Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

**Canadian Science Advisory Secretariat (CSAS)**

**Research Document 2016/031**

**Gulf Region**

# Molecular techniques for parentage analysis to assess supplementation effectiveness for Atlantic Salmon (*Salmo salar*) on the Miramichi River

Scott A. Pavey

Canada Research Chair in Aquatic Molecular Ecology and Ecological Genomics
Department of Biological Sciences & Canadian Rivers Institute
100 Tucker Park Road
University of New Brunswick Saint John
Saint John, New Brunswick, Canada, E2L 4L5

May 2016

Canada

**Foreword**

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

**Published by:**

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca


ISSN 1919-5044

**Correct citation for this publication:**

Pavey, S.A. 2016. Molecular techniques for parentage analysis to assess supplementation effectiveness for Atlantic Salmon (*Salmo salar*) on the Miramichi River. DFO Can. Sci. Advis. Sec. Res. Doc. 2016/031. v + 8 p.

## TABLE OF CONTENTS

## ABSTRACT

The current supplementation program for Atlantic Salmon (*Salmo salar*) on the Miramachi River involves capturing mature adults, fertilizing eggs, and raising young in a hatchery before releasing them back into the wild. A pilot study for a different strategy may occur on the Northwest Miramichi, where juveniles are collected, raised to mature adults in a hatchery, and released into the river to naturally spawn. Here I perform a literature review and make recommendations to monitor the effectiveness of this pilot project. The proposed technique is to use genetic markers for parentage analysis, so that both wild and supplemented adults can be compared for their contribution to the next generation. Current high-throughput genotyping techniques of both single nucleotide polymorphisms (SNPs) and microsatellites are compared with simulations. While all three techniques performed adequately, the most flexible and promising is genotyping-in-thousands (GT-Seq). Sampling design recommendations are made, such as genetically sampling an equal amount of wild and supplemented individuals.

## Techniques moléculaires pour l'analyse de parenté afin d'évaluer l'efficacité de l'ensemencement pour le saumon de l'Atlantique (*Salmo salar*) dans la rivière Miramichi


## RÉSUMÉ

Le programme d'ensemencement actuel pour le saumon de l'Atlantique (*Salmo salar*) dans la rivière Miramachi consiste à capturer des adultes matures, à faire féconder des œufs et à élever les jeunes dans une écloserie avant de les relâcher dans la nature. Une étude pilote d'une stratégie différente pourrait être réalisée dans la rivière Miramichi Nord-Ouest, consistant à capturer les juvéniles, à les élever en écloserie jusqu'à maturité, puis à les relâcher dans la rivière pour qu'ils y fraient naturellement. Ici, j'effectue une analyse documentaire et je formule des recommandations afin de surveiller l'efficacité de ce projet pilote. La technique proposée consiste à utiliser des marqueurs génétiques pour l'analyse de parenté, de façon à pouvoir comparer la contribution des populations sauvages et des adultes ensemencés à la prochaine génération. Les techniques actuelles de génotypage à haut débit des microsatellites et du polymorphisme touchant un nucléotide unique (SNP) sont comparées à l'aide de simulations. Même lorsque les trois techniques sont effectuées adéquatement, la technique la plus souple et prometteuse est le génotypage par séquençage (GT-seq). Des recommandations sur la conception de l'échantillonnage sont fournies afin que l'échantillonnage génétique touche un nombre égal d'individus sauvages et ensemencés.

## INTRODUCTION

Hatchery interventions have been used to address declines in abundance of fish and to mitigate high risks of population extinction (Fraser 2016). This usually involves the collection of mature adults from the wild, stripping of eggs and sperm, artificial fertilization and incubation of fertilized eggs and raising progeny in a hatchery setting for release back into the wild at a later life stage. This intervention circumvents periods of typically high juvenile mortality that might occur in the wild.

However, artificially reared individuals may have a different survival rate than wild individuals (Milot et al. 2013; Fraser 2016). Learned behaviors and unintended selection in the hatchery may increase mortality when released (i.e higher predation, poor feeding efficiency). Thus any stock supplementation program must evaluate the fitness of released individuals compared with their wild counterparts (Fraser 2016). Without such information it is impossible to adjust the program to improve results, or evaluate cost effectiveness.

Parentage analysis using genetic markers makes such an evaluation possible. Through parentage analysis, offspring can be attributed to specific parents. Thus number of surviving progeny from all adults used in an enhancement program can be compared to number of surviving progeny from a wild sample of adults (Milot et al. 2013).

A new stock supplementation program is being proposed for Atlantic Salmon (*Salmo salar*) in the Northwest Miramichi River. The current program consists of capturing mature adults, striping eggs and milt, incubating eggs and raising fry in a hatchery setting, and releasing juveniles into the river. The new program, referred to as smolt to adult supplementation (SAS) proposes to capture juveniles, raise them to mature adults and release them back into the river to spawn naturally. Much less is known of this supplementation strategy than the previous strategies, but it has potential for success because:

- it avoids the ocean feeding life stage where most mortality occurs;
- no wild spawning individuals are removed,
- released adults naturally spawn, allowing sexual selection to occur, and
- progeny from supplemented adults get the full experience (starting from egg fertilization) in the natural environment.

This document provides a review of the available molecular approaches for parentage analysis that would be most appropriate for this situation, considers some of the design requirements for a successful monitoring program, and discusses limitations of genetic technology, software and spatial scale of the intervention.

## PARENTAGE ANALYSIS

Parentage analysis could be used to estimate the fitness of supplemented individuals compared to their wild counterparts. Parentage analysis requires genetically screening pools of potential parents and progeny and making matches. There are five different statistical approaches for parentage analysis: exclusion, categorical allocation, fractional allocation, full probability parental analysis, and parental reconstruction. Below is a summary of these approaches, paraphrased from Jones et al (2010).

First, complete exclusion is the gold standard of parentage studies; every parentage study is implicitly striving towards this goal. Since all diploid organisms get one haploid genome from each parent, if a potential parent shares neither allele with an offspring, it can be excluded as a

possible parent. If this can be done with all parents except two, then the true parents have been found (one mother and one father). The greater the number of individuals involved in a study, the more exclusions will be necessary, and therefore the more difficult this technique will be. Also, all potential parents must be genotyped for this technique to be sufficient by itself.

The second approach is categorical allocation. The principal behind this approach is that even if all parents cannot be completely excluded, some are more likely than others to be the true parents. Potential parents can be ranked based on this likelihood, and if there is a big difference between the most and second most likely, the former is chosen as the true parent. The threshold for false positive and false negative can be adjusted. This is the most used approach in natural systems.

The third approach is fractional allocation. It is similar to categorical allocation, but a progeny could be attributed to more than one mother or father. This is a biologically impossible situation, but is sometimes statistically desirable, because the uncertainty is reflected in the fractional allocations. If you are 90% sure that father A is the true father, categorical allocation (above) would consider this to be the most likely father (and depending on thresholds) designate this individual as the father. Fractional allocation would distribute 90% of the fatherhood to father A, and distribute the remaining 10% to other possible fathers.

The fourth approach is full probability modelling. This approach is an extension of fractional allocation and allows more model complexity (e.g. unequal reproductive potential of dominant males) and the inclusion of prior information (e.g. progeny morphology similar or spatial proximate to potential parent are more likely). This is also an appropriate strategy for the current project.

Finally, the fifth approach is parental reconstruction. If progeny are known to be siblings, the parental genotypes can be reconstructed and compared to the pool of potential parents. This is not an appropriate strategy in natural settings were siblings are expected to mix around with other families, such as the current system.

## MOLECULAR APPROACHES

By far, most parentage studies to date have used microsatellites. Microsatellites are tandem repeats of usually 2-6 base pairs (Freeland et al. 2011). They are common throughout the genome in both intronic and intergenic regions. They can be highly polymorphic (>50 alleles per locus in some cases) which makes them an attractive candidate for parentage analysis. Potential disadvantages of microsatellites are they require labour intensive genotyping, there can be a high error rate of genotyping, and there is a chance of allelic dropout (one allele does not amplify due to a mutation in the primer sequence, thus a heterozygote with such an allele will be mistaken for a homozygote) (Hauser et al. 2011).

Recently, parentage analyses have used Single Nucleotide Polymorphisms (SNPs) (Hauser et al. 2011). SNPs are sequences of DNA that are variable at a single position which is the SNP. In practice, there a maximum of two alleles, thus they are nowhere near as polymorhpic as microsatellites. However this lack of polymorphism may be compensated by analyzing a large number of loci, as techniques are available that can genotype individuals at hundreds or thousands of SNPs. Additionally, the genotyping error rate is generally an order of magnitude lower than that of microsatellites. In side-by-side comparisons to date, SNPs have generally performed better than microsatellites in parentage analyses (Hauser et al. 2011; Sellars et al. 2014; Telfer et al. 2015; Weinman et al. 2015). The reason for this is primarily the typical error rate for microsatellites (2%) is much higher than for SNPs (0.5%).

New technologies and laboratory techniques have resulted in high throughput techniques for genotyping of microsatellites and SNPs. The development of Next-generation sequencing (NGS) has made powerful genotyping techniques readily available (Metzker 2010). The basic principal is that DNA from individuals are marked with sequence barcodes so that individuals can be pooled on a single NGS run and later separated based on these barcode markers. The most commonly used NGS technologies are Illumina HiSeq© and Illumina MiSeq©. HiSeq© produces approximately 100 to 150 million sequences that are 100 base pairs in length, whereas MiSeq© produces 10 to 20 million sequences, but is capable of reading longer base pair lengths (Glenn 2011). The other applicable NGS sequencing technology is Ion Torrent© by Life Technologies© (Glenn 2011). The current iteration of Ion Torrent© sequencers have lower throughput than Illumina© and suffer from higher error rate and non-random errors.

Another technology is Fluidigm© microfluidic arrays (Perkel 2008). It is used to perform a pre-determined set of assays. This facilitates high throughput Taqman© assays, which was previously capable of only low throughput, and can be used to genotype individuals for SNPs.

The proposed supplementation program for the Miramichi and the corresponding monitoring program required to assess the contributions of SAS fish to subsequent generations would involve genotyping thousands of individuals even as a pilot study and therefore requires very high-throughput techniques. Therefore only high throughput methods are discussed below.

For SNPs, two potential techniques exist. Genotyping-in-Thousands (GT-seq) (Campbell et al. 2015; Pavey 2015) provides very cost effective genotyping of thousands of individuals at hundreds of SNPs. Briefly, individuals are barcoded and sequenced in a large multiplex of pre-determined SNP primers. Sequencing occurs on an Illumina HiSeq© lane, and a publically available bioinformatic pipeline performs the genotyping. This technique is very flexible, if the number of SNPs wanted is reduced, more individuals can be placed on a lane. The technique could also be adapted to the Ion Torrent© NGS platform.

The second potential strategy is Fluidigm© microfluidic arrays. Each array can genotype 96 individuals at 96 SNPs. Access to the proprietary liquid handling platform is required to run the chips. Also, a panel of SNPs must be chosen, and a batch of chips manufactured. This strategy is less flexible as the number of SNPs and individuals per chip is less readily changeable.

A third technique has greatly streamlined microsatellite genotyping to the point where it can be considered to be high-throughput. The MEGASAT© pipeline (pers. comm. Paul Bentzen; manuscript in review) uses Illumina© technology for high throughput and automated digital genotyping of microsatellites. This technique is very similar to GT-seq except with microsatellites instead of SNPs, and less loci (dozens) are realistic (though each locus is more polymorphic).

## SIMULATIONS

I performed simulations to assess the power of these three candidate techniques in the current supplementation program evaluation. I used the application CERVUS© (Marshall et al. 1998) to perform the simulations. Using summary data on overall allele frequency data, CERVUS© constructs parent and progeny genotypes, then tests the assignment ability for each progeny. All simulations assessed the number of progeny that could be assigned to a single mother (fathers could be considered separately but would be exactly the same for the purposes of the simulation). The simulations assumed 5000 potential parents (similar to potential parents available in the NW Miramichi), or 2500 mothers. If 10% of these were genotyped wild mothers (250) and an additional 10% genotyped supplemented mothers, then 20% of all potential mothers were sampled (thus 80% remain unsampled potential parents). This amount was

chosen as a realistic proportion of all potential parents, including all supplemented individuals and a sample of the wild spawning individuals that could be sampled in the current project. A total of 10,000 progeny were simulated. A perfect technique would assign 2,000 of these to genotyped parents, as the sampled mothers represent 20% of all potential mothers. This number was not chosen as a recommended sample size for the Miramichi monitoring program, but rather for this simulation exercise to compare marker panels. Three genotyping scenarios were tested. First of all, it is important to note that each SNP effectively has only two alleles. In the case of a G/T SNP, there are three possible genotypes: G/G, G/T, and T/T. There are only two alleles, and the level of polymorphism is indicated by the minor allele frequency, or the frequency of whatever allele is less present. The major allele frequency is simply one minus the minor allele frequency. For GT-seq, individuals were genotyped at 400 SNPs that had a minor allele frequency of 0.3 (thus the major allele frequency was 0.7). This was chosen because it is reasonable to assume that a set of 400 SNPs with this level of polymorphism can be found. For the Fluidigm© SNP array, individuals were genotyped at 96 SNPs that had a minor allele frequency of 0.4 (thus the major allele frequency was 0.6). This is a higher level of variability that would be attainable for this smaller number of SNPs. For MEGASAT©, a microsatellite dataset consisting of 12 highly polymorphic microsatellites (O'Reilly et al. 1996; Paterson et al. 2004; King et al. 2005) from a panel developed by Patrick O'Reilly (DFO unpublished data) was used. The mean number of alleles for these 12 microsatellites was 33.58 and the mean heterozygosity was 0.909 (Saint John River population; P. O'Reilly; DFO unpublished data). In all cases, the genotyping error rate was set at 0.5%. This is standard for SNPs but low for microsatellites, but the MEGASAT© bioinformatic pipeline makes such a low error rate achievable for microsatellites.

Results are presented in Table 1. All three methods performed well and all would be possible approaches for the pilot study of the proposed Miramichi SAS supplementation. The GT-seq (400 SNP) and 12 highly polymorphic microsatellite methods outperformed the panel of 96 SNPs, the latter lacked the power to identify all of the true mothers (despite the nearly maximal heterozygosity). The GT-seq method with 400 SNPs is the most flexible and scalable to larger systems in the future. This simulation demonstrates that it can easily accommodate the spatial scale of the pilot study and is robust to a large proportion of unsampled wild adults.

Any marker panel should be chosen to maximize variability at each locus specific to the study. Since Atlantic Salmon has many genomic resources including a sequenced genome and many polymorphic SNPs described (Bourret et al. 2013; Davidson 2013; Bourret et al. 2014), it should be quite possible to find a panel of SNPs that will be a powerful parentage tool for the Miramichi River population. Error rate of the genotyping should be empirically estimated through technical replicates. Additional simulations could be run with the empirically estimated error rate and different sets of markers together with budgetary considerations to make a highly informed choice of the best marker panel and genotyping stategy. Regardless, GT-seq remains more flexible and scalable. Additionally, SNPs are the wave of the future, and new genotyping methods in the near future will be SNP based although high throughput techniques for microsatellites may retain microsatellites as a diagnostic tool in the future. The choice of genetic marker is important as there would be a legacy cost to choosing microsatellites versus SNPs. As new techniques are developed, the utility of historical information is at risk and rerunning archived DNA samples with the new techniques could be resource intensive.

## DATA ANALYSIS APPLICATIONS

A comprehensive review of parentage analysis software is provided by Jones et al.(2010). Applications that are appropriate for the proposed SAS supplementation of the Miramichi, including accommodation of a large proportion of unsampled parents, are CERVUS© (Marshall

et al. 1998), PASOS© (Duchesne et al. 2005), COLONY2© (Wang 2004), and MASTERBAYES© (Hadfield et al. 2006). CERVUS©, PASOS© and COLONY2© use a likelihood approaches to assign parents to offspring. MASTERBAYES© takes a Bayesian approach, and can accommodate more complicated modelling and prior information (dominance/ sub-ordinance of individuals, morphology, etc.). It is important to note that all of these applications use multiple statistical approaches of the five summarized above; usually starting with exclusion.

An application that was not covered in the above review is SNPPIT© (Hauser et al. 2011) and is also worthy of consideration. It is similar to CERVUS© but built and optimized exclusively for SNP data and is therefore more computationally efficient.

Given the software specific variations, it is recommended that all five of these applications be explored for data analysis, as each has strengths and weaknesses, different assumptions and analytical approaches. CERVUS© is the most widely used application, but does not use all available data to form relationships (Walling et al. 2010). COLONY2© does use such data, but the computational time can be excessive and unrealistic for large datasets with many loci (Hauser et al. 2011). MASTERBAYES© can incorporate spatial and other prior information but it may be overcomplicated for the SAS supplementation situation of the Miramichi. Data exploration is certainly warranted with several applications. It is likely that two or more will converge on similar results.

## ADDITIONAL DESIGN REQUIREMENT RECOMMENDATIONS

There are a number of design considerations for undertaking parentage assignment programs. All SAS supplemented adults should be tissue sampled before release. An approximately equal number of wild and supplemented potential parents should be genotyped if possible. This will result in a balance design and will facilitate direct comparisons between progeny attributed to supplemented and wild individuals, and eventually to screen for parentage of returning adults. Additional sampling effort of wild spawning individuals may be required. Sex of all sampled adults should be recorded. All genetic samples should individually labelled (no bulk collections where several individuals are put in a single container) to ensure correct tracking of tissues with phenotype information of the sampled individuals.

To assess the potential relative contribution of SAS and wild parents to future production, the earliest life stage to be sampled is smolts. The more individual smolts sampled, the better, as this will improve the precision of the estimates of the supplemented vs. wild contributions. Since the Northwest Miramichi River has produced in the order of 500,000 smolts per year, genetically sampling 5000 smolts (or 1%) would be a minimum sample to consider (sampled as randomly as possible). Each genotyped parent can have fitness quantified in the number of smolts it produced and wild vs. supplemented differences can be compared. The power of different sample sizes to assign progeny to parentage types will depend on a number of factors, including the variance in smolts produced per individual, and the mean difference between wild and supplemented fitness. At the first possibility that supplemented progeny return as adults, samples of these fish can be assessed as potential progeny with the appropriate previous year classes of potential parents, providing an indication of relative fitness of wild versus SAS supplemented progeny or the entire life cycle. In addition to overall contribution, relative fitness of pure versus interbred progeny can be assessed including both wild parents, wild mother with supplemented father, wild father with supplemented mother, and both supplemented parents.

It is critical that the error rate of the genetic markers be empirically estimated and that the number of all potential parents be estimated. These variables are explicitly included in the parentage models, and are extremely important for confident assignments (Jones et al. 2010).

Note that these numbers may be very different between male and female, due to the presence of mature male parr.

SNPs should be tested on 100-200 individuals from the Northwest Miramichi so a final panel of sufficient polymorphic, unlinked loci can be selected. Simulations in CERVUS© can be conducted with the chosen panel, error rate, and unsampled parent proportion to choose wise significance thresholds that balance false positives with false negatives, and to help determine the number of sample sizes for adequate power.

## REFERENCES CITED


Bourret, V., Dionne, M., and Bernatchez, L. 2014. Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. Mol. Ecol. 23: 4444-4457.

Primmer, C.R., Vasemägi, A., Karlsson, S., Hindar, K., Mcginnity, P., Verspoor, E., Bernatchez, L., and Lien, S. 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Mol. Ecol. 22: 532-551.

Campbell, N.R., Harmon, S.A., and Narum, S.R. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Res. 15: 855-867.

Davidson, W.S. 2013. Understanding salmonid biology from the Atlantic salmon genome. Genome 56: 548-550.

Duchesne, P., Castric, T., and Bernatchez, L. 2005. PASOS (parental allocation of singles in open systems): a computer program for individual parental allocation with missing parents. Mol. Ecol. Notes 5: 701-704.

Fraser, D.J. 2016. Risks and benefits of mitigating low marine survival in wild salmon using smolt-to-adult captive-reared supplementation. DFO Can. Sci. Advis. Sec. Res. Doc. 2016/030. v + 31 p.

Freeland, J.R., Petersen, S.D., and Kirk, H. 2011. Molecular Ecology. Wiley-Blackwell.

Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Res. 11: 759-769.

Hadfield, J., Richardson, D., and Burke, T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. Mol. Ecol. 15: 3715-3730.

Hauser, L., Baird, M., Hilborn, R., Seeb, L.W., and Seeb, J.E. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Mol. Ecol. Res. 11: 150-161.

Jones, A.G., Small, C.M., Paczolt, K.A., and Ratterman, N.L. 2010. A practical guide to methods of parentage analysis. Mol. Ecol. Res. 10: 6-30.

King, T.L., Eackles, M.S., and Letcher, B.H. 2005. Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. Mol. Ecol. Notes 5: 130-132.

Marshall, T., Slate, J., Kruuk, L., and Pemberton, J. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639-655.

Metzker, M.L. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11: 31-46.

Miller, K.M., Teffer, A., Tucker, S., Li, S., Schulze, A.D., Trudel, M., Juanes, F., Tabata, A., Kaukinen, H.K., Ginther, N.G., Ming, T.J., Cooke, S.J., Hipfner, J.M., Patterson, D.A., Hinch, S.G. 2014. Infectious disease, shifting climates, and opportunistic predators: cumulative factors potentially impacting wild salmon declines. Evol. Appl. 7: 812-855.

Milot, E., Perrier, C., Papillon, L., Dodson, J.J., and Bernatchez, L. 2013. Reduced fitness of Atlantic salmon released in the wild after one generation of captive breeding. Evol. Appl. 6: 472-485.

O'Reilly, P.T., Hamilton, L.C., McConnell, S.K., and Wright, J.M. 1996. Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Can. J. Fish. Aquat. Sci. 53: 2292-2298.

Paterson, S., Piertney, S., Knox, D., Gilbey, J., and Verspoor, E. 2004. Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Mol. Ecol. Notes 4: 160-162.

Pavey, S.A. 2015. High-throughput SNPs for all: genotyping-in-thousands. Mol. Ecol. Res. 15: 685-687.

Perkel, J. 2008. SNP genotyping: six technologies that keyed a revolution. Nature Methods 5: 447-453.

Sellars, M.J., Dierens, L., McWilliam, S., Little, B., Murphy, B., Coman, g.J., Barendse, W., and Henshall, J. 2014. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. Aquac. Res. 45: 417-426.

Telfer, E.J., Stovold, G.T., Li, Y., Silva-Junior, O.B., Grattapaglia, D.G., and Dungey, H.S. 2015. Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: a comparative analysis of marker data power and robustness. PLoS ONE 10(7): e0130601. doi:10.1371/journal.pone.0130601.

Walling, C.A., Pemberton, J.M., Hadfield, J.D., and Kruuk, L.E. 2010. Comparing parentage inference software: reanalysis of a red deer pedigree. Mol. Ecol. 19: 1914-1928.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166: 1963-1979.

Weinman, L.R., Solomon, J.W., and Rubenstein, D.R. 2015. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. Mol. Ecol. Res. 15: 502-511.

## TABLES

*Table 1: Results of simulations that compared a panel of 12 microsatellites (µSat), 96 Single Nucleotide Polymorphisms (SNPs), and 400 SNPs. Parameters considered are presented in the three columns. This simulation focused on mothers only, but the results for fathers would be the similar.*

| Panel | Percent of true mothers identified | Percent false positives (true mother sampled) | Percent false positives (true mother unsampled) |
|---|---|---|---|
| 12 uSats | 99% | 0% | 1% |
| 96 SNPs | 91% | 0% | 1% |
| 400 SNPs | 100% | 0% | 0% |