MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 267–277 doi: 10.1111/1755-0998.12590

WISEPAIR: a computer program for individual matching in genetic tracking studies

ANDREW P. ROTHSTEIN, RYAN MCLAUGHLIN, ALEJANDRO ACEVEDO-GUTIÉRREZ and DIETMAR SCHWARZ
Department of Biology, Western Washington University, Bellingham, WA 98225, USA

Abstract

Individual-based data sets tracking organisms over space and time are fundamental to answering broad questions in ecology and evolution. A 'permanent' genetic tag circumvents a need to invasively mark or tag animals, especially if there are little phenotypic differences among individuals. However, genetic tracking of individuals does not come without its limits; correctly matching genotypes and error rates associated with laboratory work can make it difficult to parse out matched individuals. In addition, defining a sampling design that effectively matches individuals in the wild can be a challenge for researchers. Here, we combine the two objectives of defining sampling design and reducing genotyping error through an efficient Python-based computer-modelling program, WISEPAIR. We describe the methods used to develop the computer program and assess its effectiveness through three empirical data sets, with and without reference genotypes. Our results show that WISEPAIR outperformed similar genotype matching programs using previously published from reference genotype data of diurnal poison frogs (*Allobates femoralis)* and without-reference (faecal) genotype sample data sets of harbour seals (*Phoca vitulina*) and Eurasian otters (*Lutra lutra)*. In addition, due to limited sampling effort in the harbour seal data, we present optimal sampling designs for future projects. WISEPAIR allows for minimal sacrifice in the available methods as it incorporates sample rerun error data, allelic pairwise comparisons and probabilistic simulations to determine matching thresholds. Our program is the lone tool available to researchers to define parameters a priori for genetic tracking studies.

Keywords: *Allobates femoralis*, genetic tracking, individual identification, *Lutra lutra*, *Phoca vitulina*, simulation software

Received 5 April 2016; revision received 23 June 2016; accepted 11 July 2016

Introduction

Tracking individuals in the wild is fundamental to answering broad questions relating to population structure, trophic interactions, behavioural patterns and life history events (Clutton-Brock & Sheldon 2010). Individual-based data sets can elucidate intraspecific differences pertaining to trophic and foraging ecology (Newsome *et al.* 2009; Arnould *et al.* 2011; Hückstädt *et al.* 2012), population dynamics (Vindenes *et al.* 2008) and disease ecology (Johnson *et al.* 2009), highlighting important patterns and processes dictating interactions among species (Bolnick *et al.* 2003; Cianciaruso *et al.* 2009). A longitudinal study is one method for investigating individual differences that may affect the broader patterns and processes, such as those described above (Bolnick *et al.* 2002). Historically, researchers have used nongenetic tags to track individuals, such as human-made coloured bands or skin brands employed in marine mammal studies (Merrick *et al.* 1996; Hazen *et al.* 2012; Walker *et al.* 2012), or photo-identification through individual morphological marks on the animal (Stevick *et al.* 2001; Speed *et al.* 2007). However, these traditional methods have limitations such as 'tag loss' where the tag is no longer on the species. In addition to loss of tags, researchers need to continuously observe individuals in the field, and, unless accompanied with supplementary data, the lack of information beyond simple individual identification limits the questions that can be investigated. A 'permanent' genetic tag potentially circumvents a need to invasively tag animals and may be the only feasible method if there are little phenotypic differences among individuals (Palsbøll 1999). These genetic tags can arise from direct sources (blood and tissue) or from noninvasive sources such as hair, faeces, sloughed or shed skin, urine and saliva (Waits & Paetkau 2005). A genetic tag fulfils many important characteristics necessary to track individuals effectively including the following: universal applicability, potential noninvasiveness,

Correspondence: Andrew P. Rothstein, Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA; E-mail: arothste24@gmail.com

no significant loss of tags, lack of ambiguity among individuals and rapid matching of tags once the method is established (Palsbøll 1999). Genetic tags therefore allow for collecting robust longitudinal data sets. However, the use of genetic tags also requires special attention to experimental design.

Major considerations regarding genetic tracking include logistics and costs. Both the sampling (number of samples/bouts) and genotyping (laboratory work/genotyping error) necessary to track individuals force researchers to make trade-offs in the design of their project (Hoban 2014). We define sampling design as the number of samples, bouts and genetic markers used to appropriately address a research question. Sampling design has been previously highlighted as an important component to improving accuracy in noninvasive population studies (Marucco *et al.* 2011). However, studies rarely mention sampling designs for genetic tracking and typical results are of population size estimation via noninvasive recaptures, which require fewer resamples than longitudinal tracking of individuals.

Generally, there are two objectives in the design and a posteriori evaluation of a genetic tracking study: the likelihood of resampling an individual in a population and estimating the effects of genotyping error on resampling confidence. These questions can be investigated by developing optimal sampling designs to ensure that a given study can obtain enough samples in the field to resample individuals. However, genotyping error, which is usually taxa and sample-quality specific (Taberlet & Luikart 1999), leads to discrepancies between genotypes of two distinct samples from the same individual. The second objective is therefore to determine whether or not the genotyping error rate observed in a study will inhibit the ability to identify those resamples. One way to address the second objective is determining through rerun PCR samples or statistical error rates in programs such as GIMLET (Valière 2002), GEMINI (Valière *et al.* 2002), CERVUS (Kalinowski *et al.* 2007), PEDANT (Johnson & Haydon 2007) or DROPOUT (McKelvey & Schwartz 2005; Schwartz *et al.* 2006), where to assign a threshold in allelic differences for individual identification. While these objectives have previously been separated (determining resamples with error and designing optimal sampling designs for resampling individuals), they are inextricably linked when it comes to genetically tracking individuals.

Developing a successful genetic tracking study thus requires considering both genotyping error and sampling design. There are few tools available to researchers to define parameters a priori for noninvasive genetic tracking. We thus propose a hybrid approach that can integrate both repeated PCRs and a computer-based approach for addressing genotyping error when matching individual samples. We combine these two objects and developed WISEPAIR, an experimental design model for individual-based ecological questions that simulates genetic tracking and genotyping error. Our goals were to develop a method to genetically track individuals and to develop an a priori optimal sampling design to assist in effective experimental designs. We describe our methods used to develop the computer program and assess its effectiveness through three empirical data sets. One data set came from Ringler *et al.* (2014) and had reference genotypes of diurnal poison frogs (*Allobates femoralis*). The two without-reference genotype data sets are from noninvasive faecal samples of harbour seals (*Phoca vitulina*) (Rothstein 2015) and Eurasian otter (*Lutra lutra*) (Lampa *et al.* 2015). The source code is available at: https://github.com/McGlock/WisePair.

Methods

Finding matches through virtual genetic tagging – a probability model

A probability model was created to address the following main objectives: (i) simulate sampling designs from virtual populations, (ii) determine resamples of individuals through allelic pairwise comparisons and (iii) optimize sampling designs for future project development. Our program consists of three main scripts: *beanbag.py*, *wisepair.py* and *optimagic.py*. The *beanbag.py* script is specifically designed to build virtual individual genotypes of a population to be used in simulated sampling. This design is based on user-supplied criteria such as number of individuals in the population, number of loci and allelic frequencies. In addition, this script incorporates genotyping error rates during sampling. The second script, *wisepair.py*, was created to determine the number of resamples within a specified data set (real or virtual) through allelic pairwise comparisons. *wisepair.py* determines the number of resamples within a virtual data set, the number of resamples within an actual data set using specified threshold simulations, estimates the number of errors for resamples and determines whether resamples can be distinguished from non-resamples. The final script, *optimagic.py,* utilizes outputs from both *beanbag.py* and *wisepair.py* to develop optimal sampling designs for individual-based studies. *beanbag.py* and *wisepair.py* were then used to produce a threshold 'score' with which we could compare samples to a field data set and subsequent simulations in *optimagic.py*. The following describes the methods for each script within the WISEPAIR suite:

beanbag.py. *beanbag.py* creates a population with simulated genotypes, followed by running a virtual sampling season on the population. It accepts a JSON file that

contains number of loci (L), number of alleles (A) and allelic frequencies (AHz) for respective alleles. From this JSON file, it creates a simulated population for user-specified number of individuals. This virtual population is used to construct genotypes for each individual using the provided AHz and a Pythonic implementation of the Mersenne Twister, a pseudo-random number generator (Matsumoto & Nishimura 1998). For each L, the following processes are run: (i) an A is randomly drawn, (ii) its AHz is compared to a continuously randomized probability value (CRPV) from 0 to 1, (iii) a particular A is assigned to an L when AHz is ≥ the probability value, (iv) these three steps are then repeated for all loci for each individual until the virtual population is completely built. From this virtual population, the script simulates a sampling season with user-provided criteria, such as number of bouts and samples per bout. For each bout, the samples are randomly chosen, without replacement, from the available individuals until the number of samples for that bout is met. The population list is refreshed for each bout.

To accurately address genetic sampling, the model incorporates simulations of allelic dropout (ADO) and false allele (FA) error rates for the genotypes sampled. Empirical error rates and user-specified error rates are used to simulate genotyping error, where we incorporate ADO and FA into the genotype for each individual using the PEDANT software suite (Johnson & Haydon 2007). PEDANT per allele error rates are compared to a CRPV from 0 to 1. If the error rate is ≥ CRPV, then an error occurs for that allele; FA generated first, followed by ADO. It is important to note that the script places an 'unknown allele' for FA as it cannot determine what allele would actually be substituted, unlike a false allele in a real data set. For matching purposes, the FA is treated as another allele and not ignored when matching genotypes. This treatment ignores potential false positive matches; however, these matches would be exceedingly rare. While this model accepts a virtual determination of error rates, it can incorporate previously determined rates by the user. Following these steps, the sampling season is saved as a comma-separated variable (.csv) format. This standard output is used in the *wisepair.py* scoring algorithm. The *beanbag.py* script can be used for implementation and simulation of virtual populations and sampling needed when no data are available or included in iterative runs of the *wisepair.py* script to determine threshold values for determining resamples in actual data sets.

wisepair.py. The second script, *wisepair.py*, either imports the standard output of *beanbag.py* or user-supplied data in.csv format. From these imports, a full list of all pairwise comparisons for every sample is assembled. The pairwise list is run through a scoring function that compares genotypes of each pair and returns a similarity score. Initially, a raw similarity score (RSS) is determined, which is the sum of allelic differences of each pairwise comparison where a lower score indicated higher similarity. A corrected similarity score (CSS) is computed to account for variable number of loci being included (as some samples may have missing data for certain loci – for our 'without-reference' data set, we removed samples with >25% missing loci) in the scoring of each pair (CSS = RSS/[# of loci used]). Each CSS is normalized (NCSS), for graphical clarity, by subtracting the overall CSS mean and then dividing by the difference of the maximum CSS and minimum CSS:

$$\mathrm{NCSS} = \frac{[\mathrm{CSS} - \overline{\mathrm{CSS}}]}{\mathrm{CSS}_{\max} - \mathrm{CSS}_{\min}}.$$

When analysing simulated data from *beanbag.py*, a 'virtsim' ID code is included. This code allows for error-free identification of individuals, even if ADO or FA introduces discrepancies between identical genotypes. Using these IDs, *wisepair.py* builds a resampled threshold range for NCSS. The threshold range is established from the lower limit of a confidence interval around the mean of the NCSS for the unpaired comparison and the upper limit of a confidence interval around the mean for the resampled comparisons (±5% of NCSS for resampling which is 95% CI for our analyses). These ranges can be applied later to empirical data sets to identify resampled individuals. The simulated NCSS are plotted onto histograms for visual inspection of the frequency distribution of resampled individuals and distinct, newly sampled individuals. The *wisepair.py* and *beanbag.py* scripts are used for both simulations in the following script and determining resample thresholds for a project's data set (ex. Figs 1–3).

optimagic.py. The third and final script used in the program is an optimization script for both threshold values and sampling designs. This script effectively optimizes potential sampling designs by iteratively running *beanbag.py* and *wisepair.py*. The possible variables that can be optimized are number of bouts, samples per bout, counts of resampled individuals and count of times an individual is resampled over a season. Given all the specified variables, *optimagic.py* performs simulations of all possible combinations of values or ranges using the previous scripts. *beanbag.py* and *wisepair.py* iterate each design and determine the number of resamples and non-resamples using the threshold model. Following these design simulations, all scoring data are parsed and resampled individuals are counted. These data are stored in two possible.csv files. If the simulation meets, the specified

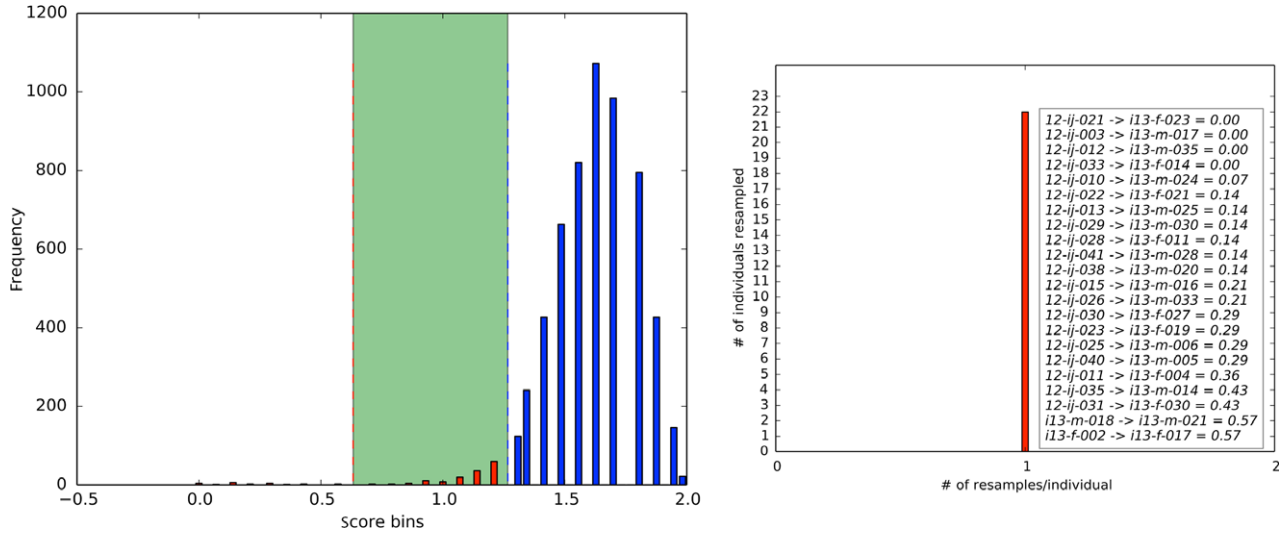

Fig. 1 Threshold histogram for tadpole data set. Simulations were iterated 100 times to determine threshold values. Blue bins represent nonresampled pairwise comparisons and red bins represent potential matched pairs. Green polygon identifies those pairwise comparisons that are overlapping comparisons with green indicating that the lower and upper bounds do not overlap (red dotted = lower, blue dotted = upper). All known matches were binned to the left of the red dotted lower threshold (0.56). The blue-dotted line signifies the upper limit of potential resamples. Second histogram shows example of detailed matches, specifically for the reduced data of juvenile-adults.

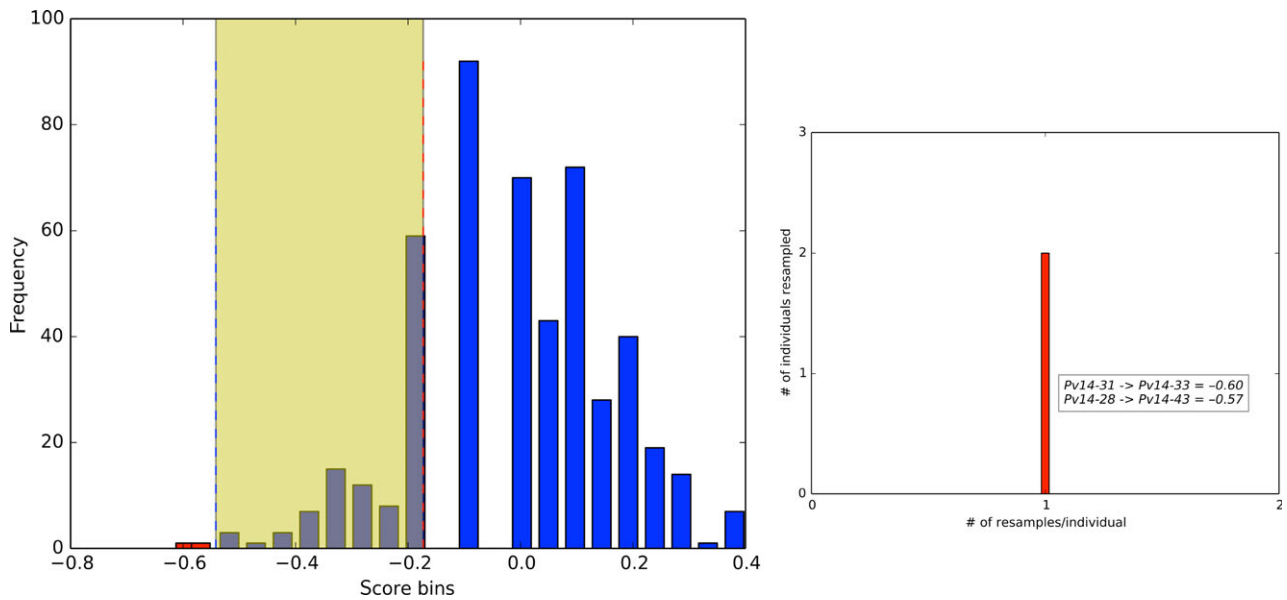

Fig. 2 Threshold histogram for harbour seal data set. In contrast to Fig. 1, the yellow polygon shows overlap of the upper and lower bounds for each confidence interval. Ambiguous or questionable resamples fall within the yellow polygon. Based on 1000 iterative simulations to determine threshold values for resampled individuals, the bounds of the corrected score were included in the histogram of pairwise. Based on the lower bound threshold, the simulations determined that there were two pairs of samples that were identified as resampled individuals (1) Pv14-28 and Pv14-43 and (2) Pv14-31 and Pv14-33, which are shown in the second graph of number of individual resampled.

resampled minimum and the mean number of times an individual is resampled then data for that sampling design are saved within the acceptable sampling file. If either of the criteria were not met for the simulations, then the sample designs failed and are placed in the unacceptable sampling file. These data can then be used to determine the best sample design for a given range of criteria (ex. Fig. 4a–c). The following sections explain how we tested these methods on empirical data sets with and without-reference genotypes.

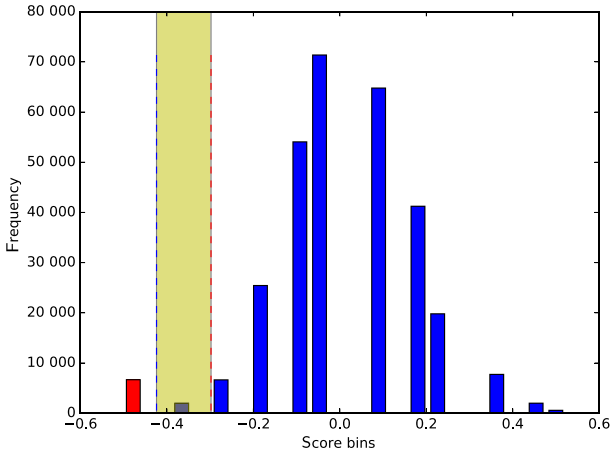

Fig. 3 Threshold histogram for Lampa *et al.* (2015) data set. Comparable to Fig. 2, positioning of the blue- and red-dotted lines as well as the yellow polygon shows overlap in matching thresholds for pairwise comparisons. Based on 100 iterative simulations, the single red bin represents potential matched individuals with blue bins representing non-resamples. All known matches were binned to the left of the blue-dotted line. Second histogram shows all resamples for all pairwise comparisons with the WISEPAIR iterative model. Removing redundancies, the number of individuals resampled is 65, with five 'false' resamples due to matching opposite sexes. The 19 singleton individual genotypes are not classified as resamples within the model.

With-reference genotype sample set

With-reference genotype samples refer to data sets that have available, known genotypes for individuals. Having with-reference genotypes allows for researchers to confidently match individuals to this reference as opposed to without-reference genotypes (typical in noninvasive studies) that leaves uncertainty to matching pairs of genotypes. The major differences in these data sets are the assumption that the with reference genotypes are error-free and without-reference are error-prone (Pompanon *et al.* 2005; Johnson & Haydon 2007). For our study, we applied data from Ringler *et al.* (2014), which consisted of tissue sampled 1800 tadpoles of diurnal poison frogs (*Allobates femoralis)* that were released into 20 artificial pools, followed by a second survey yielding 42 juvenile samples, with a final sampling of 36 males and 31 females. Individuals were matched with photograph ID from both juveniles and adult stages which ensured that genotypes matched photo ID results (juvenile to adult). Through these three sampling bouts, all samples were genotyped with 14 microsatellite loci (see Ringler *et al.* 2014). This data set was separated into two main groups: reduced and full. The full data set included all potential matches from tadpoles to juveniles to adult life stages (trios, $n = 20$), while the reduced set only included pairs from juveniles to adults (pairs, $n = 76$) (Ringler *et al.* 2014).

Without-reference genotype sample set

We collected 46 scat samples from harbour seals in three sampling periods during January–March 2014 (Jan. $n = 21$, Feb. $n = 12$, Mar. $n = 13$) from a single haul out site in Cowichan Bay, Vancouver Island, British Columbia (Rothstein 2015). The haul out is comprised of floating logs (log booms) that are available to harbour seals year-round (Cottrell 1995; Baird 2001). From the 46 scat samples collected, we successfully genotyped 32 samples through at least seven of nine microsatellite loci used. We applied loci, which were initially developed in tissue and blood, and tested them on scat samples (Burg 1996; Gemmell *et al.* 1997; Davis *et al.* 2002). We used a random number generator to identify 20% of the total samples size for reamplification and repeated genotyping. Rerun samples were analysed with PEDANT, which used a maximum-likelihood estimation of allelic dropout (ADO) and false allele (FA) rates when there is an absence of reference data (a common limitation with unknown individuals or in noninvasive genetic sampling) (Johnson & Haydon 2007). In addition, all samples were sexed using a ZFX/ZFY protein qPCR assay adapted from Matejusová *et al.* (2013), developed specifically for harbour seals. All samples were run with positive controls of known male and female scat samples acquired from captive harbour seals at Vancouver Aquarium in Vancouver, BC and Point Defiance Zoo & Aquarium, Tacoma, WA (Rothstein 2015).

Our final data set was of Eurasian otter (*Lutra lutra)* faecal samples from Lampa *et al.* (2015). This data set included 778 multilocus microsatellite (7) genotypes over a six-year period. From 2006 to 2012 (missing 2009 year), the number of genotyped samples was 121, 134, 96, 130, 138 and 159, respectively. In addition to the complete genotype, the authors used a *Lut-SRY* sex determination marker that helped in identification of individuals. Each genotype has both a sample ID and otter ID to describe matching genotypes. Due to high error rates during PCR amplification (genotyping error rates over all years were 48.9%, ADO = 45.1%, FA = 3.8%), Lampa *et al.* (2015) employed a multitubed approach with at least five amplification repeats to determine a consensus genotype. Based on their data set, they found 84 distinct individuals of the 778 genotypes.

Incorporating genotyping data into scripts

We used the WISEPAIR suite to determine the number of individuals resampled within the respective data sets. For the *wisepair.py* script, we used error rates determined in PEDANT for harbour seal data and published error rates for Ringler *et al.* (2014) and Lampa *et al.* (2015). In order to effectively and confidently identify resampled individuals, each data

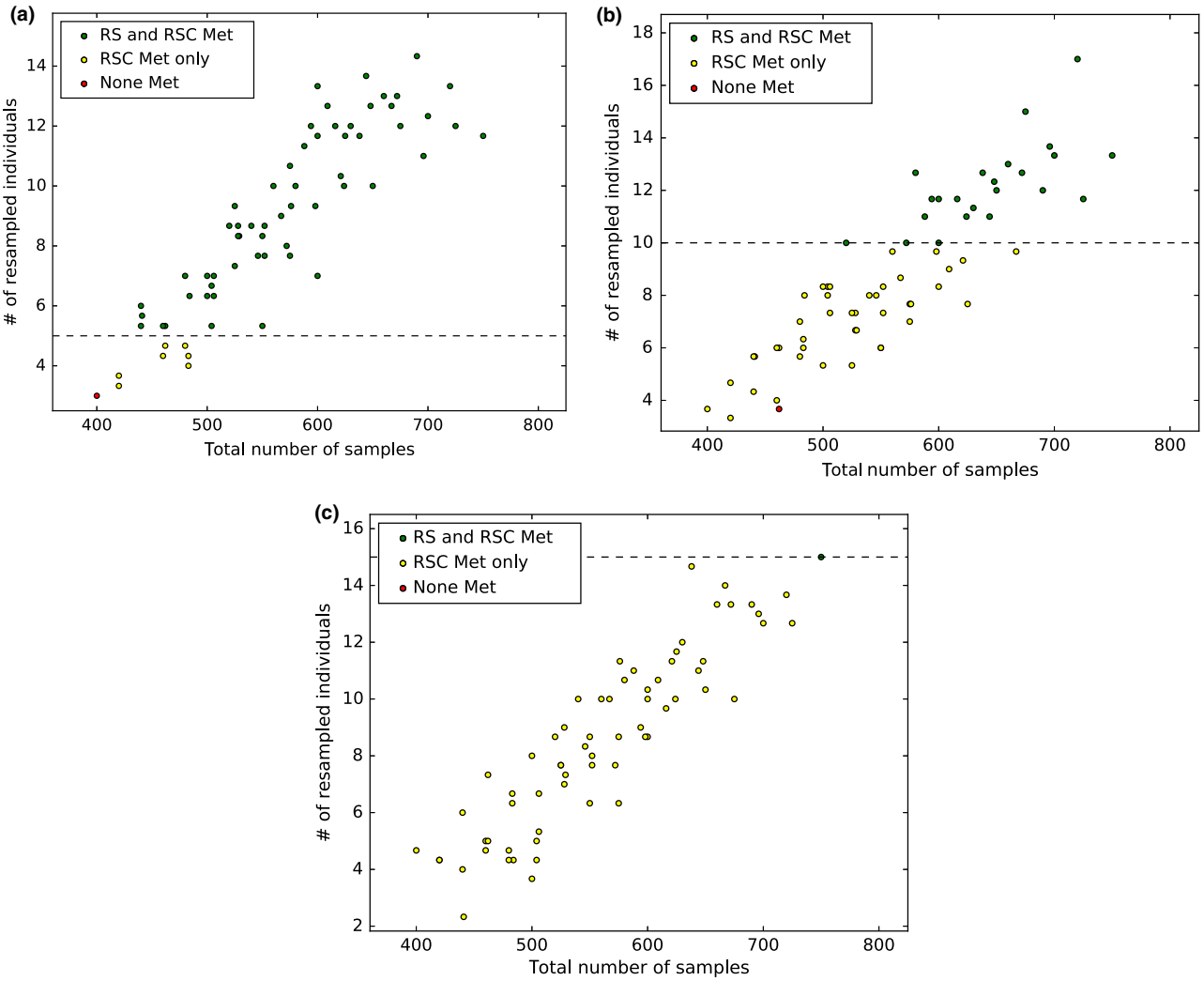

Fig. 4 optimagic.py optimal sampling schemes for a population of 100 individuals with a random 50% absent at any given bout. Criteria included a sampling effort of 20–30 scats for each visit for 20–25 bouts. (A) Depicts scheme of minimum of five individuals resampled, (B) depicts scheme of minimum of 10 individuals resampled and (C) depicts schemed of 15 individuals resampled. Dotted line represents this minimum number of resampled individuals sampled at least four times. Each scheme was iterated three times. Data points above dotted lines represent schemes that met both criteria of resampled counts (RSC) and number of individuals resampled (RS), in green. Yellow circles only met one of the criteria and red dots represent schemes that met none of the conditions. Schemes that met the input criteria ranged in sample sizes of 440–750 total samples.

set was compared through thresholds from *wisepair.py* to simulated designs under different conditions. These conditional simulations in *optimagic.py* included a population based on our data-observed allele frequencies, number of alleles and estimated error rates. *optimagic.py* was used as a means to iteratively run *beanbag.py* and *wisepair.py* for comparison purposes to the data sets.

For the Ringler *et al.*'s (2014) data, we used simulations that matched the published sampling design (1800 tadpoles, 42 juveniles and 67 adults) and iteratively ran 100 simulations to determine threshold values, similar to the harbour seal data. The difference in these simulations was that Ringler *et al.* (2014) had variable samples sizes while the harbour seal, and otter data were based on a set number of samples per bout. In addition, as per published methods (Ringler *et al.* 2014), matches were determined for both a reduced data set (juvenile–adults) and full data set (trios: tadpole–juveniles–adults).

The harbour seal data were used with 1000 iterations of a virtual population of 100 individuals, 150 total sample size and five bouts. For Lampa *et al.* (2015) data, we used 100 iterations of 21 individuals, with sampling design matching the number of genotyped samples during six bouts. The number of individuals was based on

Lampa *et al.* (2015) capture–mark–recapture population estimates. For both data sets, iteration simulations were averaged from corrected threshold values for each iteration and compiled to determine threshold values for identifying resampled individuals in the data sets.

Optimizing for future projects using optimagic.py

The final simulations determined the best sampling design for future individual-based genetic tracking studies, specifically for the haul out at Cowichan Bay. A high-frequency simulation was used based on the assumption that researchers would want to resample individuals more frequently (at least 4–6 times per individual) than in the harbour seal data set we presented. Due to permit restrictions for our collections, we were unable to use a high-frequency sampling design such as the one in this simulation. However, this circumstance provided an opportunity to demonstrate the effectiveness of *optimagic.py* as an experimental design tool. From a population of 100 individuals at Cowichan Bay (Olesiuk 2009), the *optimagic.py* script was run to fit parameters that would include a variety of potential designs. We determined designs that matched incremental number of individuals resampled: 5, 10 and 15+ individuals. In addition to the high-frequency sampling parameters, we used a population of 100 individuals with an estimate that a random 50% of individuals are absent from the haul out at any given time. Therefore, a random 50 individuals of the population are sampled without replacement during each bout. While harbour seals can be extremely variable in their haul out patterns based on life history factors (Brown & Mate 1983; Yochem *et al.* 1987; Huber *et al.* 2001) and can be locally variable (Thompson 1989), a modest estimate of 50% of seals hauled out is consistent with previously observed estimates of harbour seal behaviour (Yochem *et al.* 1987). Using this scenario, we targeted the minimum number of individuals that would be resampled based on the designs for individuals sampled ≥4 times. In addition, each design was iteratively run three times to give minimal stability to the output.

Results

With-reference genotype sample set

As highlighted above, with-reference genotypes refer to data sets that have a library of known genotypes or individuals to compare to sampled data. This type of data is especially important as it assumes to be comprised of error-free reference genotypes for matching purposes. Through model statistics described in the methods section, the main threshold scores range was 0.00 in the lower bounds and 0.57 in the upper bounds (Fig. 1). From this range, we identified 22 distinct matches between juveniles and adults and 67 distinct matches between tadpoles and adults. The additional two matches were false positives as they were matching adults-to-adults – which would be separate individuals. From the full data set, we identified 20 trio matches from tadpole–juvenile–adult and 20/20 for juvenile-adults. For comparison purposes, the results are added to the published table in Ringler *et al.* (2014) and shown in Table 1. Compared to other available programs, WISEPAIR outperformed all programs tested in the Ringler *et al.* (2014) study. Specifically, as compared to the next most accurate program (ML-RELATE), WISEPAIR correctly identified all tadpoles to adult matches, which outperformed previous comparisons (Table 1). However, our results were the only program to identify matches within sampling bouts; suggesting that in some cases, it could overestimate matches without a relatedness or sex determination control.

Table 1 Extended table from Ringler *et al.* (2014) with WISEPAIR results

	Reduced data set			Full data set				
Program	Correct	α-error	β-error	Correct	α-error	β-error	Trios	Adult-Tp
IDENTITY	4/20	0	16	n/a	n/a	n/a	n/a	n/a
GENECAP	19/20	0	1	19/20	0	1	12/20	46/67
GENALEX	20/20	0	0	n/a	n/a	n/a	n/a	n/a
ALLELEMATCH	19/20	0	1	19/20	0	1	11/20	36/67
KINGROUP	20/20	0	0	20/20	0	0	19/20	64/67
ML-RELATE	20/20	0	0	20/20	0	0	20/20	61/67
WISEPAIR	22/20*	2	0	20/20	0	0	20/20	67/67

WISEPAIR under a reduced data set (only juveniles and adults) identified 22/20 compared to other matching programs with α-error and β-error. With full data set, juveniles and adults resulted in 20/20 with trios (tadpole–juvenile–adults) and 67/67 for adult–tadpole.
*There were 22/20 matches with the excess being different adults matching – potentially due to relatedness.

Without-reference genotype sample set

Based on the simulated sampling design for the harbour seal data, the threshold range was −0.542 in the lower bounds and −0.173 in the upper bounds. This threshold value included two sets that were identified as two recaptures (Pv14-28/Pv14-43 and Pv14-31/Pv14-33) (Fig. 2). Based on this pairwise match, we were able to match two sets of samples, identifying a recapture of two individuals. These two samples stood apart from the data set due to the use of an 'outside' control marker in the sex determination (see Methods). The threshold values could become more obscure if there were no control markers in the study.

For the Lampa *et al.* (2015) data set, the threshold range was −0.321 in the lower bounds and −0.172 in the upper bounds (Fig. 3). Using the pairwise resampling data and matching to otter ID, we identified 65 individuals that were resampled at least one time, and 19 additional individuals that were singleton genotypes. In addition to the 84 matches, we had five matched samples, based on genotype, that were different sexes (Male–Female or vice versa). This mismatched result suggests that these were false matches and most likely related individuals (Table S1, Supporting information).

optimagic.py Results

Due to the limited number of individual harbour seals that were resampled, it was informative to determine the optimum sampling design for that system. Based on our three scenarios (number of individuals resampled: 5, 10 and 15 individuals, n = 100, 20–25 sampling bouts and 20–30 samples per bout), there were 58 designs that fit for five individuals resampled, 23 designs for 10 individuals and only one design fit for 15 individuals. The optimum designs for five individuals ranged from a minimum of 440 samples over 22 bouts, to 750 samples over 25 bouts (30 samples per bout) (Fig. 4a). For the minimum optimum design, there were a total of 17 individuals that were resampled, with three that were resampled ≥4 times (mean count of resamples per individual = 4.97). For 750 samples over 25 bouts, samples included 27 individuals that were resampled at least once, with 11 individuals resampled ≥4 (mean count of re-samples per individual = 6.1). For 10 resampled individuals, designs ranged from 520 to 750 samples, with 20 and 25 bouts, respectively (Fig. 4b). With a minimum of 22 individuals resampled and three individuals resampled ≥4 times (mean count of resamples per individual = 5.23). The maximum design included 27 individuals resampled, with ~4 individuals resampled ≥4 times (mean count of resamples per individual = 5.9). The final design of 15 individuals only had one optimum of 750 samples over 25 bouts with 27 individuals resampled (four individuals resampled ≥4 times; mean count of resamples per individual = 4.1) (Fig. 4c). There were a number of designs that met the minimum number of resampled individuals but did not meet the number of times those individuals would be resampled (≥4) highlighted by yellow points in Fig. 4a–c.

Discussion

While genetic tracking has been a promising technique for researchers in wildlife science, the specific laboratory challenges and few studies comparing different approaches leave opportunities for methodological advancement (Beja-Pereira *et al.* 2009). With increased availability of genetic technologies and the need for a well-planned experimental design (Schwartz & Monfort 2008, p. 240; Hoban 2014), a study that developed a new method for matching individuals and simulating experimental design for individual tracking is pertinent for the progression of this research. This is the first study to develop a combined experimental design and pairwise matching software that is specifically written for the genetic tracking of individuals in individual-based studies and is universally applicable to any taxon.

Our study matched correctly the reduced and trio data sets of Ringler *et al.* (2014) which tracked diurnal poison frogs (*Allobates femoralis)* through three life stages, testing the effectiveness of available genetic matching programs (Fig. 1, Table 1). However, for juvenile-adults, there were 22 matches compared to the published 20 matches. From the list of matches, our program matched adults-to-adults in the data sets. One explanation for this overestimation of matches is the lack of a relatedness function within the program. While there is most likely relatedness among individuals in a population, a study investigating mating systems in the diurnal poison frog showed that 82.4% of matings were between unrelated individuals, 15.2% between half-sibs and only 2.4% between full-sibs (Ringler *et al.* 2012). Another explanation for these mismatches is based on the confidence intervals during simulations. More stringent confidence intervals (>95%) may have lowered our β-error for juvenile-adult reduced data set. However, WISEPAIR was intentionally run without regard to metadata of samples (sex and age class) to test overall performance. Therefore, the overestimation could be due to WISEPAIR's nondiscriminatory pairwise comparisons between juveniles and adult genotypes. This distinction could be advantageous when researchers have metadata on individual samples, such as in Ringler *et al.* (2014), to appropriately identify psuedoreplicates versus resampled individuals. A feature that corrects for relatedness would be an added benefit to our analysis and could help elucidate the fine-scale

differences between individuals for questions pertaining to population substructure and mating systems.

For the trio matching (67/67) and adult-tadpole (20/20) in the full data set, the results matched exactly the published results of Ringler *et al.* (2014). This result highlights the ability of WISEPAIR to effectively handle large data sets for pairwise comparison, an advantageous feature for researchers that want to apply high-frequency sampling designs or large population sampling. As compared to other programs used in Ringler *et al.* (2014), WISEPAIR outperformed other comparable matching programs (Table 1). In addition, this result shows the α/β-error mismatch observed in the reduced data set does not appear to be a consistent problem of the software.

For Lampa *et al.* (2015), WISEPAIR correctly identified all 84 otter individuals. WISEPAIR identified 65 individuals that were resampled at least once and 19 individuals that were observed only one time (Table S1, Supporting information). Comparable to the results of Ringler *et al.* (2014), we identified matches that had different sexes. This result again supports that overestimation could be based on relatedness, but in this empirical case, the result is less ambiguous with the addition of a sex marker. Out result is important as it highlights WISEPAIR's ability to handle lower quality data based on high error rates – a variable that is especially relevant when working with noninvasive samples.

In contrast, for our harbour seal data, the two instances of resampled harbour seal individuals were not sufficient for parsing individual differences within a population. Previous studies have made recommendations for the number of samples needed in noninvasive studies, with some advising the number of faecal samples exceeding 2.5- to threefold the number of animals expected to be sampled (Solberg *et al.* 2006; Marucco *et al.* 2011). This number of faecal samples is based on the expectation that 20–30% of all samples sample are unable to be genotyped. With this argument, the number of samples needed to effectively track 100 harbour seals at Cowichan Bay would be 250–300 samples. However, suggested samples size is usually based on capture–recapture studies for population estimation where it is not necessary to have multiple recaptures per individual. In addition, these estimates are solely based on sample failure estimates in laboratory work and not with regard to genotyping error in subsequent postlaboratory analyses. This general estimation is not consistent with the *optimagic.py* output that recommended across all designs 440–750 samples to effectively track 5–15 individuals in a population of 100 seals at Cowichan Bay (assuming 50% are absent at any given sampling period). The *optimagic.py* provides a robust tool for planning studies requiring longitudinal sample design; a feature that was not previously available for researchers. As highlighted in the introduction, much of the previous literature uses assumptions for sampling design that fail to incorporate important facets of a noninvasive study, namely genotyping error which can affect a project's ability to match individual genotypes (Taberlet & Luikart 1999). While *optimagic.py* provides an effective starting point for researchers to determine how many samples they would need to effectively track individuals, it is important to note that models do not come without assumptions and the results of our harbour seal data may not be the definitive number of samples needed, especially with only one rematched individual.

However, many pilot studies, such as the harbour seal data, are integral parts to developing larger projects. *optimagic.py* helps support future designs by incorporating criteria such as genotyping error, number of resamples and number of times an individual will be resampled that are important parameters for projects attempting to genetically track individuals in the wild. The ranges of *optimagic.py* optimal designs (see Results) highlight the power of the model and the variety of designs that can work for researchers in a given question. The fluctuation optima are related to the combinations of bouts and sample sizes that dictate changes to whether or not it will meet an optimum. With the pseudo-random nature of some of the algorithms as stated in the methods, there will be events in the simulations that create fluctuations of optimal designs. The advantage to *optimagic.py* is that researchers can choose from the data set and rerun *optimagic.py* iteratively under one design. This output could give better insights into whether or not an individual design matches the researchers' guidelines. For example, in our harbour seal study, cost per sample was a concern due to logistics, and therefore, it may be advantageous to use the smallest sample size possible from the *optimagic.py* results. With this research limitation, the smallest samples size to ensure we could resample at least the minimum of five individuals effectively would be 440 samples over 22 bouts. It is important to note that 20–30 visits to a haul out site could be deemed invasive on harbour seals due to repeated harassment (Suryan & Harvey 1999). While typically genetic samples have been obtained through capturing an animal (tissue and blood), scat still serves as a minimally invasive option even with disturbing a haul out site.

There are some assumptions included in the pipeline described that should be addressed in future studies. One is the assumption that all samples are of the same quality (for scat, freshness based on time since defecation). There has been some work performed to determine the DNA degradation rates of scats in the field (Piggott 2005; Murphy *et al.* 2007; Brinkman *et al.* 2010). A recent study investigating faecal deposition rates and DNA degradation to optimize sampling design in Sonoran pronghorn (*Antilocapra americana sonoriensis*) determined

that a sampling interval of 4–7 days under summer conditions proved most advantageous (Woodruff *et al.* 2014). However, these rates may be site or species specific and would be important for future studies with samples in other environment (such as marine) to assess the per cent of degradation affecting the number of samples needed. This assessment would allow *optimagic.py* to appropriately buffer for these samples that may fail in the field. However, if not applicable in the *optimagic.py* program, future studies could plan to buffer for these failed samples. While genotyping error is incorporated, the failure rate of samples (i.e. number of missing data points) would be an additional parameter that would merit inclusion in future. Another assumption in this study stated that a false allele would be treated as another unique allele. In real data sets, a false allele can range in base pair length, which can increase the number of different false alleles present within a locus. For instance, one locus in the harbour seal data set, *Lc26*, had the highest false allele rate (0.25 false alleles per genotype). These false alleles differed from two base pairs up to 16 base pairs, which incorporated six new alleles categorized as false alleles. In the current simulation, false alleles present would be treated and represented as a single error instance. This assumption can overinflate the number of false alleles present and potentially hide matching genotypes in a data set or simulation. However, a false allele leading to a false 'positive' match with another individual in the population appears to be highly unlikely.

While there are a multitude of questions that can be answered with individual-level data, species-specific methodological considerations are imperative to a successful project. Genetic tracking studies can be especially successful in species with some fidelity to a location (haul outs, latrines, breeding grounds, etc.) provided an adequate sampling design is in place. Our study effectively developed a computer program tool that researchers can use for projects in individual genetic tracking by optimizing sample size through incorporating expected sampling population size, genetic error rates and sampling with and without reference genotypes. Based on cost and logistics, it is important that future studies identify the trade-offs among differing methods and apply the most robust techniques and available tools to address matching genotyping and errors associated. Nevertheless, our approach allowed for minimal sacrifice in the available methods as it incorporated sample rerun error data, allelic pairwise comparisons and probabilistic simulations to determine matching thresholds. Researchers can expect to develop more robust data sets that capture differences among individuals while addressing logistical and financial concern that can lead to prohibitive sampling designs and analyses.

Acknowledgements

We thank Dr. Austen Thomas for his collaboration and support with field sampling for our harbour seal data. Thanks to Sheena Majewski, Research Biologist at the Department of Fisheries and Oceans Canada, for field support and collections during harbour seal sampling season. We are grateful to undergraduate research assistants, Theresa Keates, Christa Kohnert and Jennifer Lopez for their assistants in genotyping laboratory work with the harbour seal data set. We also thank the two anonymous reviewers and MER Editor Dr. Eric Coissac for feedback on our manuscript. Funding was provided by Western Washington University Project Development Award, Western Washington University RSP and the Charlton Family Endowment. Collection of samples was conducted with or by Sheena Majewiski, under 'License to Study Marine Mammals for Research Purposes MML-003'.

References

Arnould J, Cherel Y, Gibbens J, White J, Littnan C (2011) Stable isotopes reveal inter-annual and inter-individual variation in the diet of female Australian fur seals. *Marine Ecology Progress Series*, **422**, 291–302.

Baird RW (2001) Status of harbour seals, Phoca vitulina, in Canada. *Canadian Field-Naturalist*, **115**, 663–675.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, **9**, 1279–1301.

Bolnick DI, Yang LH, Fordyce JA, Davis JM, Svanbäck R (2002) Measuring individual-level resource specialization. *Ecology*, **83**, 2936–2941.

Bolnick DI, Svanbäck R, Fordyce JA *et al.* (2003) The ecology of individuals: incidence and implications of individual specialization. *The American Naturalist*, **161**, 1–28.

Brinkman TJ, Schwartz MK, Person DK, Pilgrim KL, Hundertmark KJ (2010) Effects of time and rainfall on PCR success using DNA extracted from deer fecal pellets. *Conservation Genetics*, **11**, 1547–1552.

Brown RF, Mate BR (1983) Abundance, movements, and feeding habits of harbor seals, Phoca vitulina, at Netarts and Tillamook Bays, Oregon. *Fishery Bulletin*, **81**, 291–301.

Burg TM (1996) *Genetic analysis of eastern pacific harbour seals (Phoca vitulina richardsi) from British Columbia and parts of Alaska using mitochondrial DNA and microsatellites*. M.Sc., The University of British Columbia, Vancouver, British Columbia.

Cianciaruso M, Batalha M, Gaston K, Petchey O (2009) Including intraspecific variability in functional diversity. *Ecology*, **90**, 81–89.

Clutton-Brock T, Sheldon BC (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**, 562–573.

Cottrell PE (1995) *Diet, activity budgets, and movement patterns of harbour seals (Phoca vitulina) in Cowichan Bay and adjacent areas*. M.Sc., University of Victoria, Victoria, British Columbia.

Davis C, Gelatt T, Siniff D, Strobeck C (2002) Dinucleotide microsatellite markers from the Antarctic seals and their use in other pinnipeds. *Molecular Ecology Notes*, **2**, 203–208.

Gemmell N, Allen P, Goodman S, Reed J (1997) Interspecific microsatellite markers for the study of pinniped populations. *Molecular Ecology*, **6**, 661–666.

Hazen EL, Maxwell SM, Bailey H *et al.* (2012) Ontogeny in marine tagging and tracking science: technologies and data gaps. *Marine Ecology Progress Series*, **457**, 221–240.

Hoban S (2014) An overview of the utility of population simulation software in molecular ecology. *Molecular Ecology*, **23**, 2383–2401.

Huber HR, Jeffries SJ, Brown RF, Delong RL, Vanblaricom G (2001) Correcting aerial survey counts of harbor seals (Phoca vitulina richardsi) in Washington and Oregon. *Marine Mammal Science*, **17**, 276–293.

Hückstädt L, Koch P, McDonald B *et al.* (2012) Stable isotope analyses reveal individual variability in the trophic ecology of a top marine predator, the southern elephant seal. *Oecologia*, **169**, 395–406.
Johnson PC, Haydon DT (2007) Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data. *Genetics*, **175**, 827–842.
Johnson CK, Tinker MT, Estes JA *et al.* (2009) Prey choice and habitat use drive sea otter pathogen exposure in a resource-limited coastal system. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 2242–2247.
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.
Lampa S, Mihoub J-B, Gruber B, Klenke R, Henle K (2015) Non-invasive genetic mark-recapture as a means to study population sizes and marking behaviour of the elusive Eurasian Otter (*Lutra lutra*). *PLoS ONE*, **10**, e0125684.
Marucco F, Boitani L, Pletscher DH, Schwartz MK (2011) Bridging the gaps between non-invasive genetic sampling and population parameter estimation. *European Journal of Wildlife Research*, **57**, 1–13.
Matejusová I, Bland F, Hall AJ *et al.* (2013) Real-time PCR assays for the identification of harbor and gray seal species and sex: a molecular tool for ecology and management. *Marine Mammal Science*, **29**, 186–194.
Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **8**, 3–30.
McKelvey K, Schwartz M (2005) Dropout: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes*, **5**, 716–718.
Merrick RL, Calkins DG, Loughlin TR (1996) *Hot Branding: A Techniques for Long-term Marking of Pinnipeds*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, Washington.
Murphy MA, Kendall KC, Robinson A, Waits LP (2007) The impact of time and field conditions on brown bear (Ursus arctos) faecal DNA amplification. *Conservation Genetics*, **8**, 1219–1224.
Newsome SD, Tinker MT, Monson DH *et al.* (2009) Using stable isotopes to investigate individual diet specialization in California sea otters (Enhydra lutris nereis). *Ecology*, **90**, 961–974.
Olesiuk P (2009) *An Assessment of Population Trends and Abundance of Harbour Seals (Phoca vitulina) in British Columbia*. Research Document 105. Canadian Science Advisory Secretariat, Ottawa, Canada.
Palsbøll PJ (1999) Genetic tagging: contemporary molecular ecology. *Biological Journal of the Linnean Society*, **68**, 3–22.
Piggott MP (2005) Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research*, **31**, 485–493.
Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.
Ringler E, Ringler M, Jehle R, Hödl W (2012) The female perspective of mating in *A. femoralis*, a territorial frog with paternal care – a spatial and genetic analysis. *PLoS ONE*, **7**, e40237.
Ringler E, Mangione R, Ringler M (2014) Where have all the tadpoles gone? Individual genetic tracking of amphibian larvae until adulthood *Molecular Ecology Resources*, **15**, 737–746.
Rothstein AP (2015) *Non-invasive genetic tracking of Harbor Seals (Phoca vitulina)*. M.Sc., Thesis, Western Washington University, Bellingham, Washington.
Schwartz MK, Monfort SL (2008) Genetic and endocrine tools for carnivore surveys. In: *Noninvasive Survey Methods for Carnivores*(eds Long RA, MacKay P, Ray JC, Zielinski WJ), pp. 238–250. Island Press, Washington, District of Columbia.
Schwartz MK, Cushman SA, McKelvey KS, Hayden J, Engkjer C (2006) Detecting genotyping errors and describing American black bear movement in northern Idaho. *Ursus*, **17**, 138–148.
Solberg KH, Bellemain E, Drageset O-M, Taberlet P, Swenson JE (2006) An evaluation of field and non-invasive genetic methods to estimate brown bear (Ursus arctos) population size. *Biological Conservation*, **128**, 158–168.
Speed CW, Meekan MG, Bradshaw CJ (2007) Spot the match–wildlife photo-identification using information theory. *Frontiers in Zoology*, **4**, 1–11.
Stevick PT, Palsbøll PJ, Smith TD, Bravington MV, Hammond PS (2001) Errors in identification using natural markings: rates, sources, and effects on capture recapture estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1861–1870.
Suryan R, Harvey J (1999) Variability in reactions of Pacific harbor seals, Phoca vitulina richardsi, to disturbance. *Fishery Bulletin*, **97**, 332–339.
Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, **68**, 41–55.
Thompson PM (1989) Seasonal changes in the distribution and composition of common seal (Phoca vitulina) haul-out groups. *Journal of Zoology*, **217**, 281–294.
Valière N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, **2**, 377–379.
Valière N, Berthier P, Mouchiroud D, Pontier D (2002) GEMINI: software for testing the effects of genotyping errors and multitubes approach for individual identification. *Molecular Ecology Notes*, **2**, 83–86.
Vindenes Y, Engen S, Sæther BE (2008) Individual heterogeneity in vital parameters and demographic stochasticity. *The American Naturalist*, **171**, 455–467.
Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, **69**, 1419–1433.
Walker KA, Trites AW, Haulena M, Weary DM (2012) A review of the effects of different marking and tagging techniques on marine mammals. *Wildlife Research*, **39**, 15–30.
Woodruff SP, Johnson TR, Waits LP (2014) Evaluating the interaction of faecal pellet deposition rates and DNA degradation rates to optimize sampling design for DNA-based mark–recapture analysis of Sonoran pronghorn. *Molecular Ecology Resources*, **15**, 843–854.
Yochem PK, Stewart BS, DeLong RL, DeMaster DP (1987) Diel haul-out patterns and site fidelity of Harbor Seals (Phoca vitulina richardsi) on San Miguel Island, California, in autumn. *Marine Mammal Science*, **3**, 323–332.

A.P.R designed the project, analysed data and prepared and wrote the manuscript. R.M. wrote scripts for WISEPAIR suite and analysed data. A.A.G and D.S. conceived the project, assisted in analyses and edited and revised the manuscript.

Data accessibility

Program, data sets, supplement documents and README are available at: https://github.com/McGlock/WisePair.

Supporting Information

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

Table S1 Table of resampled individuals from Lampa *et al.* (2015). 65 individuals are matched as TRUE with 5 false positive matches as FALSE.