

1 **Reliability assessment of null allele detection: inconsistencies between and within**

2 **different methods**

3

4 M.J. Dąbrowski<sup>1,2</sup>, M. Pilot<sup>1,3</sup>, M. Kruczyk<sup>2,4</sup>, M. Żmihorski<sup>1</sup>, H.M. Umer<sup>2</sup>, J. Gliwicz<sup>1</sup>

5 <sup>1</sup>Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679 Warsaw,

6 Poland

7 <sup>2</sup>Department of Cell and Molecular Biology, Uppsala University, Box 596, 751 24 Uppsala,

8 Sweden

9 <sup>3</sup>School of Life Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK

10 <sup>4</sup>Postgraduate School of Molecular Medicine, Zwirki i Wigury 61, 02-091 Warsaw, Poland

11 Keywords: genotyping errors, heterozygosity, *Microtus oeconomus*, microsatellite loci, null

12 alleles; bottleneck

13 Corresponding author: M.J. Dąbrowski, Museum and Institute of Zoology, Polish Academy of

14 Sciences, Wilcza 64, 00-679 Warsaw, Poland; Fax:+48 22 6296302; [michal@miiz.waw.pl](mailto:michal@miiz.waw.pl)

15

16

17

18

19

20 Running title: Assessment of null allele estimation methods

21

22

23 Abstract

24           Microsatellite loci are widely used in population genetic studies, but the presence of  
25 null alleles may lead to biased results. Here we assessed five methods that indirectly detect  
26 null alleles, and found large inconsistencies among them. Our analysis was based on 20  
27 microsatellite loci genotyped in a natural population of *Microtus oeconomus* sampled during 8  
28 years, together with 1200 simulated populations without null alleles, but experiencing  
29 bottlenecks of varying duration and intensity, and 120 simulated populations with known null  
30 alleles. In the natural population, 29% of positive results were consistent between the methods  
31 in pairwise comparisons, and in the simulated dataset this proportion was 14%. The positive  
32 results were also inconsistent between different years in the natural population. In the null-  
33 allele-free simulated dataset, the number of false positives increased with increased bottleneck  
34 intensity and duration. We also found a low concordance in null allele detection between the  
35 original simulated populations and their 20% random subsets. In the populations simulated to  
36 include null alleles, between 22% and 42% of true null alleles remained undetected, which  
37 highlighted that detection errors are not restricted to false positives. None of the evaluated  
38 methods clearly outperformed the others when both false positive and false negative rates  
39 were considered. Accepting only the positive results consistent between at least two methods  
40 should considerably reduce the false positive rate, but this approach may increase the false  
41 negative rate. Our study demonstrates the need for novel null allele detection methods that  
42 could be reliably applied to natural populations.

43

44

## 45 **Introduction**

46        Highly polymorphic microsatellite markers are widely applied in population genetic  
47 studies since their discovery in the late 1980s. The improvement of polymerase chain reaction  
48 (PCR) and sequencing technologies allowed the use of these molecular markers to spread fast  
49 and wide into many research fields (see Guichoux et al. 2011 for review). However, the  
50 potential occurrence of "null alleles", i.e. alleles that fail to amplify during the PCR, creates a  
51 disadvantage in using these markers (Oddou-Muratorio et al. 2009). A null allele occurs when  
52 an incompatibility between any of the two locus-specific primers and its complementary  
53 target region causes the PCR amplification of an allele to fail. Such incompatibilities may be  
54 caused by mutations in the primer target region within one species, or between different  
55 species (in case of cross-species amplification) (Callen et al. 1993, Primmer et al. 1995, Jarne  
56 & Lagoda 1996). In some cases, long alleles may amplify much less efficiently than shorter  
57 ones, and therefore may appear as null alleles (Wattier et al. 1998). Low template  
58 quality/quantity can also result in the absence of amplification product and may be interpreted  
59 as the presence of a null allele (Garcia de Leon et al. 1998).

60        Null alleles have been reported in many species, e.g. humans (Callen et al. 1993),  
61 deers (Pemberton et al. 1995), bears (Paetkau & Strobeck 1995), voles (Ishibashi et al. 1996),  
62 fish (McCoy et al. 2001), crayfish (Walker et al. 2002), and oystercatchers (Van Treuren  
63 1998). The detection of null alleles is an important step in population genetic data analysis, as  
64 their presence may strongly bias the estimates of population genetics parameters (Pemberton  
65 et al. 1995, Chapuis & Estoup 2007). For example, the accuracy of assignment of individuals  
66 to populations may be reduced and  $F_{ST}$  significantly overestimated (Carlsson 2008). The

67 presence of null alleles may also lead to an incorrect exclusion of a significant number of true  
68 parents in parentage analyses (Dakin & Avise 2004). Despite this, very few studies on  
69 population genetic structure and genetic parentage report estimates of null allele frequencies  
70 in their data (see Dakin & Avise 2004 for review).

71         Several methods for null allele estimation are currently available (Dempster et al.  
72 1977, Chakraborty et al. 1992, Brookfield 1996, Summers & Amos 1997, Kalinowski & Taper  
73 2006). They are based on comparing observed and expected heterozygosity for each locus to  
74 identify loci with significant heterozygote deficit. This approach is based on the fact that a  
75 heterozygous locus with a null allele would be scored as a homozygote, since only the visible  
76 allele is detected. Crucially, all these methods assume that a population is in Hardy-Weinberg  
77 Equilibrium (HWE), and that all observed deviations towards heterozygote deficit result from  
78 the presence of null alleles. The main difference between these methods lies in the way blank  
79 results (i.e. individuals without any detectable PCR product at a particular locus) are  
80 interpreted. Some methods consider blank results as null allele homozygotes, while others  
81 classify them as PCR failures resulting from low DNA quality or human errors; some methods  
82 attempt to differentiate between these two cases (see Supplementary Material for details).  
83 Another difference lies in the approaches used for null allele frequency estimation. While the  
84 estimates of Chakraborty et al. (1992) and Brookfield (1996) are obtained analytically,  
85 estimates of Dempster et al. (1977), Summers & Amos (1997), and Kalinowski & Taper  
86 (2006) are achieved through iterative optimisation (see Supplementary Material for details).

87         The above methods showed good to moderate accuracy in estimating frequencies of  
88 known null alleles in populations simulated assuming HWE (Kalinowski & Taper 2006,

89 Chapuis & Estoup 2007). Specifically, Kalinowski & Taper (2006) demonstrated that their  
90 method performs better than the methods of Chakraborty et al. (1992) and Summers & Amos  
91 (1997), while Chapuis & Estoup (2007) showed that the method of Dempster et al. (1977)  
92 performs better than the methods of Chakraborty et al. (1992) and Brookfield (1996).  
93 However, Chapuis & Estoup (2007) also showed that the three methods they tested performed  
94 worse when applied to two empirical datasets from natural populations, where the presence of  
95 null alleles was confirmed by their successful amplification after the primers were re-  
96 designed. Moreover, in one of these populations heterozygote deficit remained significant  
97 even after the null allele was successfully amplified with the new primers. Although this result  
98 was attributed to the presence of additional null alleles (Chapuis & Estoup 2007), the  
99 observed heterozygote deficit could have resulted from other factors such as small sample  
100 size, high inbreeding levels, or immigration.

101         The assumption of HWE, common among the methods described above, can be  
102 problematic when estimating null alleles in microsatellites scored from natural populations,  
103 since natural populations never strictly comply with the assumptions of Hardy-Weinberg law  
104 (i.e. infinite size, random mating, lack of mutations, migration and natural selection).  
105 Crucially, some of the factors causing deviations from HWE also lead to heterozygote deficit,  
106 namely inbreeding, assortative mating, population structure or immigration from a genetically  
107 distinct source (Wahlund effect), and disruptive selection (Avisé 2004). Heterozygote deficit  
108 generated by such population mechanisms may be interpreted as the presence of null alleles,  
109 thus leading to false positives. On the other hand, phenomena such as disassortative mating or  
110 balancing selection can lead to heterozygote excess, which may result in failure to detect true

111 null alleles. The effect of other population genetic processes is less obvious. For example, a  
112 bottleneck leads to loss of alleles and decline in heterozygosity, but at least under some  
113 conditions it may also lead to temporary heterozygote excess (Cornuet & Luikart 1996).  
114 Fluctuations in population size, especially if associated with immigration during the growth  
115 phase, may lead to temporal fluctuations between heterozygote excess and deficit. In addition,  
116 taking a small subsample from a population (which also effectively occurs during founder  
117 events) may result in heterozygote deficit in some loci and heterozygote excess in others, due  
118 to the stochasticity of the sampling procedure. This may lead to detection of false null alleles  
119 in loci with heterozygote deficit.

120        Many population genetic studies are based on small sample sizes, and in many cases,  
121 study populations themselves are small (and therefore subject to strong drift), fluctuate in size,  
122 and exhibit considerable deviations from random mating. Such populations do not comply  
123 with the assumption that heterozygote deficit results solely from the presence of null alleles.  
124 However, the methods assuming HWE are commonly applied to such cases (e.g. see the  
125 review by Dakin & Avise 2004). In this study, we address the problem of detecting null alleles  
126 in populations that undergo demographic changes and deviate from HWE, and we assess  
127 reliability of the five widely used methods (Dempster et al. 1977, Chakraborty et al. 1992,  
128 Brookfield 1996, Summers & Amos 1997, Kalinowski & Taper 2006) in such non-equilibrium  
129 conditions. For this purpose, we apply these methods to a natural population of root vole,  
130 *Microtus oeconomus*, which was sampled over an eight-year period, and underwent  
131 substantial density fluctuations during this time. Additionally, in order to test whether  
132 population-level factors may lead to the detection of false null alleles, we analysed 1200

133 simulated populations without null alleles, but affected by a bottleneck with varying levels of  
134 intensity and duration.

135

## 136 **Materials and Methods**

### 137 *Analysed datasets*

138         We analysed 20 nuclear microsatellite loci in a population of root vole, *Microtus*  
139 *oeconomus*, which was extensively sampled over an eight-year period, and underwent a 7.7-  
140 fold change in average density during this time. Detailed information about demography and  
141 genetic variability of this population obtained from previous studies (Gliwicz & Jancewicz  
142 2004, Gliwicz & Dąbrowski 2008, Dąbrowski 2010, Pilot et al. 2010) allowed us to follow  
143 temporal changes in the estimated null allele frequencies and compare different methods of  
144 their detection.

145         In order to assess the effect of demographic changes and resulting population genetic  
146 changes on null allele detection rates under controlled conditions (i.e. with known - rather  
147 than estimated – genetic composition and demographic history), we simulated 1200  
148 populations without null alleles, but with varying level and duration of a bottleneck. The  
149 simulated data allowed us to explore the effect of demographic changes on inconsistencies in  
150 null allele detection that were observed in the natural population. In addition, we created 20%  
151 subsets of the simulated populations by random sampling, to assess the effect of population  
152 sub-sampling on null allele detection. Finally, we introduced null alleles into the earlier  
153 simulated populations in order to (1) assess the performance of each method of null allele  
154 estimation in detecting known null alleles, and (2) assess the empirical relationship between

155 the frequency of null alleles and the frequency of null allele homozygotes in non-equilibrium  
156 populations.

157         We used two general approaches for null allele detection. The first approach was based  
158 on methods assessing heterozygote deficit, as described above (Dempster et al. 1977,  
159 Chakraborty et al. 1992, Brookfield 1996, Summers & Amos 1997, Kalinowski & Taper  
160 2006). We applied this approach to both the natural population and the simulated populations.  
161 The second approach was based on the comparison of genotypes between parent-offspring  
162 pairs, and was included here as the method that does not assume HWE. However, it could be  
163 only applied to the natural population.

164

165 *Sample collection from the natural population*

166         The natural population studied was a wild population of the root vole inhabiting a river  
167 valley located in a strict reserve of Białowieża National Park in north-eastern Poland. The 1-  
168 ha field plot was situated on a vast open sedgeland, and was exposed to seasonal flooding.  
169 The root vole is a small rodent with a maximum life span of 18 months (3 months on  
170 average). In the studied population, individuals were reproductively active usually only for  
171 one breeding season. We used a catch-mark-release (CMR) method, with at least three  
172 trapping sessions carried out each year, using 100 live traps placed in a grid of 10 by 10 m.  
173 Mean trapping efficiency was over 90% of all individuals present on the plot (Pilot et al.  
174 2010). The population underwent substantial density changes over the study period, with  
175 average annual densities ranging from 9 to 69 individuals per hectare as estimated in MARK  
176 software, and no individuals trapped in 2007 (Dąbrowski 2010). Such density fluctuations



177 affected kin structure in the population and could be responsible for deviations from HWE  
178 detected in some years (Pilot et al. 2010).

179         We collected tissue samples for genetic analysis from 94% (739) of the individuals  
180 marked from 2000-2008, including 13 recaptures (originally marked in a previous year and  
181 re-trapped in the next year; these samples were not duplicated in the genetic analyses). The  
182 annual numbers of sampled individuals are presented in Table 1.

183

#### 184 Microsatellite genotyping in the wild root vole population

185 Protocols for DNA extraction and microsatellite genotyping are described in detail in the  
186 Supplementary Material. One crucial information to convey here is that there were no blank  
187 results in this dataset, i.e. no individuals had missing data at any locus. Tissue samples were  
188 obtained as biopsies and immediately stored in ethanol, which allowed us to work only with  
189 DNA of good quality. PCR amplification was done using high-quality Taq polymerase  
190 (included in QIAGEN Multiplex PCR Kit), and PCR reactions were repeated up to four times  
191 for samples that initially failed (see Supplementary Material for details). This allowed us to  
192 eliminate any missing data that could have resulted from low quality DNA, human errors and  
193 PCR reagent failures. None of these steps would, however, eliminate missing data resulting  
194 from the presence of null allele homozygotes. Given that our dataset did not contain any  
195 missing data, we can state with a high confidence that no null allele homozygotes existed in  
196 our dataset, which implies that null alleles, if present in this dataset, would only occur in low  
197 frequencies.

198

199 Genetic diversity and null allele detection in the natural population

200 Genetic diversity estimates for the root vole population, including the number of  
201 alleles per locus (N), observed (HO) and expected (HE) heterozygosity, mean polymorphic  
202 information content (PIC) and exclusion probability for the first parent (Exp(1)) were  
203 calculated in CERVUS 3.0 (Marshall et al. 1998), while departures from HWE were estimated  
204 for each locus in GENEPOP v 4.0.10 (Rousset 2008) (Supplementary Table 1).

205 We tested for the presence of null alleles for each year separately using five different  
206 methods: (1) the maximum likelihood (ML) estimator based on observed and expected  
207 heterozygosities described by Chakraborty et al. (1992) with the modification of Brookfield  
208 (1996) which accounts for the presence of null allele homozygotes, as implemented in  
209 MICRO-CHECKER 2.2.1 (van Oosterhout et al. 2004); (2) the ML estimator using chi-square  
210 goodness-of-fit, accounting for the presence of null allele homozygotes during optimization  
211 rounds (Summers & Amos 1997), as implemented in CERVUS 3.0; (3) the ML estimator  
212 accounting for genotyping errors implemented in ML-NullFreq (Kalinowski & Taper 2006);  
213 (4) the ML method using iterative EM (expectation and maximization) of Dempster et al.  
214 (1977) implemented in GENEPOP v4.0.10.; (5) a method based on the comparison of  
215 genotypes of parent-offspring pairs. The algorithms applied in each method (Dempster et al.  
216 1977, Chakraborty et al. 1992, Brookfield 1996, Summers & Amos 1997, Kalinowski & Taper  
217 2006) are described in the Supplementary Material. Hereafter the five methods will be  
218 referred to by the names of the software packages that implement them, namely MICRO-  
219 CHECKER, CERVUS, ML-NullFreq, GENEPOP, and parent-offspring method. As a result of  
220 testing our datasets using these methods, we obtained a binary response variable (presence-

221 absence of null alleles) for each of the 20 loci and each of the five methods tested. In addition,  
222 we tested for the presence of null alleles for the entire dataset with all years pooled (genotypes  
223 of recaptured individuals were not duplicated; see Supplementary Material and  
224 Supplementary Table 2).

225         The application of the parent-offspring method to the root vole population was  
226 possible since a careful reconstruction of its kin structure was available from earlier studies  
227 (Dąbrowski 2010, Pilot et al. 2010, see Supplementary Material). The program CERVUS  
228 allows for a small number of mismatches between parent and offspring genotypes, if the  
229 probability of the estimated relationship is high based on the conformity of the remaining loci.  
230 Therefore, we could use mismatching loci to detect putative null alleles. We created a list of  
231 parent-offspring pairs based on the results of the previous studies on this population  
232 (Dąbrowski 2010, Pilot et al. 2010). The average rate of mismatches between parental and  
233 offspring genotypes estimated using *error rate analysis* implemented in CERVUS was 0.08 (SD  
234 = 0.097). Presence of a null allele in a locus was reported only if the observed mismatch in a  
235 parent-offspring pair fitted the pattern expected by the presence of a null allele. For example,  
236 if a female with genotype AB at a particular locus mates with a male with genotype CN  
237 (where N is a null allele), 50% of their offspring are expected to have genotypes with this null  
238 allele (either AN or BN). Visible genotypes of the father (CC) and the offspring (AA, BB) will  
239 be inconsistent with the father-offspring relationship, therefore creating a mismatch at this  
240 locus. In contrast to the methods based on the heterozygote deficit, the parent-offspring  
241 method does not require the assumption of HWE.

242         The parent-offspring method could not be applied to the entire dataset, but only to

243 closely related individuals, which reduced the sample size to 511 parent-offspring pairs. In  
244 contrast, the other four methods were tested using all sampled individuals. Due to smaller  
245 sample size, the parent-offspring method may detect fewer null alleles compared with the  
246 other methods. However, if each method detects null alleles correctly, the null alleles detected  
247 by the parent-offspring method should be confirmed by the other methods.

248

#### 249 Generation and analysis of simulated datasets

250         Using SPAMs (Parreira et al. 2009), we simulated 1200 populations, each comprised  
251 of 100 individuals with 20 loci. The reason for generating this data was to test the way  
252 bottlenecks affect the detection of putative null alleles. In order to do this, the *one population*  
253 *size change model* with *instantaneous size change* was applied with the following options:  
254 *ancestral pop effective diploid size*: 100,000; *present pop diploid size* (six variants): (1)  
255 99,999, (2) 50,000, (3) 25,000, (4) 10,000, (5) 5,000, (6) 2,500; *duration of event* (four  
256 variants): 3, 30, 300, 3000 generations; and *mutation rate*: 0.0001. The first size change  
257 variant (from 100,000 to 99,999 individuals) was used as a control, where bottleneck effect  
258 was not present. Each combination of the population size change (from ancestral to present  
259 number of individuals) and time of this event was simulated in 50 replicates. In all 1200  
260 simulated populations the presence of putative null alleles was tested using four programs:  
261 CERVUS, GENEPOP, MICRO-CHECKER and ML-NullFreq. Genetic diversity in the simulated  
262 populations was estimated using the same methods as for the natural population of root voles  
263 (see above).

264         To test the effect of random sampling on null allele detection, we randomly selected

265 10% ( $n = 120$ ) of the simulated populations, and from each of them we randomly sampled  
266 20% of individuals. This simulated the effects of three different real life scenarios: (A)  
267 estimation of null allele frequencies based on a subset of individuals sampled from a  
268 population, (B) founder effect, or (C) sudden change in number of individuals within one  
269 breeding season. We tested for the presence of null alleles in each original population ( $n = 100$   
270 individuals) and its random subset ( $n = 20$  individuals, i.e. a 20% subset) separately. Then we  
271 considered only the loci with null alleles detected in at least one of the original populations or  
272 their subsets. For these populations, we calculated the Kendall's coefficient of concordance as  
273 a measure of similarity of null allele detection between the original and the subset  
274 populations.

275         For the next analysis, we selected 120 out of 1200 populations simulated in SPAMP, in  
276 which no null alleles (false positives) were detected by any method. This set included  
277 populations that underwent all levels of the simulated bottleneck. Then, we simulated the  
278 presence of two null alleles in each of these 120 populations using NullAlleleGenerator  
279 (<http://www.lcb.uu.se/papers/dabrowski/NullAlleleGenerator.zip>). NullAlleleGenerator randomly  
280 selected a locus (out of the 20 loci simulated) and changed one random allele into a null  
281 allele. This procedure was repeated for two loci, thus simulating two null alleles per  
282 population. Whenever genotyped in heterozygous form, the allele selected as the null allele  
283 was replaced by the other allele from that locus. Whenever occurring in a homozygous state,  
284 the simulated null allele was marked as a blank result (missing data). This way we obtained  
285 populations with true known null alleles, for which we assessed the performance of CERVUS,  
286 GENEPOP, MICRO-CHECKER and ML-NullFreq in detecting null alleles.

287 Finally, we repeated the previously described procedure of simulating two null alleles  
288 per population for the second time, but unlike in the first case, this time the 120 populations  
289 were selected randomly. In this analysis, we checked the relationship between the frequency  
290 of simulated null alleles and the frequency of null allele homozygotes. In a population under  
291 HWE, the expected frequency of a null allele homozygote is  $p^2$ , where  $p$  is the frequency of  
292 the respective null allele. However, here we simulated populations that underwent a  
293 bottleneck, and thus many of them deviated from HWE. Therefore, we checked empirically  
294 how the frequency of null allele homozygotes depended on the frequency of null alleles. This  
295 was needed for the interpretation of the lack of null allele homozygotes in the natural  
296 population we studied (which also deviated from HWE and underwent substantial  
297 demographic fluctuations).

298

### 299 Statistical analysis

300 In order to investigate the presence/absence of null alleles in a particular locus in  
301 subsequent years, we applied a generalized linear mixed model (GLMM) with binomial error  
302 distribution, and logit link function. We used this approach because several parameters can  
303 potentially affect null allele presence and these parameters need to be included within one  
304 statistical design. Moreover, as the study covers different years and different loci, both year  
305 effect and loci effect have to be included as random variables to avoid pseudoreplications. We  
306 thus used the occurrence pattern of null alleles (present vs. absent) as a dependent variable,  
307 while the method of null allele estimation (marked as 1-5) was used as a fixed categorical  
308 factor. Difference between observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity (hereafter  $H_OH_E$ ),

309 number of individuals trapped in a given year and number of alleles at a particular locus in a  
310 given year were used as three separate covariates in the model, whereas year and locus were  
311 included as random categorical factors. We implemented the GLMM using the “lmer”  
312 package (Bates et al. 2011) in R (R Development Core Team 2011).

313         The expected number of loci with null alleles was estimated as a function of the  
314 number of years analysed. For this purpose we used rarefaction curves implemented in  
315 EstimateS 800 (Colwell 2005). The curves were based on the years resampled in a random  
316 order. We constructed the curves independently for each method of null allele estimation.

317         In order to assess how similar the five methods were in their estimates of null alleles  
318 for a given locus in a particular year, we calculated a similarity index as the probability that a  
319 null allele detected by one method will be confirmed by another. This index was calculated  
320 pairwise between the methods, and visualized with 2x2 contingency tables.

321         We also conducted pairwise comparisons between individual null allele estimations,  
322 independent of year and method (8 years \* 5 methods = 40 estimations; year 2007 was  
323 removed due to lack of voles) using EstimateS 800 (Colwell 2005). In order to understand  
324 what drives similarity among randomly selected null allele estimations, we carried out the  
325 GLM analysis as follows: For each possible pair of estimations (n = 780 pairs) we computed:  
326 (1) temporal distance (ranged from 0 to 8 years), (2) pooled number of null alleles indicated  
327 by two estimations and (3) logical statement (yes or no) indicating whether both estimations  
328 for a given pair were obtained with the same method (e.g. CERVUS vs. CERVUS) or different  
329 method (CERVUS vs. GENEPOP). These three variables were used as explanatory variables,  
330 whereas the number of shared null alleles in a given pair (indicating similarity level between

331 two estimations compared, which ranged from 0 to 6, mean = 0.77, SD = 0.95) was used as  
332 response variable in GLM with Poisson error distribution and log link in R package (R  
333 Development Core Team 2011).

334 To assess whether a bottleneck influences the number of null alleles detected by each  
335 method tested, we compared the number of detected null alleles between the no-bottleneck  
336 variant with the five bottleneck variants using a  $\chi^2$  test with Bonferroni correction. In order to  
337 test whether the duration of the bottleneck affects the number of null alleles detected, we used  
338 a Kruskal-Wallis test. To assess the level of consistency in null allele detection between the  
339 tests applied to the entire population and to the corresponding randomly selected 20% subset,  
340 we used Kendall's coefficient of concordance  $W_t$  implemented in R (R Development Core  
341 Team 2011). All these calculations were carried out for each method of null allele detection.

342 Finally, using Wilcoxon Signed-Rank Test in R (R Development Core Team 2011), we  
343 assessed whether the observed frequency of null allele homozygotes in populations with  
344 simulated null alleles (see Generation and analysis of simulated datasets), was significantly  
345 different from the frequency expected under HWE.

346

## 347 **Results**

### 348 *Null allele detection in the root vole population*

349 The number of alleles per locus in the root vole population ranged from 4 to 25 (mean  
350 = 14; SD = 5.6). Expected heterozygosity ( $H_E$ ) ranged from 0.107 to 0.925 (mean = 0.780; SD  
351 = 0.218), and observed heterozygosity ( $H_O$ ) from 0.080 to 0.926 (mean = 0.762; SD = 0.215).  
352 The analysed loci had high mean polymorphic information content (mean = 0.760, SD =



353 0.220) and high exclusion probability for the first parent (mean = 0.50, SD = 0.207), which  
354 allowed us to use them successfully in parentage and kinship analyses (see Pilot et al. 2010).  
355 In all but one locus, significant deviations from HWE were detected in different years, with  
356 both heterozygosity deficit (59%) and excess (41%) being observed (Supplementary Table 1).  
357 In 60% of the loci (n = 12) the allele distribution had no missing alleles of any length within  
358 the expected range. In 15% of the loci (n = 3), the allele distribution had one missing allele  
359 length, other 15% (n = 3) had two missing allele lengths, and in the remaining 10% of the loci  
360 (n = 2), more than two allele lengths were missing. At the same time, we did not detect any  
361 null allele homozygotes, which would be indicated as a locus with no detectable product  
362 (blank result).

363         The number of loci in which putative null alleles were detected varied among years  
364 and depended on the estimation method (Table 1). The set of 20 loci was analysed for each  
365 study year separately, which resulted in 8 replicates and a total of 160 loci\*replicates (number  
366 of loci multiplied by number of years). In total, CERVUS detected the lowest number of  
367 putative null alleles at 14 loci\*replicates, while MICRO-CHECKER detected putative null  
368 alleles in 15 loci\*replicates, ML-NullFreq in 36 loci\*replicates and GenePop in 46  
369 loci\*replicates. Altogether, using these four different methods we recorded 67 loci\*replicates  
370 with putative null alleles out of the total number of 160 loci\*replicates. Among them, 68.5%  
371 (n = 46) were detected only by one out of 4 methods, 12% (n = 8) by 2 methods, 4.5% (n = 3)  
372 by 3 methods, and 15% (n = 10) by 4 methods (Table 1). Only two loci had no null alleles  
373 detected in any year. There were no loci where null alleles were detected in each year (Table  
374 1). Moreover, the number of loci with detected putative null alleles estimated for the entire

375 dataset with all years pooled also depended on the estimation method (see Supplementary  
376 Table 2). The frequency of detected putative null alleles (Table 1) was lower than 10% in 73%  
377 of cases (Figure 1). We found no correlation between the number of discontinuities in allele  
378 distribution and the number of detected null alleles within a locus. Finally, the frequency of  
379 null allele detection at loci originally developed for *M. oeconomus* did not differ from the  
380 frequency at loci originally developed for other species.

381

382 *Null allele detection based on parent-offspring genotype comparison in the root vole*  
383 *population*

384 We investigated 270 father-offspring pairs and 241 mother-offspring pairs, resulting in  
385 the detection of 18 parent-offspring pairs carrying putative null alleles. As shown in Table 1,  
386 the parent-offspring method detected null alleles in 11 loci\*replicates. Seven loci were  
387 indicated as having null alleles once (in one out of 8 years), and 2 loci were indicated twice  
388 (in two out of 8 years). Five loci\*replicates with putative null alleles detected using the  
389 parent-offspring analysis were also detected by all four ML programs tested (Table 1). Three  
390 other loci\*replicates indicated by the parent-offspring analysis were not confirmed by any of  
391 these four programs. At the same time, four other loci\*replicates with null alleles detected by  
392 all the other four programs, were not confirmed by the parent-offspring analysis.

393

394 *Null allele detection in the simulated populations*

395 In the 1200 simulated populations the number of alleles per locus ranged from 2 to 34  
396 (mean = 14; SD = 4.2). Expected heterozygosity ( $H_E$ ) ranged from 0.068 to 0.957 (mean =

397 0.861; SD = 0.081), and observed heterozygosity ( $H_O$ ) from 0.07 to 1.0 (mean = 0.861; SD =  
398 0.087). The analysed loci had high mean polymorphic information content (mean = 0.860; SD  
399 = 0.091) and high exclusion probability for the first parent (mean = 0.420, SD = 0.129) (for  
400 detailed information see Supplementary Table 3). In some loci, significant deviations from  
401 HWE were detected (Supplementary Table 4).

402       Among 24,000 loci\*replicates, the number of loci with putative null alleles detected  
403 was highest for ML-NullFreq (n = 1255 loci\*replicates; 5.2% of the total number) and  
404 GENEPOP (n = 1123 loci\*replicates; 4.7%), followed by MICRO-CHECKER (n = 500  
405 loci\*replicates; 2.1%), and it was lowest for CERVUS (n = 327 loci\*replicates; 1.4%)  
406 (Supplementary Table 5). Altogether, using the four different methods we detected 2532  
407 loci\*replicates (10.5% out of 24,000 analysed) with putative null alleles. Among them, 81%  
408 (n = 2056) were detected only by one of the 4 methods, 12% (n = 296) by 2 methods, 6% (n =  
409 163) by 3 methods, and 1% (n = 17) by all 4 methods (Supplementary Table 4).

410       All the null alleles detected in these 1200 simulated populations were false positives, as  
411 the program SPAMs used for their generation does not simulate null alleles. Therefore, we  
412 selected 120 populations where no null alleles were detected, introduced simulated null alleles  
413 by using NullAllelesGenerator, and repeated the analysis with the same four methods. In this  
414 case, MICRO-CHECKER, CERVUS and ML-NullFreq detected either none or very low  
415 frequencies of false positives (0.1% of loci that were actually free of null alleles and 1.6-1.9%  
416 of all loci with null alleles detected). In contrast, GENEPOP detected a considerable number of  
417 false positives: they were found in 9% of loci that were actually free of null alleles and  
418 constituted 55% of null alleles detected by this method (Table 2). Each of the four programs

419 produced a considerable number of false negatives (i.e. true null alleles that remained  
420 undetected) in proportions that ranged between 22% (in ML-NullFreq) and 42% (in CERVUS)  
421 of all loci with known true null alleles.

422

### 423 Statistical analysis

424 The GLMM revealed that the probability of null allele detection in the root vole  
425 population depended on the method applied. It was lowest for the parent-offspring method (as  
426 expected due to smaller sample size used for this analysis – see Materials and Methods) and  
427 highest for the GENEPOP method (Table 3). The number of null alleles detected using the  
428 parent-offspring method was about 16 times lower as compared to GENEPOP and nearly 10  
429 times lower as compared to ML-NullFreq method. Within the remaining methods, the  
430 differences were also significant: the frequency of null allele detection was higher in ML-  
431 NullFreq than in CERVUS ( $P < 0.0001$ ), and higher in GENEPOP than in CERVUS ( $P < 0.0001$ )  
432 but no differences were found between CERVUS and MICRO-CHECKER ( $P = 0.673$ ).

433 The GLMM also revealed that differences between observed and expected  
434 heterozygosity ( $H_O H_E$ ) in the root vole population had a significant influence on the detection  
435 probability of putative null alleles (Table 3). The effects of number of individuals trapped in a  
436 particular year, and the number of alleles in a given locus in a particular year were non-  
437 significant (Table 3).

438 Although the expected cumulative number of loci where putative null alleles were  
439 detected increased asymptotically with increasing sample size, the rate of increase differed  
440 between the five methods applied (Figure 2). Depending on the method, null alleles occurred

441 in 25% to 75% of loci for the all eight years cumulatively.

442        Inspection of the similarity patterns shows that a null allele detected in the root vole  
443 population by a given method is usually a very weak predictor of it being detected by another  
444 method. As a consequence, consistent estimates of null alleles by two methods were rare and  
445 ranged from 12% to 58% of method-pairs (mean = 29.05%, Figure 3). In the simulated  
446 populations, the observed similarity pattern was even lower and ranged from 1.8% to 35.4%  
447 (mean = 13.6%, Figure 3).

448        GLMs showed that the pooled number of null alleles estimated in the root vole  
449 population by two randomly selected methods explained the number of null alleles shared by  
450 these two methods (GLM,  $B = 0.19$ ,  $SE = 0.01$ ,  $z = 18.06$ ,  $P < 0.001$ ). Contrary to  
451 expectations, the number of shared putative null alleles was similar in the “between-methods”  
452 and the “within-method” pairs of estimates ( $B = 0.14$ ,  $SE = 0.10$ ,  $z = 1.36$ ,  $P = 0.174$ ). The  
453 effect of temporal distance between samples from different years was insignificant for the  
454 similarity among estimates ( $B = -0.04$ ,  $SE = 0.02$ ,  $z = 1.85$ ,  $P = 0.064$ ).

455        In the simulated populations ( $n = 100$  individuals each), a change in the population size  
456 (bottleneck effect) significantly affected the number of loci with detected null alleles (Table  
457 4). GENEPOP detected significantly higher number of loci with null alleles in four out of five  
458 comparisons, CERVUS in three, and the remaining two programs in one (Table 4). We also  
459 found that the cumulative number of loci with detected null alleles increased with the  
460 increased bottleneck duration (Figure 4). Bottleneck duration also significantly affected the  
461 number of null alleles detected by GENEPOP (Kruskal-Wallis  $H_c = 11.9$ ;  $P < 0.05$ ), although  
462 no significant correlation was detected in other programs.

463 In the “sub-sampled” simulated populations ( $n = 20$  individuals), we observed very low  
464 concordance of null alleles detected (using each of the four methods) as compared with the  
465 original population ( $n = 100$  individuals) (mean Kendall's coefficient of concordance for four  
466 programs  $W_t = 0.11$ ;  $SD = 0.06$ ) (Table 5).

467 Finally, in populations with simulated null alleles, the observed frequency of null allele  
468 homozygotes differed significantly from the expected frequency ( $V = 10070$ ;  $P < 0.001$ ;  
469 Supplementary Figure 1). We observed deviations toward both null allele homozygote  
470 deficiency and excess (Supplementary Figure 1). For null allele frequencies below 0.17 we  
471 observed cases where no null allele homozygotes occurred, but there were no such cases for  
472 null allele frequencies higher than 0.17.

473

#### 474 **Discussion**

475 Detection of null alleles using indirect methods is susceptible to errors, given that  
476 these methods are based on assumptions that are commonly violated in natural populations.  
477 Methods based on comparing observed and expected heterozygosity (Dempster et al. 1977,  
478 Chakraborty et al. 1992, Brookfield 1996, Summers & Amos 1997, Kalinowski & Taper  
479 2006) assume that null alleles can be detected based on observed deviations from HWE  
480 towards heterozygote deficit. However, natural populations may deviate from HWE because  
481 they do not meet the assumptions of the Hardy-Weinberg law, and/or because they are often  
482 studied based on a small number of samples, which may lead to random deviations from the  
483 equilibrium at different loci. The parentage method does not assume HWE, but may be prone  
484 to other types of errors, e.g. human errors with microsatellite scoring. The error rate in null

485 allele detection in natural populations is difficult to estimate, because the actual null allele  
486 frequencies are usually unknown. Our study was based on a natural population that was  
487 sampled for several consecutive years, and so the same null alleles were expected to occur  
488 throughout the entire study period. This allowed us to test the reliability of several methods of  
489 null allele estimation through the comparison of results between different years. Application  
490 of the same methods to 1200 simulated populations that underwent bottlenecks of different  
491 intensity and duration allowed us to further examine the effect of strong genetic drift on null  
492 allele detection.

493

#### 494 *Accuracy of null allele detection in the root vole population*

495       We found inconsistencies in null allele estimation both across years for each method  
496 and among different methods within each year. We also failed to find any statistically  
497 significant temporal repeatability in null allele detection at any locus. Each method detected a  
498 considerable number of null alleles in the 1200 simulated datasets, and there were significant  
499 differences in null allele estimates among the methods. However, the number of null alleles  
500 detected was positively correlated with the bottleneck size in each of the methods tested.  
501 Crucially, the simulated populations did not originally include any null alleles (the program  
502 SPAMs used for their generation does not simulate null alleles), so all the detected null alleles  
503 were false positives.

504       We thus conclude that all putative null alleles detected in the root vole population are  
505 likely to be false positives. Our conclusion is supported by the following evidence: First,  
506 given that real null alleles are derived from primer compatibility problems during PCR

507 amplification and PCR protocols did not change between years, we would expect the same  
508 null alleles to be present in each year of the study, or at least in most years (accounting for the  
509 sampling effect – see below). Yet, none of the methods tested detected such continuous  
510 presence for any of the loci. One explanation for this could be that individuals having null  
511 alleles in their genotypes were not sampled every year. In that case the number of individuals  
512 sampled in a particular year should have a significant influence on null allele detection  
513 probability. The GLMM analysis did not find such a correlation, which allows us to reject this  
514 explanation. Additionally, according to MARK estimate, over 90% of all individuals present  
515 in the study population were genotyped, and therefore the probability of omitting all  
516 individuals with a given null allele is negligible, unless this allele has a very low frequency in  
517 the population. However, the impact of null alleles with such low frequencies on results of  
518 population-level genetic analyses would be negligible.

519         Second, lack of missing data in the root vole genotype dataset is also consistent with  
520 low frequency or lack of null alleles in this population. Although the relationship between null  
521 allele frequency and the frequency of null allele homozygotes based on the Hardy-Weinberg  
522 law does not necessarily hold in non-equilibrium populations, these two parameters are  
523 always dependent as demonstrated for the simulated populations. Therefore, it is expected that  
524 in a locus with high null allele frequency, some blank results should occur. There are some  
525 cases in the root vole population where the estimated null allele frequency was above 20%  
526 (Figure 1). Under the Hardy-Weinberg law, such loci should contain over 4% of null allele  
527 homozygotes. However, we detected none, despite a large number of genotyped individuals.

528         Third, kin clustering and non-random mating have been earlier demonstrated in this



529 population (Dąbrowski 2010). We also found that significant multi-annual changes in density  
530 and random environmental events (e.g. seasonal floods) have a strong impact on rates of  
531 seasonal migration, male dispersal, and female philopatry (Pilot et al. 2010, Dąbrowski 2010).  
532 Heterozygosity could thus have been lost (and regained) from year to year due to both genetic  
533 drift and migration (Pilot et al. 2010, Dąbrowski 2010). We also found genetic signatures of  
534 bottleneck in this population (Pilot et al. 2010), and we show in this study, based on simulated  
535 data, that bottlenecks may significantly increase the frequency of false null alleles detected by  
536 each of the methods tested. Therefore, we conclude that the pattern of the putative null allele  
537 occurrence observed in the study population, is more likely to result from population genetic  
538 processes like density fluctuations, migration and non-random mating, than from factors  
539 associated with PCR amplification outcomes.

540

#### 541 *Inconsistencies among different methods of null allele detection*

542 Our study revealed large inconsistencies among the compared methods of null allele  
543 detection in both the natural root vole population and the simulated datasets. The average  
544 similarity among the methods used to detect null alleles was 29.05% for the root vole  
545 population and only 13.6% for the simulated populations. While detection of false null alleles  
546 may be explained by population genetic processes leading to deviations from HWE,  
547 inconsistencies among the methods cannot be accounted for solely by this explanation. The  
548 method based on parentage analysis relied on different assumptions and smaller pool of  
549 individuals than the heterozygosity-based methods, and the resulting differences were  
550 consistent with expectations. However, the four heterozygosity-based methods applied the

551 same general assumptions (see Supplementary Material). They differed in the way missing  
552 data was interpreted, but neither the root vole population nor the original SPAMs-generated  
553 populations included any missing data. Therefore, we conclude that the discrepancies among  
554 these methods do not result from differences in the theoretical assumptions, but rather from  
555 differences in the particular optimisation algorithms applied.

556

557 *How to combine different methods to minimise errors in null allele detection?*

558         Our study raises a question regarding whether estimates of null alleles reported in the  
559 literature, which are usually inferred using indirect methods, are always reliable. The number  
560 of null allele occurrences within different allele frequency classes calculated in this study for  
561 the root vole population (Figure 1) has a similar distribution to the one shown in Dakin &  
562 Avise (2004), based on an extensive literature review. Given that our results show that most  
563 null alleles detected in the root vole population are likely false positives (see above), this  
564 similarity raises a further question of whether the recommendation for discarding loci  
565 showing null alleles from analysed datasets (De Sousa et al. 2005) should be followed  
566 unconditionally.

567         In the case of the root vole population, several loci with putative null alleles would  
568 have to be excluded following this recommendation, with different number of loci excluded  
569 depending on the year and the detection method used. Moreover, if sampling was carried out  
570 for a longer period, we may expect that the number of loci with putative null alleles would  
571 increase with the number of study years (see Figure 2), because we observed no consistent  
572 detection pattern among years for any locus.

573           It may be thus useful to devise strategies that combine different methods to minimise  
574 errors in null allele detection based on the results from our study. We found that in the  
575 simulated populations without null alleles, 81% of false positives were detected by only one  
576 out of the four heterozygosity-based methods, while only 1% of false positives were detected  
577 by all the four methods. At the same time, in the simulated populations where null alleles  
578 were included, 58% of true null alleles were detected by all the four methods. Therefore,  
579 combining two or more methods and considering only the consistent putative null alleles  
580 should considerably reduce the detection of false positives. However, it may also result in  
581 non-detection of some true null alleles, especially if more than two methods are applied.

582           Therefore, it may be useful to assess which of the four methods tested are less error-  
583 prone. In the simulated populations without null alleles, CERVUS and MICRO-CHECKER  
584 detected less false positives (1.4% and 2.1%, respectively) as compared with the two other  
585 methods. On the other hand, in the simulated populations with null alleles, ML-NullFreq had  
586 the lowest proportion of false negatives (22%), while for CERVUS and MICRO-CHECKER  
587 this proportion was 42% and 33%, respectively. GENEPOP was the only method that still  
588 detected a considerable number of false positives in the simulated datasets (prior to the  
589 simulation of true null alleles) that were pre-selected specifically as having no false positives  
590 detected by any of the four programs. Therefore, this program seems to be particularly error-  
591 prone in terms of the detection of false null alleles. We thus suggest that the best strategy to  
592 minimise the errors in null allele detection would be the combined use of two or three of the  
593 remaining methods (ML-NullFreq, CERVUS and MICRO-CHECKER). The combination of  
594 CERVUS and MICRO-CHECKER is best for minimising the false positives' rate, while the

595 combination of ML-NullFreq and MICRO-CHECKER is best for minimising the false  
596 negatives' rate.

597         However, before applying these methods, it is important to account for the occurrence  
598 of other types of genotyping errors like allelic dropouts or false alleles (e.g. resulting from  
599 stuttering), which can be detected e.g. using MICRO-CHECKER and/or by replicating the  
600 genotyping for a number of individuals. It is also important to minimise the occurrence of  
601 missing data due to reasons other than null allele homozygotes by repeating failed PCRs at  
602 least once.

603         Because heterozygosity-based methods assume HWE, it is important to minimise  
604 errors that may result from violations of the assumptions of Hardy-Weinberg law. For  
605 example, if population genetic structure is detected, the presence of null alleles should be  
606 assessed for each sub-population separately. The parentage-based method does not assume  
607 HWE, so it may help minimising the detection of false null alleles if used in addition to the  
608 heterozygosity-based methods; however, we recognise it won't always be possible or practical  
609 to use this method, due to its reliance on a detailed reconstruction of parent-offspring  
610 relationships within the study population. Finally, our study showed that material collected  
611 from the same population during several seasons (if there is sufficient generational turnover)  
612 may help interpreting the results of null allele detection and prevent their overestimation of  
613 their numbers. Alternatively, if sample size is sufficiently large, the accuracy of null allele  
614 detection may be improved by comparing the results obtained from different random sub-sets  
615 of the entire dataset analysed.

616

## 617 *Conclusions*

618 Our study shows that many commonly used null allele detection methods exhibit low  
619 reliability and consistency when applied to non-equilibrium populations. When we account  
620 for both false null allele detection rate and non-detection rate of the true null alleles, no  
621 method can be considered as clearly superior over the others. We thus suggest the combined  
622 use of at least two methods and considering only putative null alleles detected consistently by  
623 different methods. This should considerably reduce the detection of false positives. However,  
624 this approach is compromised by an increased rate of false negatives (non-detected real null  
625 alleles), and thus provides only a sub-optimal solution. Our study demonstrates the need to  
626 develop null allele detection methods that could be applied to non-equilibrium populations  
627 without violating the model assumptions.

628

## 629 **Acknowledgements**

630 We are grateful to Dr M. Ratkiewicz for the inspiration for this study. We thank Prof. Oscar  
631 Gaggiotti, Dr Andre Moura and three anonymous reviewers for their constructive comments  
632 on the earlier version of this manuscript. We also would like to thank Dr E. Jancewicz and the  
633 students from the Faculty of Forestry, Warsaw University of Life Sciences – SGGW for  
634 assistance in the sample collection. The study was supported by a grant NN304 346339  
635 (awarded to J. Gliwicz) from Polish National Science Centre. M. Pilot was supported by a  
636 fellowship from the Foundation for Polish Science.

637 **References**

638 Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer  
639 Associates Inc., Sunderland.

640 Bates D, Maechler M, Bolker B (2011) *lme4: Linear mixed-effects models using S4 classes*. R  
641 package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>

642 Brookfield JFY (1996) A simple new method for estimating null allele frequency from  
643 heterozygote deficiency. *Molecular Ecology*, **5**, 453-455.

644 Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR  
645 (1993) Incidence and origin of null alleles in the (AC)<sub>n</sub> microsatellite markers. *American*  
646 *Journal of Human Genetics*, **52**, 922-927.

647 Carlsson J (2008) Effect of microsatellite null alleles on assignment testing. *Journal of*  
648 *Heredity*, **99**, 616-623.

649 Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote  
650 deficiencies observed in DNA typing data and their implications in forensic applications.  
651 *Annals of Human Genetics*, **56**, 45-57.

652 Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population  
653 differentiation. *Molecular Biology and Evolution*, **24**, 621-631.

654 Colwell RK (2005) EstimateS: Statistical estimation of species richness and shared species  
655 from samples. Version 7.5. User's Guide and application published at:  
656 <http://purl.oclc.org/estimates>.

657 Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent  
658 population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.

659 Dąbrowski MJ (2010) Relatedness, space use strategies and density changes in the population  
660 of the root vole *Microtus oeconomus*. PhD thesis, Museum and Institute of Zoology PAS,  
661 Warsaw.

662 Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504-  
663 509.

664 De Sousa SN, Finkeldey R, Gailing O (2005) Experimental verification of microsatellite null  
665 alleles in Norway spruce (*Picea abies* [L.] Karst.): Implications for population genetics  
666 studies. *Plant Molecular Biology Reporter*, **23**, 113-119.

667 Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the  
668 EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1-38.

669 Garcia de Leon FJ, Canonne M, Quillet E, Bonhomme F, Chatain B (1998) The application of  
670 microsatellite markers to breeding programmes in the sea bass, *Dicentrarchus labrax*.  
671 *Aquaculture*, **159**, 303-316.

672 Gliwicz J, Dąbrowski MJ (2008) Ecological factors affecting the diel activity of voles in a  
673 multi-species community. *Annales Zoologici Fennici*, **45**, 242-247.

674 Gliwicz J, Jancewicz E (2004) Voles in river valleys. In: Jędrzejewska B and Wójcik JM  
675 (eds.) Essays on Mammals of Białowieża Forest, 139-148. MRI PAS, Białowieża.

676 Guichoux E, Lagache L, Wagner S et al. (2011) Current trends in microsatellite genotyping.  
677 *Molecular Ecology Resources*, **11**, 591-611.

678 Ishibashi Y, Saitoh T, Abe S, Yoshida MC (1996) Null microsatellite alleles due to nucleotide  
679 sequence variation in the grey-sided vole *Clethrionomys rufocanus*. *Molecular Ecology*, **5**,  
680 589-590.

681 Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends*  
682 *in Ecology & Evolution*, **11**, 424-429.

683 Kalinowski ST, Taper ML (2006) Maximum likelihood estimation of the frequency of null  
684 alleles at microsatellite loci. *Conservation Genetics*, **7**, 991-995.

685 Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-  
686 based paternity inference in natural populations. *Molecular Ecology*, **7**, 639-655.

687 McCoy EE, Jones AG, Avise JC (2001) The genetic mating system and tests for cuckoldry in a  
688 pipefish species in which males fertilize eggs and brood offspring externally. *Molecular*  
689 *Ecology*, **10**, 1793-1800.

690 Oddou-Muratorio S, Vendramin GG, Buiteveld J, Fady B (2009) Population estimator of  
691 progeny test: what is the best method to assess null allele frequencies at SSR loci?  
692 *Conservation Genetics*, **10**, 1343-1347.

693 van Oosterhout C, Hutchinson WF, Wills DMP, Shipley P (2004) MICRO-CHECKER:  
694 software for identifying and correcting genotyping errors in microsatellite data. *Molecular*  
695 *Ecology Notes*, **4**, 535-538.

696 Paetkau D, Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite  
697 null allele in bears. *Molecular Ecology*, **4**, 519-520.

698 Parreira B, Trussart M, Sousa V, Hudson R and Chikhi L (2009) SPAMs: A user-friendly  
699 software to simulate population genetics data under complex demographic models. *Molecular*  
700 *Ecology Resources*, **9**, 749-753.

701 Pemberton JMS, Slate J, Bancroft DR, Barrett JA (1995) Non-amplifying alleles at  
702 microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, **4**, 249-



703 252.

704 Pilot M, Dąbrowski MJ, Jancewicz E, Schtickzelle N and Gliwicz J (2010) Temporally stable  
705 genetic variability and dynamic kinship structure in a fluctuating population of the root vole  
706 *Microtus oeconomus*. *Molecular Ecology*, **19**, 2800-2812.

707 Primmer CR, Moller AP, Ellegren H (1995) Resolving genetic relationship with microsatellite  
708 markers: a parentage testing system for the swallow *Hirundo rustica*. *Molecular Ecology*, **4**,  
709 493–498.

710 R Development Core Team (2011) *R: A language and environment for statistical computing*. R  
711 Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL  
712 <http://www.R-project.org/>.

713 Rousset F (2008) Genepop'007: a complete reimplementaion of the Genepop software for  
714 Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

715 Summers K, Amos W (1997) Behavioral, ecological, and molecular genetic analyses of  
716 reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus*.  
717 *Behavioral Ecology*, **8**, 260-267.

718 Van Treuren R (1998) Estimating null allele frequencies at a microsatellite locus in the  
719 oystercatcher (*Haematopus ostralegus*). *Molecular Ecology*, **7**, 1413-1417.

720 Walker D, Porter BA, Avise JC (2002) Genetic parentage assessment in the crayfish  
721 *Orconectes splacidus*, a high-fecundity invertebrate with extended maternal brood care.  
722 *Molecular Ecology*, **11**, 2115–2122.

723 Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a  
724 source of heterozygote deficiency at microsatellite loci: experimental evidence at the

725 dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, **7**, 1569-  
726 1573.

727

728

---

729 MJD and MP designed the project. MP and JG supervised the project. MJD and MP wrote the  
730 paper. MJD performed laboratory work, generated simulated populations and carried out null  
731 allele detection analysis. MK designed and implemented the algorithm for simulating null  
732 alleles, contributed to the statistical analyses and editing of the manuscript. MZ designed and  
733 conducted most of the statistical analyses and contributed to writing the manuscript. HMU  
734 participated in programming and automating of the process of comparing null allele detection  
735 results. This project used data collected as a result of a long-term research project on rodents  
736 in Białowieża National Park supervised by JG.

---

737

738

### 739 **Data Accessibility**

740 Data underlying this manuscript are available in the online supplemental information.

741 *Microtus oeconomus* genotypes, simulated populations genotypes and subpopulations  
742 genotypes are available on Dryad, doi: 10.5061/dryad.4p41m

743 NullAlleleGenerator, along with its full documentation and example data is available from

744 <http://www.lcb.uu.se/papers/dabrowski/NullAlleleGenerator.zip>

745

746 **Figure legends**

747 **Figure 1.** Histogram of frequencies of putative null alleles detected using different algorithms  
748 implemented in the evaluated programs: CERVUS (algorithm of Summers & Amos); MICRO-  
749 CHECKER (algorithms: Oosterhout, Chakraborty, Brookfield 1, Brookfield 2), and GENEPOP  
750 (EM algorithm of Dempster 1977).

751

752 **Figure 2.** Expected cumulative number (left axis) and percentage (right axis) of loci where  
753 putative null alleles were detected using five different methods, are presented as a function of  
754 the number of years studied. C – CERVUS; M – MICRO-CHECKER; N – ML-NullFreq; G –  
755 GENEPOP.

756

757 **Figure 3.** Similarity of null allele estimates between the methods applied for the natural root  
758 vole population and the simulated populations. The plot gives an average expectation that a  
759 null allele detected by one method will also be detected by the other method (gray – the  
760 simulated data; white – the root vole data).

761

762 **Figure 4.** Cumulative number of loci with putative null alleles within the simulated  
763 populations (n = 100 individuals each) with different bottleneck scenarios. “Time” denotes  
764 the bottleneck duration in generations. Cumulative number of loci is the sum of loci where  
765 null alleles were detected using any of the four methods.

766

767 **Table 1.** The presence of putative null alleles in the root vole population in each locus per  
 768 year, estimated using five different methods.

Study year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Sample size	150	116	130	39	70	147	84	0	16
Locus									
AV12			N	N	G	P			G
AV13	CMNG		CMNGP	CMNG	NG				
AV14				NG		N	G		
AV15	N								
Moe1		G	M	G		G	N		G
Moe2				G		G			
Moe3	G	NG	G			MNG	G		
Moe4	CMNGP	NG		CMN	N		CMNGP		N
Moe5		P			N				
Moe6	MN	G	N	G	P		N		CNG
Moe7		CMNGP	MG	G		CMNG	CMNGP		G
MSCRB4			C	C					
MSCRB6		G		G					
MSMM2									
MSMM3	N		P		N	N			
MSMM4	N			N		NG	G		
MSMM5	NG	G	G	G	G	G			
MSMM6		CMNGP		G		CMNG			G
MSMM7	NP			G					
MSMM8									

769 The presence of a putative null allele is marked by the symbol of the program (or multiple programs) that  
 770 detected it: C – CERVUS; M – MICRO-CHECKER; N – ML-NullFreq; G – GENEPOP, and P – comparison of  
 771 mismatching loci in parent-offspring genotypes. In the second row, the number of individuals sampled in each  
 772 year is shown (Sample size).

773

774 **Table 2.** Null alleles detected using MICRO-CHECKER, CERVUS, ML-NullFreq and  
 775 GENEPOP for 120 simulated populations containing two null alleles each. Loci with known  
 776 null alleles were compared with loci detected using different programs (0- loci without null  
 777 alleles; 1- loci with null alleles). Black background: true positives, grey background: true  
 778 negatives, white background: false negatives, underline value: false positives.

779

	Null Alleles Generator	
	0	1
MICRO-CHECKER	0 <u>2164</u>	78
	1 <u>2</u>	<b>156</b>
CERVUS	0 <u>2166</u>	99
	1 <u>0</u>	<b>135</b>
ML-NullFreq	0 <u>2163</u>	52
	1 <u>3</u>	<b>182</b>
GENEPOP	0 <u>1967</u>	72
	1 <u>199</u>	<b>162</b>

780 **Table 3.** Summary results of a generalized linear mixed model with binomial error  
781 distribution and logit link, explaining the presence of the null alleles as a function of the four  
782 predictors: (1) difference between observed and expected heterozygosity ( $H_OH_E$ ), (2) number  
783 of individuals trapped in a given year, (3) number of alleles in a particular locus in a given  
784 year and (4-8) the method of null allele detection. Symbols of different methods are explained  
785 in Table 1. Year and locus were included as random categorical factors in the model. For  
786 every level of each predictor the following parameters are given: estimate (B), with standard  
787 errors (SE), exponentiated estimate (Exp(B)), tests statistic (z-value), and significance ( $P$  -  
788 value).

Effect	B	SE	Exp(B)	z-value	$P$ -value
Intercept	-3.507	0.593	0.030	-5.918	0.000
(1) $H_OH_E$	-28.760	3.070	0.000	-9.368	< 0.001
(2) N of individuals	0.000	0.003	1.000	-0.013	0.990
(3) N of alleles	-0.037	0.031	0.964	-1.175	0.240
(4) Method = G	0.579	0.554	1.785	1.045	0.296
(5) Method = M	2.782	0.495	16.154	5.623	< 0.001
(6) Method = N	0.811	0.541	2.249	1.498	0.134
(7) Method = P	2.283	0.498	9.806	4.586	< 0.001
(8) Method = C	0.000		1.000		

789 **Table 4.** Pairwise comparisons of the number of loci with putative null alleles between  
 790 variant 1 of the simulation (no bottleneck) with the other five variants with different level of  
 791 bottleneck. Comparisons were made for each program separately. Values presented in the  
 792 table are results of  $\chi^2$  test. Statistically significant results are marked with (\*).

Bottleneck pair variants	CERVUS	GENEPOP	MICRO-CHECKER	ML-NullFreq
1-2	0.272	2.010	0.286	< 0.001
1-3	2.502	70.040**	3.704	4.560
1-4	9.318*	148.700**	0.008	0.011
1-5	42.760**	206.500**	1.347	1.612
1-6	114.700**	176.500**	24.940**	16.870*

793 (\*)  $P < 0.01$ ; (\*\*)  $P < 0.001$ ; Bottleneck pair variants: 1 - 99,999; 2 - 50,000; 3 - 25,000; 4 - 10,000; 5 - 5,000; 6  
 794 - 2,500. The ancestral number of individuals for all bottleneck variants was set to 100,000 individuals.

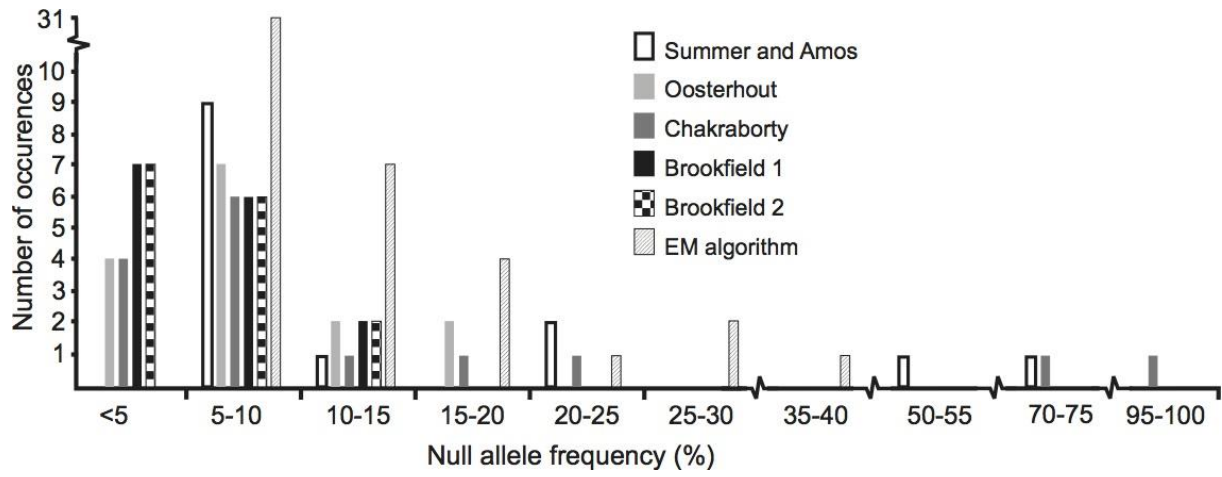
795 **Table 5.** Pairwise comparison of loci with detected null alleles in two data sets: simulated  
 796 original populations (n = 100 individuals per each population) and sub-sampled populations  
 797 (n = 20 individuals randomly selected from original population).

Variable	CERVUS	GENEPOP	MICRO-CHECKER	ML-NullFreq
Wt	0.182	0.040	0.093	0.123
N	229	223	59	214
CP	13	9	6	30
new-P	204	129	24	96
not-CP	12	85	29	88

798 Wt –Kendall's coefficient of concordance corrected for ties; N – number of loci with null alleles detected in at  
 799 least one dataset (either original or under-sampled); CP – conserved positives: loci with detected null alleles in  
 800 both sets of populations; new-P – new positives: loci with null alleles detected only in under-sampled data set;  
 801 not-CP – not conserved positives: loci with null alleles detected in the original data set which were not  
 802 confirmed within under-sampled data set.

803

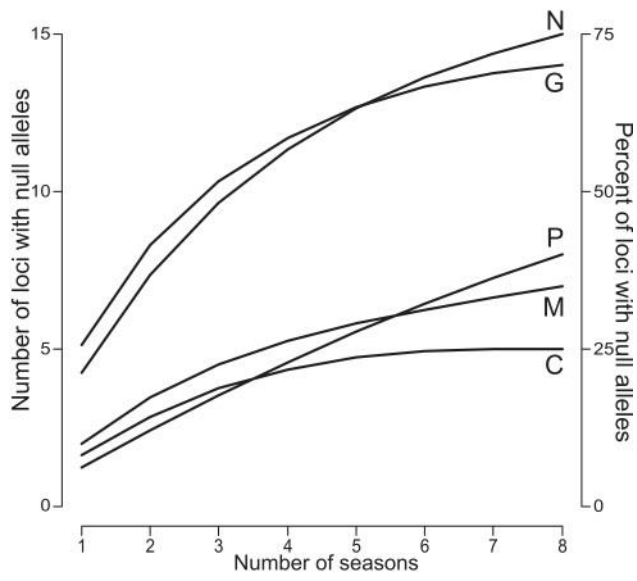




804

805 **Figure 1**

806



807

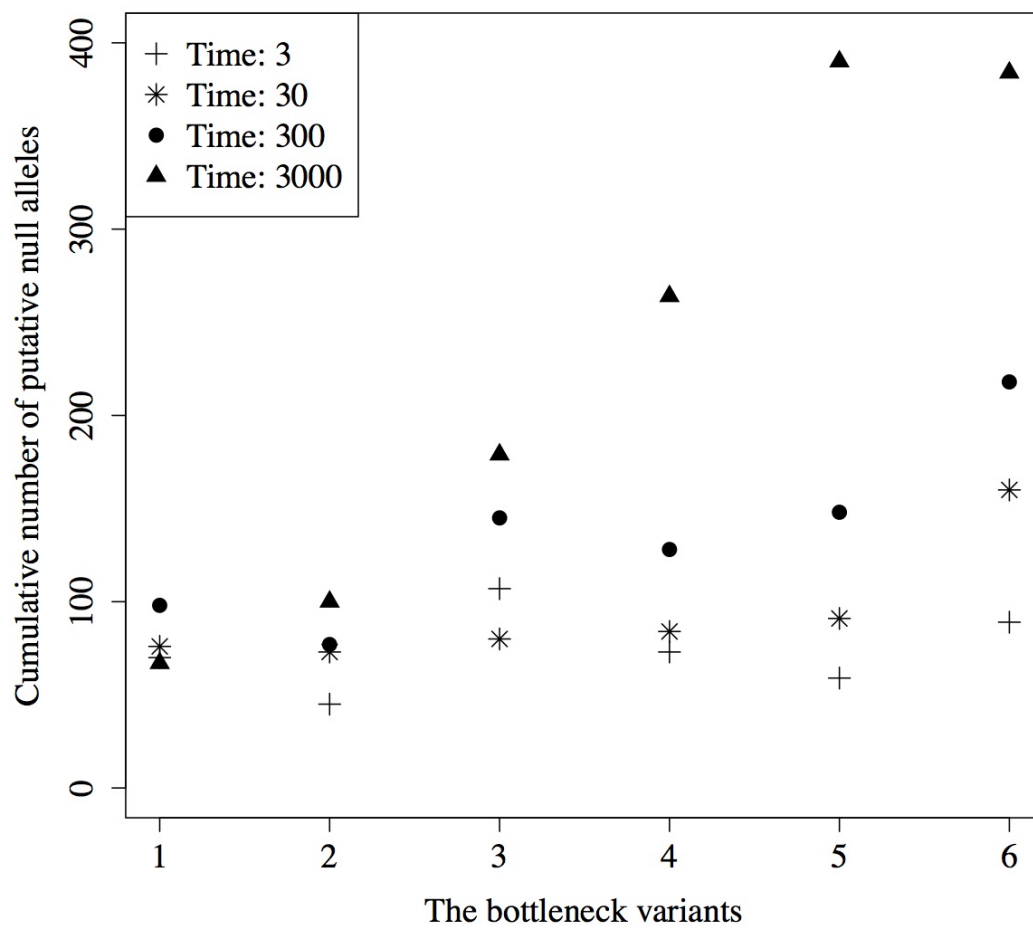
808 **Figure 2**

809

CERVUS	22.4%	57.9%	31.6%	33.3%
2.3%	GENEPOP	24.0%	28.1%	12.0%
35.4%	1.8%	MICRO-CHECKER	33.3%	30.0%
13.6%	2.9%	25.3%	ML-NullFreq	17.9%
-	-	-	-	Parent-offspring

810

811 **Figure 3**



812

813 **Figure 4**

814 **Supporting Information**

815 Additional Supporting Information may be found in the online version of this article:

816

817 **Table S1** Characteristics of 20 microsatellite loci analysed in the root vole population,  
818 organized in four multiplex PCR reactions.

819

820 **Table S2** Loci with putative null alleles confirmed for the entire dataset (root vole population,  
821 all years pooled) using all four methods.

822

823 **Table S3** Characteristics of the simulated populations. Range, mean values and SDs were  
824 computed separately for 50 replicates representing each scenario of the bottleneck size and  
825 duration (in generations).

826

827 **Table S5** Number of loci with null alleles detected using four programs in populations  
828 simulated with various levels and duration of a bottleneck.

829

830 **Fig. S1** The observed and expected frequency of homozygotes vs. null allele frequency in the  
831 loci where null alleles were simulated using NullAlleleGenerator in 120 randomly selected  
832 populations.

833

834 **Table S4** Genetic estimates and null allele detection results computed for simulated  
835 populations affected by bottleneck scenarios.