency that was possibly caused by null alleles (O'Reilly et al., 2004; Pampoulie et al., 2004; Therriault et al., 2005). In the O'Reilly et al. (2004) study, adjusting for null alleles actually led to a decrease in the number of significant pairwise tests of genetic differentiation. To examine whether taking null alleles into account might influence the level of gene flow seen in this study, we adjusted for null alleles and recalculated estimates of genetic differentiation. The finding of high-gene flow among Caribbean populations of *T. bifasciatum* remained after analysis of adjusted genotypes. In addition, we tested models for a range of numbers of distinct populations in our data in Structure 2.1. A model of a single population was strongly supported by the data. Yet deviations from HWE within populations, as seen in our study, tend to bias towards finding structure in the Bayesian clustering algorithm employed in this program and this bias is more pronounced under the correlated alleles model we employed in the analysis. These results further suggest that the lack of structure in analysis of microsatellite data is not simply owing to the presence of null alleles, although additional loci without a significant proportion of segregrating null alleles would be necessary to confirm this.

Population structure in reef fish species has now been documented on variable spatial scales with several different classes of molecular markers (Shulman and Bermingham, 1995; Bernardi et al., 2001; Muss et al., 2001; Planes et al., 2001; Rocha et al., 2002; Taylor and Hellberg, 2003; Chen et al., 2004; Ovenden et al., 2004; Thacker, 2004; Hoffman et al., 2005), yet the species studied to date represent only a small proportion of the tropical reef fish fauna. Thus it is not yet possible to reach an estimate of what proportion of species exhibit limits to dispersal over evolutionary timescales that are witnessed as genetic differentiation. In the Caribbean, reef fish species exhibiting genetically differentiated populations may be the exception, in spite of recent studies suggesting quite limited dispersal (Cowen et al., 2006). The current study, along with a previous analysis of mitochondrial variation in *Lutjanus campechanus*, suggest a possible reason for

a lack of genetic differentiation in some species, the existence of nonequilibrium gene pools owing to demographic factors, possibly obscuring current gene flow and migration patterns inferred from molecular markers. For these species and others that fall into this category, ecological scale estimates of dispersal take on added importance for the design of marine reserves.

It is as yet unclear what contribution factors such as pelagic larval duration make to genetic differentiation (Shulman and Bermingham, 1995; Fauvelot and Planes, 2002), and finding a correlation between these factors may be complicated by the influence of historical population expansion on the distribution of genetic variation. *T. bifasciatum*, with pelagic eggs and relatively long pelagic larval duration (mean of approximately 45–49 days: Victor, 1986; Philbotte, 2002), may exist at one end of a continuum of dispersal potential. Another member of the genus *Thalassoma*, *T. hardwicki*, has similar life-history parameters related to dispersal in the marine environment (pelagic eggs and a pelagic larval period of 49 days) and also shows little population structure as assessed by mitochondrial control region sequence data (Chen et al., 2004). As the accumulation of data on gene flow in reef fish continues, correlations between features of organismal life history and levels of genetic differentiation may become more apparent.

## Acknowledgements

## References

Atema J, Kingsford MJ, Gerlach G (2002). Larval reef fish could use odour for detection, retention and orientation to reefs. *Mar Ecol Prog Ser* **241**: 151–160.

Benzie JAH (1999). Genetic structure of coral reef organisms: ghosts of dispersal past. *Am Zool* **39**: 131–145.

Bernardi G, Holbrook SJ, Schmitt RJ (2001). Gene flow at three spatial scales in a coral reef fish, the three-spot dascyllus, *Dascyllus trimaculatus*. *Mar Biol* **138**: 457–465.

Birkeland C (1997). Symbiosis, fisheries and economic development on coral reefs. *Trends Ecol Evol* **12**: 364–367.

Bohonak AJ (2002). IBD (isolation by distance): a program for analyses of isolation by distance. *J Hered* **93**: 153–154.

Botsford LW, Micheli F, Hastings A (2003). Principles for the design of marine reserves. *Ecol Appl* **13**: S25–S31.

Brookfield JFY (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* **5**: 453–455.

Caley MJ, Carr MH, Hixon MA, Hughes TP, Jones GP, Menge BA (1996). Recruitment and the local dynamics of open marine populations. *Annu Rev Ecol Syst* **27**: 477–500.

Chen CA, Ablan MCA, McManus JW, Bell JD, Tuan VS, Cabanban AS et al. (2004). Population structure and genetic variability of six bar wrasse (*Thallasoma hardwicki*) in northern South China Sea revealed by mitochondrial control region sequences. *Mar Biotechnol* **6**: 312–326.

Cornuet JM, Luikart G (1996). Description and evaluation of two tests for detecting recent bottlenecks. *Genetics* **144**: 2001–2014.

Cowen RK, Paris CB, Srinivasan A (2006). Scaling of connectivity in marine populations. *Science* **311**: 522–527.

Dakin EE, Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity* **93**: 504–509.

Ehrlich PR (1975). The population biology of coral reef fishes. *Annu Rev Ecol Syst* **6**: 211–248.

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

Fauvelot C, Planes S (2002). Understanding origins of present day genetic structure in marine fish: biologically or historically driven patterns? *Mar Biol* **141**: 773–788.

Ford MJ (2002). Applications of selective neutrality tests to molecular ecology. *Mol Ecol* **11**: 1245–1262.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Gaines SD, Gaylord B, Largier JL (2003). Avoiding current oversights in marine reserve design. *Ecol Appl* **13**: S32–S46.

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463–471.

Goodman SJ (1997). RST CALC: a collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Mol Ecol* **6**: 881–885.

Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). URL http://www.unil.ch/izea/softwares/fstat.html.

Goudet J, Raymond M, Demeeus T, Rousset F (1996). Testing differentiation in diploid populations. *Genetics* **144**: 1933–1940.

Hall T (1999). Bioedit sequence alignment editor. Department of Microbiology, North Carolina State University. URL http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Harpending HC (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* **4**: 591–600.

Hoffman EA, Kolm N, Berglund A, Arguello JR, Jones AG (2005). Genetic structure in the coral-reef-associated Banggai cardinalfish, *Pterapogon kauderni*. *Mol Ecol* **14**: 1367–1375.

Hutchings JA (1996). Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. *Can J Fish Aquat Sci* **53**: 943–962.

Jones GP, Milicich MJ, Emslie MJ, Lunow C (1999). Self-recruitment in a coral reef fish population. *Nature* **402**: 802–804.

Kinlan BP, Gaines SD, Lester SE (2005). Propagule dispersal and the scales of marine community process. *Diversity Distrib* **11**: 139–148.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* **17**: 1244–1245.

Lacson JM (1992). Minimal genetic variation among samples of six species of coral reef fishes collected at La Parguera, Puerto Rico, and Discovery Bay, Jamaica. *Mar Biol* **112**: 327–331.

Lee W, Conroy J, Howell WH, Kocher TD (1995). Structure and evolution of teleost mitochondrial control region. *J Mol Evol* **41**: 54–66.

Leis JM (2002). Pacific coral-reef fishes: the implications of behaviour and ecology of larvae for biodiversity and conservation and a reassessment of the open population paradigm. *Environ Biol Fish* **65**: 199–208.

Levin SA (1992). The problem of pattern and scale in ecology. *Ecology* **73**: 1943–1967.

Mora C, Sale PF (2002). Are populations of coral reef fish open or closed? *Trends Ecol Evol* **17**: 422–428.

Muss A, Robertson DR, Stepien CA, Wirtz P, Bowen BW (2001). Phylogeography of *Ophioblennus*: the role of ocean currents and geography in reef fish evolution. *Evolution* **55**: 561–572.

Myers RA, Hutchings JA, Barrowman NJ (1997). Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecol Appl* **7**: 91–106.

O'Connell M, Wright JM (1997). Microsatellite DNA in fishes. *Rev Fish Biol Fish* **7**: 331–363.

O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004). Inverse relationship between $F_{st}$ and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Mol Ecol* **13**: 1799–1814.

Ovenden JR, Salini J, O'Connor S, Street R (2004). Pronounced genetic population structure in a potentially vagile fish species (*Pristipomoides multidens*, Teleostei: Perciformes: Lutjanidae) from the East Indies triangle. *Mol Ecol* **13**: 1991–1999.

Palumbi SR (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecol Appl* **13**: S146–S158.

Pampoulie C, Gysels ES, Maes GE, Hellemans B, Leentjes V, Jones AG *et al.* (2004). Evidence for fine-scale genetic structure and estuarine colonisation in a potential high gene flow marine goby (*Pomatoschistus minutus*). *Heredity* **92**: 434–445.

Philbotte J (2002). Pelagic larval duration of the Caribbean Wrasse, *Thalassoma bifasciatum*. *Biol Bull* **203**: 245–246.

Planes S, Doherty PJ, Bernardi G (2001). Strong genetic divergence among populations of a marine fish with limited dispersal, *Acanthochromis polyacanthus*, within the Great Barrier Reef and the Coral Sea. *Evolution* **55**: 2263–2273.

Planes S, Parroni M, Chauvet C (1998). Evidence of limited gene flow in three species of coral reef fishes in the lagoon of New Caledonia. *Mar Biol* **130**: 361–368.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pruett CL, Saillant E, Gold JR (2005). Historical population demography of red snapper (*Lutjanus campechanus*) from the northern Gulf of Mexico based on the analysis of sequences of mitochondrial DNA. *Mar Biol* **147**: 593–602.

Ramos-Onsins SE, Rozas J (2002). Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**: 2092–2100.

Rand DM (1995). Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. *Conserv Biol* **10**: 665–671.

Rand DM (2001). The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst* **32**: 415–448.

Rocha LA, Bass AL, Robertson DR, Bowen BW (2002). Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Mol Ecol* **11**: 243–252.

Rocha LA, Robertson DR, Roman J, Bowen BW (2005). Ecological speciation in tropical reef fishes. *Proc R Soc B* **272**: 573–579.

Rozas J, Rozas R (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.

Sale PF (2002). The science we need to develop for more effective management. In: Sale PF (ed) *Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem*. Academic Press: San Diego. pp 361–376.

Schiermeier Q (2002). How many more fish in the sea? *Nature* **491**: 662–665.

Schneider S, Roessli D, Excoffier L (2000). Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. URL http://lgb.unige.ch/arlequin/.

Shulman MJ, Bermingham E (1995). Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* **49**: 897–910.

Skibinski DOF (2000). DNA tests of neutral theory: applications in marine genetics. *Hydrobiologia* **420**: 137–152.

Stobutzki IC (2001). Marine reserves and the complexity of larval dispersal. *Rev Fish Biol Fish* **10**: 515–518.

Swearer SE, Caselle JE, Lea DW, Warner RR (1999). Larval retention and recruitment in an island population of a coral reef fish. *Nature* **402**: 799–802.

Taylor MS, Hellberg ME (2003). Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science* **299**: 107–109.

Thacker CE (2004). Population structure in two species of the reef goby *Gnatholepis* (Teleostei: Perciformes) among four South Pacific island groups. *Coral Reefs* **23**: 357–366.

Therriault TW, Orlova MI, Docker MF, MacIsaac HJ, Heath DD (2005). Invasion genetics of a freshwater mussel (*Dreissena rostriformis bugensis*) in eastern Europe: high gene flow and multiple introductions. *Heredity* **95**: 16–23.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004). Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.

Victor BC (1986). Duration of the planktonic larval stage of one hundred species of Pacific and Atlantic wrasses (family Labridae). *Mar Biol* **90**: 317–326.

Warner RR, Cowen RK (2002). Local retention of production in marine populations: evidence, mechanisms, and consequences. *Bull Mar Sci* **70**: 245–249.

Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Williams DA, Purcell J, Cowen RK, Hughes CR (2004). Characterization of microsatellite multiplexes for population genetic studies of bluehead wrasse (*Thalassoma bifasciatum*, Pisces: Labridae). *Mol Ecol Notes* **4**: 525–527.

Wooninck L, Strassman JE, Fleischer RC, Warner RR (1998). Characterization of microsatellite loci in a pelagic spawner: the bluehead wrasse, *Thalassoma bifasciatum*. *Mol Ecol* **7**: 1613–1621.

Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.