

TOWARD A MORE EFFICIENT APPROACH TO GENETIC TAGGING

Michael D. Tringali

Florida Fish and Wildlife Conservation Commission
Fish and Wildlife Research Institute
100 8th Avenue SE, St. Petersburg, Florida 33701 USA
mike.tringali@myfwc.com

In all stock enhancement programs, regardless of scale, at least some of the released animals should be tracked (recaptured and identified) to evaluate the success of the program. Often, marking these animals with physical tags can be impractical. Recently, genetic methods based on a concept called ‘familyprinting’ have advanced sufficiently to allow identification of parent/offspring relationships in a genetically homogeneous (unstructured) population. Unfortunately, no theoretic framework has been proposed for evaluating levels of statistical confidence in parental-pair identifications on a case-by-case basis. Investigators have instead relied on *post hoc* simulations to estimate the group-wise power of their particular loci to correctly include or exclude parentage. Such simulations are likely to 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.

Can the lack of realism in current parentage-assignment techniques blur statistical confidence in post-release identifications? Absolutely. In most cases, 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. If so, likelihoods of recapturing offspring of hatchery breeding pairs in hatchery/wild admixtures will be much higher than standard simulations would predict.

To address this problem, I used a Bayesian framework to develop a parentage-based method of identifying individuals produced in captivity and released into wild populations¹. With this approach, probabilities of correctly assigning *coparentage* (mother and father) can be computed directly for each tested individual. When the probabilities are appropriately conditioned, the need for *post hoc* power estimation is circumvented and issues involving family structure (differential hatchery/wild reproductive success) are addressed. 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, numerical analyses have shown that 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*.

To illustrate the Bayesian method, we can consider 7,756 red drum that were sampled in Tampa Bay between November 2003 and June 2005, as part of a stock enhancement study. Using data from 8 microsatellite loci, genotypes from 687 of these red drum were found to be compatible with the genotypes of hatchery breeding pairs. The remaining 7,069 were not

compatible with any hatchery breeding pair and were presumed to be products of natural recruitment. Using appropriate priors, posterior probabilities of coparentage, $\Pr(CP)$, were computed. To facilitate graphical representation, observed values were converted to posterior odds and transformed using the natural logarithm (Figure 1). The empirical distribution of posterior probabilities of coparentage for the 687 compatible individuals ranged from 0.983 to 0.99999999829984. The mean of the distribution was 0.999994. For the red drum program, we often consider a threshold at 0.999. For probabilities that exceed this threshold, we can be reasonably certain that at least 999 of 1000 of our assignments will be accurate. The majority of the probability mass for $\Pr(CP)$ exceeded 0.999 (shaded area in Fig. 1). Overall, the data indicate that sufficient genetic evidence is being obtained for highly accurate post-release identification of red drum. To date, fin clips from >20,000 field-caught red drum have been tested genetically. More than 2,470 have been identified as progeny of specific breeding pairs from the stocking program. 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.

To illustrate the effect that family-size differences between hatchery and wild breeding pairs may have on statistical precision, posterior distributions for the above specimens were constructed based on uninformed priors (Figure 1). This is analogous to the post-hoc simulation approach of standard likelihood. In all cases, coparentage probabilities values based on uninformed priors were lower than those based on informed priors, often considerably so. The majority of values were lower than the critical value of 0.999.

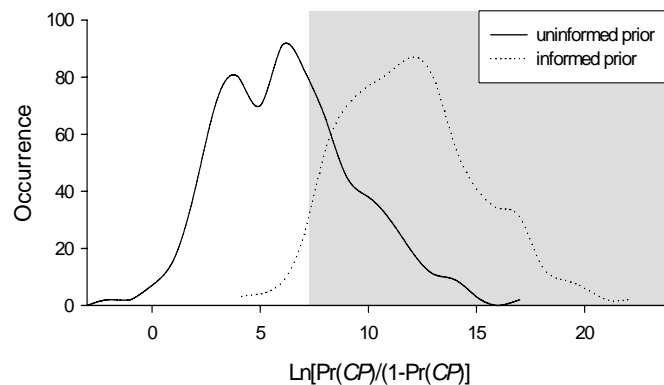


Figure 1. Distribution of Bayesian-informed and uninformed posterior probabilities of coparentage.

Conclusion – Even when many 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 perform better than standard analytical approaches when the relative contribution of hatchery offspring in the admixture is high and the number of hatchery breeding pairs is low relative to the number of breeding pairs in the wild. 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. It reduces to open-population 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 should allow investigators to optimize the number of markers used, reducing expense and the opportunity for mistyping.

¹ Tringali, M.D. 2006. A Bayesian approach for the genetic tracking of cultured and released individuals. *Fisheries Research* 77(2):159-172.