Aquaculture
April 2011, Volume 314, Issues 1-4, Pages 80-86
http://dx.doi.org/10.1016/j.aquaculture.2011.01.043

From theory to practice: Empirical evaluation of the assignment power of marker sets for pedigree analysis in fish breeding

Marc Vandeputte[a, b, *], Marie-Noëlle Rossignol[c] and Cedric Pincent[d]

[a] INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78350 Jouy-en-Josas, France
[b] Ifremer, Chemin de Maguelone, F-34250 Palavas-les-Flots, France
[c] LABOGENA, Domaine de Vilvert, F-78350 Jouy-en-Josas, France
[d] SYSAAF Section Aquacole, Campus de Beaulieu, F-35000 Rennes, France

[*] Corresponding author : M. Vandeputte, Tel.: + 33 4 67 13 04 07; fax: + 33 4 67 13 04 58, email address : marc.vandeputte@jouy.inra.fr

Abstract:

The use of marker-based pedigrees is increasing in aquaculture breeding, and obtaining high assignment rates is necessary for practical use of this methodology. In this paper, we used 12 real parentage assignment datasets from three species (European sea bass *Dicentrarchus labrax*, common carp *Cyprinus carpio* and rainbow trout *Oncorhynchus mykiss*) to investigate the relationships between theoretical, simulated and real assignment power. We found out that there was a large decrease between the theoretical and the observed values, which we modeled in four independent steps: 1) from theoretical values to population-wise simulations (− 2.8% on average), 2) from population-wise simulations to true parent set specific simulations (− 2.6% on average), 3) from true parent-set specific simulations to observed values in offspring with valid genotypes at all loci (− 0.5% on average) and 4) from observed values in offspring with valid genotypes at all loci to observed values in all offspring sampled (− 2.4% on average). For all steps, we provide a regression equation which models the loss of assignment power, or at least a maximal practical value for the loss of assignment power. Finally, equations are provided to model the expected true assignment rate from the theoretical assignment power or from the combined exclusion probability of the loci used. They show that the expected true assignment rates are considerably lower than the theoretical ones, for example achieving 99% true assignment requires a theoretical assignment power of 99.999996%, while 95% already requires 99.9989% theoretical assignment rate.

Keywords: Aquaculture; Parentage assignment; Exclusion probability; Microsatellites

1. Introduction

Efficient selective breeding and estimation of genetic parameters require the knowledge of pedigrees to be able to estimate genetic variation between families and to compute reliable breeding values. As fish are too small to be tagged at hatching, the historical method to obtain this pedigree information is separate rearing of full-sibs families, which has been successfully used for breeding programmes in many species, like salmon, rainbow trout, tilapia, rohu carp, and Pacific white shrimp (see review in Hulata, 2001). Nevertheless, this method requires the use of many tanks for separate rearing of families, which represents a high investment. This method also carries the risk to bias family values with tank effects if the rearing procedures are not standardized enough. In the 1990's, the possibility arose to use highly variable markers (microsatellites) to identify the parents of an individual, provided both the candidate and its parents were genotyped for a number of markers (Herbinger, 1995; Estoup et al., 1998). This allows to get rid of the family rearing units, or even to use mass spawning events to constitute the family structure needed. Several softwares for parentage assignment have been developed (e.g. Probmax, Danzmann, 1997; FAP, Taggart, 2007; Vitassign, Vandeputte et al., 2006; CERVUS, Kalinowski et al., 2007, for the most used in aquaculture). The feasibility of using marker-based pedigrees on a large scale has been demonstrated in several species [Norris and Cunningham (2004) in salmon, Fishback et al. (2002) in rainbow trout, Vandeputte et al. (2004) in common carp, Vandeputte et al. (2007) in European sea bass, Gheyas et al. (2009) in silver carp], and new marker sets are continuously being optimized for virtually every aquaculture species targeted for selective breeding. One of the most popular parameters to evaluate the efficiency of a given marker for parentage assignment is the exclusion probability, which is the probability of a randomly chosen parent-pair being genetically excluded as parents of a randomly chosen offspring, in case that parent pair did not produce that offspring (Dodds et al., 1996; Villanueva et al., 2002). The exclusion probability can easily be estimated for each locus using the observed frequencies of the different alleles in the population. By combining the exclusion probabilities of the different loci in the marker set, it is possible to compute the combined probability of exclusion and to predict the assignment power of the marker set in the population (Villanueva et al., 2002). However, this prediction tends to be overly optimistic, especially when parents can contribute to several full-sibs families (as in nested or factorial mating designs, Villanueva et al., 2002). In aquaculture breeding, this is generally the case, as the use of parentage assignment allows the use of factorial mating designs, which are beneficial for genetic gain and conservation of genetic variance (Dupont-Nivet et al., 2006; Busack and Knudsen, 2007). This *de facto* generates contributions of the same parents to many different full-sib families. Many other factors like relatedness of parents, unequal family sizes, selection (which can generate both relatedness and unequal family sizes), genotyping errors, null alleles, mutations, linkage of markers, random allelic associations among loci, can also contribute to lower the assignment success of a marker set (Villanueva et al., 2002; Jones and Ardren, 2003; Vandeputte et al., 2006; Matson et al., 2008). Therefore, it is recommended to use simulations to assess the assignment power of marker sets, and some softwares allow this [Vitassign, Vandeputte et al. (2006), Cervus, Kalinowski et al. (2007), FAP, Taggart (2007), P-Loci, Matson et al. (2008)]. However, these simulations do not always take into account genotyping errors (Vitassing, FAP), null alleles or linkage (Vitassign, Cervus, FAP), or they require an estimated value for null alleles and/or genotyping errors (Cervus, P-Loci) which is not always possible to obtain. Additionally, the assignment success may depend on the assignment method chosen (exclusion for Vitassign, FAP and P-Loci, maximum likelihood for Cervus).

When using parentage assignment in selective breeding, it is extremely important to obtain high and reliable assignment rates, as the cost of rearing and phenotyping candidates is high and every non assigned fish represents a net loss. In order to lower the cost of genotyping, it is often

proposed to use the minimal number of loci which are able to assign a given proportion of individuals. This is usually done using combined probabilities of exclusion (e.g. Lemos et al., 2006; Slabbert et al., 2009) or simulations (e.g. Jerry et al., 2004, 2006; Dong et al., 2006). In some cases, when tested, it appears that the observed assignment rate from real field data is well below the expected value (Jerry et al., 2004; Dong et al., 2006; Slabbert et al., 2009). Then, it would be advisable to use more reliable estimates of the real assignment power of microsatellites. The use of multiplexed marker sets permits a significant decrease in the cost of analyses, allowing the use of more markers (Renshaw et al., 2006), but on the other hand, the development of a multiplex for commercial use is more complex than genotyping individual loci, so the multiplex should give high assignment rates in all circumstances, and if possible also in populations other than the one used to develop it.
In this paper, we used accumulated published and unpublished information on genotyping of three species (rainbow trout, common carp, European sea bass), using different marker sets, in different laboratories, in different populations, to empirically study the differences between expected and real assignment rates, with the aim to propose decision rules on the number of markers to use to obtain a given level of assignment in practice. We tried to relate observed assignment power to the probability of exclusion, which has been demonstrated to be an adequate measurement of the power of a marker set in parentage assignment by exclusion (Wang, 2007)

2. Material and methods

2.1. Animals and marker frequency data

Sea bass, rainbow trout and common carp from several selection and/or genetic parameters estimation experiments were used (Table 1). All matings studied were full (FF) or partial (FS) factorial matings, with all parents known and genotyped. Depending on the experiments, animals were genotyped by different labs (commercial and research) for different sets of microsatellite markers. For each experiment, the type of broodstock (wild, domesticated, selected) is indicated (Table 1).

2.2. Estimation of assignment power

For each marker set in each population studied, allele frequencies for each marker were calculated from parental genotypes. These frequencies were used to compute parent pair exclusion probabilities at each locus, then the combined probability of exclusion across all loci for infinite numbers of parents and offspring, as done in Villanueva et al. (2002). Then, the probability P_{th} (theoretical assignment power) that an offspring is assigned to a single (exact) parent pair was computed, using *Nm*Nf* as the number of potential parent pairs (*Prob(0)* in Villanueva et al., 2002):
$P_{th}=P_{excl}^{(Nm*Nf-1)}$ (Equation 1)
Then, a simulation was run to estimate the average parentage assignment rate in a different way: for each simulation, the genotypes of *Nm* sires and *Nf* dams were randomly drawn using the observed allelic distribution (derived from parent frequencies). For each mating plan, 1000 offspring were randomly generated, considering equal probabilities to descend from any sire and dam in the mating plan. Due to sampling, family sizes were then not equal but had equal expectations, following

a Poisson distribution. For each offspring, the alleles transmitted by the sire and the dam were randomly chosen. Then, the offspring genotype was compared to all parental genotypes using the exclusion routine from VITASSIGN (Vandeputte et al., 2006) and the proportion of uniquely assigned offspring was computed. For each cross and marker set, this was repeated 100 times, so that the final assignment power estimated (P_{sim1}) was the mean of 100 simulated parent sets, each with 1000 offspring. We also recorded the minimum (P_{min1}) and maximum (P_{max1}) assignment power out of 100 repetitions for each cross and marker set.

Then, for each true parental set in Table 1, 2000 to 20000 randomly produced offspring were simulated with VITASSIGN to estimate the probability of unique assignment P_{sim2} with the true parents. Finally, the true genotyped offspring (genotyped for all loci) from the cross were assigned to their parents using VITASSIGN, either with perfect match (P_{exact}) or with one mismatch (P_{1msm}), to account for genotyping errors (Vandeputte et al., 2006; Christie, 2010). The proportion of offspring assigned to several parent pairs (P_{poly}) or unassigned when allowing for one mismatch (P_{unas}) were also computed, in order to better describe the reasons why some offspring were not assigned to a unique parent pair. If P_{poly} is high, this is indicative of a lack of power in the marker set, while if P_{unas} is high, it means that there are genotyping errors (Vandeputte et al., 2006; Christie, 2010). Finally, as not all offspring were genotyped for all loci due to amplification problems or missing samples, the proportion of assigned offspring from the full offspring sample (including those with incomplete genotypes) with one mismatch tolerated was also estimated (P_{all}).

2.3. Statistical analyses

The simulated and observed proportions of non-uniquely assigned offspring ($1\text{-}P_x$) from the different studies described in Table 1 were computed, and log-transformed using a natural log. Then, these log-transformed proportions were compared with linear regression using SAS-REG (The SAS Institute, Cary, NY). Log-transformed proportions were used to linearize proportions in the vicinity of zero, as in most studies, due to the use of appropriate marker sets, the proportions of non-uniquely offspring were low to very low.

3. Results

3.1. Theoretical and simulated values for each marker set

The combined probabilities of exclusion were very high for all marker sets, the lowest being 0.999965 for cross 5 with marker set SB5, while the highest was 0.999999999986 for cross 1b with marker set SB2 (Table2). The theoretical proportions of uniquely assigned offspring (P_{th}) were also very high, all higher than 99%. The simulated assignment rates (P_{sim1}) based on the allelic frequencies in the populations were also high in most cases, but always lower than the theoretical assignment rates (-0.0003% to -13%, -2.8% on average).

3.2 Observed values for realized crosses

The observed exact assignment rates in the crosses studied were generally high for offspring with fully exploitable genotypes (>90% in nine cases out of 12) or even very high (>95% in six cases – Table 3). The allowance for one mismatch generally increased the assignment rate, but to a limited level (less than 3% increase). The proportion of offspring assigned to several parent pairs was highly variable, from 0.02% to 18.81%. The proportion of unassigned offspring was generally low (<2%). Not all offspring had fully exploitable genotypes, and hence the assignment power when considering all offspring (P_{all}) was lower in many cases.

3.3. Comparisons between observed and predicted assignment power

Although the simulated assignment power was well below the theoretical one (from -0.0003% to -13%, -2.8% on average), the relation between both was linear (in a log-log plot) and highly significant (Figure 1):
$ln(1-P_{sim1}) = 0.547\ ln(1-P_{th})$, $R^2_{adj}=0.995$ (Equation 2)
Similarly, the simulated assignment power done by Vitassign from the true parental genotypes was generally moderately higher than the observed assignment rate in offspring with a valid genotype at each locus (on average -0.5%, varying from +1.6% to -1.6%).In this case, the observed assignment power was estimated with one mismatch tolerated, only in offspring with valid genotypes at all loci. Both figures were highly correlated (Figure 2):
$ln(1-P_{1msm}) = 0.908\ ln(1-P_{sim2})$, $R_{adj}^2=0.994$ (Equation 3)
On the contrary, the simulated assignment power done on the real parents (P_{sim2}) was not tightly connected (Figure 3) with the simulated assignment power done on the whole population (P_{sim1}) :
$ln(1-P_{sim2}) = 0.707\ ln(1-P_{sim1})$, $R^2_{adj}=0.770$ (Equation 4)
In three cases out of twelve cases P_{sim2} was higher than P_{sim1} (up to +5.1%), but it was generally lower (-2.6% on average, down to -11%)

However, in all but two cases, the minimum value simulated out of 100 replicated simulations (P_{min}) was close to or below the assignment power simulated from the true parent set (P_{sim2}), and could be well described by a linear regression on P_{sim1} (Figure 3):
$ln(1-P_{min})= 0.729\ ln(1-P_{sim1})$, $R^2_{adj}=0.991$ (Equation 5)
It should be noted that in cases where parents had been subject to phenotypic selection (crosses 1a, 1b, 6, 7, 9, 10), P_{sim2} was close to P_{min}, or even lower for crosses 1a and 1b, while when wild parents were used (crosses 2, 3, 4, 5) P_{sim2} was in general much higher than P_{min}.
In general, more than 90% of the offspring had fully exploitable genotypes (Table 3), but in some cases this proportion was lower, mostly when the marker sets had numerous loci. The causes were missing samples, lack of amplification at one or several loci, or multi-allelic loci (due to spontaneous triploids or contamination of samples). Still, some offspring with incomplete genotypes could be assigned to a single parental pair, but when considering all offspring, the assignment rate P_{all} was lower than when considering only offspring with valid genotypes at all loci (-2.4% on average, from zero to -4.9%). The regression of $ln(1-P_{all})$ on $ln(1-P_{1msm})$ was rather good but not excellent (Figure 4):
$ln(1-P_{all})= 0.737\ ln(1-P_{1msm})$, $R^2_{adj}=0.932$ (Equation 6)
Except in two cases, this regression could be considered as a maximum value for $ln(1-P_{all})$

4. Discussion

In the cases studied, the theoretical power of the marker sets used was always very high (>99%), but it appeared that the simulated and true assignment rates, while in most cases high, could also be insufficient (<95%). The first loss of power appeared when moving from the theoretical assignment power (P_{th}) to an assignment power in simulated samples (P_{sim1}) from the population. A first reason for this can be the fact that P_{excl} and P_{th} are derived assuming Hardy-Weinberg equilibrium in the populations, which may not be the case in a real population. However, this effect is usually small (Wang, 2007). Differences between P_{th} and P_{sim1} were already noticed by Vilanueva et al (2002) mostly for nested and factorial mating designs, as opposed to single pair matings, and were linked to the fact that parent pairs in such designs are not independent samples. Here, we used only factorial mating designs, as it is *de facto* the type of mating design considered behind all assignment programs. Knowledge of an *a priori* breeding scheme can exist (and be used) in the case of controlled mating designs, but assignment without restriction to the full factorial between all sires and dams is the best way to identify potential mistakes in the planned mating scheme. Moreover, when family assignment is used in mass spawning events (e.g. Perez-Enriquez et al., 1999; Chatziplis et al., 2007; Herlin et al., 2007;), there is no planned breeding scheme and the factorial is the appropriate model of analysis. Factorial designs are also the best design to keep genetic variance and enhance genetic gain for a given number of parents when the number of families is not constrained (Dupont-Nivet et al., 2006). Then, we consider that the use of marker assisted parentage assignment should be planned in the case of factorial mating designs. Estimating the simulated assignment

power in a population, based on allelic frequencies can be done either using simulation programs: the present one SimExPo (available on request from marc.vandeputte@jouy.inra.fr), CERVUS (Kalinowski et al., 2007), FAP (Taggart, 2007) or P-loci (Matson et al., 2008), or using Equation 2 to derive it from the theoretical assignment power calculated from Villanueva et al. (2002).

The second stage of loss of assignment power was between the assignment power simulated from the allelic frequencies in the population (P_{sim1}) and the assignment power simulated from the true parent set (P_{sim2}). The loss of assignment power here can be linked both to sampling variance of the parents, but also to the fact that the parents are related and share more common alleles than independent random samples would do (Villanueva et al., 2002; Matson et al., 2008). Small effective population sizes and selection of parents can both increase this phenomenon. In order to take into account the sampling variance effect, we compared P_{sim2} with the minimum and the maximum of 100 replicated simulations from parents randomly drawn in the population (Figure 3). In most cases the values of P_{sim2} fell within or close to the area between by P_{min} and P_{max}, except for two points (crosses 1a and 1b) which represent two offspring samples from the same mating design, analyzed with two marker sets. Interestingly, in this case, three quarters of the male parents were potentially sibs (derived from the same parents), and the remaining quarter was a sub-sample of their fathers. Moreover, half of the male parents had been subjected to strong phenotypic selection on growth (5% pressure, see details in Vandeputte et al., 2009), so in this case we had all ingredients to have non-random distribution of alleles among the true parents. Although it was not possible to reliably predict P_{sim2} from P_{sim1}, there was a very good linear relationship between $ln(1-P_{sim1})$ and $ln(1-P_{min})$, and except for the case of crosses 1a and 1b, P_{min} was always close to or smaller than P_{sim2}, and could then be used to estimate P_{sim2} in a conservative way (Equation 4). P_{min} was close to P_{sim2} in the case of selected parents, and was generally much lower than P_{sim2} when wild parents were used. Therefore, for aquaculture applications where parentage assignment is used mostly for selective breeding programs, the prediction using Equation 4 seems appropriate, while it might be overly conservative for wild populations studies.

The third stage of loss of assignment power is between the simulated assignment power from the real parental set (P_{sim2}) and the realized assignment power, either without (P_{exact}) or with (P_{1msm}) one mismatch tolerated. At this stage, the differences between the realized values and the simulated ones should essentially come from genotyping errors *sensu lato* (misreading of genotypes, mutations, null alleles). Genotyping errors are expected to 1) generate some unassigned offspring and 2) generate a difference between P_{exact} and P_{1msm} (Vandeputte et al., 2006; Christie, 2010). This is what we see in most cases in Table 3, with moderate increases of assignments from P_{exact} to P_{1msm}, and small proportions of unassigned offspring. This is indicative of a significant, but low level of genotyping errors (Vandeputte et al., 2006; Christie, 2010). In the cases studied, null alleles were sometimes present but at a low level, as the absence of null alleles was one of the technical criteria used to select loci for their inclusion in the marker sets. However, this may not always be possible, as in some organisms (e.g. mollusks - see Hedgecock et al., 2004) null alleles may be very frequent and variable among populations. In these cases, larger decreases between P_{sim2} and P_{exact} or P_{1msm} could be expected. At that stage, another possible source of

variation in assignment rates is the uneven distribution of family sizes (Taggart, 2007). Simulations make the hypothesis that all parents have the same expected number of offspring, which is never the case in practice (see Perez-Enriquez et al., 1999; Chatziplis et al., 2007; Herlin et al., 2007 for examples).

The last stage of assignment power loss is between the offspring which have complete genotypes at all loci and those with only partial genotypes. Amplification problems are quite frequent, but usually at a low rate as ease and reliability of amplification are selection criteria for including a locus in a parentage assignment suite. Nevertheless, this problem is expected to be more frequent when many markers are included in the suite. Other possible causes for partial or unexploitable genotypes are missing samples, contamination of samples and spontaneous triploids. Adequate procedures can reduce contamination of samples and missing samples, but amplification problems will always be possible.

Finally, we propose to combine all four sources of loss of assignment power by combining equations 2, 3, 5 and 6, using $ln(1-P_{min})$ as a maximum estimate of $ln(1-P_{sim2})$, and then set up a prediction equation for a minimum value of P_{all} based on P_{th}:

$ln(1-P_{all})= 0.908*0.737*0.735*0.547\ ln(1-P_{th})= 0.269\ ln(1-P_{th})$

Which is equivalent to $P_{all}= 1-(1-P_{th})^{0.269}$ (Equation 7)

or $P_{th}= 1-(1-P_{all})^{3.7}$ (Equation 8).

When undertaking scientific studies like heritability estimates, sub-optimal assignment rates can be used, as long as they do not generate biased family contributions, and are not too low. Some valuable results have been published with moderate assignment rates [e.g. 66.9% in Blonk et al. (2010), 75.7% in Kocour et al. (2007)], but in industry conditions, genotyping is a high operating cost and low assignment rates are not tolerated. In this case, more than 99% assignment should be targeted (see Navarro et al., 2008).

Then, the targeted true assignment rate can be fixed, and the required theoretical assignment power can be computed using Equation 8. The difference between the theoretical and the expected true assignment rate can be considerable: according to this calculation, achieving 99% true assignment requires a theoretical assignment power of 99.999996%, while 95% already requires 99.9989% theoretical assignment rate. The relationship between the targeted assignment power and the combined exclusion probability can also be derived from Equation 8 and Equation 1, and approached by $P_{excl} \sim 1 - \frac{(1-P_{all})^{3.7}}{Nm.Nf}$ (Equation 9), where *Nm* and *Nf* are the projected numbers of sires and dams in the matings to be analyzed. Similarly, if one considers that availability of samples and amplification should be technically solved, another target value can be $P_{excl} \sim 1 - \frac{(1-P_{1msm})^{2.7}}{Nm.Nf}$ (Equation 10), ignoring coefficient 0.737 from Equation 6. Using this latter relationship with the data of Navarro et al (2008), which used fully genotyped seabream offspring, we found a predicted assignment rate of 47 % for P_{excl}=0.9998 and 38% for P_{excl}=0.9997, while the observed values were 62 and 37%, respectively, which is quite satisfactory. Unfortunately, it could not be tested with higher P_{excl} values, as not enough decimals are available in the paper.

Some cases have not been studied in this paper, like highly inbred animals in very small matings, but we feel the cases studied here are a good sample of what can be done in

fish breeding, and therefore the recommendations done in this paper should help designing appropriate marker sets for future parentage assignment studies. It should be noted that P_{excl} is a population specific parameter, so if a marker set is designed to be used in several populations, it will be advisable to estimate P_{excl} in all populations to choose the appropriate number of markers. Another point is that during the course of a breeding programme, it is expected that inbreeding will increase and genetic variability will decrease. Therefore, assignment power should decrease over time (Villanueva et al., 2002). Here, some of the cases studied involve selected populations (1a, 1b, 6, 7, 9, 10) and do not seem to behave differently from the others. However, it cannot be excluded that on the long run, or in breeding programmes with low broodstock number and/or high selection intensities, assignment rates would progressively become lower than expected.

Conclusion

The analysis proposed in the present paper showed that using theoretical assignment power when designing a marker set for parentage assignment leads to overly optimistic predictions. The equations proposed allow some corrections to use assignment power values more representative of what will happen in reality. Nevertheless, as genotyping methods and multiplexing are more and more efficient, it would be advisable to use these values as a baseline, and if possible to include a few excess markers to guarantee the highest assignment rates under all circumstances.

Acknowledgements

This work was carried out using parentage assignment data from several projects (EU projects Heritabolum and Competus, national projects Qualitytruite2 and Vegeaqua, Joint INRA-Ifremer project “Genetic improvement of fish”). The authors wish to thank the following persons for providing and granting access to the data: Amandine Launay, Stéphane Mauger, Béatrice Chatain, Mathilde Dupont-Nivet, Richard Le Boucher, Laure Grima, Pierrick Haffray, Vincent Petit, Frédéric Cachelou, Alastair Hamilton, Hervé Chavanne, Katia Parati, Silvia Cenadelli, Otomar Linhart, Martin Kocour, Marie-Yvonne Boscher, Céline Chantry-Darmon, Jean-Claude Mériaux

References

Blonk, R.J.W., Komen, H., Kamstra, A., Van Arendonk, J.A.M., 2010. Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. Genetics 184, 213-219.

Busack, C., Knudsen, C.M., 2007. Using factorial mating designs to increase the effective number of breeders in fish hatcheries. Aquaculture 273, 24-32.

Chatziplis, D., Batargias, C., Tsigenopoulos, C.S., Magoulas, A., Kollias, S., Kotoulas, G., Volckaert, F.A.M., Haley, C.S., 2007. Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): The BASSMAP pilot study. Aquaculture 272, S172-S182.

Christie, M.R., 2010. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. Mol. Ecol. Resour. 10, 115-128.

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

Dodds, K.G., Tate, M.L., McEwan, J.C., Crawford, A.M., 1996. Exclusion probabilities for pedigree testing farm animals. Theor. Appl. Genet. 92, 966-975.

Dong, S.R., Kong, J., Zhang, T.S., Meng, X.H., Wang, R.C., 2006. Parentage determination of Chinese shrimp (*Fenneropenaeus chinensis*) based on microsatellite DNA markers. Aquaculture 258, 283-288.

Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., Chatain, B., 2008. Heritabilities and GxE interactions for growth in the European sea bass (*Dicentrarchus labrax* L.) using a marker-based pedigree. Aquaculture, 81-87.

Dupont-Nivet, M., Vandeputte, M., Haffray, P., Chevassus, B., 2006. Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. Aquaculture 252, 161-170.

Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P., Guyomard, R., 1998. Parentage assignment using microsatellites in turbot (*Scophtalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. Can. J. Fish. Aquat. Sci. 55, 715-725.

Fishback, A.G., Danzmann, R.G., Ferguson, M.M., Gibson, J.P., 2002. Estimates of genetic parameters and genotype by environment interactions for growth traits of the rainbow trout (*Oncorhynchus mykiss*) as inferred using molecular pedigrees. Aquaculture 206, 137-150.

Gheyas, A.A., Woolliams, J.A., Taggart, J.B., Sattar, M.A., Das, T.K., McAndrew, B.J., Penman, D.J., 2009. Heritability estimation of silver carp (*Hypophthalmichthys molitrix*) harvest traits using microsatellite based parentage assignment. Aquaculture 294, 187-193.

Grima, L., Chatain, B., Ruelle, F., Vergnet, A., Launay, A., Mambrini, M., Vandeputte, M., 2010. In search for indirect criteria to improve feed utilization efficiency in sea bass (*Dicentrarchus labrax*): Part II: Heritability of weight loss during feed deprivation and weight gain during re-feeding periods. Aquaculture 302, 169-174.

Hedgecock, D., Li, G., Hubert, S., Bucklin, K., Ribes, V., 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, Crassostrea gigas. J. Shellfish Res. 23, 379–385.

Herbinger, C.M., 1995. DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communaly reared rainbow trout. Aquaculture 137, 245-256.

Herlin, M., Taggart, J.B., McAndrew, B.J., Penman, D.J., 2007. Parentage allocation in a complex situation: A large commercial Atlantic cod (*Gadus morhua*) mass spawning tank. Aquaculture 272, S195-S203.

Hulata, G., 2001. Genetic manipulations in aquaculture: a review of stock improvement by classical and modern methodologies. Genetica 111, 155-173.

Jerry, D.R., Evans, B.S., Kenway, M., Wilson, K., 2006. Development of a microsatellite DNA parentage marker suite for black tiger shrimp *Penaeus monodon*. Aquaculture 255, 542-547.

Jerry, D.R., Preston, N.P., Crocos, P.J., Keys, S., Meadows, J.R.S., Li, Y., 2004. Parentage determination of Kuruma shrimp *Penaeus* (*Marsupenaeus*) *japonicus* using microsatellite markers. Aquaculture 235, 237-247.

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

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program Cervus accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16, 1099-1106.

Kocour, M., Mauger, S., Rodina, M., Gela, D., Linhart, O., Vandeputte, M., 2007. Heritability estimates for processing and quality traits in common carp (*Cyprinus carpio* L.) using a molecular pedigree. Aquaculture 270, 43-50.

Lemos, A., Freitas, A.I., Fernandes, A.T., Gonzalves, R., Jesus, J., Andrade, C., Brehm, A., 2006. Microsatellite variability in natural populations of the blackspot seabream *Pagellus bogaraveo* (Brünnick, 1768): a database to access parentage assignment in aquaculture. Aquacult. Res. 37, 1028-1033.

Matson, S.E., Camara, M.D., Eichert, W., BANKS, M.A., 2008. P-loci: a computer program for choosing the most efficient set of loci for parentage assignment. Mol. Ecol. Resour. 8, 765-768.

Navarro, A., Badilla, R., Zamorano, M.J., Pasamontes, V., Hildebrandt, S., S nchez, J.J., Afonso, J.M., 2008. Development of two new microsatellite multiplex PCRs for three sparid species: Gilthead seabream (*Sparus auratus* L.), red porgy (*Pagrus pagrus* L.) and redbanded seabream (*P. auriga*, Valenciennes, 1843) and their application to paternity studies. Aquaculture 285, 30-37.

Norris, A.T., Cunningham, E.P., 2004. Estimates of phenotypic and genetic parameters for flesh colour traits in farmed Atlantic salmon based on multiple trait animal model. Livest. Prod. Sci. 89, 209-222.

Perez-Enriquez, R., Takagi, M., Taniguchi, N., 1999. Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. Aquaculture 173, 413-423.

Renshaw, M.A., Saillant, E., Bradfield, S.C., Gold, J.R., 2006. Microsatellite multiplex panels for genetic studies of three species of marine fishes: red drum (*Sciaenops*

ocellatus), red snapper (*Lutjanus campechanus*), and cobia (*Rachycentron canadum*). Aquaculture 253, 731-735.

Slabbert, R., Bester, A.E., D'Amato, M.E., 2009. Analyses of genetic diversity and parentage within a South African hatchery of the abalone *Haliotis midae* Linnaeus using microsatellite markers. J. Shellfish Res. 28, 369-375.

Taggart, J.B., 2007. FAP: an exclusion-based parental assignment program with enhanced predictive functions. Mol. Ecol. Notes 7, 412-415.

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Rodina, M., Gela, D., Vallod, D., Chevassus, B., Linhart, O., 2004. Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). Aquaculture 235, 223-236.

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

Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Chatain, B., 2007. A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. Genetics 176, 1049-1057.

Vandeputte, M., Dupont-Nivet, M., Haffray, P., Chavanne, H., Cenadelli, S., Parati, K., Vidal, M.O., Vergnet, A., Chatain, B., 2009. Response to domestication and selection for growth in the European sea bass (*Dicentrarchus labrax*) in separate and mixed tanks. Aquaculture 286, 20-27.

Vandeputte, M., Kocour, M., Mauger, S., Rodina, M., Launay, A., Gela, D., Dupont-Nivet, M., Hulak, M., Linhart, O., 2008. Genetic variation for growth at one and two summers of age in the common carp (*Cyprinus carpio* L.): Heritability estimates and response to selection. Aquaculture 277, 7-13.

Villanueva, B., Verspoor, E., Visscher, P.M., 2002. Parental assignment in fish using microsatellite genetic markers with finite numbers of parents and offspring. Anim. Genet. 33, 33-41.

Wang, J., 2007. Parentage and sibship exclusions: higher statistical power with more family members. Heredity 99, 205-217.

Tables

Table 1: Summary of the characteristics of the populations studied. *Nm* = number of sires, *Nf*= number of dams. *#FSF*= number of expected full-sib families. *Noff*=number of offspring genotyped. Broodstock type: W=wild, D=domesticated, S= selected. #mk: number of microsatellite markers genotyped. Marker sets: numbers indicate different marker sets. All matings are full factorials except matings 3, 9 and 10 which are partial factorials. 1a and 1b are the same mating, but different offspring groups analyzed with different marker sets

Species	ID	Nm	Nf	#FSF	Noff	Broodstock type	#mk	Marker set	Reference
European sea bass	*1a*	76	*13*	*988*	*2760*	W/D/S	*8*	*SB1*	*Vandeputte et al., 2009*
	1b	76	*13*	*988*	*954*	W/D/S	*12*	*SB2*	*Unpublished*
	2	75	*26*	*1950*	*7300*	W	6	*SB1a**	*Unpublished*
	3	33	*23*	*253*	*7100*	W	*6*	*SB3*	*Dupont-Nivet et al., 2008*
	4	41	*8*	*328*	*1339*	W	*6*	*SB4*	*Grima et al., 2010*
	5	20	*2*	*40*	*587*	W	*5*	*SB5*	*Unpublished*
Common carp	*6*	147	*8*	*1176*	*812*	D/S	*8*	*CC1*	*Kocour et al., 2007*
	7	96	*8*	*768*	*797*	D/S	*8*	*CC2*	*Vandeputte et al., 2008*
	8	24	*10*	*240*	*550*	D	*10*	*CC3*	*Vandeputte et al., 2004*
Rainbow trout	*9*	100	*82*	*820*	*2004*	S	*12*	*RT1*	*Unpublished*
	10	100	*95*	*950*	*2045*	S	*12*	*RT1*	*Unpublished*
	11	25	*9*	*225*	*2016*	D	*12*	*RT1*	*Unpublished*

* SB1a is a subset of marker set SB1

Table 2: Summary of the characteristics of the marker sets used in different crosses of three species of fish for parentage assignment.

Species	*Cross ID*	Marker set	#loci	mean alleles/ locus	P_{excl}[a]	P_{th}[a]	P_{sim1}[b]
European sea bass	*1a*	SB1	8	20.1	0.999999998277	99.999830%	99.935%
	1b	SB2	12	17.3	0.999999999986	99.999999%	99.997%
	2	SB1a[c]	6	21.7	0.999999946540	99.989581%	99.342%
	3	SB3	6	20.8	0.999999952864	99.996272%	99.630%
	4	SB4	6	19.3	0.999999756135	99.992026%	99.280%
	5	SB5	5	16.3	0.999964954042	99.863411%	94.647%
Common carp	*6*	CC1	8	7.5	0.999992361778	99.106521%	86.036%
	7	CC2	8	7.6	0.999996803045	99.755000%	94.735%
	8	CC3	10	7.8	0.999995752297	99.861196%	97.116%
Rainbow trout	*9*	RT1	12	7.5	0.999999681489	99.682027%	97.080%
	10	RT1	12	6.8	0.999999664953	99.665547%	97.262%
	11	RT1	12	8.3	0.999999998621	99.999969%	99.966%

[a] P_{excl} and P_{th} are the combined probability of exclusion of the marker set and its exclusion power for the different crosses (detailed in Table 1), calculated from Villanueva et al. (2002). [b] P_{sim1} is the simulated assignment power in the same crosses using randomly drawn parents and the allelic frequencies in the populations.
[c] SB1a is a subset of SB1

Table3: Summary of the parentage assignment results in the different crosses tested, for offspring genotyped for all markers in the marker set.

Species	Cross ID	Marker set	P_{sim2} [a]	P_{exact} [b]	P_{1msm} [c]	P_{poly} [d]	P_{unass}	% full genotypes [e]	P_{all} [f]
European sea bass	*1a*	SB1	*97.52%*	*95.97%*	*96.08%*	*2.85%*	1.07%	*98.0%*	*94.1%*
	1b	SB2	*96.55%*	*94.17%*	*96.53%*	*3.47%*	0.00%	*75.6%*	*92.5%*
	2	SB1a	*97.83%*	*97.04%*	*98.26%*	*1.71%*	*0.03%*	*99.4%*	*96.5%*
	3	SB3	*99.94%*	*97.92%*	*99.82%*	*0.06%*	0.13%	*99.5%*	*97.8%*
	4	SB4	*99.82%*	*99.61%*	*99.69%*	*0.31%*	0.00%	*96.9%*	*99.4%*
	5	SB5							
			99.44%	*98.47%*	*98.47%*	*1.53%*	*0.00%*	*100.0%*	*98.5%*
Common carp	*6*	CC1	*79.90%*	*81.19%*	*81.19%*	*18.81%*	*0.00%*	*83.1%*	*75.7%*
	7	CC2	*84.25%*	*82.32%*	*83.13%*	*15.38%*	1.48%	*93.0%*	*80.3%*
	8	CC3							
			94.85%	*94.18%*	*94.18%*	*4.55%*	1.27%	*100.0%*	*94.2%*
Rainbow trout	*9*	RT1	*93.55%*	*90.74%*	*92.24%*	*6.18%*	1.58%	*86.8%*	*83.9%*
	10	RT1	*92.00%*	*89.60%*	*90.63%*	*8.96%*	0.41%	*83.1%*	*86.5%*
	11	RT1	*99.85%*	*98.13%*	*99.78%*	*0.02%*	*0.20%*	*90.0%*	*93.4%*

[a] P_{sim2} is the assignment power simulated for randomly generated offspring using the observed parental genotypes.
[b] P_{exact} is the proportion of offspring with a valid genotype at all loci assigned to a single pair using perfect exclusion
[c] P_{1msm} is the proportion of offspring with a valid genotype at all loci assigned to a single pair when allowing for one allelic mismatch.
[d] P_{poly} is the proportion of poly-assigned offspring when one mismatch is tolerated

[e] P_{unass} is the proportion of unassigned offspring when one mismatch is tolerated
[f]. %full genotypes is proportion of offspring with a fvalid genotype at all loci
[g] P_{all} is the proportion of offspring assigned to a single pair, with one mismatch tolerated, when all offspring (including those with a partial or unexploitable genotype) were included

Figures

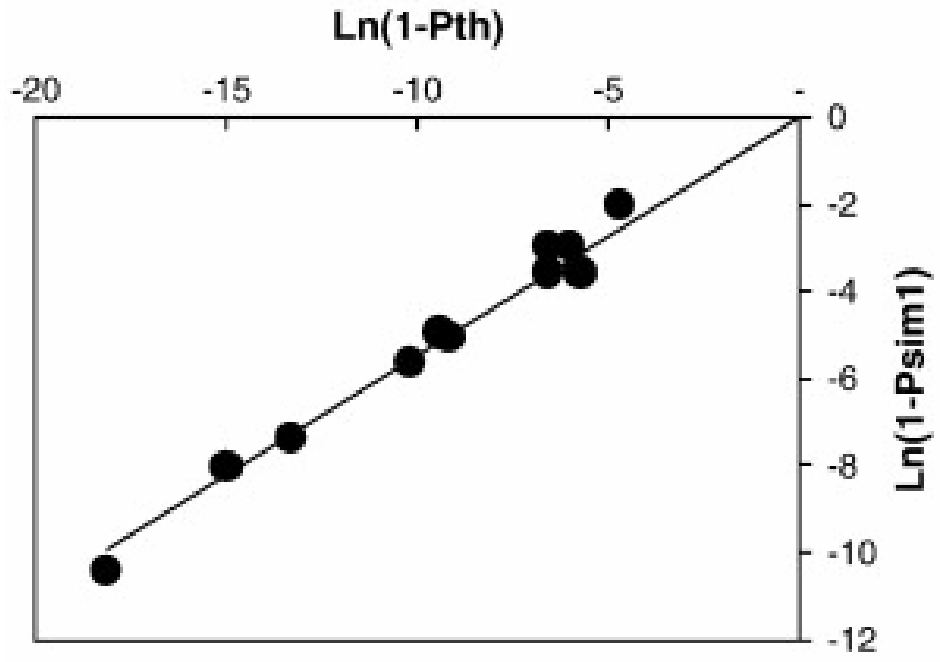

Figure 1: Logarithm of simulated proportions of not uniquely assigned offspring using alleic frequencies in the population ($1\text{-}P_{sim1}$), as a function of the logarithm of theoretical proportions of not uniquely assigned offspring ($1\text{-}P_{th}$) calculated after Villanueva et al. (2002).

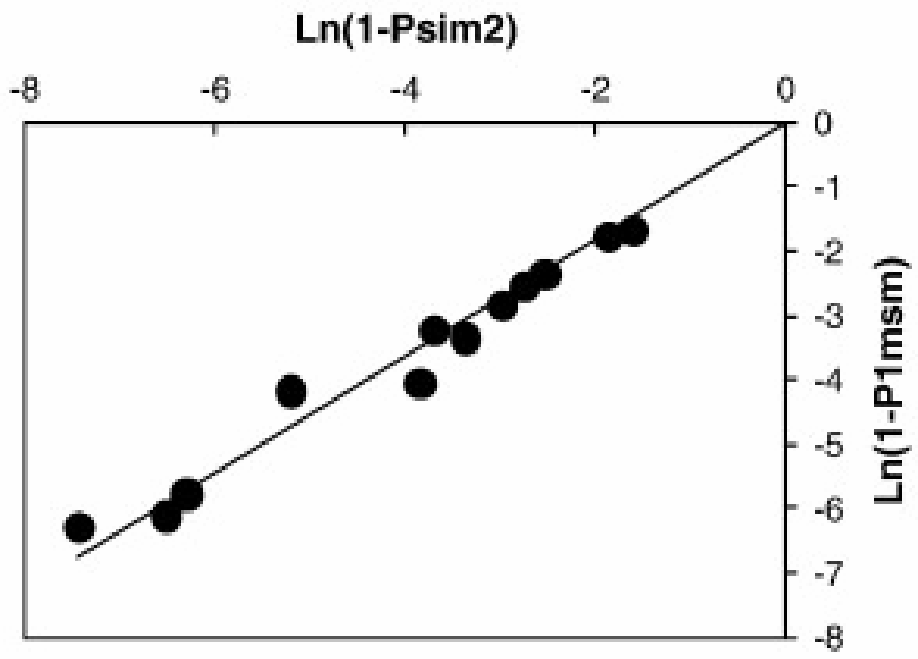

Figure 2: Logarithm of observed proportions of not uniquely assigned offspring with one mismatch tolerated ($1\text{-}P_{1msm}$), as a function of the logarithm of simulated proportions of not uniquely assigned offspring when using randomly produced offspring based on the true parental genotypes ($1\text{-}P_{sim2}$).

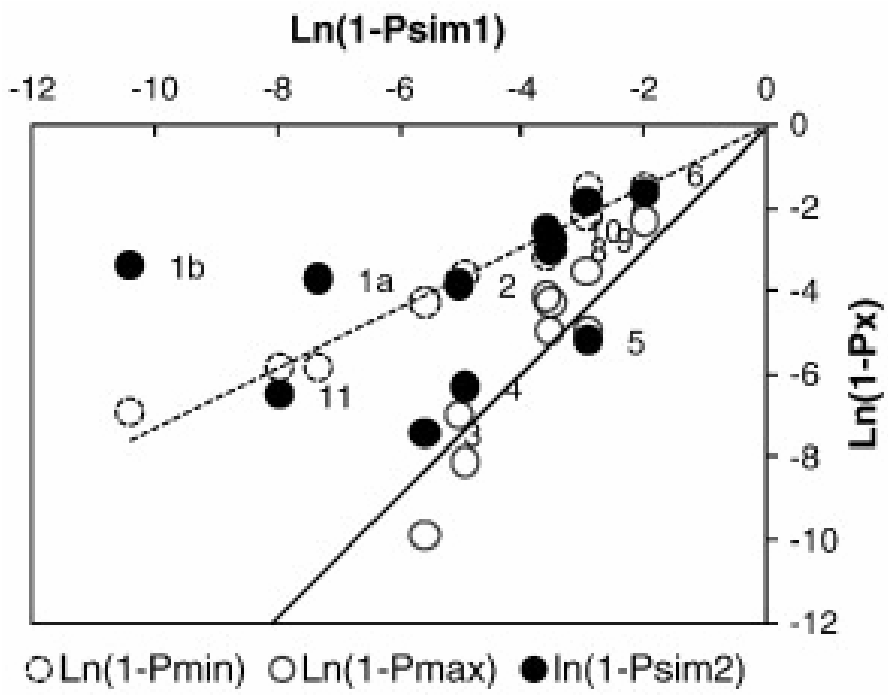

Figure 3 : Logarithms of simulated proportions of not uniquely assigned offspring, as a function of the simulated not uniquely assigned offspring based on randomly drawn parents using allelic frequencies in the population ($1\text{-}P_{sim1}$). P_{min}, P_{max}: minimum and maximum proportion of uniquely assigned offspring in 100 simulated parent sets. P_{sim2}: proportion of simulated uniquely assigned offspring from the real parental set described in Table 1, labeled with cross number. $ln(1\text{-}P_{max})$ was not plotted for the three points with $ln(1\text{-}P_{sim1})>6$, as in these cases P_{max} was equal to 1 and $\ln(1\text{-}P_{max})$ could not be computed).

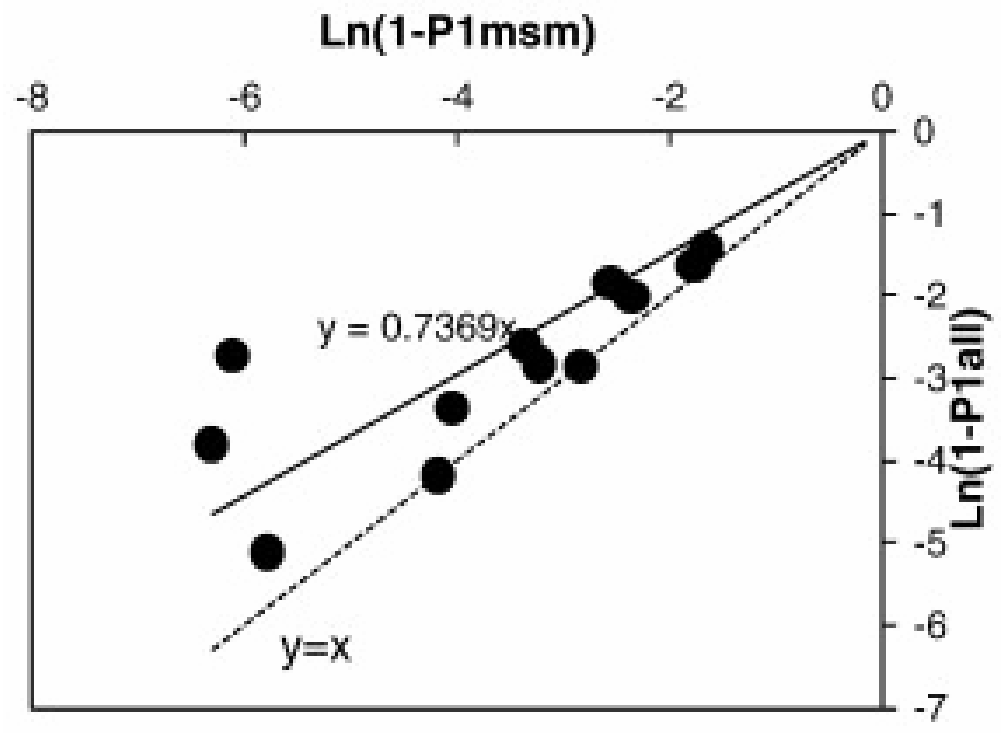

Figure 4: Logarithm of observed proportions of not uniquely assigned offspring when using all offspring ($1\text{-}P_{all}$), as a function of the logarithm of observed proportions of not uniquely assigned offspring when using fully genotyped offspring only ($1\text{-}P_{1msm}$).