

# Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations

Arnaud Estoup, Karim Gharbi, Magali SanCristobal, Claude Chevalet, Pierrick Haffray, and René Guyomard

**Abstract:** Eight turbot (*Scophthalmus maximus*) and eight rainbow trout (*Oncorhynchus mykiss*) microsatellites were selected for parentage assignment in fish-farmed populations. The number of alleles, gene diversity, polymorphic information content, and the probabilities of exclusion of these loci were 8, 0.76, 0.73, and 0.55 in turbot and 4, 0.65, 0.59 and 0.39 in rainbow trout, respectively. The power of the markers for parentage assignment was assessed by computing the frequency of good and unique decisions ( $f_{gu}$ ) in a population of genitors defined by its allele frequencies and assuming three different types of mating schemes. The eight turbot microsatellites gave larger maximal mating schemes (the largest mating structure with a  $f_{gu} \geq 0.95$ ) than the eight rainbow trout loci: 1 female ( $F$ ) mated to 520 males ( $M$ ) (paternity retrieval scheme), more than  $140F \times 140M$  (factorial scheme), and more than 15 000 independent pairs (natural population scheme) for turbot, and  $1F \times 88M$ ,  $34F \times 34M$ , and 7000 independent pairs for rainbow trout. The variation of the  $f_{gu}$  values with the number of loci confirmed that the turbot set of microsatellites was more efficient for parentage assignment than the rainbow trout markers.

**Résumé :** Huit microsatellites ont été testés pour l'assignation de parenté sur des populations de pisciculture de turbot (*Scophthalmus maximus*) et de truite arc-en-ciel (*Oncorhynchus mykiss*). Ces locus étaient caractérisés par un polymorphisme élevé avec un nombre d'allèles, une diversité génique, un contenu en polymorphisme informatif et une probabilité d'exclusion respectivement supérieurs à 8, 0.76, 0.73 et 0.55 chez le turbot et à 4, 0.65, 0.59 et 0.39 chez la truite arc-en-ciel. La puissance pour l'assignation de parenté des marqueurs a été évaluée en calculant des fréquences de décisions uniques et justes ( $f_{gu}$ ) dans une population de géniteurs définie par ses fréquences alléliques et en supposant trois types de schéma d'accouplement différents. Les huit locus turbots ont donné des schémas d'accouplement maximaux (structure d'accouplement la plus large avec un  $f_{gu} \geq 0.95$ ) supérieurs à ceux obtenus avec les huit locus truites arc-en-ciel: 1 femelle ( $F$ ) accouplée à 520 mâles ( $M$ ) (schéma de recherche de paternité), plus de  $140F \times 140M$  (schéma factoriel), et plus de 15 000 paires indépendantes (schéma de population naturelle) pour le turbot, et  $1F \times 88M$ ,  $34F \times 34M$  et 7000 paires indépendantes pour la truite arc-en-ciel. Les variations des  $f_{gu}$  avec le nombre de locus ont confirmé que le jeu de microsatellites du turbot était plus efficace pour l'assignation de parenté que les marqueurs de truite arc-en-ciel.

## Introduction

At any particular selection intensity, genetic progress is maximized by a combination of within-family and between-family selection (Falconer 1981). This combined selection requires that pedigree information is available. In fish, such information can be obtained if the different families are reared in separate places (ponds, cages, or tanks), at least until they can be physically tagged. The management of families in separate tanks is costly and labour intensive, particularly for marine fish

species where fry have to be fed with live prey until weaning. Alternatively, pedigree information on mixed families can be obtained through parentage assignment using highly polymorphic genetic markers such as mini- or micro-satellites (Harris et al. 1991; Wright and Bentzen 1994). The first advantage of this alternative is that large and high-cost facilities are not required. However, molecular typing remains relatively expensive and the cost of certain selection schemes would probably not compete significantly with that of the independent rearing of families until physical tagging. The second main advantage of molecular typing is that it allows families to be reared under competition from the beginning, so that selection can be achieved under the same conditions as farm production. It is worth mentioning that in present and future experimental fish selection schemes using molecular pedigree determination (e.g., the "walk-back selection" of Doyle and Herlinger (1994)), molecular and physical tagging methods are or will be used in a complementary way to minimize cost or simply to identify genitors selected according to their molecular pedigree.

The use of microsatellites for parentage assignment in

Received March 19, 1997. Accepted August 1, 1997.  
J13925

A. Estoup, K. Gharbi,<sup>1</sup> and R. Guyomard. Laboratoire de Génétique des Poissons, INRA, 78352 Jouy-en-Josas, France.  
M. SanCristobal and C. Chevalet. Laboratoire de Génétique Cellulaire, INRA, 31326 Castanet Tolosan, France.  
P. Haffray. SYSAF Aquacole, 65 rue de St Briec, 35042 Rennes, France.

<sup>1</sup> Author to whom all correspondence should be addressed.  
e-mail: gharbi@jouy.inra.fr

fish-farmed populations is in its infancy. Herlinger et al. (1995) used four microsatellite markers to establish pedigrees in a mixed aquaculture population of rainbow trout obtained after a complete factorial cross between 10 males and 10 females. About 65% of the fish could be traced to a single parental couple and 91% to one or two parental couples of the 100 possible couples. A statistical method recently developed by SanCristobal and Chevalet (1997) determines, for a given mating scheme, the most likely father-mother pair among a finite set of possible parental pairs; this method allows for possible typing errors or mutations. By means of random simulations, SanCristobal and Chevalet (1997) investigated general situations to characterize the joint influences of the number and polymorphism of typed loci, of the population structure and size, and of errors rates. Yet, none of these investigations were carried out using an empirical microsatellite data set, for which alleles are often far from being equiprobable as assumed in theoretical studies.

In the present study, we used the method of SanCristobal and Chevalet (1997) to assess the power for parentage assignment of actual sets of microsatellite markers developed for aquaculture stocks of two fish species of economical interest, the turbot (*Scophthalmus maximus*) and the rainbow trout (*Oncorhynchus mykiss*). The following questions were more precisely addressed. Do these sets of turbot and rainbow trout microsatellites have different powers for parentage assignment? What is the maximal mating scheme that can be analysed with the selected sets of turbot and rainbow trout microsatellites? How does the power of parentage assignment depend on the number of loci and on the mating scheme structure?

## Materials and methods

### Biological material

Forty-five turbot were sampled in a population of adult genitors used for cohort production by the fish-farm company France-Turbot (France). These fish of wild origin were captured in the Atlantic Ocean near the island of Noirmoutier (France). Thirty adult rainbow trout were sampled in a French commercial strain. Blood samples (turbot) and adipose fin samples (rainbow trout) were stored at room temperature in a TNES-urea solution (Estoup et al. 1993) and in pure ethanol, respectively. Individual DNA extractions were performed as described in Estoup et al. (1993, 1996) from blood and adipose fin samples, respectively.

### Cloning, analysis, and selection of microsatellites

Eleven-hundred recombinant clones for turbot and 1350 for rainbow trout were screened using an equal mix of (TC)<sub>10</sub>, (TG)<sub>10</sub>, (CAC)<sub>5</sub>CA, CT(CCT)<sub>5</sub>, CT(ATCT)<sub>6</sub>, and (TGTA)<sub>6</sub>TG oligonucleotides labelled with the DIG oligonucleotide tailing kit (Boehringer); the DIG nucleic acid detection kit (Boehringer) was used for detection. Basic cloning, screening, and detection steps are described in Estoup and Cornuet (1994). Detailed and optimized procedures are available at the World-Wide Web address <http://www.inap.inra.fr/dsa/microsat/microsat.htm>. Inserts of positive clones were sequenced as described in Estoup et al. (1993), except that  $\alpha^{33}\text{P}$ -dATP was used instead of  $\alpha^{35}\text{S}$ -dATP. Microsatellites were amplified using  $\gamma^{33}\text{P}$  radioactive PCR. For each primer pair, one primer was end-labeled with  $\gamma^{33}\text{P}$ -dATP. The 10- $\mu\text{L}$  end-labeling mixture contained 20 pmol of primer, 4 units of T4 polynucleotide kinase, 1  $\mu\text{L}$  of 10 $\times$  buffer (Biolabs), and 5  $\mu\text{L}$  of  $\gamma^{33}\text{P}$ -dATP. The 10- $\mu\text{L}$  PCR mixture contained 2  $\mu\text{L}$  of DNA template, 400 nM of nonradioactive primer, 0.11  $\mu\text{L}$  of radioactive

primer solution, 75  $\mu\text{M}$  of each dNTP, 0.7–1.5M  $\text{MgCl}_2$ , 20  $\mu\text{g}/\text{mL}$  bovine serum albumin, 1 $\times$  Promega reaction buffer, and 0.25 unit of *Taq* polymerase (Promega). After a denaturing step of 5 min at 96°C, samples were processed through 5 cycles consisting of 1 min at 96°C, 30 s at the optimal annealing temperature, and 1 min 15 s at 72°C followed by 20–25 cycles of 30 s at 95°C, 30 s at the optimal annealing temperature, and 1 min 15 s at 72°C. The last elongation step was lengthened to 5 min. PCRs were performed in a PTC100 machine (MJ Research) without overlaying mineral oil. After direct addition of 4  $\mu\text{L}$  of loading solution, the mixture was heated for 4 min at 90°C, and 2  $\mu\text{L}$  was electrophoresed on 6% denaturing polyacrylamide sequencing gels. The polymorphism level of 12 turbot and 20 salmonid (rainbow trout and brown trout) microsatellites was first tested by typing eight genitors of turbot and eight of rainbow trout. Tested salmonid markers were rainbow trout microsatellites cloned in the present study, those published in Morris et al. (1996) and Sakamoto et al. (1994a, 1994b, 1994c), and brown trout microsatellites cloned in Estoup et al. (1993). Microsatellites with both a high level of variability and a nonambiguous allelic pattern (with a low number of stutter bands) were selected for the typing of all remaining individuals. Because rainbow trout is of tetraploid origin, as are other salmonids (Allendorf and Thorgaard 1984), a single pair of primers sometimes amplified a second locus. This category of rainbow trout locus was not selected unless each duplicated locus gave alleles with very different ranges of size variation, as observed for the locus Omy25INRA for which the second copy of the locus was fixed for a much shorter allele (at least 18 bp).

### Statistical analyses

Deviation from Hardy-Weinberg equilibrium and populational linkage disequilibrium between all possible pairs of loci were analysed by Fisher's exact test, using the GENEPOP package (Raymond and Rousset 1995). For Hardy-Weinberg equilibrium, significance of multiple  $P$  values were tested using the Fisher's method (GENEPOP package) and applying a Bonferroni-type correction (Rice 1989). For populational linkage disequilibrium, multiple probability tests were processed applying the improved Bonferroni procedure for dependent test results (Hochberg 1988). The gene diversity was estimated for each locus by

$$H = 1 - \sum_i p_i^2$$

where  $p_i$  is the frequency of allele  $i$  estimated over  $n$  chromosomes sampled at this locus in this population (Nei 1987). The polymorphic information content (PIC) was computed for each locus according to Botstein et al. (1980). The probability of exclusion was calculated for each locus ( $E_i$ ) using the expressions given by Hanset (1975) and Smouse and Chakraborty (1986). Global probabilities of exclusion ( $E_g$ ) were calculated for different numbers of  $L$  loci as

$$E_g = 1 - \prod_{i=1}^{L-1} (1 - E_i)$$

adding loci from most to least informative (maximum global probability of exclusion,  $E_{g\text{max}}$ ) and adding loci from least to most informative (minimum global probability of exclusion,  $E_{g\text{min}}$ ).  $H$ , PIC,  $E$ , and  $E_g$  were computed using personal programs. Differences in mean number of alleles,  $H$ , PIC, and  $E$  between the two sets of turbot and rainbow trout microsatellite markers were assessed using the Mann-Whitney  $U$  test (Sprent 1989). Because the number of alleles depends on the sample size, it was adjusted for a common sample size of 90 chromosomes for the above test using Ewens's formula (Ewens 1972).

### Parentage assignment

Three types of mating scheme were considered: (i) a paternity retrieval scheme, where  $M$  males were mated to one female; (ii) a

factorial scheme, where  $M$  males were mated to  $F$  females in a crossed scheme; (iii) a natural population scheme, where  $N$  independent pairs of  $M = N$  unrelated males and  $F = N$  unrelated females mated.

For a given mating scheme, the method of SanCristobal and Chevalet (1997) determines for an offspring's genotype the most likely father-mother pair among a finite set of possible parental pair genotypes. This method allows for possible typing errors or mutations. The power of a set of markers for parentage assignment in a genitor population was tested as follows: a program generates  $M + F$  parental genotypes according to the allele frequencies in the genitor population, draws at random one parental pair and generates one offspring from this pair, allowing for errors. For each drawing, a posterior probability value is determined for each of the  $M \times F$  parental pairs and sorted. The largest one identifies the likeliest pair, which is compared to the true one. The frequency of good and unique decisions ( $f_{gu}$ ) is then calculated over 100 000 independent runs. To save computation time, the number of runs was reduced to 10 000 when the crossing schemes were higher than  $1F \times 800M$ ,  $100F \times 100M$ , and 1000 independent pairs. Cases with several pairs sharing the same likelihood are recorded as erroneous decisions. The above computation assumes Hardy-Weinberg equilibrium for the genitor population, statistical independence between loci, and equal distribution of offspring of the different families in any crossing scheme. It is worth stressing that  $f_{gu}$  are not computed from actual parental genotypes but from parental genotypes generated according to the allele frequencies estimated from a sample of the genitor population. SanCristobal and Chevalet (1997) showed that an efficient decision rule is obtained when a nonzero rate is allowed for the typing error or mutations, even if no precise data are available concerning this rate. For each locus and computation, the error rate was fixed to the high and hence conservative value of 5%.

To assess the power of the cloned loci for parentage assignment, the maximal configuration, defined as the largest one for which the  $f_{gu}$  value is still  $\geq 0.95$ , was determined in the turbot and rainbow trout samples for the three above-mentioned mating schemes. This was performed including the entire set of loci in the computation and selecting the four loci with the highest probability of exclusion. The later case (four loci) corresponds to a realistic number of loci that could be typed when considering the time and financial support attributable to a program of routine nonautomated pedigree determination. In turbot, maximal configurations with the entire set of loci were estimated in a conservative way ( $f_{gu} \gg 0.95$ ) for factorial and natural population schemes, because the parental structures involved were so large that the computation times exceeded 1 week.

For turbot and rainbow trout, the influence of the number of loci on  $f_{gu}$  values was determined for a  $10 \times 20$  factorial mating scheme. In the corresponding computation, loci were progressively removed from (i) lowest to highest  $E$  and (ii) from highest to lowest  $E$ . Procedure (i) allowed estimation of the maximum  $f_{gu}$  ( $f_{gu \max}$ ) and procedure (ii), the minimum  $f_{gu}$  ( $f_{gu \min}$ ), for different numbers of markers. To study the influence on the power of parentage assignment of the number of possible couples and of the total number of parents, the  $f_{gu \max}$  values were computed for different numbers of loci and compared in turbot for  $1 \times 20$  and  $4 \times 5$  mating schemes (20 possible couples for both schemes), and for  $1 \times 20$  and  $10 \times 11$  mating schemes (21 parents for both schemes).

## Results

### Cloning and selection of microsatellites

Averaged size of inserts was 300 and 550 bp in turbot and rainbow trout, respectively. Fifty and 100 positive clones were detected of the 1100 and 1350 recombinant clones and stored at  $-80^\circ\text{C}$  for further analyses in turbot and rainbow trout, respectively. Twenty-four (18 in turbot and 6 in rainbow trout) of the clones characterized by a positive signal of high

intensity were sequenced. This category of clones is expected to contain a microsatellite sequence composed of a high number of tandem repeats. By proceeding this way, we hoped to bias our sequencing towards highly variable markers (Weber 1990). Congruent with this, all except two clones contained a microsatellite sequence (mostly of type  $(TG)_n$ ) with the longest stretch of uninterrupted repeats ranging from 13 to 59 repeated units. Eight turbot and eight salmonid microsatellites with both high levels of variability and unambiguous allelic patterns were selected for typing the entire farm population samples. Primer sequences and PCR conditions are given for each selected locus in Table 1.

### Hardy-Weinberg equilibrium and linkage disequilibrium

No significant deviation from Hardy-Weinberg expectations (HWEs) was found at the 5% level in the eight probability tests for turbot, and only one was found out of the eight probability tests for rainbow trout (locus PuPuPy,  $P = 0.03$ ). For both species, multiple probability tests by population (Fisher's method) did not yield any significant deviation from HWEs ( $P = 0.58$  and  $P = 0.10$  for turbot and rainbow trout, respectively), and the individual probability test for the locus PuPuPy was not significant after Bonferroni-type correction. Exact tests for linkage disequilibrium between microsatellite loci gave one and five significant  $P$  values at the 5% level of 28 pairs of loci in turbot and rainbow trout, respectively. None of these pairs of loci showed significant linkage disequilibrium after the Bonferroni-type correction for dependent test results.

### Genetic diversity measures and probability of exclusion

Because the probability of exclusion  $E$  is a slightly better predictor of the true probability of correct parentage assignment decision than PIC and  $H$  (SanCristobal and Chevalet 1997), microsatellite loci were ranked in Table 2 according to their individual  $E$  values. PIC values had the same ranking as  $E$  values, but this was not the case for a small proportion of  $H$  values in rainbow trout and to a much larger extent for  $n_{\text{all}(c)}$  values in both turbot and rainbow trout. Ranking discrepancies between  $E$  and  $n_{\text{all}(c)}$  highlight the influence of the distribution of alleles in the parentage assignment power of a locus (see Appendix). For a given number of alleles, the closer they are to an equiprobable distribution, the higher is the power of the locus for parentage exclusion (Smouse and Chakraborty 1986).

All selected microsatellites displayed a high level of polymorphism with  $n_{\text{all}}$ ,  $H$ , PIC, and  $E$  greater than 8, 0.761, 0.724, and 0.547 in turbot, respectively, and to 4, 0.646, 0.589, and 0.389 in rainbow trout, respectively. Mean values of  $n_{\text{all}(c)}$ ,  $H$ , PIC, and  $E$  were higher in turbot than in rainbow trout. However, standard deviations overlapped (Table 2), and no significant difference between any of these measures were detected in the two species using the Mann-Whitney  $U$  test ( $P > 0.07$ ). For any given number of loci, minimum and maximum  $E_g$  values were higher in turbot than in rainbow trout, indicating that turbot microsatellites are more efficient for parentage assignment than rainbow trout loci (Table 2). This result is illustrated by Fig. 1 in which  $-\log(1 - E_g)$  was plotted for different numbers of loci in both species.

To test the propensity of  $E_g$  to predict the true probability of correct parentage assignment decisions on our set of experimental data, the same minimum and maximum  $E_g$  values were plotted against the corresponding minimum and maximum

**Table 1.** Core sequence of microsatellite alleles cloned from turbot (*Scophthalmus maximus*), rainbow trout (*Oncorhynchus mykiss*), and brown trout (*Salmo trutta*), primer sequences, and PCR conditions.

Locus	Core sequence	Primers (5′–3′)	PCR conditions		
			T <sub>m</sub> (°C)	MgCl <sub>2</sub> (mM)	No. of cycles
Turbot					
Sma3-10INRA	(TG) <sub>50</sub>	CTATGCAAGTGAAGACACGTAC CTTTCACCTCTGTTGTATCTAAGG	58	1.2	30
Sma3-8INRA	(GT) <sub>18</sub> *** (CT) <sub>5</sub>	CCCTCCGTCAGACAAAGAG GACGAAGTTAATGTTTCATTG	55	1.0	28
Sma3-12INRA	(TG) <sub>21</sub>	CACAATTGAATCACGAGATG GCCACCACTGCGTAACAC	58	1.0	30
Sma3-129INRA	(GT) <sub>29</sub>	GCACTGCCTTTTCATTGG CAGCTCTAGATTGTTTATCCC	58	1.5	30
Sma4-14INRA	(GT) <sub>21</sub>	GCGAGCAAATATCAGAAGG CCAGAAACAGCTCCGACTC	58	0.7	30
Sma5-111INRA	(TG) <sub>13</sub>	TCTACACTGCAGGTTGGG CTGATTATGGGCTGGACG	52	1.2	30
Sma1-152INRA	(GT) <sub>14</sub>	GGGCAGGTGTCTGTCTGG GCACAGCTCCGAGCACAG	56	1.2	30
Sma1-125INRA	(TAGA) <sub>11</sub> *** (TG) <sub>4</sub>	CACACCTGACAAAGCTCAAC GCTGAACATTTTCATGTTGATAG	58	1.0	30
Rainbow trout					
Omy18INRA	(CT) <sub>43</sub>	CGGGCTCAGATTTTCACAG GATGCATGTACTCTGTAGGTAGC	56	0.7	25
Omy25INRA	(GT) <sub>20</sub>	GATGGACGAGGACCTCAG CTACAGATACAGGACAGGGAC	55	1.0	30
Str2INRA	(CT) <sub>4</sub> (TG) <sub>31</sub>	GGTGGCCTGGGTATAGCC GGTGTCGTTTCAGCTGTAGCG	56	0.8	28
PuPuPy	(PuPuPy) <sub>n</sub>	AGGGGGAGCATGCAGCG CATTTAAGTGAAAAGACGTAACCTTACC	59	1.1	30
Omy77	(TC) <sub>25</sub> *** (TC) <sub>5</sub>	GTTCTCTACTGAGTCATGGGGTC CCAAGAATTTTCTGATCCGG	55	1.4	30
FGT2	(TG) <sub>12</sub> TT(TG) <sub>18</sub>	GTGTGGAACCCAACAGGC GAAACACGTCTGGGGTCAGT	63	0.8	25
FGT3	(TG) <sub>35</sub>	CAAGAAATTTGTGGAGCGG GAAGCCCTGTTTGACTTTTAGC	58	1.1	28
FGT4	(TG) <sub>25</sub>	CTCACAACAGCACACGCAC CATTTTCAGGTTTGGCGC	58	1.2	28

**Note:** All turbot microsatellites and the rainbow trout markers Omy18INRA and Omy25INRA were cloned in the present study. The loci PuPuPy and Omy77 were cloned and named PuPuPy and Omy77 by Morris et al. (1996), and loci FGT2, FGT3, and FGT4 were cloned by Sakamoto et al. (1994a, 1994b, 1994c). The renaming of loci PuPuPy and Omy77 is justified by the fact that new primers were designed from the published sequences of the flanking regions. Brown trout microsatellite locus Str2INRA was cloned by Estoup et al. (1993). T<sub>m</sub>, annealing temperature.

$f_{gu}$  values computed for a  $10 \times 20$  factorial scheme in turbot and rainbow trout (Fig. 2). All curves strongly overlapped especially for  $E_g$  and  $f_{gu}$  values computed with more than four loci, indicating the value of  $E_g$  as a predictor of  $f_{gu}$ , and confirming the approximate linear relationship between  $-\log(1 - E_g)$  and  $\log(f_{gu}/(1 - f_{gu}))$  for  $E_g > 0.95$  (SanCristobal and Chevalet 1997).

## Parentage assignment

### Maximal mating schemes

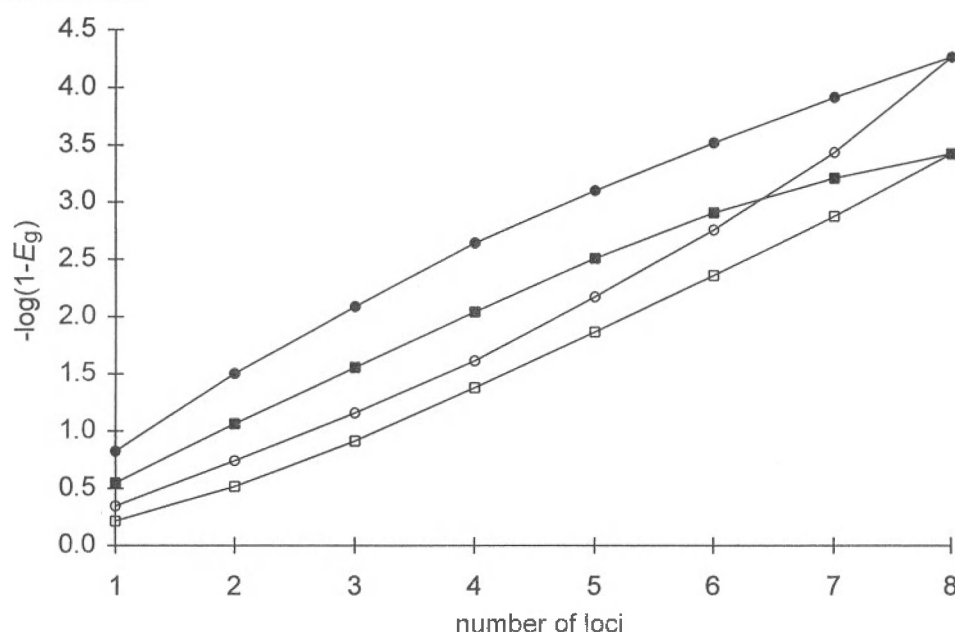
In rainbow trout, the maximal mating schemes (the largest mating structures that gave a  $f_{gu} \geq 0.95$ ) were 1 female ( $F$ ) mated by 80 males ( $M$ ) (paternity retrieval scheme),  $34F \times 34M$  (factorial scheme), and 7000 independent pairs (natural population scheme) for the entire set of eight loci. The eight

turbot microsatellites gave much larger maximal mating schemes, i.e.,  $1F \times 520M$ , more than  $140F \times 140M$ , and more than 15 000 independent pairs. Similar computations performed with the four loci with the highest  $E$  gave the following maximal mating schemes:  $1F \times 6M$ ,  $3F \times 3M$ , and 80 independent pairs for rainbow trout and  $1F \times 17M$ ,  $8F \times 8M$ , and 620 independent pairs for turbot.

### Influence of the number of loci and of the mating scheme structure

The variation of minimum and maximum  $f_{gu}$  with the number of loci (Fig. 3.) reflect the way loci were added in the computations, i.e., from most to least informative for maximum  $f_{gu}$  estimates and from least to most informative for minimum  $f_{gu}$  estimates. For any number of loci, minimum and maximum  $f_{gu}$  values in turbot were higher than minimum and maximum  $f_{gu}$

**Fig. 1.** Values of  $-\log(1 - E_g)$  for different numbers of loci in turbot and rainbow trout. Maximum and minimum global probability of exclusion ( $E_{g \max}$  and  $E_{g \min}$ , respectively) were computed adding loci from highest to lowest  $E$  and from lowest to highest  $E$ , respectively.  $E_{g \max}$  and  $E_{g \min}$  values are reported in Table 2. (●) Turbot maximum values; (■) rainbow trout maximum values; (○) turbot minimum values; (□) rainbow trout minimum values.



**Table 2.** Measures of genetic diversity and probability of exclusion for microsatellite loci in turbot and rainbow trout farmed populations.

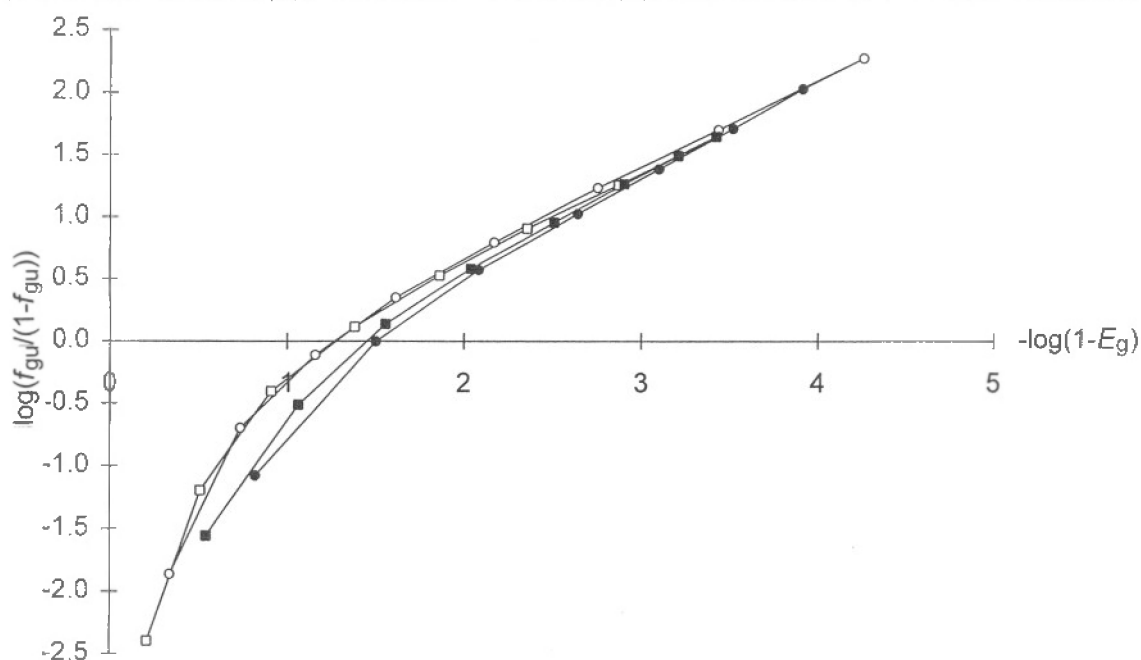
Locus	$n_{\text{all}}$	$n_{\text{all(c)}}$	$H$	PIC	$E$	$E_{g \max}$	$E_{g \min}$
<b>Turbot</b>							
Sma3-10INRA	26	26.00	0.926	0.921	0.850	0.850 000	0.999 945
Sma3-129INRA	19	19.15	0.895	0.886	0.790	0.968 500	0.999 630
Sma3-8INRA	12	12.00	0.869	0.855	0.739	0.991 778	0.998 239
Sma5-111INRA	13	13.08	0.861	0.845	0.723	0.997 723	0.993 253
Sma1-125INRA	13	13.00	0.818	0.795	0.649	0.999 201	0.975 644
Sma3-12INRA	10	10.00	0.802	0.775	0.618	0.999 695	0.930 609
Sma4-14INRA	12	12.00	0.774	0.753	0.599	0.999 878	0.818 347
Mean (SD)	14.13 (5.37)	14.25 (5.37)	0.838 (0.055)	0.819 (0.064)	0.689 (0.097)		
<b>Rainbow trout</b>							
Str2INRA	11	12.58	0.859	0.843	0.717	0.717 000	0.999 619
PuPuPy	8	9.17	0.848	0.830	0.695	0.913 685	0.998 654
FGT3	10	11.24	0.834	0.815	0.677	0.972 120	0.995 586
Omy77	9	10.16	0.835	0.815	0.673	0.990 883	0.986 333
FGT2	10	11.64	0.826	0.804	0.659	0.996 891	0.958 205
FGT4	10	12.3	0.778	0.755	0.598	0.998 750	0.877 435
Omy18INRA	5	5.74	0.733	0.691	0.501	0.999 376	0.695 111
Omy25INRA	4	4.49	0.646	0.589	0.389	0.999 619	0.389 000
Mean (SD)	8.38 (2.39)	9.64 (2.82)	0.795 (0.068)	0.768 (0.082)	0.614 (0.106)		

**Note:** Loci were ranked in both species according to their individual probability of exclusion ( $E$ ).  $n_{\text{all}}$ , observed number of alleles;  $n_{\text{all(c)}}$ , number of alleles adjusted for a common sample size of 90 chromosomes;  $H$ , gene diversity; PIC, polymorphic information content;  $E_{g \max}$ , global probability of exclusion calculated for different numbers of loci adding loci from lowest to highest  $E$ ;  $E_{g \min}$ , global probability of exclusion calculated for different numbers of loci adding loci from highest to lowest  $E$ .

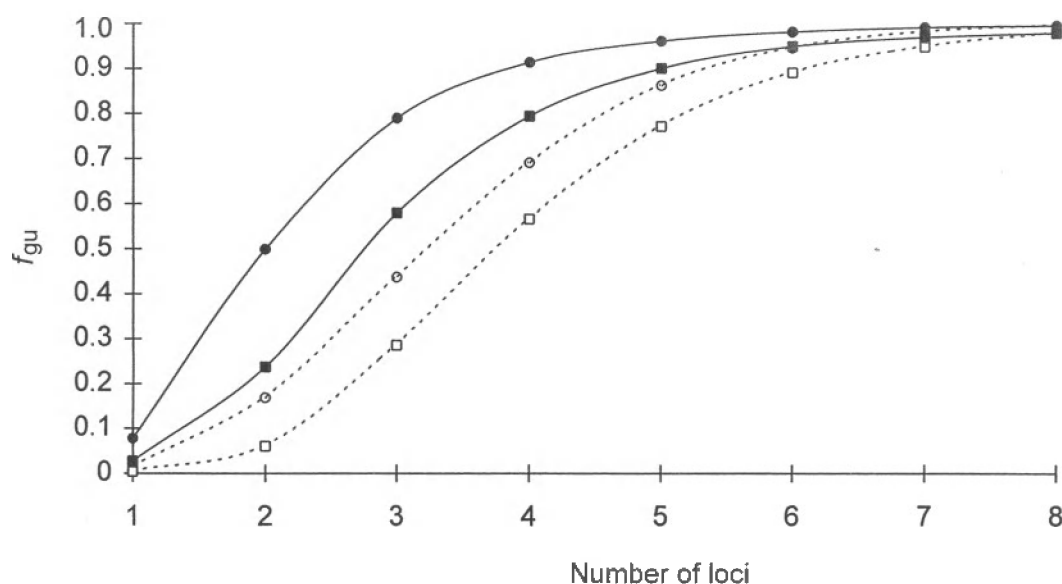
values in rainbow trout, respectively, confirming that turbot microsatellites are more efficient for parentage assignment than rainbow trout loci. This result also appeared with similar computations based on different mating schemes (computations not shown). Two additional factors are expected to

influence  $f_{gu}$  values: the number of possible couples and the total number of parents. Our computations showed that, for a given number of possible couples, the mating scheme with the highest number of parents gave the lowest  $f_{gu}$ , and, reversibly, for a given number of parents, the mating scheme with the

**Fig. 2.** Relationship between  $\log(f_{gu}/(1-f_{gu}))$  and  $-\log(1-E_g)$  computed for different numbers of loci assuming a  $10 \times 20$  factorial scheme in turbot and rainbow trout. Maximum and minimum global probability of exclusion and unique and good frequency were computed adding loci from highest to lowest  $E$  for maximum values and from lowest to highest  $E$  for minimum values.  $E_{g \max}$  and  $E_{g \min}$  values are reported in Table 2. (●) Turbot maximum values; (■) rainbow trout maximum values; (○) turbot minimum values; (□) rainbow trout minimum values.



**Fig. 3.** Maximum  $f_{gu}$  values (loci with the highest  $E$ ) and minimum  $f_{gu}$  values (loci with the lowest  $E$ ) for different numbers of markers in turbot and rainbow trout assuming a  $10 \times 20$  factorial scheme. (●) Turbot  $f_{gu \max}$ ; (■) rainbow trout  $f_{gu \max}$ ; (○) turbot  $f_{gu \min}$ ; (□) rainbow trout  $f_{gu \min}$ .



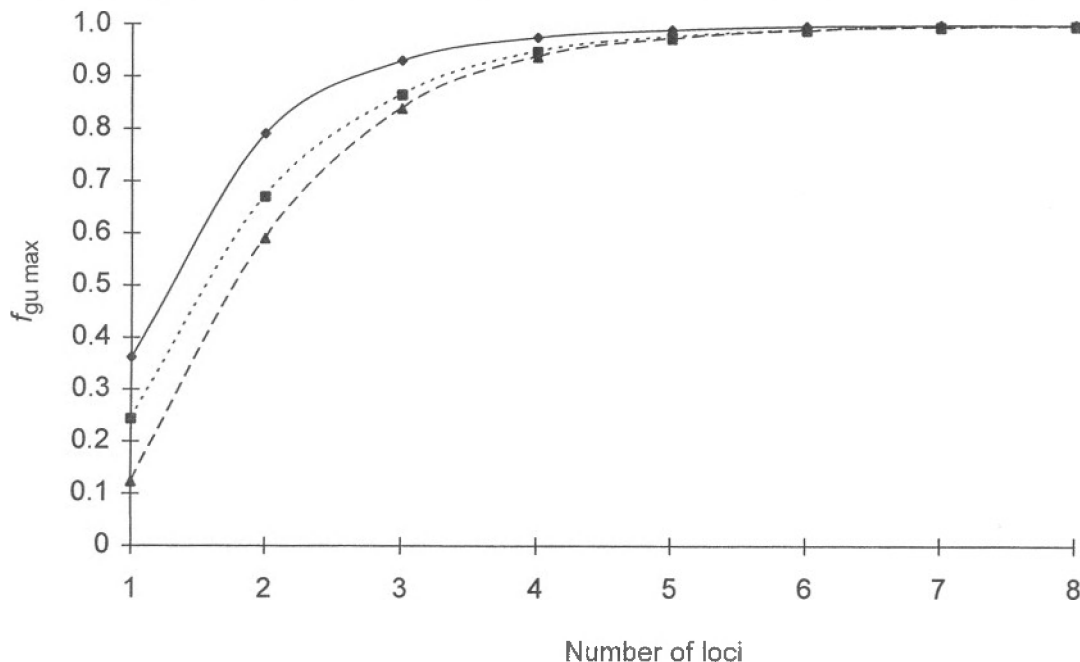
highest number of couples gave the lowest  $f_{gu}$ . For example, in turbot the curve corresponding to a  $1 \times 20$  scheme stays below the curve of a  $5 \times 4$  scheme, and the curve of a  $10 \times 11$  scheme stays below the curve of a  $1 \times 20$  scheme (Fig. 4). Similar results were obtained for additional computations based on different numbers of parents in turbot and rainbow trout (computations not shown).

## Discussion

### Higher power for parentage assignment of turbot microsatellites

The greater power of turbot microsatellites for parentage assignment is due to the higher mean level of variability at these loci in the typed population sample. Although this result did

**Fig. 4.** Influence of the number of possible couples and of the number of parents on parentage assignment. Maximum  $f_{gu}$  values (loci with the highest  $E$ ) were computed for different numbers of turbot microsatellites assuming a  $1 \times 20$  and  $5 \times 4$  crossing scheme (20 possible couples for both schemes) and a  $10 \times 11$  and  $1 \times 20$  crossing scheme (21 parents for both schemes). (◆)  $5 \times 4$ , (■)  $1 \times 20$ , (▲)  $10 \times 11$ .



not appear statistically significant, presumably because of the limited number of loci, it holds for all genetic variability measures calculated. A possible explanation could be that a reduction of polymorphism in rainbow trout has occurred because of several generations of domestication (turbot genitors are from a natural population). However, unpublished allozyme data indicated that the level of polymorphism observed in this study of domestic rainbow trout was greater than or equal to the level of polymorphism observed within natural rainbow trout populations (Allendorf 1975; Berg and Gall 1988). Therefore, we believe that this hypothesis is unlikely to explain the differences in microsatellite polymorphism observed between turbot and rainbow trout. An alternative and more likely explanation is that effective population sizes are higher in turbot than in rainbow trout populations. This hypothesis has been proposed to explain why marine fish species generally displayed higher gene diversity within populations than freshwater fish species in protein studies (Gyllenstein 1985). However, it is noteworthy that much lower intrapopulation variation was found at protein loci in natural populations of turbot (Blanquer et al. 1992; Bouza et al. 1997) than in domesticated stocks of rainbow trout (Busack et al. 1979; Guyomard 1981 and unpublished data). A possible explanation is that turbot populations have experienced a bottleneck event and that microsatellite loci have already reached steady-state values while protein loci are still far from equilibrium, because of the differences in mutation rates between the two categories of markers.

#### Basic assumptions for parentage assignment

The first basic assumption, HWE, does not concern the computation of the most likely father-mother pair. Therefore, a significant deviation from HWE of one or several loci should have no influence on the estimation of multilocus individual  $f_{gu}$  values, i.e.,  $f_{gu}$  computed for a given mating scheme with a

given set of parental genotypes. However, HWE is assumed when generating the putative parental pair's genotypes. Thus, deviation from HWE could lead to mean  $f_{gu}$  that would not necessarily correspond to the actual mean  $f_{gu}$  within the genitor population.

The second basic assumption, statistical independence between loci, can be split up into two different cases: genetic linkage at (i) the individual level and (ii) the population level. A genetic linkage at the individual level would decrease the number of efficient loci and consequently lower multilocus individual and hence mean  $f_{gu}$  values. It is worth noting that if genetic linkage at an individual level is detectable at a population level, the offspring typing of a full-sib family is necessary to demonstrate its occurrence. Genetic linkage at the population level can be the consequence of several causes (excluding the case of a genetic linkage at an individual level) such as, the mixture of genetically differentiated demes (Wahlund effect), a strong family substructuring of the population, a recent bottleneck, or it can occur by chance (Hill and Robertson 1968). This category of genetic linkage has no influence on multilocus individual  $f_{gu}$  values but could lead to biased mean  $f_{gu}$  estimates since putative parental genotypes are generated assuming a statistical independence of loci.

The third basic assumption of the parentage assignment procedure used in the present study is that offspring of different families are equally distributed in any crossing scheme. Different family survival rates certainly occur in natural and hatchery populations. This means that actual mean  $f_{gu}$  will be biased towards the  $f_{gu}$  values of the most represented families, the extent of this bias decreasing with the number and power of the loci analysed.

Tests of parentage assignment directly based on actual parental genotypes would be independent of HWE and genetic linkage disequilibrium at the population level. However, such



tests would remain potentially biased by genetic linkage at the individual level as well as by different family survival rates.

### Optimization of parentage assignment protocols

Parentage assignment protocols can be optimized at both the marker and parental level.

The first criterion that should be applied to markers is the selection of loci with a high probability of exclusion. This statistic is indeed a slightly better predictor of the true probability of correct parentage assignment decision than the polymorphic information content, the gene diversity (SanCristobal and Chevalet 1997), and particularly the number of alleles. The amplification of several loci in the same tube (multiplex PCR) and (or) the migration in a single gel lane (multiplex loading) of several of such marker sets using fluorescently labelled primers and semiautomatic genotyping systems is also of interest, because it significantly lowers the cost and time required for data acquisition (Olsen et al. 1996; Wenburg et al. 1996). In that respect, it is worth mentioning that two multiplexed loci with  $n$  alleles each are more informative, for roughly the same amount of molecular work, than a single one with  $2n$  alleles. Indeed, assuming that alleles are equiprobably distributed, two independent loci with  $n$  alleles each give  $g_1 = n^2(n+1)^2/4$  and a single locus with  $2n$  alleles gives  $g_2 = n(2n+1)$  possible genotypes, with  $g_1 > g_2$  for  $n > 2$ . The quality of the allelic pattern should also be taken into consideration because loci with a high number of stutter bands are often associated with high typing error rates. Finally, the occurrence of null alleles (alleles that do not give any PCR product) in the parental gene pool have to be avoided, because they could greatly affect pedigree determination (Callen et al. 1993; Pemberton et al. 1995; Brookfield 1996). The presence of null alleles can be suspected but not demonstrated by testing homozygote deficiency in the genitor population sample. An effective but limiting strategy for dealing with null alleles is to use only loci that are typed as heterozygous in a given set of parents or offspring (Pemberton et al. 1995) or, reversibly, only parents that are typed as heterozygous at a given set of loci. The integration into individual and mean  $f_{gu}$  computation of putative null alleles at homozygous parental genotypes would correspond to a more general strategy.

For a given set of loci, optimization of parentage assignment protocols at a parental level is mainly achievable when microsatellite typing concerns a limited number of actual genitors from which a subset will be crossed to estimate the genetic heritability of a trait or to test for differential sire-group expression of a trait, e.g., survival rate (Doyle et al. 1995). Because the difficulties of assigning parents is dependent upon how common the parent's microsatellite profile is, large variance of  $f_{gu}$  is expected depending on the subset of parents used for the crossing. To illustrate this feature, the distribution of individual  $f_{gu}$  values was computed in turbot from the actual genotypes of 12 males and 9 females assuming a  $1F \times 8M$  mating scheme and a 5% error rate with three loci. For each of the 1000 drawings of  $1F \times 8M$  parental mating schemes, a  $f_{gu}$  value was estimated over 10 000 offspring drawings. Mean  $f_{gu}$  was 0.858 with a standard deviation of 0.049. A large difference was obtained between the lowest  $f_{gu}$  (0.736) and the highest  $f_{gu}$  (0.961). Hence, by running a large number of such simulations or using an appropriate optimization algorithm on actual parental genotypes, it should be possible to optimize

parentage assignment protocols for a given set of loci by selecting parental combinations associated with high  $f_{gu}$  values.

### Acknowledgments

We are indebted to Elisabeth Lebegue, Michel Guennoc (France-Turbot), and Jean-Louis Gaignon, Loic Quemener (IFREMER) for providing turbot samples; Christian Golliot (Salmonidés d'Aquitaine), Vincent Petit (Les Aquaculteurs Landais), and Cecile Vauchez (SYSAAF) for providing rainbow trout samples; Martine Adriamanga and Agnes Leroy for technical assistance; and Bernard Chevassus for stimulating discussions. This work was supported by the program "méthodologie de gestion et conservation des ressources génétiques" of the bureau des ressources génétiques and the Biotech-EC program PL920430.

### References

- Allendorf, F.W. 1975. Genetic variability in a species possessing extensive gene duplication: genetic interpretation of duplicate loci and examination of genetic variation in populations of rainbow trout. Ph.D. thesis, University of Washington, Seattle, Wash.
- Allendorf, F.W., and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonid fishes in the evolutionary genetics of fishes. In *Evolutionary genetics of fishes*. Edited by J. Turner. Plenum Publishers, New York. pp. 1-53.
- Berg, W.J., and Gall, G.A.E. 1988. Gene flow and genetic differentiation among California coastal rainbow trout populations. *Can. J. Fish. Aquat. Sci.* **45**: 122-131.
- Blanquer, A., Alayse, J.P., Berrada-Rkhami, O., and Berrebi, P. 1992. Allozyme variation in turbot (*Psetta maxima*) and brill (*Scophthalmus rhombus*) (Osteichthyes, Pleuronectiformes, Scophthalmidae) throughout their range in Europe. *J. Fish Biol.* **41**: 725-736.
- Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* **32**: 314-331.
- Bouza, C., Sanchez, L., and Martinez, P. 1997. Gene diversity analysis in natural populations and cultured stocks of turbot (*Scophthalmus maximus* L.). *Anim. Genet.* **28**: 28-36.
- Brookfield, J.F.Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol. Ecol.* **5**: 453-455.
- Busack, C.A., Halliburton, R., and Gall, G.A.E. 1979. Electrophoretic variation and differentiation in four strains of domesticated rainbow trout (*Salmo gairdneri*). *Can. J. Genet. Cytol.* **21**: 81-94.
- Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H., Richards, R.I., Mulley, J.C., and Sutherland, G.R. 1993. Incidence and origin of "nul" alleles in the (AC) $n$  microsatellite markers. *Am. J. Hum. Genet.* **52**: 922-927.
- Doyle, R.W., and Herbinger, C. 1994. The use of DNA fingerprinting for high-intensity, within-family selection in fish breeding. In *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*. Vol. 19. Department of Animal and Poultry Sciences, University of Guelph, Guelph, Ont.
- Doyle, R.W., Herbinger, C., Taggart, C.T., and Lochmann, S. 1995. The use of DNA microsatellite polymorphism to analyze genetic correlations between hatchery and natural fitness. *Am. Fish. Soc. Symp.* **15**: 205-211.
- Estoup, A., and Cornuet, J.-M. 1994. Utilisation de sondes oligonucléotidiques marquées à la digoxigénine pour la recherche de microsatellites. *J. Biotechnol.* **10**: 5-7.
- Estoup, A., Presa, P., Krieg, F., Vaiman, D., and Guyomard, R. 1993.



- (CT)<sub>n</sub> and (GT)<sub>n</sub> microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity*, **71**: 488–496.
- Estoup, A., Largiadèr, C.R., Perrot, E., and Chourrout, D. 1996. One-tube rapid DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Mol. Mar. Biol. Biotechnol.* **5**: 295–298.
- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**: 87–112.
- Falconer, D.S. 1981. Introduction to quantitative genetics. 2nd ed. Longman, London.
- Guyomard, R. 1981. Electrophoretic variation in four French populations of domesticated rainbow trout (*Salmo gairdneri*). *Can. J. Genet. Cytol.* **23**: 33–47.
- Gyllenstein, U. 1985. The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous and freshwater species. *J. Fish Biol.* **26**: 691–699.
- Hanset, R. 1975. Probabilité d'exclusion de paternité et de monozygote, probabilité de similitude. Généralisation à *N* alleles co-dominants. *Ann. Med. Vet.* **119**: 71–80.
- Harris, A.S., Bieger, S., Doyle, R.W., and Wright, J.M. 1991. DNA fingerprinting of tilapia *Oreochromis niloticus* and its application to aquaculture genetics. *Aquaculture*, **92**: 157–163.
- Herbinger, C.M., Roger, W.D., Pitman, R.P., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M., and 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.
- Hill, W.G., and Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**: 226–231.
- Hochberg, Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**: 800–802.
- Morris, D.B., Richard, K.R., and Wright, J.M. 1996. Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for genetic study of salmonids. *Can. J. Fish. Aquat. Sci.* **53**: 120–126.
- Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
- Olsen, J.B., Wenburg, J.K., and Bentzen, O. 1996. Semiautomated multilocus genotyping of pacific salmon (*Oncorhynchus* spp.) using microsatellites. *Mol. Mar. Biol. Biotechnol.* **5**: 259–272.
- Pemberton, J.M., Slate, J., Bancroft, D.R., and Barrett, J.A. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.* **4**: 249–252.
- Raymond, M., and Rousset, F. 1995. Population genetics software for exact test and ecumenism. *J. Hered.* **86**: 248–249.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution*, **43**: 223–225.
- Sakamoto, T., Okamoto, N., and Ikeda, Y. 1994a. Dinucleotide repeat polymorphism of rainbow trout, FGT2. *J. Anim. Sci.* **72**: 2765.
- Sakamoto, T., Okamoto, N., and Ikeda, Y. 1994b. Dinucleotide repeat polymorphism of rainbow trout, FGT3. *J. Anim. Sci.* **72**: 2766.
- Sakamoto, T., Okamoto, N., and Ikeda, Y. 1994c. Dinucleotide repeat polymorphism of rainbow trout, FGT4. *J. Anim. Sci.* **72**: 2767.
- SanCristobal, M., and Chevalet, C. 1997. Error tolerant parent identification from a finite set of individuals. *Genet. Res. (Cambridge)*, **70**: 53–62.
- Smouse, P.E., and Chakraborty, R. 1986. The use of restricted fragment length polymorphisms in paternity analysis. *Am. J. Hum. Genet.* **38**: 918–939.
- Sprent, P. 1989. Applied nonparametric statistical methods. Chapman & Hall, London.
- Weber, J.L. 1990. Informativeness of human (dC–dT)<sub>n</sub>(dG–dA)<sub>n</sub> polymorphisms. *Genomics*, **7**: 524–530.
- Wenburger, J.K., Olsen, J.B., and Bentzen, P. 1996. Multiplexed systems of microsatellites for genetic analysis in costal cutthroat trout (*Oncorhynchus clarki clarki*) and steelhead (*Oncorhynchus mykiss*). *Mol. Mar. Biol. Biotechnol.* **5**: 273–283.
- Wright, J.M., and Bentzen, P. 1994. Microsatellites: genetic markers for the future. *Rev. Fish Biol. Fish.* **4**: 384–388.