Journal of Heredity 2008:99(6):616–623
doi:10.1093/jhered/esn048
Advance Access publication June 4, 2008

Effects of Microsatellite Null Alleles on Assignment Testing

Jens Carlsson

From the Duke University Marine Laboratory, Nicholas School of the Environment and Earth Sciences, 135 Duke Marine Laboratory Road, Beaufort, NC 28516-9721.

Address correspondence to J. Carlsson at the address above, or e-mail: jens.carlsson@duke.edu.

Abstract

Microsatellite null alleles are found to a varying degree across all taxa. They are problematic as they may inflate measures of genetic differentiation and create false homozygotes. Although there are several methods for correcting allele frequencies for null alleles and enable estimations of F_{ST}, much less is known about how null alleles affect assignment testing. Data presented here, based on simulations, show that the percentage of correctly assigned individuals in model-based clustering and Bayesian assignment methods were slightly, though significantly, reduced in the presence of null alleles (frequency range from 0.000 to 0.913). The bias in assignment tests caused by null alleles lead to a slight reduction in the power to correctly assigned individuals (0.2 and 1.0 percent units for STRUCTURE- and 2.4 percent units for GENECLASS-based assignment tests). Further, the presence of null alleles caused a small, however, significant overestimation of F_{ST}. Consequently, microsatellite loci affected by null alleles would probably not alter the overall outcome of assignment testing and could therefore be included in these types of studies. Nevertheless, loci prone to null alleles should be used with caution as they lower the power of assignment tests and alter the accuracy of F_{ST}, and loci less prone to null alleles should always be preferred.

Microsatellites are one of the most extensively used markers for population genetic studies because they are codominant and typically have large numbers of alleles. Microsatellites can be used for traditional population genetic analyses where the level of genetic differentiation among populations is the focus (i.e., F_{ST}). The statistical approaches associated with microsatellite analyses have, however, developed rapidly, and recent applications of microsatellites have revolutionized the field of population and conservation genetics by shifting the focus from populations to individuals. One of the most important applications is assigning individuals to, or excluding individuals from, potential source populations. The applications of assignment tests are numerous and include population differentiation (Waples and Gaggiotti 2006), detection of recent immigrants (Rannala and Mountain 1997), mixed stock analysis (Hansen et al. 2001), forensic identification of animal remains (Primmer et al. 2000), and identification of animals for conservation purposes (Nielsen et al. 2001). The ability to assign individuals is of great value for conservation efforts. For instance, assignment testing allows for assessing mixed stock fisheries and can be used to detect when individuals from populations of particular conservation values are present and permits managers to act accordingly (e.g., Shaklee et al. 1999).

There are, however, several problems associated with microsatellites including large allele dropout and slip-strand mispairing during polymerase chain reaction that can cause stutter (van Oosterhout et al. 2004). Another pervasive problem is null alleles that are caused by mutations in the primer binding region and prevent amplification of affected alleles (Pemberton et al. 1995). Null alleles are found in most taxa (Dakin and Avise 2004) but seem to be particularly common in populations with high effective population sizes (Chapuis and Estoup 2007) including insects (e.g., Lehmann et al. 1997; Chapuis et al. 2005) and mollusks (e.g., McGoldrick et al. 2000; Carlsson et al. 2006).

The presence of null alleles can sometimes be detected as an excess of homozygotes leading to deviations from Hardy–Weinberg expectations (HWEs). As null alleles create false homozygotes, they are problematic for parentage analysis (e.g., Pemberton et al. 1995; Reece et al. 2004). In addition, because they lower apparent genetic variability, they may erroneously inflate levels of genetic differentiation and affect population genetic analyses that rely on HWE (e.g., de Sousa et al. 2005; Chapuis and Estoup 2007). Although there is limited data concerning how null alleles affect measures of genetic differentiation (e.g., Chapuis and Estoup 2007), there is even less information about the effects on genetic assignment testing (Cornuet et al. 1999; Hauser et al. 2006).

There are several statistical approaches that may be used for correcting allele frequencies to account for null alleles (cf., van Oosterhout et al. 2004; Chapuis and Estoup 2007). These corrected allele frequencies can then be used to assess levels of population differentiation (i.e., F_{ST}). However, it is not as straightforward to account for null alleles when performing assignment tests as they rely on accurate information about the multilocus genotype of individuals. It is not possible through statistical approaches to assess the true multilocus genotype, and it would be necessary to sequence the entire microsatellite region (including both primer binding sites) for all individuals to identify which individuals are true homozygotes and which are heterozygotes for the null allele (i.e., individuals that have mutations in the primer binding region that prevents an allele from amplifying). As sequencing is expensive and time consuming, it is seldom a viable option for population genetic studies. Although allele frequencies for potential source populations used for assignment testing can be corrected for null alleles, samples containing unassigned individuals cannot be corrected as 1) corrections for null alleles rely on HWE (which is not expected to be found in a mixed sample with individuals from several populations, e.g., Wahlund effects) and 2) it is impossible to know (without extensive sequencing) which individuals are true homozygotes or heterozygotes for the null allele.

Even though the presence of null alleles in microsatellite data sets can be detected through statistical analyses (e.g., MICROCHECKER software, van Oosterhout et al. 2004), it is problematic to study the effect on assignment testing because the presence/absence of null alleles in specific individuals is not known. One approach that allows for quantifying the bias on results of assignment tests that are caused by null alleles is to simulate populations and introduce known null alleles. In addition, simulations would enable analysis of whether higher frequencies of null alleles lead to more pronounced biases. Such information is particularly useful for population genetic studies on taxa that are prone to null alleles. This study aims to increase our understanding of how null alleles specifically affect assignment testing by utilizing simulated population data that represent a range of realistic scenarios.

Materials and Methods

Simulated Populations

The software EASYPOP 1.7 (Balloux 2001) was used to simulate 3 data sets with varying number of microsatellite loci (4, 12, or 20). The simulations consisted of 60 population sets, each consisting of 4 populations with 150 individuals in each population and equal sex ratios. The simulation conditions included a mutation rate of 0.002 with 80% single-step mutations and 20% infinite allele mutations [mutation rate was based on published data (cf., Weber and Wong 1993)]. Gene flow followed an island model with equal zygotic migration rates among sexes. The initial number of alleles was set to 30 per locus and randomly assigned across individuals, and the simulations were run for 1000 generations. To achieve different levels of genetic differentiation among populations, the level of gene flow was set to 0.100, 0.050, 0.025, or 0.010 (low, low–medium, high–medium, and high genetic differentiation, respectively). The same simulation settings (including migration rate) were used 5 times to generate variable F_{ST} estimates at 4 different migration rates resulting in a total of 5 population sets per simulation setting. These 5 population sets are not replicates as the EASYPOP software uses considerable amount of randomizations (Balloux 2001) leading to great variability among runs. Each population set should, therefore, be considered independent. The specific aim of the study was to describe how null alleles affected assignment testing. Hence, no subsampling of populations was made as subsampling might introduce additional variation that is not related to null alleles.

Null Alleles

One allele at each locus in each population set was chosen at random to be a null allele. Consequently, all heterozygotes including the null allele were transformed to homozygotes for the alternative allele and all homozygotes for the null allele were transformed to missing data as this is how their genotypes would appear when being typed. This procedure created alternative loci data sets that contained known frequencies of null alleles. The frequency of null alleles per locus within populations varied from 0.000 to 0.913, whereas the average frequency of null alleles within population sets ranged from 0.021 to 0.202. Although it is likely that real microsatellite data may contain more than one null allele at affected loci, the true number of null alleles is rarely, if ever, known. Thus, the single null allele simulated here is appropriate because multiple null alleles would have identical effects as homoplasy, which is commonly observed in microsatellites where alleles with identical size might have different descent.

Statistics

ARLEQUIN 3.1 (Excoffier et al. 2005) was used to calculate expected and observed levels of heterozygosity and for assessing whether genotype frequencies were consistent with HWE (exact tests, Guo and Thompson 1992). The GENEPOP 3.4 software package (Raymond and Rousset 1995) was used to estimate Weir and Cockerhams's (1984) unbiased estimator of Wright's *F*-statistics (F_{ST}) among populations within each population set. The Anderson–Darling test (Stephens 1974) was used to test if the data were normally distributed. Parametric tests were used when the data were normally distributed, whereas in case where data were not normally distributed nonparametric tests were used. The sequential Bonferroni technique (Rice 1989) was used to adjust significance levels in cases with multiple tests.

The objective of this study was not to compare different software for assignment testing. However, 2 of the most commonly used approaches, as applied in STRUCTURE

and GENECLASS, were used to study how null alleles affect assignment testing. Two different approaches were used that employ the model-based clustering assignment method as implemented in the STRUCTURE 2.2 software (Pritchard et al. 2000; Falush et al. 2003, 2007). In the first approach, no correction was done to the raw data to account for null alleles, and homozygotes for null alleles were consequently coded as missing data (hereafter STRnull). The second approach utilized the feature of the latest version of STRUCTURE (ver. 2.2) that can take into account null alleles by utilizing the feature of the software to handle recessive data. Homozygotes for null alleles were accordingly coded as homozygotes for a recessive allele (as suggested by the authors of the software), and STRUCTURE was instructed that the data contained recessive alleles (STRrec).

It is somewhat problematic to perform "self-assignment" tests with STRUCTURE because the software does not give information about which clusters to compare when evaluating the assignment success in data sets unaffected and affected by null alleles. For instance, individuals from say population A would be assigned to cluster 1 when using populations sets unaffected by null alleles. However, when using population A and data affected by null alleles (the same data set, but now with null alleles), individuals might assign to cluster 2. Cluster 1 from the unaffected population set should therefore be compared with cluster 2 from the, by null allele, affected population set. However, as STRUCTURE will not give you any information about which clusters to compare (in this case, cluster 1 and cluster 2), it is not possible to compare assignment success between data sets affected and unaffected by null alleles. To overcome this problem, the first 50 individuals were removed from each population. These 50 individuals were then moved into a separate group of unassigned individuals. The remaining 100 individuals per population were used as potential source populations, making 4 potential source populations and 4 samples of unassigned individuals. The first 4 source populations were made up of 100 individuals and the following 4 populations contained 4 sets of unassigned individuals.

The accuracy at which STRUCTURE was able to assign individuals to the correct source population was estimated by using the option "USEPOPINFO" for the potential source populations, whereas this option was turned off for individuals who were being assigned. This means that STRUCTURE took into account prior population information for the potential source populations, whereas no prior information was provided for the unassigned individuals. Except for the USEPOPINFO option (and "row of recessive alleles" in the STRrec tests), all parameters were set to default and K was set to 4, and each STRUCTURE analysis consisted of a burn-in period of 10 000 followed by 100 000 replicates. A longer burn-in period of 100 000 and a larger number of replicates, 1 000 000, were initially used to assess when parameters such as likelihood, F_{ST}, and alpha stabilized (data not shown). The initial analyses showed that a burn-in of 10 000 and 100 000 replicates would be adequate for these assignment tests (data not shown). The data reported from the STRUCTURE analyses are the average proportions of individuals per population set that were correctly assigned to their original population.

Assignment testing was also performed by using the software package GENECLASS 2.0.g (Piry et al. 2004). The Bayesian method of Rannala and Mountain (1997) as implemented in GENECLASS was selected for performing self-assignment tests (leave one out procedure) to assess the assignment accuracy. This means that each individual was excluded from the population set of potential source populations and then assigned to one of the source populations. The results reported here are the proportions of correctly assigned individuals within population sets. The Bayesian assignment method was chosen as most studies use this statistical approach, and it has been shown to perform better than frequency or distance-based methods (Cornuet et al. 1999; Hansen et al. 2001).

Results

Genetic Variability and Levels of Differentiation

The number of alleles per locus in the absence of null alleles within populations ranged from 3 to 17, observed within-population heterozygosity ranged from 0.107 to 0.927, and the expected heterozygosity ranged from 0.115 to 0.904. The percentage of loci within population sets that showed deviations from HWE expectations ranged from 0.0% to 18.8%, whereas the range was from 0.0% to 12.5% after sequential Bonferroni corrections ($k = 4$, 12, and 20 for the 4, 12, and 20 loci data sets, data not shown). The number of alleles per locus, after introducing null alleles, within populations ranged from 2 to 16, and the observed heterozygosity ranged from 0.016 to 0.920, whereas the expected heterozygosity within populations varied from 0.072 to 0.899. The proportion of loci within population sets after introducing null alleles that deviated from HWE ranged from 16.7% to 87.5% before corrections for multiple tests. After sequential Bonferroni corrections ($k = 4$, 12, and 20 for the 4, 12, and 20 loci data sets), the number of loci deviating from HWE ranged from 21.3% to 81.3% (data not shown).

The degree of genetic differentiation, F_{ST}, for evaluation of the effect of null alleles on the STRUCTURE analysis was based on sets of 100 individuals ($F_{ST\text{-}100}$), whereas F_{ST} for the GENECLASS analysis was based on sets of 150 individuals ($F_{ST\text{-}150}$). Population sets affected by null alleles significantly overestimated $F_{ST\text{-}100}$ [average of 0.003 units in the 4 and 12 loci data sets and with 0.004 in the 20 loci data sets, Wilcoxon matched-pairs signed-ranks test (WMPSRT) $P < 0.05$ for all tests] and $F_{ST\text{-}150}$ (average of 0.004 for all loci data sets, WMPSRT $P < 0.05$ for all tests). The degree of overestimation of $F_{ST\text{-}100}$ and $F_{ST\text{-}150}$ increased with higher F_{ST} in all data sets ($P < 0.05$ for all tests, excluding the 4 loci $F_{ST\text{-}150}$ data set). The $F_{ST\text{-}100}$ and $F_{ST\text{-}150}$ estimates from the null-allele free population sets and the population sets affected by null alleles were highly correlated in all data

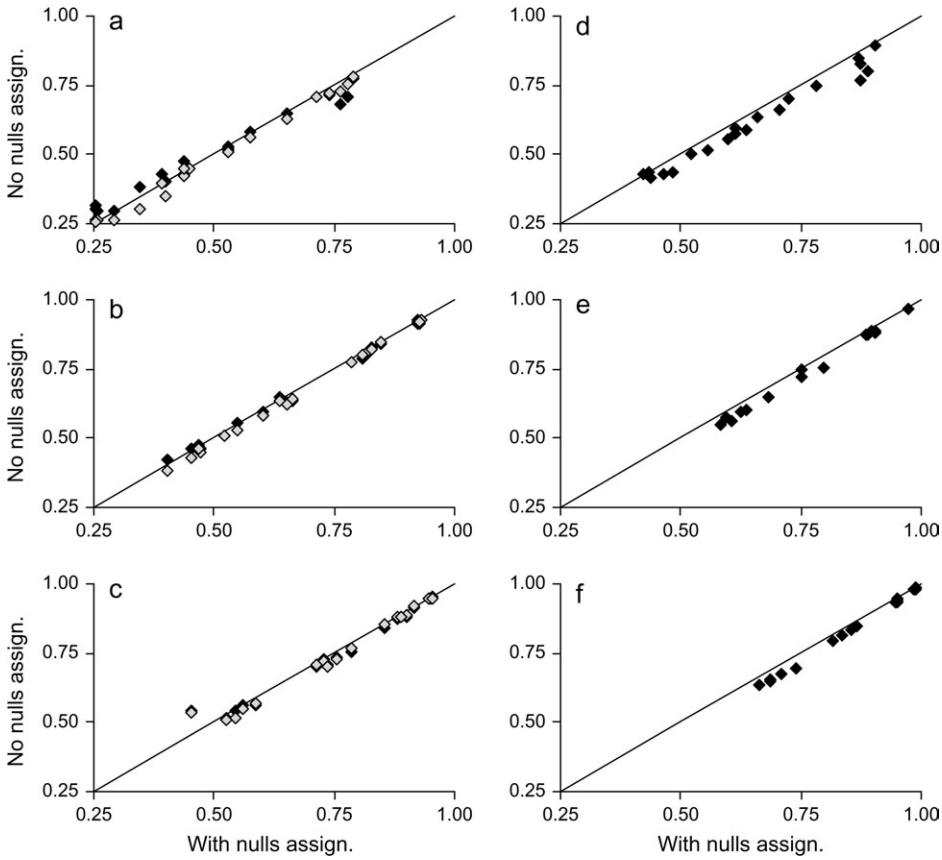

Figure 1. Plot of the proportion of correctly assigned individuals based on STRUCTURE (graphs **a**–**c**) and GENECLASS (graphs **d**–**f**) in data sets unaffected (No nulls assign.) and affected by null alleles (With nulls assign.). Filled diamonds represent data affected by null alleles, whereas gray diamonds represent results of null homozygotes being coded as recessive alleles, based on 4 (graphs **a** and **d**), 12 (graphs **b** and **e**), and 20 (graphs **c** and **f**) microsatellite loci. The diagonal lines represent 1:1 ratios. Data points below the diagonal line indicate incidents when null alleles reduced the proportion of individuals correctly assigned.

sets [Spearman rank-order correlations (SROC) $P < 0.05$, for all $F_{ST\text{-}100}$ and $F_{ST\text{-}150}$ data sets]. Higher null-allele frequencies averaged within population sets did not lead to significant larger discrepancies in $F_{ST\text{-}100}$ or $F_{ST\text{-}150}$ estimates in any of the data sets (SROC $P < 0.05$, for all tests, data not shown).

Assignment Tests

The proportion of correctly assigned individuals within population sets when using STRUCTURE (STR) ranged from 0.252 to 0.956. Although the proportion of correctly assigned individuals, when treating null alleles as missing data (STRnull), ranged from 0.262 to 0.954 and when null homozygotes were coded as recessive alleles (STRrec), it varied from 0.255 to 0.954 (Figure 1a–c).

Null alleles significantly lowered the proportion of correctly assigned individuals in all but the STRnull 4 loci data set (−0.4, 0.6, and 0.5 percent units lower in STRnull compared with STR in the 4, 12, and 20 loci data sets, WMPSRT $P = 0.452$, 0.033, and 0.002, respectively, Figure 1a–c). Similarly, a reduction in the proportion of correctly assigned individuals was noted in all STRrec data sets compared with when null alleles were absent (1.3, 1.2, and 0.6 percent units lower in STRrec compared with STR for the 4, 12, and 20 loci, WMPSRT $P = 0.004$, 0.001, and 0.003, respectively, Figure 1a–c). The 4 loci STRrec data sets showed significantly lower assignment success than did the corresponding STRnull data, whereas no differences were found among the 12 and 20 loci data sets (1.8, 0.5, and 0.1 percent units lower for STRrec than the corresponding data from the 4, 12, and 20 STRnull loci data sets, WMPSRT $P = 0.009$, 0.210, and 0.799, respectively, Figure 1a–c).

On average, the proportion of correctly assigned individuals was reduced by 0.2 percent units [standard error of the mean (SEM) = 0.003] in the presence of null alleles (STRnull) and with 1.0 percent units (SEM = 0.002) when null homozygotes were coded as recessive alleles (STRrec). Increased, average intrapopulation set, null-allele frequencies (note that the frequency of null alleles were estimated on the entire population set containing 600 individuals) lead to a significant regression between the proportion of correctly assigned individuals within in the 12 STRnull loci

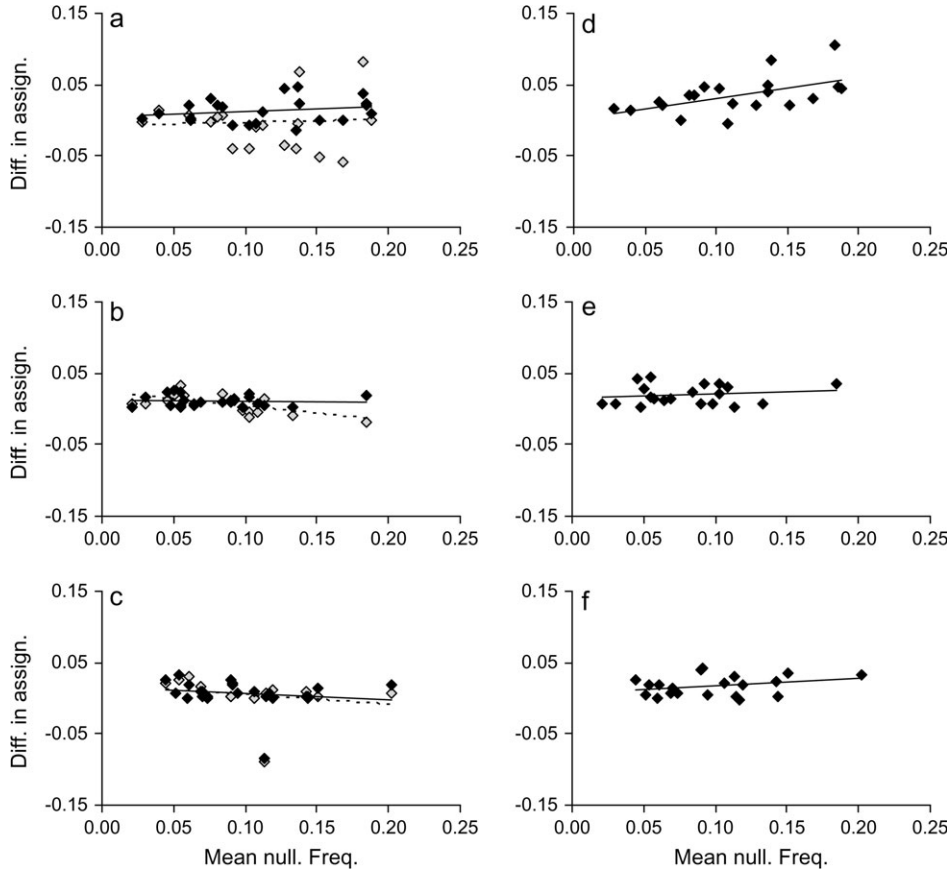

Figure 2. Differences in the proportion of correctly assigned individuals (Diff. in assign) based on STRUCTURE (graphs **a**–**c**) and GENECLASS (graphs **d**–**f**) caused by average frequency of null alleles within population sets (Mean null. freq.) based on 4 (graphs **a** and **d**), 12 (graphs **b** and **e**), and 20 (graphs **c** and **f**) microsatellite loci. Filled diamonds represent data affected by null alleles, and gray diamonds in graphs (**a**–**c)** represent data where null homozygotes were coded as recessive alleles. The solid lines indicate trendlines for data unaffected by null alleles, and the dotted lines represent data where null homozygotes are coded as recessive alleles.

data set (regression $P = 0.004$, adjusted $R^2 = 0.34$), whereas no significant correlations were found in the 4 or 20 data sets (SROC $P > 0.05$ for the 4 and 20 loci data sets, Figure 2a–c) or in the STRrec data sets (regression $P > 0.05$ for the 4 and 12 loci data sets and SROC $P > 0.05$ for the 20 loci data sets, Figure 2a–c). The proportion of correctly assigned individuals at different F_{ST} values from the alternative populations sets (STR, STRnull, and STRrec) were highly correlated in all data sets (regression $P < 0.001$, for the 4 and 12 loci data sets, SROC $P < 0.001$ for the 20 loci data sets, Figure 3a–c).

The GENECLASS-based assignment tests showed that the proportion of correctly self-assigned individuals in the absence of null alleles within population sets ranged from 0.422 to 0.988, and after introduction of null alleles, the corresponding proportion of correctly assigned individuals varied from 0.418 to 0.990 (Figure 1d–f). Null alleles significantly lowered the proportion of correctly assigned individuals in all loci data sets (WMPSRT $P < 0.001$ for all tests, Figure 1d–f).

The proportion of correctly assigned individuals was on average, across loci data sets, reduced by 2.4 percent units (SEM = 0.006) in the presence of null alleles. Higher average intrapopulation set null-allele frequencies lead to increased differences in the proportion of correctly assigned individuals within the 4 loci data set (regression $P = 0.013$, adjusted $R^2 = 0.26$), whereas no significant correlations were found in the 12 or 20 loci data sets (SROC $P > 0.05$ for all tests, Figure 2d–f). The proportion of correctly assigned individuals at different F_{ST} levels from the alternative population sets were highly correlated in all data sets (regression $P < 0.001$ for the 4 loci data set and SROC $P < 0.001$, for the 12 and 20 loci data sets, Figure 3d–f).

The proportion of correctly assigned individuals was tightly associated with the level of genetic differentiation in both STRUCTURE- and GENECLASS-based assignments tests (SROC $P < 0.001$, for all tests). Moreover, both data sets with and without null alleles (Figure 3a–f) showed a strong logarithmic regression (R^2 values were higher for a logarithmic model than for a linear model in all data sets,

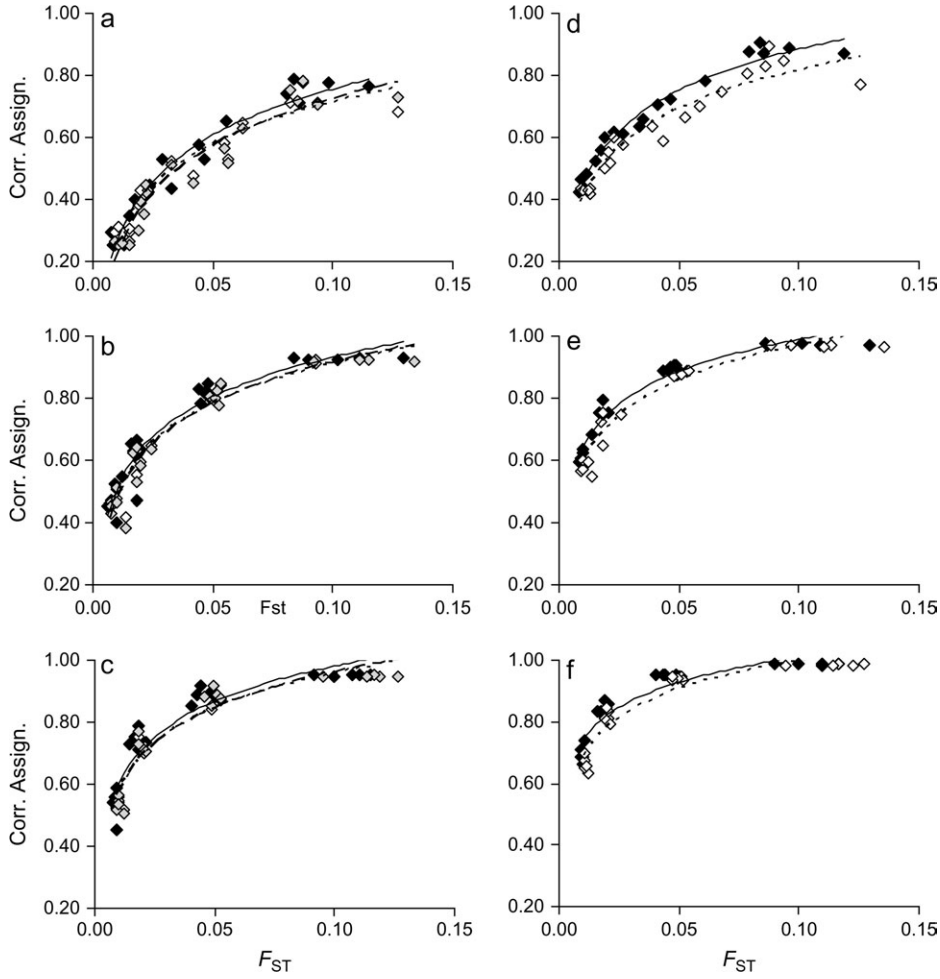

Figure 3. Regression between F_{ST} estimates ($F_{ST\text{-}100}$ and $F_{ST\text{-}150}$) and proportion of correctly assigned individuals based on STRUCTURE (graphs **a**–**c**) and GENECLASS (graphs **d**–**f**) in data sets unaffected (filled diamonds) and affected by null alleles (open diamonds), and in graphs (**a**–**c)**, the gray diamonds represent results from STRUCTURE when null homozygotes were treated as recessive alleles, based on 4 (graphs **a** and **d**), 12 (graphs **b** and **e**), and 20 (graphs **c** and **f**) microsatellite loci. Regressions of F_{ST} on the proportion of correctly assigned individuals. Solid lines indicate null-allele unaffected data, dotted lines data affected by null alleles, and the broken lines when null homozygotes were treated as recessive alleles.

note that $F_{ST\text{-}100}$ and $F_{ST\text{-}150}$ estimates were transformed, natural logarithm, prior to regression analysis, data not shown) between the level of genetic differentiation and the proportion of correctly assigned individuals. The proportion of correctly assigned individuals increased with the level of genetic differentiation and seemed to asymptote at F_{ST} levels of 0.1 and higher when maximum assignment efficiency was reached (Figure 3a–f).

Discussion

The level of genetic differentiation (i.e., F_{ST}), number of alleles, heterozygosity, sample sizes, and the number of loci used in the present simulation study are within the range that is commonly encountered in empirical studies. Under these conditions, the data suggest that null alleles will slightly reduce the proportion of correctly assigned individuals in both STRUCTURE- and GENECLASS-based assignment tests. However, the low magnitude of the effects suggests that microsatellites with null alleles can still be used for assignment testing. Even though, some populations had very high null-allele frequencies (up to 0.913 at a single locus), the effects on assignment success were moderate.

Cornuet et al. (1999) showed that likelihood-based assignment testing was robust even after introducing low frequencies (0.01) of null alleles and still outperformed distance-based methods. Hauser et al. (2006) using an empirical approach suggested that modest abuse of the assumption of absence of null alleles had only a small effect on the accuracy of assignment testing. The data presented in the current study indicate that the accuracy of assignment testing is slightly, though significantly, reduced by the

presence of null alleles (except in the STRnull 4 loci data set, cf., Figure 1a–f).

No significant relationship between the average frequency of null alleles within population sets and the difference in proportion correctly assigned individuals in STRUCTURE or GENECLASS could be detected (cf., Figure 2a–f), except for in the case of the STRrec 12 loci and the GENECLASS 4 loci data sets. As the effect of single loci are expected to impact the outcome of assignment testing more when the total numbers of loci is low, it may be that the effect will be reduced as additional loci are included. This does not, however, explain why there was a significant increase of differences in assignment tests when 12 loci were analyzed (STRrec). Though the effect of null alleles was rather moderate in all assignment tests as the proportion of correctly assigned individuals were on average only reduced by 0.2 and 1.0 percent units when analyzed with STRUCTURE- and the GENECLASS-based assignment by 2.4 percent units.

On the whole, the number of loci and level of genetic differentiation seem to have greater effects on the accuracy of assignment testing than does the presence/absence of null alleles (cf., Figure 3a–f). Assignment accuracy seems to asymptote at F_{ST} values close to 0.1 when 12 loci or more are used even in the presence of null alleles (Figure 3a–f), and it is unlikely that simulations including higher F_{ST} than used here would improve assignment success. The STRUCTURE software can take into account that some or all missing data are caused by null alleles. However, under the conditions simulated here, there was no advantage in treating null alleles as recessive alleles. In fact, in some cases (when using 12 or 20 loci), the effect of treating null alleles as recessive alleles lowered the proportion of correctly assigned individuals slightly (though significantly) compared with when null-allele homozygotes were treated as missing data.

Chapuis and Estoup (2007) documented larger bias in F_{ST} with increasing F_{ST} in the presence of null alleles. The results from the present simulation show similar results with a moderate overestimation of F_{ST} which was most pronounced when population structure was the greatest. The present study suggests that even though F_{ST} was slightly overestimated (F_{ST} increased between 0.003 and 0.004), loci affected by null alleles are still useful for population genetic studies. Nevertheless, as the effect of null alleles on estimates of effective population size, gene flow, or any other statistics that rely on the accuracy of F_{ST} estimates has not been evaluated, markers prone to null alleles should be used with caution.

Conclusions

Null alleles had a moderate effect on the accuracy of assignment testing under the simulated conditions in this study. These effects are not detrimental for assignment testing as long as limitations and caveats are taken into account. Similarly, F_{ST} was only slightly overestimated in the presence of null alleles, and it is unlikely that null alleles will have major impacts on conclusions regarding presence or absence of genetic differentiation. It seems evident that increased number of loci and degree of genetic differentiation has more significant effect on the accuracy of assignment testing than the presence of null alleles. This information is valuable for population genetic studies of taxa that are prone to null alleles as it may enable geneticists to utilize loci affected by null alleles. Microsatellite development for these species is often very time consuming and expensive because large numbers of markers have to be developed and optimized before a set of markers with low frequencies of null alleles can be achieved. This is not, however, to say that efforts should not be made to use loci that show low frequency of null alleles. On the contrary, microsatellites that are not prone to null alleles should always be preferred as they are less ambiguous and more powerful for assignment tests.

Acknowledgments

The author likes to thank Jan F. Cordes and Jan R. McDowell for helpful comments and suggestions on earlier versions of the manuscript.

References

Balloux F. 2001. EASYPOP (version 1.7): a computer program for the simulation of population genetics. J Hered. 92:301–302.

Carlsson J, Morrison CL, Reece KS. 2006. Wild and aquaculture populations of the eastern oyster compared using microsatellites. J Hered. 97:595–598.

Chapuis M-P, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol. 24:621–631.

Chapuis M-P, Loiseau A, Michalakis Y, Lecoq M, Estoup A. 2005. Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*. Mol Ecol Notes. 5:554–557.

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M. 1999. New methods employing multilocus genotypes to select or exclude populations as origin of individuals. Genetics. 153:1989–2000.

Dakin EE, Avise JC. 2004. Microsatellite null alleles in parentage analysis. Heredity. 93:504–509.

de Sousa SN, Finkeldey R, Gailing O. 2005. Experimental verification of microsatellite null alleles in Norway spruce (*Picea abies* [L.] Karst.): implications for population genetic studies. Plant Mol Biol Rep. 23:113–119.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 164:1567–1587.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes. 7:574–578.

Guo SW, Thompson EA. 1992. Performing the exact test for Hardy-Weinberg proportion for multiple alleles. Biometrics. 48:361–372.

Hansen MM, Kenchington E, Nielsen EE. 2001. Assigning individual fish to populations using microsatellite DNA markers: methods and applications. Fish Fish. 2:93–112.

Hauser L, Seamons TR, Dauer M, Naish KA, Quinn TP. 2006. An empirical verification of population assignment methods by marking and

parentage data: hatchery and wild steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA. Mol Ecol. 15:3157–3173.

Lehmann T, Besanky NJ, Hawley WA, Fahey TG, Kamau L, Collins FH. 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. Mol Ecol. 6:243–253.

McGoldrick DJ, Hedgecock D, English LJ, Baoprasertkul P, Ward RD. 2000. The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): selection and null alleles. J Shellfish Res. 19:779–788.

Nielsen EE, Hansen MM, Bach L. 2001. Looking for a needle in a haystack: discovery of indigenous salmon in heavily stocked populations. Conserv Genet. 2:219–232.

Pemberton JM, Slate J, Bancroft DR, Barrett JA. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. Mol Ecol. 4:249–252.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A. 2004. GeneClass2: a Software for genetic assignment and first-generation migrant detection. J Hered. 95:536–539.

Primmer CR, Koskinen MT, Piironen J. 2000. The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proc R Soc Lond B Biol Sci. 267:1699–1704.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Rannala B, Mountain J. 1997. Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA. 94:9197–9201.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2) population genetics software for exact test and ecumenicism. J Hered. 86:248–249.

Reece KS, Ribeiro WL, Gaffney PM, Carnegie RB, Allen SK Jr. 2004. Microsatellite marker development and analysis in the eastern oyster, *Crassostrea virginica*: confirmation of null alleles and non-Mendelian segregation ratios. J Hered. 95:355–361.

Rice WR. 1989. Analyzing tables of statistical tests. Evolution. 43: 223–225.

Shaklee JB, Beacham TD, Seeb L, White BA. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fish Res. 43:45–78.

Stephens MA. 1974. EDF statistics for goodness of fit and some comparisons. J Am Stat Assoc. 69:730–737.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 4:535–538.

Waples RS, Gaggiotti OE. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. 15:1419–1439.

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. Hum Mol Genet. 2:1123–1128.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38:1358–1370.

Received March 12, 2008
Accepted May 5, 2008

Corresponding Editor: Jose V. Lopez