

### **S1: Type I and type II errors in determining paternity with molecular markers.**

In many studies, the determination of paternity on the basis of molecular markers works is based on the following general procedure. After having obtained the allelic values (at various loci) for an offspring and a putative father, the degree of allele sharing is translated into a score that is related to the likelihood to obtain the offspring genotype given that the putative father is the genetic father. This score, to be defined below, will be called the ‘L-score’ in the following considerations. To determine paternity, the L-score is compared with a threshold value (to be specified below): if  $L \geq T$ , the putative father is considered to be the genetic father, if  $L < T$ , the putative father is discarded from being the genetic father. Let us for simplicity assume that the focus is at the social mate of the known mother. Then the offspring is considered to be a ‘within-pair young’ (WPY) if  $L \geq T$  and an ‘extra-pair young’ (EPY) if  $L < T$ .

Two types of error can be made in case of a classification as the one described above: a within-pair young can falsely be considered to be an extra-pair young (type I error), or an extra-pair young can be falsely considered to be a within-pair young (type II error). As shown in Figure A1, the magnitude of the errors can be kept in check by a proper choice of the threshold value  $T$ . The probability distribution A corresponds to the expected distribution of L-scores given that the social father is the true genetic father. Accordingly, the area  $\alpha_A$  under this distribution to the *left* of  $T$  corresponds to the probability to misclassify a WPY as being an EPY. In other words,  $\alpha_A$  is the probability of making a type I error. The probability distribution B corresponds to the expected distribution of L-scores of males that are not the genetic father. Now the area  $\alpha_B$  under this distribution to the *right* of  $T$  corresponds to the probability of making a type II error.

The magnitude of  $\alpha_A$  and  $\alpha_B$  reflects the degree of overlap of distribution A and B, and hence the discriminatory power of the set of molecular markers used. In case of highly variable markers, the distributions will be more separated allowing to achieve low values of both types of errors. Given the distributions A and B, the choice of  $T$  determines the relative magnitude of  $\alpha_A$  and  $\alpha_B$ . This means that the choice of  $T$  should reflect the relative importance one wants to give to either error, which may strongly depend on the underlying research question. In our case, we consider both types of error equally important and therefore choose  $T$  such that  $\alpha_A = \alpha_B$ .

In practise, we used the program Cervus 3.0 (Kalinowski et al., 2007) to determine our ‘L-scores’ and the distributions A and B. For any given year, we entered the allele frequencies at the marker loci of all adult individuals found in the study population, allowing Cervus to take into account differences in discriminatory power between the markers. The L-scores calculated by Cervus are log-likelihood ratios, also called LOD-scores. Given the allelic pattern of the known mother, Cervus calculates two likelihoods (for details see (Kalinowski et al., 2007)): (1) the likelihood that the offspring genotype is obtained given that the social male is the true father; and (2) the likelihood that the offspring pattern is obtained from a random male in the population. The LOD-score then corresponds to the natural logarithm of the ratio of these likelihoods. The distributions A and B are obtained by a simulation approach. First, the parental population is generated by producing a large number of adults, whose genotype frequencies reflect the known allele frequencies at the marker loci found in our study population. Then, a large number ( $n= 100.000$ ) of offspring is produced by randomly pairing a male and a female. The offspring genotypes are derived from the parental genotypes in a Mendelian way. To mimic the real data as closely as possible, the allelic values of the offspring were altered with a small probability (chosen to be 0.01) corresponding to typing errors and other mistakes. In this way, we obtained the LOD-scores of the ‘genetic father’-offspring pairs, which correspond to distribution A in Figure A1. Similarly, LOD-scores of ‘random male’-offspring pairs were generated, corresponding to distribution B.

We applied this procedure separately for the three study years, 2002, 2003 and 2004. For the year 2004, the results are represented in Figure A2. In this case, the threshold value separating WPY and EPY is chosen such that the probabilities of making a type I and a type II error turns out to be  $\alpha_A = \alpha_B = 0.024$ . In other words, the chance to falsely assign a WPY to be an EPY is 2.4%, and this is equal to falsely assign an EPY to be a WPY. The distribution of LOD-scores was very similar in the other two years. Type I and type II errors were limited to  $\alpha= 0.022$  (in 2002) and  $\alpha= 0.024$  (in 2003), respectively.

## References

- Kalinowski ST, Taper ML, Marshall TC, 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16:1099-1106.

Figure A1. Probability of errors of type I ( $\alpha_A$ ) and type II ( $\alpha_B$ ) in paternity analysis. Curve A is the probability distribution of L-scores given that the social mate of the known mother is the genetic father, while curve B represents the distribution where the social mate is not the genetic father. If paternity is determined on the basis of a threshold value T, two errors can occur: with probability  $\alpha_A$ , a within-pair young is considered extra-pair (type I error); with probability  $\alpha_B$ , an extra-pair young is considered within-pair (type II error).

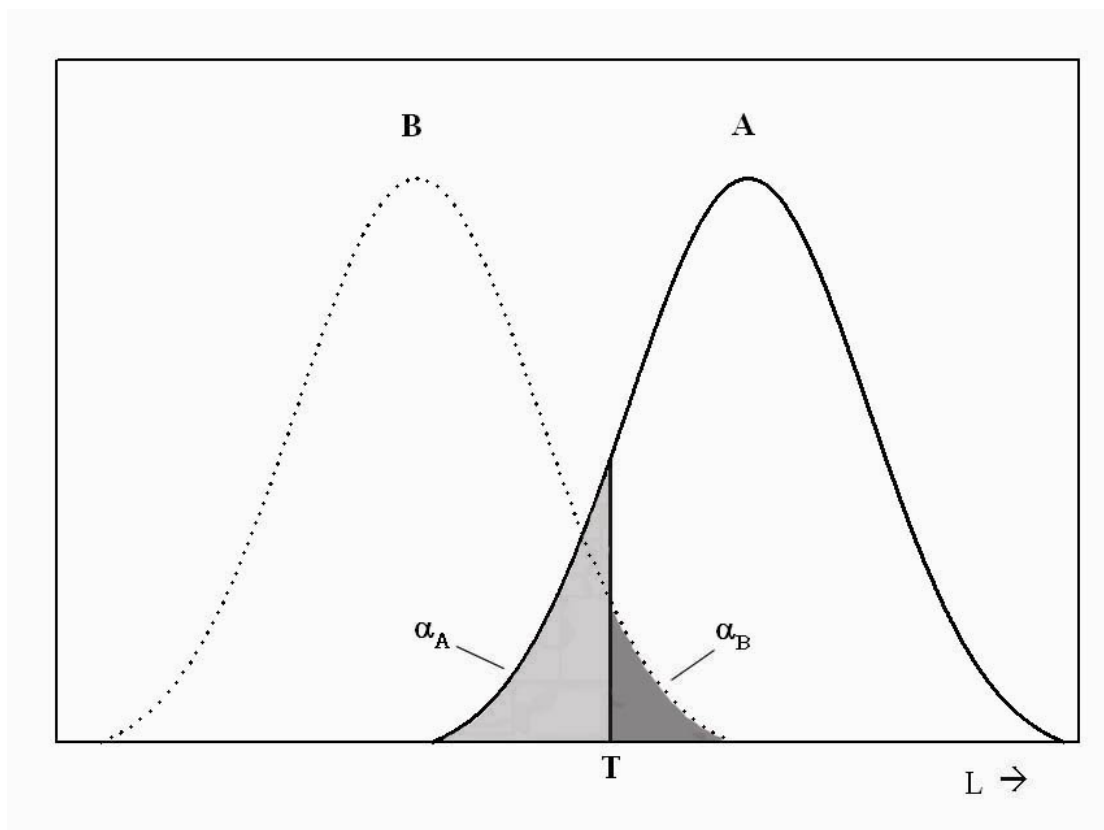


Figure A2. The distribution of LOD-scores generated by Cervus for ‘genetic father’-offspring pairs (curve A) and ‘random male’-offspring pairs (curve B). These curves correspond to the probability distributions A and B in Figure A1. The threshold value (here  $T = 0.176$ ) was chosen in such a way that the probability of type I and type II errors was equal (here  $\alpha_A = \alpha_B = 0.024$ ).

NB: The seven peaks in distribution B reflect the fact that we had markers at seven loci. The peaks correspond to mismatches between random male and offspring at  $k = 1, \dots, 7$  loci.

