NOTE / NOTE

A method to improve confidence in paternity assignment in an open mating system

M.M. Kasumovic, L.M. Ratcliffe, and P.T. Boag

Abstract: Molecular techniques have allowed researchers studying mating systems to determine the identity of extra-pair sires, providing more accurate measures of individual realized reproductive success. Yet, an existing problem in such studies is the inability to assign paternity to individuals that have not been captured. This frequently arises when only a proportion of the population is sampled or when visitors from outside the study area have access to the breeding population. It is therefore difficult to assign paternity in situations where not all candidate sires are sampled because some assignments may be incorrect, especially when using a likelihood-based approach. This study outlines a method that combines two different programs, GERUD 1.0 and CERVUS 2.0, to increase confidence in paternity assignment. The benefit of using these programs in conjunction is that GERUD 1.0 can reconstruct genotypes of males that are not sampled in families where the female was sampled, and CERVUS 2.0 can use this information to better assign paternity because more information is provided. We show how applying this method to Least Flycatchers (*Empidonax minimus*), a sub-oscine bird with an open mating system, substantially increases confidence in paternity assignments.

Résumé : Des techniques moléculaires ont permis aux chercheurs qui étudient les systèmes d'accouplement de déterminer l'identité des pères hors couple, ce qui fournit des mesures plus précises du succès reproducteur individuel réalisé. Néanmoins, un des problèmes dans de telles études est l'impossibilité d'attribuer la paternité à des individus qui n'ont pas été capturés. Une telle situation se produit lorsque seulement une partie de la population est échantillonnée ou lorsque des visiteurs provenant de l'extérieur de la zone d'étude ont accès à la population qui est en train de se reproduire. Il est ainsi difficile d'attribuer la paternité dans des situations où tous les pères possibles ne sont pas échantillonnés, car alors certaines attributions seront erronées, particulièrement si on utilise une méthode basée sur la vraisemblance. Nous avons mis au point une méthode qui combine deux logiciels, GERUD 1.0 et CERVUS 2.0, pour augmenter la confiance des attributions de paternité. L'avantage d'utiliser ces deux logiciels conjointement est que GERUD 1.0 peut reconstruire le génotype de mâles non échantillonnés dans les familles où la femelle a été échantillonnée et que CERVUS 2.0 peut utiliser cette information pour mieux attribuer la paternité, puisque plus de données sont disponibles. Nous montrons comment l'application de la méthode au moucherolle tchébec (*Empidonax minimus*), un oiseau suboscine à système d'accouplement ouvert, augmente de façon significative la fiabilité des attributions de paternité.

[Traduit par la Rédaction]

Introduction

Current behavioural research relies heavily on molecular paternity analyses because molecular techniques are able to inform researchers about behaviours that are difficult to observe. With the advent of high-resolution DNA markers, a suite of different analytical approaches has been developed to analyze molecular information and infer paternity (Danzmann 1997; Marshall et al. 1998; Neff et al. 2000). Identifying whether a litter or brood contains extra-pair offspring is relatively simple if there are clear mismatches between genotypes, but identifying the parents of such extra-pair offspring can be challenging. Without accurate paternity assignment, it is difficult to determine individual realized reproductive success. This problem can be particularly pronounced in systems where social pair bonds are weak or where copulations outside the pair bond are common and include individuals from outside the population under study (open breeding system).

Measuring realized reproductive success is particularly challenging in bird species with open breeding systems where individuals are difficult to capture or are highly mobile, visiting several sites in a single day. For example, determining the number of candidate sires is not straightfor-

Received 17 March 2003. Accepted 15 October 2003. Published on the NRC Research Press Web site at http://cjz.nrc.ca on 23 January 2004.

M.M. Kasumovic,[1,2] L.M. Ratcliffe, and P.T. Boag. Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada.

[1]Corresponding author (e-mail: m.kasumovic@utoronto.ca).
[2]Present address: Division of Life Sciences, University of Toronto at Scarborough, Toronto, ON M1C 1A4, Canada.

 doi: 10.1139/Z03-198

ward if many individuals visiting a population are not observed or if observed individuals are not sampled. This study demonstrates a method to increase the confidence in paternity assignment in an open system using already available analytical techniques. We show that combining two different paternity programs, GERUD 1.0 (Jones 2001) and CERVUS 2.0 (Marshall et al. 1998), substantially improves paternity assignment in the Least Flycatcher (*Empidonax minimus*), a migratory sub-oscine with an open mating system.

Methods

Paternity data were collected from two separate Least Flycatcher populations (one site in 2000 and another site 18 km away in 2001) at the Queen's University Biological Station at Chaffey's Lock, Ont. (44°34′N, 76°19′W). The Least Flycatcher is a socially monogamous bird species that settles contiguously on small all-purpose territories, forming clusters that resemble leks (Briskie 1994; Tarof 2001). Least Flycatcher pairs maintain their territories throughout the breeding season, but individuals from other sites will visit clusters throughout the breeding season, resulting in an open mating system. Previous research shows that 61.9% of broods contain extra-pair offspring (Tarof 2001). In eastern Ontario, Least Flycatchers breed in the upper canopy and are difficult to capture, typically resulting in incomplete sampling within a cluster.

We studied a total of two clusters. In 2000, 75% (15/20) of the clustered individuals and two other nonterritorial males were sampled. In 2001, 79% (15/19) of the clustered individuals and one other nonterritorial male were sampled. Nonterritorial males were considered to be visitors to the clusters, since they were never seen before or after the day of capture. Thirty-seven offspring were sampled from 10 nests in 2000: in 5 nests, both parents were sampled; in 3 nests, only the males were sampled; and in 2 nests, only the female was sampled. In the 2001 cluster, 42 offspring were sampled from 11 nests: in 7 nests, both parents were sampled; in 2 nests, only the male was sampled; and in 2 nests, only the female was sampled.

Species-specific microsatellites were used for paternity analysis (Tarof et al. 2001). Table 1 presents levels of heterozygosity and allele numbers for both populations. The total exclusion probability with all four loci and both parents known, as calculated by CERVUS 2.0, was 99.2% in 2000 and 98.5% in 2001. We assumed no egg dumping occurred, since there were no genotype mismatches between mothers and their offspring. GERUD 1.0 was used first to analyze the microsatellite data and then CERVUS 2.0 was used to assign paternity.

GERUD 1.0 is a computer program that reconstructs unknown parental genotypes by subtracting the known parental genotype from the known progeny array (Jones 2001). It can be downloaded from http://www.bcc.orst.edu/~jonesa/. Since the female is the known parent in this study, the unknown reconstructed genotypes are those of males. After determining the possible paternal genotypes, the program uses information from patterns of segregation of alleles in the progeny as well as allelic frequencies in the adult population to determine which of the genotypic combinations are most probable. Using this program, paternal genotypes can be reconstructed when the maternal genotype is known and then compared with the observed genotypes of males within the cluster to determine paternity. Furthermore, GERUD 1.0 can reconstruct the genotypes of males in the cluster that have not been sampled. Care must be taken, since GERUD 1.0 assumes perfect genotyping of all individuals.

Table 1. The expected heterozygosity and allele number of sampled individuals from two clusters of Least Flycatchers (*Empidonax minimus*) breeding in 2000 and 2001.

	2000		2001	
Locus	H_e	N_a	H_e	N_a
Z27	0.938	14	0.881	11
Z1	0.854	9	0.873	9
C23	0.829	9	0.851	7
D46	0.848	8	0.756	7

Note: H_e is the expected heterozygosity and N_a is the number of alleles at the locus. Only adults were used in this calculation. n = 17 in 2000 and n = 16 in 2001.

Since GERUD 1.0 determines extra-pair paternity through clutch information, the program was initially used to examine each clutch and determine the minimum number of fathers required to sire the offspring within the nest and therefore the number of nests that contained extra-pair young. Secondly, GERUD 1.0 was used to reconstruct all the paternal genotypes in nests within the cluster where the female was sampled. The reconstructed genotypes were then compared with the actual genotypes of sampled males. GERUD 1.0 also reconstructed the genotypes of males that sired offspring in nests where the social male was not sampled. In these cases, we assumed that the male for which the genotype was reconstructed was the unsampled social male, since the reconstructed genotypes did not match those of other sampled males and it is more likely that the social male, rather than an unsampled, unobserved visiting male, was the father.

After completing the analysis using GERUD 1.0, CERVUS 2.0 was used to assign paternity. CERVUS 2.0 is a paternity assignment program that uses a likelihood-based approach and allelic frequencies to determine which one of the sampled individuals is the most likely paternal sire (Marshall et al. 1998). CERVUS 2.0 can be downloaded from http://helios.bto.ed.ac.uk/evolgen/cervus/cervus.html. Paternity assignments occur at the individual level rather than the family level, i.e., each offspring is treated as an individual parentage test. Unlike GERUD 1.0, CERVUS 2.0 takes account of scoring errors and mutations and requires information only from candidate parents, though assignments may have greater confidence if the known parent is used. CERVUS 2.0 performs well in situations where the majority of potential sires is known, though confidence in assignments decreases as the proportion of candidate sires that are sampled also decreases (Slate et al. 2000).

CERVUS 2.0 was used in two different trials. The first trial was performed to analyze the paternity using only sampled individuals from the clusters. Paternity was then analyzed using the genotypes of sampled individuals plus all the genotypes reconstructed by GERUD 1.0 (i.e., genotypes of those individuals that were not sampled). Using the recon-

structed genotypes of social males, we could assign paternity to nests and also determine whether the male gained any extra-pair paternity from other nests within the cluster. For each trial, the number of candidate sires was calculated as the number of males settled in the cluster plus the number of unknown males seen visiting the cluster and the number of males seen within 500 m of the cluster. Therefore, the proportion of males sampled was calculated by dividing the number of males sampled by the number of candidate sires, giving conservative estimates of 65% in 2000 and 50% in 2001. We assumed a 1% mistyping rate and assigned paternity at both a strict confidence level of 95% and a relaxed confidence level of 80% (Marshall et al. 1998). Results of the two trials performed with CERVUS 2.0 were compared between trials and with the results from GERUD 1.0. Assignments were then checked manually to insure there were no mismatches.

Results

Over both years, a total of 12 nests had both the male and the female sampled, and the results of the analysis performed with GERUD 1.0 agreed with the paternity assignments obtained with CERVUS 2.0 for all of the 48 offspring from these 12 nests (four of which were extra-pair) both with and without using reconstructed genotypes. In the five nests where only males were sampled, CERVUS 2.0 assigned paternity only to sampled males, even when all reconstructed male genotypes were used. Paternity was assigned to the social males for 14 of 18 offspring from these five nests and to neighbouring males for the other 4 offspring.

In the other four nests, only the females were sampled and data from social males were not available. When CERVUS 2.0 alone was used to assign paternity, paternity was assigned to different males throughout the cluster at low confidence (i.e., below 80%) for 9 of the 13 offspring. The remaining four offspring (from two different nests) were assigned to neighbouring males at 95% confidence. Since CERVUS 2.0 cannot assign paternity to males that have not been sampled, GERUD 1.0 was used to reconstruct the genotypes of the social males. When the reconstructed genotypes were used, CERVUS 2.0 assigned paternity of two of the four nests (six offspring) solely to the reconstructed males with 95% confidence. For the other two nests (four offspring), paternity was still assigned to the two sampled neighbours with 95% confidence. Two of the three remaining offspring were assigned to the reconstructed social males with 95% confidence. For the final assignment, the genotype that GERUD 1.0 reconstructed did not match the genotype of any sampled individual or any of the reconstructed genotypes of males from other families. Therefore, the final assignment was assumed to be a visiting male that was not sampled or observed. The CERVUS 2.0 analysis supported this assignment when all reconstructed genotypes were used. Since having a single social sire or two sires (the unsampled social male and a neighbouring sampled male) is more parsimonious than having multiple neighbouring sires, and since the GERUD 1.0 assignments agreed with the CERVUS 2.0 assignments, paternity was assigned to the social males that were not sampled.

Discussion

All reconstructions of genotypes of sampled males by GERUD 1.0 matched the CERVUS 2.0 assignments. Genotypes of a total of three males that were not sampled but were observed as social males were reconstructed, and one male that was neither sampled nor observed was classified as a visitor to the cluster. CERVUS 2.0 correctly assigned paternity to all the males for which genotypes were reconstructed when the reconstructed genotypes from GERUD 1.0 were used. CERVUS 2.0 also assigned paternity to five social males whose female partner was not sampled and could therefore not be used in GERUD 1.0. Without the use of GERUD 1.0, paternity could not be assigned to four nests (13 offspring; 4 extra-pair) and without the use of CERVUS 2.0, paternity could not be assigned to five nests (18 offspring; 4 extra-pair). Therefore, by combining both programs, the success rate of paternity assignment increased by 16.5% compared with the use of GERUD 1.0 alone and by 22.8% compared with the use of CERVUS 2.0 alone. These results suggest that these particular breeding clusters of Least Flycatchers exhibit moderate levels of extra-pair paternity (16.5% EP nestlings) and that EPY are sired predominantly (92.3%), though not exclusively (7.7%), by other males residing in the same cluster.

Clearly, these specific values should be interpreted with caution, since they are contingent on our small sample sizes and specified assumptions. Nevertheless, we conclude that combining both programs substantially improved confidence in paternity assignment, since results obtained with both programs were in agreement and clarified paternity assignments in situations where one of two social parents was not sampled. Using both programs worked better than simply using either one alone, since GERUD 1.0 alone could not assign paternity in cases where the female was not sampled, and CERVUS 2.0 itself lacked data from males that were not sampled and therefore assigned paternity incorrectly to neighbouring males at low confidence. Using both programs together allowed CERVUS 2.0 to use the data made available by GERUD 1.0 to better estimate the reproductive success of both sampled and unsampled males.

In addition, there are factors that could increase the validity of assignments while using this approach. Confidence in male genotypic reconstructions and paternity assignments will increase if a greater number of molecular markers are used and (or) a greater number of offspring are sampled per brood, since this increases the accuracy of paternal genotypic reconstructions and decreases the probability of assigning paternity incorrectly. However, even with this approach, it is still difficult to assign paternity in situations where both parents are unknown, since male genotypes cannot be reconstructed and paternity is therefore assigned at very low confidence.

Overall, we suggest that this approach can provide better information on the distribution of paternity throughout populations, where incomplete sampling of parents is common and unavoidable. This improved information can aid in the determination of the level of extra-pair paternity within a population and the proportion of offspring that are sired by visiting males, rather than residents. Overall, this approach may improve the understanding of variability in realized reproductive success between males.

Acknowledgments

We would like to especially thank Liam Colgan, Rachel Gardner, Celine Muis, and Scott Tarof for helping to collect data, and Adam Jones, Steve Lougheed, Tristan Marshall, and two anonymous reviewers for helpful comments on the manuscript. This research was supported by a student award from the American Ornithologists' Union and the Taverner Award from the Society of Canadian Ornithologists to M.M.K., and grants from the Natural Sciences and Engineering Research Council of Canada to L.M.R. and P.T.B.

References

Briskie, J.V. 1994. Least Flycatcher (*Empidonax minimus*), no. 99. *In* The birds of North America. *Edited by* A. Poole and F. Gill. The Academy of Natural Sciences, Philadelphia, Pa.; The American Ornithologists' Union, Washington, D.C. pp. 1–21.

Danzmann, R.G. 1997. PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from known pools of parents and progeny. J. Hered. **88**: 333–334.

Jones, A.G. 2001. GERUD1.0: a computer program for the reconstruction of parental genotypes from progeny arrays using multilocus DNA data. Mol. Ecol. Notes, **1**: 215–218.

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.

Neff, B.D., Repka, J., and Gross, M.R. 2000. Parentage analysis with incomplete sampling of candidate parents and offspring. Mol. Ecol. **9**: 515–528.

Slate, J., Marshall, T., and Pemberton, J. 2000. A retrospective assessment of the accuracy of the paternity inference program CERVUS. Mol. Ecol. **9**: 801–808.

Tarof, S.A. 2001. Spatial dynamics and evolutionary explanations of clustering behaviour in least flycatchers *Empidonax minimus*. Ph.D. thesis, Queen's University, Kingston, Ont.

Tarof, S.A., Ratcliffe, L.M., and Boag, P.T. 2001. Polymorphic microsatellite loci for assigning parentage in least flycatchers (*Empidonax minimus*). Mol. Ecol. Notes, **1**: 146–148.