

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Accuracy of pairwise methods in the reconstruction of family relationships, using molecular information from turbot (*Scophthalmus maximus*)

Silvia Teresa Rodríguez-Ramilo ^{a,*}, Miguel Ángel Toro ^a, Paulino Martínez ^b,
Jaime Castro ^b, Carmen Bouza ^b, Jesús Fernández ^a

^a Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA),
Ctra. A Coruña Km. 7,5, 28040 Madrid, Spain

^b Departamento de Genética, Facultad de Veterinaria, Universidad de Santiago de Compostela, 27002 Lugo, Spain

Received 12 June 2007; received in revised form 4 October 2007; accepted 10 October 2007

Abstract

Many estimators and algorithms have been developed to infer the genealogical relationships from molecular marker data when there is no pedigree information. Most pairwise methods provide estimates in a continuous range that needs to be converted into genealogical relationships (namely full-sibs, half-sibs and unrelated) if there is a previous knowledge of the population structure. Transformations are usually based on arbitrary thresholds, but the possibility of correcting the coancestry estimates *via* explicit pedigree reconstruction methods has been suggested. Using molecular data for ten highly polymorphic microsatellite loci on a population of turbot (*Scophthalmus maximus*) with a known genealogy, the behaviour of four pairwise marker-based coancestry estimators and the molecular coancestry has been evaluated. The population consisted on 138 families with 4 full-sib individuals each and one family with 8 full-sib individuals (up to 15 half-sib families were also present due to the sharing of parents between some of the full-sibs families). Our results suggested that transforming the pairwise estimators and the molecular coancestry to family relationships through the establishment of thresholds performs slightly better than the explicit pedigree reconstruction method, when accuracy is measured in a pairwise basis. However, if joint relationships between more than two individuals were tested at a time, the threshold methods led to a high percentage of incongruous triads of full-sib individuals, with Mendelian incompatibilities appearing in congruous full-sib families (more than 70% and 60% in our study, respectively). The explicit pedigree reconstruction approach, due to its nature, is free from such problems. Therefore, the pedigree reconstruction approach seems to be a valuable tool to provide a congruent and compatible pedigree when the pairwise marker-based coancestry matrices or the molecular coancestry need to be transformed.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Molecular marker; Microsatellites; Pairwise; Pedigree reconstruction; Coancestry estimator; *Scophthalmus maximus*

1. Introduction

In many areas of population biology and genetics the knowledge of the pedigree structure or the coancestries between individuals of a population is important, for

* Corresponding author. Tel.: +34 913473759; fax: +34 913572293.
E-mail address: ramilo@inia.es (S.T. Rodríguez-Ramilo).

example in the estimation of genetic parameters (heritabilities and genetic correlations) and breeding values (Lynch and Walsh, 1998). However, in most natural populations and many captive ones the genealogical information is unavailable. This is especially true in the field of aquaculture species.

During the last years a lot of microsatellite data has arisen that allows for parentage and coancestry analysis in fish species (Jackson et al., 2003; Sekino et al., 2004; Castro et al., 2006). Family assignment is widely applied when parents and offspring genotypes are available. However, coancestry analysis is necessary when unstudied hatchery populations, commercial batches or a group of wild fishes are used to constitute a broodstock or start a breeding program. In this situation, the main objectives should be the avoidance of inbreeding and the increase of the response to selection in a breeding program.

Among the approaches developed to infer coancestry from molecular marker data, there is a group of methods involving an explicit reconstruction of genealogies [see Butler et al. (2004) and Fernández and Toro (2006) for a revision]. Particularly, the method proposed by Fernández and Toro (2006) inputs a predefined coancestry matrix (either molecular coancestries or estimated ones) to generate a genealogy with the highest correlation with the initial matrix. Another category includes pairwise methods focussed on the calculation of the coancestries between pairs of individuals and, thus, they do not imply a pedigree reconstruction.

Besides the assumptions of Hardy–Weinberg and linkage equilibrium for the markers used and the dependence of the estimators on the knowledge of the true allelic frequencies of the base population considered (a problem also for most genealogy reconstruction estimators), a main concern when a genealogy is reconstructed using pairwise methodology is that most estimators provide a continuous measure of coancestry. Consequently, thresholds are necessary to convert coancestry estimates in discrete relationships. For example, in this study, only three relationships are considered: full-sibs, half-sibs and unrelated, and pairs of individuals must be classified into one of these predefined groups based on the estimates of coancestry.

Another inconvenience of pairwise methods is that, once transformed into discrete classes, they can lead to incongruous assignments. This occurs because only two individuals are taken into account at a time, and for example, individuals i and j could be classified as full-sibs, j and k as full-sibs, but the estimated relationship between i and k could be any relationship other than full-sibs (Thomas and Hill, 2000). Moreover, in this

pairwise comparison approach, the inclusion of individuals with genotypes that do not comply with Mendelian segregation rules into the same full-sib family is possible (Thomas and Hill, 2002). For example, if three individuals homozygous for three different alleles at the same locus (AA, BB and CC) are available, a pairwise comparison basis could estimate that the three individuals are full-sibs, which is nonsense.

Fernández and Toro (2006) proposed the use of their estimator to perform the conversion between continuous measures of coancestry to discrete categories by using the pairwise estimated matrix as the input of their method.

In this report, we evaluate the accuracy of four pairwise marker-based coancestry estimators to reconstruct the pedigree in a population of turbot (*Scophthalmus maximus*) with known genealogy, especially in terms of the number of incongruous/incompatible assignments. Transformations based on thresholds as well as the possibility of correcting the coancestry estimates *via* an explicit pedigree reconstruction approach were tested.

2. Materials and methods

2.1. Biological data set

The turbot (*S. maximus*; Scophthalmidae; Pleuronectiformes) is a species with high commercial value (Castro et al., 2004). The samples analysed were collected at Stolt Sea Farm facilities, a highly specialized company on turbot production. One hundred and thirty-eight families with four full-sib individuals each and one family with eight full-sib individuals were evaluated. Because parents were shared between some families, there were 15 half-sib families in our sample, involving 27 of the 139 full-sib families.

2.2. Molecular information

The 560 individuals constituting the commercial population were genotyped for ten highly polymorphic microsatellite loci (*Smax-01*, *Smax-02*, *Smax-04*, *Sma3-8INRA*, *Sma5-111INRA*, *Sma3-12INRA*, *Sma4-14INRA*, *Sma1-125INRA*, *Sma3-129INRA* and *Sma1-152INRA*), previously characterised in *S. maximus* (Bouza et al., 2002; Castro et al., 2004). The original data set was corrected using actual family data and all genotyping errors were eliminated (Castro et al., 2004). The four mutations detected were corrected following a conservative criterion, replacing the mutant allele by the closest in size to the segregant parent. Measures of genetic diversity and deviation from Hardy–Weinberg equilibrium were obtained with the computer program CERVUS 3.0 (Marshall et al., 1998). Bonferroni correction was applied for multiple tests.

2.3. Coancestry

The molecular coancestry (f_M) was calculated applying the Malécot's (1948) definition to the marker loci. Thus, the

molecular coancestry at a particular locus between individuals i and j was calculated as the probability that two alleles taken at random, one from each individual, were equal (identical by state, IBS). Throughout several markers, the molecular coancestry was obtained as the arithmetic mean over marker loci.

From the above molecular coancestry definition, different marker-based coancestry estimators have been proposed to infer the genealogical coancestries. In the present study the following pairwise coancestry estimators have been analysed: f_{QG} (Queller and Goodnight, 1989), f_R (Ritland, 1996), f_{LR} (Lynch and Ritland, 1999) and f_W (Wang, 2002). These pairwise marker-based coancestry estimators were implemented on free software available on the web (Kinship 1.2, Queller and Goodnight 1989; SPAGeDi 1.2, Hardy and Vekemans 2002; MER, Wang 2002).

2.4. Transformation of continuous values

In order to get genealogical relationships from the pairwise estimates of coancestry, two thresholds were established: 0.0625 to differentiate between unrelated (UR) and half-sib (HS) pairs of individuals, and 0.1875 to move from HS to full-sib (FS) category. These are the mid-points between the expectations of the three considered categories (0, 0.125 and 0.25 for UR, HS and FS, respectively). The corresponding thresholds for the molecular coancestry (f_M) can be inferred from the following expression:

$$1 - f_M = H_e(1 - f_G)$$

where H_e is the expected heterozygosity of the base population (Toro et al., 2002) and f_G is the genealogical coancestry. Using the actual allele frequencies to calculate the expected heterozygosity, the inferred threshold between UR and HS was 0.2272 and between HS and FS was 0.3303.

The pedigree reconstruction method of Fernández and Toro (2006), f_{FT} thereafter, was also evaluated. This algorithm generates the pedigree coancestry matrix as the highest correlated with any provided coancestry matrix as a reference (originally the molecular coancestry matrix). The pedigree reconstruction approach was implemented on free software available on the web (MOL_COANC, Fernández and Toro 2006). In the present study f_{FT} method was implemented on the f_M as well as on the coancestry matrices obtained from each of the pairwise estimators.

2.5. Measures of accuracy

2.5.1. Descriptive statistics

Squared error (the squared difference between the marker-based coancestry and the genealogical coancestry), correlation and regression coefficients (of the genealogical coancestry on the marker-based coancestry). These statistics were calculated using SPSS 12.0 on the threshold transformed values for all possible pairs of individuals.

2.5.2. Correct and wrong assignments (Smith et al., 2001; Thomas and Hill, 2002)

As we only considered individuals to be FS, HS or UR, there were nine situations depending on the true and the estimated relationship. Three of them corresponded to correctly estimated relationships, another three were overestimations (estimated FS when really HS or UR and, estimated HS when really UR) and the remaining three were underestimations (the opposite situations). The advantage of such a measure is that it allows to specify the direction (i.e. it is not equivalent to classify a FS pair as UR or to estimate UR couples as FS) and the magnitude of errors (i.e. it is not equivalent to classify a FS pair as HS or UR).

2.5.3. Incongruous triads of full-sibs individuals (Thomas and Hill, 2000)

The detection of cases where individuals i and j have been classified as full-sibs, j and k have been classified as full-sibs too, but the estimated relationship between i and k was other than full-sibs.

2.5.4. Mendelian segregation incompatibilities (for a description see Fernández and Toro, 2006)

Every group with more than two FS individuals that did not conform to the well known rules: (i) no more than four different alleles and genotypes can exist in a FS family; (ii) if there are four alleles, a particular allele can be heterozygous with two other alleles at most; (iii) if there are four alleles, no homozygotes can be found; (iv) if there are three alleles, only one type of homozygote can exist. No tests were performed to check for the congruency of other type of relatives (e.g. half-sib groups of individuals) because the number of possible configurations is huge and the probability of detecting an incompatibility is very low.

3. Results

3.1. Genetic variability

The linkage group and position of the markers according to Bouza et al. (in press), and the measures of genetic diversity in the analysed microsatellite loci are given in Table 1. All loci showed high diversity values. The mean number of alleles was 15.10, and ranged from a minimum of eight in *Sma1-152 INRA* to a maximum of 21 in *Sma3-129INRA*. The expected heterozygosity and the polymorphic information content ranged from 0.745 and 0.706 to 0.907 and 0.899 with an average over loci of 0.825 and 0.804, respectively. The expected and observed heterozygosity showed similar values, with only a significant deviation from Hardy–Weinberg equilibrium observed at *Sma3-129INRA* after applying Bonferroni correction ($P = 0.002$).

3.2. Descriptive statistics

The relative frequency of the estimated pairwise coancestry values before the conversion to family relationships is

Table 1
Linkage group, position and estimates of genetic diversity in each evaluated locus

Linkage group	Position	Locus	n_a	H_e	H_o	PIC	HW
15	43.5	<i>Smax-01</i>	15	0.811	0.796	0.788	ns
17	11.7	<i>Smax-02</i>	18	0.840	0.855	0.828	ns
3	74.3	<i>Smax-04</i>	19	0.808	0.809	0.783	ns
16	49.4	<i>Sma3-8INRA</i>	14	0.852	0.857	0.835	ns
22	0	<i>Sma5-111INRA</i>	17	0.880	0.889	0.868	ns
6	9.6	<i>Sma3-12INRA</i>	12	0.836	0.868	0.815	ns
7	0	<i>Sma4-14INRA</i>	17	0.796	0.777	0.777	ns
13	1.1	<i>Sma1-125INRA</i>	10	0.776	0.777	0.745	ns
17	46.5	<i>Sma3-129INRA</i>	21	0.907	0.868	0.899	*
5	35.5	<i>Sma1-152INRA</i>	8	0.745	0.734	0.706	ns
	Mean		15.10	0.825	0.823	0.804	

n_a : number of alleles; H_e : expected heterozygosity; H_o : observed heterozygosity; PIC: polymorphic information content; HW: significance of deviation from Hardy–Weinberg equilibrium; ns: not significant. *: significant at the 5% level.

illustrated separately for each type of relationship in Fig. 1, and the mean and variance for each coancestry estimator are shown in Table 2. The evaluated methods underestimated the averaged coancestry values in the three types of data set (the expected values were 0, 0.125 and 0.25 for unrelated, half-sib and full-sib individuals, respectively). This underestimation was particularly important in the half-sib comparison. The highest variances were observed for the f_R and f_{LR} estimators in the full-sibs. The behaviour between the coancestry estimators for the total number of pairs of individuals (156,520) was very similar to the unrelated comparison. This is the consequence that most comparisons were between unrelated individuals (155,408), due to the particular structure of the population.

The statistics (squared error and the regression and correlation coefficients) obtained for each pairwise marker-based coancestry estimator and the molecular coancestry are shown in Table 3. Calculations were performed on the estimated coancestry matrix once transformed to discrete categories for the total number of pairs of individuals and also for the full-sib and half-sib individuals. Concerning to the total number of pairs of individuals, the estimators f_R and f_{LR} showed the most accurate performance with a reduced squared error (Table 3a) and a higher correlation and regression coefficients, although all values were very low (highest correlation of 0.37). When values were transformed via the f_{FT} method (Table 3b), f_M , f_{QG} and f_W showed a further reduction in the correlation and the regression coefficients. However, these values increased both for f_R and f_{LR} . In addition, the squared error decreased in both pairwise estimators. Low correlation coefficient values were also due to the structure of the population comprising more than 99% of UR pairs. That is, an increased correlation and regression coefficients could be expected if a more balanced population structure was available (33% for each kind of relationship). When only full-sibs and half-sibs were considered, f_{QG} and f_W showed the best behaviour (Table 3a), although after the explicit pedigree

reconstruction (Table 3b) both f_R and f_{LR} maintained the best performance, and all the estimators showed an increased error.

3.3. Inferred assignments

Table 4 shows the number of correct and wrong assignments when the conversion to genealogical relationships was made according to thresholds (a), and under the explicit pedigree reconstruction method (b). Reduced percentages of subestimates were observed because the most frequent relationship between the evaluated individuals was UR (99.3%). In general, pairwise marker-based coancestry estimators showed a higher percentage of correct assignments and a reduced percentage of overestimates, whatever the transformation method. The highest percentage of correct assignments corresponded to f_R with 94% and 98% for threshold and explicit pedigree reconstruction, respectively. This was mainly due to the better behaviour of Ritland's (f_R) estimator in the UR assignment. The remaining estimators behaved better in assigning correctly HS and FS relationships, especially the f_{QG} estimator, which showed the best performance both in HS and FS assignments under the thresholds transformation. It is worth noting that f_M yielded similar results to any other estimator and even sometimes performed better (e.g. in the assignment of HS and FS after the explicit pedigree reconstruction).

3.4. Triads assignments

Table 5 shows the total number of triads of full-sib individuals constructed under each estimator and the percentage of incongruous ones. The data contained 608 triads of, naturally, congruous full-sib individuals. After the conversion to genealogical relationships according to established thresholds (Table 5a) the pairwise coancestry estimators and f_M overestimated the total number of triads of full-sib individuals with an extremely high percentage of incongruences (> 73%). Following the explicit pedigree reconstruction strategy, the total number of full-sib triads was underestimated in accordance with the lower number of estimated full-sib relationships (Table 4b), but no incongruent triads of full-sib individuals were found, due to the nature of the f_{FT} estimator (Table 5b).

3.5. Mendelian segregation incompatibilities

The incompatibilities (at least in one locus) regarding to Mendelian segregation and the percentage respect to the total number of congruent sets of full-sib families (when more than two individuals are considered) after the conversion to genealogical relationships using thresholds are given in Table 6. All pairwise coancestry estimators as well as the f_M showed a high percentage of Mendelian incompatibilities (higher than 62%). Contrarily, due to its nature, the f_{FT} algorithm did not allow for any Mendelian incompatibility when used to transform the continuous estimates of coancestry (data not shown).

Another way to test the accuracy of each procedure is to count the number of individuals correctly allocated to their true

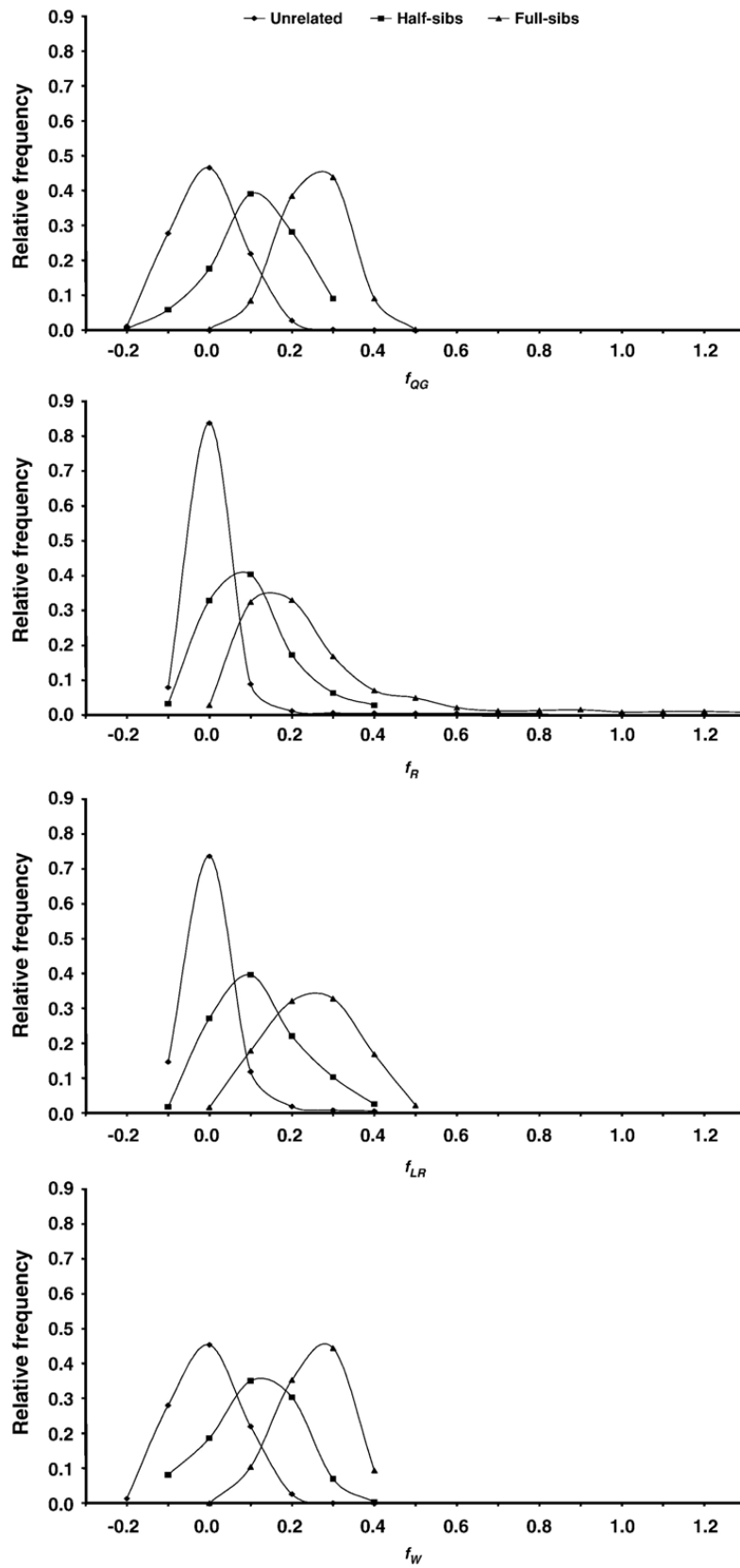


Fig. 1. Relative frequency of the pairwise marker-based coancestry values before the conversion to genealogical relationships. Total number of pairs of unrelated individuals: 155,408. Total number of pairs of half-sibs: 256. Total number of full-sibs: 856. f_{QG} : Queller and Goodnight (1989) estimator. f_R : Ritland (1996) estimator. f_{LR} : Lynch and Ritland (1999) estimator. f_W : Wang (2002) estimator.

Table 2
Mean and variance of the coancestry estimators for each type of genealogical relationship

Pairwise comparisons		Mean	Variance
Unrelated	f_{QG}	-0.002	0.005
	f_R	-0.002	0.002
	f_{LR}	-0.002	0.002
	f_W	-0.003	0.005
Half-sibs	f_{QG}	0.108	0.009
	f_R	0.096	0.009
	f_{LR}	0.111	0.009
	f_W	0.104	0.010
Full-sibs	f_{QG}	0.248	0.005
	f_R	0.239	0.040
	f_{LR}	0.246	0.010
	f_W	0.248	0.006

Total number of pairs of unrelated individuals: 155,408. Total number of pairs of half-sibs: 256. Total number of full-sibs: 856. f_{QG} : Queller and Goodnight (1989) estimator. f_R : Ritland (1996) estimator. f_{LR} : Lynch and Ritland (1999) estimator. f_W : Wang (2002) estimator.

family. Fig. 2 shows the frequency of the different possibilities when considering congruent full-sib families of more than two individuals.

When thresholds were used, there were a high probability of grouping together at least two real FS but mixed with other individuals for all estimators (white bars, categories with asterisk except 1*). However, most of such groups were incompatible with Mendelian segregation (grey bars). Obviously, when three or four actual FS were grouped alone (categories 3 and 4) the families resulted compatible. When f_{FT} was the transforming method (black bars) all families were congruous and compatible and the distribution between categories was very similar to the threshold classification method.

4. Discussion

Many estimators and algorithms have been developed to infer the genealogical relationships from molecular marker data when there is a lack of pedigree information. A broad classification of approaches (Butler et al., 2004) distinguishes between those reconstructing a complete population structure and pairwise methods that do not involve an explicit reconstruction of genealogies. Depending on the assumptions made by each method and the underlying methodology, most of them suffer from one or several different limitations. The most general concerns refer to (1) the assumption that molecular markers are in Hardy–Weinberg and linkage equilibrium; and (2) the dependence on the knowledge of the true allelic frequencies in the base population. Pairwise methods yield estimates along a continuous range including negative values, which implies the necessity

to define arbitrary thresholds to establish family relationships (full-sibs, half-sibs and unrelated), as in the case of the population of turbot evaluated in this study. But some authors (e.g. Thomas and Hill, 2000) have pointed out that this strategy may lead to incongruous assignments and Mendelian segregation incompatibilities, as they account only for two individuals at a time. Fernández and Toro (2006) proposed a method to adjust continuous values to genealogical relationships providing the estimated matrix as the input of the algorithm.

Our results suggest that, as found in other studies, when comparisons are made in terms of pairs of individuals assigned using both molecular and genealogical methods (squared error, correlation and regression coefficients and assignments) all the estimators showed a high percentage of correct assignments (f_R and f_{LR} being the best), but with a reduced correlation and regression coefficients of the genealogical coancestry on the marker-based coancestry. When explicit pedigree reconstruction (Fernández and Toro, 2006) is the implemented method, the aforementioned parameters increased in f_R and f_{LR} , although they were still far from 1, probably due to the population structure with an elevated percentage of UR individuals. In populations with a more balanced proportion of UR, HS and FS individuals, the f_{QG} and f_W estimators should behave best (unpublished data).

Table 3
Statistics for each pairwise marker-based coancestry estimator and the molecular coancestry

		f_M	f_{QG}	f_R	f_{LR}	f_W
<i>a</i>						
Total	Mean squared error	0.003	0.003	0.001	0.002	0.003
	Correlation coefficient	0.269	0.280	0.373	0.360	0.278
	Regression coefficient	0.096	0.099	0.203	0.163	0.098
HS or FS	Mean squared error	0.006	0.005	0.010	0.006	0.005
	Correlation coefficient	0.546	0.590	0.428	0.504	0.594
	Regression coefficient	0.341	0.399	0.256	0.316	0.398
<i>b</i>						
Total	Mean squared error	0.003	0.002	0.000	0.001	0.003
	Correlation coefficient	0.168	0.206	0.462	0.408	0.153
	Regression coefficient	0.066	0.099	0.490	0.336	0.061
HS or FS	Mean squared error	0.020	0.022	0.023	0.019	0.022
	Correlation coefficient	0.326	0.338	0.383	0.388	0.337
	Regression coefficient	0.182	0.187	0.233	0.225	0.188

(a) Thresholds transformation. (b) Explicit pedigree reconstruction. See Table 2 for abbreviations. f_M : molecular coancestry. Total: the total number of pairs of individuals is 156,560. HS or FS: the number of pairs of individuals being half-sibs or full-sibs is 1112.

Table 4
Number of correct (i. e. genealogical and estimated coancestry coincide) and wrong assignments

a		Real				b		Real			
f_M		UR	HS	FS	%	f_M		UR	HS	FS	%
Estimated	UR	130,336	96	18	83.9	Estimated	UR	129,191	141	232	83.1
	HS	23,172	111	228	43.4		HS	26,076	113	357	44.1
	FS	1,900	49	610	71.3		FS	141	2	267	31.2
	T				83.7		T				82.8
f_{QG}		UR	HS	FS	%	f_{QG}		UR	HS	FS	%
Estimated	UR	127,235	75	5	81.9	Estimated	UR	139,571	166	256	89.8
	HS	26,773	128	174	50.0		HS	15,698	84	356	32.8
	FS	1,400	53	677	79.1		FS	139	6	244	28.5
	T				81.8		T				89.4
f_R		UR	HS	FS	%	f_R		UR	HS	FS	%
Estimated	UR	146,648	115	50	94.4	Estimated	UR	153,375	189	241	98.7
	HS	8,082	97	381	37.9		HS	2,006	63	453	24.6
	FS	678	44	425	49.6		FS	27	4	162	18.9
	T				94.0		T				98.1
f_{LR}		UR	HS	FS	%	f_{LR}		UR	HS	FS	%
Estimated	UR	142,206	87	25	91.5	Estimated	UR	151,484	165	198	97.5
	HS	12,097	112	218	43.8		HS	3,868	87	426	34.0
	FS	1,105	57	613	71.6		FS	56	4	232	27.1
	T				91.3		T				97.0
f_W		UR	HS	FS	%	f_W		UR	HS	FS	%
Estimated	UR	127,028	80	4	81.7	Estimated	UR	129,711	171	264	83.5
	HS	26,998	123	179	48.0		HS	25,536	79	360	30.9
	FS	1,382	53	673	78.6		FS	161	6	232	27.1
	T				81.7		T				83.1

(a) Thresholds transformation. (b) Explicit pedigree reconstruction. True (genealogical) relationship is indicated in columns and estimated relationship in rows. Pairs of individuals genealogically being: UR (unrelated)=155,408; HS (half-sibs)=256 and FS (full-sibs)=856. Bold %: percentage of UR, HS, FS and T (total) assigned correctly. See Table 2 for abbreviations. f_M : molecular coancestry.

It has been suggested (Thomas and Hill, 2000, 2002; Fernández and Toro, 2006) that if more than two individuals are taken into account at a time, pairwise coancestry approaches suffer from different limitations. In this study, the number of incongruous triads of full-sib individuals and the number of Mendelian segregation incompatibilities at least in one locus for all possible combinations of congruent full-sib families were evaluated. The pairwise approaches studied (Queller and Goodnight, 1989; Ritland, 1996; Lynch and Ritland, 1999; Wang, 2002) showed a high percentage of incongruous triads of full-sib individuals (more than 70%) and also a high number of Mendelian segregation incompatibilities (more than 60%), with their negative consequences in the management of the population.

In this study we have been able to detect the incongruent associations and the incompatible familiar groups. But even in this case, often there are different ways to solve the incompatibility, and it is not

Table 5
Total number of triads of full-sib individuals and percentage of incongruous full-sib triads

	f_M	f_{QG}	f_R	f_{LR}	f_W
a					
Total number of triads of full-sib individuals	33,070	14,817	4,632	8,414	15,392
% of incongruous triads of full-sibs	89.9	86.8	73.7	78.5	87.9
b					
Total number of triads of full-sib individuals	201	211	66	92	163
% of incongruous triads of full-sibs	0.0	0.0	0.0	0.0	0.0

(a) Thresholds transformation. (b) Explicit pedigree reconstruction. See Table 2 for abbreviations. f_M : molecular coancestry. The true number of full-sib individuals from the genealogical coancestry is 608.

Table 6
Number of Mendelian segregation incompatibilities (at least in one locus) and percentage respect to the total number of congruent sets of full-sib families after the conversion to genealogical relationships using thresholds

	f_M	f_{QG}	f_R	f_{LR}	f_W
More than 4 alleles	110	83	37	72	49
4 alleles					
More than 2 heterozygotes for 1 allele	87	110	45	93	148
Homozygotes	57	26	23	37	28
3 alleles					
More than 1 homozygote	8	3	1	9	4
Percentage	79.4	73.3	61.6	74.8	71.8

Only sets with more than two individuals were considered. See Table 2 for abbreviations. f_M : molecular coancestry.

straightforward to decide the correct one. Moreover, the discarded individuals from a full-sib family could remain as HS of the present family or could be included in a new FS family and, therefore, it would be UR with the evaluated family.

Contrarily, the explicit pedigree reconstruction approach (Fernández and Toro, 2006), due to its nature, does not allow the build-up of any incongruous triad of full-sib individuals or full-sib families with Mendelian segregation incompatibilities. The absence of incongruous triads of full-sibs could be helpful in situations where it is important to separate full-sib individuals, for example, in a specific number of tanks in aquaculture. Therefore, using pairwise marker-based coancestry matrices as input files of the pedigree reconstruction method seems to be a good approach to assign genealogical relationships.

5. Conclusion

The results presented here show that transforming the pairwise marker-based coancestry estimators and the molecular coancestry to family relationships through the establishment of thresholds performs slightly better than the explicit pedigree reconstruction method when only pairs of individuals are evaluated at a time. However, if more than two individuals are tested at a time, the threshold methods lead to a high percentage of incongruous triads of full-sib individuals and Mendelian segregation incompatibilities in congruous full-sib families. The explicit pedigree reconstruction approach, due to its nature, is free from such problems. Therefore, the pedigree reconstruction approach seems a valuable tool to provide a congruent and compatible pedigree when the pairwise marker-

based coancestry matrices or the molecular coancestry need to be transformed.

Note: FORTRAN code (or a compiled file) to detect incongruous triads of full-sib individuals and Mendelian segregation incompatibilities in congruous full-sib families will be available on the following web site (<http://www.uvigo.es/webs/c03/webc03/XENETICA/XB2/Jesus/Fernandez.htm>).

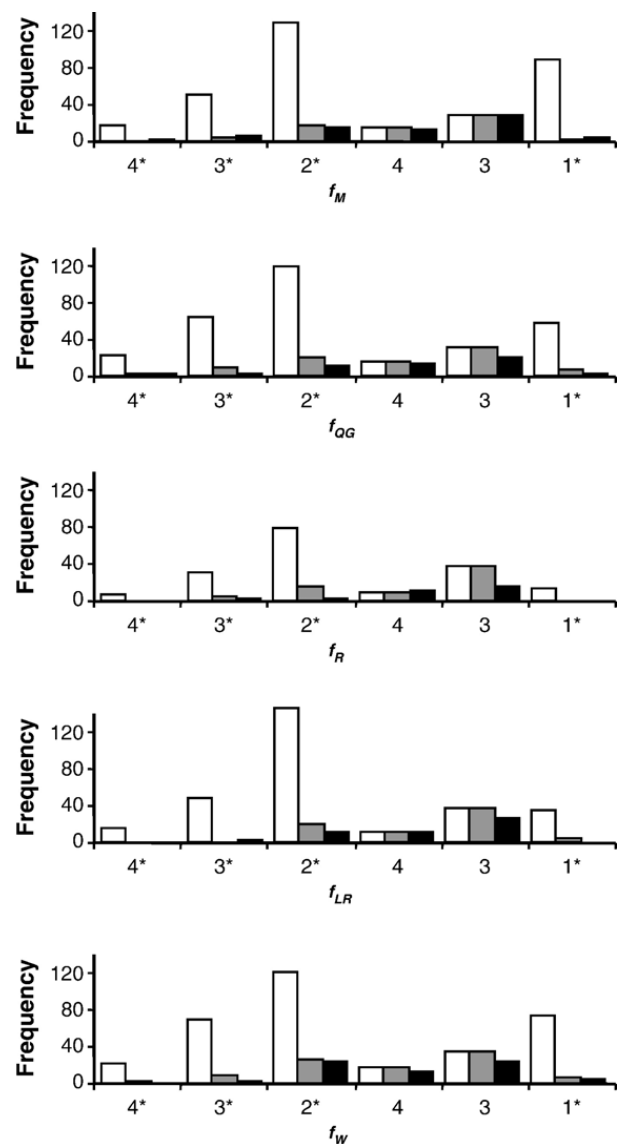


Fig. 2. Frequency of the different categories observed for the number of actual full-sibs grouped into the same family. Only families with more than two individuals were considered. Asterisks denote families with other individuals included together with the actual full-sibs. 1*: constructed groups where all individuals belonged to different families (i. e., none of the couples was actual full-sibs). White bars: thresholds transformation, total number of congruous full-sib families. Grey bars: thresholds transformation, number of congruous and compatible families. Black bars: explicit pedigree reconstruction, number of congruous and compatible families. Abbreviations as in Table 2. f_M : molecular coancestry.

Acknowledgments

We thank two anonymous referees for helpful comments on the manuscript. The authors want to thank Stolt Sea Farm S.A. for providing the samples. This work was supported by a project of the Plan Estratégico del INIA (CPE03-004-C2) and a project of the Plan Nacional del MEC (CGL2006-13445-C02-01).

References

- Bouza, C., Presa, P., Castro, J., Sánchez, L., Martínez, P., 2002. Allozyme and microsatellite diversity in natural and domestic populations of turbot (*Scophthalmus maximus*) in comparison with other Pleuronectiformes. *Canadian Journal of Fisheries and Aquatic Sciences* 59, 1460–1473.
- Bouza, C., Hermida, M., Pardo, B.G., Fernández, C., Fortes, G.G., Castro, J., Sánchez, L., Presa, P., Pérez, M., Sanjuan, A., de Carlos, A., Álvarez-Dios, J.A., Ezcurra, S., Cal, R.M., Piferrer, F., Martínez, P., in press. A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics* 177 (4).
- Butler, K., Field, C., Herbinger, C.M., Smith, B.R., 2004. Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. *Molecular Ecology* 13, 1589–1600.
- Castro, J., Bouza, C., Presa, P., Pino-Querido, A., Rianza, A., Ferreiro, I., Sánchez, L., Martínez, P., 2004. Potential sources of error in parentage assessment of turbot (*Scophthalmus maximus*) using microsatellite loci. *Aquaculture* 242, 119–135.
- Castro, J., Pino, A., Hermida, M., Bouza, C., Rianza, A., Ferreiro, I., Sanchez, L., Martínez, P., 2006. A microsatellite marker tool for parentage analysis in Senegal sole (*Solea senegalensis*): genotyping errors, null alleles and conformance to theoretical assumptions. *Aquaculture* 261, 1194–1203.
- Fernández, J., Toro, M.A., 2006. A new method to estimate relatedness from molecular markers. *Molecular Ecology* 15, 1657–1667.
- Hardy, O.J., Vekemans, X., 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2, 618–620.
- Jackson, T.R., Martin-Robichaud, D.J., Reith, M.E., 2003. Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. *Aquaculture* 220, 245–259.
- Lynch, M., Ritland, K., 1999. Estimation of pairwise relatedness with molecular markers. *Genetics* 152, 1753–1766.
- Lynch, M., Walsh, B., 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- Malécot, G., 1948. *Les Mathématiques de l'hérédité*. Masson, Paris.
- 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, 639–655.
- Queller, D.C., Goodnight, K.F., 1989. Estimating relatedness using molecular markers. *Evolution* 43, 258–275.
- Ritland, K., 1996. Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research* 67, 175–185.
- Sekino, M., Sugaya, T., Hara, M., Taniguchi, N., 2004. Relatedness inferred from microsatellite genotypes as a tool for broodstock management of Japanese flounder *Paralichthys olivaceus*. *Aquaculture* 233, 163–172.
- Smith, B.R., Herbinger, C.M., Merry, H.R., 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. *Genetics* 158, 1329–1338.
- Thomas, S.C., Hill, W.G., 2000. Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics* 155, 1961–1972.
- Thomas, S.C., Hill, W.G., 2002. Sibship reconstruction in hierarchical population structures using Markov chain Monte Carlo techniques. *Genetical Research* 79, 227–234.
- Toro, M., Barragán, C., Óvilo, C., Rodríguez, J., Rodríguez, C., Silió, L., 2002. Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetics* 3, 309–320.
- Wang, J.L., 2002. An estimator for pairwise relatedness using molecular markers. *Genetics* 160, 1203–1215.