MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) doi: 10.1111/1755-0998.12589

A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird

SARA A. KAISER,*† SCOTT A. TAYLOR,‡§ NANCY CHEN,‡¶ T. SCOTT SILLETT,† ELIANA R. BONDRA¶ and MICHAEL S. WEBSTER*

Macaulay Library, Cornell Lab of Ornithology, 159 Sapsucker Woods Rd, Ithaca, NY 14850, USA, †Migratory Bird Center, Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoological Park, MRC 5503, Washington DC, 20013, USA, ‡Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, 159 Sapsucker Woods Rd, Ithaca, NY 14850, USA, §Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, 1900 Pleasant Street, 334 UCB, Boulder, CO 80309, USA, ¶Department of Ecology and Evolutionary Biology, Cornell University, E145 Corson Hall, 215 Tower Road, Ithaca, NY 14853, USA

Abstract

Single-nucleotide polymorphisms (SNPs) are preferred over microsatellite markers in many evolutionary studies, but have only recently been applied to studies of parentage. Evaluations of SNPs and microsatellites for assigning parentage have mostly focused on special cases that require a relatively large number of heterozygous loci, such as species with low genetic diversity or with complex social structures. We developed 120 SNP markers from a transcriptome assembled using RNA-sequencing of a songbird with the most common avian mating system—social monogamy. We compared the effectiveness of 97 novel SNPs and six previously described microsatellites for assigning paternity in the black-throated blue warbler, *Setophaga caerulescens*. We show that the full panel of 97 SNPs (mean H_o = 0.19) was as powerful for assigning paternity as the panel of multiallelic microsatellites (mean H_o = 0.86). Paternity assignments using the two marker types were in agreement for 92% of the offspring. Filtering individual samples by a 50% call rate and SNPs by a 75% call rate maximized the number of offspring assigned with 95% confidence using SNPs. We also found that the 40 most heterozygous SNPs (mean H_o = 0.37) had similar power to assign paternity as the full panel of 97 SNPs. These findings demonstrate that a relatively small number of variable SNPs can be effective for parentage analyses in a socially monogamous species. We suggest that the development of SNP markers is advantageous for studies that require high-throughput genotyping or that plan to address a range of ecological and evolutionary questions.

Keywords: microsatellite, paternity, RNA-seq, single-nucleotide polymorphism, social monogamy, transcriptome

Received 20 May 2016; revision received 14 July 2016; accepted 19 July 2016

Introduction

The application of molecular markers has revolutionized population ecology (Avise 1994; Freeland *et al.* 2011). Their application to natural populations has transformed our view of mating systems in most taxa (Avise *et al.* 2002; Griffith *et al.* 2002; Uller & Olsson 2008; Roberts & Byme 2011) and enabled the estimation and comparison of genetic variation and gene flow (dispersal) among populations (Brumfield *et al.* 2003; Morin *et al.* 2004; Garvin *et al.* 2010; Helyar *et al.* 2011). The insights gained from their application in studies of parentage and population structure continue to deepen as the technology to isolate and develop more informative molecular markers, such as microsatellites and single-nucleotide polymorphisms (SNPs), becomes more efficient and cost-effective (Morin *et al.* 2004; Anderson & Garza 2006; Abdelkrim *et al.* 2009; Dawson *et al.* 2013). Despite technological improvements and decreasing costs, microsatellite and SNP marker development is labour-intensive and requires substantial time to validate (Morin *et al.* 2004). It is therefore advantageous to invest in the development of markers suitable for addressing a wide range of ecological and evolutionary questions in population studies.

For over a decade, microsatellites have been the most popular molecular marker for studies of parentage and kinship in wild populations of mammals, reptiles, fishes and birds (Vignal *et al.* 2002; Uller & Olsson 2008;

Correspondence: Sara A. Kaiser, Fax: 202-673-0040;
E-mail: KaiserS@si.edu

Coleman & Jones 2011; Dawson *et al.* 2013). Microsatellites are highly polymorphic at each locus because of their high mutation rates, which means that only a few microsatellite loci are necessary to distinguish individuals and estimate relatedness. However, microsatellites can also be difficult to score accurately (high genotyping error), requiring methods to quantify and minimize genotyping error rates in parentage analyses (Pompanon *et al.* 2005; Kalinowski *et al.* 2007). In contrast, SNPs present a more attractive marker for many population-level questions due to their abundance and broader genome coverage (coding and noncoding regions; Brumfield *et al.* 2003), ease of scoring (low genotyping error; Ranade *et al.* 2001), high-throughput and low-cost genotyping (Anderson & Garza 2006) and known mutational processes (Brumfield *et al.* 2003; Ellegren 2004). For studies of parentage, however, the power to exclude nonparents may be lower in SNPs because these markers are mostly biallelic compared to multiallelic microsatellites, and their allele frequencies are often skewed, resulting in low heterozygosity (Marth *et al.* 2001; Glaubitz *et al.* 2003). This power issue can be resolved by using a larger set of SNP markers (Glaubitz *et al.* 2003; Anderson & Garza 2006). Comparisons between microsatellites and SNPs for use in parentage studies are mostly limited to model species with reference genomes (Anderson & Garza 2006), domesticated and farmed species (Heaton *et al.* 2002; Rengmark *et al.* 2006; Sellars 2014) and species with low genetic diversity (Tokarska *et al.* 2009; Fernández *et al.* 2013; but see Hauser *et al.* 2011). Comparisons of these marker types are lacking for wild populations of most taxa and for species with a range of mating systems (Weinman *et al.* 2014).

Birds have been important for studies of extra-pair paternity across the diversity of animal mating systems (Westneat *et al.* 1990; Griffith *et al.* 2002; Westneat & Stewart 2003; Wan *et al.* 2013), but the effectiveness of SNPs for parentage has not been comprehensively examined. Some bird studies have used SNPs in combination with microsatellites for parentage analysis when using microsatellites alone was not informative (Backström *et al.* 2008; Cramer *et al.* 2011; Labuschagne *et al.* 2015). One study of a cooperatively breeding bird found that SNPs performed as well as microsatellites at assigning paternity (Weinman *et al.* 2014). However, only a small fraction (9%) of bird species are cooperative breeders compared to 90% that are socially monogamous (Griffith *et al.* 2002; Cockburn 2006; Jetz & Rubenstein 2011). Complex kin structure in cooperatively breeding mating systems should require a larger number and higher variability of SNPs to assign paternity because of higher relatedness among potential fathers than in socially monogamous mating systems (Weinman *et al.* 2014). It is unclear how many polymorphic SNPs would be adequate for assigning paternity in a socially monogamous species with a large number of potential fathers.

We address this knowledge gap by comparing the power of SNPs and microsatellites for paternity assignment in the socially monogamous black-throated blue warbler, *Setophaga caerulescens*. We have studied the mating system of this migratory bird at a site in the northern portion of its breeding range for over 20 years (Holmes 2007, 2011). Paternity analyses using genotypes generated from six microsatellites have revealed that extra-pair fertilizations are common in this population, occurring in 56% of nests (Webster *et al.* 2001; Kaiser *et al.* 2015). For this study, we developed SNP markers from a transcriptome assembled using RNA-sequencing (RNA-seq) of black-throated blue warblers sampled in northern populations. We genotyped individuals with each set of markers and compared the results of paternity analyses using the most widely used parentage program, CERVUS version 3.0 (Kalinowski *et al.* 2007). Missing data are a challenge in studies that use reduced representation genomic approaches combined with high-throughput sequencing methods to identify and genotype SNPs (Nielsen *et al.* 2011; Toews *et al.* 2015). To inform filtering decisions of missing data, we assessed how adjustments in the variant filtering parameters affected the confidence of paternity assignments for subsets of SNPs. We then determined the number and variability of SNPs that were sufficient to obtain similar parentage assignment success to those generated from analyses using the full panel of microsatellites or SNPs. This comprehensive assessment provides valuable information for investigators developing SNP markers from the transcriptomes of nonmodel organisms for use in population studies of socially monogamous species.

Materials and methods

Population sampling

We conducted a long-term parentage study of a marked population of black-throated blue warblers breeding in the 3160-ha Hubbard Brook Experimental Forest, Woodstock, New Hampshire, USA (43°56′N, 71°45′W); 1995–2015. The study area is in the northern portion of their breeding range, which extends from the north-eastern United States and southern Canada southward along the higher elevations of the Appalachian Mountains (Holmes *et al.* 2005). In each breeding season, we captured, colour-banded and collected blood samples from adults, and monitored nesting attempts by social parents (those defending territories and providing parental care). We banded and collected blood samples from nestlings 6 days posthatching (mean clutch size = 3.6, range = 2–5 eggs; Holmes *et al.* 2005). All blood samples were stored

in lysis buffer (White & Densmore 1992) and genomic DNA extracted using Qiagen DNeasy Blood and Tissue kits (Qiagen). We quantified DNA yields using a QUBIT 2.0 Fluorometer (Life Technologies) to check that final concentrations were at least 10 ng/μL for genotyping. We followed protocols approved by our Institutional Animal Care and Use Committees to capture, handle, mark and sample black-throated blue warblers (Cornell University, 2009-0133; Smithsonian National Zoological Park, 08-11, 12-12; Wellesley College, 1304).

Transcriptome assembly from RNA-sequencing

RNA-sequencing and transcriptome assembly were carried out following Mason & Taylor (2015). We constructed two barcoded RNA-seq libraries for transcriptome assembly and SNP discovery. The first RNA-seq library was generated by combining 15 mg each of liver, muscle and brain (45 mg total tissue) from one adult male black-throated blue warbler that we collected and preserved in RNAlater (Qiagen) during the 2013 breeding season in the Adirondack Mountains, NY. We used a TissueRuptor (Qiagen) to homogenize the individual tissue pool. The second RNA-seq library was generated by pooling 25 μL of whole blood preserved in RNAlater obtained from 10 individuals sampled during fall migration in 2013 at the Braddock Bay Bird Observatory, NY. We followed the 'standard' mRNA extraction protocol as detailed in the Dynabeads® mRNA DIRECT™ kit (Invitrogen). We performed the mRNA extraction protocol twice to maximize the removal of rRNA from our extraction before constructing cDNA libraries. We converted mRNA into cDNA libraries using the NEBNext RNA First Strand Synthesis Module (New England BioLabs). We then performed second strand cDNA synthesis, end repair, dA-tailing and adaptor ligation for each cDNA library. Following adapter ligation and library purification, we performed 12 cycles of the 'denaturation annealing extension' step during the index PCR. We assessed the quality and quantity of cDNA using the QFC algorithm and an Agilent Bioanalyzer 2100. The two barcoded cDNA libraries were pooled in equimolar ratio and sequenced on a single lane using 100-bp single-end reads on an Illumina HiSeq 2000 at the Cornell University Core Laboratories Center. Raw, demultiplexed reads are available through the NCBI Short Read Archive (SRA Accession nos: SRR3164929, SRR3165127; PRJNA311681).We demultiplexed 166 622 978 barcoded reads, filtered the reads (min phred-scale quality score = 20) and trimmed low-quality sequences and sequences including contamination from Illumina adapters with TRIMMOMATIC version 0.27 (Lohse *et al.* 2012), resulting in 155 550 748 single-end reads. We conducted a *de novo* assembly of the reference transcriptome using both cDNA libraries with TRINITY version 2013-02-25 (Grabherr *et al.* 2011; Haas *et al.* 2013). We filtered our raw transcriptome assembly by removing contigs that failed to meet a 2-TPM threshold in at least one individual and retained only the longest isoform of each transcript. The resulting assembly consisted of 16 261 transcripts with an average contig length of 977 bp and N50 of 2107 bp.

SNP marker development

We aligned trimmed reads from all 11 individuals to the transcriptome assembly using the BURROWS–WHEELER ALIGNER (BWA-backtrack, version 0.7.13; Li & Durbin 2009) with default parameters. We used Picard tools (http://broadinstitute.github.io/picard) to sort, clean and merge the resulting binary alignment files (BAM) files and to mark PCR duplicates before indel realignment and SNP calling with UnifiedGenotyper in the GENOME ANALYSIS TOOLKIT (GATK version 3.1; DePristo *et al.* 2011). We followed the GATK best practices for SNP calling, skipping the base quality recalibration and variant quality recalibration steps because we did not have a set of known variants. After removing monomorphic sites and filtering based on a number of standard quality metrics (Fisher Strand FS <30, read depth >10, site quality >50, RMS Mapping Quality MQ >35, Qual by Depth QD >3), we found 122 680 SNPs and 7056 indels. From this panel of genomewide SNPs, we selected the highest quality 5000 biallelic SNPs spaced at least 150 bp apart (read depth >45, site quality >90, QD >10, MQ >35). We checked for potential splice sites in the 150-bp flanking regions of each SNP by aligning the 301-bp fragments to the zebra finch (*Taeniopygia guttata*) genome (assembly 3.2.4) using standalone BLAST (blast-n algorithm, default parameters; Camacho *et al.* 2009) and removed any SNPs that aligned to more than one location. Because avian genomes have high synteny (Ellegren 2010), we could assign each SNP a putative chromosome location based on the best BLAST hit based on e-value. We submitted the remaining 1325 SNPs and their flanking regions to RepeatMasker (Smit *et al.* 2015) and removed 614 SNPs with repetitive sequence in the flanking regions. The remaining 1229 SNPs were submitted to DNA LandMarks Inc. (Saint-Jean-sur-Richelieu, Québec, Canada) for primer design.

DNA LandMarks Inc. designed the PCR and iPLEX extension primers for SNP genotyping on the Sequenom MassArray using ASSAY DESIGN SUITE version 1.0, with EXTEND PRIMER ASSAY DESIGN version 4.1.0.17. From the 1229 SNPs identified, 17 SNP panels were randomly generated. Each panel was designed to multiplex 60 primers per well and reduce cross-binding among multiplexed primers to increase primer specificity and to reduce

genotyping errors (typically <1%; Gabriel *et al.* 2009). We selected two wells containing 120 SNPs distributed across chromosomes, including the Z chromosome (birds: males ZZ/ females ZW) for our SNP assay (Table S1, Supporting information).

Microsatellite and SNP genotyping

During our 20-yr study, we genotyped 4413 offspring from 1346 nests at six polymorphic microsatellite loci previously isolated from black-throated blue warbler and related passerines (Table S2, Supporting information). We amplified 1 μL of genomic DNA from each individual at each locus in two multiplexed PCRs (*Dca*28, *Dpu*01 and *Cuu*04; *Dca*32, *Dpu*16 and *Vecr*08) following previously described methods (Kaiser *et al.* 2015). The microsatellite PCR products were size-sorted on a 3730 DNA Analyzer (Applied Biosystems [ABI], Carlsbad, CA, USA) and run with the GeneScan 500 base pair LIZ internal size standard (ABI). We scored alleles at each locus by visually confirming automated scores generated by GENEMAPPER version 3.7 (ABI) and repeated PCRs at homozygous loci to verify genotypes.

For comparison of microsatellites, we selected a subset of individuals for SNP genotyping that we sampled in 2 years where 95% of the candidate fathers in the population were sampled (134 offspring from 38 nests in 2011, 104 offspring from 33 nests in 2012). SNP genotyping was conducted by DNA LandMarks Inc. using the iPLEX assay on a Sequenom MassArray. SNP loci were amplified from genomic DNA using standard PCR followed by the iPLEX single base extension reaction, which yielded allele-specific differences in mass for each SNP locus (Gabriel *et al.* 2009). Samples were analysed with MALDI-TOF mass spectrometry and TYPER software generated genotypes for each locus. Of the 120 SNP markers selected for genotyping, 16 SNPs were removed from the assay due to their low performance (call rate <0.5) and 7 SNPs were removed based on low minor allele frequency (MAF <0.0009). The final data set included 97 SNPs (Table S1, Supporting information). The mean call rate for the 97 loci was 94% (range: 64–99%).

Parentage analysis

We conducted separate paternity analyses with microsatellite and SNP genotype data for each breeding season using CERVUS version 3.0, which uses a maximum-likelihood-based approach to infer parentage (Marshall *et al.* 1998; Kalinowski *et al.* 2007). CERVUS calculates the natural logarithm of the likelihood ratio (LOD score), which provides the likelihood of paternity of each candidate male relative to a random male in the population for each offspring. CERVUS uses simulations of the allele frequencies of adults in the population to calculate critical differences in LOD scores between the most likely father and all other candidate fathers to assign paternity at either 80% or 95% confidence. We simulated paternity for 10 000 offspring to determine the critical LOD scores for the assignment of paternity for each marker and subsets of the full panel of 97 SNPs (simulation parameters given in Table S3, Supporting information). In each paternity analysis, we included known mothers (confirmed via focal nest observations and later by examining the frequency of mother–offspring mismatches from microsatellite-based parentage analyses). We included broods only if the social father (i.e. the male paired to the known mother) was sampled. CERVUS assignments of the most likely fathers (>95% candidate fathers sampled in the population) were made using trio LOD scores, which statistically distinguishes among nonexcluded candidate males while considering the genotypes of the known mother and potential mistyping errors.

For microsatellite-based paternity analyses, we evaluated CERVUS assignments using trio likelihood scores and decision rules previously developed for this species (Smith *et al.* 2005). Briefly, if the social father had a high likelihood score but mismatched the nestling at one or two loci, we investigated the possibility of null alleles or mistyping by repeating PCRs at mismatched loci and rescoring the alleles. We accepted the CERVUS assignment of the social father as the genetic sire of a nestling if he had ≤1 mismatches. A nestling was considered sired by an extra-pair male if the social father was not listed as a candidate father by CERVUS (because of a negative trio LOD score) or had ≥2 mismatches. In these cases, we identified the extra-pair father as the male that had ≤1 mismatches and that had the highest trio LOD score. In some cases, no candidate male matched the nestling's genotype (i.e. all had >2 mismatches with the nestling). We considered these nestlings to have been sired by an unsampled male (5% of males on study plot).

For SNP-based paternity analyses, we evaluated CERVUS assignments using trio LOD scores and reported the average number of father–offspring mismatches given the number of loci typed for assignments with 80% and 95% confidence. Based on these analyses, we developed decision rules for paternity analyses using SNP markers (see Results).

Data sets

To compare the efficacy of SNP and microsatellite markers for assigning parentage, we filtered SNPs and individuals based on completeness (loci and sample call rate >50%). To avoid biases in the marker comparison caused by missing data, we removed 34 offspring from the

microsatellite data set that were filtered from the SNP data set. This resulted in 410 individuals genotyped at 97 polymorphic SNP loci in our baseline SNP data set. We examined differences between SNPs and microsatellites in the number of offspring assigned at two CERVUS confidence levels (80% and 95%) and in the identities of assigned genetic sires (sample size = 238 offspring).

We conducted 18 additional paternity analyses with subsets of SNP loci to examine the effect of missing data on assignments of paternity and to determine the number and variability of SNP loci that would be sufficient to match assignments of paternity using the full panel of SNP loci with 95% confidence. We compared SNP data sets that differed in completeness in call rates of SNPs (loci call rate >50%, 75% and 90%), individuals (sample call rate >50%, 75% and 90%) and both SNPs and individuals (call rates >50%, 75% and 90%). We quantified the proportion of paternity assignments that matched the full panel of 97 SNPs. Next, we selected from the full panel of 97 SNP loci 1) the most heterozygous SNP loci and 2) the SNP loci with the highest call rates, and incrementally reduced the number of SNP loci included in both analyses (n = 97, 77, 57, 37, 27, 17, 7) to compare their power and efficiency for assigning paternity. Each comparison was restricted to 198 offspring assigned to social or extra-pair fathers with 95% confidence using the baseline SNP data set (loci and sample call rate >50%, 97 SNPs).

Of the 97 SNP loci, six deviated significantly from Hardy–Weinberg equilibrium (HWE; Table S1, Supporting information). We included these SNP loci in parentage analyses because they deviated from HWE using the baseline SNP data set (loci call rate >50%) from either 2011 or 2012, but not both years. However, all but two of these loci were filtered from the reduced SNP data sets (two filtered from loci call rate >75% and four from loci call rate >90%).

Results

Microsatellite and SNP marker characteristics

The microsatellite panel was more polymorphic than the SNP panel, but the two markers had comparable power to assign the genetic father when the mother was known (Table 1). The mean number of alleles per locus, mean observed and expected heterozygosity and polymorphic information content (PIC) were greater for microsatellite loci. The combined nonexclusion probabilities for the father were lower for SNPs, but the combined nonexclusion probabilities for the parent pair were similar for microsatellites and SNPs. In all cases, we knew the identity of the mother from behavioural observations at the nest and our goal was to identify the genetic father of each genotyped offspring from the sampled parents by excluding candidate fathers when the mother was known.

Comparison of microsatellite and SNP paternity assignments

The full panel of 97 SNPs was as effective at assigning paternity as the panel of six microsatellites. Of the paternity assignments without the known mother, 206 of 238 (87%) offspring were assigned with 95% confidence using the full SNP panel, 219 of 238 (92%) offspring were assigned with 95% confidence using microsatellites, and all offspring were assigned with 80% confidence using either marker type (Table S4, Supporting information). Of the paternity assignments given the known mother, 238 of 238 (100%) offspring were assigned with 95% confidence using the full SNP panel, 237 of 238 (99%) offspring were assigned with 95% confidence using microsatellites, and all offspring were assigned with 80% confidence using either marker (Table S4, Supporting information).

Offspring were mostly assigned to the same social fathers and extra-pair fathers using microsatellites and SNPs (Table 2). Given known mothers, 194 of 238 (82%) offspring were assigned to the same father by the two markers with 95% confidence in CERVUS (highest trio LOD score). Of these assignments, 126 of 195 (65%) offspring were assigned to their social father and 68 of 195 (35%) offspring were assigned to the same extra-pair father. The average number of father–offspring mismatches for assignments with 95% confidence was 0.10 for microsatellites and 1.02 for SNPs. Two offspring were assigned the same extra-pair father by the two markers, but at different confidence levels. The microsatellites assigned extra-pair fathers to these offspring with 95% confidence and SNPs assigned the same extra-pair father despite negative LOD scores. In these cases, either the extra-pair father or offspring had an incomplete SNP genotype (68 of 97 loci and 62 of 97 loci, respectively). Six offspring were assigned to their social father using microsatellites with 95% confidence when SNPs ranked the social father second after an extra-pair father genotyped at fewer loci (lower sample call rate) with a similar LOD score. We considered these six offspring to be sired by their social father and rejected the CERVUS assignment using SNPs.

For 20 of 238 (8%) offspring, we found no consensus between microsatellite and SNP assignments (Table 2). Microsatellites assigned three offspring to their social father with 95% confidence when SNPs failed to assign fathers because of incomplete offspring SNP genotypes (fewer loci compared between candidate males and offspring). Microsatellites assigned four offspring to

Table 1 Comparison of marker characteristics* for microsatellites and single-nucleotide polymorphisms (SNPs) used for CERVUS paternity analyses from 172 adult black-throated blue warblers sampled during the 2011 and 2012 breeding seasons at Hubbard Brook

Marker panel	Number of loci	Mean proportion loci typed	Mean alleles per locus	Mean H_e	Mean H_o	Mean PIC	Nonexclusion probability (first parent)	Nonexclusion probability (second parent)	Nonexclusion probability (parent pair)
Microsatellites	6	1.00	18.92	0.87	0.86	0.86	1.9×10^{-3}	1.1×10^{-4}	1.3×10^{-7}
SNPs	97	0.96	2.00	0.22	0.19	0.18	1.9×10^{-2}	9.0×10^{-5}	1.6×10^{-7}

*Mean alleles per locus, mean expected (H_e) and observed (H_o) heterozygosities, mean polymorphic information content (PIC) and combined nonexclusion probabilities.

Table 2 Comparison of the parentage assignments for 238 offspring using microsatellites or single-nucleotide polymorphisms (SNPs) and the corresponding mean number of locus mismatches. Assignments are separated by cases for which the two marker types were in agreement, disagreement or could not be resolved

		Mean number of locus mismatches	
CERVUS assignments	*n*	Microsatellites	SNPs
Agreements (same father)	200	0.10	1.02
Disagreements			
Social father vs. no assignment	3	<2	–
Extra-pair father vs. no assignment (SNPs)	4	<2	–
Extra-pair father (SNPs) vs. no assignment	9	–	<2
Different extra-pair father	4	<2	<2
Unresolved	18	2.18	2.33

extra-pair fathers when SNPs failed to assign fathers. In these cases, either the extra-pair father assigned by microsatellites had an incomplete SNP genotype or was ranked by SNPs after other extra-pair males, all with negative LOD scores. SNPs assigned nine offspring to extra-pair fathers with 95% confidence when microsatellites failed to assign fathers. In each case, all father–offspring mismatches at microsatellite loci were >2 and all candidate males had negative LOD scores. Behavioural data indicated that SNPs identified the most likely extra-pair father. Two offspring were assigned different extra-pair fathers by microsatellites and SNPs. The most likely identity of the extra-pair father was resolved using behavioural data for one offspring that was assigned with SNPs. For the second offspring, behavioural data could not resolve whether the microsatellite or SNP assignment was more likely.

For 18 of 238 (8%) offspring, no candidate male was a good genetic match with either marker type using our paternity assignment criteria (Table 2). We assumed these offspring to be sired by an unsampled extra-pair male. The average number of father–offspring mismatches for top-ranked fathers of unassigned offspring was 2.18 for microsatellites and 2.33 for SNPs.

Overall, CERVUS paternity assignments of 220 of 238 (92%) were in agreement using microsatellites and SNPs. In nine cases where microsatellites found no genetic match, SNPs were able to resolve the fathers of offspring with 95% confidence that we otherwise would have assumed to be sired by unsampled males. However, in seven cases where SNPs failed to assign fathers because of incomplete offspring or candidate male genotypes, microsatellites assigned fathers with 95% confidence. Thus, the performance of the full panel of 97 SNPs compared to the microsatellites depended on sample call rates across SNP loci.

Comparison of paternity assignments with missing SNP data

Missing data in the SNP data sets influenced the assignment of paternity and the number of offspring compared in the paternity analyses. Increasing the filtering threshold of SNPs by call rate decreased the number of SNP loci compared and the mean error rate across loci (Table S3, Supporting information) with only minor reductions in the assignment of paternity with 95% confidence (Fig. 1). Increasing the filtering threshold of sample call rates by 75% and 90% reduced the number of offspring typed and compared by 5 and 14 offspring, respectively (Table S3, Supporting information), but increased concordance in parentage assignments between microsatellites and SNPs.

Comparison of paternity assignments with different numbers and variability of SNP loci

A small number of heterozygous SNPs were adequate for assigning paternity, but relatively more SNP loci were necessary as the call rates of SNP loci decreased. A subset of 40 SNPs with a mean heterozygosity of 0.37 or 80 SNPs with a mean call rate of 0.97 had sufficient power for assigning paternity to 95% of offspring with known mothers (Fig. 2). Thus, a reduced SNP panel

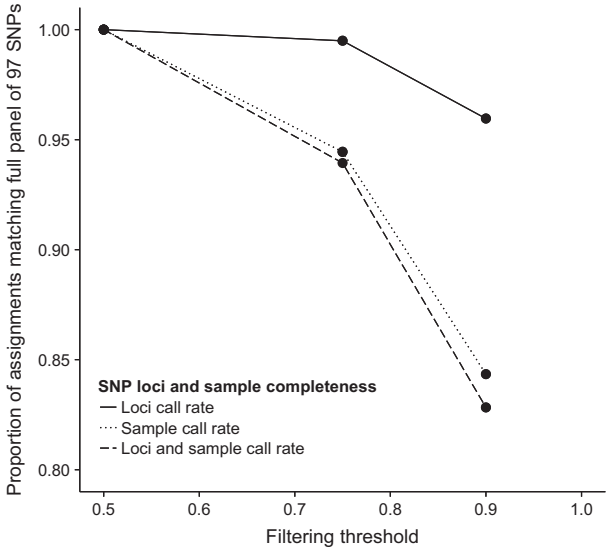

Fig. 1 Proportion of paternity assignments from SNP data sets that differed in completeness (loci call rate, solid line; sample call rate, dotted line; loci and sample call rates, dashed line) that matched assignments of 198 black-throated blue warbler offspring using the full panel of 97 SNPs with 95% confidence.

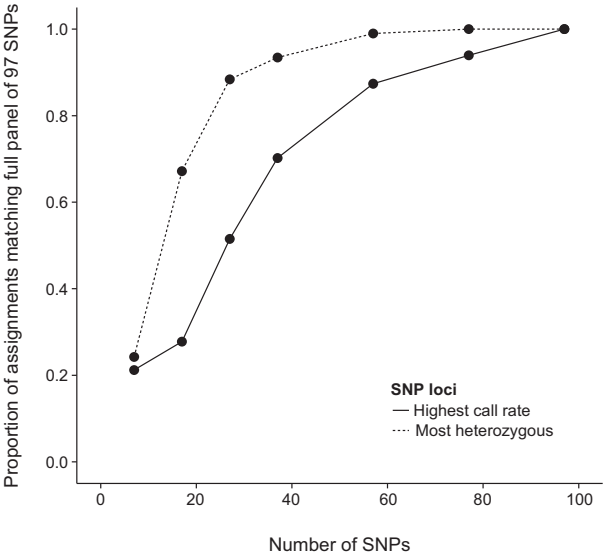

Fig. 2 Proportion of paternity assignments from decreasing numbers of single-nucleotide polymorphisms (SNPs) with the highest heterozygosity (dashed line) and highest call rates (solid line) that matched assignments of 198 black-throated blue warbler offspring using the full panel of 97 SNPs with 95% confidence.

could be used for parentage analyses when the heterozygosity of SNPs is optimized over the call rate of SNPs.

Discussion

Single-nucleotide polymorphisms are preferred over microsatellite markers in studies inferring the evolutionary history of populations because of their known mutation patterns and low genotyping error rates (Brumfield *et al.* 2003). Yet, SNPs have only recently been applied to studies of parentage (e.g. Hess *et al.* 2015; Liu *et al.* 2016; Strucken *et al.* 2016). The power to assign parentage depends on the number of loci examined and their heterozygosity; more loci are necessary as heterozygosity declines (Miller *et al.* 2002; Morin *et al.* 2004). The few studies comparing the characteristics of SNPs and microsatellites for parentage assignment have mostly focused on domesticated species with low genetic diversity (e.g. Tokarska *et al.* 2009; Sellars 2014) or species with complex social and kin structures (e.g. Weinman *et al.* 2014). These systems are special cases that require a relatively large number of heterozygous loci to estimate parentage and relatedness and are not broadly applicable to socially monogamous systems—the most common mating system in birds and one observed in most taxa (Gowaty 1996). We show that 97 SNPs (mean H_o = 0.19, mean alleles per locus = 2.00) had similar power to assign paternity as six microsatellites (mean H_o = 0.86, mean alleles per locus = 18.92) in the socially monogamous black-throated blue warbler. Paternity assignments using SNPs or microsatellites were in agreement for 92% of the offspring. SNPs were marginally more effective at resolving parentage than microsatellites, but their performance decreased with missing data. We also show that by selecting the most heterozygous SNPs, 40 SNPs (mean H_o = 0.37) were sufficient for assigning paternity to the same proportion of offspring as the full panel of 97 SNPs.

Selecting SNP loci with high heterozygosity provides the greatest power for assigning parentage relative to microsatellite-based analyses (Krawczak 1999; Morin *et al.* 2004; Anderson & Garza 2006). For example, in a simulation study, Morin *et al.* (2004) found that 40–100 SNPs with a mean heterozygosity of 0.3 were needed to achieve a similar probability of paternal exclusion as 7–14 microsatellite loci with a mean heterozygosity of 0.75. Simulations of parentage in a population of European bison (*Bison bonasus*) with extremely low genetic diversity required 50–100 SNPs with a mean heterozygosity of 0.5 to assign paternity with 95% confidence as 17 microsatellite loci with a low mean heterozygosity of 0.3 (Tokarska *et al.* 2009). An empirical comparison of microsatellites and SNPs in a population of superb starlings (*Lamprotornis superbus*) with high kin structure found that 60 SNPs with a mean heterozygosity of 0.4 were sufficient for assigning paternity among highly related candidate fathers, but only when the social parents were identified in the analyses (Weinman *et al.* 2014). We examined a more general case for birds and

found that 40 SNPs with a mean heterozygosity of 0.4 were adequate for assigning paternity in a socially monogamous bird population with known social parents. These results demonstrate that a small number of SNPs can be just as powerful as multiallelic microsatellites for assigning paternity when the heterozygosity of loci is considered.

The main factor reducing the accuracy of paternity assignments using SNP loci was missing data, which resulted in incomplete genotypes of offspring or candidate fathers. Missing data are a challenge in any application of high-throughput SNP genotyping (Nielsen *et al.* 2011; Toews *et al.* 2015). We developed a custom SNP genotyping assay to multiplex 60 loci per well using the Sequenom MassARRAY platform. This work was conducted in a specialized laboratory equipped with Sequenom technology; it was not possible to rerun failed individual PCRs for specific loci that were not called for a set of individuals. Comparison of paternity assignments made with increasing completeness of SNP loci indicated that filtering loci by a 75% call rate was sufficient for assigning paternity to the same proportion of offspring as the full panel of 97 SNPs. However, the proportion of offspring assigned and the accuracy of those assignments decreased when both SNP loci and samples were filtered by just a 50% call rate. Therefore, to maximize the number of offspring assigned with 95% confidence, we suggest first filtering samples by a 50% call rate and then filtering SNP loci by a 75% call rate.

The choice and methods to develop SNPs or microsatellites for studies of parentage should be made based on the scope of the study, while considering the advantages and limitations of each marker type and method of marker development. The initial costs are similar for isolating microsatellites using 454 pyrosequencing and for SNP discovery using next-generation sequencing (Abdelkrim *et al.* 2009; Helyar *et al.* 2011). The overall development time of microsatellites, including primer design and testing, screening for polymorphism, PCR optimization and genotyping by multiplexing, is substantially longer than the development time of a SNP genotyping assay with a panel of 100–200 SNPs. The disadvantage of SNP development is that SNP discovery requires bioinformatics training. The disadvantage of designing a SNP genotyping array is that it requires access to genotyping platforms such as Sequenom or Illumina BeadChips. We worked with a private facility to develop our Sequenom SNP panel, but academic core facilities also provide access to specialized software for assay design and equipment for SNP genotyping. Long-term studies in which new individuals will be added each year and studies that generate large sample sizes, such as selective breeding programmes, would benefit from the development of a SNP screening panel because of their high-throughput, repeatable genotyping. Studies that involve multiple laboratories would also benefit from SNP marker development because of their reproducibility and low genotyping error rates compared to microsatellites.

Several high-throughput sequencing methodologies now allow high-throughput and simultaneous SNP discovery and genotyping in any organism (Davey *et al.* 2011) for use in paternity and relatedness studies, including genotyping by sequencing (GBS; Dodds *et al.* 2015), restriction site-associated DNA sequencing (RAD-seq; Senn *et al.* 2013; Kess *et al.* 2016), transcriptome sequencing (Schunter *et al.* 2014), sequence capture (Mamanova *et al.* 2010), and whole genome resequencing. Appropriate SNPs can subsequently be used in high-throughput genotyping with or without the generation of a genotyping assay (Toews *et al.* 2015). We developed SNP markers from a transcriptome assembled using RNA-seq rather than these alternative methods because of the availability of fresh blood and tissue samples (necessary for RNA extraction), the absence of a reference genome for *S. caerulescens* and our desire for a repeatable SNP genotyping panel. Methods such as RAD-seq and GBS could provide a sufficient number of high-quality SNPs for a paternity project like the one we describe here; however, these methods do not necessarily sample the same loci across all individuals and require specialized library preps for each sample. RAD-seq and GBS methods generate data with high variation in coverage across individuals and loci, which can increase uncertainty in SNP calls and lower the repeatability in genotyping (Davey *et al.* 2011; Chen *et al.* 2014). To generate a SNP genotyping panel like the one described here, SNPs of interest must be surrounded by adequate flanking sequence for primer or probe design; such sequence is usually not generated by RAD-seq or GBS unless paired-end sequencing is conducted. Additionally, this limitation can be overcome if a reference genome, to which sequence reads can be mapped, is available. Given that reference genomes are available for a limited number of taxa, RNA-seq and transcriptome assembly or paired-end sequencing of RAD-seq or GBS data are currently the best options for SNP discovery for paternity studies.

The primary advantage of SNPs is that their mutation and evolution are better understood than microsatellites, which makes SNP markers useful for applications in population genetics (Brumfield *et al.* 2003; Morin *et al.* 2004). The mutational process for microsatellites is complex, making microsatellites unsuitable for many population genetic analyses (Ellegren 2000; Schlötterer 2000). In addition, given current technologies, SNPs can be genotyped in a more cost-effective and high-throughput

manner. However, because SNP discovery is typically performed with a smaller sample of individuals, studies using SNP genotyping arrays need to consider ascertainment bias (Brumfield *et al.* 2003; Clark *et al.* 2005). For example, using SNPs discovered in individuals from the northern population of black-throated blue warblers could downwardly bias estimates of genetic diversity in the southern population, which may have expanded from a separate glacial refugium (Davis *et al.* 2006; Grus *et al.* 2009). Ascertainment bias is not a big issue for paternity analyses, but is a nontrivial consideration if a single panel of SNP markers is applied to multiple types of studies, especially those that use allele frequencies to infer demography or selection (Morin *et al.* 2004). Fortunately, it is often possible to correct for ascertainment bias (Nielsen *et al.* 2004).

SNPs are rapidly replacing microsatellites in population studies as their development becomes more efficient and less expensive for nonmodel species. We generated a large panel of genomewide SNPs from the transcriptome of the black-throated blue warbler assembled using RNA-seq and developed a high-throughput SNP genotyping assay on the Sequenom MassArray. We demonstrate that a small number of heterozygous SNPs are as powerful as polymorphic microsatellites for assigning paternity in a socially monogamous songbird. The main drawback to the SNP genotyping assay used in this study was that missing data resulted in incomplete genotypes, which affected the ability of SNPs to resolve parentage in a few cases. This could be overcome by using an alternative genotyping platform or by genotyping a larger number of SNPs, especially as methods to increase the number of multiplexed SNP loci improve. The development of SNP markers is advantageous for simultaneously addressing a variety of ecological and evolutionary questions. However, practical considerations, such as the method of SNP development and ascertainment bias, should be carefully considered if a single panel of markers is applied to multiple populations.

Acknowledgements

We are grateful to D. Bonter, A. Patterson, N. Mason, volunteers at the Braddock Bay Bird Observatory and the field technicians at Hubbard Brook who contributed their efforts to this study. We thank B. Butcher and K. Grabenstein for assistance in the laboratory and I. Lovette and L. Stenzler for technical advice on microsatellite and SNP genotyping. This research was supported by US National Science Foundation grants awarded to Cornell University (0640470), the Smithsonian Institution (0640732) and Wellesley College (064082300). This research is a contribution of the Hubbard Brook Ecosystem Study, part of the Long-Term Ecological Research network supported by the US National Science Foundation.

References

Abdelkrim J, Robertson BC, Stanton J-AL, Gemmell NJ (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, **46**, 185–192.

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, **172**, 2567–2582.

Avise JC (ed.) (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Avise JC, Jones AG, Walker D, DeWoody JA (2002) Genetic mating systems and reproductive natural histories of fishes: lessons for ecology and evolution. *Annual Review of Genetics*, **36**, 19–45.

Backström N, Fagerberg S, Ellegren H (2008) Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology*, **17**, 964–980.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Chen N, Van Hout CV, Gottipati S, Clark AG (2014) Using Mendelian inheritance to improve high-throughput SNP discovery. *Genetics*, **198**, 847–857.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, **15**, 1496–1502.

Cockburn A (2006) Prevalence of different modes of parental care in birds. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1375–1383.

Coleman SW, Jones AG (2011) Patterns of multiple paternity and maternity in fishes. *Biological Journal of the Linnean Society*, **103**, 735–760.

Cramer ERA, Hall ML, de Kort SR, Lovette IJ, Vehrencamp SL (2011) Infrequent extra-pair paternity in the banded wren, a synchronously breeding tropical passerine. *Condor*, **113**, 637–645.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Davis LA, Roalson EH, Cornell KL, McClanahan KD, Webster MS (2006) Genetic divergence and migration patterns in a North American passerine bird: implications for evolution and conservation. *Molecular Ecology*, **15**, 2141–2152.

Dawson DA, Ball AD, Spurgin LG *et al.* (2013) High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC Genomics*, **14**, 176–197.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Dodds KG, McEwan JC, Brauning R *et al.* (2015) Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genomics*, **16**, 1047–1061.

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **16**, 551–558.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Ellegren H (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution*, **25**, 283–291.

Fernández ME, Goszczynski DE, Lirón JP *et al.* (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and Molecular Biology*, **36**, 185–191.

Freeland JR, Petersen SD, Kirk H (eds.) (2011) *Molecular Ecology*. John Wiley & Sons, Chichester.

Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current Protocols in Human Genetics*, **60**, 2.12.1–2.12.16.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources*, **10**, 915–934.

Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology*, **12**, 1039–1047.

Gowaty PA (1996) Battles of the sexes and origins of monogamy. In: *Partnership in Birds: The Study of Monogamy* (ed. Black JM), pp. 21–52. Oxford University Press, Oxford.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Griffith SC, Owens IPF, Thuman KA (2002) Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology*, **11**, 2195–2212.

Grus WE, Graves GR, Glenn TC (2009) Geographic variation in the mitochondrial control region of black-throated blue warblers (Dendroica caerulescens). *Auk*, **126**, 198–210.

Haas BJ, Papanicolaou A, Yassour M *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494–1512.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (Oncorhynchus nerka) population. *Molecular Ecology Resources*, **11**, 150–161.

Heaton MP, Harhay GP, Bennett GL *et al.* (2002) Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome*, **13**, 272–281.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, **11**, 123–136.

Hess JE, Campbell NR, Docker MF *et al.* (2015) Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, **15**, 187–202.

Holmes RT (2007) Understanding population change in migratory songbirds: long-term and experimental studies of Neotropical migrants in breeding and wintering areas. *Ibis*, **149**, 2–13.

Holmes RT (2011) Avian population and community processes in forest ecosystems: long-term research in the Hubbard Brook Experimental Forest. *Forest Ecology and Management*, **262**, 20–32.

Holmes RT, Rodenhouse NL, Sillett TS (2005) Black-throated blue warbler (Setophaga caerulescens), No. 87. In: *The Birds of North America Online* (ed. Poole A). The Cornell Lab of Ornithology, Ithaca.

Jetz W, Rubenstein DR (2011) Environmental uncertainty and the global biogeography of cooperative breeding in birds. *Current Biology*, **21**, 72–78.

Kaiser SA, Sillett TS, Risk BB, Webster MS (2015) Experimental food supplementation reveals habitat-dependent male reproductive investment in a migratory bird. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142523.

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.

Kess T, Gross J, Harper F, Boulding EG (2016) Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle Littorina saxatilis. *Journal of Molluscan Studies*, **82**, 104–109.

Krawczak M (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis*, **20**, 1676–1681.

Labuschagne C, Nupen L, Kotzé A, Grobler PJ, Dalton DL (2015) Assessment of microsatellite and SNP markers for parentage assignment in ex situ African Penguin (*Spheniscus demersus*) populations. *Ecology and Evolution*, **5**, 4389–4399.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Liu S, Palti Y, Gao G, Rexroad CE (2016) Development and validation of a SNP panel for parentage assignment in rainbox trout. *Aquaculture*, **452**, 178–182.

Lohse M, Bolger AM, Nagel A *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, **40**, W622–W627.

Mamanova L, Coffey AJ, Scott CE *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7**, 111–118.

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

Marth G, Yeh R, Minton M *et al.* (2001) Single-nucleotide polymorphisms in the public domain: how useful are they? *Nature Genetics*, **27**, 371–372.

Mason NA, Taylor SA (2015) Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, **24**, 3009–3025.

Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, **160**, 357–366.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Nielsen R, Hubisz MJ, Clark AG (2004) Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*, **168**, 2373–2382.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, **12**, 443–451.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

Ranade K, Chang MS, Ting CT *et al.* (2001) High-throughput genotyping with single nucleotide polymorphisms. *Genome Research*, **11**, 1262–1268.

Rengmark AH, Slettan A, Skaala Ø, Lie Ø, Lingaas F (2006) Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture*, **253**, 229–237.

Roberts JD, Byme PG (2011) Polyandry, sperm competition, and the evolution of anuran amphibians. *Advances in the Study of Behavior*, **43**, 1–53.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.

Schunter C, Garza JC, MacPherson E, Pascual M (2014) SNP development from RNA-seq data in a nonmodel fish: how many individuals are needed for accurate allele frequency prediction? *Molecular Ecology Resources*, **14**, 157–165.

Sellars MJ (2014) Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp. *Aquaculture*, **45**, 417–426.

Senn H, Ogden R, Cezard T *et al.* (2013) Reference-free SNP discovery for the Eurasian beaver from restriction site–associated DNA paired-end data. *Molecular Ecology*, **22**, 3141–3150.

Smit AFA, Hubley R, Green P (2015) *RepeatMasker Open-4.0*. Available from http://www.repeatmasker.org.

Smith SB, Webster MS, Holmes RT (2005) The heterozygosity theory of extra-pair mate choice in birds: a test and a cautionary note. *Journal of Avian Biology*, **36**, 146–154.

Strucken EM, Lee SH, Lee HK *et al.* (2016) How many markers are enough? Factors influencing parentage testing in different livestock populations. *Journal of Animal Breeding and Genetics*, **133**, 13–23.

Toews DPL, Campagna L, Taylor SA *et al.* (2015) Genomic approaches to understanding population divergence and speciation in birds. *Auk*, **133**, 13–30.

Tokarska M, Marshall T, Kowalczyk R *et al.* (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. *Heredity*, **103**, 326–332.

Uller T, Olsson M (2008) Multiple paternity in reptiles: patterns and processes. *Molecular Ecology*, **17**, 2566–2580.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, **34**, 275–305.

Wan D, Chang P, Yin J (2013) Causes of extra-pair paternity and its interspecific variation in socially monogamous birds. *Acta Ecologica Sinica*, **33**, 158–166.

Webster MS, Chuang-Dobbs HC, Holmes RT (2001) Microsatellite identification of extrapair sires in a socially monogamous warbler. *Behavioral Ecology*, **12**, 439–446.

Weinman LR, Solomon JW, Rubenstein DR (2014) A comparison of SNP and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, **15**, 502–511.

Westneat DF, Stewart IRK (2003) Extra-pair paternity in birds: causes, correlates, and conflict. *Annual Review of Ecology and Systematics*, **34**, 365–396.

Westneat DF, Sherman PW, Morton ML (1990) The ecology and evolution of extra-pair copulations in birds. *Current Ornithology*, **7**, 331–370.

White PS, Densmore LD (1992) Mitochondrial DNA isolation. In: *Molecular Genetic Analysis of Populations: A Practical Approach* (ed. Hoelzel AR), pp. 50–51. Oxford University Press, New York, NY.

S.A.K. designed the study, collected and analysed the data and drafted the manuscript. S.A.T. designed the study, assembled the transcriptome, prepared sequences for NCBI submission and helped to draft the manuscript. N.C. designed the study, performed SNP discovery, prepared sequences for NCBI submission and helped to draft the manuscript. T.S.S. helped design the study, secured funding and provided comments on the manuscript. E.R.B. helped with library preparation and provided comments on the manuscript. M.S.W. helped design the study, secured funding and provided comments on the manuscript. All authors read and approved the final manuscript.

Data accessibility

SNP-containing DNA sequences: NCBI dbSNP Accession nos (ss1996900653–ss1996900749). Raw (adapter and quality trimmed) Illumina transcriptome sequences: NCBI Short Read Archive (SRA Accession nos: SRR3164929, SRR3165127; PRJNA311681). Microsatellite and SNP genotypes: Dryad Digital Repository: doi:10.5061/dryad.1vn46.

Supporting Information

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

Table S1 List of forward, reverse, and extend primer sequences for 97 SNPs.

Table S2 Characteristics of six microsatellite markers used for parentage analyses.

Table S3 Input parameters for parentage analyses.

Table S4 Number of observed and expected CERVUS paternity assignments for all microsatellite and SNP panels.