Journal of Avian Biology

J5572
Woltmann, S., Sherry, T. W. and Kreiser, B. R. 2012. A genetic approach to estimating natal dispersal distances and self-recruitment in resident rainforest birds. – J. Avian Biol. 43: 33–42.

Supplementary material

Appendix 1

Methods

Parentage Analyses – empirical data

ML-RELATE (Kalinowski et al. 2006) is a widely used program that allows for some of the inherent limitations of our study system: (1) there is no way of estimating the proportion of candidate parents sampled, and (2) individuals can only be aged as < or ≥ 1-year-old. The first point precludes the use of another widely used program CERVUS (Marshall et al. 1998), which requires an estimate of the proportion of candidate parents sampled. Unfortunately, when the focal animals occupy large or poorly-defined areas, defining which individuals could reasonably be considered candidate parents requires prior knowledge of typical dispersal distances. For example, the chestnut-backed antbird occurs more or less contiguously from Nicaragua to Panama, yet few would consider a bird sampled in Panama as a "potential candidate" for parentage of an individual sampled in northern Costa Rica. The second point precludes effective use of approaches like FAMOZ (Gerber et al. 2003) or a more recent approach proposed by Christie (2010), in which juveniles (i.e. offspring to be assigned to parents) are separated from adults (potential parents) in the dataset. Chestnut-backed antbirds are relatively long-lived, and acquire "adult" (definitive) plumage by year one (Wolfe et al. 2009, Woltmann and Sherry 2011); parent-offspring relationships could well exist among birds classified as adult. Not testing for PO dyads within our adult samples could thus lead to an underestimation of self-recruitment.

Multiple approaches can be used to quantify and mitigate errors in parentage analyses. We follow Christie (2010) in considering Type I error as falsely concluding a relationship exists (i.e. false-positives), and Type II error as falsely concluding a relationship does not exist. Type I

errors can be caused by linkage disequilibrium and limited marker power (too few loci, low allele richness). We minimized Type I errors by selecting a suite of loci that showed no evidence of linkage disequilibrium, and each locus had at least 4 alleles. We tested whether false-positive relationships were likely in our datasets via simulations (see below). Null alleles, scoring error, and mutation can cause Type II errors. We did not include loci with estimates of null allele rates $\geq$ 2% (MICRO-CHECKER; van Oosterhout et al. 2004), and quantified genotyping error by reamplifying and rescoring a subset of individuals. ML-RELATE does not accommodate mutation or scoring error, and thus a single mismatch precludes the possibility of PO (Marshall et al. 1998, Gerber et al. 2003). To allow for mutation, we tentatively considered two individuals parent-offspring if they were incompatible at only a single locus (out of 15). For these dyads, we treated the data at the mismatched locus from one of the individuals as missing, and re-ran the analysis to determine whether the program considered the dyad PO. For each putative PO dyad, we used the "Specific Hypothesis Test" feature of ML-RELATE to test the relative likelihood of the dyad actually being FS (the relationship class most likely to be misidentified as PO; Blouin et al. 1996). We used 10,000 permutations to evaluate significance, and accepted $P < 0.10$ as significant.

Parentage Analyses – Monte Carlo simulations

To assess the likelihood of our dataset containing false-positive relationships, we quantified the probability of any one dyad's relationship arising by chance sharing of alleles (indicating the true parent or offspring was unsampled). We used GENETIX (Belkhir et al. 2004) to create simulated populations with identical allele frequencies as our data. This is similar in principle to the approach used by the program FAMOZ (Gerber et al. 2003). These simulated populations were

analysed with ML-RELATE to quantify how often various "relationships" (i.e. false positives) could arise by chance in our dataset.

Based on initial trials, simulated datasets of at least two orders of magnitude larger (in terms of number of dyads) were needed to detect measurable false positive PO dyads. For the simulated pooled (5-population) dataset, we created 10 populations of 500 individuals (1,247,000 dyads). For the simulated individual site datasets we created 5 populations each of either 200 (sites PL, QG, TI, RF; 99,500 dyads) or 400 individuals (site LS; 399,000 dyads). The size of the simulated datasets reflected a balance between the need to generate a large enough dataset to detect false positives and the need to keep computing time reasonable. All simulated datasets were analysed with ML-RELATE, and we averaged the number of randomly generated PO, FS and HS dyads. Mean per-dyad error rates were calculated from the simulated datasets, and were applied to our observed data to estimate the expected number of false-positive dyads in our analyses.

Results

Parentage Analyses – Error rates based on Monte Carlo simulations

Although ML-RELATE identified no cross-site PO dyads, the program identified a few cross-site FS, and many cross-site HS in the pooled dataset. In simulations, putative FS and HS dyads were detected considerably more frequently than PO, indicating greatly reduced (FS, mean error ~ 32%) or zero (HS, mean error ~ 160%) confidence in our ability to correctly identify these relationships with these data and this program (Table 2). These are not entirely unexpected results - the ability to identify non-PO relationships consistently and correctly typically requires much more genetic information than in our study (Blouin 2003). Given our field sampling effort and limited genealogical information about individuals, we conclude that our genetic data had

sufficient power to identify PO dyads, but could not reliably identify FS or HS relationships. We thus consider the few (9 of 99 putative FS identified) cross-site putative FS dyads in the analysis of our pooled dataset inconclusive evidence of between-site migration, and consider all putative HS dyads completely uninformative.

Supplemental References

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme, F. 2004. GENETIX, logiciel sous WindowsTM pour la génétique des populations, Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Blouin, M. S., Parsons, M., Lacaille, V. and Lotz, S. 1996. Use of microsatellite loci to classify individuals by relatedness. - Mol. Ecol. 5: 393-401.

Christie, M. R. 2010. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. - Mol. Ecol. Res. 10: 115-128.

Gerber, S., Charbrier, P. and Kremer, A. 2003. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. - Mol. Ecol. Notes 3: 479-481.

Kalinowski, S. T., Wagner, A. P. and Taper, M. L. 2006. ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. - Mol. Ecol. Notes 6: 576-579.

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

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. - Mol. Ecol. Notes 4: 535-538.

Wolfe, J. D., Pyle, P. and Ralph, C. J. 2009. Breeding seasons, molt patterns, and gender and age criteria for selected northeastern Costa Rican resident landbirds. - Wilson J. Ornithol.121: 556-567.

Woltmann, S. and Sherry, T. W. 2011. High apparent survival and stable territory dynamics of Chestnut-backed Antbird (*Myrmeciza exsul*) in a large rain forest preserve. - Wilson J. Ornithol. 123: 15-23.

Table A1. Number of alleles (N_A) and observed heterozygosity (H_O) of 15 microsatellite loci in chestnut-backed antbird at five sites in Costa Rica. Sample sizes (individuals) are given below each site name.

		H_O				
		La Selva	Tirimbina	Plastico	Quebrada Gonzales	Rio Frio
locus	N_A	n = 137	n = 34	n = 29	n = 20	n = 35
HyNa06	18	0.934	0.971	1.000	0.850	0.943
Mex034	12	0.898	0.882	0.828	0.900	0.829
Mex080	6	0.781	0.676	0.862	0.600	0.629
Mex140	16	0.883	0.882	0.897	1.000	0.914
Mex162	7	0.745	0.706	0.690	0.850	0.743
Mex176	15	0.774	0.735	0.690	0.850	0.800
Mex178	9	0.810	0.824	0.821	0.850	0.771
Mex191	12	0.847	0.824	0.828	0.950	0.829
MyEx19	17	0.876	0.941	0.862	0.900	0.886
MyEx20	4	0.693	0.765	0.759	0.600	0.600
MyEx27	12	0.686	0.471	0.690	0.700	0.657
MyEx41	10	0.831	0.824	0.862	0.900	0.629
MyEx46	4	0.610	0.706	0.724	0.750	0.514
MyEx52	26	0.933	0.912	0.828	0.950	0.912
MyEx61	11	0.832	0.882	0.862	0.650	0.886