AN ABSTRACT OF THE DISSERTATION OF

Mark R. Christie for the degree of Doctor of Philosophy in Zoology presented on September 4, 2009.

Title: Larval Dispersal in Marine Fishes: Novel Methods Reveal Patterns of Self-Recruitment and Population Connectivity

Abstract approved:

Mark A. Hixon

Many marine fish populations are severely declining due to over-fishing, loss of both juvenile and adult habitats, and accelerating environmental degradation. Fisheries management and the implementation of marine protected areas (MPAs) and other conservation tools are currently hindered by large gaps in knowledge about larval dispersal and its subsequent effects on population dynamics and regulation. This lack of knowledge is due to the inherent difficulty associated with tracking miniscule marine fish larvae. Population genetics approaches are particularly promising, but current methods have been of limited use for inferring ecologically relevant rates of population connectivity because of the large population sizes and high amounts of gene flow present in most marine species.

To address these issues, I developed novel genetic methods of identifying parent-offspring pairs to directly track the origin and settlement of larvae in natural populations. These parentage methods fully account for large numbers of pair-wise

comparisons and do not require any demographic assumptions or observational data. Furthermore, these methods can be used when only a small proportion of candidate parents can be sampled, which is often the case in large marine populations. I also employed Bayes' theorem to take into account the frequencies of shared alleles in putative parent-offspring pairs, which can maximize statistical power when faced with fixed numbers of loci. I accounted for genotyping errors by introducing a quantitative method to determine the number of loci to allow to mismatch based upon study-specific error rates.

These novel parentage methods were applied to yellow tang (*Zebrasoma flavescens,* Acanthuridae) sampled around the Island of Hawai'i (measuring 140 km by 129 km) during the summer of 2006. We identified four parent-offspring pairs, which documented dispersal distances ranging from 15 to 184 kilometers. Two of the parents were located within MPAs and their offspring dispersed to unprotected areas. This observation provided direct evidence that MPAs can successfully seed unprotected sites with larvae that survive to become established juveniles. All four offspring were found to the north of their parents and a detailed oceanographic analysis from relevant time periods demonstrated that passive transport initially explained the documented dispersal patterns. However, passive dispersal could not explain how larvae eventually settled on the same island from which they were spawned, indicating a role for larval behavior interacting with fine-scale oceanographic features. Two findings together suggested that sampled reefs did not contribute equally to successful recruitment: (1) low levels of genetic differentiation

among all recruit samples, and (2) the fact that the 4 documented parents occurred at only 2 sites. These findings empirically demonstrated the effectiveness of MPAs as useful conservation and management tools and highlighted the value of identifying both the sources and successful settlement sites of marine larvae.

I next examined patterns of larval dispersal in bicolor damselfish (*Stegastes partitus*, Pomacentridae) collected during the summers of 2004 and 2005 from reefs lining the Exuma Sound, Bahamas (measuring 205 km by 85 km). Parentage analysis directly documented two parent-offspring pairs located within the two northern-most sites, which indicated self-recruitment at these sites. Multivariate analyses of pair-wise relatedness values confirmed that self-recruitment was common at all sampled populations. I also found evidence of "sweepstakes events", whereby only a small proportion of mature adults contributed to subsequent generations. Independent sweepstakes events were indentified in both space and time, bolstering the direct observations of self-recruitment and suggesting a role for sweepstakes analyses to identify the scale of larval dispersal events.

This dissertation provides insights into the patterns of larval dispersal in coral-reef fishes. The coupling of direct (e.g., parentage) and indirect (e.g., assignment methods, sweepstakes analyses) methods in conjunction with continued technological and methodological advances will soon provide large-scale, ecologically relevant, rates of larval exchange. By uncovering the dynamics of these enigmatic processes, the implementation of conservation and management strategies for marine fishes in general will undoubtedly experience greater success.

Larval Dispersal in Marine Fishes: Novel Methods Reveal Patterns of Self-Recruitment and Population Connectivity

by
Mark R. Christie

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 4, 2009
Commencement June 2010

Doctor of Philosophy dissertation of Mark R. Christie presented on September 4, 2009.

APPROVED:

Major Professor, representing Zoology

Chair of the Department of Zoology

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Mark R. Christie, Author

ACKNOWLEDGEMENTS

I first thank my advisor, Mark Hixon, for his continual encouragement and generous academic and financial support. Mark has an amazing ability to foster both professional and personal growth, to which I am deeply indebted. I also wish to thank Mark for recognizing the importance of this work and believing in my abilities. I truly could not have hoped for a better mentor.

I thank my graduate committee (in alphabetical order) for their valuable advice and support. Steve Arnold warrants immense gratitude for providing laboratory space and equipment. I also wish to thank Steve for his great sense of humor and inclusive spirit; I have often felt like an honorary member of the Arnold lab. Michael Banks introduced me to population genetics analyses and I am grateful for his continual encouragement, helpful feedback, and contagious enthusiasm. I thank Mike Blouin for his generous advice and feedback and I look forward to our future collaborations. Jerri Bartholomew provided great service as a graduate council representative. I am especially grateful to Selina Heppell for stepping in at the last moment to replace Jerri and for providing valuable comments on my dissertation.

I also wish to thank Hixon lab members (in chronological order) for support, advice, and, most importantly, friendship. Chris Stallings showed me the ropes of graduate student life, and for that I am truly grateful. Darren Johnson has kept daily life entertaining and has always been a great source for ideas and advice - I look forward to continued collaborations and friendship. Mark Albins has been the ideal office mate, a reliable source of feedback, and a great friend. Timothy Pusack has

been a valuable addition to the lab, despite being a Yankees fan. He is also not named Mark, which is an advantage. To all of you, I am truly grateful for the thousands of fish tissue samples that were professionally acquired.

I thank all of my co-authors, not previously mentioned, for all of their hard work and valuable intellectual contributions: Jim Beets, Yanli Jia, Stosh Thompson and Brian Tissot. I also wish to acknowledge the field work and/or intellectual contributions from: Hitoshi Araki, Kevin Buch, Jeremy Claisse, Jeff Eble, Alex Krupkin, Dellise Ortiz, Kim Page-Albins, Ivan Phillipsen, John Purcell, Jacob Tennessen, and Bill Walsh. Financial support was generously provided by a Thompson Fellowship to me, and Conservation International, National Science Foundation, and National Undersea Research Program grants to Mark Hixon.

The OSU Zoology department has provided me with an ideal personal and professional community. I would like to thank the office staff, Tara Bevandich, Mary Crafts, Sarah Cain, Traci Durrell-Khalife, and Torri Shrock, for their tireless efforts. I also wish to thank Joe Beatty for teaching assistantships and graduate student support. I am truly thankful for all the great friendships that I have made with other zoology graduate students and associates. A simple list cannot do our friendships justice, nevertheless I wish to acknowledge: Betsy and Rob Bancroft, Lindsay and Jason Biga, Mike Bogan, Kaitlin Bonner, Angela Brandt, Dafne Eerkes-Medrano, Sarah Eddy, Chris Friesen, Margot Hessing-Lewis, Jacob and Monica Jacobsen-Tennessen, Karen Kiemnec-Tyburczy, Laura McMullen, Sean Moore, Kim Page-Albins, Rocky Parker, Ivan Phillipsen, Marla Ranelleti, Christy Schnitzler, Joe Tyburczy, Josef Uyeda,

Elyse Vaccaro, Luis Vinueza, Mike Westphal Askao Yamamuro, and Phoebe and Jay Zarnetske.

The work in this dissertation could not have been accomplished without the love and dedication of my parents, James and Heidi Christie, for which I am truly grateful. I also thank my sister, Jacqueline Henze, for being a great friend and source of moral support. I lovingly thank Catherine Searle for her help and encouragement and for simply being my best friend.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF APPENDIX FIGURES

Larval Dispersal in Marine Fishes: Novel Methods Reveal Patterns of Self-Recruitment and Population Connectivity

Chapter 1: General Introduction

Understanding patterns of marine larval dispersal is vital to improve knowledge of marine metapopulation dynamics, aid in the management of exploited populations, and allow for informed reserve design to protect suitable habitats, biodiversity, and spawning populations. Therefore, determining whether and how populations of marine fish are connected by larval dispersal is of great importance in ecology, fisheries, and conservation. The majority of marine fish species have a bipartite life cycle consisting of a dispersive larval stage, and a comparatively sedentary juvenile/adult stage. However, the extent to which larvae move between local populations versus settle within natal populations is largely unknown. If the majority of larvae settle within their natal population, a process known as self-recruitment, a protected area will facilitate only populations within its boundaries. Alternatively, if there are high levels of larval connectivity among sites, then a well-designed network of marine reserves will promote persistence and resilience (Hughes et al. 2005). Additionally, identifying the relative reproductive contribution of sites (i.e., sources vs. sinks) over both spatial and temporal scales is vital for appropriate reserve design (Berkeley et al. 2004).

Surprisingly few definitive studies of larval dispersal have been completed. This paucity in data is largely due to the difficulty of observing and tracking marine larvae (Mora & Sale 2002). Currently two main approaches are used to determine

larval connectivity: otolith analysis and genetics. Otoliths (calcified ear stones) are used in two different ways to determine patterns of larval dispersal: microchemical analysis and direct tagging. Otolith microchemistry involves examining concentrations of naturally occurring trace elements within the otoliths of fishes (Swearer et al. 1999). For results to be meaningful, elemental concentrations must vary among local environments (White et al. 2008) and larvae must experience different dispersal pathways thereby spending time in the elementally different areas (Thorrold et al. 2002). Otolith tagging involves bathing demersal eggs or injecting gravid females with a chemical (most commonly tetracycline or radioisotopes) that creates a detectable mark on the otoliths of developing embryos (Jones et al. 1999, Thorrold et al. 2006, Almany et al. 2007). After the pelagic larval phase, recruits are collected and their otoliths are examined for the presence of the tag.

Genetic methods provide a particularly promising approach to elucidating patterns of larval dispersal. Most early studies of coral-reef fish dispersal used allozymes or mitochondrial DNA to determine genetic differentiation among populations (Doherty et al. 1995, Shulman and Bermingham 1995). A review of early genetics studies reported that most coral-reef fish populations, including populations that are separated by thousands of kilometers, have low levels of genetic differentiation (Shulman 1998). This observation is not unique to coral-reef systems; most studies of marine fishes in general document similar patterns (O'Connell and Wright 1997). However, most signals of genetic differentiation could be eliminated by low levels of gene flow among populations, which may be ecologically

insignificant (Mills and Allendorf 1996, Wang 2004). Therefore, it remains vital to reconcile the extent to which larval retention versus population connectivity occur at ecological time scales. Recent studies of coral-reef fishes have found evidence for high levels of self-recruitment within populations (Jones et al. 1999, Taylor and Hellberg 2003, Almany et al. 2007, Planes et al. 2009), though limited sampling designs and methodological constraints limit the generality of these results. Therefore, the complete characterization of larval exchange among populations depends upon the improvement of methodological approaches.

One underexplored method for directly determining patterns of larval dispersal is parentage. Because most juvenile and adult fish are relatively sedentary, substantial spatial distances between parents and offspring can be attributed directly to larval dispersal. Chapter 2 (Christie 2009) introduces a novel approach for detecting parent-offspring pairs in large natural populations, and is well suited for the majority of marine fishes. The methods introduced here have several distinct advantages over other commonly used parentage approaches: (1) the majority of candidate parents do not have to be sampled, (2) no genealogical information is required, and (3) Bayesian methods are introduced for situations with limited exclusionary power. This chapter further introduces novel methods for handling genotyping errors and provides a quantitative framework for determining how many loci to allow to mismatch based upon study-specific error rates.

Chapter 3 employs these novel parentage methods to determine patterns of larval dispersal in yellow tang (*Zebrasoma flavescens,* Acanthuridae). During the

summer of 2006, yellow tang were sampled from 10 coral reefs around the Island of Hawai'i (measuring 140km by 129km), three of which were marine protected areas (MPAs). Despite small sample sizes in comparison to the large populations, we identified four parent-offspring pairs, directly documenting dispersal distances ranging from 15 to 184 kilometers. All offspring were found at sites to the north of their parents. Mean ocean surface flow and advection-diffusion models of virtual drifting larvae showed that dispersal is initially well explained by passive processes, yet retention to the Island of Hawai'i would require an interplay of larval behavior and complex oceanographic processes. Further population-genetic analyses indicated that not all sampled populations contributed equally to successful reproduction. This study is the first to directly demonstrate that MPAs successfully seed areas outside their borders with larval fish that survive to become established juveniles.

Chapter 4 tackles the same questions as chapter 3, but in a different ocean for a different species. During the summers of 2004 and 2005, we sampled bicolor damselfish (*Stegastes partitus*, Pomacentridae) from coral reefs lining the Exuma Sound, Bahamas (measuring 205 km by 85 km). Despite limited sampling, two parent-offspring pairs were identified, but unlike the yellow tang in chapter 3, these individuals directly documented self-recruitment. A multivariate analysis of pair-wise relatedness values suggested that self-recruitment occurred at the majority of sampled sites. There was also convincing evidence of spatially and temporally independent sweepstakes effects (Hedgecock 1994a, 1994b), which further bolstered the self-

recruitment analyses and provided mechanistic insight into processes affecting marine larval dispersal.

With continued rapid advances in both theory and methodology (genetic and otherwise), the near future will be a remarkable and exciting time for uncovering community-wide patterns of larval connectivity. The findings of this dissertation provide novel insights into patterns of larval dispersal in marine fishes. I demonstrated that powerful parentage methods can successfully identify parent-offspring pairs from large natural populations. These methods reveal that MPAs effectively supply fished sites with larvae and that marine fish larvae may recruit back to their natal populations (bicolor damselfish) or travel almost two hundred kilometers away (yellow tang). Incorporating these studies into a broader socio-political context will allow for more effective conservation and management decisions.

Chapter 2

Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets

Mark R. Christie

Molecular Ecology Resources

In Press: DOI 10.1111/j.1755-0998.2009.02687.x

Abstract

Parentage analysis in natural populations presents a valuable yet unique challenge due to large numbers of pair-wise comparisons, marker set limitations, and few sampled true parent-offspring pairs. These limitations can result in the incorrect assignment of false parent-offspring pairs that share alleles across multi-locus genotypes by chance alone. I first define a probability, $\Pr(\delta)$, to estimate the expected number of false parent-offspring pairs within a data set. This probability can be used to determine whether one can accept all putative parent-offspring pairs with strict exclusion. I next define the probability $\Pr(\phi \mid \lambda)$, which employs Bayes' theorem to determine the probability of a putative parent-offspring pair being false given the frequencies of shared alleles. This probability can be used to separate true parent-offspring pairs from false pairs that occur by chance when a data set lacks sufficient numbers of loci to accept all putative parent-offspring pairs. Lastly, I propose a method to quantitatively determine how many loci to let mismatch for study-specific error rates and demonstrate that few data sets should need to allow more than two loci to mismatch. I test all theoretical predictions with simulated data and find that, first, $\Pr(\delta)$ and $\Pr(\phi \mid \lambda)$ have very low bias, and second, that power increases with lower sample sizes, uniform allele frequency distributions, and higher numbers of loci and alleles per locus. Comparisons of $\Pr(\phi \mid \lambda)$ to strict exclusion and CERVUS demonstrate that this method may be most appropriate for large natural populations when supplemental data (e.g., genealogies, candidate parents) are absent.

Introduction

Parentage analysis is a precise form of assignment testing (Manel et al. 2005) and can be particularly useful for detecting ecological and evolutionary patterns in systems with high levels of gene flow. Such systems have limited genetic differentiation, which severely restricts the utility of population-level assignment methods. Therefore, parentage analyses may allow for the inference of gene flow and dispersal at ecologically relevant time scales. A challenge to employing parentage analysis in natural populations is that large population sizes, variable dispersal distances, and high rates of mortality may severely constrain the number of sampled parent-offspring pairs. These challenges are amplified in systems where patterns of dispersal are unobservable, such as the larval dispersal stage in the majority of marine fishes and invertebrates (Palumbi et al. 1997; Hixon et al.; 2002; Leis 2006), where propagules are too small to track directly (but see Jones et al. 1999; Thorrold et al. 2006). Additionally, due to a lack of pragmatic methods, long-distance dispersal events are often ignored or remain undetected in many species of plants (Nathan 2006), fungi (Kauserud et al. 2006), and animals that are cryptic or have complex life histories (Derycke et al. 2008). Large genotypic data sets may be used to uncover some of these enigmatic processes, and parentage analysis can be a powerful tool for the direct detection of patterns of dispersal and population connectivity.

Several studies have successfully employed parentage analyses to address questions of gene flow and dispersal. For example, parentage analysis has revealed patterns of dispersal in rodents (Telfer et al. 2003; Waser et al. 2006; Nutt 2008),

insects (Tentelier et al. 2008), and fishes with dispersive larvae (Jones et al. 2005). Especially promising are recent attempts to estimate dispersal kernels with mean parent-offspring distances determined via parentage methods (Oddou-Muratorio et al. 2003; Robledo-Arnuncio and Garcia 2007). Additionally, parentage analyses could be coupled with population-level techniques in novel and effective ways. Direct estimates of parent-offspring dispersal could be incorporated as priors into Bayesian assignment methods or incorporated into a landscape genetics (Manel et al. 2003) or circuit-theory framework (McRae and Beier 2007). As parentage methods become more powerful and population-level methods increasingly detect fine-scale genetic structure, synergistic approaches hold great promise for accurate dispersal estimates.

The majority of studies using parentage methods to determine patterns of dispersal have relied upon likelihood-based approaches (Thompson 1975; Thompson, 1976; Meagher 1986; Thompson and Meagher 1987). Several approaches have been suggested to evaluate the significance of likelihood ratios (Gerber et al. 2003; Anderson and Garza 2006) with CERVUS being the most commonly used program (Marshall et al. 1998; Kalinowski et al. 2007). Unfortunately, these methods of evaluating significance require estimates of demographic parameters often difficult or impossible to obtain from natural populations. The program CERVUS, for example, requires precise estimates of the number of candidate parents per offspring and the proportion of candidate parents sampled (Kalinowski et al. 2007). These parameters, along with a direct setting of the confidence level, serve to control type I and type II errors. However, in many cases, this process obfuscates parentage analyses because it

is unclear how sensitive CERVUS is to estimates of these parameters. Therefore, this approach may not be appropriate for many natural populations, when the probability of finding a parent is low and where reliable observational data are difficult to obtain.

For natural populations with few sampled parents, strict exclusion or kinship techniques are the preferred analytical approaches for parentage assignment (Jones and Ardren 2003). Kinship methods are restrictive because they determine only whether a data set has more related individuals than expected by chance (Queller et al. 2000), but often cannot identify which individuals those are. Strict exclusion, which is the process of excluding dyads through Mendelian incompatibility, is a powerful method. However, one must first determine whether their data set has enough polymorphic markers to minimize the occurrence of false pairs (i.e., adults that share an allele with a putative offspring by chance). As a consequence, many exclusion probabilities have been developed for a variety of applications. Some approaches focus on data sets where the genotypes of the mother and putative sire, or at least one parent, are available (Chakraborty et al. 1988; Jamieson and Taylor 1997), whereas other exclusion methods focus on excluding only a handful of candidate parents (Dodds et al. 1996). One exclusion probability is appropriate for situations when neither parent is known was first described by Garber and Morris (1983) and later expressed in terms of homozygotes (Jamieson and Taylor 1997). Here, I show that this exclusion probability is biased when there are differences in allele frequencies between samples of adults and juveniles and recommend an unbiased alternative.

When applied correctly, exclusion is a powerful parentage method because it fully accounts for the uniqueness of the parent-offspring relationship (Milligan 2003) without any assumptions. It is this strength, however, that is often the greatest drawback to exclusion-based approaches because a genotyping error at a single locus can invalidate a true parent-offspring pair. In contrast to likelihood methods, it has proven difficult to incorporate genotyping error into exclusion-based methods. Thus the majority of exclusion-based studies usually allow for a certain number of loci to mismatch (e.g., Vandeputte 2006; McLean et al. 2008). This simply means that if a locus for a putative parent-offspring pair does not share an allele, then that locus is dropped for the analyses of that particular putative parent-offspring pair. Therefore, in order to fully account for genotyping error, it is necessary to start a project with a few more loci than the minimum required for sufficient exclusionary power. One major concern is deciding how many loci should be allowed to mismatch, and to date, this has largely been a subjective process (Hoffman and Amos 2005). If too many loci are allowed to mismatch, one runs the risk of falsely assigning parent-offspring pairs. If too few loci are allowed to mismatch, then one runs the risk of not identifying true parent-offspring pairs. Thus, I propose a quantitative approach to determine how many loci to allow to mismatch based upon study-specific estimates of genotyping error.

In this paper, I define the probability $\Pr(\delta)$, which is an unbiased exclusion probability that can be applied when one or both parents are absent. This probability can simply be multiplied by the total number of pair-wise comparisons to estimate the

number of false parent-offspring pairs within that data set. If a data set contains insufficient numbers of loci, such that it generates an unacceptable probability of containing false parent-offspring pairs, it is still possible to separate true from false parent-offspring pairs. To do so, I define a second probability, $\Pr(\phi \mid \lambda)$, that determines the probability of a putative parent-offspring pair being false given the frequencies of shared alleles. This novel approach allows researchers to identify true parent-offspring pairs when there is insufficient power for strict exclusion and, importantly, does not require any estimates of demographic parameters. I then describe an approach to determine how many loci to let mismatch based upon study-specific error rates. Software to implement all analyses presented here are available at http://sites.google.com/site/parentagemethods/. In what follows, I first describe these methods and subsequently validate them by measuring bias in simulated data sets and by drawing comparisons between existing methods.

Materials and Methods

False parent-offspring pairs

Here, I describe the probability of false parent-offspring pairs occurring within a data set. This probability can determine whether the information content of one's data set is sufficient to accept all putative parent-offspring pairs with simple Mendelian incompatibility. This framework is developed assuming the use of co-dominant markers in diploid organisms. I also include a table that provides explicit

definitions of terms used throughout this paper, as terminology varies across studies (Table 2.1).

The probability of a randomly selected dyad from a particular locus sharing an allele equals:

$$\Pr(Z) = \sum_{i=1}^{Na}\left(2z_{1i} - {z_{1i}}^2\right)\left(2z_{2i} - {z_{2i}}^2\right) - \sum_{i=1}^{Na-1}\sum_{g=i+1}^{Na}\left(2z_{1i}q_{1g}\right)\left(2z_{2i}q_{2g}\right) \quad (1)$$

where Na, equals the total number of alleles at a locus, z_1 equals the allele frequency for allele i in the sample of adults, and z_2 equals the allele frequency for allele i in the sample of juveniles. Thus, ${z_1}^2$ and ${z_2}^2$ equal the frequency of homozygotes containing allele i in samples of adults and juveniles, respectively, assuming Hardy-Weinberg Equilibrium (HWE). Alleles occurring in only one sample (i.e., adults or juveniles) will not be included in the above expression because the product equals zero. Notice that the expected number of homozygotes for an allele is subtracted from the total number of times the same allele occurs to prevent dyads that are homozygous for the same allele from being counted twice. Likewise, it is important to only count dyads that are heterozygous for the same alleles only once. Therefore, I subtract a double summation where q equals the frequencies of alleles $i+1:Na$ and where z_1q_1 and z_2q_2 equal the HWE expected genotype frequencies of unique heterozygotes in samples of adults and juveniles, respectively.

Under some circumstances it may be desirable to use an equation that does not employ HWE estimates of genotype frequencies. One example would be if genotype frequency estimates have high accuracy yet do not conform to HWE expectations. The equation that does not assume HWE is:

$$\Pr(Z_G) = \sum_{i=1}^{Na} (2z_{1i} - zz_{1i})(2z_{2i} - zz_{2i}) - \sum_{i=1}^{Ng} (zq_{1i})(zq_{2i}) \quad (2)$$

where zz_1 and zz_2 equal the observed frequencies of homozygotes containing allele i in the samples of adults and juveniles, respectively, and zq_1 and zq_2 equal the observed frequencies of all unique heterozygotes, Ng, in the samples of adults and juveniles, respectively.

To expand this approach to multiple loci, it is assumed throughout this paper that loci are in linkage equilibrium and are thus independent of one another. However, linked loci could be incorporated by explicitly accounting for the dependence between loci provided that estimates of recombination rates can be obtained (see methods in Thompson & Meagher 1998). If the assumption of linkage equilibrium is valid, it is possible to multiply probabilities across loci such that:

$$\Pr(\delta) = \prod_{i=1}^{L} \Pr(Z)_i \quad (3)$$

where L equals the total number of loci. To determine the approximate number of false parent-offspring pairs, $Fpairs$, for a given data set, $\Pr(\delta)$ should be multiplied by the total number of pair-wise comparisons:

$$Fpairs = \Pr(\delta) \cdot n_1 \cdot n_2 \quad (4)$$

where and n_1 equals the number of adults, n_2 equals the number of juveniles. It is important to keep in mind that this is a probability, and that variance due to sampling will cause slight deviations from this quantity. However, on average, these equations predict the number of false pairs very accurately (Figure 2.1, Table 2.1). Notice that equation 1 bears some similarity to the exclusion equations described in Jamieson and Taylor (1997). However, the exclusion equations presented by Jamieson and Taylor use allele frequencies from the combined samples of adults and juveniles, which results in positively biased estimates if there are only slight differences in allele frequencies between samples of adults and juveniles (Figure 2.2, Table 2.2).

The importance of minimizing the number of false parent-offspring pairs depends on the study, although the utility and accuracy of any parentage analysis obviously deteriorates as the number of false parent-offspring pairs increases. If the expected number of false parent-offspring pairs is negligible (i.e., near 0), then strict exclusion can be safely used. Here, the probability of any particular putative parent-offspring pair being false, when using strict exclusion, equals:

$$\Pr(\phi) = \frac{Fpairs}{N_P} \quad (5)$$

where N_P equals the observed number of putative parent-offspring pairs, which is simply calculated by summing the number of dyads that share at least one allele at all loci. N_P is also equal to the total number of false parent-offspring pairs plus the total number of true parent-offspring pairs. Because $\Pr(\phi)$ equals the probability of any putative parent-offspring pair being false, one should strive to minimize this value by employing many polymorphic loci. Additionally, it may be useful to obtain an *a priori* estimate of $\Pr(\delta)$, decide upon an acceptable number of false pairs, and solve for the maximum sample size for a particular marker set. Such *a priori* estimates can aid in determining whether more loci should be developed before performing parentage analyses.

Putative parent-offspring pairs

If the probability of type I error for strict exclusion is unacceptably high, such that it is unwise to accept all putative parent-offspring pairs as true pairs, it is often possible to determine whether some putative parent-offspring pairs are true pairs. This is achieved by calculating the probability of a putative parent-offspring pair being false given the frequencies of shared alleles, which using Bayes' theorem equals:

$$\Pr(\phi \mid \lambda) = \frac{\Pr(\lambda \mid \phi) \cdot \Pr(\phi)}{\Pr(\lambda)} \tag{6}$$

where $\Pr(\phi)$ equals the probability of a putative parent-offspring pair occurring by chance and $\Pr(\lambda)$ equals the probability of observing the shared alleles. $\Pr(\phi)$ is defined by equation 5, yet we still need to define $\Pr(\lambda \mid \phi)$, the probability of observing the shared alleles given that the putative parent-offspring pair is false.

To calculate $\Pr(\lambda \mid \phi)$, one must first calculate a measure of the shared allele frequencies in a putative parent-offspring pair and secondly create a distribution of similar values generated from false pairs for comparison. It is important to note that it does not matter what measure of shared allele frequencies is used. Here, I employ an approach similar to equation one to calculate an overall measure of shared allele frequencies, but one could just as easily use common likelihood methods (e.g. Thompson, 1976), as the results would be identical. As before, each locus is treated independently. Thus the measure of shared allele frequencies employed here equals:

$$\Pr(\tilde{Z}) = \sum_{i=1}^{\tilde{N}a} \left(2z_{1i} - {z_{1i}}^2\right)\left(2z_{2i} - {z_{2i}}^2\right) - \sum_{i=1}^{\tilde{N}a-1} \sum_{g=i+1}^{\tilde{N}a} \left(2z_{1i}q_{1g}\right)\left(2z_{2i}q_{2g}\right) \tag{7}$$

where all symbols are the same as equation 1, except that $\tilde{N}a$ equals the number of alleles, including the shared allele, that occur with a frequency less than or equal to that of the shared allele. Because this approach only examines the frequency of the

one shared allele in accordance with Mendelian inheritance, if a putative parent offspring pair happens to be heterozygous for the same alleles, it is appropriate to employ the rarer of the two alleles in the above framework. This probability can once again be combined across all loci, assuming linkage equilibrium, such that:

$$\Pr(\widetilde{\delta}) = \prod_{i=1}^{L} \Pr(\widetilde{Z})_i \tag{8}$$

where this equation represents the probability of observing a dyad, not a putative parent-offspring pair, that shares equally or less common alleles.

To determine $\Pr(\lambda \mid \phi)$, a distribution of $\Pr(\widetilde{\delta})$ from false parent-offspring pairs must be created. This is achieved by creating data sets (hereafter referred to as null sets) with the same allele frequencies, sample sizes and number of loci as the real data set of interest. These null sets contain no true parent-offspring pairs, thus all simulated adults and juveniles that share at least one allele across all loci do so by chance alone (i.e., all putative pairs are false parent-offspring pairs). For every null data set, $\Pr(\widetilde{\delta})$ is calculated for every false parent-offspring pair. These values are used to create a distribution of false parent-offspring pairs. To reduce bias, at least 10,000 individual false $\Pr(\widetilde{\delta})$ values should be generated from a minimum of 100 null sets, which was found to be more than sufficient under all conditions tested. Notice that the number of values used to create this distribution does not come from any assumptions about the data. It is only necessary to ensure that this distribution is

representative of the true distribution of false pairs, with more calculated values creating a more accurate description of the distribution. The mean and variance of this distribution depends upon the power of the data set used to create it. Figure 2.3a shows an example of such a distribution created from 10,000 false pairs.

To calculate $\Pr(\lambda \mid \phi)$, the value of $\Pr(\widetilde{\delta})$ for the putative pair under consideration, $\Pr(\widetilde{\delta})_i$, is compared to the distribution of values generated by false parent-offspring pairs, $\Pr(\widetilde{\delta})_F$, such that:

$$\Pr(\lambda \mid \phi) = \frac{N\left(\Pr(\widetilde{\delta})_F \le \Pr(\widetilde{\delta})_i\right)}{N(\Pr(\widetilde{\delta})_F)} \tag{9}$$

where the numerator equals the total number of false values, $\Pr(\widetilde{\delta})_F$, generated from null sets, that are less than or equal to $\Pr(\widetilde{\delta})_i$ and the denominator equals the total number of false $\Pr(\widetilde{\delta})_F$ values used to create the distribution. For example, if 100 $\Pr(\widetilde{\delta})_F$ values out of a distribution of 10,000, were found to be less than or equal to $\Pr(\widetilde{\delta})_i$ then $\Pr(\lambda \mid \phi)$ would equal 0.01.

The remaining probability needed to satisfy equation 6 is $\Pr(\lambda)$, the probability of observing the shared alleles. This is easily obtained by noticing that Bayes' theorem is often restated (see Sokal and Rohlf 1995 for a general treatment; Carlin and Louis 2000 for a detailed treatment):

$$\Pr(\phi \mid \lambda) = \frac{\Pr(\lambda \mid \phi) \cdot \Pr(\phi)}{\Pr(\lambda \mid \phi) \cdot \Pr(\phi) + \Pr(\lambda \mid \phi^C) \cdot \Pr(\phi^C)} \quad (10)$$

where $\Pr(\phi^C)$ equals the complement of $\Pr(\phi)$ and where $\Pr(\lambda \mid \phi^C)$ equals unity. This is because $\Pr(\lambda \mid \phi^C)$ equals the probability of observing the shared alleles given that the putative parent-offspring pair is true. There should be no reason for a true parent-offspring pair to be constrained to any particular set of alleles, thus this value should equal unity unless there is selection for or against alleles. Notice that if $\Pr(\lambda \mid \phi)$ equals one, meaning that the putative parent-offspring pair shares the most common alleles at all loci, then $\Pr(\phi \mid \lambda) = \Pr(\phi)$. Additionally, it is clear that if rarer alleles are shared, then it becomes less likely that a putative parent-offspring pair is false (Figure 2.3b).

As with most statistical methods, an arbitrary cutoff value can be decided upon *a priori*. Choosing a cutoff value is largely a matter of convenience and may depend on the goals of the study. The interpretation of $\Pr(\phi \mid \lambda)$ is straightforward: it represents the probability of a putative parent-offspring pair being false given the frequencies of shared alleles. It is important to recognize that a large probability does not mean that a dyad is a false parent-offspring pair, but rather that more power is needed in the form of additional loci to determine whether it is a true parent-offspring pair.

Genotyping error

If a putative parent-offspring pair does not share an allele at a single locus, then $\Pr(\phi \mid \lambda)$ remains undefined. To allow for genotyping errors, null alleles, and mutations, it is necessary to quantitatively estimate how many loci should be allowed to mismatch based upon the study-specific error rate. First, one must perform a second independent analysis on a subset of genotyped samples across all loci. The study-specific error rate, ε, is then defined as the quotient of the number of alleles that differ after the second analysis to the total number of alleles compared (*sensu* BONIN et al. 2004). To quantitatively estimate the number of loci that should be allowed to mismatch, one must first determine the probability of observing at least one error in a multi-locus genotype. I use a simplification of Bonin et al.'s (2004) formula:

$$P = 1 - (1 - \varepsilon)^{2L} \tag{11}$$

where L is equal to the total number of loci employed in the study. This probability comes from solving the binomial for the proportion of multi-locus genotypes with no errors and subtracting the result from unity to account for all errors. Alternatively, one can solve for the proportion of multi-locus genotypes that have exactly *i* errors:

$$P_i = \binom{2L}{i} (\varepsilon)^i (1 - \varepsilon)^{2L-i} \tag{12}$$

I extend this probability to determine the probability of observing at least one error in a multi-locus pair-wise comparison. This probability is equal to:

$$P' = 2P - P^2 \tag{13}$$

where P' equals the proportion of pair-wise comparisons (i.e., dyads) that will have at least one error at a locus. Notice that there is no solution for L, without the use of imaginary numbers. However, one can iteratively determine the proportion of dyads that would have at least one error given a number of mismatching loci:

$$P_i' = 2\left(P - \sum_{i=1}^{M} P_i\right) - \left(P - \sum_{i=1}^{M} P_i\right)^2 \tag{14}$$

where M equals the number of loci allowed to mismatch and must be an integer greater than 0. Thus P_i' equals the number of dyads with at least one error given that M loci are allowed to mismatch. Not all errors will cause a mismatch because the majority of dyads will not be parent-offspring pairs and additionally the majority of positions where an error occurs will not cause a mismatch. Thus choosing a cutoff value for P_i' is somewhat subjective and should be reported along with the number of loci allowed to mismatch. As a general rule of thumb, a P_i' between 0.05 and 0.1 is likely have no putative pairs with a mismatch-causing error. The advantage to this method becomes quickly apparent when one notices how quickly the error rate drops

by allowing a single locus to mismatch. Therefore, while a P_i' near 0.05 may be unnecessarily conservative, in most cases it is approached by simply allowing one to two loci to mismatch (Figure 2.5). Notice that this approach can also be applied to methods that determine the probability of identity among genotypes and that one can additionally account for null alleles, missing data, and mutation simply by adding estimates of those rates to ε.

Validation

The above methodology was tested with simulated data sets to determine whether the theoretical predictions matched actual occurrences of false parent-offspring pairs and to compare with existing methods. All data simulation and probability calculations were implemented in R version 2.5.1 (R Development Core Team, 2007). Simulated data sets were created using a set of alleles whose frequencies were determined by the equation:

$$z = \frac{i}{(Na+1)-i} \cdot \sum_{i=1}^{Na} \frac{i}{(Na+1)-i} \tag{15}$$

where Na equals the total number of alleles and i equals allele i in the set 1: Na. This distribution is fairly conservative as it results in several fairly common alleles (Bernatchez & Duchesne, 2000). Allele frequency distributions had no effect on the precision and accuracy of these methods, but in all cases a uniform distribution (i.e.,

equal allele frequencies) resulted in the greatest power. Once the population allele frequencies were determined, 100,000 genotypes were created in accordance with HWE. This pool of genotypes was randomly sampled and placed into a group of either adults or juveniles. This process was repeated for each locus. True parent-offspring pairs were created by randomly sampling 1 individual from both the sample of adults and the sample of juveniles. The two individuals were aligned, locus by locus, and at each locus, a randomly chosen allele was copied from the adult to the offspring. This procedure was executed regardless of whether the pair already shared an allele at that locus, which simulated the occasional, but realistic, occurrence of dyads being homozygous or heterozygous for the same alleles. This procedure also has the benefit of making the distribution of shared alleles equal to that of the overall sample distribution, which is expected with a random sample.

I first tested whether the theoretical predicted number of false pairs, as calculated by $\Pr(\delta)$, matched the actual number of false parent-offspring pairs. Simulated data sets were created with varied numbers of loci, sample sizes, alleles, and true parent-offspring pairs. All simulated data sets had equal sample sizes of adults and juveniles, which maximizes the number of pair-wise comparisons. One thousand simulated data sets were created for each combination of variables. For each data set, theoretical estimates of the expected number of false parent-offspring pairs were calculated using equation 4 and compared to the actual number of false pairs observed. The bias, root mean square error, and variance of the predicted number of false pairs were calculated from all 1000 simulated data sets.

I next tested whether $\Pr(\delta)$ was less biased than the multi-locus approach presented by Jamieson and Taylor (1997), represented in their paper with equations two through four and hereafter denoted as P. I manipulated F_{ST} between the samples of adults and juveniles from 0 to 0.1. Here, and only for this section, a uniform allele frequency distribution was employed to simplify the creation of different F_{ST} values. Because identical F_{ST} values can be created with different combinations of allele frequencies, I adjusted the frequencies such that the mean allele frequencies for adults and juveniles always equaled the starting allele frequencies when F_{ST} equaled zero. For example, in the two allele case, F_{ST} equaled 0 when both alleles were set to 0.5. However, if one allele was increased to 0.6 in adults, then the same allele was decreased to 0.4 in juveniles so that the mean allele frequency remained 0.5. This process creates a standardized approach to creating F_{ST} values and highlights the differences between the two equations. The differences in estimates were both plotted and presented in terms of bias. I additionally report a standardized bias, which equals the bias divided by the total number of false parent-offspring pairs when F_{ST} equals zero. This was to demonstrate that although the bias appears to decrease with increasing power, this is only an artifact of there being far fewer false parent-offspring pairs, and that proportionally the bias is much greater in data sets with greater power.

I first validated $\Pr(\phi \mid \lambda)$ by ensuring that the chosen type I error rate matched the actual type I error rate. To do this, I created simulated data sets with varying

probabilities of a putative parent-offspring pair being false, $\Pr(\phi)$. All data sets were tested with 1, 30, and 60 true parent-offspring pairs and 1000 simulated data sets were created for each combination of values. The type I error rate, α, was set at 0.001, 0.01, 0.05 and 0.09 for each data set. Bias was measured by comparing the difference between the observed error rate and the set error rate, over all simulated data sets. For all analyses, the larger of 100 null sets or 10,000 $\Pr(\widetilde{\delta})_F$ values was used for the calculation of $\Pr(\lambda \mid \phi)$.

I next compared $\Pr(\phi \mid \lambda)$ to CERVUS v. 3.0, again with data sets created with varied $\Pr(\phi)$ values. To accomplish this, sample size, numbers of loci and alleles per locus, and the numbers of true parent-offspring pairs were varied. Both the proportion of true parent-offspring pairs correctly assigned as true pairs and the proportion of false parent-offspring pairs incorrectly assigned as true pairs were recorded for both methods. Direct comparisons to CERVUS are difficult to make because CERVUS requires the estimates of two parameters that $\Pr(\phi \mid \lambda)$ does not require: the number of candidate parents and the proportion of candidate parents sampled. For all comparisons, the number of candidate parents was set to either 500 or 1000 and the strict confidence level of 95% was used. The proportion of candidate parents was set to the true parameter value. This is equitable because comparisons with larger numbers of candidate parents, and inaccurate estimates of the proportion of candidate parents sampled resulted in poor performance by CERVUS (see Figure 4 of Marshall et

al. 1998). Putative parent-offspring pairs were accepted as true parent-offspring pairs if $\Pr(\phi \mid \lambda)$ was less than or equal to 0.05.

Lastly, I test equation 14, which predicts the proportion of dyads with at least one error for a given error rate and number of loci allowed to mismatch. I first demonstrate how the proportion of dyads with at least one error is affected by varying error rates and numbers of loci. I next created simulated data sets with 15 and 20 loci, each with error rates of 0.01 and 0.015. The number of loci allowed to mismatch was varied from one to four, and the proportion of true parent-offspring pairs correctly assigned was recorded. Agreement between the predicted proportion of dyads with at least one error and the proportion of true parent-offspring pairs correctly assigned is further evaluated in the discussion.

Results

The theoretical predictions for the number of false parent-offspring pairs, as determined by equation 4, match very closely to the observed number of false parent-offspring pairs from the simulated data sets (Figure 2.1, Table 2.2). Not surprisingly, the number of false parent-offspring pairs increases as sample size increases. Also, there are fewer false parent-offspring pairs in data sets with more loci. The rate of increase in false parent-offspring pairs is identical in data sets with different numbers of loci, but identical allele frequency distributions. Table 2.2 demonstrates that increasing numbers of true parent-offspring pairs does not affect the predictive performance of $\Pr(\delta)$. The bias is slightly larger in the data sets with smaller sample

size, which is due to increased variance in the actual number of false parent-offspring pairs. Numbers of loci or alleles do not influence bias, provided that the sample size is large enough to accurately estimate allele frequencies. Overall, the methods presented here predict the actual number of false parent-offspring pairs with high accuracy and precision.

$\Pr(\delta)$ is unbiased when there are differences in allele frequencies between the samples of adults and juveniles (Table 2.3). However, the approach employed by Jamieson and Taylor (1997) reveal that even small differences in allele frequencies between adults and juveniles can result in a large overestimation in the number of false parent-offspring pairs (Figure 2.2). The bias in their method increases with increasing genetic differentiation, whereas $\Pr(\delta)$ remains unbiased regardless of the level of genetic differentiation. Additionally, while the bias in their approach decreases with increasing number of alleles and loci, the standardized bias increases with increasing numbers of alleles and loci. This demonstrates that although the bias appears to decrease with increasing power, this is only an artifact of there being far fewer false parent-offspring pairs, and that proportionally the bias is much greater in data sets with greater exclusionary power.

The bias in type I error for $\Pr(\phi \mid \lambda)$ was very low across all tested levels of α, regardless of the number of true parent-offspring pairs or the value of $\Pr(\phi)$ (Table 2.4). The bias in type I error does not appear to follow any trends with number of true pairs, or type of data set. Additionally, the bias was negative in all cases meaning that the actual type I error rate was minutely smaller than the set type I error rate, thus

making this approach, if anything, conservative. Overall, the very low rates of bias suggest that $\Pr(\phi \mid \lambda)$ can be used confidently to determine the probability of a putative parent-offspring pair sharing alleles by chance. The comparisons of $\Pr(\phi \mid \lambda)$ to strict exclusion demonstrate the utility of each method (Figure 2.4). Not surprisingly, strict exclusion had a lower type II error than $\Pr(\phi \mid \lambda)$, correctly identifying all true parent-offspring pairs. In fact, with no genotyping error, strict exclusion has a type II error rate of 0. However, there is no mechanism for strict exclusion to determine the difference between real and false parent-offspring pairs, such that its type I error rate always equals unity. Thus, if there are any false parent-offspring pairs in the data sets, they will be assigned as true pairs. This highlights the need to only use strict exclusion after sufficient power has been quantitatively determined (e.g., equation 5). The observed patterns of type II error for $\Pr(\phi \mid \lambda)$ were as expected. As $\Pr(\phi)$ increases, the proportion of true parent-offspring pairs that were detected decreases.

Comparisons with CERVUS were also informative (Figure 2.4). The proportion of true parent-offspring pairs successfully assigned was lower for CERVUS than for $\Pr(\phi \mid \lambda)$, except when $\Pr(\phi)$ equaled 0, where they were equivalent. Not surprisingly, the proportion of parents successfully assigned with CERVUS was lower when the number of candidate parents was set to 1000 as opposed to 500, which highlights the problems of using CERVUS when there are potentially large numbers of candidate parents. However, CERVUS performs as well or better than $\Pr(\phi \mid \lambda)$ when the number of candidate parents is low or when the number of candidate parents is close to the true number of parents in the sample (data not shown). Here, we used the parameter

value for the proportion of candidate parents sampled, however, inaccurate estimates of this parameter can have large effects on both the type I and type II error. For both methods, the type I error was within acceptable limits, though it appeared to be slightly lower for $\Pr(\phi \mid \lambda)$.

Lastly, I tested the effect of allowing loci to mismatch on the number of true parent-offspring pairs correctly assigned. All data sets with greater numbers of loci and higher error rates had a larger proportion of dyads with at least one error (Figure 2.5). The proportional rate of decrease in dyads with at least one error was much greater for data sets with more loci and higher error rates. This pattern is reflected in the simulated data sets, where data sets with more loci and higher error rates had a greater increase in power by allowing loci to mismatch (Figure 2.6). The data sets with 15 loci and an error rate of 0.01 only required one locus to mismatch before proportion of true parent-offspring pairs correctly assigned equaled 1. All other values required two loci to mismatch before the proportion of true parent-offspring pairs correctly assigned equaled 1. Despite the relatively high error rates and number of loci, no data set required more than two loci to mismatch before all of the true parent-offspring pairs were correctly identified.

Discussion

This paper introduces novel approaches for determining parentage in natural populations. The theoretical predictions of the number of false parent-offspring pairs calculated with $\Pr(\delta)$ are matched very closely by the actual number of false parent-

offspring pairs from simulated data sets. $\Pr(\delta)$ remains unbiased when there are differences in allele frequencies between adults and juveniles, to which prior methods are susceptible. There may still be occasions, however, when it is better to use allele frequencies from the combined sample of adults and offspring, such as with small sample sizes or samples with inaccurate allele frequency estimates. As expected, the number of false parent-offspring pairs decreases with increasing numbers of loci and alleles. The number of false parent-offspring pairs increases with increasing sample size for a given set of loci, thus the larger the sample, the more loci that are required to accurately detect parentage. It is worthwhile noting that even data sets with 20 loci had false parent-offspring pairs across all tested sample sizes suggesting that many studies employing strict exclusion may be plagued by false parent-offspring pairs. In fact this highlights the need for any study employing Mendelian incompatibility to report some measure of exclusionary power. Since the theoretical predictions of $\Pr(\delta)$ and the simulated data match well, and with little bias, this approach can be used confidently to determine how many false parent-offspring pairs are likely to exist in large data sets from natural populations.

Conveniently, one can conduct *a priori* analyses for a given marker set to determine the maximum sample size before type I error becomes problematic. Alternatively, with a simple rearrangement of equation 4, one can determine the exclusionary power required for a desired sample size and number of false pairs. This approach is strongly recommended for any project in its infancy. However, in many

cases the data are already collected or there are only a limited number of loci available, at which point $\Pr(\phi \mid \lambda)$ should be used.

The analyses demonstrate that $\Pr(\phi \mid \lambda)$ performs well for identifying true parent-offspring pairs without compromising the type I error. The importance of this result is particularly noticeable when one considers the type I error rate (i.e., the proportion of false parent-offspring pairs identified as true parent-offspring pairs) for strict exclusion equals unity. These results also show that, even in data sets where there is a fairly high ratio of false pairs to putative pairs, $\Pr(\phi \mid \lambda)$ is able to correctly identify the majority of true parent-offspring pairs. This result is encouraging because it indicates that this approach works well with data sets that do not have quite enough power to employ strict exclusion. It is important to point out that the average shared value of allele frequencies among true parent-offspring pairs is often much less than the average value among false pairs. This is because, as false parent-offspring pairs arise in data sets, they initially share the most common alleles. As false parent-offspring pairs become more prevalent (e.g., with increasing sample size), they gradually begin to share rarer alleles. It is this pattern, in part, that makes $\Pr(\phi \mid \lambda)$ a powerful method because it exploits the subtle differences in allele frequencies that exist between some true and false parent-offspring pairs. Furthermore, the intuitive results from Figure 2.3, suggest that $\Pr(\phi \mid \lambda)$ may still be used to discern a few true parent-offspring pairs from data sets with very large numbers of false parent-offspring pairs, provided that the true parent-offspring pairs share rare alleles.

The methods presented here also have some distinct advantages over other commonly used methods, such as the likelihood approaches as implemented by CERVUS (Kalinowski et al., 2007), because no estimates of the number of candidate parents or the proportion of candidate parents sampled are needed. The results from data sets analyzed by CERVUS demonstrate that power decreases rapidly for large numbers of candidate parents. Nonetheless, CERVUS works remarkably well for its intended application, and when many of the putative parents are sampled and accurate estimates of the required parameters are available, CERVUS is highly recommended. However, as is the case in many natural populations, when the number of true parent-offspring pairs within a large sample is low and the numbers of candidate parents are large, alternative approaches are required. Thus, new methods, such as the probability $\Pr(\phi \mid \lambda)$ introduced here, are needed to detect parent-offspring pairs in large data sets collected from natural populations with little to no genealogical information.

It is important to determine the effects of genotyping error for any parentage method, as even low levels of genotyping error may play a significant role in increasing type II error for large data sets (Slate et al. 2000). However, the majority of errors will not cause a mismatch because: (1) most of the dyads are not true parent-offspring pairs and (2) most positions for an error to occur will not cause a mismatch. I demonstrate here that error has the largest effect on power when no loci are allowed to mismatch. However, the proportion of dyads with at least one error drops rapidly when just one locus is allowed to mismatch. For all the combinations of loci and error rates tested here, no data sets had detectable type II error after two loci were allowed

to mismatch (exclusionary power was not an issue), which suggests that many studies may need only to allow one or two loci to mismatch. Interestingly, the initial rate of decrease in the proportion of dyads with a genotyping error is greater for data sets with more loci or higher error rates, meaning that these types of data sets have the most to gain by allowing loci to mismatch. Nevertheless, sufficient increases in power can be obtained even for data sets with 10 loci and an average error rate of 0.01. The methods presented here could be improved by determining the proportion of true parent-offspring pairs, as opposed to dyads, with at least one error. This challenging task would involve determining how often an error would occur between shared alleles after taking into account the allelic state of the individuals. As it stands, a cutoff level for the proportion of dyads with at least one genotyping error of between 0.05 and 0.1 should result in the maximum increase in power provided that there are sufficient numbers of loci.

With the rapid increases in DNA-based technology, an overwhelming number of markers may soon become available. However, given a set of equally polymorphic loci, one can see that the each additional locus provides diminishing returns (e.g., see equation 3). The possible waste of time, money, and effort associated with negligible gains in exclusionary power can again be avoided by performing *a priori* analyses for given marker sets and study designs. Additionally, employing too many loci may be counterproductive due to increases in the study-wide error rate. At some point diminishing returns will be reached between the addition of loci and the number of loci needed to mismatch to accommodate error. In light of recent trends toward

employing large numbers of markers, I reiterate the importance of quantitatively determining the number of loci to allow to mismatch to avoid dropping loci that truly invalidate a putative parent-offspring relationship.

The parentage methods presented here suggest several promising avenues for further investigation. These methods could benefit from further testing and refinement under conditions that are rare, but present in certain natural populations. It is possible, for example, that highly skewed reproductive success or large sampling variances could create bias. It would also be useful to examine the effects of inbreeding and population substructure as was recently performed for relatedness measures (Anderson and Weir 2007). Additionally, where the presence of first-degree relatives other than parents and offspring may be an issue, I advise calculating the probability of related individuals sharing alleles at all loci for a given marker set and, if necessary, adding more loci (see Blouin 2003 and references therein). Furthermore, it would be beneficial to quantitatively test a broader array of parentage methods to determine which methods are most appropriate for the large variety of conditions likely to be faced when determining parentage in natural populations. Lastly, for data sets with very low numbers of false parent-offspring pairs (e.g. <1) or data sets with very large numbers of false parent-offspring pairs (e.g. >10000) the simulations used to calculate $\Pr(\lambda \mid \phi)$ can be time consuming, such that alternative approaches for calculating $\Pr(\lambda \mid \phi)$ may be more efficient than the simulation-based procedure presented here.

The fields of parentage and kinship analysis have been growing rapidly since the development of hypervariable markers (Blouin 2003; Jones and Ardren 2003), yet there is a growing need to apply these approaches to large populations in order to uncover patterns of dispersal and gene flow. Many organisms have propagules that are difficult to track directly, and patterns of dispersal in these systems are not well understood. Parentage provides an approach that is analogous to mark-recapture studies and therefore may be especially useful in describing gene flow and dispersal at shorter time scales (e.g., among cohorts, years, or seasons) and among populations with little genetic differentiation. Moreover, with species that are relatively long-lived one could construct a dynamic "library" of potential parents with which to compare putative offspring year after year. Samples collected over subsequent years may provide clearer insights into patterns of dispersal. The direct dispersal information gained from parentage analysis can also be used to inform population-level analyses of dispersal and may be particularly useful when incorporated into population genetics (Manel et al. 2003; McRae and Beier 2007 or coupled with ecological or remote-sensing data to complement, validate and enhance non-genetic approaches for estimating dispersal. Parent-offspring information could also be coupled with population density estimates and isolation-by distance analyses to provide more accurate estimates of species dispersal kernels (Rousset 1997). The theory and methods presented here have been developed in an effort to expand our capabilities of assessing gene flow and dispersal in large natural populations. Of course, these methods are also suited to answer a broad array of questions, such as determining

associations between phenotypes or domestication and fitness (e.g., DeWoody 2005, Araki et al. 2007). Thus, parentage analysis in natural populations is vital for the advancement of both ecology (e.g., appropriate reserve design and spacing, description of meta-population structure) and evolution (e.g., gene-flow estimation, selection, speciation).

Literature Cited

Anderson, A. D., and B. S. Weir. 2007. A maximum-likelihood method for the estimation of pairwise relatedness in structured populations. Genetics 176:421-440.

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567-2582.

Araki, H., B. Cooper, and M. S. Blouin. 2007. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science 318:100-103.

Blouin, M. S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology & Evolution 18:503-511.

Carlin, B. P., and T. A. Louis. 2000. Bayes and empirical Bayes methods for data analysis. Chapman and Hall, Boca Raton.

Chakraborty, R. 1991. Inclusion of Data on Relatives for Estimation of Allele Frequencies. American Journal of Human Genetics 49:242-243.

Chakraborty, R., T. R. Meagher, and P. E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527-536.

Derycke, S., T. Remerie, T. Backeljau, A. Vierstraete, J. Vanfleteren, M. Vincx, and T. Moens. 2008. Phylogeography of the Rhabditis (Pellioditis) marina species complex: evidence for long-distance dispersal, and for range expansions and restricted gene flow in the northeast Atlantic. Molecular Ecology 17:3306-3322.

DeWoody, J. A. 2005. Molecular approaches to the study of parentage, relatedness, and fitness: Practical applications for wild animals. Journal of Wildlife Management 69:1400-1418.

Dodds, K. G., M. L. Tate, J. C. McEwan, and A. M. Crawford. 1996. Exclusion probabilities for pedigree testing farm animals. Theoretical and Applied Genetics 92:966-975.

Garber, R. A., and J. W. Morris. 1983. General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and in no-parent cases of disputed parentage. Pages 277-280 *in* R.H.Walker, editor. Inclusion Probabilities in Parentage Testing. American Association of Blood Banks, Arlington, VA.

Gerber, S., P. Chabrier, and A. Kremer. 2003. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Molecular Ecology Notes 3:479-481.

Hixon, M. A., S. W. Pacala, and S. A. Sandin. 2002. Population regulation: historical context and contemporary challenges of open vs. closed systems. Ecology 83:1490-1508.

Hoffman, J. I., and W. Amos. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology 14:599-612.

Jamieson, A., and S. S. Taylor. 1997. Comparisons of three probability formulae for parentage exclusion. Animal Genetics 28:397-400.

Jones, A. G., and W. R. Ardren. 2003. Methods of parentage analysis in natural populations. Molecular Ecology 12:2511-2523.

Jones, G. P., M. J. Milicich, M. J. Emslie, and C. Lunow. 1999. Self-recruitment in a coral reef fish population. Nature 402:802-804.

Jones, G. P., S. Planes, and S. R. Thorrold. 2005. Coral reef fish larvae settle close to home. Current Biology 15:1314-1318.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099-1106.

Kauserud, H., O. Stensrud, C. Decock, K. Shalchian-Tabrizi, and T. Schumacher. 2006. Multiple gene genealogies and AFLPs suggest cryptic speciation and

long-distance dispersal in the basidiomycete *Serpula himantioides* (Boletales). Molecular Ecology 15:421-431.

Leis, J. M. 2006. Are larvae of demersal fishes plankton or nekton? Advances in Marine Biology, Vol 51 51:57-141.

Manel, S., O. E. Gaggiotti, and R. S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology & Evolution 20:136-142.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639-655.

McLean, J. E., T. R. Seamons, M. B. Dauer, P. Bentzen, and T. P. Quinn. 2008. Variation in reproductive success and effective number of breeders in a hatchery population of steelhead trout (*Oncorhynchus mykiss*): examination by microsatellite-based parentage analysis. Conservation Genetics 9:295-304.

McRae, B. H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences of the United States of America 104:19885-19890.

Meagher, T. R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of the most-likely male parents. American Naturalist 128:199-215.

Milligan, B. G. 2003. Maximum-likelihood estimation of relatedness. Genetics 163:1153-1167.

Nathan, R. 2006. Long-distance dispersal of plants. Science 313:786-788.

Nutt, K. J. 2008. A comparison of techniques for assessing dispersal behaviour in gundis: revealing dispersal patterns in the absence of observed dispersal behaviour. Molecular Ecology 17:3541-3556.

Oddou-Muratorio, S., M. L. Houot, B. Demesure-Musch, and F. Austerlitz. 2003. Pollen flow in the wildservice tree, Sorbus torminalis (L.) Crantz. I. Evaluating the paternity analysis procedure in continuous populations. Molecular Ecology 12:3427-3439.

Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead, E. Vyse, R. Ward, and C. Strobeck. 1998. Variation in genetic diversity across the range of North American brown bears. Conservation Biology 12:418-429.

Palumbi, S. R., G. Grabowsky, T. Duda, L. Geyer, and N. Tachino. 1997. Speciation and population genetic structure in tropical Pacific Sea urchins. Evolution 51:1506-1517.

Queller, D. C., F. Zacchi, R. Cervo, S. Turillazzi, M. T. Henshaw, L. A. Santorelli, and J. E. Strassmann. 2000. Unrelated helpers in a social insect. Nature 405:784-787.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria:http://www.R-project.org.

Robledo-Arnuncio, J. J., and C. Garcia. 2007. Estimation of the seed dispersal kernel from exact identification of source plants. Molecular Ecology 16:5098-5109.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228.

Slate, J., T. Marshall, and J. Pemberton. 2000. A retrospective assessment of the accuracy of the paternity inference program CERVUS. Molecular Ecology 9:801-808.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry 3rd ed. W.H. Freeman and Company, New York.

Telfer, S., S. B. Piertney, J. F. Dallas, W. A. Stewart, F. Marshall, J. L. Gow, and X. Lambin. 2003. Parentage assignment detects frequent and large-scale dispersal in water voles. Molecular Ecology 12:1939-1949.

Tentelier, C., T. Guillemaud, S. Ferry, and X. Fauvergue. 2008. Microsatellite-based parentage analysis reveals non-ideal free distribution in a parasitoid population. Molecular Ecology 17:2300-2309.

Thompson, E. A. 1975. The estimation of pair-wise relationship. Annals of Human Genetics 39:173-188.

Thompson, E. A. 1976. Inference of genealogical structure. Social Science Information 15:477-526.

Thompson, E. A., and T. R. Meagher. 1987. Parental and sib likelihoods in genealogy reconstruction. Biometrics 43:585-600.

Thompson, E. A., and T. R. Meagher. 1998. Genetic linkage in the estimation of pairwise relationship. Theoretical and Applied Genetics 97:857-864.

Thorrold, S. R., G. P. Jones, S. Planes, and J. A. Hare. 2006. Transgenerational marking of embryonic otoliths in marine fishes using barium stable isotopes. Canadian Journal of Fisheries and Aquatic Sciences 63:1193-1197.

Vandeputte, M., S. Mauger, and M. Dupont-Nivet. 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Molecular Ecology Notes 6:265-267.

Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology 10:249-256.

Waser, P. M., J. D. Busch, C. R. McCormick, and J. A. DeWoody. 2006. Parentage analysis detects cryptic precapture dispersal in a philopatric rodent. Molecular Ecology 15:1929-1937.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. Evolution 38:1358-1370.

Acknowledgements

This paper is a chapter from my doctoral dissertation and I thank my graduate committee for their reviews and support: Mark Hixon (chair), Steve Arnold, Michael Banks, and Mike Blouin. I also wish to thank Hitoshi Araki, Mark Albins, Darren Johnson, Catherine Searle, Jacob Tennessen, the editors, and four anonymous reviewers for helpful comments that greatly improved this manuscript. This work was supported by grants from Conservation International and the National Science Foundation (05-50709) to M.A. Hixon.

Table 2.1. Definitions of terms used throughout the paper.

Term	Definition
Adult	Any individual from a sample of sexually mature individuals
Parent	The true mother or father of an individual in a data set
Juvenile	Any individual from a sample of sexually immature individuals
Offspring	An individual that has a parent within the sample of adults
Dyad	Any pair-wise comparison between an adult and a juvenile
False parent-offspring pair	A dyad that shares at least one allele at all loci by chance due to large sample sizes or insufficient numbers of loci or alleles per locus.
True parent-offspring pair	A dyad that shares at least one allele at all loci due to direct Mendelian transmission.
Putative parent-offspring pair	A dyad that shares at least one allele at all loci but that has yet to be assigned as a true or false parent-offspring pair.
Type I error	A dyad that shares alleles across all loci by chance and is falsely determined to be a true parent-offspring pair.
Type II error	A true parent-offspring pair that is not identified as such. Is often, along with type one error, expressed as a rate.
Power	One minus the type two error rate. The proportion of true parent-offspring pairs that are correctly assigned.

Table 2.2. Mean predicted number of false parent-offspring pairs and the bias between predicted and the actual number of false pairs. True pairs equals the number of true parent-offspring pairs, and NL, NA, and N equal the number of loci, the number of alleles per locus, and sample size, respectively, where N equals the number of adults plus juveniles. Values were calculated from 1000 simulated data sets.

True pairs	NL	NA	N	False pairs	Bias	$\sqrt{MSE}$
0	10	10	250	221.67	-0.0002760	0.00219
15	10	10	250	221.82	-0.0001316	0.00216
30	10	10	250	223.59	-0.0000362	0.00216
60	10	10	250	225.61	0.0002410	0.00221
0	10	15	250	57.01	-0.0001562	0.00222
15	10	15	250	57.53	-0.0000553	0.00220
30	10	15	250	57.56	-0.0000487	0.00217
60	10	15	250	57.31	-0.0000801	0.00216
0	15	10	250	28.59	-0.0000627	0.00223
15	15	10	250	29.92	0.0000353	0.00249
30	15	10	250	29.01	0.0000103	0.00228
60	15	10	250	28.35	-0.0000802	0.00228
0	15	10	500	108.54	-0.0000011	0.00220
15	15	10	500	108.53	-0.0000091	0.00224
30	15	10	500	108.67	0.0000057	0.00225
60	15	10	500	108.55	-0.0000058	0.00223
0	15	15	500	13.13	-0.0000038	0.00233
15	15	15	500	13.12	-0.0000047	0.00230
30	15	15	500	13.2	-0.0000014	0.00218
60	15	15	500	13.2	0.0000012	0.00224

Table 2.3. Bias between the predicted exclusion probability and the actual exclusion probability used to calculate the expected number of false parent-offspring pairs. Two exclusion probabilities were compared: Jamieson and Taylor's (1997) equation 2, denoted as P, and $\Pr(\delta)$. For P, two measures of bias are employed: the absolute bias, Bias P, which equals the predicted exclusion probability minus the actual exclusion probability, and the standardized bias, Bias^S P, which is the bias divided by P when F_{ST} equals 0. This measure was used to demonstrate that although the bias can be very small, this is only an artifact of the exclusion probability being very small.

Alleles	Loci	FST	Bias P	Bias^S P	Bias $\Pr(\delta)$
2	10	0	0	0	0
2	10	0.005	0.011	0.041	0
2	10	0.01	0.022	0.083	0
2	10	0.05	0.098	0.374	0
2	10	0.1	0.160	0.607	0
2	15	0	0	0	0
2	15	0.005	0.008	0.061	0
2	15	0.01	0.016	0.121	0
2	15	0.05	0.068	0.505	0
2	15	0.1	0.102	0.754	0
10	10	0	0	0	0
10	10	0.005	8.69×10^{-6}	0.386	0
10	10	0.01	1.23×10^{-5}	0.544	0
10	10	0.05	2.18×10^{-5}	0.966	0
10	10	0.1	2.25×10^{-5}	1.000	0
10	15	0	0	0	0
10	15	0.005	5.55×10^{-8}	0.519	0
10	15	0.01	7.41×10^{-8}	0.692	0
10	15	0.05	1.06×10^{-7}	0.994	0
10	15	0.1	1.07×10^{-7}	1.000	0

Table 2.4. Bias between the set type I error rate and the observed type I error rate for $\Pr(\phi \mid \lambda)$. The type I error rate, α, was set at 0.001, 0.01, 0.05 and 0.09 for each data set. True pairs equals the number of true parent-offspring pairs, and $\Pr(\phi)$ equals the probability of a putative parent-offspring pair being false and decreases with increasing number of alleles and loci or decreasing sample size (see text for details). Each value was calculated from 1000 simulated data sets.

		α			
$\Pr(\phi)$	*True pairs*	0.001	0.01	0.05	0.09
0.9	1	-0.00065	-0.00069	-0.00071	-0.00075
0.9	30	-0.00061	-0.00066	-0.00073	-0.00075
0.9	60	-0.00062	-0.00064	-0.00069	-0.00071
0.5	1	-0.00052	-0.00058	-0.00061	-0.00065
0.5	30	-0.0005	-0.00053	-0.00057	-0.00059
0.5	60	-0.00059	-0.00062	-0.00068	-0.00072
0.25	1	-0.00043	-0.00047	-0.00049	-0.00053
0.25	30	-0.00042	-0.00045	-0.00052	-0.00057
0.25	60	-0.00038	-0.00046	-0.00048	-0.00051
0.1	1	-0.00028	-0.00029	-0.00032	-0.00057
0.1	30	-0.00013	-0.00019	-0.00023	-0.00048
0.1	60	-0.00024	-0.00036	-0.00038	-0.00053

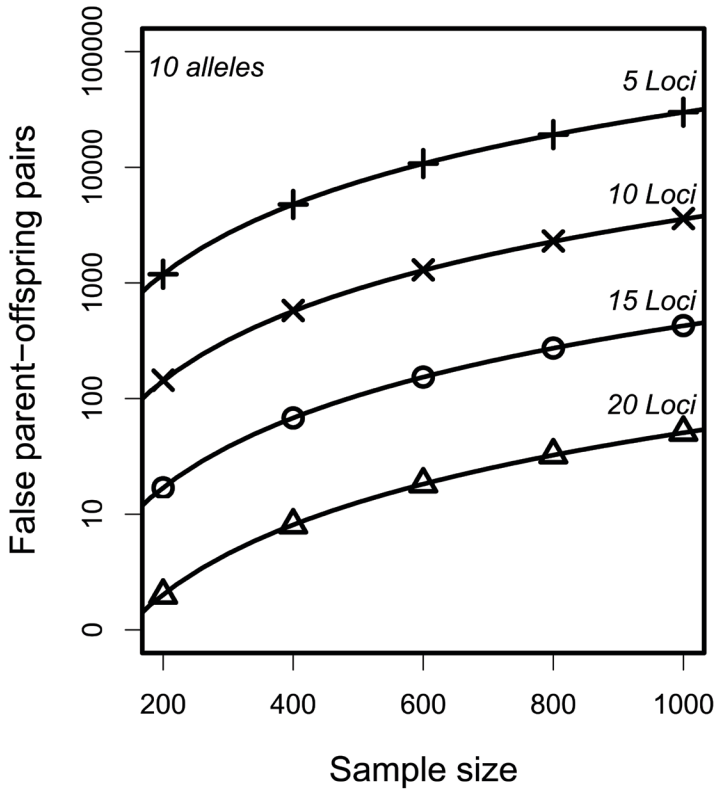


Figure 2.1. Actual number of false parent-offspring pairs as measured by simulated data sets and as predicted by *Fpairs* . Each value was determined from 1000 simulated data sets with 10 alleles per locus. The actual number of false parent-offspring pairs are shown as + for data sets with 5 loci, × for data sets with 10 loci, ○ for data sets with 15 loci, and as △ for data sets with 20 loci. The black lines are the predicted values of the number of false parent-offspring pairs, *Fpairs* , with each line calculated using the same number of loci as the symbols that lay on it. Sample size equals the number of adults plus the number of juveniles, both of which are of equal size.

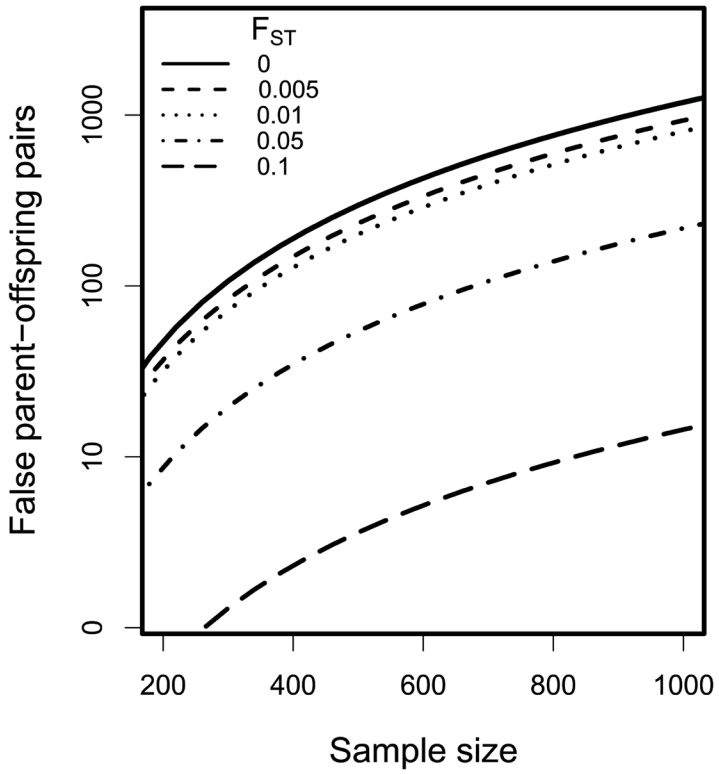
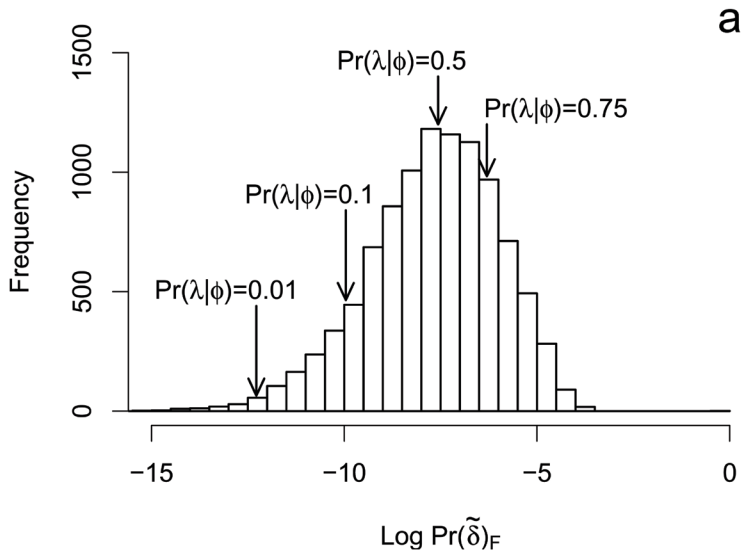


Figure 2.2. Actual and predicted number of false parent-offspring pairs as measured by simulated data sets and as predicted by *Fpairs* . The lines represent the actual and predicted values of the number of false parent-offspring pairs, *Fpairs* , from simulated data sets with 10 loci and10 alleles per locus. A uniform allele frequency distribution was employed. Actual values were measured for combined sample sizes of 200, 400, 600, 800, and 1000 individuals. Actual values lay directly over the predicted values and symbols were omitted for clarity. F_{ST} was varied from 0 to 0.1. As F_{ST} increases there is a decrease in the actual number of false parent-offspring pairs, which is matched precisely by *Fpairs* . The solid black line represents $F_{ST} = 0$, but also equals the predicted number of false parent-offspring pairs for methods that do not take into account the allele frequencies of adults and juveniles separately, regardless of the value of F_{ST}, and are thus positively biased.

a



b

	Pr(λ|ϕ)				
Pr(ϕ)	0.01	0.1	0.25	0.5	0.75
0.90	0.083	0.474	0.692	0.818	0.871
0.80	0.039	0.286	0.500	0.667	0.750
0.60	0.015	0.130	0.273	0.429	0.529
0.40	0.007	0.063	0.143	0.250	0.333
0.20	0.003	0.024	0.059	0.111	0.158
0.10	0.001	0.011	0.027	0.053	0.077
0.05	0.001	0.005	0.013	0.026	0.038
0.01	0.000	0.001	0.003	0.005	0.008

Figure 2.3. (a) Distribution of $\Pr(\widetilde{\delta})_F$ values created from 10,000 false parent-offspring pairs generated from simulated null data sets with 15 loci, 15 alleles per locus, 500 individuals, and no true parent-offspring pairs. Various probabilities of observing a set of shared alleles given that a putative parent-offspring pair is false, $\Pr(\lambda \mid \phi)$, are indicated with arrows. (b) A table of $\Pr(\phi \mid \lambda)$ values, the probability of a putative parent-offspring pair being false given the frequencies of alleles that it shares. Notice that the value of $\Pr(\phi \mid \lambda)$ is dictated by two values, $\Pr(\phi)$ and $\Pr(\lambda \mid \phi)$. Thus $\Pr(\phi \mid \lambda)$ will be small if the putative parent-offspring pair shares rare alleles (i.e $\Pr(\lambda \mid \phi)$ is small), or if $\Pr(\phi)$ is small, or both.

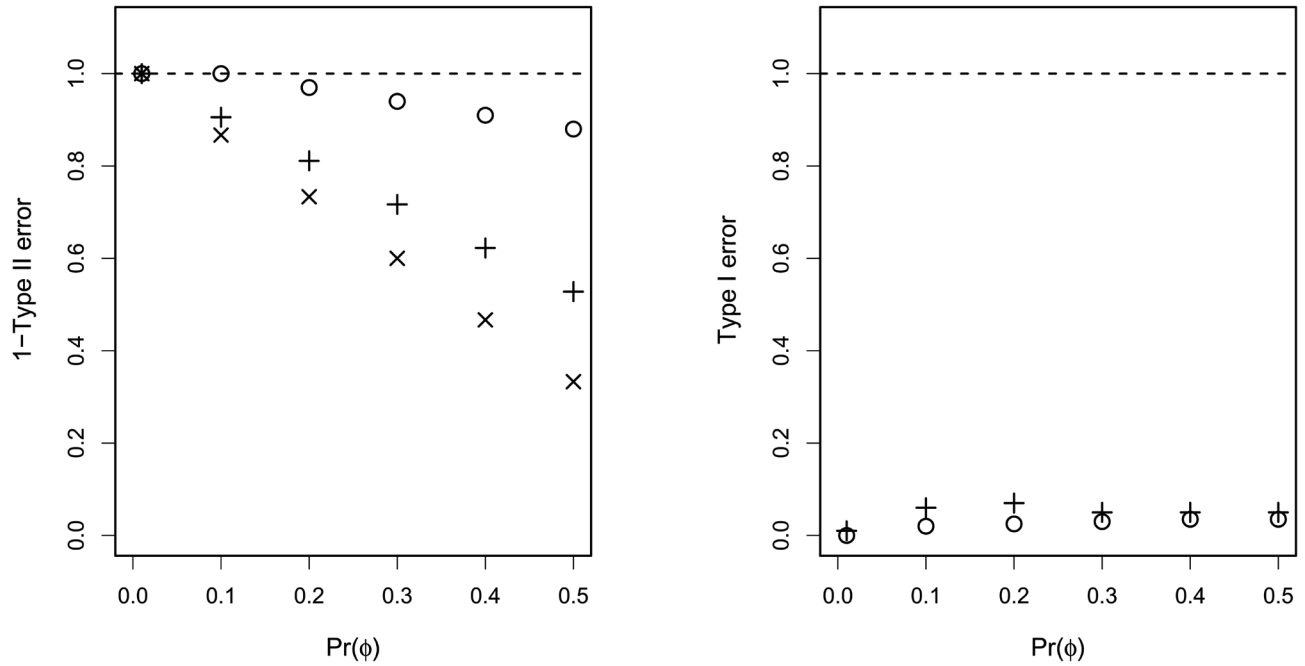


Figure 2.4. Comparisons of $\Pr(\phi \mid \lambda)$ to strict exclusion and CERVUS across varied levels of $\Pr(\phi)$. The left plot examines the proportion of true parent-offspring pairs that are correctly identified as true parent offspring pairs, 1 minus type II error, which is equivalent to the power. The right plot examines the proportion of false parent-offspring pairs that are incorrectly assigned as true parent-offspring pairs, type I error. Results for strict exclusion are shown with a dashed line. Results for $\Pr(\phi \mid \lambda)$ are shown with ○. Results for CERVUS are shown with a + for data sets where the number of candidate parents was set to 500, and with an × where the number of candidate parents was set to 1000. Type I error for CERVUS was only reported for data sets where the number of candidate parents was set to 500, though the results were nearly identical for the data sets with 1000 candidate parents.

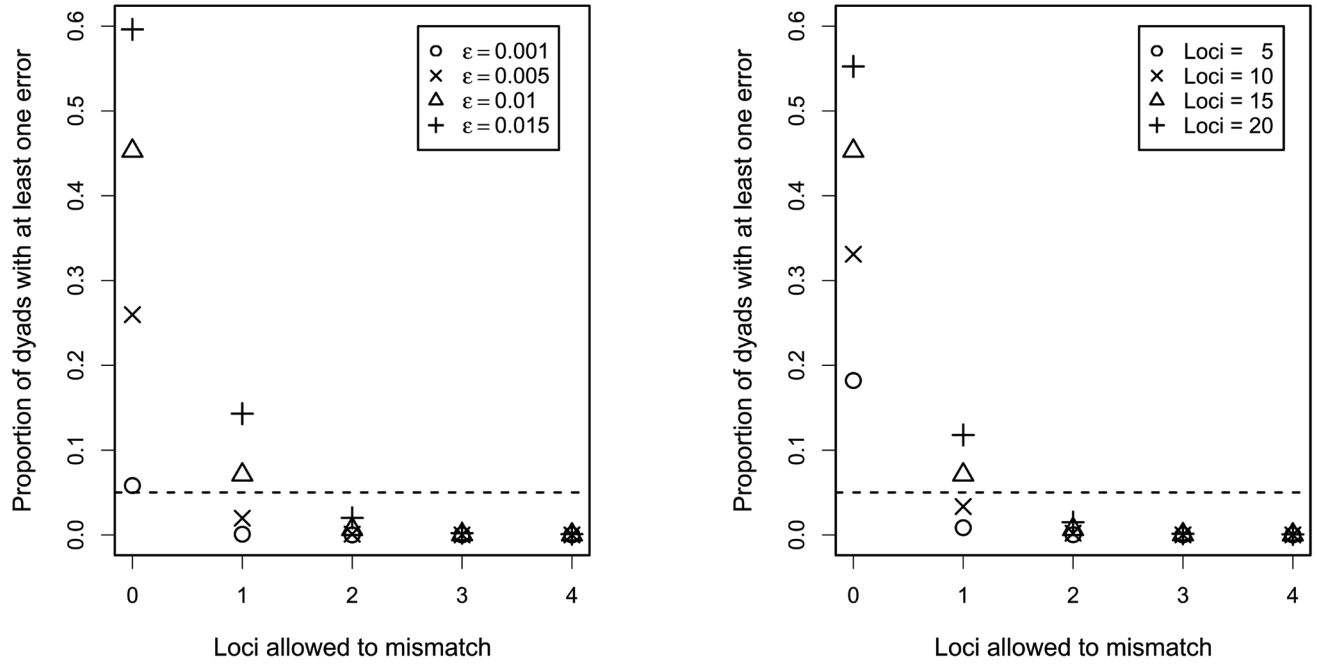


Figure 2.5. The proportion of dyads with at least one error as a result of allowing different numbers of loci to mismatch. The proportion of dyads with at least one error was calculated with equation 14 and is dependent upon the error rate and number of loci. The plot on the left examines the effect of varied error rates (ε), while holding the number of loci at 15. The plot on the right examines the effect of varied numbers of loci, while holding the error rate at 0.01.

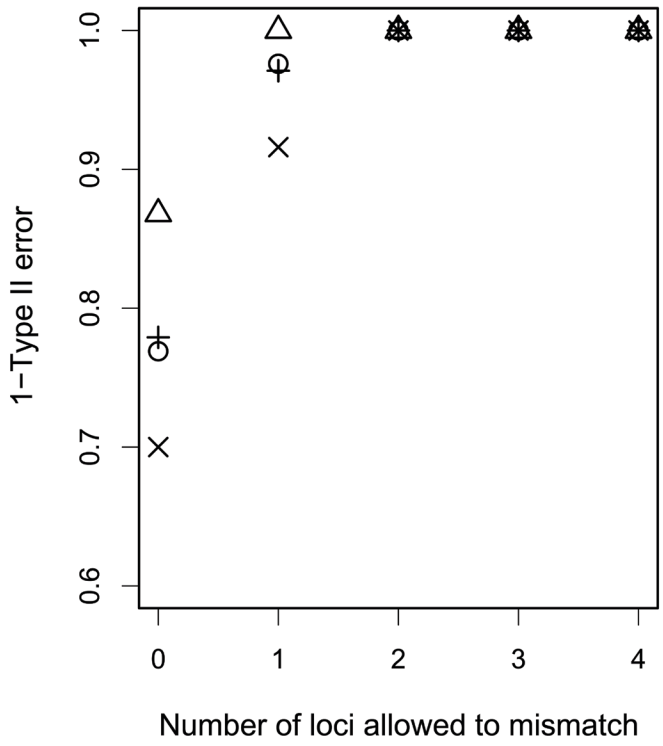


Figure 2.6**.** The proportion of true parent-offspring pairs correctly assigned, 1minus type II error (or power), as a function of the number of loci allowed to mismatch. Simulated data sets were created as follows: △ 15 loci, 0.01 error rate, + 15 loci, 0.015 error rate, ○ 20 loci, 0.01 error rate, × 20 loci, 0.015 error rate. Notice that even for a high error rate and relatively large numbers of loci, allowing for 2 loci to mismatch results in the correct assignment of all true parent-offspring pairs.

Chapter 3

Larval connectivity in an effective network of marine protected areas

Mark R. Christie, Brian N. Tissot, Mark A. Albins, James P. Beets, Yanli Jia,

Stephen E. Thompson & Mark A. Hixon

Ecology
In Review

Abstract

Acceptance of marine protected areas (MPAs) as fishery and conservation tools has been hampered by lack of direct evidence that MPAs successfully seed unprotected areas with larvae of targeted species. For the first time, we present direct evidence of large-scale population connectivity within an existing and effective network of MPAs. A novel parentage analysis identified parent-offspring pairs from a large, exploited population of the coral-reef fish *Zebrasoma flavescens* in Hawai'i, revealing larval dispersal distances ranging from 15 to 184 km. In half the cases, successful dispersal was from an MPA to unprotected areas. All offspring recruited to reefs to the north of where they were spawned, as predicted by prevailing surface currents. However, observed retention of offspring near shore was not fully predicted by simulations of passively dispersing virtual drifters, suggesting a possible role for larval behavior. Two findings together suggest that sampled sites did not contribute equally to successful recruitment: (1) comparatively low levels of genetic differentiation among all recruit samples, and (2) the fact that the 4 documented parents occurred at only 2 sites. These findings empirically demonstrate the effectiveness of MPAs as useful conservation and management tools, and highlight novel methods for identifying source populations of marine larvae.

Introduction

Connectivity in marine metapopulations is characterized by the dispersal of planktonic larvae among local populations (Kritzer and Sale 2004). Recent empirical

efforts to track larval dispersal have demonstrated localized self-recruitment (Jones et al. 1999, Swearer et al. 1999, Jones et al. 2005, Almany et al. 2007). However, connectivity within an established network of reserves has not yet been documented, though small-scale connectivity within a proposed network has recently been demonstrated for an anemonefish with short larval duration (Planes et al. 2009). The utility of MPAs as management and conservation tools for replenishing marine populations outside MPA borders depends on two processes (Russ 2002): spillover of mobile juveniles and adults into adjacent unprotected habitat, and successful seeding of unprotected sites with larvae spawned within MPAs. While there is mounting evidence for localized spillover (Mumby and Steneck 2008, Williams et al. 2009), there have been no empirically documented cases of MPAs seeding unprotected sites, which has inhibited acceptance of this management tool (Gell and Roberts 2003).

Lack of data demonstrating seeding is due to the miniscule sizes of planktonic larvae, a life history stage possessed by the majority of marine species, which makes direct tracking of dispersal extremely difficult (Ehrlich 1975). Nevertheless, marine reserves will better meet their goals with appropriate design decisions (e.g., size, spacing, location) that are guided by knowledge of dispersal patterns (Gell and Roberts 2003, Palumbi 2003). Additionally, marine metapopulations may be characterized by source-sink dynamics (Mora and Sale 2002) that can vary greatly across small spatial scales (Berkeley et al. 2004). Thus, there is a need to identify both the patterns of larval dispersal and the relative reproductive contribution of local populations in both proposed and existing networks of MPAs.

Patterns of marine larval dispersal have been examined with both indirect and direct methods (Hedgecock et al. 2007). For fishes, indirect estimation of larval connectivity has focused on microchemical signatures in otoliths (ear bones) (Thorrold et al. 2007) or population genetics inference, including genotypic assignment tests (e.g., Gerlach et al. 2007). Because these methods rely on strong spatial differences in microchemical signatures of seawater (Thorrold et al. 2007) or high levels of genetic differentiation (Manel et al. 2005), respectively, they are of limited utility in many marine systems. In comparison, direct methods focus on individual larvae. While substantial advances have been made in directly documenting dispersal with mark/recapture techniques (Jones et al. 1999, 2005, Almany et al. 2007), most methods involve the marking of otoliths, an expensive and time consuming process.

The use of genetic parentage analyses to document dispersal directly presents a largely unexplored, yet promising alternative. To date, parentage analyses have been methodologically constrained to environments where fishes with short pelagic larval durations occupy locations where all or most of the adults can be sampled (Jones et al. 2005, Planes et al. 2009). Here, we show that difficulties associated with applying parentage methods to large natural populations have been overcome by a novel Bayesian parentage method that fully accounts for large numbers of pair-wise comparisons and low probabilities of finding true parent-offspring pairs (Christie 2009). This genetic method is well suited for a broad range of systems where only a small proportion of candidate parents can be sampled (e.g., the majority of marine species). Ultimately, the merging of direct and indirect genetic methods will provide

rapid advances towards determining large-scale, real-time patterns of marine larval dispersal (Christie 2009).

On coral reefs of the Island of Hawai'i, yellow tang (*Zebrasoma flavescens)* serve an important ecological role as abundant herbivores (Walsh 1987). Yellow tang are also of substantial economic importance as they comprise 80% by number and 70% by value of all fish collected by the aquarium trade in West Hawai'i (Williams et al. 2009). Commercial collections of approximately 0.5 million fish per year were shown to have reduced adult abundance by up to 46% at unprotected sites during the 1990s (Tissot and Hallacher 2003). To sustain the aquarium fishery, a network of 9 MPAs was established along the Kohala-Kona coast in 1999, resulting in the prohibition of commercial aquarium collection along 35% of the 150 km coastline of West Hawai'i (Williams et al. 2009). Before-after comparisons following 7 years of protection clearly demonstrated the success of this network, with yellow tang abundance increasing by 72% within MPA boundaries (Tissot and Hallacher 2003). Furthermore, evidence for spillover of adults was documented near the boundaries of these MPAs (Williams et al. 2009).

These finding demonstrate that this network of MPAs is indeed effective, but do not resolve the most pressing question of whether these MPAs seed other sites with larvae. To address this question, we employed novel genetic parentage methods to directly track successful dispersal events. We coupled these results with oceanographic analyses to determine the extent to which these events are predictable and to search for deviations from a null model of passive dispersal. Lastly, we used

population genetic methods to reveal that not all of our sampled sites contributed equally to successful recruitment. These multidisciplinary approaches provide novel insights into patterns of marine larval dispersal that have profound and broad ranging implications for conservation and management decisions.

Materials and Methods

Yellow tang possess life history characteristics similar to many fishery species: large local population sizes, broadcast spawning, a pelagic larval duration exceeding a month (approximately 54 days), and a long-lived demersal life stage, up to at least 41 years (Claisse et al. 2009). Importantly, yellow tang have relatively small home ranges as both juveniles (Claisse et al. 2009) and adults (Claisse 2009), such that distances between parents and offspring greater than 1 km can be attributed to larval dispersal rather than local movements of juveniles and adults.

Adult $(n = 532)$ and recently settled $(n = 541)$ yellow tang were collected from June through August 2006 from 10 sites located around the Island of Hawai'i (Fig. 3.1, Table A1). Yellow tang were collected by SCUBA and taken to the surface, where they were measured and had a sample of their dorsal fin tissue clipped for genetic analyses. Samples were amplified at 15 microsatellite loci (Christie and Eble 2009). The genotypes of all adults were compared to the genotypes of all recruits to identify putative parent-offspring pairs that shared at least one allele at all loci. These putative pairs were genotyped at 5 additional microsatellite loci and were re-analyzed, from extraction through scoring, at all 20 loci to minimize the possibility of any

laboratory errors. A sample of 95 individuals was also genotyped at the additional 5 loci employed for parentage analysis to provide unbiased estimates of allele frequencies. For each putative parent-offspring pair, the probability of being false given the frequencies of shared alleles, $\Pr(\phi \mid \lambda)$, was calculated (Christie 2009). A prior study directly corroborated parentage approaches by tagging and tracking offspring with chemically labelled otoliths (Planes et al. 2009).

Samples were genotyped on an ABI 3100 capillary sequencer and scored with GENOTYPER software (Applied Biosystems). Data were scored, binned, and subsequently re-scored to check for errors. Independent scoring by two observers of approximately 65% of the genotypes resulted in a discordance rate of less than 0.1%. Study-specific genotyping error rates were calculated by re-genotyping 96 randomly chosen individuals at all loci. Average observed heterozygosity was 0.764 with an average allelic richness of 14.328 (range: 4 to 28). All loci were in both Hardy-Weinberg equilibrium (Table A1) and linkage equilibrium.

For parentage analyses, allele frequencies were estimated from the entire sample of adults and the entire sample of recruits. Simulations required for the calculation of $\Pr(\phi \mid \lambda)$ were conducted with 10,000 false pairs generated from over 100 null data sets (Christie 2009). None of the putative parent-offspring pairs had missing data. The study specific error-rate of 0.008 allowed for up to 2 loci to mismatch (Christie 2009), though 2 documented parent-offspring pairs matched at all 20 loci and 2 pairs mismatched at only one locus. The possibility of parent-offspring pairs actually being some other first-order relative (i.e., full sibs) was eliminated by

calculating the probability of simulated full sibs sharing an allele at 19 ($p < 0.009$) or all 20 loci ($p < 0.003$). Simulated full sibs were created in KINGROUP (Konovalov et al. 2004) with the observed yellow tang allele frequencies.

We next examined potential patterns of passive dispersal around Hawai'i during the pelagic larval duration of the four tracked offspring using the Hybrid Coordinate Ocean Model (HYCOM) (Bleck 2002). HYCOM is a fully three-dimensional ocean circulation model, which we implemented for the region that encompasses the eight main Hawaiian Islands. The model has a horizontal resolution of approximately 4 km. HYCOM model output was integrated from the earliest spawning date to the latest settlement date of tracked offspring to illustrate the mean surface (0-3m) flow patterns during dispersal. To estimate the spawning and settlement dates of the offspring, we calculated their age, in days, using a species-specific linear growth equation empirically derived for new recruits from Hawai'i (see details in Appendix A). Additionally, virtual drifters were released as close to the location of a sampling site as possible, but sufficiently far from land to prevent the drifters immediately drifting to shore. HYCOM advection-diffusion models were initiated with 961 (31 x 31) particles evenly distributed over rectangular patches of 0.03 degrees in width and length located at each site. All drifters were released on estimated spawning dates at a depth of 1.5 m. Particle positions were sampled periodically until completion of the 54-day pelagic larval duration. The drifters were permitted to take steps in a random manner to simulate the effects of sub-grid scale

processes that were not resolved by the model. The size of the steps equates to a diffusion coefficient of 10 m^2s^{-1}.

Results

We identified four independent parent-offspring pairs (Fig. 3.1, Table 3.1), which is quite remarkable given the large population sizes of yellow tang (Williams et al. 2009). None of the sampled adults mated with each other to produce the tracked offspring (i.e., each offspring was assigned to only one parent). Two parents inhabited the MPA at Miloli'i, and each of their respective offspring dispersed to unprotected reefs to the north. This is the first empirical documentation of an MPA successfully seeding unprotected areas. The two other parents were sampled at an unprotected reef near Punalu'u, and each of their respective offspring colonized separate MPAs to the north. Dispersal distances ranged from 15 to 184 km (Table 3.1).

HYCOM model output integrated from the earliest spawning date to the latest settlement date of the four tracked offspring revealed that passive surface drift would have initially carried larvae northward (Fig. 3.2A), consistent with the observation that all four offspring were found to the north of their parents. Initial trajectories of 3 of the 4 drifter simulations were northward, consistent with mean surface flow and the observed distributions of parents and offspring, whereas drifters from the fourth simulation were carried to shore within two days of release (Figs. 3.2B and 3.2C, Fig. A1). For all but the fourth simulation, the model predicted eventual offshore or up-island drift, such that passive surface dispersal would not have delivered larvae back

to the Island of Hawai'i at the time of settlement (Figs. 3.2B and 3.2C). These deviations from direct observations likely occurred because the model did not incorporate larval behavior (Cowen et al. 2000) or adequately capture near-shore oceanographic processes. Representative trajectories from virtual drifters released at all 10 of the sampled sites indicate that northern sites on the Island of Hawai'i could be sources of larvae for the remainder of the main Hawaiian archipelago (Fig. A2).

Genetic differentiation among all sampled sites was low ($F_{ST} < 0.001$) and non-significant, indicating substantial connectivity among sites over evolutionary time scales. Bolstering this conclusion, an isolation-by-distance analysis with adults and recruits as separate samples revealed no pattern (Fig. 3.3). Recruit-versus-recruit sample comparisons had lower average pair-wise F_{ST} values than expected by chance ($p < 0.004$), suggesting that recruits were more similar to each other genetically than they were to adults, and also that recruits were more similar to each other than adults were to adults (Fig. 3.3). Using a randomization test assuming equal reproductive success among sites, it is unlikely that two parents would have been detected at each of only two sites given our sampling design ($p < 0.037$). Because the sampled recruits were more similar genetically to one another than expected by chance, and because observing the four documented parents at only two sites was unlikely, we conclude that the sampled sites did not contribute equally to the successful recruitment of yellow tang (see Appendix A for statistical details).

Discussion

Our observation of larval connectivity between protected and unprotected reefs provides the first direct evidence of marine protected areas (MPAs) successfully seeding unprotected areas with larval fish. Lack of unequivocal evidence for this hypothesized seeding effect has long impeded acceptance of MPAs as useful tools for marine fisheries management and conservation. Furthermore, the small proportion of sampled adults coupled with the fact that the tracked larvae survived to become established juveniles, indicates high rates of ecologically meaningful population connectivity among these sites. We additionally demonstrated that the observed patterns of larval dispersal shortly after spawning are well explained by oceanographic processes, but that passive dispersal alone cannot account for the final settlement locations of larvae.

The combination of parentage and population genetics analyses presented here also suggests that not all of our sampled populations contributed equally to successful recruitment. If these documented patterns persist through time, then it may be prudent to protect sites located to the south of the existing network (e.g., Punalu'u). Given that the life history of yellow tang is similar to that of many fishes, including fishery species, it is likely that such source-sink dynamics are prevalent within marine metapopulations (Kritzer and Sale 2004). Understanding the intensity and temporal consistency of these sources and sinks is crucial for effective conservation and management measures, including marine reserve design (Mora and Sale 2002, Palumbi 2003).

In addition to demonstrating the seeding effect of MPAs, documenting connectivity among marine populations also has an important social and economic role. The identification of connectivity between distant reef fish populations on the island of Hawai'i demonstrates that human coastal communities are also linked: management in one part of the ocean affects people who use another part of the ocean. Understanding connections at all levels is the foundation for truly effective ecosystem-based management (Francis et al. 2007).

Acknowledgments

We thank K. Page-Albins, L. Armitage, J. Gomes, D. Jayewardene, E. Perez, T. Stevenson, and especially D. Ortiz for help in the field, A. Harris, J. Claisse, J. Eble, and A. Krupkin for helpful analytical contributions, I. Phillipsen for illustrations, D. Johnson, C. Searle, and T. Pusack for helpful reviews, and M.A. Banks, M.S. Blouin, W.J. Walsh, and S.J. Arnold for lab facilities, materials and advice. Supported by grants from Conservation International (MAH), Marine Environmental Research (MAH, JPB, and BNT), and the National Science Foundation (05-50709, 08-51162 to MAH). YJ was supported by the National Oceanic and Atmospheric Association through a grant to the International Pacific Research Center.

Author contributions

MRC, MAH, and BNT designed the project, MAH secured funding, SET provided logistic support, MAA coordinated field logistics, MAA, SET, BNT, JPB, and MAH

collected samples, MRC performed molecular and data analyses, YJ performed oceanographic modelling, and MRC and MAH drafted the manuscript.

Literature Cited

Almany, G. R., M. L. Berumen, S. R. Thorrold, S. Planes, and G. P. Jones. 2007. Local replenishment of coral reef fish populations in a marine reserve. Science 316:742-744.

Berkeley, S. A., M. A. Hixon, R. J. Larson, and M. S. Love. 2004. Fisheries Sustainability via protection of age structure and spatial distribution of fish populations. Fisheries 29:23-32.

Bleck, R. 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modeling 4:55-88.

Christie, M. R. 2009. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. Molecular Ecology Resources *in press*.

Christie, M. R., and J. A. Eble. 2009. Isolation and characterization of 23 microsatellite loci in the yellow tang, *Zebrasoma flavescens* (Pisces: Acanthuridae). Molecular Ecology Resources 9:544-546.

Claisse, J. T. 2009. The life history and movements of yellow tang, *Zebrasoma flavescens*, with implications for fisheries management using protected areas. *PhD Dissertation*. University of Hawaii.

Claisse, J. T., S. A. Mctee, and J. D. Parrish. 2009. Effects of age, size, and density on natural survival for an important coral reef fishery species, yellow tang, *Zebrasoma flavescens*. Coral Reefs 28:95-105.

Cowen, R. K., K. M. M. Lwiza, S. Sponaugle, C. B. Paris, and D. B. Olson. 2000. Connectivity of marine populations: Open or closed? Science 287:857-859.

Ehrlich, P. 1975. The population biology of coral reef fishes. Annual Review of Ecology and Systematics 6:211-247.

Francis, R. C., M. A. Hixon, M. E. Clarke, S. A. Murawski, and S. Ralston. 2007. Ten commandments for ecosystem-based fisheries management. Fisheries 32:217-233.

Gell, F. R., and C. M. Roberts. 2003. Benefits beyond boundaries: the fishery effects of marine reserves. Trends in Ecology & Evolution 18:448-455.

Gerlach, G., J. Atema, M. J. Kingsford, K. P. Black, and V. Miller-Sims. 2007. Smelling home can prevent dispersal of reef fish larvae. Proceedings of the National Academy of Sciences of the United States of America 104:858-863.

Hedgecock, D., P. H. Barber, and S. Edmands. 2007. Genetic approaches to measuring connectivity. Oceanography 20:70-79.

Jones, G. P., M. J. Milicich, M. J. Emslie, and C. Lunow. 1999. Self-recruitment in a coral reef fish population. Nature 402:802-804.

Jones, G. P., S. Planes, and S. R. Thorrold. 2005. Coral reef fish larvae settle close to home. Current Biology 15:1314-1318.

Konovalov, D. A., C. Manning, and M. T. Henshaw. 2004. KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. Molecular Ecology Notes 4:779-782.

Kritzer, J. P., and P. F. Sale. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries 5:131-140.

Manel, S., O. E. Gaggiotti, and R. S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology & Evolution 20:136-142.

Mora, C., and P. F. Sale. 2002. Are populations of coral reef fish open or closed? Trends in Ecology and Evolution 17:422-428.

Mumby, P. J., and R. S. Steneck. 2008. Coral reef management and conservation in light of rapidly evolving ecological paradigms. Trends in Ecology & Evolution 23:555-563.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications 13:S146-S158.

Planes, S., G. P. Jones, and S. R. Thorrold. 2009. Larval dispersal connects fish populations in a network of marine protected areas. Proceedings of the National Academy of Sciences of the United States of America 106:5693-5697.

Russ, G. R. 2002. Yet another review of marine reserves as reef fishery management tools. Pages 421-443 *in* P. F. Sale, editor. Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem. Academic Press, San Diego.

Swearer, S. E., J. E. Caselle, D. W. Lea, and R. R. Warner. 1999. Larval retention and recruitment in an island population of coral-reef fish. Nature 402:799-802.

Thorrold, S. R., D. C. Zacherl, and L. A. Levin. 2007. Population Connectivity and Larval Dispersal Using Geochemical Signatures in Calcified Structures. Oceanography 20:80-89.

Tissot, B. N., and L. E. Hallacher. 2003. Effects of Aquarium Collectors on Coral Reef Fishes in Kona, Hawaii. Conservation Biology 17:1759.

Walsh, W. J. 1987. Patterns of recruitment and spawning in Hawaiian reef fishes. Environmental Biology of Fishes. 18:257-276.

Williams, I. D., W. J. Walsh, J. T. Claisse, B. N. Tissot, and K. A. Stamoulis. 2009. Impacts of a Hawaiian marine protected area network on the abundance and fishery sustainability of the yellow tang, *Zebrasoma flavescens*. Biological Conservation 142:1066-1073.

Table 3.1. Probabilities and dispersal distances of parent-offspring pairs (see Fig. 3.1).

Sample site		False pair probability	Dispersal distance
Parent	Offspring	$\Pr(\phi \mid \lambda)$	(km)
Miloli'i	Ho'okena	0.0038	15.4
Miloli'i	Wawaloli	0.0013	64.9
Punalu'u	Honokohau	0.0272	140.1
Punalu'u	Anaeho'omalu	0.0109	184.2

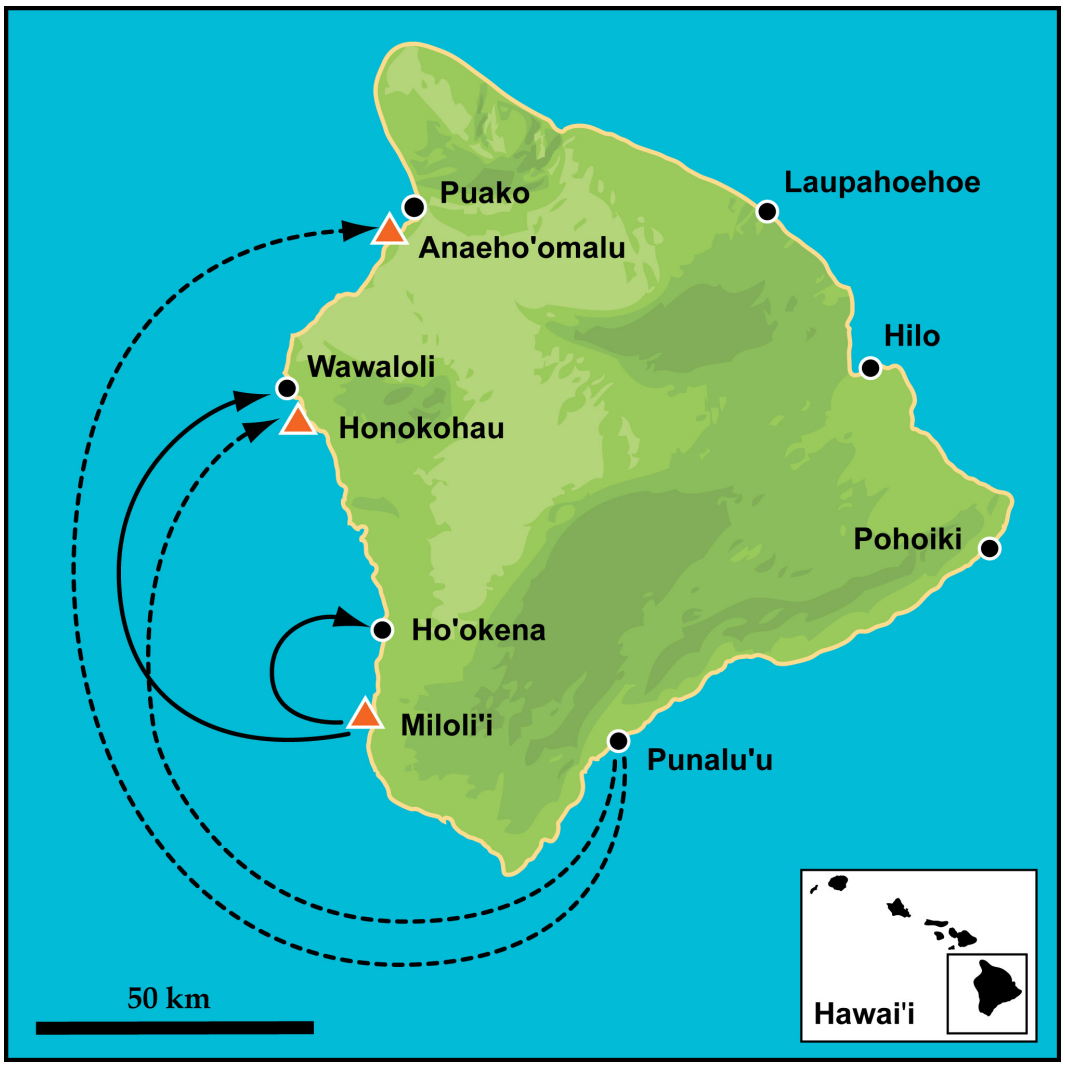


Figure 3.1. Patterns of larval connectivity in yellow tang off the Island of Hawai'i as directly determined by four independent parent-offspring pairs. Ten sample sites are indicated by triangles and circles, where triangles represent marine protected areas (MPAs) and circles represent unprotected areas. Solid lines indicate the first unequivocal evidence of an MPA seeding unprotected sites, and dashed lines indicate larval dispersal from an unprotected site to MPAs.

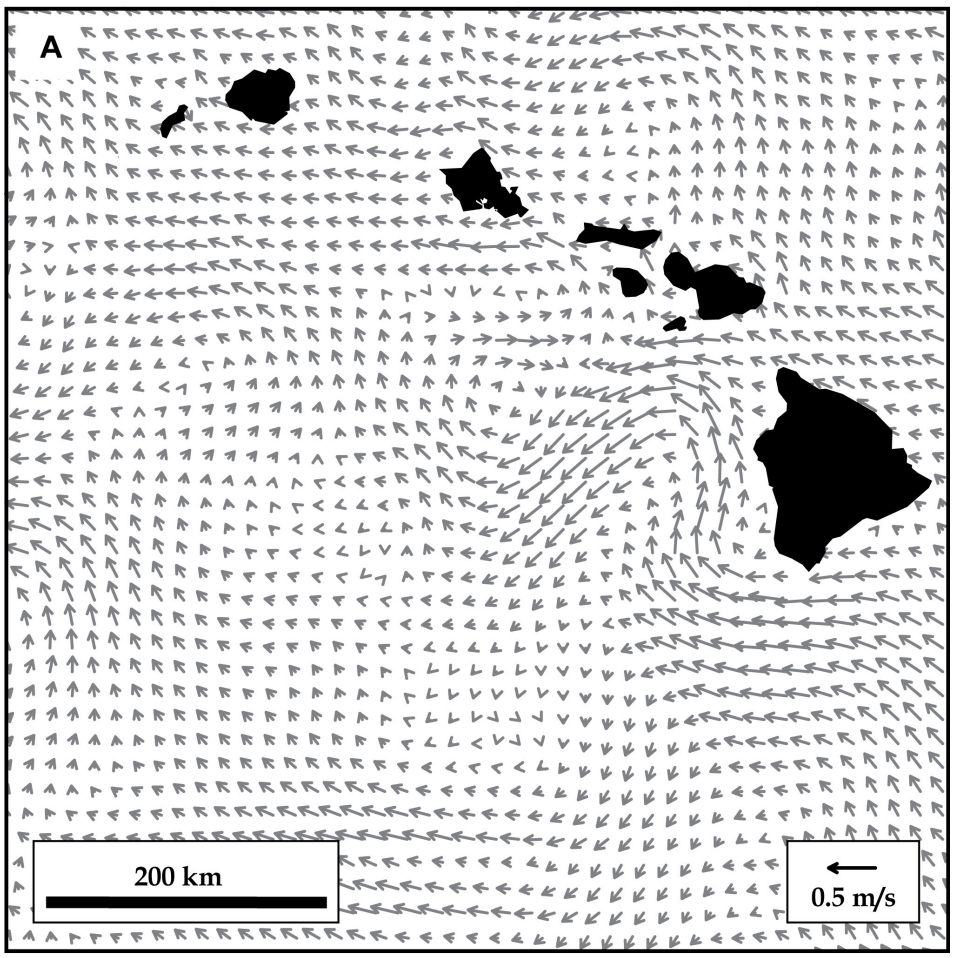


Figure 3.2. Surface currents and simulated patterns of dispersal of passive virtual drifters off the Island of Hawai'i. (**A**) Mean currents over the larval dispersal period of the 4 documented offspring (18 April to 6 July 2006).

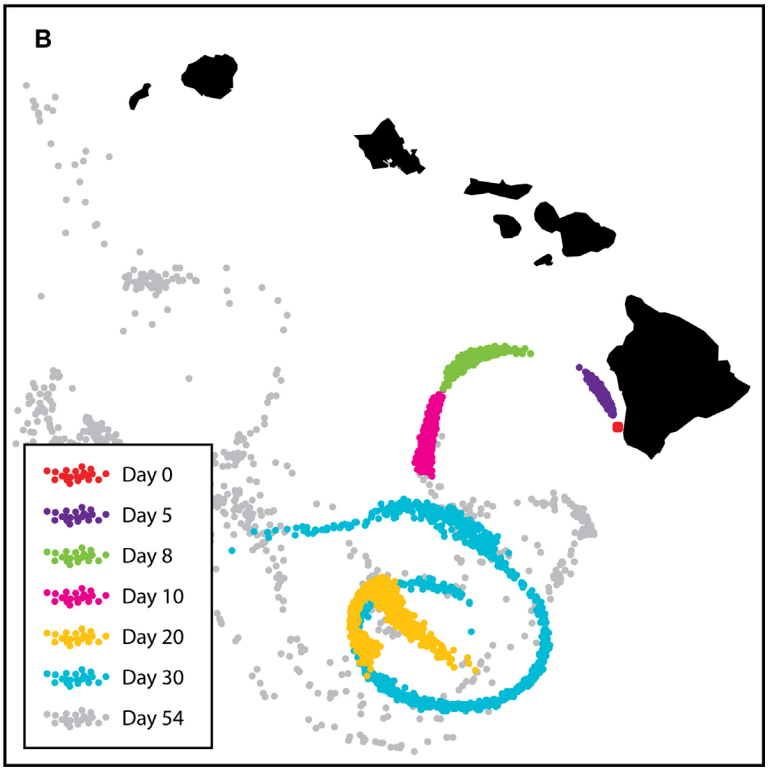


Figure 3.2 (Continued). (**B**) Passive dispersal of drifters released from Miloli'i at the date of spawning of a documented offspring (26 April 2006).

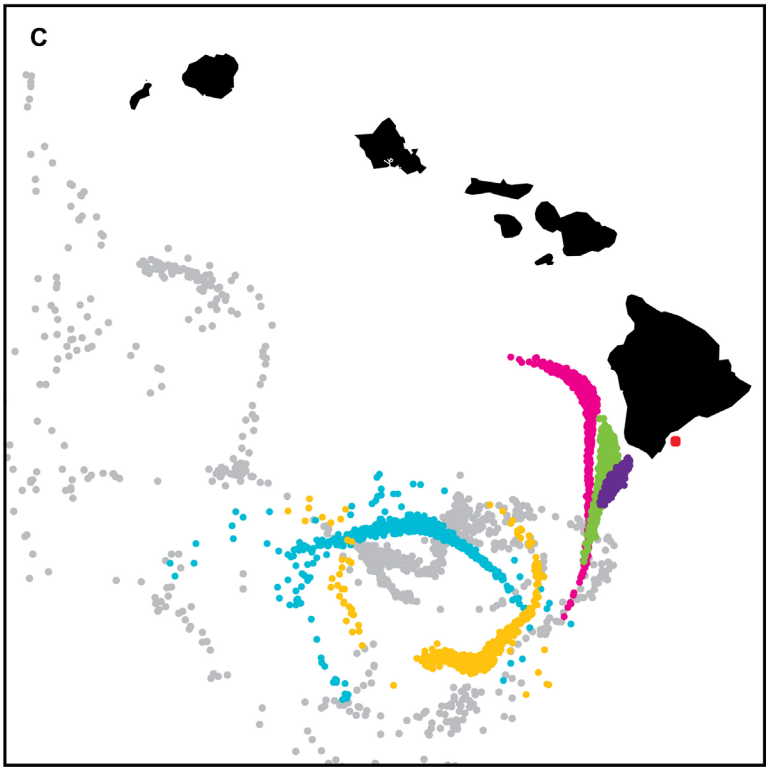


Figure 3.2 (Continued). (**C**) Passive dispersal of drifters released from Punalu'u at the date of spawning of another documented offspring (24 April 2006). Note that, in both cases, initial post-spawning dispersal was northward, but subsequent dispersal for the remainder of the pelagic larval period was away from the Island of Hawai'i. See supporting information for model outputs of the other two documented offspring.

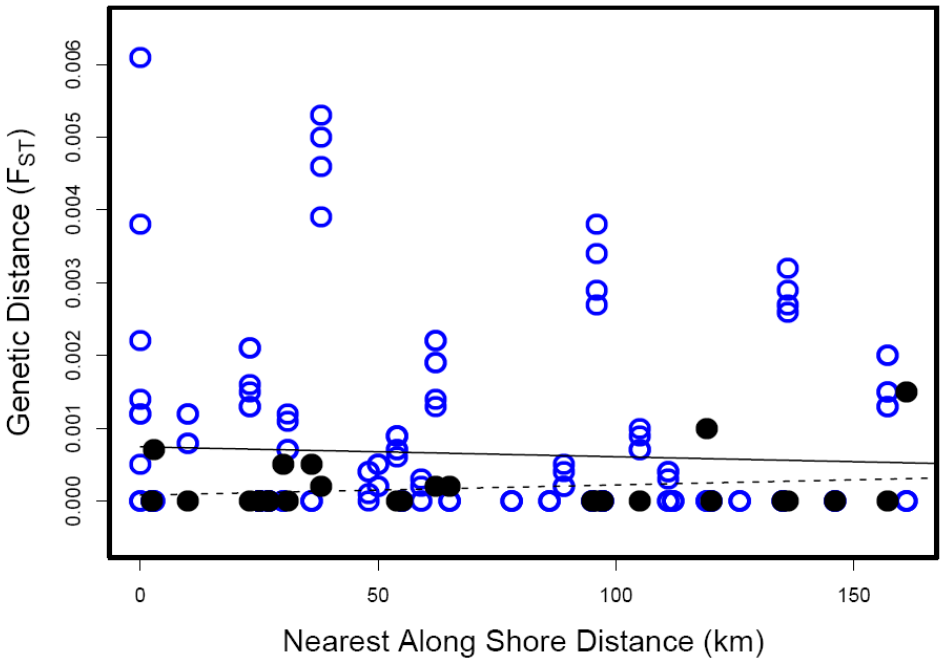


Figure 3.3. Pattern of isolation-by-distance among yellow tang genetic samples collected at the Island of Hawai'i. The nearest alongshore distance among sample sites is plotted against genetic distance as measured by F_{ST}. Adult and recruit collections are treated as separate populations. Recruit-versus-recruit comparisons are noted with filled black circles and all other comparisons are noted with open blue circles. There is no pattern of isolation-by-distance, which is expected with high connectivity among sites. Furthermore, F_{ST} values among recruits are lower than expected by chance ($P < 0.004$).

Chapter 4

Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish

Mark R. Christie, Darren W. Johnson, Christopher D. Stallings & Mark A. Hixon

Molecular Ecology
Submitted

Abstract

Identifying patterns of larval dispersal within marine metapopulations is vital for effective fisheries management, appropriate reserve design, and conservation efforts. We employed genetic markers (microsatellites) to determine dispersal patterns in bicolor damselfish (Pomacentridae: *Stegastes partitus*). Tissue samples of 771 fish were collected in 2004 and 2005 from 11 sites encompassing the Exuma Sound, Bahamas. Bayesian parentage analysis identified two parent-offspring pairs, which directly documented self-recruitment at the two northern-most sites. Principal coordinates analyses of pair-wise relatedness values further indicated that self-recruitment was common in all sampled populations. Nevertheless, measures of genetic differentiation (F_{ST}) and results from assignment methods suggested high levels of gene flow among populations. Comparisons of heterozygosity and relatedness among samples of adults and recruits indicated spatially and temporally independent sweepstakes events. The novel results presented here reveal that self-recruitment and sweepstakes reproduction are the predominant, ecologically-relevant processes that shape patterns of larval dispersal in this system.

Introduction

The vast majority of marine organisms have a planktonic larval stage. How far and to what extent larvae disperse from their natal sites remains a pressing question in marine ecology, conservation biology, and fisheries biology. Answers to these questions have vast ramifications for understanding metapopulation dynamics (Hixon *et al.* 2002;

Kritzer & Sale 2004), enhancing marine reserve design (Botsford *et al.* 2003; Palumbi 2003,), and facilitating fisheries management (Gell & Roberts 2003; Francis *et al.* 2007). Because the larvae of most marine organisms are miniscule, it is extremely difficult to observe and track them *in situ*. Consequently, early approaches for determining dispersal patterns focused on predictive models of passive larval transport (e.g., Roberts 1997). Results from such studies resulted in the common assumption that the vast majority of marine populations were demographically open and characterized by high levels of larval connectivity (Cowen *et al.* 2000). More recent work, using both genetic and microchemical analyses, have demonstrated that self-recruitment – the return of larvae to their natal population – may be more common than previously thought (Jones *et al*. 2005; Almany *et al.* 2007).

Reconciling these conflicting patterns of larval dispersal remains challenging because most marine populations cannot be simply categorized as closed or open (Cowen *et al.* 2000), but rather occur along a dynamic continuum of self-recruitment and population connectivity. Understanding the full complexity of dispersal patterns requires sampling of multiple cohorts (i.e. multiple dispersal events) both spatially and temporally. Furthermore, the majority of marine species have high rates of gene flow over evolutionary time scales (Hedgecock *et al.* 2007). Determining the extent to which populations are connected, in spite of high gene flow, remains the single greatest challenge for revealing ecologically meaningful patterns of larval dispersal.

Spatial patterns of larval dispersal can be detected with either direct or indirect methods (Hedgecock *et al.* 2007). Indirect methods focus on population-level analyses and often require theoretical assumptions (e.g., drift-mutation equilibrium). Such methods are often plagued by a lack of statistical power for detecting ecologically relevant patterns of connectivity when faced with moderate to high levels of gene flow (Wang 2004). Nevertheless, when the appropriate conditions are met, certain indirect methods can effectively reveal broad patterns of larval dispersal (Manel *et al.* 2005; Saenz-Agudelo *et al.* 2009). Direct methods, on the other hand, focus on tracking individual larvae from birth to settlement usually via mark/recapture methods. For example, fairly elaborate methods have been developed to tag the otoliths (ear stones) of fishes with various elemental markers (Thorrold *et al.* 2006). However, such tagging methods are often quite expensive and can be limited by logistical constraints such as limited mark duration and the need for multiple collection trips.

One underexplored direct method of tracking marine larvae is parentage analysis (Christie 2009). To date, parentage analyses have been used to determine dispersal patterns only in fishes with short pelagic larval durations in populations where all of the adults can be sampled (Jones *et al.* 2005; Planes *et al.* 2009). Here, we overcame difficulties of applying parentage methods to large natural populations by employing a novel Bayesian parentage method that fully accounts for large numbers of pair-wise comparisons and small or unknown proportions of sampled parents (Christie 2009). Given the large population sizes and potentially vast dispersal

distances of many marine species, it remains likely that even large data sets may record few direct observations of larval dispersal. Thus the coupling of both direct and indirect methods will likely reveal greater insights than either approach alone.

Besides parentage, other tests of relatedness within marine species hold much promise. Analyses that focus on cohorts of settling recruits can yield important spatial and mechanistic insight into patterns of larval dispersal (Selkoe *et al.* 2006). One important process is the "sweepstakes effect," where a small proportion of the available gene pool successfully contributes to the replenishment of the population (Hedgecock 1994a, b). Because the majority of adults do not successfully reproduce, the characteristic signatures of a sweepstakes effect include reduced genetic diversity and increased levels of relatedness in cohorts of recruits when compared to adults. The sweepstake hypothesis further predicts that recruits should have less within-cohort but greater among-cohort genetic diversity than adults (Hedgecock *et al.* 2007). While most studies indicate that sweepstakes effects are likely caused by stochastic larval mortality, a similar pattern could be created before the pelagic larval stage if select adults (e.g., the largest individuals) produce either more offspring (Berkeley *et al.* 2004) or offspring that are more likely to survive (D.W. Johnson *et al.* unpublished data). Regardless of the mechanisms underlying sweepstakes effects, documenting such patterns over both spatial and temporal scales can reveal detailed insights into the patterns of larval dispersal (Hedgecock 1994b).

The bicolor damselfish (*Stegastes partitus*) is an ideal species for studying patterns of larval dispersal because they are ubiquitously distributed throughout the

Bahamas and Caribbean. Furthermore, very little post-settlement movement occurs (adult movement is on the order of meters), such that any geographic distances between parents and offspring can be attributed solely to larval dispersal. Nevertheless, bicolor damselfish possess large population sizes and high rates of gene flow typical of most targeted marine species. Bicolor damselfish males guard demersal eggs, with males often simultaneously defending several clutches (Knapp *et al.* 1995). The eggs hatch 3.5 days after spawning, and the larvae are planktonic for approximately 28 days (Wilson & McCormick 1999).

A large-scale population-genetics study of bicolor damselfish revealed significant isolation-by-distance at spatial scales around 1000km (Purcell *et al.* 2009), suggesting that there is little gene flow among distant sites. Region-wide genetic comparisons indicated that Exuma Sound, Bahamas - our study system - is isolated from other sites in the Caribbean (Purcell *et al.* 2009). Additionally, the Grand Bahama Bank, a wide but shallow limestone shelf that encompasses the Exuma Sound, likely acts as a barrier to larval dispersal both into and out of the sound (Gutierrez-Rodriguez & Lasker 2004; Stoner & Davis 1997). Within the Exuma Sound, complex oceanographic patterns likely influence patterns of larval dispersal. Seasonal mesoscale gyres could entrain larvae and provide a mechanism for larval transport between reefs located on different sides of the sound (Hickey, *in preparation*). Furthermore, general northwesterly surface currents could result in the along-shore transport of larvae from southern to northern reefs (Colin, 1995).

Here, we address two pressing questions regarding ecologically relevant patterns of dispersal in bicolor damselfish: (1) to what extent do larvae return to their natal populations (self-recruitment) versus disperse among local populations, and (2) to what spatial and temporal extent do sweepstakes effects occur? We conclude that self-recruitment and local sweepstakes events are the central processes that influence patterns larval dispersal at short time scales.

Materials and Methods

Sample collection

Tissue samples were gathered from 771 *Stegastes partitus* collected from 11 sites within the Exuma Sound, Bahamas, during 2004 and 2005 (Figure 4.1, Table B.1). Adults (>5cm total length) and recruits (<2.5cm total length) were collected via hand nets by pairs of SCUBA divers. A solution of 10% quinaldine to 90% methanol was used to anesthetize the damselfish before live capture. Tissue was clipped from the pelvic fins of adults and placed in a urea-based storage solution consisting of 10mM Tris, 125mM NaCl, 10mM EDTA, 1% SDS, 8M urea, pH adjusted to 7.5 with HCL (J.F.H. Purcell personal communication). After sampling, adults were returned unharmed to their original collection location on the reef. Caudal fin tissue was collected from recruits, which were preserved for future analyses.

In 2004, sampling was concentrated along the western edge of the Exuma Sound (Figure 4.1, Table B.1). Tissue was collected from 315 fish from six sites. In 2005, sampling was expanded to include sites located around the entire Exuma Sound.

Approximately fifty adults and fifty recruits were collected at five sites for a total of 456 fish. For both years, sampling was conducted from June to August, which encompasses the peak spawning and recruitment period for bicolor damselfish.

DNA extraction and microsatellite typing

DNA was extracted using a protocol optimized for samples stored in urea-based buffer (J.F.H. Purcell personal communication). Tissue was incubated in extraction buffer (75mM NaCl, 25mM EDTA, 1% SDS) along with proteinase K (2µL of 20mg/ml) at 55°C for 2 hours. After incubation, one half volume of ammonium acetate (7.5M) was added. Samples were centrifuged and genomic DNA was precipitated from the resulting supernatant with standard isopropanol and ethanol washes (Sambrook & Russell 2001).

Samples were genotyped at seven microsatellite loci originally described by Williams *et al.* (2003). The seven loci employed in this study were *Sp*GATA16, *Sp*GATA40, *Sp*AAT40, *Sp*AAC33, *Sp*AAC41, *Sp* AAC42, and *Sp*AAC47. PCR reactions contained 1.5mM $MgCl_2$, 0.2mM dNTPs, 0.2 U *Taq* DNA polymerase (Promega), 10 µM of each primer, and 2.0µL of approximately 100ng/µL template in a total reaction volume of 15 µL. Thermocycling profiles consisted of an initial denature at 94°C for 4 minutes followed by 35 cycles of 30 seconds at 94°C, 45 seconds at 52°C, and 45 seconds at 72°C. All loci had an optimal annealing temperature of 52°C, except for *Sp*GATA40 (60°C), and *Sp*AAC41, *Sp* AAC42 (55°C). PCR products were screened on an ABI 3100 automated sequencer (Applied

Biosystems). Allele sizes were determined with the fragment analysis software GENOTYPER 3.7. Approximately 5% of individuals were re-processed through the entire procedure to remedy difficulties with scoring alleles and to regenotype individuals that were homozygous at the most polymorphic loci (see methods in Morin *et al.* 2009). A further 96 individuals were re-processed to calculate a study-specific error rate. After this process, a total of 268 out of 12138 scored alleles (2.21%) remained as missing data.

All data were tested for departure from Hardy-Weinberg equilibrium (HWE) within each population by locus and over all loci using GENEPOP 3.4 (Raymond & Rousset 1995). A total of 1000 batches and 5000 iterations per batch were employed to reduce the standard errors below 0.01. GENEPOP was additionally used to calculate observed and expected heterozygosities. The mean number of alleles per locus, mean allelic richness, and observed number of alleles were calculated with FSTAT version 2.9.3.2 (Goudet 2001). Additionally, randomization tests (21000 randomizations) were conducted using FSTAT to detect significant F_{IS}. MICRO-CHECKER was employed to determine whether any deviations from HWE were due to null alleles or large allele drop-out, as well as to check for stuttering (Van Oosterhout *et al.* 2004). Both GENEPOP (1000 batches, 2000 iterations) and GENETIX (5000 permutations) (Belkhir *et al.* 2002) were employed to test for linkage disequilibrium at all locus pairs and over all populations.

Parentage methods

The multi-locus genotypes of all adult damselfish were compared to the multi-locus genotypes of all recruits. The study-specific genotyping error rate of approximately 0.014 allowed for up to 1 locus to mismatch (see methods in Christie 2009). All pairs that shared at least one allele at six out of seven loci were considered putative parent-offspring pairs. The putative pairs were entirely reanalyzed, from extraction through scoring, to minimize the possibility of laboratory error. Due to a substantial decrease in power by allowing 1 locus to mismatch, all putative parent-offspring pairs that continued to mismatch at one locus were discarded. None of the putative parent-offspring pairs had missing data. For all other individuals, missing data were coded as the most common allele, which is a conservative approach. For each putative parent-offspring pair, we calculated the probability of the pair being false given the frequencies of shared alleles, $\Pr(\phi \mid \lambda)$ (Christie 2009). This method employs Bayes' theorem to fully account for the exclusion probability of each locus, while also accounting for the frequencies of shared alleles. Within this framework, shared rare alleles decrease the probability of a putative pair being false because it is an unlikely event. Furthermore this method, unlike many commonly implemented approaches, is not affected by differences in allele frequencies between adult and recruit samples. Simulations required for the calculation of $\Pr(\phi \mid \lambda)$ were conducted with 10,000 false pairs generated from over 100 null data sets.

To assess the possibility of parent-offspring pairs being a different first-order relative (i.e., full sibs), we calculated the probability of full-sib pairs being indistinguishable from parent-offspring pairs ($p < 0.025$) (Goodnight & Queller 1999).

To generate this p-value we created 10,000 simulated full sibs in KINGROUP (Konovalov *et al.* 2004) with the observed damselfish allele frequencies. Additionally, it is unlikely that two siblings of such vastly different sizes (adult vs. recruit) would be alive at the same time. Both parents of such full-sibs would have to be over 2 years old, given known bicolor damselfish growth rates and average size at maturity. Because, on average, each individual damselfish had less than an 8% chance of surviving to two years, this event was quite unlikely (M.A. Hixon *et al.* unpublished data).

Population structure

We performed multiple tests for the presence of population genetic structure. Adults and recruits were treated as separate samples, though pooling did not alter findings. Global and pair-wise F_{ST} values among all populations were calculated with FSTAT. Exact tests for allelic differentiation among populations were performed in GENEPOP with 1000 batches and 5000 iterations per batch. We also employed assignment methods to search for fine-scale population structure by using STRUCTURE (Pritchard *et al.* 2000), BAPS (Corander *et al.* 2004), and GENECLASS (Piry *et al.* 2004).

To examine patterns of dispersal by comparing shared alleles, we calculated pair-wise relatedness values among all 771 individuals using Queller and Goodnight's (1989) relatedness metric as implemented in GENALEX (Peakall & Smouse 2006). This relatedness metric simply describes the number of shared alleles between pairs of

individuals and standardizes this value based on the individual's allelic state (e.g., heterozygous) and on the frequency of the alleles in the reference population. To visualize the results of this analysis, we conducted a principal coordinates analysis (PCoA) on the pair-wise relatedness matrix. Individuals that share identical alleles would occupy the same location in multivariate space, while individuals with different and rare alleles would be occupy distant locations in multivariate space. PCoA performs well with a wide variety of distance measures (McCune & Grace 2002; Jombart *et al.* 2009) and is well suited for a pair-wise relatedness matrix. We repeated this analysis with pair-wise relatedness matrices calculated with other relatedness metrics (e.g., Lynch & Ritland 1999), which produced similar patterns but with more outliers.

To statistically evaluate whether our sample groups (adults or recruits collected from different sites) occupied different regions of multivariate space we performed multi-response permutation procedures (MRPP). This method calculates the average multivariate distance within each group and determines whether the average within-group distance is significantly smaller than the average within-group distances generated by random assignment of individuals to groups (Mielke and Berry 2001; McCune & Grace 2002). We used Euclidean distances and 10,000 permutations for each comparison. Analyses were performed within the R statistical software environment (scripts available from corresponding author upon request) (R Development Core Team 2009). Test statistics were compared to a Pearson Type III distribution with mean, variance and skewness calculated from permuted datasets

(McCune & Grace 2002). We performed MRPP for all PCoA groups as well as for each between-site comparison to determine whether the observed pattern was different than expected by chance.

Lastly, we examined our data for temporal differences between 2004 and 2005 at Lee Stocking Island (LSI), the only site sampled both years, using exact tests. We further examined sweepstakes recruitment in our 2005 data by conducting a principal coordinates analysis (PCoA) on pair-wise F_{ST} values using GENALEX. Note that the analysis F_{ST} identifies differences in allele frequencies between all sampled populations, while the PCoA on pair-wise relatedness values (see above paragraphs) examines shared alleles among all individuals. We also calculated within-population levels of heterozygosity and relatedness. For both measures, we calculated the mean across populations, but within groups (adults or recruits), and calculated 95% confidence intervals using a *t* multiplier with 4 degrees of freedom.

Results

General genetic patterns

The mean number of alleles per locus ranged from 20.9 to 30.0 across populations. Allelic richness over all loci, calculated from a minimum sample size of 26 individuals, ranged from 18.5 to 20.5 (Table 4.1). The observed heterozygosity over all populations and loci was 0.89 and ranged from 0.84 to 0.93. Loci *sp*GATA16 and *sp*GATA40 were approximately twice as polymorphic as the other five loci, with 70 and 78 alleles, respectively, sampled throughout the entire Exuma Sound. The number

of private alleles per population ranged from zero to seven, with the most being found in Cat Island adults.

There was no evidence for large-allele drop-out or stuttering at any locus or population, as determined by MICRO-CHECKER. Null alleles were suggested as a possible cause for departure from HWE for two loci at four of the eighteen sample locations. This problem was resolved after homozygous individuals were re-genotyped (see Methods). None of the seven loci had more than two significant departures from HWE across all 18 populations, suggesting no problems with null alleles. Most of the occurrences of loci being out of HWE occurred in populations of recruits (Table 4.1), which is suggestive of sweepstakes effects (see Discussion). Seventeen of the eighteen populations showed no support for evidence of linkage disequilibrium (Table 4.1). One population, LSI recruits in 2005, had a significant percentage of loci in linkage disequilibrium.

Self-recruitment

Remarkably, two parent-offspring pairs were identified, directly documenting self-recruitment at the two northern-most sites in Exuma Sound (Figure 4.1). One pair was sampled at Eleuthera, ($\Pr(\phi \mid \lambda) = 0.036$), and the other pair was sampled at the Land and Sea Park, ($\Pr(\phi \mid \lambda) = 0.011$). No parent-offspring pairs between any two sites were detected. Given the relatively small sample sizes, it is remarkable that any parent-offspring pairs were identified and is suggestive of high rates of self-recruitment at the two northern sites.

Evidence for self-recruitment within bicolor damselfish populations located in the Exuma Sound was further bolstered by results from PCoA of pair-wise relatedness values (Figure 4.2), where the first principal coordinate explained 23.19% of the total variation and the second principal coordinate explained 18.62% of the total variation. MRPP analysis revealed that it is very unlikely to have observed this overall pattern by chance ($T = -12.36,\ p < 0.001$). Although the analysis was performed for all individuals jointly, figure 4.2 is displayed by population for graphical clarity. Thus, the relative positions in two-dimensional ordinate space of all individuals within and among populations are accurately depicted. The adults and recruits within each population demonstrate extensive overlap (Figure 4.2), which is highly suggestive of self-recruitment within each population and further supports the results from parentage analysis. Additionally, the pair-wise MRPP analyses reveal that no adults and recruits from the same site were significantly different from one another and all within site comparisons had low effect sizes (Table 4.2, Appendix C), where effect size measures the strength of the difference between the two groups (McCune & Grace 2002). Each sample of recruits had a lower effect size when compared to adults from the same sample site than when compared to the average effect size of adults from all other populations (Table 4.2), which suggested that recruits shared more alleles with adults from their own sampling location than any other location. Additionally, the majority of pair-wise comparisons among sample sites were significant (Appendix C), indicating differences among sites.

Pair-wise F_{ST} values among all samples were low (range: 0 to 0.013) and only one out of 153 pair-wise comparisons were significantly greater than 0 (Appendix D). Furthermore, Bayesian clustering programs did not identify more than one population. Assignment tests lacked power to reliably assign recruits to parental populations. All of these results are indicative of substantial gene flow among populations at evolutionary and possibly ecological timescales.

Sweepstakes reproduction

Recruit allele frequencies at Lee Stocking Island were significantly different between 2004 and 2005 (exact tests, Table 4.3). No within-year comparisons of recruits were significant. The principal coordinate analysis of pair-wise F_{ST} values, for which both axes explained 44.34% of the total variation, provided further evidence suggestive of a sweepstakes effect (Figure 4.3). All adult populations were much more similar to one another than the recruit populations were to each other. Additionally, no recruit populations clustered near the adult populations, suggesting that recruit populations had different allele frequencies from one another and from the adult populations. These differences in allele frequencies are best explained as a striking consequence of spatially independent sweepstakes effects as opposed to recruits coming from distinct natal sources (see below and discussion).

Further signatures of a sweepstakes effect were indicated by examining differences in average heterozygosity and average relatedness among adults and recruits. The average level of heterozygosity among adult populations was

significantly higher than the average heterozygosity among recruits (Figure B.1), which is expected if only a portion of the adults successfully reproduced. Additionally, average relatedness was significantly higher among recruits than adults (Figure B.1), which is a strong indicator of sweepstakes patterns as recruits coming from a subset of adults would be expected to share more alleles. When all recruit samples were pooled, relatedness was negative (mean: -0.0010, 95% CI: -0.0012 to -0.0008). This pattern was not surprising given the dissimilarity among recruit samples displayed in Figure 4.3 and suggests that each sample of recruits came from independent sweepstakes events. Three trends provide additional evidence of sweepstakes reproduction: (1) recruit samples tended to have higher levels of linkage disequilibrium among pairs of loci, (2) greater numbers of alleles were found in adult samples than recruit samples from the same site, and (3) less within-cohort but greater among-cohort genetic diversity was observed in recruits compared to adults.

Discussion

Despite small sample sizes, two parent-offspring pairs were identified at two separate sites. Given the large population sizes of bicolor damselfish, these results strongly suggest that there are high rates of self-recruitment at these two sites. Note that both of the sites with documented parent-offspring pairs occurred in the northern Exuma Sound, suggesting that there may be oceanographic features that facilitate self-recruitment in this region. Indeed, large gyres are known to form in this region during the summer recruitment season (B.M. Hickey, in preparation)

Pair-wise relatedness analyses further indicated that self-recruitment may have been prevalent among all sites located within the Exuma Sound. All of the recruits clustered in the same multivariate space as adults from the same location, which is the pattern that would be expected if there were high levels self-recruitment. Furthermore, for 32 out of 36 possible comparisons, samples of recruits were more closely related to nearby adults than to adults at other sample sites, suggesting a high incidence of self-recruitment. Because the PCoA analysis clusters individuals based on shared alleles, this method can be used as a fine-scale assignment test. As such, large effect sizes reveal clear demarcations between both adults and recruits from Lee Stocking Island vs. the Land and Sea Park and Lee Stocking Island vs. Long Island. Also, comparisons of individuals (both adults and recruits) from Long Island vs. Eleuthera and Cat vs. Long reveal large effect sizes, and thus few shared alleles. Such results are consistent with high levels of self-recruitment, and would eventually create subtle differences among populations due to genetic drift. Furthermore, these results complement studies indicating behavioral mechanisms that would facilitate larval retention within this species (e.g., Cowen *et al.* 2000).

Nonetheless, results indicating self-recruitment must be considered within the context of our indirect analyses. Genetic differentiation as measured by F_{ST}, which employs allele frequencies as opposed to shared alleles, was low and only one pair-wise comparison was significantly different from 0. Fine-scale clustering tools such as STRUCTURE were unable to identify more than one population. These results indicate that there likely are large amounts of gene flow at evolutionary time scales.

This interpretation is further bolstered by the relatively low amount of genetic differentiation detected across a range-wide study (Purcell *et al.* 2009). Given these patterns, we conclude that there is a background of gene flow to neighboring populations that, over evolutionary time scales, has led to homogeneity in allele frequencies among sites. While it is difficult to determine how this homogeneity translates to connectivity at ecological time scales, the parentage and relatedness results suggest that connectivity among populations may occur less frequently than self-recruitment. Adding to the complexity of this system is the fact that highly polymorphic markers, while ideal for parentage analyses, make it difficult to isolate a signal for genetic differentiation from noise (Waples 1998). Additionally, homoplasy and large population sizes may further weaken any signal (Purcell *et al.* 2009). Nevertheless, the novel and coupled analyses presented here clearly demonstrate that self-recruitment is occurring amidst a background of high gene flow.

We also found that recruitment to bicolor damselfish populations is influenced by sweepstakes effects. Our analysis of F_{ST} among sample sites in 2005 (Figure 4.3) indicate that each population experiences separate sweepstakes effects, which further supports the observation of self-recruitment within each population. It is unlikely that genetic differences among recruit samples result from distant or unsampled source populations. This interpretation would require that each sample of recruits came from a genetically distinct source population, which is highly unlikely given the low overall levels of genetic differentiation observed in this study and across the entire Caribbean (Purcell *et al.* 2009). Additionally, the parentage and relatedness analyses document

high rates of self-recruitment. We therefore conclude that the striking differences in F_{ST} between adults and recruits are the result of differential reproductive success among adults within each population (i.e., sweepstakes).

Additional evidence of independent sweepstakes events comes from the observation that the average relatedness value of among recruit samples was positive, while the overall relatedness of all pooled recruits was negative. While we did detect a clear signature of a sweepstakes effect in this study system, we do not believe that the magnitude of this effect equals that of other published studies. For example, Hedgecock (2007) estimated that only 10 to 20 adult oysters produced over 185 sampled offspring. Using our genetic markers, estimates of the effective number of breeders routinely include infinity as the upper confidence limit. This outcome is likely due to the large number of alleles per locus and because such methods are not effective at accurately estimating large effective population sizes (Waples 2006). Nevertheless, we conclude that only a small portion of the potential parents contribute to subsequent generations of bicolor damselfish in Exuma Sound.

Our demographic data from these same sites reveal that larger male bicolor damselfish have more mates and greater numbers of eggs per clutch than smaller males (Johnson *et al.* in preparation). Such skewed reproductive success among males could contribute to the observed patterns of few adults contributing successfully to subsequent generations. Nevertheless, we cannot eliminate stochastic larval or post-settlement mortality as the underlying mechanism. However, it is unlikely that selection on our genetic markers caused the observed patterns because selection would

not occur across all unlinked markers and is unlikely to act differentially at the small spatial scale of this study.

Comparisons at the same site (Lee Stocking Island) from different years yielded significant differences in allelic frequencies among recruit samples. Thus, it appears that sweepstake effects occurred within sites among years as well as among sites within years. Further evidence for this effect is illustrated in Figure 4.2, where the recruits from 2004 cluster in the lower left quadrant whereas recruits from 2005 tend to cluster in the upper left quadrant. This pattern does not rule out the possibility of the largest males producing the most offspring, as there is such high inter-year mortality and turnover of reproductive adults that the largest males would likely have been replaced. Future work should focus on determining the relative importance of temporal versus spatial variation in patterns of larval dispersal.

The combination of novel direct and indirect methods used in this study provides much greater insight into patterns of marine larval dispersal than previously identified. Specifically, we have shown that self-recruitment, sweepstakes effects, and gene flow all play a role in characterizing the patterns of larval dispersal in bicolor damselfish. Detailed knowledge of both among and within population dispersal patterns is vital for improving marine conservation efforts, informing fisheries management, and advancing marine metapopulation theory (Botsford *et al.* 2009). Incorporating this knowledge into a broader theoretical and socio-political framework will also provide measurable advances towards conservation and management goals.

Acknowledgements

This work was supported by grants from the National Science Foundation (00-93976, 05-50709, 08-51162) to M.A. Hixon. J.F.H. Purcell warrants special thanks for detailed advice regarding storage, extraction and amplification of damselfish samples. We thank S.J. Arnold, M.A. Banks, and M.S. Blouin for laboratory space, valuable advice, and moral support. We also acknowledge M. Albins, K. Buch, C. Dahlgren, A. Krupkin, J. Noell, T. Pusack, C. Searle, and R. Watts for field and laboratory assistance as well as comments that greatly improved this manuscript.

Literature Cited

Almany, G. R., M. L. Berumen, S. R. Thorrold, S. Planes, and G. P. Jones. 2007. Local replenishment of coral reef fish populations in a marine reserve. Science 316:742-744.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2002. Genetix 4.03, Logiciel Sous Window TM Pour la Genetique Des Populations. Laboratoire Genome, Populations, Interactions, CNR SUMR 5000, Universite de Montpellier II, Montpellier, France.

Berkeley, S. A., M. A. Hixon, R. J. Larson, and M. S. Love. 2004. Fisheries Sustainability via protection of age structure and spatial distribution of fish populations. Fisheries 29:23-32.

Botsford, L. W., F. Micheli, and A. Hastings. 2003. Principles for the design of marine reserves. Ecological Applications 13:S25-S31.

Botsford, L. W., J. W. White, M. A. Coffroth, C. B. Paris, S. Planes, T. L. Shearer, S. R. Thorrold, and G. P. Jones. 2009. Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. Coral Reefs 28:327-337.

Christie, M. R. 2009. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. Molecular Ecology Resources *in press*.

Colin, P. L. 1995. Surface currents in Exuma Sound, Bahamas, and adjacent areas with reference to potential larval transport. Bulletin of Marine Science 56:48-57.

Corander, J., P. Waldmann, P. Marttinen, and M. J. Sillanpaa. 2004. BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20:2363-2369.

Cowen, R. K., K. M. M. Lwiza, S. Sponaugle, C. B. Paris, and D. B. Olson. 2000. Connectivity of marine populations: Open or closed? Science 287:857-859.

Francis, R. C., M. A. Hixon, M. E. Clarke, S. A. Murawski, and S. Ralston. 2007. Ten commandments for ecosystem-based fisheries management. Fisheries 32:217-233.

Gell, F. R., and C. M. Roberts. 2003. Benefits beyond boundaries: the fishery effects of marine reserves. Trends in Ecology and Evolution 18:448-455.

Goodnight, K. F., and D. C. Queller. 1999. Computer software for performing likelihood tests of pedigree relationship using genetic markers. Molecular Ecology 8:1231-1234.

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

Gutierrez-Rodriguez, C., and H. R. Lasker. 2004. Microsatellite variation reveals high levels of genetic variability and population structure in the gorgonian coral *Pseudopterogorgia elisabethae* across the Bahamas. Molecular Ecology 13:2211-2221.

Hedgecock, D. 1994a. Does variance in reproductive success limit effective population size of marine organisms? In: *Genetics and evolution of aquatic organisms* (ed. Beaumont, A.R.). Chapman and Hall, London, pp. 122-135.

Hedgecock, D. 1994b. Temporal and spatial genetic structure of marine animal populations in the California Current. CalCOFI, 35:73-81.

Hedgecock, D., P. H. Barber, and S. Edmands. 2007. Genetic approaches to measuring connectivity. Oceanography 20:70-79.

Hedgecock, D., S. Launey, A. I. Pudovkin, Y. Naciri, S. Lapegue, and F. Bonhomme. 2007. Small effective number of parents (Nb) inferred for a naturally spawned

cohort of juvenile European flat oysters *Ostrea edulis*. Marine Biology 150:1173-1182.

Hixon, M. A., S. W. Pacala, and S. A. Sandin. 2002. Population regulation: historical context and contemporary challenges of open vs. closed systems. Ecology 83:1490-1508.

Jombart, T., D. Pontier, and A. B. Dufour. 2009. Genetic markers in the playground of multivariate analysis. Heredity 102:330-341.

Jones, G. P., S. Planes, and S. R. Thorrold. 2005. Coral reef fish larvae settle close to home. Current Biology 15:1314-1318.

Knapp, R. A., P. C. Sikkel, and V. T. Vredenburg. 1995. Age of Clutches in Nests and the Within-Nest Spawning-Site Preferences of Three Damselfish Species (Pomacentridae). Copeia 1995:78-88.

Konovalov, D. A., C. Manning, and M. T. Henshaw. 2004. KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. Molecular Ecology Notes 4:779-782.

Kritzer, J. P., and P. F. Sale. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries 5, 131-140.

Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152:1753-1766.

Manel, S., O. E. Gaggiotti, and R. S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution 20:136-142.

McCune, B., and J. Grace. 2002. *Analysis of ecological communities.* 2nd edn. MjM, Gleneden Beach, OR.

Mielke, P. J., and K. Berry. 2001. *Permutation methods: A distance function approach*. Springer-Verlag, New York.

Morin, P., R. Leduc, and et al. 2009. Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources* 9:498-594.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications 13:S146-S158.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95:536-539.

Planes, S., G. P. Jones, and S. R. Thorrold. 2009. Larval dispersal connects fish populations in a network of marine protected areas. Proceedings of the National Academy of the Sciences 106:5693-5697.

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

Purcell, J. F. H., R. K. Cowen, C. R. Hughes, and D. A. Williams. 2009. Population structure in a common Caribbean coral-reef fish: implications for larval dispersal and early life-history traits. Journal of Fish Biology 74:403-417.

Queller, D. C., and K. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution 43:258-275.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria:http://www.R-project.org.

Raymond, M., and F. Rousset. 1995. GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Roberts, C. M. 1997. Connectivity and Management of Caribbean Coral Reefs. Science 278:1454-1457.

Saenz-Agudelo, P., G. P. Jones, S. R. Thorrold, and S. Planes. 2009. Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. Molecular Ecology 18:1765-1776.

Sambrook, J., and D. W. Russell. (2001). *Molecular Cloning: A Laboratory Manual.* 3rd edn. Cold Springs Harbor Laboratory Press, Cold Springs Harbor, N.Y.

Selkoe, K. A., S. D. Gaines, J. E. Caselle, and R. R. Warner. 2006. Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* 87:3082-3094.

Stoner, A., and M. Davis. 1997. Abundance and distribution of queen conch veligers (*Strombus gigas* Linne) in the central Bahamas. 1. Horizontal patterns in relation to reproductive and nursery grounds. Journal of Shellfish Research 16:7-18.

Thorrold, S. R., G. P. Jones, S. Planes, and J. A. Hare. 2006. Transgenerational marking of embryonic otoliths in marine fishes using barium stable isotopes. Canadian Journal of Fisheries and Aquatic Sciences 63:1193-1197.

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

Wang, J. L. 2004. Application of the one-migrant-per-generation rule to conservation and management. Conservation Biology 18:332-343.

Waples, R. S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89:438-450.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics 7:167-184.

Williams, D. A., J. Purcell, C. R. Hughes, and R. K. Cowen. 2003. Polymorphic microsatellite loci for population studies of the bicolor damselfish, *Stegastes partitus* (Pomacentridae). Molecular Ecology Notes, 3:547-549.

Wilson, D. T., and M. I. McCormick. 1999. Microstructure of settlement-marks in the otoliths of tropical reef fishes. Marine Biology 134:29-41.

Table 4.1**.** Summary statistics for each sample site averaged among all 7 loci. Observations include: mean number of alleles per locus, mean allelic richness, mean inbreeding coefficient (F_{IS}), observed heterozygosity (H_O), expected heterozygosity (H_E), among-loci HWE *P* value with number of loci out of HWE, percentage of loci pairs in linkage disequilibrium, and the number of private alleles.

Sample Locality	Mean # alleles/locus	Mean allelic richness	F_{IS}	H_O	H_E	HWE *P* value (# loci out)	% loci in LD†	Private alleles
Adults								
Lee Stocking Island	27.7	18.7	0.053	0.919	0.934	0.5016 (0)	0	0
Land and Sea Park	23.1	19.7	0.026	0.926	0.930	0.4950 (0)	0	1
Eleuthera	23.3	19.5	0.051	0.927	0.936	0.0826 (0)	0	2
Cat Island	25.0	20.1	0.098	0.919	0.929	0.0000* (2)	4.76	7
Long Island	22.9	19.8	0.041	0.925	0.938	0.1858 (0)	4.76	3
Compass Cay	22.0	18.7	0.061	0.903	0.935	0.0004 (2)	0	2
Bock Rock	20.9	19.4	0.031	0.901	0.93	0.2719 (0)	0	1
String Bean Cay	30.0	19.7	0.121	0.916	0.924	0.3192 (0)	0	4
Big Point	25.0	18.5	0.045	0.911	0.932	0.3984 (0)	0	3
Three Sisters Reef	21.0	20.1	0.058	0.894	0.929	0.0609 (0)	0	0
South Reef	22.3	18.7	0.077	0.885	0.927	0.0145 (1)	0	2
Recruits								
Lee Stocking Island	27.1	19.3	0.121	0.916	0.926	0.0000* (3)	28.57*	4
Land and Sea Park	23.0	19.1	0.011	0.919	0.928	0.1116 (1)	4.76	5
Eleuthera	22.9	20.5	0.046	0.915	0.928	0.0141 (2)	0	3
Cat Island	22.7	18.8	0.099	0.917	0.928	0.0000* (2)	4.76	2
Long Island	22.6	18.7	0.061	0.914	0.929	0.6336 (0)	0	3
String Bean Cay	22.6	19.1	0.022	0.856	0.908	0.3222 (0)	9.52	0
Big Point	27.3	19.4	0.081	0.856	0.931	0.0004* (2)	0	3

* Significant after a Bonferroni correction

† Calculated as number of loci pairs in linkage disequilibrium (P< 0.05) divided by the total number of comparisons

Table 4.2. Comparisons of effect sizes between recruits and adults collected from the same site (in bold) and between recruits and adults collected from all other sites. Lower effect sizes between adults and recruits collected from the same site suggest self-recruitment. Effect sizes averaged over all between-site comparisons, $\overline{A}$, reveal greater differences between recruits and adults from different sites. Effect sizes were calculated following multivariate analyses of pair-wise relatedness values.

	Adult sample						
Recruit sample	Cat	Long	Eleuthera	Park	LSI 2004	LSI 2005	$\overline{A}$
Cat	**0.025**	0.244	0.129	0.244	0.021	0.089	0.145
Long	0.117	**0.046**	0.242	0.027	0.208	0.276	0.174
Eleuthera	0.126	0.258	**0.029**	0.275	-0.047	0.032	0.129
Park	0.117	0.088	0.234	**0.049**	0.211	0.279	0.186
LSI 2004	0.169	0.337	0.111	0.366	**-0.057**	0.029	0.202
LSI 2005	0.251	0.316	0.274	0.353	0.111	**0.142**	0.261

Table 4.3. Patterns of genetic differentiation for sites located near Lee Stocking Island in 2004 (BP, BR, SB) and 2005 (LSI). Pair-wise F_{ST} values are below the diagonal. P values for exact tests of allelic differentiation are above the diagonal. Significant tests, after a Bonferroni correction, are indicated in bold. Notice that no within-year comparisons are significant, while all between-year recruit comparisons are significant.

	BP adults	BP recruits	BR adults	SB recruits	SB adults	LSI recruits	LSI adults
BP adults	--	0.342	0.219	0.132	0.046	0.099	0.991
BP recruits	0.007	--	0.016	0.286	0.000	**0.005**	0.637
BR adults	0.010	0.009	--	0.117	0.351	**0.000**	0.117
SB recruits	0.010	0.008	0.010	--	0.048	**0.004**	0.487
SB adults	0.009	0.007	0.006	0.008	--	**0.000**	0.126
LSI recruits	0.007	0.006	0.010	0.009	0.008	--	0.187
LSI adults	0.009	0.009	0.009	0.010	0.010	0.010	--

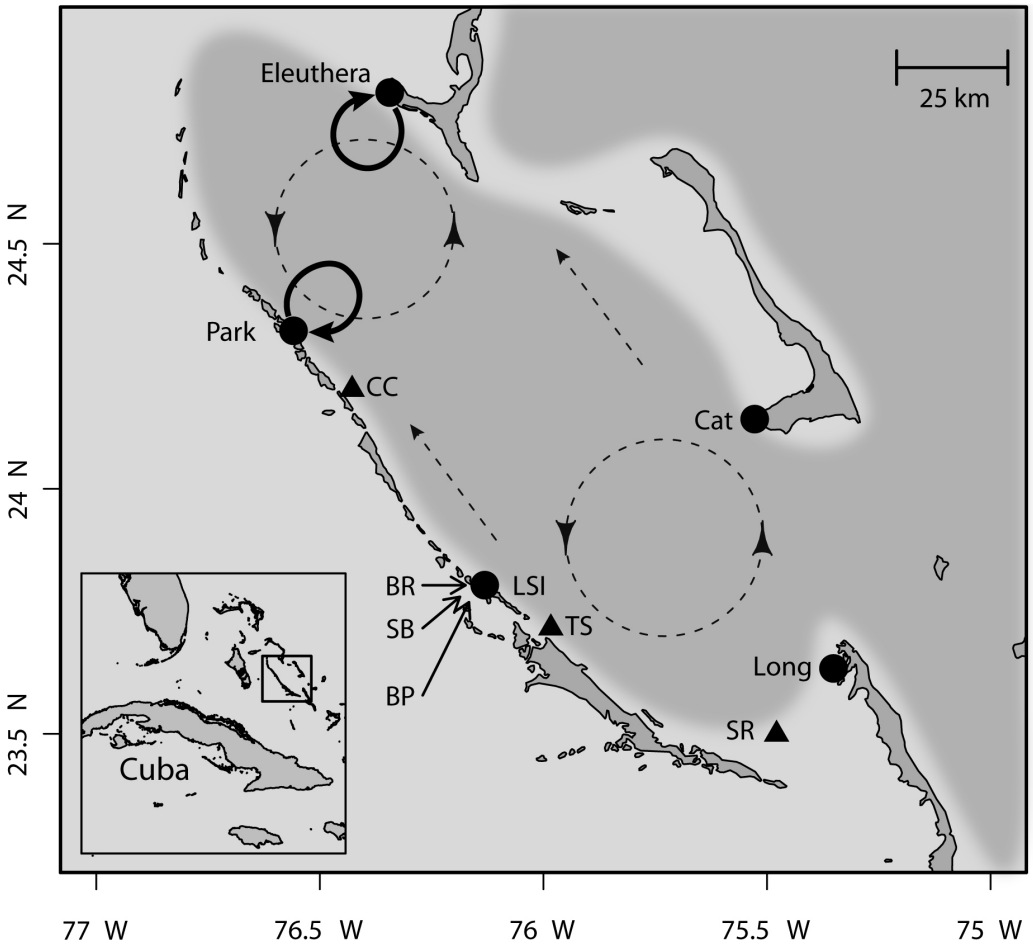


Figure 4.1. Sample sites and prevailing surface currents (dashed arrows) for bicolor damselfish collected in Exuma Sound, Bahamas. Parentage analysis identified two parent-offspring pairs (solid arrows), which directly documents self-recruitment at the two northern-most sites. Light seafloor indicates the shallow (mostly <3 m deep) Great Bahama Bank, whereas dark seafloor indicates the sound and nearby open ocean (mostly >1500 m deep). Triangles and straight arrows indicate 2004 sample sites, and filled circles indicate 2005 sample sites. Site abbreviations are as follows: Compass Cay (CC), Bock Rock (BR), String Bean Cay (SB), Big Point (BP), Lee Stocking Island (LSI), Three Sisters Reef (TS), and South Reef (SR).

Figure 4.2. Principal coordinates analysis (PCoA) on all pair-wise relatedness values of sampled bicolor damselfish, with results separated by sampling location for clarity. Adults are represented by filled circles and recruits are represented by open circles. Both axes combined explain 42% of the total variation. Note that (1) all recruits cluster in the same multivariate space as adults from the same sampling location, suggesting self-recruitment; and (2) sites such as Park and LSI, and Eleuthera and Long occupy different quadrants suggesting little larval connectivity. LSI 2004 comprised samples collected at String Bean Cay and Big Point (see Fig. 1). Results from multi-response permutation procedure for all sites indicate that the observed distribution of individuals in multivariate space is unlikely to occur by chance ($p < 0.0001$).

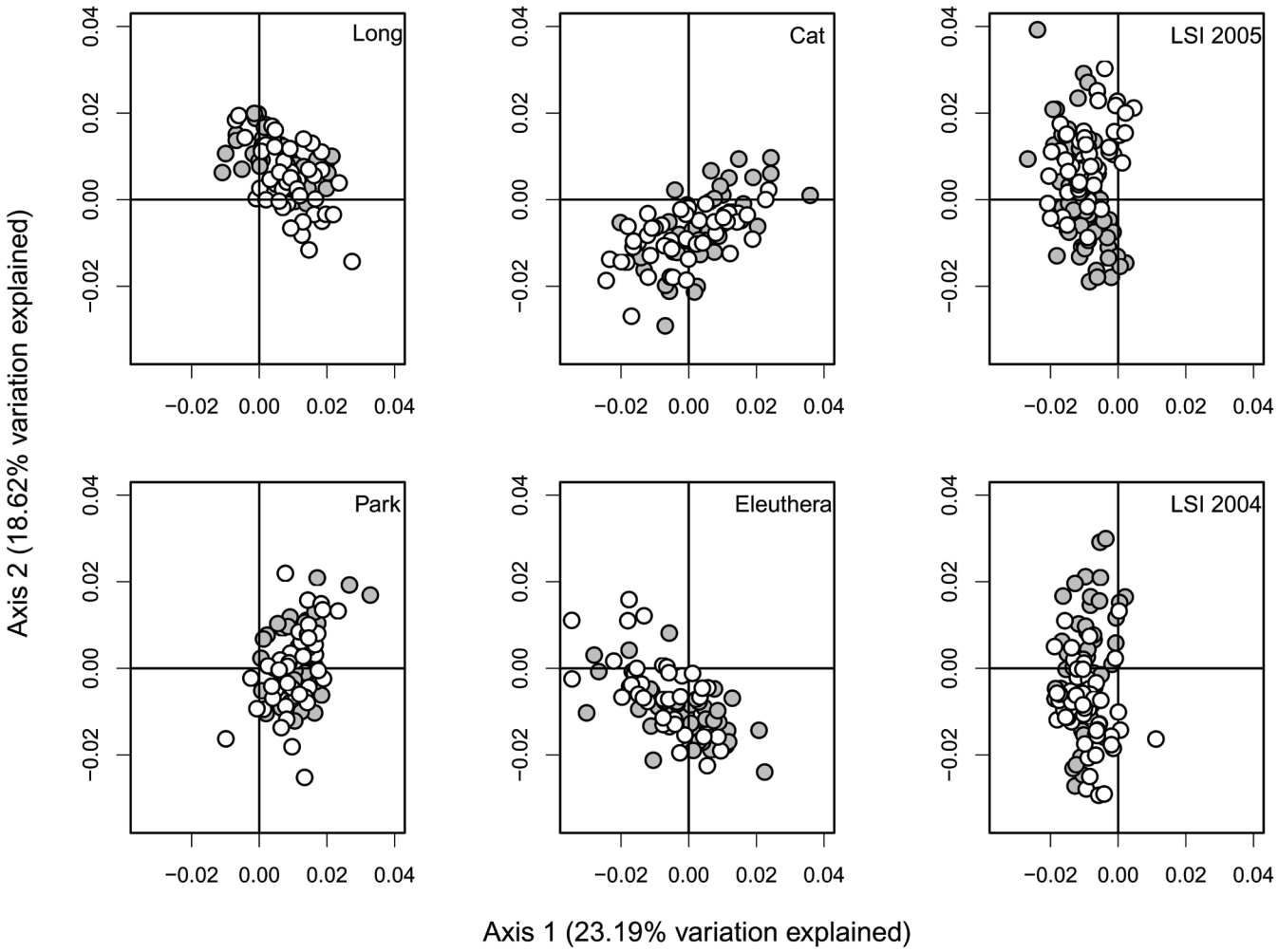

Long
Cat
LSI 2005
Park
Eleuthera
LSI 2004
Axis 2 (18.62% variation explained)
Axis 1 (23.19% variation explained)
−0.02
0.00
0.02
0.04

Figure 4.2

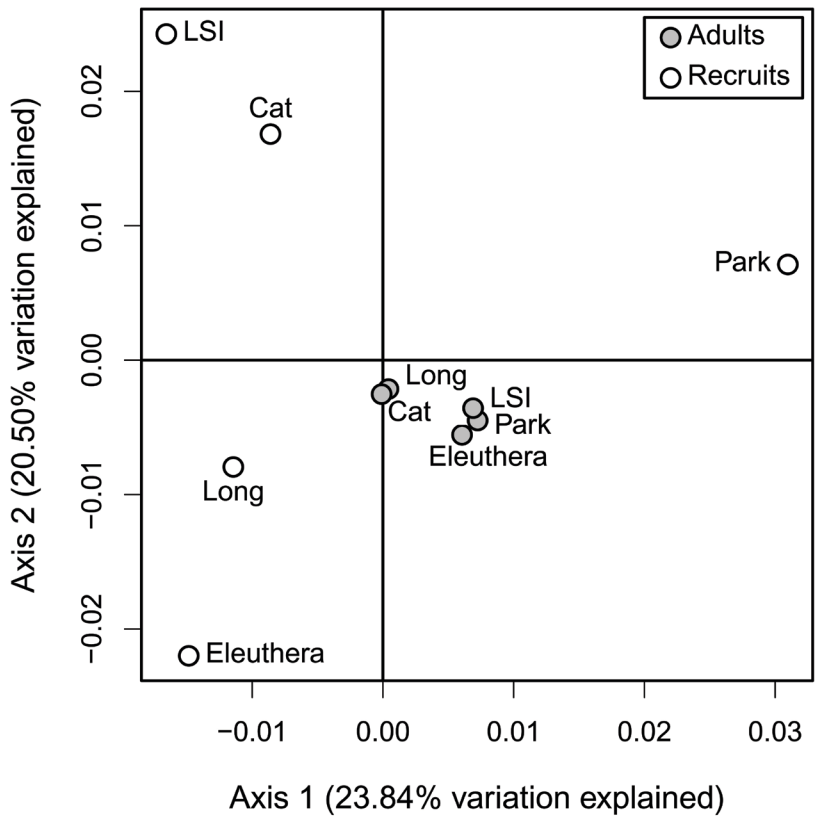


Figure 4.3 Principal coordinates analysis (PCoA) on all 2005 pair-wise F_{ST} values. Adults are represented by filled circles and recruits are represented by open circles. Both axes combined explain 44% of the total variation. Note that (1) all adults cluster together indicating greater genetic similarity to other adult samples than to the recruit samples; and (2) all recruit samples are both different from other recruit samples and from adult samples, which is indicative of separate sweepstakes events.

Chapter 5: General Conclusions

Understanding patterns of marine larval dispersal is important for advancing marine metapopulation theory, informing fisheries management, and improving conservation efforts. In an effort to identify these patterns I have: (1) provided novel parentage methods to directly determine patterns of marine dispersal, and (2) applied a broad array of novel and well-established population genetic methods to two different coral-reef fishes. Out of a combined total of 1844 yellow tang and damselfish samples, parentage methods were able to identify 6 parent-offspring pairs. While this number may seem trivial, it represents a large leap forward in elucidating patterns of marine larval dispersal. Because only a very small proportion of the total population of both study-species was sampled, the documented parent-offspring pairs necessarily represent high rates of population connectivity (yellow tang) and self-recruitment (bicolor damselfish). Quantifying these rates of dispersal, perhaps with novel mark-recapture methods in conjunction with good demographic estimates, will be invaluable for guiding management and conservation policy.

There are many informative similarities between the yellow tang and bicolor damselfish studies. Both fishes, like the vast majority of marine fishes, have pelagic larval stages and small home ranges as demersal adults. Both species also exhibit very little genetic-differentiation as measured by F_{ST} and other genetic distance metrics. This pattern of spatial homogeneity implies that both species experience moderate to high levels of gene-flow over evolutionary and perhaps ecological time scales. These

low levels of genetic differentiation make commonly implemented approaches, such as assignment tests (Manel et al. 2005), unreliable because such methods exploit differences in allele frequencies among populations to assign individuals to source populations. Even other assignment methods that do not directly rely upon allele frequencies *per se* (e.g., Pritchard et al. 2000) were ineffective at detecting more than one population in both species. Such results demonstrate that, in both study systems, gene flow occurs frequently enough to erode any differences in allele frequencies acquired through mutation and drift. This pattern makes it challenging to detect patterns of gene-flow at ecological time scales and requires that other, novel methods (e.g., parentage), be employed.

A related similarity between the two study species is an absence of genetic isolation-by-distance, regardless of which distance measure was used (e.g., Euclidean, along-shore). This observation is likely due to the relatively small spatial scale of sampling (compared to species ranges) as well as gene flow and mutation (homoplasy) eroding any differences in allele frequencies. Both species also had very high levels of heterozygosity, number of alleles per locus, and within-population diversity, which is common in many marine fishes (O'Connell and Wright 1997). Such observations may also be related to effective population sizes, which, for these two species, included infinity as the upper confidence limit, and are likely quite large.

Despite the similarities between these two species of coral-reef fish, I also found some striking differences. In yellow tang, parentage analyses revealed that larvae can travel from 15 to 184 kilometers. Furthermore, oceanographic analyses

revealed that the larvae were, at least initially, passively transported by along-shore currents. However, for bicolor damselfish, parentage analysis revealed that populations likely experience high rates of self-recruitment. These results were further bolstered by principal coordinates analysis of pair-wise relatedness values, showing extensive overlap between adults and recruits from the same site. Thus, it is likely that yellow tang have high rates of ecologically-relevant larval connectivity among sites, while bicolor damselfish populations may experience higher levels of larval retention. On one hand, yellow tang are broad-cast spawners and their larvae remain planktonic for approximately 54 days. Bicolor damselfish, on the other hand, guard demersal eggs and their larvae are planktonic for only 28 days. Correlations between spawning behavior, pelagic larval duration, and dispersal distance have been of great interest for many years (Doherty et al. 1995), and the results presented here may contribute to future meta-analyses. Furthermore, differences in larval development and behavior between the two species may influence retention and dispersal (Cowen et al. 2000, Leis et al. 2009). Such differences in life histories may explain some of the documented differences in larval dispersal between these two species.

We also found important genetic differences between the two species that likely reflect the different rates of larval connectivity. In yellow tang, pair-wise genetic differences among samples of recruits were smaller than expected by chance. In bicolor damselfish, the opposite pattern was observed, with large differences in allele frequencies among samples of recruits observed, particularly when compared

with adults. Furthermore, recruit samples of bicolor damselfish had much greater levels of average within-sample relatedness than for all pooled samples, which is indicative of independent sweepstakes effects. Such differences in samples of recruits may hold the key to determining the ecologically-relevant spatial scale of larval connectivity among populations of marine fish. In general, the differences between the two species highlight the need to determine community-wide patterns of larval dispersal before designing marine reserves.

Future studies should focus on estimating connectivity matrices among sampled populations (Botsford et al. 2009). The rapid advancement of genome-wide technology means that we will soon see an abundance of molecular data. As such, new methods that focus on maximizing the signal to noise ratio will be required. Parentage will likely remain a valuable tool across a broad array of natural populations. Incorporating parentage data into other methods, perhaps by using it as priors in Bayesian assignment methods, could be a powerful approach. Additionally, other relatedness approaches, such as the multivariate approaches used with bicolor damselfish, could be quite useful in species with high rates of gene flow and warrant further examination.

In conclusion, genetic methods have great potential to detect fine-scale, ecologically relevant patterns of larval dispersal in marine fishes. Applying such fine-scale methods to two species of coral-reef fishes, we directly documented both self-recruitment and population connectivity. Importantly, we demonstrated for the first time that MPAs can seed fished sites with larvae that survive to become juveniles.

Such documentation of the long-hypothesized “seeding effect” demonstrates that MPAs can be useful as both management and conservation tools. Information about patterns of dispersal acquired from these and similar studies must be fully incorporated into a broader theoretical and soci-political framework to provide a scientifically informed foundation for ecosystem-based management of the oceans.

Bibliography

Almany, G. R., M. L. Berumen, S. R. Thorrold, S. Planes, and G. P. Jones. 2007. Local replenishment of coral reef fish populations in a marine reserve. Science 316:742-744.

Anderson, A. D., and B. S. Weir. 2007. A maximum-likelihood method for the estimation of pairwise relatedness in structured populations. Genetics 176:421-440.

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567-2582.

Araki, H., B. Cooper, and M. S. Blouin. 2007. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science 318:100-103.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2002. Genetix 4.03, Logiciel Sous Window TM Pour la Genetique Des Populations. Laboratoire Genome, Populations, Interactions, CNR SUMR 5000, Universite de Montpellier II, Montpellier, France.

Berkeley, S. A., M. A. Hixon, R. J. Larson, and M. S. Love. 2004. Fisheries Sustainability via protection of age structure and spatial distribution of fish populations. Fisheries 29:23-32.

Bleck, R. 2002. An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. Ocean Modeling 4:55-88.

Blouin, M. S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology & Evolution 18:503-511.

Botsford, L. W., F. Micheli, and A. Hastings. 2003. Principles for the design of marine reserves. Ecological Applications 13:S25-S31.

Botsford, L. W., J. W. White, M. A. Coffroth, C. B. Paris, S. Planes, T. L. Shearer, S. R. Thorrold, and G. P. Jones. 2009. Connectivity and resilience of coral reef metapopulations in marine protected areas: matching empirical efforts to predictive needs. Coral Reefs 28:327-337.

Carlin, B. P., and T. A. Louis. 2000. Bayes and empirical Bayes methods for data analysis. Chapman and Hall, Boca Raton.

Chakraborty, R. 1991. Inclusion of Data on Relatives for Estimation of Allele Frequencies. American Journal of Human Genetics 49:242-243.

Chakraborty, R., T. R. Meagher, and P. E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527-536.

Christie, M. R. 2009. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. Molecular Ecology Resources *in press*.

Christie, M. R., and J. A. Eble. 2009. Isolation and characterization of 23 microsatellite loci in the yellow tang, *Zebrasoma flavescens* (Pisces: Acanthuridae). Molecular Ecology Resources 9:544-546.

Claisse, J. T. 2009. The life history and movements of yellow tang, *Zebrasoma flavescens*, with implications for fisheries management using protected areas. *PhD Dissertation*. University of Hawaii.

Claisse, J. T., S. A. Mctee, and J. D. Parrish. 2009. Effects of age, size, and density on natural survival for an important coral reef fishery species, yellow tang, *Zebrasoma flavescens*. Coral Reefs 28:95-105.

Colin, P. L. 1995. Surface currents in Exuma Sound, Bahamas, and adjacent areas with reference to potential larval transport. Bulletin of Marine Science 56:48-57.

Corander, J., P. Waldmann, P. Marttinen, and M. J. Sillanpaa. 2004. BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20:2363-2369.

Cowen, R. K., K. M. M. Lwiza, S. Sponaugle, C. B. Paris, and D. B. Olson. 2000. Connectivity of marine populations: Open or closed? Science 287:857-859.

Derycke, S., T. Remerie, T. Backeljau, A. Vierstraete, J. Vanfleteren, M. Vincx, and T. Moens. 2008. Phylogeography of the *Rhabditis (Pellioditis) marina* species complex: evidence for long-distance dispersal, and for range expansions and restricted gene flow in the northeast Atlantic. Molecular Ecology 17:3306-3322.

DeWoody, J. A. 2005. Molecular approaches to the study of parentage, relatedness, and fitness: Practical applications for wild animals. Journal of Wildlife Management 69:1400-1418.

Dodds, K. G., M. L. Tate, J. C. McEwan, and A. M. Crawford. 1996. Exclusion probabilities for pedigree testing farm animals. Theoretical and Applied Genetics 92:966-975.

Doherty, P. J., S. Planes, and P. Mather. 1995. Gene flow and larval duration in seven species of fish from the Great Barrier Reef. Ecology 76:2373-2391.

Ehrlich, P. 1975. The population biology of coral reef fishes. Annual Review of Ecology and Systematics 6:211-247.

Francis, R. C., M. A. Hixon, M. E. Clarke, S. A. Murawski, and S. Ralston. 2007. Ten commandments for ecosystem-based fisheries management. Fisheries 32:217-233.

Garber, R. A., and J. W. Morris. 1983. General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and in no-parent cases of disputed parentage. Pages 277-280 *in* R.H.Walker, editor. Inclusion Probabilities in Parentage Testing. American Association of Blood Banks, Arlington, VA.

Gell, F. R., and C. M. Roberts. 2003. Benefits beyond boundaries: the fishery effects of marine reserves. Trends in Ecology & Evolution 18:448-455.

Gerber, S., P. Chabrier, and A. Kremer. 2003. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Molecular Ecology Notes 3:479-481.

Gerlach, G., J. Atema, M. J. Kingsford, K. P. Black, and V. Miller-Sims. 2007. Smelling home can prevent dispersal of reef fish larvae. Proceedings of the National Academy of Sciences of the United States of America 104:858-863.

Goodnight, K. F., and D. C. Queller. 1999. Computer software for performing likelihood tests of pedigree relationship using genetic markers. Molecular Ecology 8:1231-1234.

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

Gutierrez-Rodriguez, C., and H. R. Lasker. 2004. Microsatellite variation reveals high levels of genetic variability and population structure in the gorgonian coral *Pseudopterogorgia elisabethae* across the Bahamas. Molecular Ecology 13:2211-2221.

Hedgecock, D. 1994a. Does variance in reproductive success limit effective population size of marine organisms? In: *Genetics and evolution of aquatic organisms* (ed. Beaumont, A.R.). Chapman and Hall, London, pp. 122-135.

Hedgecock, D. 1994b. Does variance in reproductive success limit effective population size of marine organisms? *in* Genetics and evolution of aquatic organismsBeaumont, A.R. Chapman and Hall, London.

Hedgecock, D., P. H. Barber, and S. Edmands. 2007. Genetic approaches to measuring connectivity. Oceanography 20:70-79.

Hedgecock, D., S. Launey, A. I. Pudovkin, Y. Naciri, S. Lapegue, and F. Bonhomme. 2007. Small effective number of parents (Nb) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. Marine Biology 150:1173-1182.

Hixon, M. A., S. W. Pacala, and S. A. Sandin. 2002. Population regulation: historical context and contemporary challenges of open vs. closed systems. Ecology 83:1490-1508.

Hoffman, J. I., and W. Amos. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology 14:599-612.

Hughes, T. P., D. R. Bellwood, C. Folke, R. S. Steneck, and J. Wilson. 2005. New paradigms for supporting the resilience of marine ecosystems. Trends in Ecology & Evolution 20:380-386.

Jamieson, A., and S. S. Taylor. 1997. Comparisons of three probability formulae for parentage exclusion. Animal Genetics 28:397-400.

Jombart, T., D. Pontier, and A. B. Dufour. 2009. Genetic markers in the playground of multivariate analysis. Heredity 102:330-341.

Jones, A. G., and W. R. Ardren. 2003. Methods of parentage analysis in natural populations. Molecular Ecology 12:2511-2523.

Jones, G. P., M. J. Milicich, M. J. Emslie, and C. Lunow. 1999. Self-recruitment in a coral reef fish population. Nature 402:802-804.

Jones, G. P., S. Planes, and S. R. Thorrold. 2005. Coral reef fish larvae settle close to home. Current Biology 15:1314-1318.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099-1106.

Kauserud, H., O. Stensrud, C. Decock, K. Shalchian-Tabrizi, and T. Schumacher. 2006. Multiple gene genealogies and AFLPs suggest cryptic speciation and long-distance dispersal in the basidiomycete *Serpula himantioides* (Boletales). Molecular Ecology 15:421-431.

Knapp, R. A., P. C. Sikkel, and V. T. Vredenburg. 1995. Age of clutches in nests and the within-west spawning-site preferences of three damselfish species (Pomacentridae). Copeia 1995:78-88.

Konovalov, D. A., C. Manning, and M. T. Henshaw. 2004. KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. Molecular Ecology Notes 4:779-782.

Kritzer, J. P., and P. F. Sale. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries 5:131-140.

Leis, J. M. 2006. Are larvae of demersal fishes plankton or nekton? Advances in Marine Biology, Vol 51 51:57-141.

Leis, J., A. Hay, and G. Howarth. 2009. Ontogeny of in situ behaviours relevant to dispersal and population connectivity in larvae of coral-reef fishes. Marine Ecology Progress Series 379:163-179.

Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152:1753-1766.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18:189-197.

Manel, S., O. E. Gaggiotti, and R. S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution 20:136-142.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639-655.

McCune, B., and J. Grace. 2002. *Analysis of ecological communities.* 2nd edn. MjM, Gleneden Beach, OR.

McLean, J. E., T. R. Seamons, M. B. Dauer, P. Bentzen, and T. P. Quinn. 2008. Variation in reproductive success and effective number of breeders in a hatchery population of steelhead trout (*Oncorhynchus mykiss*): examination by microsatellite-based parentage analysis. Conservation Genetics 9:295-304.

McRae, B. H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proceedings of the National Academy of Sciences of the United States of America 104:19885-19890.

Meagher, T. R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of the most-likely male parents. American Naturalist 128:199-215.

Mielke, P. J., and K. Berry. 2001. *Permutation methods: A distance function approach*. Springer-Verlag, New York.

Milligan, B. G. 2003. Maximum-likelihood estimation of relatedness. Genetics 163:1153-1167.

Mills, L. S., and F. W. Allendorf. 1996. The one-migrant-per-generation rule in conservation and management. Conservation Biology 10:1509-1518.

Mora, C., and P. F. Sale. 2002. Are populations of coral reef fish open or closed? Trends in Ecology and Evolution 17:422-428.

Morin, P., R. Leduc, and et al. 2009. Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. *Molecular Ecology Resources* 9:498-594.

Mumby, P. J., and R. S. Steneck. 2008. Coral reef management and conservation in light of rapidly evolving ecological paradigms. Trends in Ecology & Evolution 23:555-563.

Nathan, R. 2006. Long-distance dispersal of plants. Science 313:786-788.

Nutt, K. J. 2008. A comparison of techniques for assessing dispersal behaviour in gundis: revealing dispersal patterns in the absence of observed dispersal behaviour. Molecular Ecology 17:3541-3556.

O'Connell, M., and J. Wright. 1997. Microsatellite DNA in fishes. Reviews in Fish Biology and Fisheries 7:331-363.

Oddou-Muratorio, S., M. L. Houot, B. Demesure-Musch, and F. Austerlitz. 2003. Pollen flow in the wildservice tree, *Sorbus torminalis (L.)* Crantz. I. Evaluating the paternity analysis procedure in continuous populations. Molecular Ecology 12:3427-3439.

Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead, E. Vyse, R. Ward, and C. Strobeck. 1998. Variation in genetic diversity across the range of North American brown bears. Conservation Biology 12:418-429.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications 13:S146-S158.

Palumbi, S. R., G. Grabowsky, T. Duda, L. Geyer, and N. Tachino. 1997. Speciation and population genetic structure in tropical Pacific Sea urchins. Evolution 51:1506-1517.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95:536-539.

Planes, S., G. P. Jones, and S. R. Thorrold. 2009. Larval dispersal connects fish populations in a network of marine protected areas. Proceedings of the National Academy of the Sciences 106:5693-5697.

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

Purcell, J. F. H., R. K. Cowen, C. R. Hughes, and D. A. Williams. 2009. Population structure in a common Caribbean coral-reef fish: implications for larval dispersal and early life-history traits. Journal of Fish Biology 74:403-417.

Queller, D. C., and K. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution 43:258-275.

Queller, D. C., F. Zacchi, R. Cervo, S. Turillazzi, M. T. Henshaw, L. A. Santorelli, and J. E. Strassmann. 2000. Unrelated helpers in a social insect. Nature 405:784-787.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria: http://www.R-project.org.

Raymond, M., and F. Rousset. 1995. GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Roberts, C. M. 1997. Connectivity and Management of Caribbean Coral Reefs. Science 278:1454-1457.

Robledo-Arnuncio, J. J., and C. Garcia. 2007. Estimation of the seed dispersal kernel from exact identification of source plants. Molecular Ecology 16:5098-5109.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228.

Russ, G. R. 2002. Yet another review of marine reserves as reef fishery management tools. Pages 421-443 *in* P. F. Sale, editor. Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem. Academic Press, San Diego.

Saenz-Agudelo, P., G. P. Jones, S. R. Thorrold, and S. Planes. 2009. Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. Molecular Ecology 18:1765-1776.

Sambrook, J., and D. W. Russell. (2001). *Molecular Cloning: A Laboratory Manual.* 3rd edn. Cold Springs Harbor Laboratory Press, Cold Springs Harbor, N.Y.

Selkoe, K. A., S. D. Gaines, J. E. Caselle, and R. R. Warner. 2006. Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* 87:3082-3094.

Shulman, M. J. 1998. What can population genetics tell us about dispersal and biogeographic history of coral-reef fishes? Australian Journal of Ecology. 23:216-225.

Shulman, M. J., and E. Bermingham. 1995. Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. Evolution 49:897-910.

Slate, J., T. Marshall, and J. Pemberton. 2000. A retrospective assessment of the accuracy of the paternity inference program CERVUS. Molecular Ecology 9:801-808.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry 3rd ed. W.H. Freeman and Company, New York.

Stoner, A., and M. Davis. 1997. Abundance and distribution of queen conch veligers (*Strombus gigas* Linne) in the central Bahamas. 1. Horizontal patterns in relation to reproductive and nursery grounds. Journal of Shellfish Research 16:7-18.

Swearer, S. E., J. E. Caselle, D. W. Lea, and R. R. Warner. 1999. Larval retention and recruitment in an island population of coral-reef fish. Nature 402:799-802.

Taylor, M. S., and M. E. Hellberg. 2003. Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. Science 299:107-109.

Telfer, S., S. B. Piertney, J. F. Dallas, W. A. Stewart, F. Marshall, J. L. Gow, and X. Lambin. 2003. Parentage assignment detects frequent and large-scale dispersal in water voles. Molecular Ecology 12:1939-1949.

Tentelier, C., T. Guillemaud, S. Ferry, and X. Fauvergue. 2008. Microsatellite-based parentage analysis reveals non-ideal free distribution in a parasitoid population. Molecular Ecology 17:2300-2309.

Thompson, E. A. 1975. The estimation of pair-wise relationship. Annals of Human Genetics 39:173-188.

Thompson, E. A. 1976. Inference of genealogical structure. Social Science Information 15:477-526.

Thompson, E. A., and T. R. Meagher. 1987. Parental and sib likelihoods in genealogy reconstruction. Biometrics 43:585-600.

Thompson, E. A., and T. R. Meagher. 1998. Genetic linkage in the estimation of pairwise relationship. Theoretical and Applied Genetics 97:857-864.

Thorrold, S. R., D. C. Zacherl, and L. A. Levin. 2007. Population connectivity and larval dispersal using geochemical signatures in calcified structures. Oceanography 20:80-89.

Thorrold, S. R., G. P. Jones, M. E. Hellberg, R. S. Burton, S. E. Swearer, J. E. Neigel, S. G. Morgan, and R. R. Warner. 2002. Quantifying larval retention and connectivity in marine populations with artificial and natural markers. Bulletin of Marine Science 70:291-308.

Thorrold, S. R., G. P. Jones, S. Planes, and J. A. Hare. 2006. Transgenerational marking of embryonic otoliths in marine fishes using barium stable isotopes. Canadian Journal of Fisheries and Aquatic Sciences 63:1193-1197.

Tissot, B. N., and L. E. Hallacher. 2003. Effects of Aquarium Collectors on Coral Reef Fishes in Kona, Hawaii. Conservation Biology 17:1759.

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

Vandeputte, M., S. Mauger, and M. Dupont-Nivet. 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. Molecular Ecology Notes 6:265-267.

Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Molecular Ecology 10:249-256.

Walsh, W. J. 1987. Patterns of recruitment and spawning in Hawaiian reef fishes. Environmental Biology of Fishes. 18:257-276.

Wang, J. L. 2004. Application of the one-migrant-per-generation rule to conservation and management. Conservation Biology 18:332-343.

Waples, R. S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89:438-450.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics 7:167-184.

Waser, P. M., J. D. Busch, C. R. McCormick, and J. A. DeWoody. 2006. Parentage analysis detects cryptic precapture dispersal in a philopatric rodent. Molecular Ecology 15:1929-1937.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the analysis of population structure. Evolution 38:1358-1370.

White, J. W., J. D. Standish, S. R. Thorrold, and R. R. Warner. 2008. Markov Chain Monte Carlo methods for assigning larvae to natal sites using natural geochemical tags. Ecological Applications 18:1901-1913.

Williams, D. A., J. Purcell, C. R. Hughes, and R. K. Cowen. 2003. Polymorphic microsatellite loci for population studies of the bicolor damselfish, *Stegastes partitus* (Pomacentridae). Molecular Ecology Notes, 3:547-549.

Williams, I. D., W. J. Walsh, J. T. Claisse, B. N. Tissot, and K. A. Stamoulis. 2009. Impacts of a Hawaiian marine protected area network on the abundance and fishery sustainability of the yellow tang, *Zebrasoma flavescens*. Biological Conservation 142:1066-1073.

Wilson, D. T., and M. I. McCormick. 1999. Microstructure of settlement-marks in the otoliths of tropical reef fishes. Marine Biology 134:29-41.

APPENDICES

APPENDIX A. Supplementary Materials, Chapter 4

Supporting Information

Supplementary Text:

Sampling. Adults were collected from July through August 2006 (see Table A1 for sample sizes). Recruits and juveniles were collected from June through August 2006, the annual settlement season, with monthly collections at sites located on the west (Kohala-Kona) coast of the Island of Hawai'i.

DNA extraction, amplification, and scoring. All fin-clip samples were stored at -20°C in 1.5mL microcentrifuge tubes filled with 95% non-denatured ethanol. DNA was extracted using a standard Chelex® (Biorad Laboratories) and proteinase K protocol. PCR details are available elsewhere (Christie and Eble 2009). Alleles were binned both by eye and with the use of FLEXIBIN (Amos et al. 2007). Bins were never greater than one half of the repeat motif on either side of the allele.

Genetic analyses. Exploratory analyses and summary statistics were calculated with GENALEX (Peakall and Smouse 2006). Loci were tested for deviations from Hardy-Weinberg equilibrium (HWE) in GENEPOP (Raymond and Rousset 1995) with 10,000 batches and 5,000 iterations per batch. All loci conformed to HWE after a Bonferroni correction, though there were a few deviations in some populations (Table A1). These populations never had more than one locus that was out of HWE and the loci differed among populations. There was no evidence for linkage disequilibrium among all pairs of loci after tests were conducted in both GENETIX (Belkhir et al.

2002) (5,000 permutations) and GENEPOP (5,000 batches, 10,000 iterations per batch). Global and pair-wise F_{ST} values were calculated with FSTAT (Goudet 2001). For the complete data set, the amount of missing data equalled 0.6 % (197 out of 33,150 scored alleles). For parentage analysis, missing data was coded as the most common allele, which is a conservative approach. Mantel tests for isolation by distance analysis were calculated with ISOLDE as implemented in GENEPOP (Raymond and Rousset 199).

Offspring aging. To estimate the dates that the documented offspring were spawned, we calculated the ages of the offspring on their collection date using a species-specific linear growth equation $(r^2 = 0.79)$. This equation was derived by comparing the total length of 56 yellow tang recruits collected from Hawai'i (size range 30 to 47mm) to their nearest age (in days) as determined by otolith daily growth rings (David J. Shafer, University of Hawai'i, *in preparation*). The total lengths of the four identified offspring ranged from 34 to 44 mm, well within the range of available data. We subtracted the age from the collection date to determine the approximate date of spawning. To calculate the approximate settlement date we added the pelagic larval duration of yellow tang, 54 days, to the spawning date.

Oceanography. The regional HYCOM model configuration was initialized with the ocean state of a global HYCOM simulation (HYCOM consortium 2009) on January 1 2006 and integrated to August 31 2006, the year of this study. Ocean currents,

temperature and salinity from the global HYCOM simulation were applied at the lateral boundaries of the region. At the surface, the ocean was forced with QuikScat winds (NASA 2009), among other atmospheric forcing fields. Satellite altimeter data (CLS 2009) indicated a cyclonic eddy in the lee of the Island of Hawai'i, which persisted throughout the pelagic larval duration of the identified yellow tang offspring (data available upon request). This eddy was captured by the HYCOM simulation but it propagated away from shore about a month earlier than that indicated by satellite altimeter data. Had the HYCOM model better resolved the duration of this eddy, it is possible that more drifters would have been retained near-shore at the time of settlement**.**

Statistical analyses. Parentage and statistical analyses were calculated using the *R* statistical software environment (R Development Core Team 2009). A randomisation test determined whether the F_{ST} values among recruit samples were smaller than expected by chance. All pair-wise F_{ST} values, including recruit-versus-recruit comparisons, were randomly sampled without replacement 28 times (the number of observed recruit-vs.-recruit comparisons) and the mean value recorded. This process was repeated 10,000 times to create a distribution of mean F_{ST} values. The observed mean F_{ST} value for recruit-versus-recruit comparisons was compared to this distribution to calculate the probability of occurring by chance. Pohoiki was not included in this analysis because we were able to sample only 18 recruits at that site (Table A1).

To determine the probability of observing two parents at each of two sites, we created simulated individuals for each population, corresponding to the sample sizes of adult samples (see Table S1). Four individuals were then randomly sampled with replacement (to simulate sampling the 4 observed parents) and the location of all simulated parents categorically noted (e.g., 4 parents at 4 different sites). This process was repeated 10,000 times to create a distribution of sampling probabilities. To calculate the probability of observing 2 parents at 2 sites, we added the number of events as or more rare (e.g., all 4 parents sampled at 1 site) than observing 2 parents at each of 2 sites, and divided by 10,000.

Supporting References

Amos, W., J. I. Hoffman, A. Frodsham, L. Zhang, S. Best, and A. V. S. Hill. 2007. Automated binning of microsatellite alleles: problems and solutions. Molecular Ecology Notes 7:10-14.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2002. Genetix 4.03, Logiciel Sous Window TM Pour la Genetique Des Populations. Laboratoire Genome, Populations, Interactions, CNR SUMR 5000, Universite de Montpellier II, Montpellier, France.

CLS Space Oceanography Division. 2009. ftp://ftp.cls.fr/pub/oceano/enact/msla/readme_MSLA_ENACT.htm.

Christie, M. R., and J. A. Eble. 2009. Isolation and characterization of 23 microsatellite loci in the yellow tang, *Zebrasoma flavescens* (Pisces: Acanthuridae). Molecular Ecology Resources 9:544-546.

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

HYCOM consortium data products and services. 2009. http://hycom.rsmas.miami.edu/dataserver.

NASA Ocean Vector Winds Science Team and Remote Sensing Systems. 2009. www.remess.com.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria:http://www.R-project.org.

Raymond, M., and F. Rousset. 1995. GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Table A.1. Sample sizes and population genetics summary.

Sample site*	Sample size	Na†	Pa‡	H_O§	H_E§	F°
Anaeho'omalu adults	49	11.200	6	**0.718**	0.796	0.103
Anaeho'omalu juveniles	25	9.667	2	0.751	0.779	0.045
Anaeho'omalu recruits	57	11.133	3	0.753	0.790	0.043
Hilo adults	49	12.333	1	0.778	0.791	0.018
Hilo recruits	42	11.800	1	0.779	0.798	0.015
Honokohau adults	73	10.000	2	0.763	0.776	0.012
Honokohau recruits	109	12.800	5	0.767	0.799	0.038
Ho'okena adults	65	11.800	1	**0.737**	0.793	0.072
Ho'okena recruits	68	11.933	4	0.787	0.806	0.019
Miloli'i adults	60	11.800	0	0.781	0.796	0.029
Miloli'i recruits	67	12.000	1	**0.754**	0.809	0.075
Pohoiki adults	51	11.400	1	0.729	0.792	0.074
Pohoiki recruits	18	9.467	1	0.811	0.796	-0.028
Puako adults	66	11.933	1	0.791	0.794	-0.001
Puako recruits	48	11.400	0	0.758	0.795	0.048
Punalu'u adults	43	11.333	2	0.783	0.803	0.027
Punalu'u juveniles**	49	11.533	1	0.789	0.798	0.009
Wawaloli adults	50	11.467	1	**0.743**	0.790	0.048
Wawaloli recruits	83	12.467	2	**0.747**	0.792	0.051
Laupahoehoe**	1	2.000	0	n/a	n/a	n/a

*Age-size categories: recruits < 49 mm TL, 50 mm TL < juveniles < 149 mm TL, adults > 150 mm TL

***Zebrasoma flavescens* were present at extremely low densities at Laupahoehoe and no recruits were found at Punalu'u.

† Average number of alleles per locus.

‡ Number of private alleles.

§ Observed (Ho) and expected (He) heterozygosities averaged across all loci. Populations where one locus deviated from HWE after a Bonferroni correction are indicated in bold.

°Fixation index averaged across loci and calculated with unbiased estimates of heterozygosity.

Supporting Figures

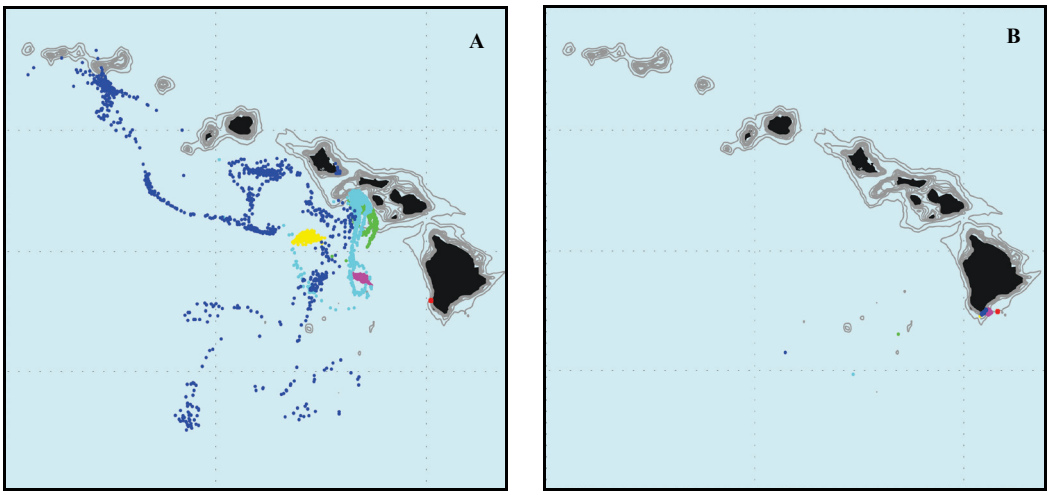


Figure A.1. Simulated patterns of passive drifter dispersal off the Island of Hawai'i. A total of 961 simulated drifters were sampled at 0 (red), 5 (magenta), 10 (yellow), 20 (green), 30 (light blue), and 54 (dark blue) days after release. (**A**) Passive dispersal of drifters released from Miloli'i at the estimated date of spawning of a documented offspring (May 13 2006). (**B**) Passive dispersal of drifters released from Punalu'u at the estimated date of spawning of a documented offspring (April 18 2006). Note that these virtual drifters ran aground shortly after they were spawned, unlike those simulated for other documented offspring spawned at this site (Fig. 2*C* in main text).

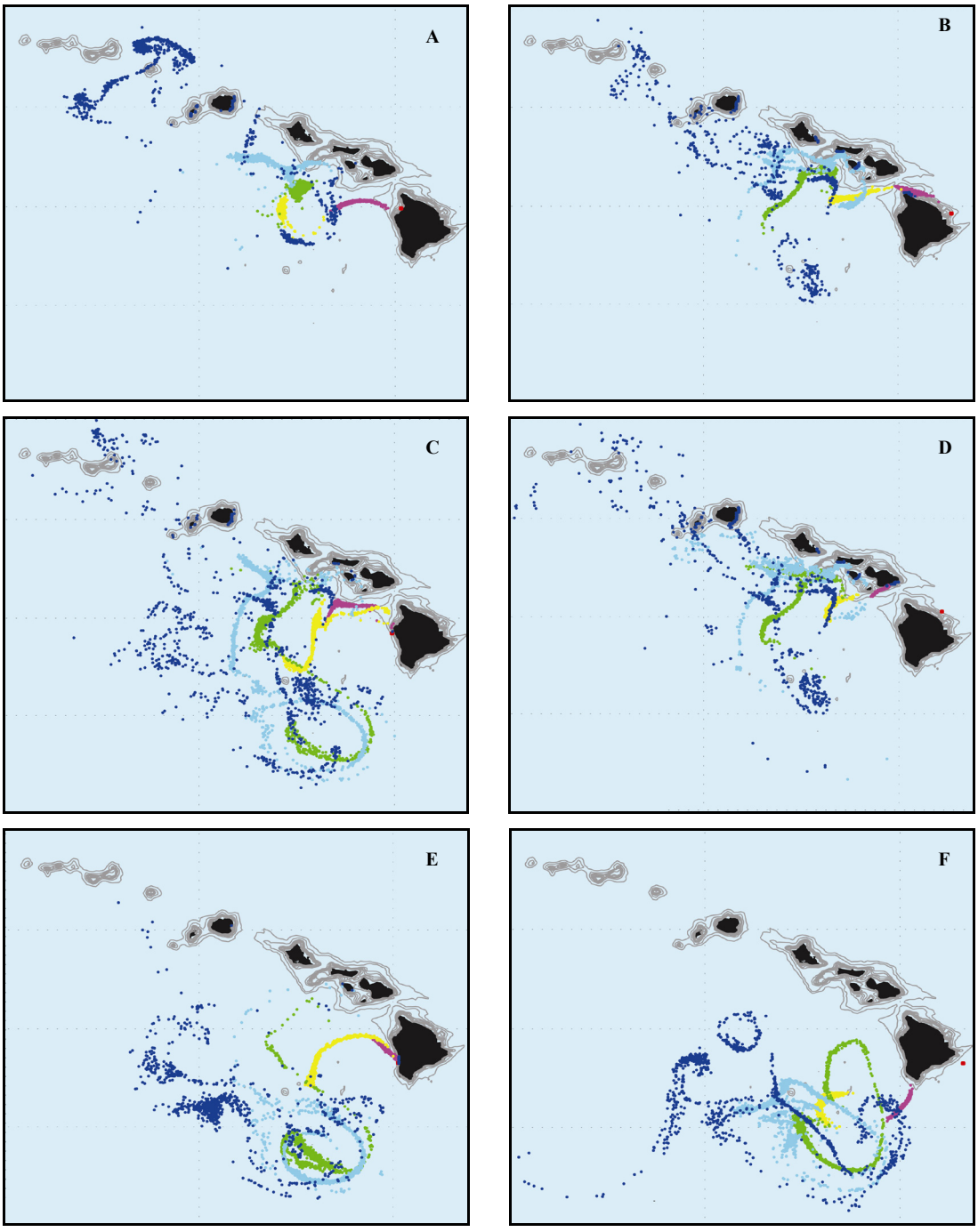


Figure A.2. Simulated patterns of passive drifter dispersal off Hawai'i. A total of 961 simulated drifters were sampled at 0 (red), 5 (magenta), 10 (yellow), 20 (green), 30 (light blue), and 54 (dark blue) days after release. Example drifter release locations and dates include (**A**) Anaeho'omalu (April 18 2006), (**B**) Hilo (April 26 2006) (**C**) Honokohau (April 26 2006), (**D**) Laupahoehoe (April 24 2006), (**E**) Ho'okena (May 13 2006) and (**F**) Pohoiki (April 18 2006).

APPENDIX B. Supplementary Materials, Chapter 5

Table B.1. Details of fish sample collections including the number of adults and recruits sampled at each site. Site code refers to abbreviations used in Fig. 1 and throughout the text.

Sample Locality	Site code	Latitude(N)	Longitude(W)	# Adults	# Recruits	Collection Date
Lee Stocking Island	LSI	23°48'13.49"	76° 7'54.17"	42	61	June-July 2005
Land and Sea Park	Park	24°19'19.24"	76°33'32.58"	44	45	July 2005
Eleuthera	Eleuthera	24°48'29.09"	76°20'36.49"	49	37	July 2005
Cat Island	Cat	24° 8'30.65"	75°31'36.84"	46	44	July 2005
Long Island	Long	23°34'24.66"	75°20'10.26"	41	47	July 2005
Compass Cay	CC	24°12'10.00"	76°25'40.00"	39	7	July 2004
Bock Rock	BR	23°48'11.41"	76° 9'14.72"	28	4	June 2004
String Bean Cay	SP	23°47'43.97"	76° 8'11.39"	56	31	June-July 2004
Big Point	BP	23°46'42.77"	76° 7'35.13"	28	49	June-August 2004
Three Sisters Reef	TS	23°42'56.99"	75°58'59.98"	25	4	July 2004
South Reef	SR	23°29'55.99"	75°28'42.99"	39	5	July 2004
Total				437	334	

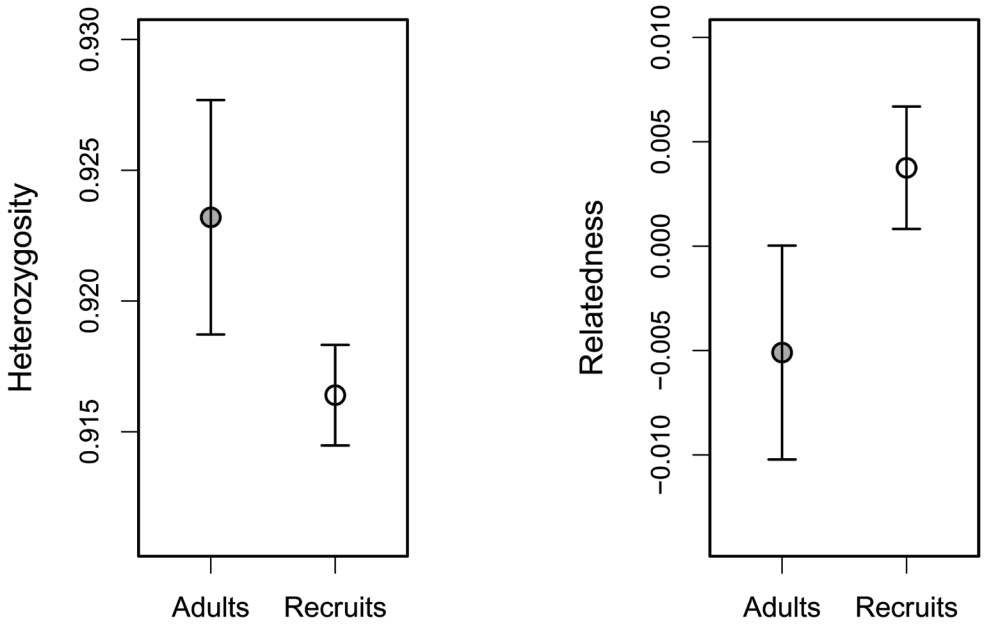


Figure B.1. Mean levels of observed heterozygosity and relatedness among 2005 sample sites. Bars represent 95% confidence intervals and do not overlap. The pattern of reduced genetic diversity and increased relatedness within recruit samples is a distinctive signature of a sweepstakes effect.

Appendix C. Multi-response permutation procedure (MRPP) results for each pair-wise sample locality comparison. MRPP was calculated on results from PCoA analysis of pair-wise relatedness values (Figure 3) with 10,000 permutations. Observed and expected δ equal the observed and expected weighted mean within group distance. The probability of departure from the null hypothesis of no difference, p, and corresponding effect size, A, are also reported.

Sample locality		Observed δ	Expected δ	p	A
Long adults	Long recruits	0.013	0.014	0.19629	0.046
Long adults	Park adults	0.013	0.014	0.08208	0.077
Long adults	Park recruits	0.013	0.015	0.06283*	0.088
Long adults	Cat adults	0.017	0.019	0.03681*	0.149
Long adults	Cat recruits	0.015	0.020	0.00000**	0.244
Long adults	Eleuthera adults	0.014	0.019	0.00000**	0.272
Long adults	Eleuthera recruits	0.015	0.020	0.00001**	0.258
Long adults	LSI 2005 adults	0.014	0.019	0.00000**	0.269
Long adults	LSI 2005 recruits	0.013	0.019	0.00000**	0.316
Long adults	LSI 2004 adults	0.016	0.020	0.00007**	0.196
Long adults	LSI 2004 recruits	0.012	0.019	0.00000**	0.337
Long recruits	Park adults	0.013	0.014	0.32366	0.027
Long recruits	Park recruits	0.014	0.014	0.22518	0.042
Long recruits	Cat adults	0.017	0.019	0.09565	0.117
Long recruits	Cat recruits	0.015	0.019	0.00000**	0.219
Long recruits	Eleuthera adults	0.014	0.019	0.00000**	0.242
Long recruits	Eleuthera recruits	0.015	0.020	0.00001**	0.244
Long recruits	LSI 2005 adults	0.014	0.020	0.00000**	0.276
Long recruits	LSI 2005 recruits	0.013	0.019	0.00000**	0.317
Long recruits	LSI 2004 adults	0.016	0.021	0.00000**	0.208
Long recruits	LSI 2004 recruits	0.013	0.019	0.00000**	0.333
Park adults	Park recruits	0.013	0.014	0.19063	0.049
Park adults	Cat adults	0.016	0.019	0.05516	0.138
Park adults	Cat recruits	0.015	0.020	0.00001**	0.244
Park adults	Eleuthera adults	0.014	0.019	0.00000**	0.269
Park adults	Eleuthera recruits	0.015	0.020	0.00000**	0.275
Park adults	LSI 2005 adults	0.014	0.020	0.00000**	0.311
Park adults	LSI 2005 recruits	0.013	0.020	0.00000**	0.353
Park adults	LSI 2004 adults	0.016	0.021	0.00000**	0.241
Park adults	LSI 2004 recruits	0.013	0.020	0.00000**	0.366
Park recruits	Cat adults	0.017	0.019	0.07772	0.117
Park recruits	Cat recruits	0.015	0.020	0.00001**	0.213
Park recruits	Eleuthera adults	0.014	0.019	0.00001**	0.234
Park recruits	Eleuthera recruits	0.015	0.020	0.00002**	0.241
Park recruits	LSI 2005 adults	0.014	0.020	0.00000**	0.279
Park recruits	LSI 2005 recruits	0.013	0.020	0.00000**	0.332
Park recruits	LSI 2004 adults	0.016	0.021	0.00000**	0.210
Park recruits	LSI 2004 recruits	0.013	0.019	0.00000**	0.327

Cat adults	Cat recruits	0.019	0.020	0.09833	0.025
Cat adults	Eleuthera adults	0.017	0.020	0.04793*	0.137
Cat adults	Eleuthera recruits	0.019	0.021	0.09782	0.126
Cat adults	LSI 2005 adults	0.016	0.020	0.00916*	0.157
Cat adults	LSI 2005 recruits	0.016	0.021	0.00002**	0.251
Cat adults	LSI 2004 adults	0.019	0.021	0.15478	0.093
Cat adults	LSI 2004 recruits	0.016	0.019	0.02038*	0.169
Cat recruits	Eleuthera adults	0.016	0.018	0.01330*	0.120
Cat recruits	Eleuthera recruits	0.017	0.018	0.19879	0.055
Cat recruits	LSI 2005 adults	0.015	0.017	0.06603	0.089
Cat recruits	LSI 2005 recruits	0.015	0.019	0.00000**	0.237
Cat recruits	LSI 2004 adults	0.018	0.018	0.39213	0.021
Cat recruits	LSI 2004 recruits	0.014	0.015	0.20457	0.059
Eleuthera adults	Eleuthera recruits	0.016	0.017	0.06504	0.029
Eleuthera adults	LSI 2005 adults	0.015	0.017	0.00686*	0.131
Eleuthera adults	LSI 2005 recruits	0.014	0.019	0.00000**	0.274
Eleuthera adults	LSI 2004 adults	0.017	0.018	0.14267	0.070
Eleuthera adults	LSI 2004 recruits	0.014	0.015	0.03503*	0.111
Eleuthera recruits	LSI 2005 adults	0.015	0.016	0.29469	0.032
Eleuthera recruits	LSI 2005 recruits	0.014	0.018	0.00005**	0.209
Eleuthera recruits	LSI 2004 adults	0.018	0.017	0.74778	-0.047
Eleuthera recruits	LSI 2004 recruits	0.014	0.014	0.66135	-0.028
LSI 2005 adults	LSI 2005 recruits	0.014	0.016	0.09098	0.142
LSI 2005 adults	LSI 2004 adults	0.016	0.016	0.62418	-0.017
LSI 2005 adults	LSI 2004 recruits	0.014	0.014	0.30045	0.029
LSI 2005 recruits	LSI 2004 adults	0.016	0.018	0.03598*	0.110
LSI 2005 recruits	LSI 2004 recruits	0.013	0.016	0.00000**	0.222
LSI 2004 adults	LSI 2004 recruits	0.015	0.015	0.78843	-0.057
All samples		0.014	0.021	<0.0001**	0.288

* Significant at the 0.05 level

** Significant after a Bonferroni correction

Appendix D. Pair-wise F_{ST} values for all sample sites (below diagonal) and corresponding p-value after 10,000 randomizations (above diagonal). Similar p-values were obtained with exact tests. Significant tests after a Bonferroni correction are indicated in bold. Recruit samples are indicated with an asterisk (*).

	Long	Long*	Park	Park*	Cat	Cat*	Eleuthera	Eleuthera*	TS
Long		0.352	0.453	0.040	0.533	0.646	0.578	0.365	0.572
Long*	0.006		0.847	0.038	0.758	0.309	0.145	0.501	0.392
Park	0.005	0.006		0.769	0.918	0.717	0.847	0.716	0.790
Park*	0.008	0.009	0.006		0.759	0.532	0.189	0.041	0.401
Cat	0.006	0.006	0.006	0.007		0.739	0.184	0.769	0.586
Cat*	0.006	0.007	0.007	0.008	0.006		0.565	0.637	0.110
Eleuthera	0.005	0.007	0.005	0.006	0.006	0.006		0.899	0.554
Eleuthera*	0.007	0.006	0.007	0.010	0.006	0.007	0.006		0.183
TS	0.007	0.008	0.007	0.008	0.008	0.010	0.007	0.010	
BP	0.007	0.007	0.006	0.009	0.008	0.007	0.008	0.009	0.011
BP*	0.006	0.007	0.006	0.007	0.006	0.006	0.006	0.008	0.009
BR	0.007	0.007	0.008	0.009	0.009	0.008	0.007	0.008	0.009
SB*	0.007	0.009	0.009	0.009	0.007	0.009	0.008	0.010	0.009
CC	0.007	0.007	0.006	0.006	0.006	0.007	0.005	0.008	0.008
SR	0.006	0.005	0.005	0.007	0.005	0.007	0.005	0.007	0.008
SB	0.005	0.006	0.006	0.006	0.006	0.006	0.006	0.009	0.006
LSI*	0.007	0.007	0.007	0.009	0.006	0.005	0.007	0.008	0.009
LSI	0.010	0.008	0.009	0.010	0.009	0.011	0.010	0.011	0.010

	BP	BP*	BR	SB*	CC	SR	SB	LSI*	LSI
Long	0.569	0.588	0.412	0.318	0.025	0.506	0.074	0.052	0.900
Long*	0.275	0.200	0.416	0.194	0.168	0.888	0.147	0.035	0.994
Park	0.997	0.815	0.594	0.253	0.558	0.962	0.355	0.175	0.985
Park*	0.341	0.269	0.079	0.231	0.251	0.184	0.065	0.030	0.657
Cat	0.617	0.599	0.331	0.900	0.492	0.941	0.075	0.221	0.947
Cat*	0.761	0.345	0.450	0.368	0.096	0.471	0.250	0.582	0.778

Eleuthera	0.419	0.193	0.260	0.279	0.167	0.880	0.093	0.083	0.838
Eleuthera*	0.644	0.481	0.666	0.216	0.242	0.852	0.034	0.015	0.929
TS	0.279	0.685	0.672	0.522	0.249	0.285	0.279	0.119	0.980
BP		0.604	0.208	0.254	0.210	0.767	0.023	0.456	0.995
BP*	0.007		0.115	0.397	0.130	0.316	0.006	0.087	0.889
BR	0.010	0.009		0.245	0.186	0.339	0.467	0.002	0.526
SB*	0.010	0.008	0.010		0.136	0.710	0.094	0.125	0.712
CC	0.007	0.006	0.008	0.008		0.287	0.002	0.012	0.358
SR	0.007	0.006	0.008	0.007	0.006		0.062	0.009	0.955
SB	0.009	0.007	0.006	0.008	0.007	0.006		**0.000**	0.323
LSI*	0.007	0.006	0.010	0.009	0.008	0.007	0.008		0.788
LSI	0.009	0.009	0.013	0.010	0.010	0.009	0.010	0.010	