Available online at www.sciencedirect.com
SCIENCE DIRECT
Fisheries Research 77 (2006) 159–172
FISHERIES RESEARCH
www.elsevier.com/locate/fishres

A Bayesian approach for the genetic tracking of cultured and released individuals

Michael D. Tringali *

Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, 100 Eighth Avenue, S.E., St. Petersburg, FL 33701, USA

Received 28 March 2005; received in revised form 19 September 2005; accepted 12 October 2005

Abstract

In all supplemental stocking programs, regardless of scale, at least some of the released animals should be tracked (recaptured and identified) to evaluate and quantify the effect of the release on wild stocks. Often, marking these animals extrinsically can be impractical. Here, a parentage-based (familyprinting), Bayesian approach is presented for genetically tracking individuals produced in captivity and released among wild conspecifics. Any class of autosomal, codominant, molecular markers may be used, provided that loci are independent and population genotype frequencies conform to Mendelian expectations for diploid systems. Incorporating reference allele-frequency data from the recipient stock and genotype data from the captive parents, parentage of tested individuals can be established via likelihood ratios that compare the probability of the genetic evidence for coparentage to the probability for coincidence for individuals whose genotypes are compatible with parental pairs. Given a sufficient number of variable loci, products of these likelihood ratios and appropriate prior probabilities yield sufficiently large posterior probabilities of coparentage, i.e., very low expectations for false-positive assignment. Thus, post-release differences in growth, survivorship, or performance traits may be evaluated among groups, among families, or among genotypes and various stocking practices (e.g., size-at-release, release location) can be studied in vivo. The principal benefit of the approach occurs when family sizes of hatchery breeding pairs are considerably larger than those of wild pairs in the stocked population, as expected during successful enhancement. An application of the method to a large-scale stocking program is described, including results of blind performance testing and mutation rate analyses to investigate program error rates.

Keywords: Stock enhancement; Genetic tag; Familyprinting; Red drum; *Sciaenops ocellatus*

1. Introduction

Supplemental stocking has been practiced worldwide at various levels for centuries, but evaluation of post-release survival of stocked fish and their contribution to the fishery is, for the most part, an emergent component of such activities (Leber and Lee, 1997). The need to track released fish with minimal influence on their behavior, health, or survival has led to the development of a variety of extrinsic tags (Guy et al., 1996). Among these are coded wire tags (CWT), passive integrated transponders (PIT tags), body-cavity tags, anchor-type tags, and visible implants. Each tag type has advantages and disadvantages in given applications. For example, PIT tags can be repeatedly 'sampled' without harm to the fish but are expensive; CWTs are comparatively inexpensive but must be extracted from fish to be read. All of the above-listed tags share at least two common disadvantages—there is a lower limit to the size of fish into which they may be safely inserted and a practical upper limit to the number of fish that can be tagged. Because it may be more cost-effective to release large numbers of small fish than small numbers of large fish (e.g., Kent et al., 1995; Wilson et al., 1998), stocking programs may benefit from a method of tracking that is not constrained by the number of fish to be released or their size at release.

Molecular genetic markers have been used extensively to identify and monitor hatchery fish in supplemented stocks (e.g., Murphy et al., 1983; Taggart and Ferguson, 1986;

* Tel.: +1 727 896 8626; fax: +1 727 823 0166.
E-mail address: mike.tringali@myfwc.com.

0165-7836/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.fishres.2005.10.007

Gharett and Seeb, 1990; King et al., 1993; Crozier and Moffett, 1995; Hansen et al., 1995; Tessier et al., 1997; Wilson et al., 1997; Norris et al., 1999; Perez-Enriquez and Nobuhiko, 1999). Mixed-stock and population-assignment analyses (Millar, 1987; Pella and Masuda, 2001) have been used to estimate the relative contributions of hatchery and wild fishes in admixtures (e.g., Hansen et al., 1995; Kamonrat, 1996) and to assign individuals to hatchery or wild stocks (e.g., Hansen et al., 2001; Koskinen et al., 2002). However, because these analyses require sufficiently high levels of genetic heterogeneity between hatchery and wild stocks for precision, they are not readily applicable to all stocking programs, especially those in which broodfish are randomly sampled from the wild each generation.

An alternative approach – familyprinting – has been suggested (Letcher and King, 1999). Familyprinting has been defined as identifying (assigning) the parentage of tested individuals. Like most mixed-stock analyses, familyprinting involves the use of multilocus genotype data. Computer simulations have indicated that familyprinting could potentially be used to determine the parental pairs of progeny sampled from a genetically homogeneous (unstructured) population (Letcher and King, 1999; Bernatchez and Duchesne, 2000; Eldridge et al., 2002). Unfortunately, no theoretic framework has been proposed for evaluating levels of confidence in parental-pair assignments on a case-by-case basis. Investigators have instead relied on post hoc simulations to estimate the group-wise power of their loci to correctly include or exclude parentage. As reviewed by Jones and Ardren (2003), such simulations do not take advantage of all available information and may be biologically unrealistic because they rely on assumptions of random mating (e.g., within and among hatchery and wild breeders) and binomial variance in family size. They also require and are sensitive to an estimate of the total (hatchery and wild) number of breeding pairs per generation interval, N_{HW}, in the system, which, in the case of marine stock enhancement, is expected to be quite large ($>10^4$). When system-wide random mating is assumed, statistical power for hatchery parental assignments declines rapidly as N_{HW} increases.

In most cases, however, hatchery breeders are segregated from wild breeders. When the stocking program is relatively effective, the reproductive successes of hatchery breeding pairs are considerably greater on average than those of wild breeding pairs in the system. Consequently, parentage probabilities for offspring of hatchery breeding pairs in hatchery/wild admixtures are expected to be higher than standard simulations would predict. Here, I present a Bayesian framework for a parentage-based method of tracking individuals produced in captivity and released into wild populations wherein probabilities of correctly assigning parentage can be computed directly for each tested individual. When the probabilities are appropriately conditioned, the need for post hoc power estimation is circumvented and relevant issues involving family structure are addressed. General application of the method to the post-release monitoring of captive-bred fish is discussed and illustrated via a case study of an ongoing stocking program for red drum (*Sciaenops ocellatus*).

1.1. Marker-based parentage testing

Bayesian methods may be used to examine the probability that a hypothesis ($H1$) is true given the observed data, relative to one or more competing hypotheses (e.g., $H2$). A *prior probability* – the probability that a hypothesis is true prior to consideration of the observed data – may be specified. The prior probability is based on prior or conditioning information (I). The term *likelihood* is used to describe the conditional probability of observing the new data (D) given a particular hypothesis. The likelihood ratio (L) is the ratio of two probabilities of obtaining D under competing hypotheses. The term *posterior probability* refers to the probability that $H1$ is true given D and I. Bayes' well-known theorem (Bayes, 1763) states that

$$\frac{\Pr(H1|D, I)}{\Pr(H2|D, I)} = \frac{\Pr(D|H1, I)}{\Pr(D|H2, I)} \frac{\Pr(H1|I)}{\Pr(H2|I)}, \quad (1)$$

or

$$\text{posterior odds} = \text{likelihood ratio} \times \text{prior odds},$$

where the odds for two events is the ratio of their probabilities. When alternative hypotheses are evaluated, $\Pr(H2|D, I) = 1 - \Pr(H1|D, I)$ and $\Pr(H2|I) = 1 - \Pr(H1|I)$; thus, posterior odds may be converted to posterior probabilities by rewriting Eq. (1) to give

$$\Pr(H1|D, I) = \frac{\Pr(H1|I) \times L}{[1 - \Pr(H1|I)] + \Pr(H1|I) \times L}. \quad (2)$$

Because likelihood ratios are proportional to probabilities, the multiplicative law may be applied over multiple, independent sets of data (e.g., multiple, unlinked loci) to obtain a likelihood ratio on the combined data (Edwards, 1992).

The statistical approach for genetic tracking is based on computation of the joint probability of maternity and paternity, which, in Bayesian terms, may be described as the *posterior probability of coparentage*, $\Pr(\text{CP}|D, I)$. In other words, we seek to determine if a tested individual is the offspring of a specific parental pair (mother and father), whose multilocus genotypes are known. For brevity, the terms D and I will be hereafter omitted from posterior probabilities. The posterior probability Pr(CP) may be referred to generically as an *assignment probability* in that it may be used to assign a tested individual to a parental pair. To do so, multilocus genotypes for the tested individual, the putative mother and the putative father are examined. For each locus, the conditional likelihood ratio for coparentage (L_{CP}) may be taken as the quotient X_{CP}/Y_{CP}, where X_{CP} and Y_{CP} specify the following probabilities:

- $X_{CP} = \Pr\{$observing the tested individual's genotype when the putative mother and father are the actual parents$\}$.

- Y_{CP} = Pr{observing the tested individual's genotype when the actual mother and father are random, unrelated breeders from the relevant population}.

Genotype-dependent formulas for this likelihood ratio (LRF_{CP}) may be derived as follows. Let G_i, G_p, and G_m represent the genotypes of the tested individual, putative father, and putative mother, respectively, at a given locus. Let $H1$ correspond to the probability of observing G_i, assuming G_m and G_p are from the actual parents; X_{CP} can be expressed as

$$X_{CP} = \Pr(G_m, G_p, G_i|H1, I) = \tau(G_i|G_m, G_p)\delta(G_m)\delta(G_p), \quad (3)$$

where $\tau(G_i|G_m, G_p)$ is the Mendelian segregation probability of observing the individual's genotype given the putative mother's and father's genotypes and $\delta(G_m)$ and $\delta(G_p)$ are the frequencies of the putative mother's and father's genotypes in the reference population. Let $H2$ correspond to the probability of observing G_i, when the actual mother and father are random, unrelated breeders; Y_{CP} is

$$Y_{CP} = \Pr(G_m, G_p, G_i|H2, I) = \delta(G_i)\delta(G_m)\delta(G_p), \quad (4)$$

where $\delta(G_i)$ is the frequency of the tested individual's genotype in the reference population. Thus,

$$\frac{X_{CP}}{Y_{CP}} = \frac{\Pr(G_m, G_p, G_i\,|H1, I)}{\Pr(G_m, G_p, G_i|H2, I)} = \frac{\tau(G_i|G_m, G_p)\delta(G_m)\delta(G_p)}{\delta(G_i)\delta(G_m)\delta(G_p)} = \frac{\tau(G_i|G_m, G_p)}{\delta(G_i)}, \quad (5)$$

which may be taken as L_{CP} and provides the basis for LRF_{CP} given in Table 1. Finally, for k loci and a prior $\Pr(SCP|I_S, I_{CP})$, the posterior probability of coparentage becomes

$$\Pr(CP) = \frac{\Pr(SCP|I_S, I_{CP})\prod_{i=1}^{k} L_{CP,i}}{[(1 - \Pr(SCP\,|I_S, I_{CP}\,)) + \Pr(SCP|I_S, I_{CP})\prod_{i=1}^{k} L_{CP,i}]} = \frac{\Pr(SCP|I_S, I_{CP}) \times CPI}{(1 - \Pr(SCP|I_S, I_{CP})) + \Pr(SCP|I_S, I_{CP}) \times CPI}, \quad (6)$$

where CPI is defined as the cumulative *coparentage index*. A mathematical proof of the coparentage index (Eq. (5)), based on the well-known single-parent indices, was given elsewhere (Tringali, 2005, see also Meagher and Thompson, 1986). Because L_{CP} depends in part on $\delta(G_i)$, the assumption of Hardy–Weinberg equilibrium (HWE) applies. Reference-sample allele frequencies are assumed to represent population allele frequencies. It is further assumed that linkage disequilibrium, mutations, and genotyping errors do not occur. A method for incorporating a presumptive genotyping error rate into the coparentage index is given in Appendix 1. An approach to the development of an appropriate experimental prior, $\Pr(SCP|I_S, I_{CP})$, for typical culture-release situations is discussed in the following section.

Table 1
Single-locus likelihood-ratio formulae (LRF_{CP}) for computing L_{CP}

G_i[a]	G_m[a]	G_p[a]	$\tau(G_i\|G_m, G_p)$[b]	$\delta(G_i)$[c]	LRF_{CP}
qq	***pq***	***qq***	1/2	q^2	$1/2q^2$
pq	***pp***	***qq***	1	$2pq$	$1/2pq$
pq	***pr***	***qq***	1/2	$2pq$	1/4pq
qq	***qq***	***qq***	1	q^2	$1/q^2$
pq	***pp***	***qr***	1/2	$2pq$	$1/4pq$
pq	***pr*** or ***ps***	***qr***	1/4	$2pq$	$1/8pq$
qq	***pq***	***qr***	1/4	q^2	$1/4q^2$
qq	***qq***	***qr***	1/2	q^2	$1/2q^2$
pq	***pq***	***pq***	1/2	$2pq$	$1/4pq$
pq	***pq***	***qq***	1/2	$2pq$	$1/4pq$
pq	***pq***	***qr***	1/4	$2pq$	$1/8pq$
pq	***pp***	***pq***	1/2	$2pq$	$1/4pq$
pq	***pr***	***pq***	1/4	$2pq$	$1/8pq$
qq	***pq***	***pq***	1/4	q^2	$1/4q^2$

[a] Genotypes for the tested individual (G_i), putative maternal (G_m), and paternal (G_p) parents.
[b] $\tau(G_i|G_m, G_p)$ is the probability of observing the tested individual's genotype given the putative mother's and putative father's genotypes.
[c] $\delta(G_i)$ is genotype frequency of the tested individual in the reference population. The quantities p and q represent population frequencies for alleles ***p*** and ***q***.

1.2. Testing multiple individuals over a subset of potential breeding pairs

In the present application, parentage tests involving a known subset of parental pairs (S) from a mixed-generation population are to be performed for many individuals collected from the entire population. Let S be restricted to the group of possible breeding pairs (not individual breeders) that will be removed from the wild population and mated in captivity to produce fish for stocking. When attempting to identify stocked fish from a cultured fish + wild fish admixture, some of the tested individuals may be assigned to one (or more) of the parental pairs in S and some may be genotypically incompatible with all pairs in S, having at least one parent not in the subset. For k loci, the probability that a tested individual will be found that is compatible with a parental pair in S strictly due to chance – i.e., the probability that $H2$ obtains – increases with each additional individual that is tested and with the number of parental pairs in S it is compared to.

A Bayesian framework may be adopted whereby individual assignment probabilities are updated with a prior that represents the probability that the nth tested individual will belong to a given parental pair in S prior to consideration of its multilocus genotype. In this framework, individuals may be tested sequentially until a compatible individual (one that might be the offspring of a pair in S) is found. The probability that a tested individual will be the offspring of any one of the parental pairs in subset S (i.e., that it originated from the hatchery) may be then approximated by $\Pr(S|I_S) \cong [(CPI \times 1/N_{HW})/((1 - 1/N_{HW}) + (CPI \times N_{HW}))]/n_T$, where n_T is the total number of individuals tested thus far. The reader

may recognize that the term $[(\mathrm{CPI} \times 1/N_{\mathrm{HW}})/((1 - 1/N_{\mathrm{HW}}) + (\mathrm{CPI} \times 1/N_{\mathrm{HW}}))]$ equates to Eq. (6) when there is no prior information regarding relative hatchery contribution; thus, we assume sufficient genetic evidence exists, such that $[(\mathrm{CPI} \times 1/N_{\mathrm{HW}})/((1 - 1/N_{\mathrm{HW}}) + (\mathrm{CPI} \times 1/N_{\mathrm{HW}}))] \cong 1$. Numerical inspection reveals that this condition is satisfied when CPI is approximately five times greater than N_{HW}.

Let $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$ represent the prior probability that a tested individual is the offspring of a particular pair in the subset, given that it is the offspring of one of the pairs in that subset. If it may be initially assumed that each breeding pair in S has an equivalent chance of being the parents of a tested individual given that the individual is the offspring of one of the pairs in the subset, then $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}}) = 1/N_{\mathrm{H}}$, where N_{H} is the number of mated or possibly mated parental pairs in S. By the multiplicative law, the probability that the tested individual will belong to a given parental pair in subset, $\Pr(\mathrm{SCP}|I_{\mathrm{S}}, I_{\mathrm{CP}})$, is equal to the product of $\Pr(S|I_{\mathrm{S}})$ and $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$. Therefore, taking $\Pr(S|I_{\mathrm{S}})$ and $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$ as $[(\mathrm{CPI} \times 1/N_{\mathrm{HW}})/((1 - 1/N_{\mathrm{HW}}) + (\mathrm{CPI} \times 1/N_{\mathrm{HW}}))]/n_{\mathrm{T}}$ and $1/N_{\mathrm{H}}$, respectively, $\Pr(\mathrm{SCP}|I_{\mathrm{S}}, I_{\mathrm{CP}}) \cong [(\mathrm{CPI} \times 1/N_{\mathrm{HW}})/((1 - 1/N_{\mathrm{HW}}) + (\mathrm{CPI} \times 1/N_{\mathrm{HW}}))]/n_{\mathrm{T}}N_{\mathrm{H}}$. When post-release specimens represent a simple-random sample of the hatchery-wild admixture, the prior $\Pr(S|I_{\mathrm{S}})$ may then be updated with each new individual tested, becoming $x_{\mathrm{S}}/n_{\mathrm{T}}$, where x_{S} may be approximated by

$$x_{\mathrm{S}} \cong \sum_{i=1}^{n_{\mathrm{T}}} \mathrm{PR(CP)}_i \tag{7}$$

Thus, when n_{T} is sufficiently large, $\Pr(\mathrm{SCP}|I_{\mathrm{S}}, I_{\mathrm{CP}}) \cong x_{\mathrm{S}}/n_{\mathrm{T}}N_{\mathrm{H}}$.

There is likely to be variance, sometimes considerable, in family sizes of hatchlings at release in fish stocking programs. Thus, the uniform prior may not accurately describe $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$. When $1/N_{\mathrm{H}}$ overestimates the relative contribution of a given breeding pair, $\Pr(\mathrm{SCP}|I_{\mathrm{S}}, I_{\mathrm{CP}})$ will, in effect, be liberal because posterior probabilities involving that pair will be overstated. Conversely, when $1/N_{\mathrm{H}}$ underestimates the relative contribution of a given pair, $\Pr(\mathrm{SCP}|I_{\mathrm{S}}, I_{\mathrm{CP}})$ will be conservative and posterior probabilities involving that pair will be understated. If it is available, information relating to the numbers of progeny released per breeding pair can initially be developed into a prior for $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$. When sufficient recapture data become available that allow specification of the relative post-release contribution of each breeding pair, $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$ can be updated. An iterative process that enables refined estimates of $\Pr(\mathrm{SCP}|I_{\mathrm{S}}, I_{\mathrm{CP}})$ for cases when CPI values are modest will be given elsewhere.

In some cases, genotype data may only be available for a single parent. If so, the single-locus likelihood ratio ($X'_{\mathrm{M}}/Y'_{\mathrm{M}}$, Appendix 1) for the fatherless maternity index (MI) may be used, which reduces to L_{M} given in Tringali (2005) when $e = 0$. The expression for motherless paternity (PI) is equivalent, albeit parameterized for alleged fathers. To obtain posterior probabilities $\Pr(M)$ for maternity, the respective indices require appropriate conditioning. For single-parent assignment, let S now refer strictly to the group of potential female breeders (male breeders if paternity is to be established). Again, a two-part prior is employed where $\Pr(S|I_{\mathrm{S}}) \cong [(\mathrm{MI} \times 1/N_{\mathrm{F,HW}})/((1 - 1/N_{\mathrm{F,HW}}) + (\mathrm{CPI} \times 1/N_{\mathrm{F,HW}}))]/n_{\mathrm{T}}$ is the first component, where $N_{\mathrm{F,HW}}$ is the total number of female breeders in the system. In lieu of $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}})$, a single-parent version of the prior is needed. When maternity is to be tested and maternal variance in S is binomial, the prior probability that a tested individual was derived from a particular female in S given that one of the females in S is the mother may be taken as $\Pr(M|S, I_{\mathrm{M}}) = 1/N_{\mathrm{F,H}}$, where $N_{\mathrm{F,H}}$ is the total number of female breeders in S. As before, this prior may be updated with information regarding relative individual contributions. Analogous priors and likelihood ratios may be similarly developed for tests of paternity.

1.3. Case study: red drum stock enhancement

The method described above is being used to track released red drum in an experimental stocking program conducted by the Florida Fish and Wildlife Research Institute (FWRI, St. Petersburg, FL) and Mote Marine Laboratory (Sarasota, FL). Stocking began in Tampa Bay, Florida (Fig. 1), in March 2000 and is expected to continue through 2005, during which time a projected four million red drum will have been produced and released. Genetic tracking will continue at least through 2008. The experimental design for stocking hatchery-reared red drum, replicated over time, includes the following variables: river system, distance of release site from river mouth, synchronicity or asynchronicity with respect to natural production, and size-at-release. Most of the hatchlings (95%) stocked into Tampa Bay were released at sizes considered to be too small for extrinsic tagging (<65 mm standard length, SL). At the time of this article's submission, coparentage assays for >16,000 red drum from Tampa Bay have been completed. Of these, >1800 hatchery-derived red drum were identified (assigned to hatchery breeding pairs). The remaining specimens were inferred to be products of natural recruitment. Studies involving optimal size-at-release, release season, and release location are still ongoing and results will be reported at a later time. Here, I describe marker performance with respect to program objectives. Also, to investigate false-negative (improper rejection of $H1$) and false-positive (improper acceptance of $H1$) error rates in genetic assignment, blind performance testing and mutation screening was conducted.

2. Materials and methods

Post-release (test) specimens were provided to FWRI's Molecular Genetics Laboratory (MGL) for identification via the following sources: (1) fishery-independent stratified-random sampling (McMichael, 2000) and other sampling,

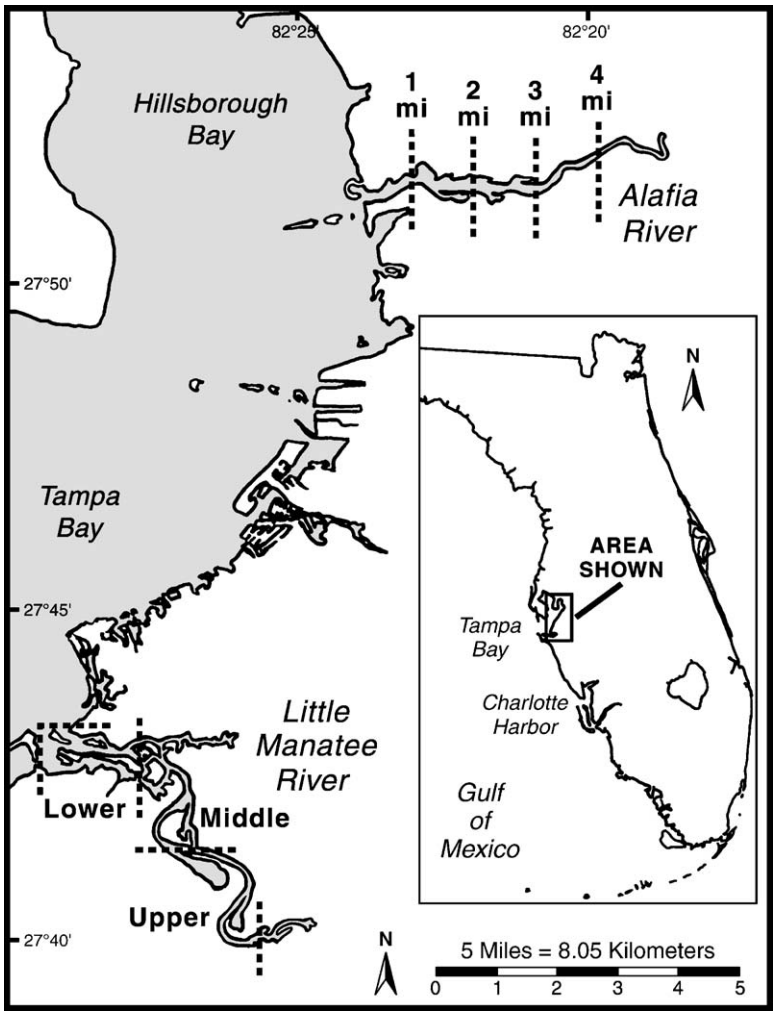

Fig. 1. Location of red drum study area. Insert shows Florida coast, including Charlotte Harbor, where wild specimens were collected for false-positive error testing. Detail map of Tampa Bay indicates boundaries for release sites in the Alafia and Little Manatee rivers.

(2) creel surveys, and (3) voluntary angler returns. Data collected with each specimen included capture date and location and standard and/or total length. Upon receipt, all tissues (predominantly fin clips) or whole juvenile specimens were maintained frozen at −80 °C until use. When whole fish were obtained, small pieces of somatic tissue were excised for assay. Broodfish for the stocking program were collected from the subadult (3–4 years old; Murphy and Taylor, 1990) red drum population inside Tampa Bay. Fin clips were obtained from all male and female broodfish at the time of their capture for use in DNA typing. MGL personnel have optimized eight sciaenid microsatellite loci (Turner et al., 1998) in multiplex PCR assays for use in the ongoing genetic-tracking study (Table 2). Alleles at loci *Soc*49, *Soc*85, *Soc*99, and *Soc*243 were co-amplified and assayed simultaneously, as were alleles at loci *Soc*83, *Soc*129, *Soc*133, and *Soc*276. Electropherograms were each scored by two readers independently; one attempt was usually made to resolve disputed or unscorable results via reassay.

2.1. Brood-stock information and general testing strategy

At the time of their collection, all members of the captive red drum brood stock ($n = 206$) were made individually identifiable via PIT tagging. From October 1999 to November 2004, 169 broodfish were subdivided into 34 small spawning groups (three to six individuals per group; 211 possible breeding-pair combinations) and conditioned for production (Table 3). Detailed records of spawning-group composition were maintained; breeders were not reused in alternate spawning groups or in subsequent production years. Each broodfish that potentially contributed to production was independently genotyped twice for all loci and scores were evaluated for consistency. Microsatellite DNA profiles, spawning-group membership, and breeding and release information were maintained in a parental database. The 206 broodfish were considered to be a random sample of wild red drum prior to any possible recruitment of stocked fish into the adult-subadult population. For these fish, the numbers of

Table 2
Eight microsatellite DNA loci used in post-release genetic tracking of hatchery-reared red drum (*Sciaenops ocellatus*)

Locus[a]	Repeat unit	PCR label[b]	No. of alleles	Size range of alleles[c]	H_0	HWE, *P*-value	S.E.
*Soc*49	CA	NED	20	209–249	0.91	0.2084	0.0272
*Soc*83	TG	6-FAM	18	114–152	0.81	0.9624	0.0122
*Soc*85	AC	HEX	17	77–121	0.85	0.8648	0.0221
*Soc*99	CA	HEX	25	157–211	0.92	0.6244	0.0370
*Soc*129	TATC	HEX	23	109–215	0.89	0.6561	0.0356
*Soc*133	TGC	6-FAM	6	190–205	0.51	0.8688	0.0105
*Soc*243	CCT	6-FAM	7	85–103	0.71	0.9512	0.0055
*Soc*276	CA	HEX	13	97–129	0.73	0.3601	0.0321

[a] Primer sequences given in Turner et al. (1998). The same reaction profile was used for all loci: 95 °C for 15 min (to activate HotStar Taq DNA polymerase), 32× (94 °C for 30 s, 56 °C for 30 s, 72 °C for 90 s), and 72 °C for 15 min. Negative-control reactions were performed for all grouped assays.
[b] The ABI flourescent dye used to 5′-end label one of the PCR primer pairs.
[c] Allele sizes, in base pairs, based on ABI 3100 Genetic Analyzer electrophoretic mobilities.

alleles, allele size ranges, allele frequencies, and observed and average heterozygosities were determined for all loci. Exact tests for Hardy–Weinberg equilibrium were conducted using GENEPOP Version 3.3 (Guo and Thompson, 1992; Raymond and Rousset, 1995). Exact tests of linkage disequilibrium among pairs of loci were conducted using LinkDos (Garnier-Gere and Dillmann, 1992); probabilities were analyzed for table-wide significance (Rice, 1989) at the 95% level using the sequential Bonferroni procedure ($k = 28$; $P \leq 0.0018$).

Table 3
Spawning-group composition and group-specific priori probabilities for broodfish used in the Tampa Bay red drum stocking program between November 1999 and July 2004

Spawning group(s)[a]	n_f[b]	n_m[b]	N_H	Size range at release[c]	Number released	Pr(CP|*S*, I_{CP})[d]
99-7	3	3	9	II, III	52357	0.00145
00-4A	2	1	2	I	40102	0.00498
00-4B, 5, 8, 10	7	11	19	III	39551	0.00052
00-6	1	3	3	I	48695	0.00403
00-7	2	1	2	I	108505	0.01348
00-9A	1	3	3	I	20725	0.00172
00-9B	2	1	2	I	83952	0.01043
01-4	2	2	4	II	40221	0.00250
01-5	2	2	4	I	91032	0.00566
01-6[e]	2	3	6	I	678013	0.02808
01-9	1	3	3	I	81162	0.00672
02-1	3	2	6	I	102896	0.00426
02-3	3	3	9	I	104976	0.00290
02-5	3	2	6	I	59862	0.00248
02-6A[e]	2	2	4	I	412066	0.02560
02-6B, 7	4	7	14	III	32256	0.00057
02-8	2	2	4	II	58209	0.00362
02-9	2	4	8	I	366015	0.01137
03-1	3	3	9	I	106316	0.00294
03-4A[e]	2	2	4	I	326068	0.02026
03-4B, 5	6	6	18	II	75496	0.00104
03-6	3	3	9	I	103329	0.00285
03-7	3	3	9	I	372158	0.01028
03-8	3	3	9	I	103488	0.00286
03-9	3	3	9	I	108780	0.00300
04-1	3	3	9	I	103000	0.00284
04-3	3	3	9	I	102058	0.00282
04-8	3	3	9	I	102690	0.00284
04-9	3	3	9	I	100005	0.00276
Total	79	90	211		4023983	

[a] The ##-# format indicates the production year and spawning tank(s); discrete groups occupying a spawning tank sequentially during a production year are denoted with A and B. In three cases, progeny from two or more spawning groups were combined for release.
[b] n_f = no. of female broodfish in the spawning tank; n_m = no. of males in the spawning tank.
[c] I = 25–65 mm SL; II = 66–135 mm SL; III > 135 mm SL.
[d] In this case (number released/total number released)/N_H. Current as of June 2005.
[e] Parents of hatchlings used in blind test.

The experimental design of the stocking program necessitated the testing of numerous red drum specimens. To increase the speed at which specimens could be processed, a screening procedure was initiated in November of 2001. That is, a four-locus (*Soc*49, *Soc*85, *Soc*99, and *Soc*243) exclusion-based assay was first performed for each test specimen. A spreadsheet program (MS Excel) was used to rapidly analyze specimens for genotypic compatibility with all possible parental pairs over the four loci; compatible specimens were referred to as 'candidates'. The program included a built-in CHECK function that identified (flagged) specimens that would have been compatible with a parental pair except that (1) both alleles at one locus differed by +1 or −1 repeat unit (potential mobility shift or scoring error), or (2) one allele at one locus differed by +1 or −1 repeat unit (potential scoring error or mutation). Specimens not identified as candidates were considered to be products of natural reproduction and not tested further. Parent–offspring relationships for candidates and flagged specimens were then further explored based on typing results at the four remaining loci (*Soc*83, *Soc*129, *Soc*133, and *Soc*276). A second spreadsheet program was used to analyze candidates for genotypic compatibility with breeding pairs over all eight loci and to assign values of Pr(CP) to those subsequently matched to a breeding pair. Electropherograms for flagged specimens were reexamined and/or rerun; scores were corrected if necessary.

2.2. Locus performance, blind testing, and empirical mutation rate analyses

Multilocus genotypes from 687 individuals identified as being compatible with the genotypes of hatchery breeding pairs were used to examine locus performance. These individuals were captured in Tampa Bay through fishery-independent sampling between November 2003 and June 2005. For these, single-locus likelihood ratios and their sample statistics [per-locus sample means ($\bar{L}_{CP}$) and standard errors] were calculated. Because the magnitude of $\bar{L}_{CP}$ is directly related to the amount of genetic information provided by the locus, these values were used to rank project loci. Next, observed CPI values for these offspring were transformed into log of the odds (LOD) scores (Meagher, 1986) using the natural logarithm. LOD score distributions were plotted for various subsets of loci. Using all loci and appropriate priors, posterior odds of coparentage were computed and compared to posterior odds conditioned with uninformed priors ($1/N_{HW}$). Because N_{HW} is not well known for the stocked red drum system, values of 10,000, 100,000, and 200,000 were considered.

The objective of blind performance testing was to determine if MGL staff could (1) correctly identify known hatchery-derived specimens and assign them to their known parental pairs and (2) correctly identify known wild specimens as such. Hatchery personnel routinely collect samples (n = 75–100 hatchlings) prior to release events and provide these to MGL staff for various analyses. A total of 100 hatchlings produced by spawning groups 01–6 (n = 40), 02–6 (n = 20), and 03–4 (n = 40) (Table 3) were packaged and labeled to appear as post-release (field-caught) specimens from the fishery-independent sampling program; these were submitted to MGL staff in conjunction with a monthly field-caught sample. Hatchling sizes were within the range of sizes observed in fishery-independent field-caught samples. One hundred wild young-of-the-year red drum collected from Charlotte Harbor (Fig. 1) were similarly packaged and submitted for genetic testing. All genotyping assays were performed as described previously; the two lab members conducting the assays were not aware that they were being tested or that such a test was a possibility.

The objective of the empirical mutation-rate analysis was to derive an expected rate of false-negative errors attributable to mutations for the loci used. This was accomplished using available genotype data (Tringali et al., 2004) from 700 non-released hatchling red drum progeny derived from seven spawning groups (n = 100 hatchlings per spawning group). Based on genotype compatibility, hatchlings were assigned to a parental pair within a known spawning group. In cases where all pairs within the known group of possible parents were incompatible with a hatchling, parentage was inferred according to the most parsimonious mutational explanation. Genotype data from the set of 100 captive-bred 'blind-test' hatchlings and their parents were combined with data from the above 700 hatchlings and their parents. Per-locus and per-individual mutation rates were estimated from presumptive mutations observed in the 800 hatchlings.

3. Results

For the 206 broodfish, the maximum number of alleles observed at a locus was 25 (*Soc*99) and the minimum was six (*Soc*133); the average number of alleles per locus was 16 (Table 2). Observed heterozygosities ranged from 0.51 to 0.92; average heterozygosity over all loci was 0.79. Genotype frequencies were consistent with Hardy–Weinberg equilibrium expectations at all loci. Probabilities for pairwise estimates of linkage disequilibrium ranged from 0.02 to 0.99; following sequential Bonferroni adjustment to exact test results, there was no table-wide support for a hypothesis of nonindependence between any locus-pair. With the exception of *Soc*129 (the tetranucleotide marker), red drum markers had irregular or multimodal allele-frequency distributions (Fig. 2).

In part, due to the nature of the allele distributions for red drum, the relationship between $\bar{L}_{CP}$ and allele number was not linear (Fig. 3). On average, high mean L_{CP} values were observed in 687 hatchery-compatible red drum for loci having ≥17 alleles. Two loci having six and seven alleles (*Soc*133 and *Soc*243, respectively) provided limited but useful information. As expected, LOD scores improved with the addition of loci (Fig. 4A). Considering all loci, the mean of the LOD score distribution was 19.89 (CPI = 434,627,543).

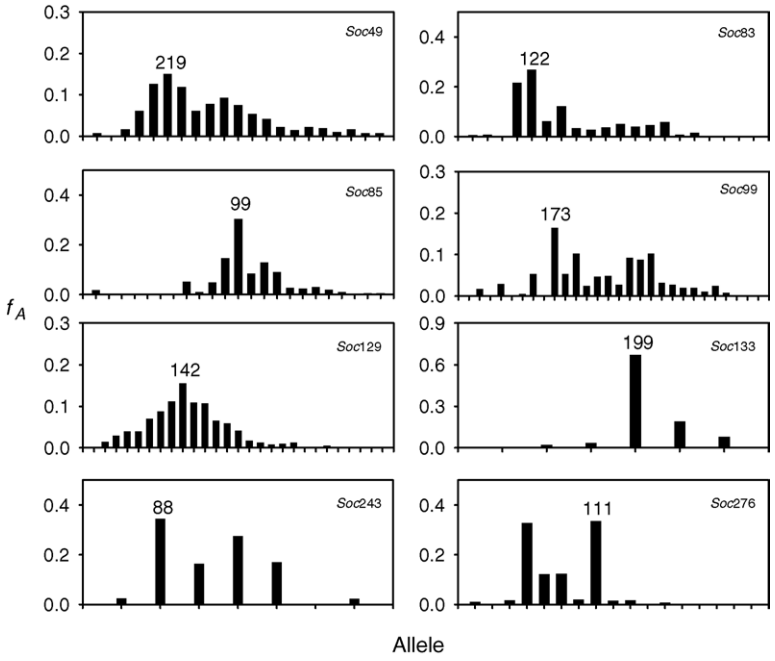

Fig. 2. Empirical allele frequencies (f_A) for nine microsatellite DNA loci used in the PTB genetic tracking. Incremental marks on the *x*-axis are indicative of single repeat units (alleles); the fragment size of the most common allele is identified for each locus.

To facilitate graphical representation, observed Pr(CP) values were converted to posterior odds and transformed using the natural logarithm (Fig. 4B). Upon conditioning with priors that included spawning-group specific Pr(CP|S, I_{CP}) (see Table 3), the empirical distribution of posterior probabilities of coparentage for the 687 compatible individuals ranged from 0.983 to 0.999999999829984. The mean of the distribution was 0.999994. Overall, these data indicate that sufficient genetic evidence is being obtained for highly accurate post-release identification of red drum.

To illustrate the effect that failure to consider family-size differences between hatchery and wild breeding pairs may have on statistical precision, posterior distributions for

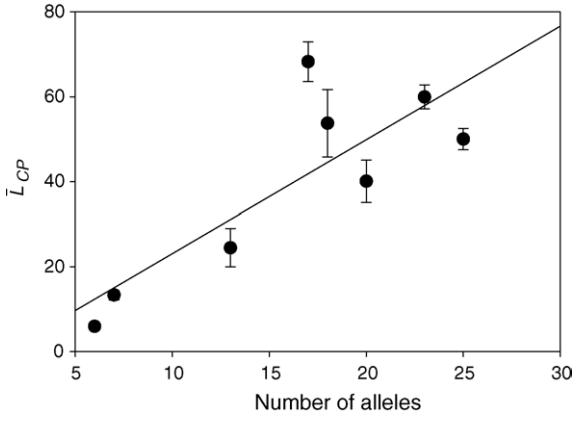

Fig. 3. Empirical means of single-locus likelihood ratios for 687 released and recaptured red drum as a function of the number of alleles per locus. Error bars represent one S.E. above and below the mean; the solid line depicts the linear regression for the data shown.

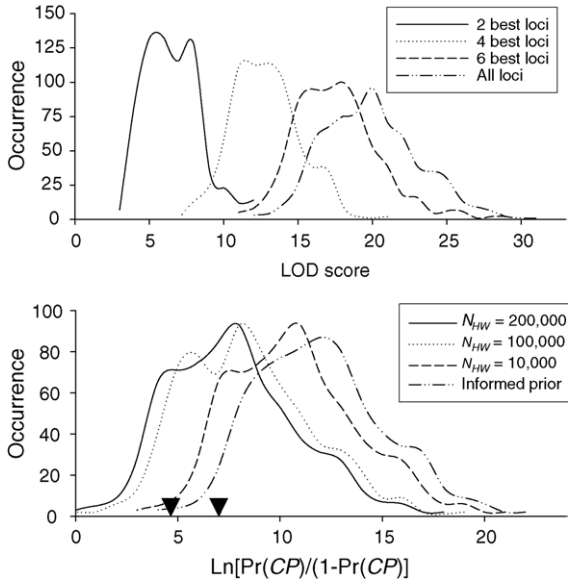

Fig. 4. Empirical distributions of LOD scores and posterior assignment probabilities for 687 released and recaptured red drum. (A) Project loci were ranked in descending order according to $\bar{L}_{CP}$, where *Soc*85 > *Soc*129 > *Soc*83 > *Soc*99 > *Soc*49 > *Soc*276 > *Soc*243 > *Soc*133. Loci were grouped into 2-, 4-, 6-, and 8-best (all loci) combinations to compute LOD scores. (B) Posterior probabilities of coparentage were converted to posterior odds and transformed using the natural logarithm. For uninformed priors, randomly mating populations having 10,000, 100,000, and 200,000 breeding pairs were considered. The informed prior was constructed as described in the text. The leftmost arrow indicates a critical threshold for Pr(CP) of 0.99; the rightmost arrow indicates a critical threshold of 0.999.

the above specimens were constructed based on uninformed priors—i.e., $\Pr(SCP|I_S, I_{CP}) = 1/N_{HW}$ (Fig. 4B). Because the true total number of red drum breeding pairs is difficult to estimate, a range of values was explored. In all cases, Pr(CP) values based on uninformed priors were lower than those based on informed priors. When the number of breeding pairs was large ($N_{HW} > 100{,}000$), much of the probability mass for uninformed Pr(CP) was below a threshold of 0.999, whereas the majority of the probability mass for informed Pr(CP) exceeded it.

For the blind test, 92 of the 100 known wild specimens were excluded via the initial four-locus screening process from assignment to any of the parental pairs from the spawning groups listed in Table 3; eight were compatible with at least one breeding pair. These eight wild specimens were advanced for additional testing and correctly excluded upon consideration of the genotype data at all eight loci. Of the known hatchery-derived specimens, 99 of 100 were identified as being compatible with a parental pair in the appropriate spawning group after the initial four-locus screening process; these were advanced for additional testing and were correctly assigned to a parent in the appropriate spawning group. The genotype of one known hatchery specimen was flagged as being one repeat unit removed from compatibility with a parental pair at locus *Soc*49 after the initial four-locus assay. This specimen was also advanced for additional testing and found to be otherwise compatible with the pair at all eight loci. Upon repetition of testing, the discrepancy at *Soc*49 could not be attributed to a scoring or binning mistake, and the specimen was classified as a possible mutant—the correct decision, given the known parentage of the specimen. Thus, no false-positive or false-negative errors were observed during the performance test.

In addition to the mutation observed at locus *Soc*49 in the blind test, two additional mutations were observed among the 700 prerelease hatchlings tested. These occurred at *Soc*85 and *Soc*99 and involved single unit changes. Accordingly, per-individual and per-locus mutation rates were 0.00375 (three mutations per 800 individuals) and 0.00047 (three mutations per 800 individuals per eight loci), respectively, for this set of red drum.

4. Discussion

Applying the genetic laws for the transmission of alleles from parent to offspring, genotype data from potential parents and tested individuals can be interpreted through likelihood ratios that compare the probability of the evidence for coparentage to its probability for coincidence. This method of inference, which is akin to a "missing persons" case in human forensic applications (Evett and Weir, 1998), provides a statistical basis for genetic tracking when probabilities are properly conditioned with prior information for a given experiment. Thus, if sufficiently informative genetic data can be obtained from broodfish, released hatchlings will be identifiable upon recapture from a mixture of hatchery and wild conspecifics. Genotype data are not required from hatchlings prior to their release. Importantly, genotypes of released organisms do not require manipulation via selective breeding (e.g., Gharett and Seeb, 1990; King et al., 1993), and the captive brood stock need not be genetically divergent as a unit from the recipient wild stock (e.g., Hansen et al., 2001; Koskinen et al., 2002).

For familyprinting, the unit of identification is the immediate family—mother, father, and offspring. As such, progeny groups from single families, extended family groups (full- and half-sib mixtures), and/or multiple groups of families can be established during hatchery production and tracked as discrete units upon release (Letcher and King, 1999). Post-release differences in growth, survivorship (e.g., Eldridge et al., 2002), or performance traits may be evaluated among groups, families, or individual breeders. If estimates of survivorship or performance differences among the release units are not confounded by interfamilial genetic effects, various stocking practices (e.g., size-at-release, stocking density, release location; Drawbridge, 2002) can be empirically examined. By mitigating for interfamilial effects through family mixing and replication, the scientific method can be applied (Leber, 1999) to improve or optimize stocking procedures.

Genetic circumstances will be unique for every stocking program. The recipient wild stock might or might not be substructured or inbred. The captive brood stock and wild stock might be genetically similar or divergent. Broodfish might be closely related to each other (Kincaid, 1995; Tessier et al., 1997; Tringali, 2003) or essentially unrelated. Here, I treated the simplest case, assuming that captive breeders were a randomly collected subset of the wild breeding population and were, on average, no more or no less related to each other than they were to wild breeders. Further, I assumed that the recipient population was not highly inbred, that the duration of stocking did not exceed a generation interval, and that the broodfish did not mate with others in subset *S* while in the wild prior to their collection. However, the basic approach is readily extendable to more complicated circumstances with appropriate attention to alternate hypotheses. For example, L_{CP} can be modeled as a latent variable under appropriate circumstances (inbreeding, Wahlund phenomena) and employed in Eq. (5).

4.1. Genetic systems and baseline population sampling

Genetic tracking may be accomplished as described with any class of molecular markers, provided they are autosomal, codominant, independent of among-locus associations, and reasonably polymorphic. This method can easily be extended to include dominant markers [e.g., RAPD (Williams et al., 1990) or AFLP (Voss et al., 1995)] with appropriate modification of L_{CP}. Microsatellite DNA markers are treated explicitly herein because they generally meet the above criteria (Wright, 1993; Wright and Bentzen, 1994) and are widely available for many fish species. Moreover, because microsatellite markers

are resolved via PCR assays, invasive or destructive sampling of organisms is not required and lab work can be expedited by assaying multiple markers simultaneously (e.g., O'Reilly et al., 1996). Two critical factors that should be considered during marker development are reliability and variability. Regarding marker variability, inspection of the likelihood-ratio formulas (Table 1) reveals that uninformative likelihood ratios ($L_{CP} < 1$) will not occur, even when the tested individual and potential parents share common alleles. For coparentage assignment, highly informative ratios should be achievable when numbers of alleles exceed ~15, frequencies of the most common allele do not exceed 0.25, and allele distributions are broad.

Representative population allele frequencies form the basis for $H2$ probability estimates in parentage testing. Populations of relevance are defined explicitly by the probability Y_{CP}; i.e., a reference sample should be drawn randomly from among the population of adults who could be considered to be possible contributors of gametes to the tested individual(s). Thus, non-Bayesian methods that simulate genotype distributions using genetic information from test individuals themselves introduce a source of bias when family sizes of hatchery breeders are larger than those of wild breeders. At the inception of a program, it is recommended that a random sample of genotypes from the breeding population be taken prior to stocking and that the associated allele frequencies be employed in likelihood-ratio computations during the first generation of stocking. Genotype frequencies from the reference population should be examined for conformance to Hardy–Weinberg equilibrium expectations. If the captive brood stock comprises a sufficient random sample of the breeding population, broodfish alleles may be used or included with those from other baseline samples to generate allele distributions. When broodfish have rare alleles, a slight bias may occur in posterior probability estimates because the frequencies of those alleles are likely to be over-represented with respect to the reference population. In this case, the direction of the bias is such as to make the assignment probability conservative. Fortunately, because all likelihood ratios will be ≥ 1, conservative allele frequency estimates will not lead to "running up a down escalator" phenomena (Brenner, 1997) with the addition of more markers, as they often do in more complex pedigree problems.

4.2. Prior probabilities

Generally, all posterior probabilities may be viewed as conditional (Evett and Weir, 1998), and conditioning information and circumstances for each tested individual may differ. For example, if a tested individual was caught in a sampling event that contained other released individuals or if it was obtained from the vicinity of the release location shortly after release, one might expect the prior odds that it was a released fish to be comparatively high. The conditioning approach described here assumes simple random sampling of a well-mixed cultured fish + wild fish admixture. Overdispersion (e.g., patchiness in the spatial distribution of released fish), however, may lead to individual assignment probabilities that are liberal or conservative depending on the direction of sampling bias. Program managers should be mindful of this when designing sampling protocols and during analysis. An important benefit of the Bayesian approach is that it allows specification of alternate priors for different testing conditions. For example, in the red drum project, separate priors are employed for fishery-independent sampling (largely, collections of juveniles near release sites) and fishery-dependent sampling. As expected, informed fishery-independent priors for the program have proven to be considerably higher than informed priors for fishery-dependent sampling. When uncertainty exists over the accuracy of prior probabilities, it may sometimes suffice to report posterior probabilities over a range of plausible priors (Dickey, 1973).

Low priors have the effect of reducing certainty in the genetic information reflected by CPI. When many broodfish are synchronously mated in mass spawning events, the number of possible parental pairs and, thus, number of required comparisons, will increase and the prior $\Pr(\mathrm{CP}|S, I_{CP})$ will be reduced. Information regarding all *possible* parental pairs (e.g., broodfish sharing a spawning tank) should be recorded. In some cases, such as when broodfish are segregated into discrete spawning groups during production, some male/female pairings would not be possible. Accordingly, $\Pr(\mathrm{CP}|S, I_{CP})$ can be kept as high as possible during analysis by avoiding artificial comparisons—e.g., testing over a pair of broodfish that did not share a spawning tank. When fish are released over multiple years, broodstock are not reused, and tested individuals are not aged, the number of potential parental pairs will increase and the resolution of the genetic-tracking technique will decrease. As additional breeding pairs are employed in the red drum stocking program and more compatible individuals are observed, the addition/substitution of one or more loci will be required to maintain the robust false-positive error rate.

4.3. Programmatic error rate and assignment decision

During the process of genetic tracking, assignment errors may occur in various ways, e.g., mutation, null alleles, erroneous binning or scoring of raw data, or transcription error. These examples represent *false-negative errors*—errors that typically lead to the rejection of $H1$ when it is true. Errors of this type may be systemic when a parent has been incorrectly typed at any locus, negating the potential that any of its offspring will be identified. Systemic errors may result in downwardly biased estimates of hatchery:wild fish ratios and confound program comparisons involving progeny of particular breeding pairs or spawning groups. These can be greatly reduced by typing each potential breeder twice. Nonsystemic false-negative errors (mutation, mistyped test individuals) have an effect similar to tag loss in extrinsic tagging studies, i.e., they lead to biased estimates of ratios of hatchery and wild fish. Therefore, programmatic genotyping error rates

should be investigated. No typing errors were noted in the blind test for red drum, suggestive of an error rate lower than 1.0% per individual tested. More performance tests will have to be conducted to determine the actual rate.

For most animal species, microsatellite mutation rates are assumed to range between 10^{-3} and 10^{-4} per locus per generation (Dallas, 1992; Zhivotovsky and Feldmen, 1995). The overwhelming majority of these mutations occur in the form of single-unit increases or decreases (which are about equally likely), although multiple-unit increases have been documented (Henderson and Petes, 1992; Weber and Wong, 1993; Amos et al., 1996). The estimated mutation rate for red drum (4×10^3) was within the reported range for other animals. As for the potential impact of mutation on red drum tracking, approximately one mutation-related incompatibility should be expected for every 270 hatchery-derived individuals so identified during testing.

As in the red drum study, it is advisable to employ built-in quality control procedures to mitigate potential mutation-based or typing errors both at the bench and during analysis. With a known or presumptive typing error rate, each single-locus likelihood ratio may be corrected as described herein [see also SanCristobal and Chevalet, 1997; Garant et al., 2001]. Although there was no evidence of null allelism in tests involving known red drum specimens, explicit computations (Tringali, 2005) and indirect approaches (e.g., PROBMAX, Version 1.2; Danzmann, 1997) allow for treatment of null alleles.

Because probability statements are used in this approach, a decision must be made regarding a critical value of Pr(CP) that will be accepted as sufficient for the positive classification of compatible individuals. For red drum specimens from fishery-independent monitoring, a threshold of 99.9% has been adopted—i.e., when tested individuals are assigned to hatchery parent pairs, we wish to be correct in 999 of 1000 of those assignments. Having so identified ~1800 hatchery-compatible fish to date, we can reasonably expect that approximately two false-positive errors may have occurred. No false-positive erroneous classifications were observed among the 100 known wild specimens in the blind test. Whereas this limited analysis lacked power to address the false-positive error rate robustly, it was useful in a broad sense in demonstrating the capabilities of genetic tracking.

4.4. Concluding remarks

In conclusion, even when multiple breeding pairs are used and many individuals are tested, Bayesian posterior assignment probabilities for coparentage can be sufficiently high as to be robust for genetic tracking provided that a suitable number of markers are employed. Whereas false-positive errors then become unlikely, false-negative error rates will be largely governed by quality control in individual programs. The approach described here is expected to be especially useful when the relative contribution of hatchery offspring in the admixture is high and the number of hatchery breeding pairs is low relative to N_{HW}. When hatchery offspring account for 100% of the fish in the system, the approach becomes analogous with a closed-population maximum-likelihood approach where all possible parents have been genotyped. Unlike existing open-population maximum-likelihood methods (reviewed in Jones and Ardren (2003)), the Bayesian approach is relatively insensitive to estimates of N_{HW} given reasonable genetic evidence. However, it reduces to those methods, which are based implicitly on an uninformed prior, when released fish are no more or less likely than wild fish to be captured. Depending on testing conditions, statistical gains in precision made by considering all information within a Bayesian framework may allow investigators to reduce the number of markers used after a period of time, reducing expense and the opportunity for mistyping.

Acknowledgements

I am very grateful to Dr. Charles Brenner for advice regarding construction of the probability for coparentage and multi-test priors. Jamie Sullivan and Michelle Davis ably conducted genotyping assays during the blind-tests. I thank current and former members of the FWRI Molecular Genetics Lab and all other collaborators in the red drum stocking program. The manuscript benefited from comments by Jim Quinn, Judy Leiby, and two anonymous reviewers. Financial support was provided by the State of Florida and the Federal Sport Fish Restoration Act, Department of the Interior, U.S. Fish and Wildlife Service, Grant number F-69, Study II.

Appendix 1

Marshall et al. (1998) described a variable, e, defined as the probability of an error in genotyping (scoring error, mutation, or null allele). I use this variable here to derive a term for the likelihood ratio L_{CP} which incorporates an estimated or presumptive error rate. For a tested individual i, possible mother m, and possible father p, the following error types can occur at locus k: no typing errors in any of the three genotypes [probability $(1-e)^3$], one typing error in either i, m, or p [probability $e(1-e)^2$], two typing errors, i.e., in i and m, i and p, or m and p [probability $e^2(1-e)$], or three typing errors, i.e., one in i, m, and p [probability e^3].

Again defining $H1$ as the hypothesis corresponding to the probability of observing G_i assuming G_m and G_p are from the actual parents, and incorporating the error term as described in Marshall et al. (1998), X'_{CP} (hereafter, a prime superscript indicates inclusion of the error term) can be expressed as

$$\begin{aligned} X'_{CP} &= \Pr(G_m, G_p, G_i | H1, I) \\ &= [(1-e)^3 \tau(G_i|G_m, G_p)\delta(G_m)\delta(G_p)] \\ &\quad + [e(1-e)^2(\tau(G_i|G_m)\delta(G_m) \end{aligned}$$

$$+\tau(G_i|G_p)\delta(G_p)+\delta(G_m)\delta(G_p))]$$
$$+[e^2(1-e)(\delta(G_p)+\delta(G_m)+\delta(G_i))]+e^3, \quad (A.1)$$

where $\tau(G_i|G_m)$ is the Mendelian segregation probability of observing the individual's genotype given the potential mother's genotype and $\tau(G_i|G_p)$ is the Mendelian segregation probability of observing the individual's genotype given the potential father's genotype. Again defining $H2$ as the hypothesis corresponding to the probability of observing G_i when the actual mother and father are random, unrelated breeders, Y'_{CP} can be expressed as

$$\begin{aligned} Y'_{\mathrm{CP}} &= \Pr(G_m, G_p, G_i|H2, I) \\ &= [(1-e)^3(\delta G_i)\delta(G_m)\delta(G_p)] \\ &\quad +[e(1-e)^2((\delta G_i)\delta(G_m) \\ &\quad +\delta(G_i)\delta(G_p)+\delta(G_m)\delta(G_p))] \\ &\quad +[e^2(1-e)(\delta(G_p)+\delta(G_m)+\delta(G_i))]+e^3. \end{aligned} \quad (A.2)$$

These two terms may be substituted into Eqs. (3) and (4), respectively.

Fatherless maternity index—In this case, for a tested individual i and its mother m, the following error types can occur at locus k: no typing errors in either of the two genotypes [probability $(1-e)^2$], one typing error in either i or m [probability $e(1-e)$], two typing errors, i.e., in both i and m [probability e^2]. Let $X_{\mathrm{M}} = \Pr\{$observing the tested individual's genotype when the father's genotype is unknown and the putative mother is the actual mother$\}$; let $Y_{\mathrm{M}} = \Pr\{$observing the tested individual's genotype when the father's genotype is unknown and the actual mother is a random, unrelated female from the relevant population$\}$. For each locus, the single-locus likelihood ratio for fatherless maternity (L_{M}) is then taken to be the quotient $X_{\mathrm{M}}/Y_{\mathrm{M}}$. Now identifying $H1$ as the hypothesis corresponding to the probability of observing G_i when the putative mother is the actual mother, X'_{M} can be expressed as

$$\begin{aligned} X'_{\mathrm{M}} &= \Pr(G_m, G_i|H1, I) \\ &= [(1-e)^2\tau(G_i|G_m)\delta(G_m)] \\ &\quad +[e(1-e)(\delta(G_m)+\delta(G_i))]+e^2. \end{aligned} \quad (A.3)$$

Identifying $H2$ as the hypothesis corresponding to the probability of observing G_i when the actual mother is a random, unrelated female, Y'_{M} can be expressed as

$$\begin{aligned} Y'_{\mathrm{M}} &= \Pr(G_m, G_i|H2, I) \\ &= [(1-e)^2\delta(G_i)\delta(G_m)] \\ &\quad +[e(1-e)(\delta(G_m)+\delta(G_i))]+e^2. \end{aligned} \quad (A.4)$$

The cumulative fatherless maternity index becomes $\mathrm{MI} = \prod_{i=1}^{k} L_{M,i}$ for k loci.

Appendix 2

Below, a numerical example for the computation of the coparentage index CPI is provided based on the following information. The putative maternal genotype for locus *Soc*49

Locus	G_m	G_p	G_i	f_A	$\mathrm{LRF_{CP}}$	L_{CP}
*Soc*49	217	227	217	0.126	1/8*pq*	10.783
	239	229	227	0.092		
*Soc*83	*126*	*122*	*122*	0.229	1/2*pq*	15.361
	126	*122*	*126*	0.121		
*Soc*85	*95*	*99*	*99*	0.303	1/8*pq*	4.635
	105	*105*	*105*	0.089		
*Soc*99	*163*	*173*	*163*	0.029	1/8*pq*	26.124
	173	*197*	*173*	0.165		
*Soc*129	*129*	*121*	*121*	0.039	1/8*pq*	20.678
	142	*142*	*142*	0.155		
*Soc*133	*199*	*199*	*199*	0.674	$1/q^2$	2.201
	199	*199*	*199*	0.674		
*Soc*243	*91*	*88*	*91*	0.163	1/8*pq*	2.799
	97	*94*	*94*	0.274		
*Soc*276	*103*	*111*	*103*	0.327	1/4*pq*	2.282
	111	*111*	*111*	0.335		

corresponds generically to a maternal ***ps*** genotype and the putative paternal genotype corresponds to a paternal ***qr*** genotype. Thus, the genotype of the tested individual may be taken as ***pq*** and the applicable likelihood-ratio formula is 1/8*pq* (Table 1). Given the population frequencies, f_A, for alleles ***p*** and ***q***, $L_{\mathrm{CP}} = 10.783$. This process is repeated for the remaining loci. Then, from Eq. (6), $\mathrm{CPI} = \prod_{i=1}^{k} L_{\mathrm{CP},i}$ for k loci, the cumulative likelihood ratio CPI = 5,830,363.337. Computation of Pr(CP) for this specimen requires specification of priors. In this case, the compatible breeding pair came from spawning group 03–7 (Table 3), for which $\Pr(\mathrm{CP}|S, I_{\mathrm{CP}}) = 0.01028$. Further, 11% of the specimens tested prior to this individual were assigned to one of the possible breeding pairs ($\Pr(S|I_{\mathrm{S}}) = 0.11$). Thus from Eq. (6), Pr(CP) = 0.99985. Recalling that the odds for two events are the ratio of their individual probabilities and that $\Pr(H2) = 1 - \Pr(H1)$, it follows that the posterior odds against false-positive identification for this specimen were 6600 to one, and it was concluded that this specimen was likely derived from the pair of parents analyzed.

References

Amos, W., Sawcer, S.J., Feates, R., Rubinsztein, D.C., 1996. Microsatellites show mutational bias and heterozygote instability. Nat. Genet. 13, 390–391.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. Phil. Trans. R. Soc. 53, 370–418 (Reprinted in: Studies in the history of probability and statistics. IX. Thomas Bayes' essay towards solving a problem in the doctrine of chances. Biometrika 45, 293-315).

Bernatchez, L., Duchesne, P., 2000. Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? Can. J. Fish. Aquat. Sci. 57, 1–12.

Brenner, C.H., 1997. Symbolic kinship program. Genetics 145, 535–542.

Crozier, W.W., Moffett, I.J.J., 1995. Application of low frequency genetic marking at *GPI*-3 and *MDH-B*1, 2 loci to access supplementary stocking of Atlantic salmon, *Salmo salar* L., in a northern Irish stream. Fish. Manage. Ecol. 2, 27–36.

Dallas, J.F., 1992. Estimation of microsatellite mutation rate in recombinant inbred strains of mice. Mamm. Genome 3, 452–456.

Danzmann, R.G., 1997. PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. J. Hered. 88, 333.

Dickey, J.M., 1973. Scientific reporting. J. R. Statist. Soc. Ser. B 35, 285–305.

Drawbridge, M.A., 2002. The role of aquaculture in the restoration of coastal fisheries. In: Costa-Pierce, B.A. (Ed.), Ecological Aquaculture: The Evolution of the Blue Revolution. Blackwell Sciences, Ltd., Oxford, UK, pp. 314–336.

Edwards, A.W.F., 1992. Likelihood: Expanded Edition. John Hopkins University Press, London, UK, 275 pp.

Eldridge, W.H., Bacigalupi, M.D., Adelman, I.R., Miller, L.M., Kapuscinski, A.R., 2002. Determination of relative survival of two stocked walleye populations and resident natural-origin fish by microsatellite DNA parentage assignment. Can. J. Fish. Aquat. Sci. 59, 282–290.

Evett, I.W., Weir, B.S., 1998. Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinauer Associates, Inc., Sunderland, MA, 278 pp.

Garant, D., Dodson, J.J., Bernatchez, P., 2001. A genetic evaluation of mating system and determinants of individual reproductive success in Atlantic salmon (*Salmo salar* L.). J. Hered. 92 (2), 137–145.

Garnier-Gere, P., Dillmann, C., 1992. A computer program for testing pairwise linkage disequilibria in subdivided populations. J. Hered. 83, 239.

Gharett, A.J., Seeb, J.E., 1990. Practical and theoretical guidelines for genetically marking fish populations. Am. Fish. Soc. Symp. 7, 407–417.

Guo, S.W., Thompson, E.A., 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48, 359.

Guy, C.S., Blankenship, L.H., Nielsen, L.A., 1996. Tagging and marking. In: Murphy, B.R., Willis, D. (Eds.), Fisheries Techniques, 2nd ed. American Fisheries Society, Bethesda, MA, pp. 353–383.

Hansen, M.M., Hynes, R.A., Loeschcke, V., Rasmussen, G., 1995. Assessment of the stocked or wild origin of anadromous brown trout (*Salmo trutta* L.) in a Danish river system, using mitochondrial DNA RFLP analysis. Mol. Ecol. 4, 189–198.

Hansen, M.M., Ruzzante, D.E., Nielsen, E.E., Mensberg, K.-L.D., 2001. Microsatellite and mitochondrial DNA polymorphism reveals life-history dependent interbreeding between hatchery and wild brown trout (*Salmo trutta* L.). Mol. Ecol. 9, 583–594.

Henderson, S.T., Petes, T.D., 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell Biol. 12, 2749–2757.

Jones, A.G., Ardren, W.R., 2003. Methods of parentage analysis in natural populations. Mol. Ecol. 12, 2511–2523.

Kamonrat, W., 1996. Spatial genetic structure of Thai silver barb *Puntius gonionotus* (Bleeker) populations in Thailand. Ph.D. Dissertation. Dalhousie University, Halifax, NS, Canada 193 pp.

Kent, D.B., Drawbridge, M.A., Ford, R.F., 1995. Accomplishments and roadblocks of a marine stock enhancement program for white seabass in California. Am. Fish. Soc. Symp. 15, 492–498.

Kincaid, H.L., 1995. An evaluation of inbreeding and effective population size in salmonid broodstocks in federal and state hatcheries. In: Schramm Jr., H.L., Piper, R.G. (Eds.), Uses and Effects of Cultured Fishes in Aquatic Ecosystems. Am. Fish. Soc. Symp. 15, pp. 193–204.

King, T.L., Ward, R., Blandon, I.R., 1993. Gene marking: a viable assessment method. Fisheries (Bethesda) 18, 4–5.

Koskinen, M.T., Sundell, P., Piironen, J., Primmer, C.R., 2002. Genetic assessment of spatiotemporal evolutionary relationships and stocking effects in grayling (*Thymallus thymallus*, Salmonidae). Ecol. Lett. 5, 193–205.

Leber, K.M., 1999. Rationale for an experimental approach to stock enhancement. In: Howell, B.R., Moksness, E., Svåsand, T. (Eds.), Stock Enhancement and Sea Ranching. Fishing News Books/Blackwell Science, Ltd., Osney Mead/Oxford, UK, pp. 63–75.

Leber, K.M., Lee, C.-S., 1997. Marine stock-enhancement potential with striped mullet, *Mugil cephalus*, in Hawaii. Bull. Nat. Res. Inst. Aquacult. 3, 117–134.

Letcher, B.H., King, T.L., 1999. Targeted stock identification using multilocus genotype 'familyprinting'. Fish. Res. 43, 99–111.

Marshall, T.C., Slate, J., Kruk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7, 639–655.

McMichael Jr., R.H., 2000. Fisheries independent monitoring program operating procedure manual. Rep. No. IHR2000-005. Florida Marine Research Institute, St. Petersburg, FL, 256 pp.

Meagher, T.R., 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. 1. Identification of most-likely male parents. Am. Nat. 128, 199–215.

Meagher, T.R., Thompson, E., 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor. Pop. Biol. 29, 87–106.

Millar, R.B., 1987. Maximum likelihood estimation of mixed stock fishery composition. Can. J. Fish. Aquat. Sci. 44, 583–590.

Murphy, B.R., Nielsen, L.A., Turner, B.J., 1983. Use of genetic tags to evaluate stocking success for reservoir walleyes, *Stizostedium vitreum vitreum*. Trans. Am. Fish. Soc. 112, 457–463.

Murphy, M.D., Taylor, R.G., 1990. Reproduction, growth, and mortality of red drum *Sciaenops occelatus* in Florida waters. Fish. Bull. U.S. 88, 531–542.

Norris, A.T., Bradley, D.G., Cunningham, E.P., 1999. Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. Aquaculture 180, 247–264.

O'Reilly, P., Hamilton, L., McConnell, S., Wright, J.M., 1996. Rapid detection of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and trinucleotide microsatellites. Can. J. Fish. Aquat. Sci. 53, 2292–2298.

Pella, J., Masuda, M., 2001. Bayesian methods for analysis of stock mixtures from genetic characters. Fish. Bull. 99, 151–167.

Perez-Enriquez, R., Nobuhiko, T., 1999. Use of microsatellite DNA as genetic tags for the assessment of a stock enhancement program of red sea bream. Fish. Sci. (Tokyo) 65, 374–379.

Raymond, M., Rousset, F., 1995. GENEPOP (Version 1.2): a population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.

Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223–225.

SanCristobal, M., Chevalet, C., 1997. Error tolerant parent identification from a finite set of individuals. Genet. Res. 70, 53–62.

Taggart, J.B., Ferguson, A., 1986. Electrophoretic evaluation of a supplemental stocking program for brown trout (*Salmo trutta*). Aquacult. Fish. Manage. 17, 155–162.

Tessier, N.T., Bernatchez, L.B., Wright, J.M., 1997. Population structure and impact of supportive breeding program inferred from mitochondrial DNA and microsatellite DNA analyses in sympatric landlocked Atlantic salmon populations (*Salmo salar* L.). Mol. Ecol. 6, 750–753.

Tringali, M.D., 2003. Genetic changes in natural populations caused by the release of cultured fishes. Ph.D. Dissertation. University of South Florida, Tampa, FL, 228 pp.

Tringali, M.D., 2005. Derivation of likelihood-based co-parentage probabilities for familyprinting. IHR2004-17, Florida Fish and Wildlife Research Institute, St. Petersburg, FL, 16 pp. http://research.myfwc.com/publications/publication_info.asp?id=49433.

Tringali, M.D., Sullivan, J.G., Davis, M., 2004. Family-size variance and effective number of breeders in a captive-bred Sciaenops ocellatus. Florida Fish and Wildlife Research Institute. Unpublished.

Turner, T.E., Richardson, L.R., Gold, J.R., 1998. Polymorphic microsatellite DNA markers in red drum (*Sciaenops ocellatus*). Mol. Ecol. 7, 1771–1788.

Voss, P., Hogers, R., Bleeker, M., et al., 1995. AFLP: a new technique for DNA fingerprinting. Nucl. Acid Res. 23, 4407–4414.

Weber, J.L., Wong, C., 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. 2, 1123–1128.

Williams, G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acid Res. 18, 6531–6535.

Wilson, J.A., Langton, R.W., Van Orsdel, C., 1998. A model for the preliminary analysis of the economic feasibility of Atlantic cod enhancement in the Gulf of Maine (USA). Bull. Mar. Sci. 62 (2), 675–687.

Wilson Jr., R.R., Donaldson, K.A., Frischer, M.E., Young, T.B., 1997. Mitochondrial DNA control region of common snook for use as a genetic tag. Trans. Am. Fish. Soc. 126, 594–606.

Wright, J.M., 1993. DNA fingerprinting of fishes. In: Hochachka, P.W., Mommsen, T.P. (Eds.), Molecular Biology of Fishes, vol. 2. Elsevier, New York, NY, pp. 57–91.

Wright, J.M., Bentzen, P., 1994. Microsatellites: genetic markers for the future. Rev. Fish Biol. Fish. 4, 384–388.

Zhivotovsky, L.A., Feldmen, M.W., 1995. Microsatellite variability and genetic distances. Proc. Natl. Acad. Sci. U.S.A. 92, 11549–11552.