ELECTRONIC APPENDIX A

Pair-wise relatedness between potential troop sires was determined as the maximum likelihood estimate of the kinship coefficient, using genetic markers. Likelihood maximisation was performed using comprehensive allele frequency data from the population involved. For a pair of animals genotyped at n loci, let a_{ijk} be the kth allele (k=1..2) at the jth locus (j=1..n) in the ith animal (i=1..2); let $f(a_{ijk})$ denote the population frequency of that allele. If h is the kinship coefficient of the pair, then the posterior probability δ(h,a,a*) of an allele a, drawn from the first individual, and allele a*, randomly drawn from the second individual, to be identical by descent is

$$\delta(h,a,a*) = 1_{a,a*} \cdot \frac{h}{(1-h) \cdot f(a) + h} \; .$$

Here $1_{a,a*}$ equals unity if a and a* are identical by state, and equals zero otherwise. If h is unknown, it can be estimated from the genotype data using an EM algorithm (Dempster et al. 1977). Let h_t denote the estimate resulting from iteration t. Then, by evaluating δ for all possible pairs of alleles at all loci, a revised estimate of h is obtained in iteration t+1 as

$$h_{t+1} = \sum_{j=1}^{n} \sum_{k_1=1}^{2} \sum_{k_2=1}^{2} \frac{\delta(h_t, a_{1jk_1}, a_{2jk_2})}{4n}$$

This process generally converges to a local maximum of the likelihood curve; convergence to the global maximum was ensured by 1000-fold replication, using randomly selected start values of h.