



**HAL**  
open science

## An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion

M. Vandeputte, S. Mauger, M. Dupont-Nivet

### ► To cite this version:

M. Vandeputte, S. Mauger, M. Dupont-Nivet. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by exclusion. *Molecular Ecology Notes*, 2005, 6 (1), pp.265-267. 10.1111/j.1471-8286.2005.01167.x . hal-04297870

**HAL Id: hal-04297870**

**<https://hal.science/hal-04297870v1>**

Submitted on 22 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

1 **An evaluation of allowing for mismatches as a way to manage**  
2 **genotyping errors in parentage assignment by exclusion**

3 M. VANDEPUTTE, S. MAUGER and M. DUPONT-NIVET

4 INRA Génétique des Poissons, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

5

6 Correspondence: Marc Vandeputte, Fax: +33 1 34 65 23 90; E-mail: Marc.Vandeputte@jouy.inra.fr

7

8 **Abstract**

9 In parentage assignment by exclusion, using multiple and very polymorphic loci, genotyping errors are  
10 a major cause of non-assignment. Using stochastic simulations, we tested the possibility to allow for  
11 mismatches at one or more allele as a way to recover assignment power. This was very efficient provided  
12 the set of loci used had a high assignment power (> 99%) and the error rate was not too high (below 3–  
13 4%). In these cases, most of the theoretical assignment power could be recovered. We also showed the  
14 efficiency of the method in a practical experiment with rainbow trout

15 .

16 **Keywords:** exclusion, microsatellites, parentage assignment.

17

18 Using multilocus genotypes to ascertain parentage for genetic studies is a widely used method in natural  
19 populations (Milner et al. 2000; Garant & Kruuk 2005) or in experimental populations for which matings  
20 are not fully controlled or individuals cannot be tagged (e.g., fish, Vandeputte et al. 2004). Exclusion is  
21 a simple and efficient method for assigning parents to an offspring that uses incompatibilities (regarding  
22 Mendelian inheritance rules) between parents and offspring. A major drawback of exclusion is that a  
23 single mismatch between parent and offspring genotypes is enough to exclude a potential parent, thus  
24 making this technique extremely sensitive to genotyping errors or mutations (Jones & Ardren 2003).  
25 There is a wide range of variation for genotyping errors, often 1% or less (Hoffman & Amos 2005), but  
26 up to 17–29% in some cases (Maudet et al. 2004). For the management of such errors, several programs  
27 give the possibility to allow for a given number of mismatches between the offspring and its parents.  
28 Probmax allows for mismatches in a stepwise mutation/error approach (Danzmann 1997), whereas

29 newpat (Worthington Wilmer et al. 1999) and parente (Cercueil et al. 2002) only allow the user to define  
30 the number of mismatches authorized. In addition, parente calculates a probability for each triplet with  
31 mismatches, taking into account the genotype, but also the error rate and the sampling rate in the  
32 population. However, no evaluation is done of the global efficiency of allowing for mismatches for  
33 managing typing errors or mutations.

34 Here, we use stochastic simulations to study (i) the impact of genotyping errors, and (ii) the effect of  
35 allowing for a small number of mismatches on the efficiency of parentage assignment by exclusion.

36 We simulated loci with five equally distributed alleles, corresponding to an individual offspring  
37 exclusion probability of 0.77 (0.60 for paternity or maternity exclusion probability, Doddset al. 1996),  
38 which is the usual range of exclusion probabilities for microsatellites (e.g., Lemes et al. 2002; Bessert  
39 & Orti 2003; Castro et al. 2004).

40 For each combination of parameters (number of loci: 6–12; number of parents: 20–200; genotyping error  
41 rate: 0–10%), we generated 30 parental sets, and 1000 offspring/ parental set. For each offspring, the  
42 sire and the dam, as well as the alleles they transmitted, were randomly chosen. Genotyping errors were  
43 simulated by substituting one or more parental allele by another randomly drawn allele (error frequency  
44 = error rate per allele). We assigned offspring to their parents using an exclusion program running in  
45 Visual Basic for Excel ( vitassign, available onrequest to Marc.Vandeputte@jouy.inra.fr), and the  
46 different types of assignment (single correct, single incorrect, multiple, unassigned) were recorded.  
47 Single correct (SC) assignments are offspring assigned to the right parental couple, single incorrect (SI)  
48 are those assigned to an incorrect single parental pair, multiple are assigned to several possible parental  
49 pairs, and unassigned are not assigned to any parental pair. Assignment power is defined as the  
50 proportion of SC assignments when the error rate is set to zero. The unassigned offspring were then  
51 tested allowing for one mismatch (one incompatible allele allowed), and the remaining unassigned were  
52 tested allowing for two mismatches.

53 We also applied this procedure to the assignment of 390 rainbow trout offspring from a 45 males \* 2  
54 females factorial cross, genotyped for five loci (estimated single assignment power in this cross: 81.7%).  
55 The results using the very first reading of five loci allowing or not for mismatches were compared to the  
56 ‘true’ results where litigious genotypes had been verified, and where up to five additional markers (total  
57 assignment power: 98.7%) were genotyped when needed. We used these first readings because we knew

58 they had a high genotyping error rate, for testing in a real condition, the efficiency of allowing for  
59 mismatches. These results were compared with those of a simulation with 45 sires, 2 dams and 5 loci  
60 with 6 equally distributed alleles (giving a global exclusion probability similar to that of the true set of  
61 loci), with error rates from 0 to 15%.

62 The impact of genotyping errors on assignment rates was very high, and was highest when the  
63 assignment power of the set of loci used was high, which may lead to less assigned offspring with more  
64 powerful sets of loci (e.g. in the case of 20 parents and 1% error rate, 87.9% correct assignment is  
65 observed with eight 5-allele loci, vs. 84.0% with twelve 5-allele loci). In Fig. 1, we show the case of 50  
66 parents, which is a good illustration of the general picture. With the most powerful set of loci (ten 5-  
67 allele loci, 99.5% assignment power), allowing for one mismatch allows to assign correctly (SC) 81.0%  
68 of the offspring with 5% genotyping error rate, vs. 46.9% only with perfect exclusion. Allowing for two  
69 mismatches yields further improvement, with 90.2% of the offspring correctly assigned. Similarly, the  
70 proportion of offspring assigned to an incorrect single parental pair (SI) raises from 0.2% with perfect  
71 exclusion to 0.7% with one mismatch allowed and to 1.2% with two mismatches allowed. However, the  
72 total rate of false assignment remains very low. The picture is quite different with a less powerful set of  
73 loci (seven 5-allele loci, 90.9% assignment without genotyping error). In this case, using one mismatch  
74 allows to assign correctly only 67.8% of the offspring (with 5% error rate) vs. 53.5% with perfect  
75 exclusion, but allowing for one mismatch more yields no improvement (68.2% reassignment). The false  
76 assignment rate is also much higher (2.2% for perfect exclusion, 4.3% with one mismatch), but not  
77 modified (4.4%) by allowing for two mismatches.

78 The differences seem to lie essentially in the assignment power of the set of loci in a given size of  
79 crossing design with a given error rate. For the same assignment power, the number of parents and/or  
80 the number of loci have no visible impact neither on the loss of power caused by genotyping errors nor  
81 on the efficiency of mismatches to recover assignment power. For example, in all combinations tested  
82 with 1% error rate, it appears that an assignment power over 99% guarantees the correct assignment of  
83 more than 98% of the offspring using one mismatch, whereas only 83–88% would be correctly assigned  
84 using perfect exclusion. It also appears that the rate of false assignment with one mismatch is low in all  
85 these cases, and in most cases hardly higher than the rate of false assignment with perfect exclusion.  
86 The danger of false assignment seems to be more present with sets of loci with moderately high power

87 (90%), but very high error rates (Fig. 1), where as much as 12% of the offspring may be assigned to a  
88 false single couple.

89 In the 'real life' example with rainbow trout (Table 1), allowing for two mismatches on the first run of  
90 genotypes raised the number of correct unique assignments from 137 (35.1%) to 223 (57.1%), with only  
91 10 (2.6%) false unique assignments. Multiple assignments remain high due to the limited assignment  
92 power of the markers set used (87.1% without genotyping error). The simulated values with 12.5% error  
93 rate gave comparable results (35.0% unique assignments with 0 mismatches, 53.9% with 1 mismatch  
94 and 56.3% with 2 mismatches). However, the simulation results point out theoretical values of 5.4, 10.1  
95 and 10.9% false assignments with 0, 1 and 2 mismatches, respectively, which is much higher than what  
96 is seen here. Probably, some assignments considered as exact are false (offspring that matched one single  
97 couple with five loci were not regenotyped, and among those some 5% should be false assignments).  
98 Using one mismatch should have generated 4.7% more false assignments, and we find 2.6%, which is  
99 quite close given the sample size. Therefore, the predictions fit well with real life, in this example with  
100 voluntary huge genotyping error.

101 Genotyping errors or mutations have been identified as a major cause of reduced assignment power with  
102 exclusion methods, although other causes exist (e.g., null alleles, missing parental samples). It has been  
103 shown previously that maximum-likelihood assignment methods could manage genotyping errors  
104 (SanCristobal & Chevalet 1997). Here, we have evaluated a simple and efficient alternative using  
105 exclusion methodology. Provided the assignment power of the set of loci is high (> 99%, which is  
106 anyway a necessity for parentage assignment) and the error rate is reasonable (below 3–4%), allowing  
107 for mismatches allows to recover most of the assignment power without generating  
108 more than 1% of false assignments, which should not be harmful in subsequent analyses, especially in  
109 quantitative genetics where 95% confidence in the pedigree seems appropriate (Milner et al. 2000).

110

## 111 **References**

- 112 Bessert ML, Orti G (2003) Microsatellite loci for paternity analysis in the fathead minnow,  
113 *Pimephales promelas* (Teleostei: Cyprinidae). *Molecular Ecology Notes*, 3, 532–534.
- 114 Castro J, Bouza C, Presa P et al. (2004) Potential sources of error in parentage assessment of turbot  
115 (*Scophthalmus maximus*) using microsatellite loci. *Aquaculture*, 242, 119–135.

116 Cercueil A, Bellemain E, Manel S (2002) parente: computer program for parentage analysis. Journal  
117 of Heredity, 93, 458–459.

118 Danzmann RG (1997) probmax: a computer program for assigning unknown parentage in pedigree  
119 analysis from known genotypic pools of parents and progeny. Journal of Heredity, 88, 333.

120 Dodds KG, Tate ML, McEwan JC, Crawford AM (1996) Exclusion probabilities for pedigree testing  
121 farm animals. Theoretical and Applied Genetics, 92, 966–975.

122 Garant D, Kruuk LEB (2005) How to use molecular marker data to measure evolutionary parameters  
123 in wild populations. Molecular Ecology, 14, 1843–1859.

124 Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources  
125 and consequences for paternal exclusion. Molecular Ecology, 14, 599–612.

126 Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. Molecular  
127 Ecology, 12, 2511–2523.

128 Lemes MR, Brondani RPV, Grattapaglia D (2002) Multiplexed systems of microsatellite markers for  
129 genetic analysis of mahogany, *Swietenia macrophylla* King (Meliaceae), a threatened Neotropical  
130 timber species. Journal of Heredity, 93, 287–290.

131 Maudet C, Luikart G, Dubray D, Von Hardenberg A, Taberlet P (2004) Low genotyping error rates in  
132 wild ungulate faeces sampled in winter. Molecular Ecology Notes, 4, 772–775.

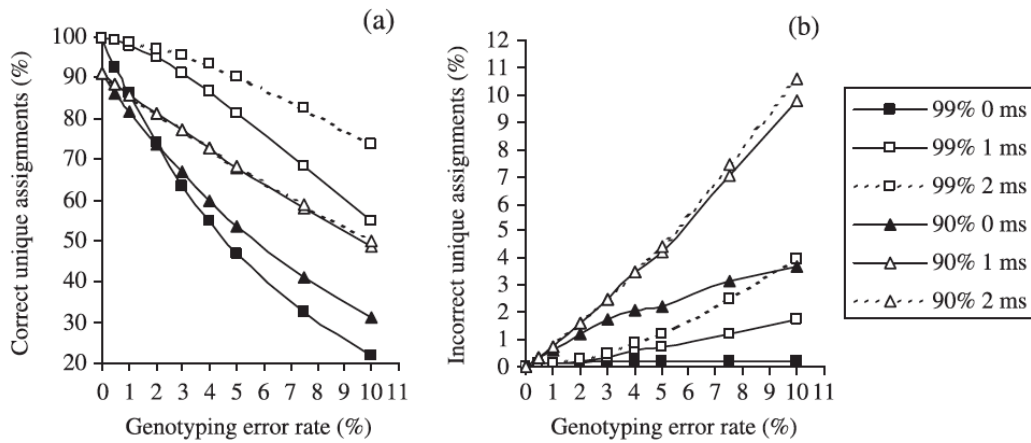
133 Milner JM, Pemberton JM, Brotherstone S, Albon SD (2000) Estimating variance components and  
134 heritabilities in the wild: a case study using the ‘animal model’ approach. Journal of Evolutionary  
135 Biology, 13, 804–813.

136 SanCristobal M, Chevalet C (1997) Error tolerant parent identification from a finite set of individuals.  
137 Genetical Research, 70, 53–62.

138 Vandeputte M, Kocour M, Mauger S et al. (2004) Heritability estimates for growth-related traits using  
139 microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). Aquaculture, 235,  
140 223–236.

141 Worthington Wilmer J, Allen PJ, Pomeroy PP, Twiss SD, Amos W (1999) Where have all the fathers  
142 gone? An extensive microsatellite analysis of paternity in the grey seal (*Halichoerus grypus*).  
143 Molecular Ecology, 8, 1417–1429.

144



146

147 **Fig. 1** Correct (a) and incorrect (b) unique parentage assignment rates in simulated offspring from 25  
 148 sires and 25 dams for different genotyping error rates using 7 loci with 5 alleles (90% assignment  
 149 power) or 8 alleles (99% assignment power), allowing for 0, 1 or 2 mismatches (ms).

150

Kind of assignments observed	Maximum number of mismatches authorized			
	0	1	2	3
Single correct	137	212	223	223
Single incorrect	0	9	10	10
Multiple	26	114	150	157
Unassigned	227	55	7	0

151

152 **Table 1** Numbers and accuracy of parental pairs identified with five microsatellites when allowing  
 153 mismatches or not in 390 fish from a 45 males\*2 females factorial cross in rainbow trout