



# Estimation of parentage and relatedness in the polyploid white sturgeon (*Acipenser transmontanus*) using a dominant marker approach for duplicated microsatellite loci

Jeff A. Rodzen<sup>a,\*</sup>, Thomas R. Famula<sup>b</sup>, Bernie May<sup>b</sup>

<sup>a</sup> California Department of Fish and Game, Wildlife Forensics Laboratory, 1701 Nimbus Rd., Ste D, Rancho Cordova, CA 95670, USA

<sup>b</sup> Department of Animal Science, Meyer Hall, University of California, Davis, One Shields Ave., Davis, CA 95616, USA

Received 8 November 2002; received in revised form 19 June 2003; accepted 19 June 2003

## Abstract

We investigated the usefulness of microsatellite loci in the white sturgeon, *Acipenser transmontanus*, for estimating parentage and relatedness using algorithms developed for dominant markers. Microsatellite alleles were scored as independent dominant markers, and the resultant data were used with likelihood ratio statistics to estimate parentage. We tested three parentage scenarios: (1) assignment to parent pairs, (2) assignment to sires and dams independently when all possible parents were included (three sires and three dams), and (3) assignment to sires and dams when the true sires and dams were combined with a broodstock of 157 fish. Accuracy of assignment to parent-pairs exceeded 99% (scenario 1). Assignment accuracy to sires and dams independently was lower (scenario 2) but we present methods to improve this accuracy even when the number of possible sires and dams is large (scenario 3). Using LOD scores in conjunction with  $\delta$  values, a statistic quantifying confidence in the LOD scores, accuracy was maintained at a high level even when the number of sires and dams is large ( $n=90$  and  $73$ , respectively). Pairwise relatedness estimates were also made for fish from known sib groups to test if the correct sib groups could be identified using the dominant marker approach. The use of unweighted pair-group method using arithmetic mean (UPGMA) clustering algorithm to identify sib groups from a matrix of relatedness values proved to be successful for identifying all full-sib groups. The results suggest our approach of scoring

\* Corresponding author. Tel.: +1-916-358-2815; fax: +1-916-358-2817.

E-mail address: [jrodzen@cal.net](mailto:jrodzen@cal.net) (J.A. Rodzen).

microsatellite alleles as dominant markers can be used to accurately estimate parentage and relatedness.

© 2004 Elsevier B.V. All rights reserved.

*Keywords:* *Acipenser transmontanus*; White sturgeon; Microsatellite; Parentage; Relatedness; Dominant marker

---

## 1. Introduction

The white sturgeon is grown commercially in the Central Valley of California (USA), primarily by Stolt Sea Farm California LLC, a company that has pioneered the commercial culture of this species over the past two decades for meat and caviar and is the largest producer in the US of juvenile white sturgeon. The establishment of a breeding program for this species is complicated by the nature of their life cycle where sexual maturation occurs at about 7 years of age for females and 3 years for males, thus yielding many overlapping year classes and the potential for accidental parent–offspring matings. Each year class is typically made from three to five separate mating events using one to three females, each yielding 150,000 or more eggs, crossed with four to eight males. The different sib groups are then pooled together either as early post-hatch larvae or at the first size grading. The result is that any pedigree information is quickly lost and the producers have no method for estimating the number of sib groups and genetic diversity within a year class.

Genetic improvement of aquaculture stocks has been the subject of ongoing research as producers seek more efficient ways to produce animals and maximize profits. Paralleling efforts made by commercial growers in other agricultural sectors, substantial gains in efficiency have been made through traditional animal breeding. Such gains in production efficiency can be attained by selecting traits with definitive, measurable economic effects. Methods to achieve optimal gains in genetic merit over time require the use of proper genetic evaluation methods. Given the reported heritabilities of traits important to aquaculture, most authors agree the best method to achieve genetic gain will be a combination of individual and family selection (e.g. Gjedrem, 1997, 1983; Gall, 1990; Refstie, 1990). Hatchery facilities for genetic evaluation of many separate families are limited in commercial production, where only a small number of families can be reared. Physical marking technology limitations often mean families must be reared separately for sib groups to be positively identified. This limitation confounds genetic merit with environmental effects in each family (Knibb, 2000; Ferguson and Danzmann, 1998; Gjoen and Bentsen, 1997; Refstie, 1990; Gjedrem, 1983).

As mentioned above, an important component of a breeding program is pedigree information on all broodstock. Pedigree information is necessary for the estimation of breeding values and the management of inbreeding. Inbreeding can be problematic in any closed breeding population because it can have notable deleterious effects on the animals (e.g. Kincaid, 1983, 1976). If the breeder has knowledge of the genetic relationships between individuals selected for broodstock, then inadvertent mating of close relatives can be avoided. This problem may be more prevalent than is immediately obvious since individuals selected as broodstock based on superior phenotypes are more likely to be

genetically related than if they were selected at random. Since genetically related animals have a phenotypic correlation, phenotypic mass selection should select more groups of related individuals than purely random selection. Examples of this phenotypic correlation between relatives are (excluding common environmental, maternal, and epistatic effects):  $1/4 \sigma_A^2$  for half sibs and  $1/2 \sigma_A^2 + 1/4 \sigma_D^2$  for full sibs, with  $\sigma_A^2$  being the additive genetic variance and  $\sigma_D^2$  being the dominance genetic variance, with such information being easily referenced in animal breeding texts.

Genetic markers may offer solutions to many of these problems. They allow families to be mixed into common environments at hatching and minimize confounding of environmental effects with genetic effects. Pooling families into the same tank also maximizes the number of families available for evaluation, thus increasing the potential selection differential while allowing a combination of individual and family selection. The use of genetic markers for parentage testing and pedigree reconstruction in aquaculture situations has also been suggested by many authors (e.g. Hara and Sekino, 2003; Sekino et al., 2003; Cunningham, 1999; Ferguson and Danzmann, 1998; Gjoen and Bentsen, 1997).

Microsatellite loci have become popular as the optimal marker system for determining parentage in aquaculture broodstocks. They have been used successfully to reassign progeny from mixed pools to their parents in several species, including sea bass *Dicentrarchus labrax* (Garcia de Leon et al., 1998), turbot *Scophthalmus maximus* (Estoup et al., 1998), channel catfish *Ictalurus punctatus* (Waldbeiser and Wolters, 1999), rainbow trout *Oncorhynchus mykiss* (Estoup et al., 1998; Ferguson and Danzmann, 1998; Herbinger et al., 1995), and Atlantic salmon *Salmo salar* (Norris et al., 2000; O'Reilly et al., 1998). Many of these studies demonstrated over 90% correct reassignment to parent pairs when the number of possible parental pairs was over 100.

The application of microsatellite loci for estimating relatedness in aquaculture has received less attention to date than the estimation of parentage. Norris et al. (2000) successfully used 15 microsatellite loci in Atlantic salmon to estimate the genetic relatedness of individuals from a pool of full-sibs, half-sibs, and unrelated fish for which pedigree structure was known *a priori*. They report Type II error rates of 0.05 to 0.00 for estimating full-sibs to be unrelated and 0.22 to 0.01 for half-sibs as unrelated using the Kinship software package (Queller et al., 1993) to estimate relatedness. Kinship uses relatedness algorithms developed by Queller and Goodnight (1989) to estimate relatedness of individuals from codominant marker data.

Microsatellite loci are available for white sturgeon and are described in Rodzen and May (2002), and they recommend against scoring the loci as codominant loci, but as groups of dominant markers as detailed in Appendix A. Evidence exists that a large amount of genome duplication has occurred leading to the belief that the species is an ancestral octoploid (Blacklidge and Bidwell, 1993). Comparisons of DNA content and chromosome number across sturgeon species in the order Acipenseriformes worldwide show that genomes vary from 4N (e.g. Birstein et al., 1993, Fontana, 1976) to 16N (Birstein et al., 1993; Blacklidge and Bidwell, 1993), with most species appearing to be 4N (Birstein et al., 1997). The authors found evidence of disomic, tetrasomic, and higher levels of inheritance of codominant markers in this species, but concluded that the octoploid-derivative genome is too complex to analyze variation at a single locus.

## 2. Methods and materials

In this study, we test parentage estimation formulae derived for AFLP loci (Gerber et al., 2000). In addition, we investigate the utility of the Lynch and Milligan (1994) method for estimating relatedness of individuals in a mixed lot of nine families resulting from one  $3 \times 3$  mating design. Once a matrix of relatedness values was calculated, we tested the utility of two clustering algorithms to analyze relatedness estimates for identifying and correctly clustering full-sib and half-sib groups. Finally, we apply these methods in an existing commercial operation to estimate the number of full-sib groups comprising 4-year classes of white sturgeon.

### 2.1. PCR profiles and primer sequences

DNA was extracted from either fin tissue or whole sac fry as described in Rodzen and May (2002). PCR protocols, primer sequences, and modes of inheritance of the nine microsatellite loci are described in detail in Rodzen and May (2002). PCR primers used in this study were 5' fluorescein labeled. PCR products were separated using denaturing polyacrylamide gel electrophoresis (PAGE) with gels of 5% acrylamide and 7.5 M urea. Products were visualized using a Molecular Dynamics model 595 fluorimager (Molecular Dynamics, Sunnyvale, CA, USA).

### 2.2. Gel scoring and data management

Phenotypes were scored at each locus by loading an allelic ladder on the same gel. Scoring of phenotypes was performed using Molecular Dynamics' Fragment Analysis software package. True genotypes typically can be discerned for this octoploid species because of the high level of duplication (Rodzen and May, 2002). For instance, an individual amplified five bands at locus that yields up to eight bands cannot be assigned a true genotype that accurately accounts for gene dosage and heterozygosity. We refer to alleles as bands since we do not know the true allelic relationships at the microsatellite loci. One does not know if that individual has null alleles at that locus or if some bands are present in multiple doses to add up to eight total doses. Results of the automated scoring were manually checked for errors, particularly when gels ran unevenly. Because of the highly duplicated nature of the white sturgeon genome, we have previously recommended that each microsatellite band in this particular species be scored as an individual dominant locus (Rodzen and May, 2002). Phenotypes were converted to a series of dominant loci as detailed in Appendix A.

### 2.3. Experimental design

#### 2.3.1. Fish of known pedigree (standards)

Three sires and three dams from a commercial aquaculture broodstock were crossed factorially, yielding nine full-sib families and three paternal and maternal half-sib families. One full-sib family failed to develop, thus eight full-sib families were used in this study. Fifty fish per family were collected at approximately 24–48-h post-hatch. DNA from each

fish was extracted, amplified, and scored for the nine microsatellite loci (Rodzen and May, 2002) used in this study.

### 2.3.2. Year-class family structure

Fish were sampled from four different year classes at Stolt Sea Farm California LLC: 1992 ( $n=138$ ), 1993 ( $n=135$ ), 1995 ( $n=100$ ), and 1996 ( $n=100$ ), plus broodstock individuals ( $n=157$ ). The 1992 and 1993 year class fish each represent combined samples from two Stolt Sea Farm sources: those being processed for caviar ( $n=88$  for each year class) and those that were randomly sampled from one growout facility ( $n=50$  for each year class). All 1995 and 1996 year class fish were randomly sampled from the growout facility. The broodstock consisted of 157 living sires and dams, some of which have been previously spawned to create year classes. The 1992 and 1993 year classes were created by crossing wild-caught (Sacramento River delta, California, USA) dams with captive held males. All dams used to make the 1992 and 1993 year class were returned to the place of capture while the sires may have been sampled with the broodstock. A standard one-hole paper punch was used to remove three plugs of tissue from thin, fleshy sections of the caudal or pelvic fins. The tissue plugs were subsequently stored in storage buffer consisting of 10% ddH<sub>2</sub>O, 90% DMSO, saturated with NaCl.

### 2.3.3. Parentage

Methods for estimating parentage using codominant loci are well known (e.g. Evett and Weir, 1998; Weir, 1996). Given that we converted microsatellite bands to dominant loci, the algorithms derived for codominant loci were inappropriate. Instead, we used the likelihood ratio statistics developed for amplified fragment length polymorphism (AFLP) markers detailed in Gerber et al. (2000). To estimate the error rates of the likelihood statistics described in Gerber et al. (2000) given our marker system and numbers of possible parents, LOD scores were calculated using the formula:

$$\text{LOD score} = \ln \left( \frac{H_1: \text{animal } k \text{ is parent of animal } j}{H_0: \text{animal } k \text{ is not the parent of animal } j} \right)$$

The probability statements for  $H_1$  and  $H_0$  are described in Gerber et al. (2000) for calculating likelihoods for parent pairs and individual parents. Three different scenarios were evaluated. Scenario 1: we tested the ability of the microsatellite loci to assign progeny to a specific cross (a parent pair). Scenario 2: progeny from the controlled crosses were reassigned to sires and dams independently with no pairs assumed *a priori*. Scenario 3: progeny of the controlled crosses were assigned to sires and dams when molecular data on the sires and dams from the controlled crosses was combined with data from the 157 broodstock.

In the first scenario, we tested the ability of the microsatellite loci alone to reassign progeny from the  $3 \times 3$  cross back to a particular cross of sires and dams (a parent pair). Progeny were omitted from analysis if more than half of the microsatellite loci failed to amplify in that individual. This situation occurred in less than 20% of typed progeny and was consistent across families, probably due to extraction problems. Percent correct

assignment was calculated for each of the nine microsatellite loci by dividing the number of correct assignments to the parent pair by the number of offspring scored. Data were then merged from the three microsatellite loci with the highest individual percent correct assignment. Three loci were used since this is the number of loci that can currently be used in automated genotyping with three fluorochromes used for the three loci and a fourth fluorochrome used for an internal size standard.

The second scenario used only the sires and dams from the controlled crosses, but this time sires and dams were analyzed independently of one another. Individual microsatellite loci were not analyzed one at a time as in the first scenario, but data from all nine loci were merged before analysis, resulting in a total of 82 bands, or 82 dominant markers by our scoring method. The microsatellite loci were not analyzed separately since assignment to individual parents is substantially less powerful than scenario 1. For each offspring, LOD scores were calculated for each sire and then sires were ranked by LOD scores. Percent correct assignment was calculated by dividing the number of times the sire with the highest LOD score was the true sire by the total number of progeny tested. The procedure was repeated for dams.

In the third scenario, sires and dams were also analyzed independently. The molecular data from the sires and dams from the controlled crosses were merged with the data from the other 157 broodstock and the assignment of 350 progeny to the correct sire or correct dam was evaluated. Since the broodstock sampled represented approximately half of the broodstock that existed in 1992 when the first sampled year class was created, we had to establish a criterion for deciding whether the most likely parent (highest LOD score) was the actual parent. The use of delta ( $\delta$ ) to compare relative likelihoods of different animals being the true parent is described in [Marshall et al. \(1998\)](#):

$$\delta = (\text{LOD score of most probable parent}) \\ - (\text{LOD score of second most probable parent})$$

LOD scores and  $\delta$  values for these 350 fish were calculated and the number of correct and incorrect parental assignments was recorded. Delta ( $\delta$ ) is a measure of how much more likely a given fish is to be the true parent than the fish closest to it in rank, basically a statistic to quantify how much more confidence can be placed on the ranking of LOD scores of the possible parents. The determination of what critical value of  $\delta$  should be used had to be determined empirically using results from this study and is discussed in Results.

#### 2.3.4. Relatedness

The equations used to estimate relatedness between all pairwise combinations of offspring are described in [Lynch and Milligan \(1994\)](#). For  $n$  individuals, this yielded an  $n$  by  $n$  symmetrical matrix of relatedness estimates. The relationship matrix was then used as input to the PHYLIP package (v3.57c, Joseph Felsenstein, Department of Genetics, University of Washington), which contains the Neighbor and Fitch programs for estimating the structure of a dendrogram from a distance matrix. Neighbor uses both the unweighted pair-group method using arithmetic mean (UPGMA) and the

neighbor-joining algorithms to estimate the structure of and branch lengths of a dendrogram. Fitch (PHYLIP v3.57c) uses the Fitch–Margoliash and least squares methods to estimate the best dendrogram. UPGMA was used in a previous study (Blouin et al., 1996) to correctly identify sib groups from relatedness values. The relatedness values calculated in that study were from codominant microsatellite loci, unlike our approach of converting bands at codominant loci to dominant loci. Combinations of input options in both Neighbor and Fitch were tested to determine which options consistently produced a dendrogram where full sibs and half sibs clustered together. Both the UPGMA and neighbor-joining algorithms were used to construct dendrograms with Neighbor. Additionally, both Neighbor and Fitch were used with and without setting the options to randomize the input order of individuals in the data set.

Ten fish from each full sib group from the controlled crosses (total  $n=80$ ) were used to create three 80 by 80 matrices of estimated pairwise relatedness values (hereafter referred to as  $\mathbf{R}$ ), with three different combinations of microsatellite loci. Only individuals with no missing data were used. The first matrix was calculated using molecular data from the locus that gave the highest percent correct parent-pair assignment (*Atr-109*). The second matrix was calculated by merging molecular data from the three best loci for estimating parent-pair (*Atr-107*, *109*, *113*). Lastly, the third matrix was calculated by merging data from all nine microsatellite loci. Combining data from all nine loci yielded 82 markers. Since the PHYLIP programs required a distance matrix as input, each matrix  $\mathbf{R}$  was converted to  $1 - \mathbf{R}$  by subtracting each element from 1.

For each year class, UPGMA was performed on the relatedness values calculated from 128 microsatellite bands, the total number of scored bands from the amplified loci. While the nine families used for testing the accuracy of the methods yielded a total of 82 markers, we observed and scored a total of 128 markers across the year classes. A relatedness matrix was also calculated for the 80 fish with known genetic relationship used above to test the different clustering methods of UPGMA and neighbor-joining. A  $40 \times 40$  “standardization” matrix,  $\mathbf{S}$ , was made from pairwise relatedness estimates from one paternal half-sib group of two 20-fish full-sib groups. The matrix of relatedness values ( $\mathbf{R}$ ) for each year class was then diagonally concatenated with  $\mathbf{S}$  prior to performing UPGMA to yield a matrix arbitrarily called  $\mathbf{T}$ , with the remaining cells set to  $-0.25$  to ensure outgrouping of the standards during UPGMA. Like with  $\mathbf{R}$ , we used  $1 - \mathbf{T}$  to convert to a distance matrix prior to using UPGMA to construct a dendrogram.  $\mathbf{T}$  has the form:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R}_{n \times n} & -0.25_{n \times 40} \\ -0.25_{40 \times n} & \mathbf{S}_{40 \times 40} \end{bmatrix}$$

where  $\mathbf{R}$  is of dimensions  $n \times n$  and  $\mathbf{S}$  is of dimensions  $40 \times 40$ .

Thus,  $\mathbf{S}$  was included in the dendrogram along with the fish of unknown relatedness to create a standard for calling the presence of a full-sib group. The arbitrary value used in  $\mathbf{T}$



does affect how well **R** and **S** would be separated visually in the resultant dendrogram. We chose a value for **T** that was effective for keeping the standards (the values in **S**) separate from the values in **R** in the dendrogram. The reason for using **T**, instead of analyzing **R** and **S** separately, was to get one dendrogram that had both the unknown animals and the animals of known relationship but still visually separated these two groups. A vertical line was then drawn across the dendrogram at the maximum value of  $1 - R_{XY}$  where all the full

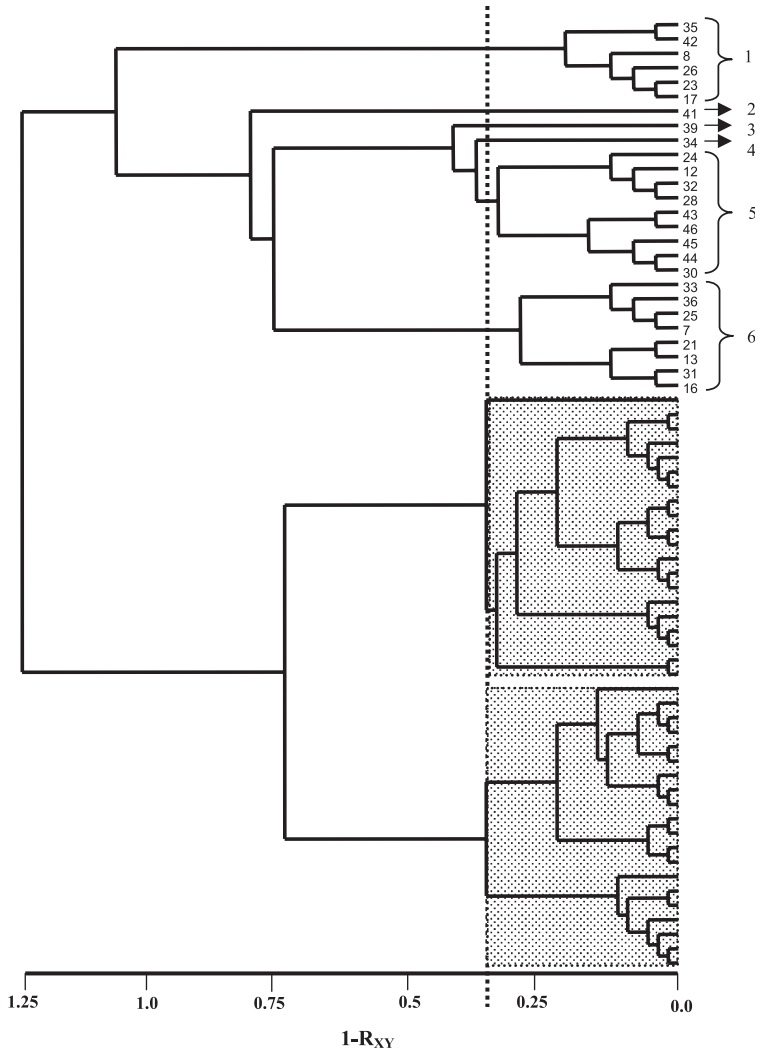


Fig. 1. Dendrogram using UPGMA of relatedness estimates for subset of 1992 Laguna fish. In grey are the two groups of full sibs (20 fish each) used to place the vertical line to estimate groups of full sibs. In this example, six full-sib groups are estimated to comprise the subset of animals of unknown relatedness. Bottom scale is  $1 - R_{XY}$  for each pairwise comparison between individuals.



sibs of the standards matrix coalesced. This is shown diagrammatically in Fig. 1. A subset of the 1992 Laguna fish is shown for illustration purposes.

#### 2.4. Combining relatedness and parentage information: establishing sib groups

Relatedness information and parentage estimates were used to estimate which fish within a year class actually formed a full-sib group. Relatedness information was used by employing the above approach of performing UPGMA on the matrix **T** that included fish of unknown relationship and **S**. A cluster of animals of unknown genetic relationship were declared to be a full-sib group if their cluster had a coalescent point less than the full-sib groups in **S** (see Fig. 1). For example, in Fig. 1, animals numbered 35, 42, 8, 26, 23, and 17 are estimated to comprise one full-sib group, as their coalescent point is less than that of the two full-sib groups shaded in grey (their coalescent point is to the right of the vertical line).

After the full-sib groupings had been estimated, the parentage estimates for the fish comprising a given sib group were examined. The five most likely sires for each fish in the 1992 and 1993 year classes, and five most likely sires and dams for each fish in the 1995 and 1996 year classes were used. Only sires were examined for the 1992 and 1993 year classes since the dams used in those years were returned to the wild after spawning. The decision to use five sires and five dams was arbitrary. Delta ( $\delta$ ) values were calculated for each of the most likely sires and dams. If the majority of the fish in the estimated sib group had a most likely sire or dam in common, it was concluded that the common sire or dam was in fact a true parent of that sib group. In many cases, no single sire or dam met the criterion set to assign it as the true sire or dam of that sib group. While this may appear an arbitrary value, it does correspond to the intermediate relatedness values in Fig. 2 separating full sibs and half sibs.

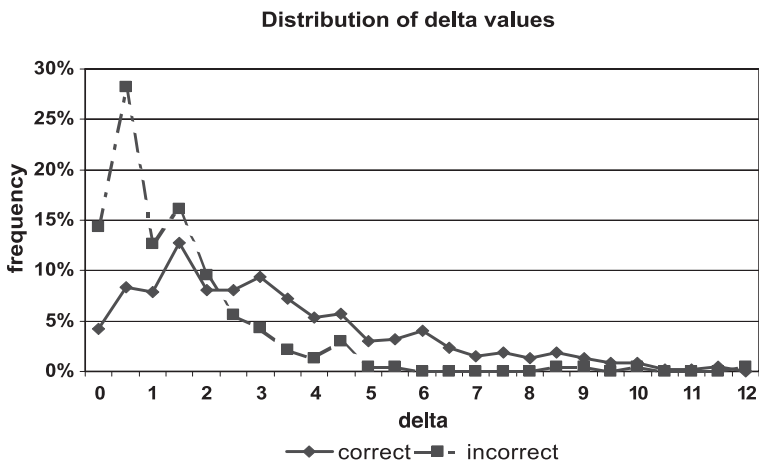


Fig. 2. Distribution of  $\delta$  values. Shown are the distributions for correct ( $n=470$ ) and incorrect ( $n=240$ ) assignments of sires and dams. The two distributions intersect at a  $\delta$  value of approximately 2.5.

### 3. Results

#### 3.1. Parentage assignment

The ability of the microsatellite loci to correctly assign parentage was evaluated in three scenarios. First, we examined the accuracy of assigning progeny to specific parent pairs from a  $3 \times 3$  cross. Percent correct parent-pair assignment ranged from 37% to 97% for single microsatellite loci (Table 1). When marker data from the three loci giving the highest individual percent correct assignment (*Atr*-107, 109, 113) were combined, the percentage increased to 99%.

Second, we examined the ability of the microsatellite loci to correctly identify individual parents. In this case, it is known that the actual sires and dams have been sampled and characterized from molecular data. There were 670 correct sire or dam assignments to 350 animals, yielding an accuracy of 670/700, or 96%. The third scenario was similar to the second in that we tested the accuracy of the microsatellite loci to correctly identify individual sires and dams. However, the marker data from the true parents was merged with 157 non-parent broodstock individuals (87 sires and 70 dams). For each progeny, the number of possible sires and dams was now 90 and 73, respectively. Under this scenario, there were 470 correct sire or dam assignments to 350 progeny, yielding an accuracy of 470/700, or 67%.

Table 1  
Parentage assignment power to parent pairs and independent sires and dams

Locus	Mode of inheritance	Scored bands	Correct progeny ( <i>n</i> )	Percent correct assignment	
<i>Scenario 1: assignment of progeny to nine possible parent-pairs</i>					
<i>Atr</i> -100	Tetrasomic	6	114/300	38	
<i>Atr</i> -105	Tetrasomic	5	115/311	37	
<i>Atr</i> -107	Hexa/octosomic	11	231/335	69	
<i>Atr</i> -109	Octosomic	19	318/328	97	
<i>Atr</i> -113	Unknown	10	254/330	77	
<i>Atr</i> -114	Octosomic	10	205/331	62	
<i>Atr</i> -117	Octosomic	12	128/291	44	
<i>Atr</i> -1101	Tetrasomic	5	154/321	48	
<i>Atr</i> -1173	Hexa/octosomic	10	168/316	53	
Three best <sup>a</sup>	N/A	40	306/310	99	
<i>Scenario 2: assignment of progeny to three sires and three dams independently without broodstock</i>					
All nine loci	N/A	82	670/700	96	
<i>Scenario 3: assignment of progeny to independent sires and dams with broodstock</i>					
All nine loci	N/A	128	470/700	67	without using $\delta$
			271/301	90	using $\delta > 2.5$

Shown is the accuracy of parentage assignment under three scenarios. The total number of bands in the parents is referred to as scored bands. Correct progeny is the number of progeny correctly reassigned to both sire and dam using their genotypes at the given locus. Mode of inheritance is from Rodzen and May (2002).

<sup>a</sup> Three best loci were *Atr*-109, 113, and 107.

Since the accuracy of the third scenario was low, we established a minimum  $\delta$  value criterion to decide if the most likely sire or dam was the true sire or dam. The distribution of  $\delta$  values is shown in Fig. 2. Correct  $\delta$  values are those for which the sire or dam with the highest LOD score was the correct sire or dam of that individual fish ( $n=470$ ). The distributions intersect at a  $\delta$  value of approximately 2.5. Using  $\delta$  values, the accuracy of the third scenario was recalculated. For each offspring, a sire or dam was considered the most likely sire or dam if it had the highest LOD score and a  $\delta$  value of at least 2.5. Three hundred and one offspring had sire or dam assignments with a  $\delta > 2.5$  (43%), and of these, 271 sire or dam assignments were correct. The accuracy of sire and dam assignments to those offspring increased to 90% using the restricted criteria.

### 3.2. Relatedness

Initially, we tested the utility of a single locus (*Atr-109*) to estimate relatedness. The means of estimated relatedness values and sample variances for pairs of individuals grouped by genetic relationship (full sib versus half sib versus unrelated) were also calculated and did not differ significantly from normality. As expected, full-sib pairs had the highest relatedness ( $\mu=0.42$ ,  $\sigma^2=0.102$ ), followed by half-sibs ( $\mu=0.12$ ,  $\sigma^2=0.090$ ), and unrelated pairs ( $\mu=-0.19$ ,  $\sigma^2=0.096$ ). However, using these relatedness values, neither Fitch nor Neighbor was able to accurately reassemble groups of full sibs into visually distinct groups in a dendrogram. Data from the three most variable microsatellite loci (*Atr-107*, 109, 113,  $n=40$  bands) were merged to estimate relatedness. While the accuracy of Fitch and Neighbor increased with data from three loci, the full-sib groups were not positively or unambiguously defined.

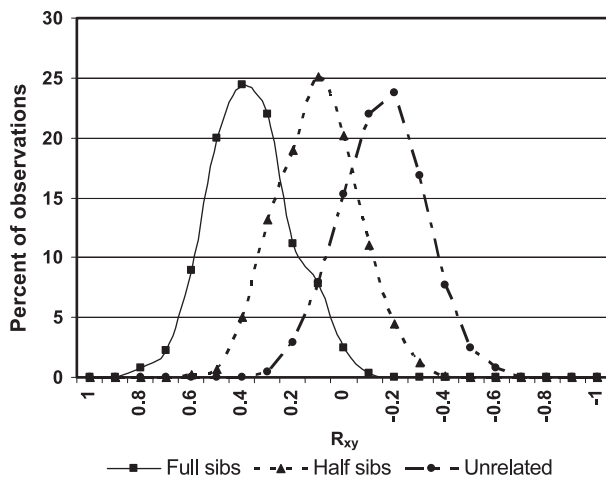


Fig. 3. Probability density of relatedness estimates given true genetic relationship. Densities reflect the distribution of pairwise relatedness estimates for individuals given the true genetic relationship of the pair. Data from all nine microsatellite systems with total of 82 bands was used to calculate relatedness values.

Data from all nine microsatellite loci were then merged yielding 82 dominant loci (bands). The probability densities for each class of genetic relationship are shown in Fig. 3. Means of estimated relatedness values were again similar to those for both a single microsatellite locus and the pool of three loci ( $\mu_{FS} = 0.38$ ,  $\mu_{HS} = 0.12$ ,  $\mu_{unrelated} = -0.14$ ). However, sample variances were much smaller ( $\sigma_{FS}^2 = 0.024$ ,  $\sigma_{HS}^2 = 0.025$ ,  $\sigma_{unrelated}^2 = 0.026$ ). The smaller variances were reflected in the ability of the clustering algorithms to accurately construct the dendrograms. Neighbor using UPGMA and randomized input

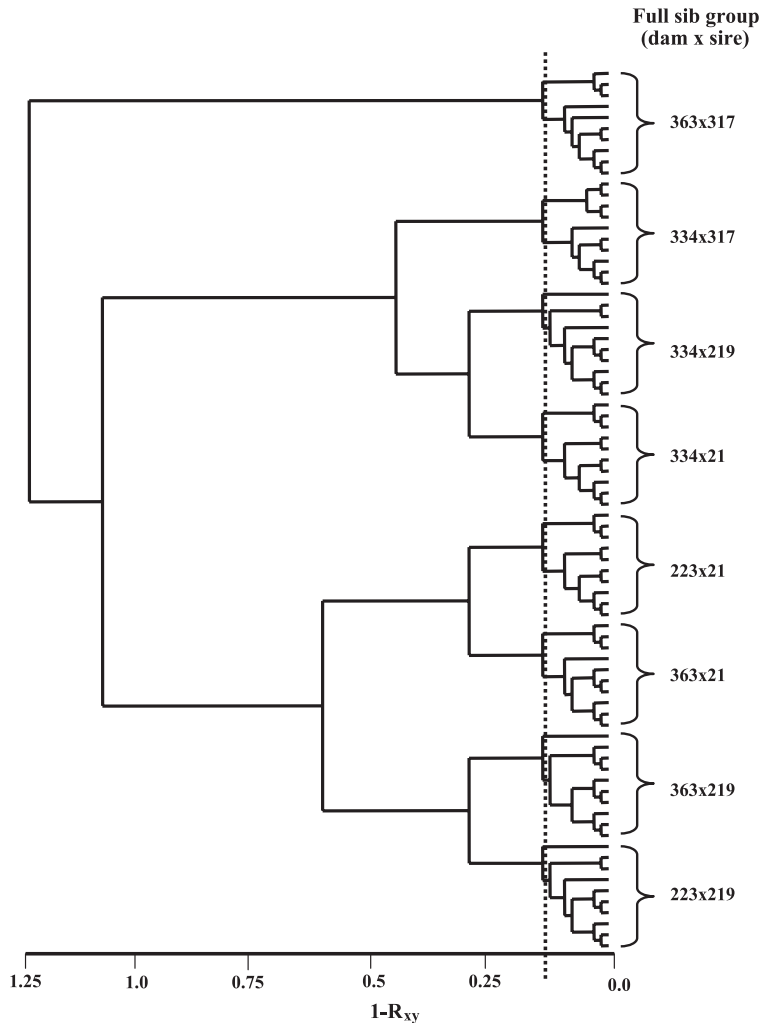


Fig. 4. Dendrogram using UPGMA of relatedness estimates from combining data from all nine loci. The eight full-sib groups, marked with brackets, formed easily identifiable clusters. The bottom scale is  $1 - R_{xy}$  for each pairwise comparison between individuals.

Table 2  
Fish sampled (*n*) and number of estimated full-sib groups for each year class

Year class	Fish sampled ( <i>n</i> )	Estimated FS groups
1992 caviar	88	14
1992 Laguna	50	10
1993 caviar	88	17
1993 Laguna	50	14
1995 Laguna	98	13
1996 Laguna	99	20
Broodstock sires	87	15
Broodstock dams	70	16

Sample sizes and estimated number of full-sib groups within each year class of fish, including the sampled sires and dams.

order of individuals yielded 100% correct grouping of full sibs into families (Fig. 4). Although Fig. 3 shows considerable overlap of distributions of pairwise relatedness estimates between the three classes of true genetic relationship, UPGMA successfully classed all the full-sib groups into separate groups. However, this method did not identify half-sib groups, which is likely the result of the factorial-mating scheme used. Each full-sib group member is equally likely to belong to two different half-sib groups, the maternal and paternal half-sib groups. This problem could likely be corrected if mtDNA markers were used in conjunction with nuclear markers.

Thus, reliable assignment into sib groups was obtained by using 82 microsatellite bands to calculate the relationship matrix **R**. This matrix was best analyzed by using UPGMA to construct dendrogram from  $1 - \mathbf{R}$  and randomizing the order of individuals in Neighbor, which correctly identified all groups of full sibs in our test data set.

### 3.3. Broodstock study

Parentage and relatedness estimates were calculated for each individual fish from the 4-year classes sampled. There was no general temporal trend in the number of full-sib groups, the number of full-sibs groups did not decline with the age of the year class. Results are presented in Table 2. The number of full-sib groups ranged from 10 (1992 Laguna) to 20 (1996 Laguna). There were also several full-sib groups of sires and dams in the 157 broodstock. The relatedness estimates and dendrograms constructed from UPGMA suggest the presence of 15 full-sib groups of sires and 16 full-sib groups of dams within the broodstock.

## 4. Discussion

The parentage results obtained in this study are comparable to those reported in current literature (Norris et al., 2000; Waldbeiser and Wolters, 1999; Estoup et al., 1998; Ferguson and Danzmann, 1998; Garcia de Leon et al., 1998; O'Reilly et al., 1998; Herbinger et al., 1995). Those studies used microsatellite loci scored traditionally as codominant indepen-

dent loci, contrary to our method of scoring microsatellite bands as dominant markers. Many of those studies demonstrated greater than a 90% assignment accuracy to parent-pairs in a variety of mating schemes and tested for correct assignment to hundreds of possible parental crosses, reflective of the mating schemes used in production of the respective species.

To contrast, the parent-pair test crosses used in this study was a three by three factorial cross, reflective of the mating scheme used in white sturgeon aquaculture. While we obtained a 99% correct reassignment rate to three sires and three dams with only three loci, we also demonstrated that a high percentage of parental reassignment in a mating system using several more sires and dams could be obtained if more loci are employed. This is similar to results in other reported studies. O'Reilly et al. (1998) observed a 99.5% correct assignment rate to 12 full-sib groups using four microsatellite loci, and the assignment rate fell to 80% when 144 possible crosses were tested using the same four loci. Herbing et al. (1995) observed a 91% assignment rate to 1 or 2 of 100 possible parental couples using only five loci. Ferguson and Danzmann (1998) obtained a 90% correct assignment rate to 135 possible families using 14 microsatellite loci. Literature regarding reassignment to individual independent parents is scarcer, which is likely reflective of the fact that most parentage studies have molecular data on all possible parents and the true crosses are known. We observed a 97% correct reassignment rate to sires or dams when the three sires and three dams from the test crosses were analyzed as independent parents.

While parentage testing has received much attention, estimation of relatedness in aquaculture populations has been largely untested. This may be the result of investigators taking advantage of the ability to obtain genetic samples from all possible parents. However, situations frequently arise where this is not possible, for instance, when a facility has to combine multiple families for which the parents were not sampled and no longer exist. Norris et al. (2000) examined this problem in Atlantic salmon by using *Kinship Queller et al. (1993)* to estimate relatedness between full sibs, half sibs, and unrelated individuals. Their results demonstrated 15 loci were needed to correctly classify pairs of individuals with a low ( $<0.05$ ) Type II error rate.

Our estimates of relatedness showed a correct regrouping of full sibs using a novel combination of mathematical approaches. Relatedness estimates were calculated from algorithms in Lynch and Milligan (1994). We then employed a phylogenetics inference tool, unweighted pair group method using arithmetic mean (UPGMA), to identify groups of individuals having similar relatedness values. This yielded 100% correct regrouping of full sibs in a dendrogram. This dendrogram did not reveal half-sib relationships in our factorial mating design and may only work when a multiple single-pair mating scheme is used. This is because the maternal half-sib groups and the paternal half-sib groups are confounded by one another and the expected value  $R_{xy}$  is the same for paternal and maternal half-sib groups. Mitochondrial data might alleviate this problem by first separating out the maternal half-sib groups.

It must be noted that the procedure we employed for estimating the number of sib groups must use animals of known pedigree *and* of similar genetic background (heterozygosity, allelic diversity, etc.) to those of unknown pedigree. This is because the calculated values of  $R$  are dependent on heterozygosity and the distribution of allele frequencies in the pool of animals for which  $R$  is being calculated. In addition, we did not

try any methods for constructing dendrograms other than the UPGMA and neighbor-joining of the matrix **T**. If someone is to use this approach, we recommend they also try using bootstrapping, permutations, or some sort of resampling of the **T** matrix in order to assign confidence intervals to the nodes on the dendrogram. One may also want to examine the effectiveness of using  $R_{xy}$  to identify parents as well in a mixed pool. In this study, we had a physically separate pool of parents and thus used parentage estimation, but one could also use  $R_{xy}$  if the parents and offspring might be mixed together.

The year class family structure revealed by UPGMA suggested that the number of full-sib groups comprising a year class has not declined significantly with increasing age of a year class. It might be expected that a year class would lose families over time due to disease outbreaks and random mortalities. Genetic diversity within the year classes appears to be well conserved, as the number of full-sib groups identified within each year class was consistent with the number of matings that produced that year class (Struffenegger, personal communication, Stolt Sea Farm California). Similarly, there is a large number of full-sib families comprising the current stock of sires and dams being retained for breeding purposes. Results from this study will be useful to the aquaculture production facility in that information is provided to indicate individual sires and dams that can be crossed from each year class to minimize the chance of an accidental backcrossing.

When selecting loci to use for estimating parentage and relatedness, it appears that those loci with the most number of alleles and alleles of similar frequencies would be the best choice, as is the case of individual identification and parentage exclusion. As is evident in [Table 1](#), the number of alleles is not necessarily correlated with the accuracy of parentage exclusion. Instead, one should account for both the number of alleles and their frequencies and estimate the accuracy of those individual loci for parentage identification or exclusion (methods detailed in [Evelt and Weir, 1998](#)). Also, computer software is available for disomic loci that will estimate the accuracy of parentage estimation from experimental data via simulation analyses (e.g. “Cervus” in [Marshall et al., 1998](#)). The actual genotypes of the parents of the known sib groups and segregation of alleles in the sib groups used for testing the accuracy of our parentage and relatedness estimation methods are presented in [Tables 3 and 5 of Rodzen and May \(2002\)](#), which provides extensive detail about the modes of inheritance of the loci used in this study. Given the duplicated nature of the white sturgeon genome and complex modes of inheritance of the loci used, it is difficult to provide a clear and consistent guideline as to selecting new loci to use in this species.

This study is unique in that microsatellite loci were not scored as typical codominant, disomically inherited loci. Instead, each microsatellite locus was scored as though it was a group of multiple independent dominant loci in which each band is scored as a dominant locus. This approach was suggested in [Rodzen and May \(2002\)](#) as a method to accommodate the highly duplicated nature of the white sturgeon genome. Allelic frequencies and relatedness values were calculated using the methods of [Lynch and Milligan \(1994\)](#). Parentage estimation was performed using methods and algorithms derived for AFLP markers and detailed in [Gerber et al. \(2000\)](#). The high percentages of correct parental assignment and identification of sib groups from relatedness estimates suggest our approach of scoring individual microsatellite bands as dominant markers is appropriate and reliable for estimating parentage and relatedness in the white sturgeon.



## Acknowledgements

This publication was supported in part by the National Sea Grant College Program of the U.S. Department of Commerce's National Oceanic and Atmospheric Administration under NOAA Grant No. NA06RG0142, project No. R/A-109, through the California Sea Grant College Program; and in part by the California State Resources Agency. The views expressed herein do not necessarily reflect the views of any of those organizations. J. Rodzen was supported on a Sea Grant traineeship and a USDA Biotechnology training grant. Special thanks to Stolt Sea Farm California personnel for their assistance in rearing fish and field sampling of their broodstock and to the anonymous reviewers who provided many helpful suggestions for improving this manuscript. The research presented herein provided valuable information to Stolt Sea Farm in their continuing efforts to develop a broodstock management program.

## Appendix A. Conversion of microsatellite bands to dominant marker loci

Phenotypic data were converted to dominant loci using the following method. Consider a microsatellite locus with six bands that amplify in a given population. Each band is assigned an arbitrary band ladder number that corresponds to its size in base pairs. There are six bands in this hypothetical population, represented as follows.

Allelic ladder Number ( <i>i</i> )	Band size (bp)
1	178
2	174
3	170
4	166
5	162
6	158

Now consider an animal with a phenotype of bands 174, 170, and 162 bp. This animal has a phenotype of 174, 170, 162, or using the bands numbered in the allelic ladder, 2, 3, 5.

Each animal's phenotype is converted into a  $1 \times n$  vector where  $n$  is the number of bands at the locus. For each band  $i$  in the system, the  $i$ th element in the vector is assigned a value of 1 if that band is present or a value of 0 if that band is absent in that animal's phenotype. Thus, the 2, 3, 5 phenotype is converted to a  $1 \times n$  vector of [0 1 1 0 1 0], yielding six dominant markers.

For data with more than one microsatellite locus, the process is repeated for each locus  $j$ . This produces a  $1 \times n_j$  matrix for each locus, where  $n_j$  is the number of bands at locus  $j$ . These matrices are then horizontally concatenated, resulting in a single  $1 \times n_T$  matrix where  $n_T$  is the total number of bands across all loci.

Finally, data for all  $k$  individuals in the data set are concatenated in a  $k \times n_T$  matrix. This matrix now contains dominant marker data for all individuals and all microsatellite loci.

## References

- Birstein, V.J., Poletav, A.I., Goncharov, B.F., 1993. The DNA content in Eurasian sturgeon species determined by flow cytometry. *Cytometry* 14, 377–383.
- Birstein, V.J., Hanner, R., DeSalle, R., 1997. Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. *Environmental Biology of Fishes* 48, 127–155.
- Blackledge, K.H., Bidwell, C.A., 1993. Three ploidy levels indicated by genome quantification in Acipenseriformes of North America. *Journal of Heredity* 84, 427–430.
- Blouin, M.S., Parsons, M., Lacaillie, V., Lotz, S., 1996. Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* 5, 393–401.
- Cunningham, E.P., 1999. The application of biotechnologies to enhance animal production in different farming systems. *Livestock Production Science* 58, 1–24.
- Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P., Guomard, R., 1998. Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Canadian Journal of Fisheries and Aquatic Sciences* 55, 715–725.
- Evett, I.W., Weir, B.S., 1998. *Interpreting DNA Evidence*. Sinauer Associates, Sunderland, MA, USA.
- Ferguson, M.M., Danzmann, R.G., 1998. Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Sciences* 55, 1553–1563.
- Gall, G.A.E., 1990. Basis for evaluating breeding plans. *Aquaculture* 85, 125–142.
- Garcia de Leon, F.J., Canonne, M., Quillet, E., Bonhomme, F., Chatain, B., 1998. The application of microsatellite markers to breeding programs in the sea bass, *Dicentrarchus labrax*. *Aquaculture* 159, 303–316.
- Gerber, S., Mariette, S., Streiff, R., Bodenes, C., Kremer, A., 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Molecular Ecology* 9, 1037–1048.
- Gjedrem, T., 1983. Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture* 33, 51–72.
- Gjedrem, T., 1997. Flesh quality improvement in fish through breeding. *Aquaculture International* 5, 197–206.
- Gjoen, H.M., Bentsen, H.B., 1997. Past, present, and future of genetic improvement in salmon aquaculture. *ICES Journal of Marine Science* 54, 1009–1014.
- Hara, M., Sekino, M., 2003. Efficient detection of parentage in a cultured Japanese flounder *Paralichthys olivaceus* using microsatellite DNA marker. *Aquaculture* 217, 107–114.
- Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M., Cook, D., 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* 137, 245–256.
- Kincaid, H.L., 1976. Effects of inbreeding on rainbow trout populations. *Transactions of the American Fisheries Society* 2, 273–280.
- Kincaid, H.L., 1983. Inbreeding in fish populations used for aquaculture. *Aquaculture* 33, 215–227.
- Knibb, W., 2000. Genetic improvement of marine fish—which method for industry? *Aquaculture Research* 31, 11–23.
- Lynch, M., Milligan, B.G., 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3, 91–99.
- Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7, 639–653.
- Norris, A.T., Bradley, D.G., Cunningham, E.P., 2000. Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers. *Aquaculture* 182, 73–83.
- O'Reilly, P.T., Herbinger, C., Wright, J.M., 1998. Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites. *Animal Genetics* 29, 363–370.
- Queller, D.C., Goodnight, K.F., 1989. Estimating relatedness using genetic markers. *Evolution* 43, 258–275.
- Queller, D.C., Strassmann, J.E., Hughs, C.R., 1993. Microsatellites and kinship. *Trends in Ecology and Evolution* 8, 285–288.
- Refstie, T., 1990. Applications of breeding schemes. *Aquaculture* 85, 163–169.
- Rodzen, J.A., May, B., 2002. Inheritance of microsatellite loci in the white sturgeon *Acipenser transmontanus*. *Genome* 45, 1064–1076.
- Sekino, M., Saito, K., Yamada, T., Kumagai, A., Hara, M., Yamashita, Y., 2003. Microsatellite-based pedigree

tracking in the Japanese flounder *Paralichthys olivaceus* hatchery strain: implications of hatchery management related to stock enhancement program. *Aquaculture* 221, 255–263.

Waldbeiser, G.C., Wolters, W.R., 1999. Application of polymorphic microsatellite loci in a channel catfish *Ictalurus punctatus* breeding program. *Journal of the World Aquaculture Society* 30, 256–262.

Weir, B.S., 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA, USA.