The Journal of Wildlife Management 76(1):88–94; 2012; DOI: 10.1002/jwmg.241

Management and Conservation

Cost-Efficient Selection of a Marker Panel in Genetic Studies

JAMIE S. SANDERLIN,[1,2] *Department of Statistics and Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA*

NICOLE LAZAR, *Department of Statistics, University of Georgia, Athens, GA 30602, USA*

MICHAEL J. CONROY, *Warnell School of Forestry and Natural Resources, Athens, GA 30602, USA*

JAXK REEVES, *Department of Statistics, University of Georgia, Athens, GA 30602, USA*

ABSTRACT Genetic techniques are frequently used to sample and monitor wildlife populations. The goal of these studies is to maximize the ability to distinguish individuals for various genetic inference applications, a process which is often complicated by genotyping error. However, wildlife studies usually have fixed budgets, which limit the number of genetic markers available for inclusion in a study marker panel. Prior to our study, a formal algorithm for selecting a marker panel that included genotyping error, laboratory costs, and ability to distinguish individuals did not exist. We developed a constrained nonlinear programming optimization algorithm to determine the optimal number of markers for a marker panel, initially applied to a pilot study designed to estimate black bear abundance in central Georgia. We extend the algorithm to other genetic applications (e.g., parentage or population assignment) and incorporate possible null alleles. Our algorithm can be used in wildlife pilot studies to assess the feasibility of genetic sampling for multiple genetic inference applications. © 2011 The Wildlife Society.

KEY WORDS abundance estimation, black bear, cost-efficient, genotyping error, Georgia, optimal marker panel, parentage, population assignment, *Ursus americanus*.

Molecular methods using genetic markers (e.g., microsatellites, single nucleotide polymorphisms [SNPs]) are important tools for wildlife managers in the conservation and management of populations (Selkoe and Toonen 2006, Schwartz et al. 2007). For any molecular study, selection of a marker panel is vital as the first step. Important components for marker panel selection include budgetary constraints, genotyping error, and the ability to distinguish individuals. Marker panel selection is especially important in noninvasive studies. Noninvasive genetic samples (i.e., shed hairs, feathers, feces) are often limited and contain degraded DNA. Some marker panel selection algorithms exist but they do not incorporate all ingredients for an optimal marker panel under budgetary constraints. Current practice is to use all available markers for a study species or a subset of the most informative markers (Smouse and Chevillon 1998, Paetkau 2003, Waits and Paetkau 2005). However, some loci are inherently more informative than others, and an objective means is needed to determine the most informative ones to meet project objectives, subject to time and cost constraints.

Population genotype frequencies are the core metrics used to determine how informative markers would be for abundance, parentage, and population assignment applications. Application-specific formulae all contain genotype frequencies as parameters, but differ in how they are combined based on the application. Probability of identity (PID), the probability that 2 randomly chosen individuals in a population will have identical genotypes (Paetkau and Strobeck 1994), is used in population abundance applications. Similar to PID, parentage studies use the probability of exclusion (P_{ex}; Jamieson 1979, 1994; Jamieson and Taylor 1997) to determine how informative markers will be based on the ability to distinguish among false pedigrees. The ability to correctly allocate individuals to populations for population assignment also relies on genotype frequencies with genetic distance between populations (D) and several distance measures exist (e.g., Nei 1972, Smouse and Chevillon 1998).

A marker panel that minimizes PID or P_{ex} is an important step; however, minimizing genotyping error would also improve data quality and integrity. Molecular methods can be costly in time and monetary resources and be prone to genotyping error. Genotyping error from allelic dropout (i.e., 1 or both of the 2 alleles are not amplified during the analysis) and/or false alleles (i.e., misprinting or the addition of an allele) may lead to misclassification of individuals within a population and lead to errors with other types of genetic inference. Regardless of the genetic inference problem, reduction of genotyping error is important for marker panel selection. General guidelines to reduce genotyping error for marker panel selection in genetic studies exist for population abundance applications (e.g., Waits and Paetkau

Received: 25 May 2010; Accepted: 17 April 2011;
Published: 16 September 2011

[1]*E-mail: jamie.sanderlin@gmail.com*
[2]*Present Address: United States Forest Service Rocky Mountain Research Station, 2500 S Pine Knoll Drive, Flagstaff, AZ 86001, USA.*

2005, Selkoe and Toonen 2006), but formal techniques or algorithms for choosing a marker set do not. For population assignment and parentage applications, algorithms for marker panel selection exist (e.g., Bromaghin 2008, Matson et al. 2008), but they do not incorporate cost constraints or genotyping error. Unfortunately, findings on the optimal number of markers for a panel subject to genotyping error have been contradictory. Suggestions range from using few highly polymorphic loci with low PID (Waits and Leberg 2000, Creel et al. 2003) to not using these heterozygous loci with more alleles due to more stutter bands (i.e., minor bands that usually differ from major bands by 2 nucleotides; Hoffman and Amos 2005).

To improve upon current methods, we present an algorithm that incorporates costs and genotyping error for marker panel selection for population assignment, parentage, and population abundance applications. Our original motivation for this problem came from selecting a marker panel for the central Georgia black bear (*Ursus americanus*) population using a pilot study to assess genotyping error (Sanderlin 2009*a*) for abundance estimation (Sanderlin 2009*b*). Our initial objective was to optimize the number of markers within a marker panel with minimal probability of identity and genotyping error at a fixed cost. We also recognized that other genetic inference problems, such as population or parentage assignment, could benefit from a formal optimization algorithm under these constraints. We discuss the algorithm for optimal marker panel selection using our initial objective for abundance estimation, and describe optimal marker panel algorithm modifications for other genetic inference applications.

METHODS

Optimization Algorithm

We cast the problem of marker allocation in a constrained nonlinear programming optimization framework. In general, constrained optimization has 3 main components: 1) decision variables, 2) an objective function, and 3) constraints (Taha 1976). Our decision variables (x_i, where $i = 1, \ldots, L$) are the identities of loci in a proposed marker panel (where x_i is binary with 1 indicating the locus is in the marker panel and 0 indicating the locus is not in the marker panel). The number of potential loci (L) from a marker panel will vary by species and population. The sum of x_i is the number of loci in the proposed marker panel. We present a general optimization algorithm that includes the following constraints: probability of identity among siblings (PID_{sib}; Evett and Weir 1998), probability of allelic dropout (ADO), probability of false alleles (FA), and cost ($C^{(a)}$).

For population abundance applications, a minimum threshold for number of loci is selected to reduce the shadow effect (Mills et al. 2000). The shadow effect occurs when 2 or more individuals are identified as 1 individual because they have identical genetic tags, often because too few loci or loci with low heterozygosity are used. We chose PID_{sib} as a more conservative metric for this constraint (Evett and Weir 1998).

We were interested in optimizing a marker panel with expected mean probabilities of genotyping error for the entire marker panel, instead of each locus individually. Genotyping error results from different processes depending on the genotyping error type. Allelic dropout is often caused by laboratory sampling stochasticity (random laboratory sampling of fragmented DNA within an individual sample) and/or amplification of small amounts of DNA (Goossens et al. 1998, Taberlet et al. 1999). False alleles often occur with polymerase chain reaction (PCR) amplification artifacts from dinucleotide microsatellites (Goossens et al. 1998, Taberlet et al. 1999) or with sample contamination. We used mean probability of allelic dropout (MDO) and mean probability of false alleles (MFA) to capture the mean probability of genotyping error across the whole marker panel.

A simplified cost function for the cost component ($C^{(a)}$) would include a fixed overhead cost (C_0), and an additional per locus cost (C_1) for each locus in the marker panel:

$$C^{(a)} = C_0 + C_1 \sum_{i=1}^{L} x_i.$$

However, if loci can be grouped into multiplexes (i.e., samples can be genotyped with multiple loci simultaneously using different fluorescent labels and marker size ranges), a more descriptive cost function is needed. The cost function for $C^{(b)}$ would now include the number of multiplexes with at least 1 locus selected M:

$$C^{(b)} = C_0 + C_1 \times M,$$

where there is a fixed overhead cost (C_0) and an additional multiplex cost (C_1) for loci in the marker panel grouped together in multiplexes. The identities of loci within these multiplexes would be study-, species-, and population-specific.

Our initial constraints included a maximum overall probability of identity among siblings (PID_{sib}), maximum allowable mean estimates of both types of error (MDO and MFA), and maximum number of loci based on a fixed cost for the genetic analysis. We summarized our objective function as follows:

$$\text{Minimize } C^{(a)} = C_0 + C_1 \sum_{i=1}^{L} x_i,$$

subject to:

$$\text{PID}_{\text{sib}} = \prod_{i=1}^{L} (\text{PID}_{\text{sib},i} \times x_i + 1 - x_i) \leq f,$$

$$\text{MDO} = \frac{\sum_{i=1}^{L} (\text{ADO}_{\text{median},i} \times x_i)}{\sum_{i=1}^{L} x_i} \leq g,$$

and

$$\text{MFA} = \frac{\sum_{i=1}^{L} (\text{FA}_{\text{median},i} \times x_i)}{\sum_{i=1}^{L} x_i} \leq h,$$

where L was the total number of loci in the marker panel, and $\text{ADO}_{\text{median},i}$ and $\text{FA}_{\text{median},i}$ were the posterior median values of allelic dropout and false allele probabilities at locus i, respectively. We specified the user-defined constraints on PID_{sib}, MDO, and MFA in general terms of f, g, and h, respectively, as upper bounds. We calculated values for PID_{sib} from allele frequencies of collected samples and used Bayesian posterior estimates of genotyping error, but any estimates of locus-specific error probabilities would be acceptable for other studies.

We used the objective function to find the optimal solution to the model, which we obtained when the "corresponding values of the decision variables yield the best value of the objective function while satisfying all the constraints" (Taha 1976:6). We accomplished minimization graphically, instead of analytically due to nonlinearity in some of the constraints. We plotted PID_{sib} and each type of genotyping error separately using different cost symbols in each plot based on the number of loci in each potential marker panel (e.g., Fig. 1 for

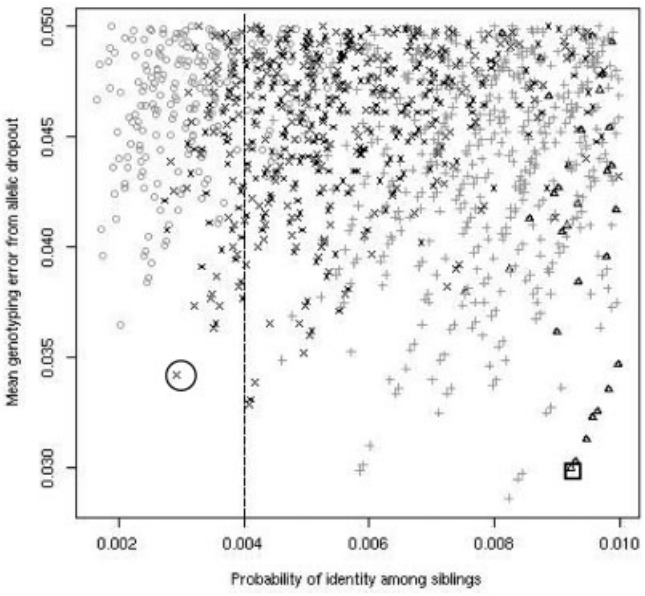

Figure 1. Example from the central Georgia American black bear population pilot study from 2003 to 2006 (Sanderlin 2009*a*) of graphically selecting an optimal marker panel. The overall probability of identity among siblings, PID_{sib}, and mean genotyping error from allelic dropout, ADO, based on median posteriors for all possible marker panels are graphically represented using symbols for the number of loci in each panel of: 7 (△), 8 (+), 9 (×), or 10 (○). We did not present the identities of all panel loci graphically, but they are available from the authors in table format. The optimal marker panel is the panel with minimized cost (i.e., smallest number of loci) given specified constraints. All marker panels with 9 loci to the left of the dashed line satisfy the objective function constraints when $\text{PID}_{\text{sib}} \leq 0.004$. We further optimized these 9 loci marker panels by selecting the panel (circled) within this subset with minimum mean genotyping error from ADO. All marker panels with 7 loci in this graph satisfy the objective function constraints when $\text{PID}_{\text{sib}} \leq 0.01$. We further optimized these 7 loci marker panels by selecting the panel (boxed) within this subset with minimum mean genotyping error from ADO.

PID_{sib}, MDO, and number of loci). Both types of genotyping error (MDO and MFA), PID_{sib}, and number of loci can be plotted simultaneously using a 3-dimensional plot. However, for simplicity, we present the constraints with only 1 type of genotyping error (MDO).

After we selected an optimal marker panel, we tested for genotypic linkage disequilibrium (LD) and Hardy–Weinberg equilibrium (HWE) among loci in the marker panels. Linkage disequilibrium (gametic phase imbalance) occurs when alleles at 2 or more distinctive loci appear in gametes more frequently than expected. Evidence of genotypic linkage disequilibrium between pairs of loci violates assumptions of independence among loci and is not optimal in a marker panel, thus only panels without evidence of LD between all loci pairs were considered (i.e., $\text{LD}_{\text{panel}} = 0$ if there was no statistical evidence of genotypic LD for all loci pairs in the panel after Bonferroni correction and 1 otherwise). We do not present a constraint that includes linked loci because genetic inference on probability of identity, average pairwise genetic distance between populations, and probability of exclusion values would be biased as these quantities do not adjust for nonindependent loci.

Extensions to the Optimization Algorithm

Violations of HWE, most often from heterozygote deficiency, could indicate nonrandom mating, selection, limited population size, random genetic drift, or mutations in the population (Hartl 2000). Heterozygote deficiency may be caused by null alleles (i.e., alleles that do not amplify in the PCR process because of mutations in the flanking regions of primers). Null alleles may affect genetic metrics, such as reduction of within-population genetic diversity (e.g., Paetkau and Strobeck 1995), and lead to overestimation of the proportion of genetic variance in the total population due to variation among subpopulations (F_{st}) and genetic distance (Chapuis and Estoup 2007). Null allele frequency estimators ($\hat{r}$) can be used to assess the presence of null alleles (Chakraborty et al. 1992, Brookfield 1996, vanOosterhout et al. 2004, Kalinowski and Taper 2006). If null alleles are suspected in the study population, we suggest either to: 1) add an additional constraint to the system where only marker panel subsets without null alleles (i.e., $null_i = 0$ if there were no null alleles at locus i and 1 otherwise) are considered:

$$\sum_{i=1}^{L} (null_i \times x_i) = 0,$$

or 2) minimize the loci with null alleles, using a predetermined average null allele frequency (y):

$$\frac{\sum_{i=1}^{L} (\hat{r}_i \times x_i)}{\sum_{i=1}^{L} x_i} \leq y$$

and adjust the genetic metrics (i.e., allele and genotype frequencies) accordingly (e.g., Roques et al. 1999).

Depending on the particular genetic inference problem, additional constraints could include: D (Nei's standard

genetic distance between populations; Nei 1972) for population assignment applications and P_{ex} (probability of exclusion; Jamieson 1979, 1994; Jamieson and Taylor 1997) for parentage assignment applications. For parentage applications, the additional restraint maximizes P_{ex} over multiple loci at a level above the predetermined probability of exclusion z:

$$1-\prod_{i=1}^{L}(1-(P_{\mathrm{ex},i}\times x_i))\geq z.$$

For population assignment, multiple genetic distance methods exist. We use D (Nei 1972) to illustrate how to use an additional metric and restriction to the system. The metric D is defined as:

$$D=-\log_e I,$$

where I is the normalized probabilities of identity between populations Y and Z for all loci (e.g., $I=J_{YZ}/\sqrt{J_Y J_Z}$) and J_{YZ}, J_Y, and J_Z are the arithmetic means over all loci using j_{YZ} ($j_{YZ}=\sum y_i z_i$), j_Y ($j_Y=\sum y_i^2$), and j_Z ($j_Z=\sum z_i^2$) computed from y_i and z_i frequencies of ith alleles in populations Y and Z. The additional restraint maximizes D at a predetermined level of genetic distance w over all loci in the marker panel $D\geq w$, where:

$$J_{YZ}=\frac{\sum_{i=1}^{L}\left(x_i\left(\sum_{k=1}^{\max(k_i)}y_k z_k\right)\right)}{\sum_{i=1}^{L}x_i},$$

$$J_Y=\frac{\sum_{i=1}^{L}\left(x_i\left(\sum_{k=1}^{\max(k_i)}y_k^2\right)\right)}{\sum_{i=1}^{L}x_i},$$

$$J_Z=\frac{\sum_{i=1}^{L}\left(x_i\left(\sum_{k=1}^{\max(k_i)}z_k^2\right)\right)}{\sum_{i=1}^{L}x_i},$$

and $\max(k_i)$ is the maximum number of alleles k at locus i. We suggest solving the system with alternate genetic applications (parentage and population assignment) graphically, instead of analytically, due to nonlinearity in those additional constraints.

Case Study

We applied the optimization algorithm to select the optimal number of markers and a marker panel set to estimate black bear abundance in central Georgia, USA (Sanderlin 2009*a*). We captured and immobilized bears (University of Georgia Institutional Animal Care and Use Committee approval numbers: A2003-10148, A2003-10148-ml) in trapping seasons extending May through August (2003–2006) on Ocmulgee and Oaky Woods Wildlife Management Areas (Bleckley, Bibb, Houston, Pulaski, and Twigg Counties). We used blood, tissue, and hair samples from captured bears and some road and capture mortalities ($n=84$ bears) in an analysis to assess genotyping error (see Sanderlin 2009*a* for complete description of field and laboratory methods and data summary) using 16 tetranucleotide loci (Sanderlin et al. 2009). We removed 8 bear hair samples from the analysis since we classified them as bad samples (e.g., less than half of the loci positively amplified).

We used the multi-locus genotypes from bear tissue samples to calculate allele frequencies and observed and expected heterozygosities at each locus with Cervus 2.0 (Marshall et al. 1998). We wrote program optimal-marker-panel (OMP) in Python (Python Software Foundation version 2.5.2, http://python.org, accessed 28 Feb 2009) to output a data table for graphical optimal marker panel evaluation (available at http://code.google.com/p/optimal-marker-panel/). In this program, we evaluate PID_{sib} using allele frequencies from tissue samples. The program also has the ability to evaluate P_{ex} and cost ($C^{(a)}$) or number of loci. We also included posterior median estimates of genotyping error using hair and tissue samples from a pilot study (Sanderlin 2009*a*). We used the following constraints on PID_{sib}, MDO, MFA, and number of loci: {0.01, 0.05, 0.01, 10} and {0.004, 0.05, 0.01, 10}. To evaluate the effect of not including genotyping error as a constraint, we only used constraints for PID_{sib} and number of loci: {0.01, 10} and {0.004, 10}. We used an alternate constraint on PID_{sib} (0.004) because it is more restrictive, and hence, more conservative with marker panel selection. After we selected an optimal marker panel, we used GENEPOP 3.4 (Raymond and Rousset 1995) to test for genotypic linkage disequilibrium and Hardy–Weinberg equilibrium using a posteriori sequential Bonferroni correction (Rice 1989) among loci in these 4 different sets of marker panels.

RESULTS

For the black bear central Georgia population case study, optimal marker panel sets with genotyping error constraints and without genotyping error constraints had the same number of loci when the PID_{sib} constraint was: 1) $\leq$0.01 (optimal solution was 7 loci), and 2) $\leq$0.004 (optimal solution was 9 loci; Figs. 1 and 2). Moreover, the marker panels were identical except for 1 out of 7 loci with $PID_{sib}\leq 0.01$, and 2 out of 9 loci when $PID_{sib}\leq 0.004$. Marker identities for these optimal panels are listed in Sanderlin (2009*a*).

DISCUSSION

Optimal selection of a marker panel ultimately depends on both project goals and amount of time and money available. Our techniques in this study provide formal procedures for choosing a marker panel set for estimating population abundance, using restrictions of cost, genotyping error, and ability to distinguish among individuals. We also discuss application-specific modifications to the algorithm and how to adjust the algorithm when null alleles are suspected.

Contrary to Waits and Leberg (2000) and Creel et al. (2003), we did not always observe higher mean expected genotyping error with more loci in marker panels (Fig. 1; Sanderlin 2009*a*). Depending on the specific loci in each

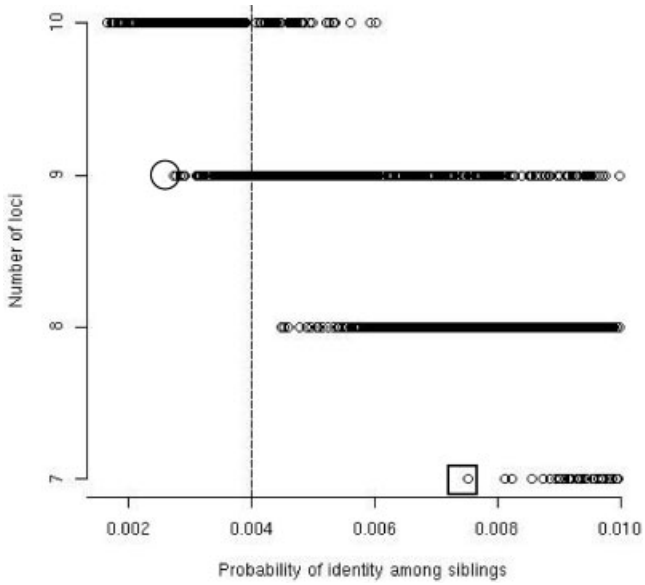

Figure 2. Optimal marker panel for the central Georgia American black bear population pilot study from 2003 to 2006 (Sanderlin 2009*a*) using constraints of probability of identity among siblings, PID_{sib}, and number of loci, but not genotyping error. We graphically represent all marker panels using the symbol (◯). We do not present the identities of all panel loci graphically, but they are available from the authors in table format. The optimal number of loci was 7 when $PID_{sib} \leq 0.01$ and 9 when $PID_{sib} \leq 0.004$. The optimal marker panel is the panel with minimized cost (i.e., smallest number of loci) given specified constraints. The circled point represents the optimal marker panel for minimum PID_{sib} when $PID_{sib} \leq 0.004$. The point within the box represents the optimal marker panel for minimum PID_{sib} when $PID_{sib} \leq 0.01$.

panel, some potential panels had lower expected mean genotyping error than panels with fewer loci. This has implications in genetic marker panel selection, particularly with noninvasive genetic studies. Although the optimal number of loci may be identical with or without genotyping error, panel composition may differ slightly.

Alternative laboratory, field sampling, and/or analytical and model estimation approaches can be used to reduce costs in noninvasive studies. Cost-effective laboratory procedures include optimizing the number of PCR replicates used in genetic analyses (e.g., Frantz et al. 2003). The cost of field sampling methods can be reduced by selecting a subsample of all field-collected DNA for genetic analysis (e.g., Tredick et al. 2007) or using optimal field sampling designs (e.g., Williams et al. 2002, Field et al. 2005). We are unaware of optimal genetic sampling designs that combine both field and laboratory costs. These have wide potential as genetic monitoring methods become more prevalent. Costs of analytical methods can be reduced by testing if a sample contains genotyping errors (e.g., McKelvey and Schwartz 2004) and using model estimation approaches for inference parameters, like population size (e.g., Knapp et al. 2009, Wright et al. 2009). Knapp et al. (2009) and Wright et al. (2009) both incorporate genotyping error into statistical models for estimating population abundance. These approaches have a direct influence on laboratory procedures by reducing the number of amplifications needed per genotype (1 for Knapp et al. 2009, 2 for Wright et al. 2009), instead of multiple amplifications (e.g., Taberlet et al. 1996). Both methods used a fixed number of loci, used all study samples, and reduced costs with the number of amplifications. They did not include explicit algorithms for marker panel selection before all study samples were genotyped. Substantial effort may be saved by choosing an optimal marker panel after a pilot study using our algorithm, and then incorporating genotyping error into population estimates with all samples using methods described in Wright et al. (2009).

Our algorithm relies on selecting a random sample representative of the population for the pilot study. It is ideal to have known parent–offspring pairs in this sample (Paetkau 2003, Selkoe and Toonen 2006), so the genetic inference quantities reflect the ability to distinguish between related individuals. As individual relatedness, population size, or degree of isolation is unknown, it may be difficult to choose an optimal number of markers (Paetkau 2004, Waits and Paetkau 2005). In addition, acceptable PID values are dependent on how many individuals may be sampled (Waits and Paetkau 2005), which is also likely true for probability of exclusion and pairwise genetic distance. The ability to detect null alleles in a population is also dependent on the pilot sample size. Given the above information, further work and simulation trials are needed for determining pilot sample size. Conversely, estimates of allelic dropout and false alleles should be robust to the number of samples or the proportion of the population in the pilot study. Although not the focus of this study, our Bayesian estimation algorithm for allelic dropout and false alleles had high Bayesian credible interval coverage (near nominal 0.95 probability) with simulation trials for multiple sample sizes (25, 50, 75) and proportions of the population (0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.50, 0.75) in the pilot study (Sanderlin 2009*a*).

Some restrictions to our system could be considered subjective, as there are few formal guidelines for our constraints. Depending on research goals, constraints within our algorithm can be modified accordingly to be more stringent or less conservative. For example, Lukacs and Burnham (2005) suggest keeping levels of genotyping error at less than 5% for population abundance studies. Project goals could also influence weighting 1 measure more heavily than another. Additional time and laboratory costs could also be included in the cost function. Further improvements to our algorithm could include a simulation component after optimal marker panel selection to assess predictive power (e.g., Banks et al. 2003). Although we present an algorithm for selecting a marker panel given genotyping error and probability of identity at fixed budgets, the objective can be changed to minimize probability of identity or genotyping error, subject to budget and other constraints, depending on study goals.

MANAGEMENT IMPLICATIONS

Optimal selection of marker panels will improve both data quality and integrity within wildlife management and conservation, especially given budget constraints. Our techniques provide formal optimization procedures and possible extensions (i.e., combining field and laboratory costs).

Resources misallocated to suboptimal marker designs could be more effectively spent on other aspects of study design, such as improving spatial replication. We recommend researchers carefully consider and implement constraints for optimal marker panels during the pilot study stage of genetic studies.

ACKNOWLEDGMENTS

This research was supported by: Georgia Department of Natural Resources (cooperative agreement number: 1434-02HQRU1551); International Bear Association Research & Conservation grant and the John Sheldon Bevins Memorial Foundation; Georgia Cooperative Fish and Wildlife Research Unit, jointly sponsored by U.S. Geological Survey, U.S. Fish and Wildlife Service, University of Georgia, Georgia Department of Natural Resources, and the Wildlife Management Institute; Daniel B. Warnell School of Forestry and Natural Resources; and McIntire-Stennis project GEO-0100-MS and GEO-0136-MS. This work was performed with the support of the Georgia Genomics Facility at the University of Georgia. C. J. Nairn, J. P. Carroll, and B. Faircloth contributed to efforts in the genetics laboratory. Use of trade names does not imply U.S. Government endorsement of commercial products. We also thank E. Latch and 2 anonymous reviewers for comments on earlier manuscript versions.

LITERATURE CITED

Banks, M. A., W. Eichert, and J. B. Olsen. 2003. Which genetic loci have greater population assignment power? Bioinformatics 19:1436–1438.

Bromaghin, J. F. 2008. BELS: backward elimination locus selection for studies of mixture composition or individual assignment. Molecular Ecology Resources 8:568–571.

Brookfield, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 5: 453–455.

Chakraborty, R., M. De Andrade, S. P. Daiger, and B. Budowle. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Annals of Human Genetics 56:45–57.

Chapuis, M.-P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24:621–631.

Creel, S., G. Spong, J. L. Sands, J. Rotella, J. Zeigle, L. Joe, K. M. Murphy, and D. Smith. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Molecular Ecology 12:2003–2009.

Evett, I. W., and B. S. Weir. 1998. Interpreting DNA evidence: Statistical genetics for forensic scientists. Sinauer, Sunderland, Massachusetts, USA.

Field, S. A., A. J. Tyre, and H. P. Possingham. 2005. Optimizing allocation of monitoring effort under economic and observational constraints. Journal of Wildlife Management 69:473–482.

Frantz, A. C., L. C. Pope, P. J. Carpenter, T. J. Roper, G. J. Wilson, R. J. Delahay, and T. Burke. 2003. Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. Molecular Ecology 12:1649–1661.

Goossens, B., L. P. Waits, and P. Taberlet. 1998. Plucked hair samples as a source of DNA: reliability of dinuccleotide microsatellite genotyping. Molecular Ecology 7:1237–1241.

Hartl, D. L. 2000. A primer of population genetics. Sinauer Associates, Inc, Sunderland, Massachusetts, USA.

Hoffman, J. I., and W. Amos. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology 14:599–612.

Jamieson, A. 1979. Electromorphs and erroneous pedigrees. Page 27 *in* Proceedings of the XVIth International Conference on Animal Blood Groups and Biochemical Polymorphism. The National Committee of the USSR, 1978, Leningrad, Russia.

Jamieson, A. 1994. The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. Animal Genetics 25:37–44.

Jamieson, A., and St. C. S. Taylor. 1997. Comparisons of three probability formulae for parentage exclusion. Animal Genetics 28:397–400.

Kalinowski, S. T., and M. L. Taper. 2006. Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. Conservation Genetics 7:991–995.

Knapp, S. M., B. A. Craig, and L. P. Waits. 2009. Incorporating genotyping error into non-invasive DNA-based mark-recapture population estimates. Journal of Wildlife Management 73:598–604.

Lukacs, P. M., and K. P. Burnham. 2005. Estimating population size from DNA-based closed capture-recapture data incorporating genotyping error. Journal of Wildlife Management 69:396–403.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639–655.

Matson, S. E., M. D. Camara, W. Eichert, and M. A. Banks. 2008. P-LOCI: a computer program for choosing the most efficient set of loci for parentage assignment. Molecular Ecology Resources 8:765–768.

McKelvey, K. S., and M. K. Schwartz. 2004. Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. Journal of Wildlife Management 68:439–448.

Mills, L. S., J. J. Citta, K. P. Lair, M. K. Schwartz, and D. A. Tallmon. 2000. Estimating animal abundance using noninvasive DNA sampling: promises and pitfalls. Ecological Applications 10:283–294.

Nei, M. 1972. Genetic distance between populations. The American Naturalist 106:283–292.

Paetkau, D., and C. Strobeck. 1994. Microsatellite analysis of genetic variation in black bear populations. Molecular Ecology 3:489–495.

Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. Molecular Ecology 4:519–520.

Paetkau, D. 2003. An empirical exploration of data quality in DNA-based population inventories. Molecular Ecology 12:1375–1387.

Paetkau, D. 2004. The optimal number of markers in genetic capture-mark-recapture studies. Journal of Wildlife Management 68:449–452.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2.): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248–249.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

Roques, S., P. Duchesne, and L. Bernatchez. 1999. Potential of microsatellites for individual assignment: the North Atlantic redfish (genus *Sebastes*) species complex as a case study. Molecular Ecology 8:1703–1718.

Sanderlin, J. S. 2009*a*. Misidentification error in non-invasive genetic mark-recapture sampling: case study with the central Georgia black bear population. Thesis, University of Georgia, Athens, USA.

Sanderlin, J. S. 2009*b*. Integrated demographic modeling and estimation of the central Georgia, USA, black bear population. Dissertation, University of Georgia, Athens, USA.

Sanderlin, J. S., B. C. Faircloth, B. Shamblin, and M. J. Conroy. 2009. Tetranucleotide microsatellite loci from the black bear (*Ursus americanus*). Molecular Ecology Resources 9:288–291.

Schwartz, M. K., G. Luikart, and R. S. Waples. 2007. Genetic monitoring as a promising tool for conservation and management. Trends in Ecology and Evolution 22:25–33.

Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615–629.

Smouse, P. E., and C. Chevillon. 1998. Analytical aspects of population-specific DNA fingerprinting for individuals. Journal of Heredity 89:143–150.

Taberlet, P., S. Griffin, B. Goossens, S. Questiau, V. Manceau, N. Escaravage, L. P. Waits, and J. Bouvet. 1996. Reliable regenotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research 24:3189–3194.

Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: look before you leap. Trends in Ecology and Evolution 14: 323–327.

Taha, H. A. 1976. Operations research, an introduction. Second edition. Macmillan Publishing Co., Inc, New York, New York, USA.

Tredick, C. A., M. R. Vaughan, D. F. Stauffer, S. L. Simek, and T. Eason. 2007. Sub-sampling genetic data to estimate black bear population size: a case study. Ursus 18:179–188.

vanOosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535–538.

Waits, J. L., and P. L. Leberg. 2000. Biases associated with population estimation using molecular tagging. Animal Conservation 3:191–199.

Waits, L. P., and D. Paetkau. 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.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, New York, New York, USA.

Wright, J. A., R. J. Barker, M. R. Schofield, A. C. Frantz, A. E. Byrom, and D. M. Gleeson. 2009. Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. Biometrics 65:833–840.

Associate Editor: Emily Latch.