Relationship type affects the reliability of dispersal distance estimated using pedigree inferences in partially sampled populations: a case study involving invasive American mink in Scotland

Melero, Y. ORCID: https://orcid.org/0000-0002-4337-1448, Oliver, M. K. and Lambin, X. (2017) Relationship type affects the reliability of dispersal distance estimated using pedigree inferences in partially sampled populations: a case study involving invasive American mink in Scotland. Molecular Ecology, 26 (15). pp. 4059-4071. ISSN 0962-1083 doi: 10.1111/mec.14154 Available at https://centaur.reading.ac.uk/83097/

To link to this article DOI: http://dx.doi.org/10.1111/mec.14154

Publisher: Wiley

Relationship type affects the reliability of dispersal distance estimated using pedigree inferences in partially sampled populations: a case study involving invasive American mink in Scotland

Y. Melero[1,2], M.K. Oliver[1], X. Lambin[1]

[1]School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK

[2]CREAF (Centre de Recerca Ecològica i Aplicacions Forestals), Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Corresponding author: Yolanda Melero, Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Universitat Autònoma de Barcelona, Campus UAB. Edifici C. 08193. Bellaterra Barcelona, Spain. y.melero@creaf.uab.cat; +34 935814677

Running title: Testing pedigree-based estimates of dispersal

Abstract

Estimating dispersal - a key parameter for population ecology and management - is notoriously difficult. The use of pedigree assignments, aided by likelihood-based softwares, has become popular to estimate dispersal rate and distance. However, the partial sampling of populations may produce false assignments. Further, it is unknown how the accuracy of assignment is affected by the genealogical relationships of individuals and is reflected by software-derived assignment probabilities. Inspired by a project managing invasive American mink (*Neovison vison*), we estimated individual dispersal distances using inferred pairwise relationships of culled individuals. Additionally, we simulated scenarios to investigate the accuracy of pairwise inferences. Estimates of dispersal distance varied greatly when derived from different inferred pairwise relationships, with mother-offspring relationship being the shortest (average = 21 km) and the most accurate. Pairs assigned as maternal half-siblings were inaccurate, with 64-97% falsely assigned, implying that estimates for these relationships in the wild population were unreliable. The false assignment rate was unrelated to the software-derived assignment probabilities at high dispersal rates. Assignments were more accurate when the inferred parents were older and immigrants and when dispersal rates between subpopulations were low (1 and 2%). Using 30 instead of 15 loci increased pairwise reliability, but half-sibling assignments were still inaccurate (> 59% falsely assigned). The most reliable approach when using inferred pairwise relationships in polygamous species would be not to use half-sibling relationship types. Our simulation approach provides guidance for the application of pedigree inferences under partial sampling and is applicable to other systems where pedigree assignments are used for ecological inference.

Keywords: Dispersal distance, pedigree inference, genetic markers, simulations, polygamous species

Introduction

Dispersal distance, defined as the distance travelled by an individual dispersing from the natal or breeding patch to a new settlement location, is a key parameter in many ecological models and critical for the successful management of populations (Sutherland *et al.* 2013). Whether for native or introduced species, dispersal plays a central role in population expansion and persistence by influencing connectivity between fragmented habitat patches or driving range shifts. Dispersal distance therefore defines the relevant spatial scale for management actions (Schaub *et al*. 2010). However, obtaining accurate estimates of dispersal distance is notoriously challenging.

Classical direct methods for inferring dispersal distance include field-based observations, radio-tracking, or physical tagging, which often have high associated costs, whilst producing relatively little data (Dingle 1996). The accessibility of genetic markers, such as microsatellite loci and single nucleotide polymorphisms, has made it possible to complement, or substitute, these methods by inferring pedigree relationships among the individuals in a population (Wilson & Rannala 2003; Guichoux *et al.* 2011; Norman & Spong 2015). When combining inferred pedigree relationships with location data, natal dispersal distance can be inferred. The relationships most frequently used to infer dispersal distances are pairwise relationships because breeding site-faithful parents and/or non-dispersed siblings provide information on the natal location of the individuals (e.g., García *et al.* 2007; Planes *et al.* 2009; Christie *et al.* 2010; Lambin *et al.* 2012; Almany *et al.* 2013; Norman & Spong 2015).

A major common challenge with pedigree analyses is the potential for false assignments that occur in data sets collected for natural wild populations which unavoidably results from the partial sampling of individuals and of their genome (Koch *et al.* 2008; Leonarduzzi *et al.* 2012). Where the identity of neither parent is known a-priori, a large number of pairwise comparisons of genotypes must be made between putative relatives, inflating the potential for assignment errors, especially when inbreeding and polygamy occur (Wang 2012, 2014a). Another issue arises when age cannot be determined; hence parent-offspring relationships are unclear. Given these problems, any potential false relationship assignments will necessarily result in inaccurate, and at times biased, estimations of dispersal distance, with important consequences for both the understanding of spatial dynamics as well as the efficacy of species management actions.

There are several software packages available for assessing kinship including: COLONY (Jones & Wang 2009), CERVUS (Kalinowski *et al.* 2007), FAMOZ (Gerber *et al.* 2003) and MASTERBAYES (Hadfield *et al.* 2006). These programs simultaneously assign sibships and parentage using maximum likelihood based on the allele frequencies within the pool of candidate parents and offspring. Inferred clusters of related individuals and pairwise putative relationships are produced, along with a measure of assignment certainty (as a critical log-likelihood score), confidence intervals, or assignment probabilities. From a statistical point of view, the values of these measurements could be used as criteria to eliminate false assignments, as they are all based on likelihoods given the observed data (i.e., the population sample). Both the confidence intervals and the assignment probabilities assume that the sample and input parameters are a precise representation of the actual population. However, the accuracy of the obtained likelihood of an assignment may not reflect uncertainty associated with the ecological complexity of the population (e.g., population spatial structure, level of inbreeding, mating system), as well as the partial, and potentially biased, sampling of wild populations (e.g., spatially aggregated or unevenly sampled cohorts). Likelihood-based measures of assignment accuracy may thus be statistically, but not ecologically, reliable. In this study, we used simulations to investigate the influence of key ecological parameters on the accuracy of pedigree assignments inferred by COLONY for a partially sampled wild population of a mobile mammalian predator. We used the full-likelihood algorithm in COLONY, as it is widely used and was shown to out-perform the pairwise-likelihood approaches of both CERVUS and FAMOZ, and was similar to MASTERBAYES (Walling *et al.* 2010; Karaket & Poompuang 2012; Harrison *et al.* 2013).

The American mink (*Neovison vison*, hereafter mink) is a representative example of a highly mobile species under management. The mink is widely distributed as an invasive non-native species and the focus of a substantial control effort worldwide (see Bonesi & Palazón 2007; Genovesi *et al.* 2012; Santulli *et al.* 2014). In northern Scotland, a large-scale participatory project to control American mink has been underway since 2006 (Bryce *et al.* 2011). Central to the mink control strategy are understanding the scale of dispersal, hence scope for reinvasion, and identifying 'hot spots' in the landscape that may disproportionately contribute immigrants. In a previous study, we used pedigree analyses to understand patterns of mink dispersal and to infer levels of compensatory immigration in response to culling (Oliver *et al.* 2016). Whilst analyses provided useful insights about changes in immigration rate following culling, we had the potential to provide more precise information about actual genetic relationships and used them to infer specific dispersal movements. However, when using COLONY, we observed notably different results in the estimate of dispersal distance depending on the nature of the mating system assumed (i.e., monogamy or polygamy). Estimates of dispersal distance also varied depending on the parental relationship of pairs of inferred siblings, with estimates derived from full-siblings being substantially longer than those from maternal half-siblings. This difference occurred despite inferences being well supported by assignment probabilities. Rather than being specific to the present study, this problem could arise in multiple systems, therefore broadly affecting inferences on ecological processes based on dispersal estimates (e.g., Saenz-Agudelo *et al.* 2011; López-Sepulcre *et al.* 2013).

With this study, we aimed to improve the standard approach to ecological inferences based on pedigree analyses that use field ecological data with incomplete sampling. We first investigated whether dispersal distances estimated from inferred pedigrees varied in relation to the relationship type (e.g., full- *vs* half-siblings) and also to the assignment probability estimated by COLONY. Secondly, we analysed the accuracy of inferred pairwise assignments (i.e., whether or not they were correct) in relation to the above-mentioned factors, as well as the age and source of the true parents (i.e., immigrant versus local), the dispersal rate between populations and the number of microsatellite loci used. We used the mink population from NE Scotland as a study system for the analyses of the estimates of dispersal distance and also as the basis for a set of simulation scenarios for testing the accuracy of inferred pairwise assignments.

Materials and methods

Study species

Mink are diploid, have overlapping generations, female and male polygamy, and frequent multiple paternities (Yamaguchi *et al.* 2004). Thus, a single litter can consist of full-siblings (same mother and father) or maternal half-siblings (same mother but different father). Besides, the offspring of different litters may also be paternal half-siblings (same father but different mother). Mating takes place once per year in March-April, with a single litter of typically 4-5 offspring (up to 12) born in May (Dunstone 1993; Melero *et al.* 2015). Adult males and females have separate, though potentially overlapping territories (Melero *et al.* 2008). Siblings stay in the maternal territory under the mother's protection until dispersal (circa September at five months old), at which point most juveniles leave the maternal territory to seek a vacant territory, where they will typically settle in solitude by the end of the year when they are sexually mature at eight months old (Dunstone 1993).

Age determination, DNA extraction and genotyping

The age of culled mink from the NE Scotland population was estimated by X-raying canine teeth (Helldin, 1997) and subsequently, for those individuals deemed 10 months or older, using tooth cementum analyses performed by Matson's Laboratory LLC (Manhattan, Montana, USA). DNA was extracted using the DNeasy kit (Qiagen, US) following the manufacturer's protocol. For all mink, genotyping was performed at 15 microsatellite loci developed for mustelids: Mer009, Mer022, Mer041, Mvi054, Mvi057, Mvi232, Mvi111, Mvi1321, Mvi1381, Mvi1843, Mvis022, Mvis072, Mvi4001, Mvi4031, Mvi4058 (O'Connell *et al.* 1996; Anistoroaei *et al.* 2006; Vincent *et al.* 2011). Polymerase chain reaction (PCR) amplifications were performed following Oliver *et al.* (2016).

Pedigree-based analysis

The reconstruction of litters was performed by pedigree analysis using COLONY 2.0.47 (Wang 2008). Individuals were categorised as putative mothers, fathers and offspring for each generation (year) following age and sex determination. Input parameters were set to account for mink biology: female and male polygamous mating systems without inbreeding avoidance, as is suspected to be the case for mink. Polygamy creates a far more complex problem of pedigree elucidation than monogamy. Thus, we selected the most stringent likelihood settings for pedigree reconstruction. Genotyping error rate was set as 0% to improve comparisons between the input genotypes and reduce the model computing time. Allele scoring was performed by one or two independent observers and those individuals whose genotype was ambiguous (< 1% of the total sampling) were removed and/or re-genotyped. The probabilities for mothers and fathers being present in the sample are not inferred by COLONY, but are rather set as an input parameter. We set them both at 50% in the absence of other prior information.

Analysis of dispersal distance

Individual dispersal distances were estimated based upon the Euclidean distance between the natal territory and the capture location of those genotyped mink from the empirical NE Scotland population assumed to be post-dispersal at the time of the capture (i.e., > eight months old). The approximate locations of natal territories were ascertained from the capture location of the pedigree-inferred mother, as female mink tend not to disperse once they are reproductively mature (Dunstone 1993). When the mother was not sampled, the natal location was estimated from the capture location of inferred full-siblings or maternal half-siblings that were younger than 5 months old and therefore likely pre-dispersal (i.e., siblings likely located in the maternal territory). Inferred fathers were not used to inform natal locations, as the settled location of males might change after roaming during the mating season (Dunstone 1993).

Simulated population scenarios

To test the accuracy of the inferred pairwise relationships, we used a set of simulations depicting three scenarios. In Scenario 1 (S1), we simulated a population mimicking the empirical NE Scotland mink population and its demographic and genetic dynamics over three years. The NE Scotland mink population was previously identified combining analyses of genetic clusters with analyses of the levels of relatedness and gene flow among populations recognised by geographic location (Fraser *et al*. 2013). This scenario was used to estimate the accuracy of inferred pairwise relationships for the specific case of the empirical NE Scotland mink population. In Scenario 2, we simulated over three years three subpopulations that were connected via a set of four different dispersal rates defined as the proportion of each population moving into another population (1%, 2%, 10% or 20%), therefore yielding four scenarios (S2.1-S2.4). The aim here was to understand the effect of the dispersal rate between subpopulations on the accuracy of inferred pairwise relationships. The true dispersal rate between subpopulations in the empirical NE Scotland mink population is unknown, although Oliver *et al*. (2016) estimated that, on average, 30% of captured individuals were immigrants (having no kin within 10 km of the capture location) based on kinship analysis. In Scenario 3, we repeated those most contrasting simulations from Scenario 2 (S2.1 and S2.4) while increasing the number of microsatellite loci from 15 to first 20 and then 30 for each simulation. The aim was to test weather increasing the number of loci used increased the accuracy of inferred pairwise relationships.

Scenario 1: Mimicking the Scottish mink population

We sought to mimic the empirical NE Scotland mink population using the spatial and genetic structure described by Fraser *et al.* (2013). This comprised a central population surrounded by two peripheral populations to the West and South, with little genetic divergence between the three (Fig. 1a). Peripheral areas were previously shown to have been a source of immigrants to the central population (Fraser *et al.* 2013). We therefore simulated a central population (P0 hereafter) and two peripheral populations ($IP0_i$), where i stands for each of the two peripheral populations. Initial population sizes were set as the carrying capacities of the corresponding empirical populations, calculated as the maximum number of occupied mink territories during the duration of the project (approach and details in Melero *et al.* 2015). This resulted in 520 simulated individuals for P0 and 60 for $IP0_i$.

All simulated individuals were given a unique identifier and a 15-microsatellite loci genotype, randomly generated from the allele frequencies of their corresponding empirical population, using the software NOOKIE in MS2, a C program which simulates Mendelian inheritance of markers from specified mating occasions and populations (Anderson & Dunham 2008; https://github.com/eriqande/nookie).

Individuals simulated in P0 and IP0$_i$ were crossed (bred) independently once in NOOKIE to establish a gene pool that was in Hardy-Weinberg equilibrium. Once crossed, we discarded all individuals in P0 and IP0$_i$ and retained their offspring OF0 and OF-I0$_i$, maintaining the initial population sizes. Population dynamics based on set breeding and survival parameters were then simulated for three years using mink life-history data obtained from the literature, as well as the empirical data collected from the NE Scotland population (Fig. 2, Table 1; see simulation scheme in Fig. S1a).

We allowed some individuals from the simulated peripheral populations OF-I0$_i$ and their subsequent offspring OF-I1$_i$ and OF-I2$_i$ to disperse into the simulated central population each year (OF0-OF3). We assumed that 25% of juvenile survivors (i.e., offspring of each year that survived until being reproductively active one year after birth) of the two peripheral populations successfully settled in the simulated central population (Fig. S1a), based on earlier estimates of the probability of an individual being an immigrant in the central population (Oliver *et al*. 2016) and on the genetic differentiation of the populations (Fraser *et al.* 2013). Throughout the simulation process, the resulting true genealogy (or pedigree), the age, and the source and settlement locations (or populations) of individuals were recorded to inform and to be compared with the subsequent inferences derived from pedigree analysis with COLONY.

Scenario 2: Simulating different dispersal rates between subpopulations

The spatial and genetic structure of the three simulated subpopulations ($P0_i$) in this scenario were based on three areas within the range of the empirical NE Scotland mink population (areas Dee, Spey and Tay; Fig. 1a). These areas were chosen as being the most genetically distinct based on global and pairwise measures of absolute genetic differentiation using Jost's D (D_{ST}) per year and along all years calculated using DEMEtics (Gerlach *et al.* 2010; mean D_{ST} values Dee-Spey = 0.26 (0.08-0.37), Dee-Tay = 0.11 (0.04-0.15), Spey-Tay = 0.13 (0.06-0.28)).

The simulated subpopulation sizes and individual genotypes were obtained separately for the three subpopulations following the same procedures as in Scenario 1, producing approximately 60 individuals per population. As in Scenario 1, the three simulated populations were crossed independently to allow allele and genotype ratios to equilibrate for each OF0 population, until measurements of pairwise $D_{ST} > 0.6$ (Table S1). Once crossed, we also discarded all individuals in $P0_i$ but retained their offspring $OF0_i$ ($n_{total} = 800$).

Using the three OF_i populations, we simulated their demographic and population dynamics using the same life-history parameters as Scenario 1 (Table 1, see simulation scheme in Fig. S1b). Four scenarios, with varying dispersal rates between the three OF_i populations, were defined as: S2.1 with the lowest dispersal rate of $r_d = 1\%$; S2.2 with $r_d = 2\%$; S2.3 with $r_d = 10\%$; and S2.4 with the highest rate of $r_d = 20\%$.

Scenario 3: Increasing the number of loci used

We repeated the procedure of Scenario S2.1 and S2.4 as those with most differing reliability of their inferences (see results section), but using for each scenario first 20 and then 30 loci instead of 15. The new loci were created based on randomly selecting allele frequencies of the known 15 microsatellites. Thus, our simulation explored the change in power based on the number of loci with similar allele variability to that observed in our empirical study without the confounding effect of variability in allele frequencies (see levels of variability for S1, S2 and S3 in Table S2).

Subsampling the simulated populations: mimicking the partial sampling of wild populations

In each scenario, we mimicked the partial sampling of a wild population by subsampling a set of individuals from the resulting simulated population at year three. For each scenario, the resulting simulated population was composed of the pool of parents $OF0_{S2}$, $OF1_{S1}$ and OF2, and their juvenile offspring OF3, with all individuals identified and of known sex (Fig. 2 and Fig. S1). The proportions subsampled were 67% of the juvenile offspring (OF3), 52% of adult females (females in $OF0_{S2}$, $OF1_{S1}$ and OF2) and 52% of adult males (males in $OF0_{S2}$, $OF1_{S1}$ and OF2). These proportions were chosen to reflect the proportion of captures in the empirical wild NE Scotland mink population; calculated as the proportion of captures in the control year with the highest number of captures in relation to corresponding initial population sizes (maximum number of territories; approach and details in Melero *et al.* 2015).

The genotypes of the subsampled simulated individuals then were used as the candidate parents and offspring for the COLONY input files, with inference procedures run using the same conditions as described above for the analyses of the empirical NE Scotland population.

Statistical analyses

We used a Gaussian linear model to investigate whether the inferred relationship types or likelihood-based assignment probability (*P* hereafter) explained variation in dispersal distance. Inferred relationship types were categorised as mother-offspring, full-siblings, and maternal half-siblings. We deemed father-offspring relationships uninformative about dispersal.

To quantify the accuracy of the inferred relationships, we noted the rate of discrepancy between the known simulated relationships and those inferred by COLONY per simulation, which we refer to as assignment error rate hereafter. Assignment error rate included two types of false pairwise assignments: false positives (Type I error, defined as inferred pairwise relationships that were not true despite associated values of $P \geq 0.8$) and false negatives (Type II error, defined as inferred pairwise relationships that were correct despite associated values of $P < 0.5$). We included both error types because, whilst the first leads to false estimates of ecological parameters such as, e.g., dispersal distance (given false relationships, and therefore also natal locations, are inferred), the second adds error to the estimation of true values (by overlooking true relationships), potentially biasing results at the population level (e.g., defining the risk of reinvasion). The assignment error rate was evaluated in relation to the inferred relationship type (mother-offspring, father-offspring, full-siblings, maternal and paternal half-siblings), *P*, and the age and origin (immigrant or local) of the true parents, since we expected older and/or immigrant parents to be more genetically distinct than younger and/or local parents and therefore easier to assign. The influence of dispersal rate between subpopulations was also tested by comparing simulations of Scenario 2, because higher assignment error rates are expected with low hetereozygosity, and increased dispersal should lead to higher heterozygosity within subpopulations (Saenz-Agudelo *et al.* 2009; Wang 2014b). Lastly, the effect of the number of loci was also tested comparing the simulations of Scenario 3. Analyses were performed using generalised linear models (GLM), where assignment error rate was set as the dependent variable fitted using a binomial distribution (one for false assignments) and a complementary cloglog link to account for the amount of zeros (Zuur *et al.* 2009).

Results

How far do real mink disperse?

The estimated dispersal distances for the empirical NE Scotland mink population for all individuals genotyped had a median = 27.7 km, mean = 37.1 km, min = 0 km, 1st quartile = 10.6 km, 3rd quartile = 54.9 km and max = 162.4 km. The mean dispersal distance was nearly twice as large when estimated using inferred maternal half-siblings (39 km, n = 756) for inferring the natal location than when estimated using mother-offspring relationships (21 km, n = 312), and the value of the third quartile increased 1.5-fold (40 to 58 km; Fig. 1b-d). The distance estimates using mother-offspring and full-sibling relationship types (mean = 28 km, n = 38) were equivalent and statistically significantly shorter than the estimates for half-siblings (half-siblings *vs* mother-offspring, HS *vs* MO, $F_{2, 478}$ = 4.11; p-value < 0.001). The estimated dispersal distance decreased significantly as the assignment probability increased, with a 2-fold decrease for inferred relationships with $P \geq 0.8$ relative to those with $P \leq 0.1$ (Estimate$_p$ = -31.83 km; $F_{1, 479}$ = -4.53; p-value < 0.001; Fig. 3); demonstrating that choice of P would affect biological inference.

Scenario 1: What influences assignment error rate in the mimicked Scottish mink population?

Fifty-five per cent of all of inferred pairwise relationships under Scenario 1 were false positive assignments, which can be clearly seen in Fig. 4. Of these, 90% were given an assignment probability value of $P > 0.9$ by COLONY. A small percentage (0.7%) were false negatives with $P < 0.5$ (details in Table S3).

The inferred relationship type had a clear and significant impact on assignment error rate, being lowest for full-siblings (5.2%) and highest for maternal and paternal half-siblings (64.5 and 71.9%, respectively; Table S3). Inferred mother- and father-offspring relationships had equal reliability with an average assignment error rate of 13.2%, but were less reliable than full-siblings (FS *vs* MO, $\text{Estimate}_{\text{fs}}$ = -2.04, SD_{fs} = 1.01, $F_{3,\,833}$ = 4.11; *p*-value = 0.04; Binomial GLM). However, the assignment error rate for half-siblings was significantly higher than for both mother- and father-offspring relationships (MHS *vs* MO, $\text{Estimate}_{\text{hs}}$ = 0.86, SD_{hs} = 0.18, $F_{3,\,833}$ = 4.65; *p*-value < 0.001; Figs. 4 and 5). The assignment error rate decreased as P increased (Estimate = -2.20, SD = 0.68, $F_{1,\,835}$ = -3.24; *p*-value > 0.019, r^2 = 0.17) for all of the inferred relationship types (*p*-values $_{P\text{:relationship type}}$ > 0.53, interaction dismissed during model selection), although model predictions differed between them (Fig. 4). Whilst the origin of each parent (local or immigrant) did not affect the assignment error rate (averages of 10.3% and 20. 3%, respectively; $F_{1,\,835}$ = 0.03; *p*-values > 0.98), parental age did, with assignment error rate being approximately 1.6 times lower when at least one parent was older than one year (from an average error rate of 38.7% to 24.2% and 16.7% when at least one parent was three, two and one year old, respectively; $F_{2,\,833}$ = -2.17 and -4.34, *p*-values = 0.007 and < 0.001, respectively).

Scenario 2: What influences assignment error rate?

The assignment error rate increased with increasing dispersal rate, being similar for simulations S2.1 and S2.2 at 40% ($r_{S2.1} = 0.01$ *vs* $r_{S2.2} = 0.02$; $F_{3,\,11061} = -0.59$; *p*-value = 0.55), but being 1.15- and 2-fold higher in simulations S2.3 and S3.4, respectively, at 46 and 85% approximately ($r_{S2.3}=0.1$ and $r_{S2.4}=0.2$; $F_{3,\,11061} = 4.05$ and 33.24, *p*-values < 0.001; Fig. 5 and Table S3).

Mother- and father-offspring relationships consistently had the highest assignment accuracies, with an error range of 17-56.6% for S2.1 and S2.4, whereas half-sibling assignments were always the least reliable, with a range of 61.8-97% for S2.2 and S2.4 (Table 2; Fig. S2 and Table S3). The usefulness of *P* as a predictor of assignment error rate decreased with the dispersal rate. When dispersal rate was low in S2.1 and S2.2, the assignment error rates were estimated to decrease 1.2- and 1.5-fold, respectively, while *P* increased from ≤ 0.2 to ≥ 0.8. This was not the case in S2.3 and S2.4, the scenarios with highest dispersal rates, where no relationship between *P* and assignment error was detected, although *P* in S2.3 had similar patterns to those seen in S2.1 and S2.2 (Table 2 and Fig. S2). Both the age and origins of parents were also influential. Having at least one older parent (two and three years old) reduced the assignment error rate by 1.7-2.3 times compared to having one-year-old parents in S2.1 and S2.2 (from 40 to 23 and 17% on average), a pattern to which we return in the discussion. However, this effect was not found in S2.3 or S2.4. Having immigrant parents reduced the assignment error rate by 1.6 times (from 44 to 28% on average), but only in S2.1 and S2.2 (Table 2).

Scenario 3: What is the impact of increasing the number of loci on assignment error rate?

The overall assignment error rate of simulations with low dispersal rate ($r_{S2.1} = 0.01$) did not differ when using 15 or 20 loci with an average error rate of 46 and 43%, respectively; however, it decreased 1.5-fold to 30% when using 30 loci ($F_{3,\,9305} = 0.10$, *p*-values = 0.03; Table S4). However, when dispersal rate was high ($r_{S2.4} = 0.2$), the error rate decreased to *ca*. half from ca. 85% when using 15 loci to 40% and 39% when using 20 and 30 loci, respectively ($F_{3,\,7843}$ = -2.00 and -20.2, *p*-values < 0.001; Table S4).

Increasing the number of loci from 15 to 20 and 30 decreased the error rate for mother- and father-offspring relationships, as well as of that of full-siblings for all simulations (range 0.7-11.3%; Table S4) with these relationship types being equally reliable. Half-siblings assignments were still not reliable, with error rates ranging from 59.4-97.3% and 59.9-84% when using 20 and 30 loci, respectively (Table 3 and Table S4). The error rate decreased 1.2- and 1.7-fold when P increased from ≤ 0.2 to ≥ 0.8 for both dispersal rates when using 30 loci (Table 3), but it was unreliable for half-siblings in all simulations (Fig. S3). Having parents older than one year old reduced the error rate by 3.7 and 6.4 times (from 58% to 15% and 9% on average, respectively, for one-to-three years old) when dispersal rate was low ($r_{S2.1} = 0.01$), but not when it was high ($r_{S2.4} = 0.2$) using either 20 or 30 loci. The same pattern occurred in relation to the origins of the parents; individuals with at least one immigrant parent were 1.22 times more likely to be properly linked to a relative than when their parents were local (from 42 to 34% on average) if dispersal rate was low using 20 or 30 loci (Table 3).

Discussion

Using simulations, we assessed the accuracy of inferences of different pairwise relationships derived from pedigree analysis using COLONY with data consisting of multiple microsatellite genotypes for individuals from partially sampled wild populations, a system that is typical of field-based ecological studies. We used a data set on invasive American mink populations in NE Scotland to parameterize our models. We found that different types of pairwise relationships were reconstructed with variable assignment error rates, and hence that inference of dispersal distances based on pedigree reconstruction would be strongly affected by the type of relationship being examined. Mother-offspring relationships and relationships involving full-siblings yielded the most reliable relationship reconstructions. In contrast, inferred maternal half-sibling pairs were nearly always unreliable, illustrating the perceived difficulty of making pedigree inferences in species with polygamy and multiple paternity (e.g., rainforest birds, Woltmann *et al.* 2012). However, the ecological reality is that multiple paternities are commonplace in many species, but the accuracy of pedigree inferences in their presence are frequently overlooked (e.g., water voles, Telfer *et al.* 2003; capercaillie, Kormann *et al.* 2012; Roanoke logperch, Roberts *et al.* 2016). The assumed rate of dispersal in the simulated populations, the age of the parent and whether they were themselves dispersers - parameters themselves rarely known with certainty in field studies - also impacted the accuracy of pedigree reconstruction. However, the impact of each of these varied depending on the context of population structure and dispersal rate. Additionally, the likelihood-based assignment probability (P) provided by the software was a good predictor of accuracy when dispersal rates between subpopulations were low (< 20%) or when number of loci used was higher than in most field studies (e.g. 30), but not otherwise.

Estimates of dispersal distance

The impact of the low reliability of inferred maternal half-sibling pairs was evident in the analyses of estimated individual dispersal distances for the empirical NE Scotland population, which was the focal ecological question underpinning our application of the pedigree-based approach. Estimates of dispersal distance averaged 21-28 km when using mother-offspring and full-siblings, shown by simulation to be the most reliable relationships. The estimate increased to 37.1 km when including maternal half-siblings (39 km when using maternal half-siblings only), the least reliable relationship type. The difference was larger still when considering upper quartiles of the dispersal distance distributions. We interpret this difference as reflecting the fact that inferred maternal half-sibling relationships include numerous falsely assigned relationships. *In extremis*, false assignments imply falsely pairing individuals that are randomly distributed in space. Thus, in a hypothetical situation in which all pairwise relationships were false, estimated dispersal distance would approach the mean pairwise distance between all individuals, which is bound to upwardly bias estimates relative to real dispersal wherever the study area is larger than dispersal distance. Applications of the pedigree reconstruction approach should therefore ensure that inferences are robust to the use of different relationship types; otherwise, as observed here, the consequences for our ecological understanding could be substantial.

Estimated dispersal distance gradually shrank with the increase of the assignment probability provided by COLONY. This is an indication of the potential usefulness of P to predict the accuracy of inferred estimates for the specific case of the empirical NE Scotland population. We note however that in the case of maternal half-siblings, the predicted assignment error rate at $P = 1$ was still 50% and therefore useless or, worse, misleading.

Factors affecting assignment error rate

The assignment error rate of the inferred pairwise relationship types increased with increasing simulated dispersal rate between neighbouring subpopulations, with an increase from 40% in simulations with 1 and 2% dispersal rate (Scenarios S2.1 and S2.3) to approximately 85% in Scenario S2.4 with 20% dispersal. Excluding half-siblings reduced the assignment error rate in the scenarios with lower dispersal rate, resulting in 21% error rate in S2.1 and S2.2 and in 31% in S2.3 when excluding half-siblings, but not in the scenario with the highest dispersal rate (76% error rate in S2.4). The increased dispersal rate in the simulated scenarios led to reduced population genetic differentiation between and increased heterozygosity within the three subpopulations (Table S1), which negatively affected the accuracy of parentage assignments (Cornuet *et al.* 1999; Wang 2002; Saenz-Agudelo *et al.* 2009). Likewise, assignment error rate was lower when specifically involving immigrant parents, but only when dispersal rates were low (1 and 2%) such that immigrants originated from more genetically differentiated populations.

At lower dispersal rates, the accuracy of inferences was not improved by increasing from 15 to 20 loci, but did when 30 loci were used (from 40% using 15 and 20 loci, to 29% using 30 loci). While at high dispersal rates, error decreased from 85% using 15 loci, to 40% using 20 and 30 loci. In all cases, excluding half-siblings reduced the error to < 10% using 20 loci and < 6% when using 30. This indicates the improvement provided by increasing the number of loci at both low and high dispersal rates for all relationship types except for half-siblings.

Overall, mother-offspring and full-sibling relationships were consistently the most reliably assigned relationship type in most simulations, although assignment reliability at the highest dispersal rate (20%, S2.4) was poor (assignment error rate = 57 and 72% for MO and FS respectively) when using only 15 loci, in contrast to the very poor assignment error rate of maternal half-siblings, which was consistently higher than 56% in all simulations. Given that our findings are extendable to the numerous examples of partially sampled and polygamous natural systems, we advocate that polyandry should be allowed when COLONY, or equivalent software, is used, but that the pairwise assignment involving half-siblings should be considered separately from the other, more reliable, assignments, if at all.

The reliability of assignment probability provided by COLONY to reflect the confidence in assignment depended on the ecological circumstances mimicked by the simulations (i.e., increased dispersal rate reduced its reliability). Because the probability is constrained by the data provided, as the data become more complex – as happens with the increase of dispersal rate - then the likelihood that these *P*-values reflect reality diminishes. This ecological context dependence is an undesirable property that undermines the value of the metric for practical applications where the true ecological context is not known. Although the uncertainty in assignment is not easily resolvable, our simulations support the suggestion that increasing the number of loci improves the confidence of the assignment probability (Harrison *et al.* 2013). This improvement occurred for all parent-offspring and full sibling relationships with error rates < 10% for all simulations with 20 or 30 loci independently of *P*. The increase of the number of loci also increased the reliability of *P* for half-siblings but only when dispersal rate was high, with average increases of ca. 1.12 and 1.1 times, respectively, for 20 and 30 loci. However, the fact that all error rates of half-siblings were > ca. 60% independently of the values of *P* (Fig. S3) indicates a lack of accuracy of these relationship types despite the increase of the number of loci.

As the age of parents increased, the assignment error rate decreased for all relationship types. Erroneous assignments are more likely to occur between related individuals (e.g., aunt-offspring relationships) than between random individuals, since the former are genetically more distinct. Therefore, we interpret this outcome as reflecting the fact that, probabilistically, older individuals should have fewer close relatives alive amongst the potential parent sample. Assignments involving older parents had lower error rates for all but the two scenarios with the highest dispersal rates (S2.3 and S2.4) independently of the number of loci used, which is likely due to the confounding influence of lower genetic differentiation between the three subpopulations. Although the age of the parents is a difficult variable to measure in field situations, it is of particular interest for harvested or culled species, such as mink, that are under long-term control. Indeed, that population age structure is driven downwards to consist primarily of juveniles and subadults with increasing years of control (Melero *et al.* 2015). This process ultimately could lead to a decrease in the accuracy of pedigree inferences through the duration of a control programme.

Nevertheless, despite clear limitations and the salient grounds for greater caution than has hitherto been applied (e.g., by Lambin *et al.* 2012; Kormann *et al.* 2012), the parentage reconstruction approach scrutinised in this paper nevertheless provides information on dispersal distance that would otherwise be unobtainable. The importance of the error rate, and resulting potential bias in estimating dispersal parameters, obviously depends on the specific application, since error rates that may be intolerable in a heritability or animal model analysis may make little difference in inferences about some populations-level variables such as dispersal rate along gradients of density. Artificially inflating the tail of a dispersal distribution, a consequence of assignment error rate, may lead to overestimations of range spread (Kot *et al.* 1996) and predicted recolonisation rate, but our analyses imply that sub-setting dispersal distance data by relationship type would provide a way to detect any such bias, if present.

Conclusions

Overall, although the simulations that we implemented could easily be repeated or extended to mimic specific study systems, many of our findings on the accuracy of parentage relationships assignments are applicable to other systems and could be used as rules-of-thumb for situations where ecological information is limited. Additionally, the inherent weakness identified here apply not only to the use of inferred pedigree assignments, but also to other current methodologies aiming to address similar questions in partially sampled populations, even though the specific sources of error and bias may vary. The use of large numbers of SNPs for example, provides for greater subsampling of the genome than microsatellites, but in most realistic field scenarios, the limitations of partial sampling of the pool of potential parents and of the number of available loci will remain (e.g., Norman & Spong 2015). Our findings provide an alternative solution to identify and reduce the limitations of ecological inferences from pairwise analyses of wild populations.

Alternatively, when using inferred pairwise relationships from partially sampled populations to estimate ecological parameters, such as dispersal distance, a conservative approach is to use only mother-offspring and full-sibling relationships, and not use inferred half-sibling relationships. Whilst this approach could potentially be wasteful of data, under no scenario were these relationships sufficiently reliable to inform, rather than bias, data sets. Since the utility of P and age and origin of the parents depends on the dispersal rate, an exercise such as that demonstrated in Scenario 1 (i.e., simulating the population of interest) should help to identify the most reliable relationship types and their levels of influence.

Acknowledgements

YM was funded by a Marie Curie FP7-PEOPLE-2011-IEF 300288-Project Depensation and XL and MO by NERC Grant NE/J01396X/1. We thank the Scottish Mink Initiative, staff, funders and multiple mink volunteers for the continued effort, samples and data; Andrzej Zalewski and Gillian Murray-Dickson for assisting with genotyping samples; Kamil Barton for his help getting started with NOOKIE programming; Stuart Piertney and Joan Pino for their helpful comments on the manuscript; the three anonymous reviewers and the editor Paul Bentzen for their useful comments.

References

Almany GR, Hamilton RJ, Bode M *et al.* (2013) Dispersal of Grouper Larvae Drives Local Resource Sharing in a Coral Reef Fishery. *Current Biology*, **23**, 626–630.

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program Structure. *Molecular Ecology Resources*, **8**, 1219–29.

Anistoroaei R, Farid A, Benkel B, Cirera S, Christensen K (2006) Isolation and characterization of 79 microsatellite markers from the American mink (*Mustela vison*). *Animal Genetics*, **37**, 185–188.

Bonesi L, Palazón S (2007) The American mink in Europe: Status, impacts, and control. *Biological Conservation*, **134**, 470–483.

Bryce R, Oliver MK, Davies L *et al.* (2011) Turning back the tide of American mink invasion at an unprecedented scale through community participation and adaptive management.

Biological Conservation, **144**, 575–583.

Christie MR, Tissot BN, Albins MA *et al.* (2010) Larval connectivity in an effective network of marine protected areas. (Goldstien SJ, Ed,). *PLoS ONE*, **5**, e15715.

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

Dingle H (1996) *Migration: The Biology of Life on the Move*. Oxford University Press, New York, USA.

Dunstone N (1993) *The Mink*. T and A D Poyserd, Ltd. London.

Dunstone N, Birks JDS (1985) The comparative ecology of coastal riverine and lacustrine mink *Mustela vison* in Britain. *Zeitschrift fuer Angewandte Zoologie*, **72**, 59–70.

Fraser EJ, Macdonald DW, Oliver MK, Piertney S, Lambin X (2013) Using population genetic structure of an invasive mammal to target control efforts – An example of the American mink in Scotland. *Biological Conservation*, **167**, 35–42.

García C, Jordano P, Godoy JA (2007) Contemporary pollen and seed dispersal in a *Prunus mahaleb* population: patterns in distance and direction. *Molecular Ecology*, **16**, 1947–55.

Genovesi P, Carnevali L, Alonzi A, Scalera R (2012) Alien mammals in Europe: updated numbers and trends, and assessment of the effects on biodiversity. *Integrative Zoology*, **7**, 247–253.

Gerber S, Chabrier P, Kremer A (2003) FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes*, **3**, 479–481.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics!

Molecular Ecology, **19**, 3845–52.

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

Hadfield JD, Richardson DS, Burke T (2006) Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, **15**, 3715–30.

Harrison HB, Saenz-Agudelo P, Planes S, Jones GP, Berumen ML (2013) Relative accuracy of three common methods of parentage analysis in natural populations. *Molecular Ecology*, **22**, 1158–70.

Jones OR, Wang J (2009) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551–555.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology*, **16**, 1099–106.

Karaket T, Poompuang S (2012) CERVUS vs. COLONY for successful parentage and sibship determinations in freshwater prawn *Macrobrachium rosenbergii* de Man. *Aquaculture*, **324–325**, 307–311.

Koch M, Hadfield JD, Sefc KM, Sturmbauer C (2008) Pedigree reconstruction in wild cichlid fish populations. *Molecular Ecology*, **17**, 4500–11.

Kormann U, Gugerli F, Ray N, Excoffier L, Bollmann K (2012) Parsimony-based pedigree analysis and individual-based landscape genetics suggest topography to restrict dispersal and connectivity in the endangered capercaillie. *Biological Conservation*, **152**, 241–252.

Kot M, Lewis MA, van den Driessche P (1996) Dispersal data and the spread of invading organisms. *Ecology*, **77**, 2027.

Lambin X, Le Bouille D, Oliver MK *et al.* (2012) High connectivity despite high fragmentation: smart iterated dispersal in a vertebrate metapopulation. In: *Dispersal Ecology and Evolution* (Clobert J, Baguette M, Benton TG, Bullock JM, Eds.). Oxford University Press.

Leonarduzzi C, Leonardi S, Menozzi P, Piotti A (2012) Towards an optimal sampling effort for paternity analysis in forest trees: what do the raw numbers tell us? *iForest - Biogeosciences and Forestry*, **5**, 18–25.

López-Sepulcre A, Gordon SP, Paterson IG, Bentzen P, Reznick DN (2013) Beyond lifetime reproductive success: the posthumous reproductive dynamics of male Trinidadian guppies. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20131116.

Melero Y, Palazón S (2011) Visón americano-*Neovison vison* (Schreber, 1777). In: *Enciclopedia Virtual de los Vertebrados Españoles* (Salvador A, Cassinello J, Eds.).

Melero Y, Palazón S, Revilla E, Martelo J, Gosàlbez J (2008) Space use and habitat preferences of the invasive American mink (*Mustela vison*) in a Mediterranean area. *European Journal of Wildlife Research*, **54**, 609–617.

Melero Y, Robinson E, Lambin X (2015) Density- and age-dependent reproduction partially compensates culling efforts of invasive non-native American mink. *Biological Invasions*, **17**, 2645–2657.

Norman AJ, Spong G (2015) Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. *Ecology and Evolution*, **5**, 3056–3065.

O’Connell M, Wright JM, Farid A (1996) Development of PCR primers for nine polymorphic American mink *Mustela vison* microsatellite loci. *Molecular Ecology*, **5**, 311–312.

Oliver MK, Piertney SB, Zalewski A, Melero Y, Lambin X (2016) The compensatory potential of increased immigration following intensive American mink population control is diluted by male-biased dispersal. *Biological Invasions*, **18**, 3047–3061.

Planes S, Jones GP, Thorrold S (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5693–5697.

Roberts JH, Angermeier PL, Hallerman EM (2016) Extensive dispersal of Roanoke logperch (*Percina rex*) inferred from genetic marker data. *Ecology of Freshwater Fish*, **25**, 1–16.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology*, **18**, 1765–76.

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2011) Connectivity dominates larval replenishment in a coastal reef fish metapopulation. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 2954–2961.

Santulli G, Palazón S, Melero Y, Gosálbez J, Lambin X (2014) Multi-season occupancy analysis reveals large scale competitive exclusion of the critically endangered European mink by the invasive non-native American mink in Spain. *Biological Conservation*, **176**, 21–29.

Sutherland WJ, Freckleton RP, Godfray HCJ *et al.* (2013) Identification of 100 fundamental ecological questions. *Journal of Ecology*, **101**, 58–67.

Telfer S, Piertney SB, Dallas JF *et al.* (2003) Parentage assignment detects frequent and large-scale dispersal in water voles. *Molecular Ecology*, **12**, 1939–1949.

Vincent IR, Farid A, Otieno CJ (2011) Variability of thirteen microsatellite markers in American mink (*Mustela vison*). *Canadian Journal of Animal Science*, **83**, 597–599.

Walling CA, Pemberton JM, Hadfield JD, Kruuk LEB (2010) Comparing parentage inference software: reanalysis of a red deer pedigree. *Molecular Ecology*, **19**, 1914–28.

Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–15.

Wang J (2008) Colony version 2.0.4. www.zsl.org/science/software/colony.

Wang J (2012) Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics*, **191**, 183–94.

Wang J (2014a) Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology*, **27**, 518–530.

Wang J (2014b) Estimation of migration rates from marker based parentage analysis. *Molecular Ecology*, **23**, 3191–3213.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Woltmann S, Sherry TW, Kreiser BR (2012) A genetic approach to estimating natal dispersal distances and self-recruitment in resident rainforest birds. *Journal of Avian Biology*, **43**, 33–42.

Yamaguchi N, Sarno RJ, Johnson WE, O’Brien SJ, Macdonald DW (2004) Multiple paternity an reproductive tactics of free-ranging American minks, *Mustela vison*. *Journal of Mammalogy*, **85**, 432–439.

Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) *Mixed Effects Models and Extensions in Ecology with R* (Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, Eds.). Springer, New York, USA.

Data Accessibility

Empirical field data of the culled American mink in NE Scotland (ID, sex, age, year) and their genotypes at 15 microsatellites loci; initial simulation data for S1 (P0 and $IP0_i$), and S2 and S3 ($P0_{Dee}$, $P0_{Spey}$ and $P0_{Tay}$); and R codes for the mink life-history parameters used for the simulations can be found archived in Dryad. DOI: http://dx.doi.org/10.5061/dryad.cj0k2

The Nookie software (Anderson & Dunham 2008) and its source code is freely available at https://github.com/eriqande/nookie.

Author Contributions

Y.M., M.K.O. and X.L. conceived the study questions and design. Y.M. and M.K.O. genotyped the culled mink, performed the pedigree analyses and estimated dispersal distance. Y.M. performed the simulations and the statistical analyses. X.L. supervised the research. Y.M., M.K.O. and X.L. interpreted the results and wrote the manuscript.

Fig. 1 Empirical data from the NE Scotland mink population. (a) The spatial structure: a central population (dots, $n = 728$ sampled individuals) and two peripheral populations (white squares, $n = 117$ and yellow squares, $n = 134$), as used for Scenario 1. For Scenario 2 we used mink data from three main river catchments: the Spey (white squares), Tay (yellow squares) and Dee (red dots, $n = 189$). Estimated natal capture locations (orange triangles), offspring post-dispersal capture locations (blue dots), estimated dispersal distance with $P < 0.8$ (blue dashed lines) and ≥ 0.8 (green solid lines) and corresponding histograms derived from inferred: (b) mother-offspring, (c) full-sibling and (d) maternal half-sibling relationships.

Fig. 2 Simplified scheme of simulated mink reproduction, not including immigration from peripheral areas (full details in Fig. S1a). P0 stands for the central population. OF*j* stands for the offspring of each j_{th} generation 0-3, where OF0 is in Hardy-Weinberg equilibrium. Light grey shows offspring, medium grey for one-year survivors, and dark grey for two-year survivors. Individuals inside a box show the pool of parents reproducing that year, solid arrows link surviving individuals, dashed arrows link parents and offspring. Subsampling was done in year three, when three cohorts and their offspring are present.

Fig. 3 Estimated individual dispersal distance (in m, denoted with circles) derived from inferred pairwise assignments, in relation to the assignment probability (*P*) of these assignments. Lines relate to the best fit linear model predictions of the relation between distance and P for mother-offspring (black outlines unfilled, $n = 312$), full-siblings (full grey circles and dark grey lines, $n = 38$) and maternal half-siblings (unfilled light grey outlines and light grey lines, $n = 756$) relationship types. Shaded areas bounded by dotted lines show the 95 % CIs. Observations with P ~1 on the *x*-axis have had a slight lateral offset applied to aid visualisation.

Fig. 4 Assignment error rate (false positives are circles scoring 0% error and $P \geq 0.8$; false negatives are dots scoring 100% and $P < 0.5$), and its model-derived predictions (lines) in relation to the assignment probability *P* under Scenario 1, for: (a) mother-offspring (black circles and lines, $n = 312$) and father-offspring (grey circles and grey shadowed lines, $n = 239$), and (b) full-siblings (black circles and lines, $n = 38$) and maternal half-sibling relationship type (light grey circles and grey shaded area, $n = 756$). Continuous lines relate to the estimated fit setting values for other parameters as origin = local and age = two years old (median); dashed lines denote the 95 % CIs.

Fig. 5 Assignment error rate for each inferred pairwise relationship type in the simulations for Scenario 1 and Scenario 2 (S2.1-S2.4) with r_d standing for the dispersal rate.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Full simulation scheme of: (a) Scenario 1 and (b) Scenario 2.

Fig. S2 Assignment error rate and its model-derived predictions in relation to the assignment probability *P* for mother-offspring, father-offspring, full-siblings and half-sibling relationship type for simulations S2.1-S2.4 all with 15 loci.

Fig. S3 Assignment error rate and its model-derived predictions in relation to the assignment probability *P* for mother-offspring, father-offspring, full-siblings and half-sibling relationship type for simulations S2.1 and S2.4 all with 20 and 30 loci.

Table S1 Pairwise genetic differentiation and heterozygosity of the simulated populations at Year 1 for Scenario 2.

Table S2 Number of alleles, allelic richness and heterozygosity per locus and population of the simulated populations at Year 1 for (a) Scenario 1, (b) Scenario 2 and (c) Scenario 3.

Table S3 Percentage of assignment error rate of inferred assignments per relationship type categorised as false positives and true assignments in relation to the assignment probability (*P*) for Scenario 1 and Scenario 2.

Table S4 Percentage of assignment error rate of inferred assignments per relationship type categorised as false positives and true assignments in relation to the assignment probability (*P*) for Scenario 3.