J. Avian Biol. 39: 252–256, 2008
doi: 10.1111/j.2008.0908-8857.04094.x

Received 18 October 2006, accepted 18 June 2007

The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*

Christin L. Pruett and Kevin Winker

C. L. Pruett (correspondence) and K. Winker, University of Alaska Museum and Institute of Arctic Biology, 907 Yukon Dr., Fairbanks, Alaska 99775, USA. Present address of C. L. P.: Sutton Avian Research Center, University of Oklahoma, PO Box 2007, Bartlesville, Oklahoma, 74005, USA. E-mail: cpruett@ou.edu

To empirically determine the effects of sample size on commonly used measures of average genetic diversity, we genotyped 200 song sparrows *Melospiza melodia* from two populations, one genetically depauperate (n = 100) and the other genetically diverse (n = 100), using eight microsatellite loci. These genotypes were used to randomly create 10,000 datasets of differing sizes (5 to 50) for each population to determine what the effects of sample size might be on several estimates of genetic diversity (number of alleles per locus, average observed heterozygosity, and unbiased average expected heterozygosity) in natural populations of conservation concern. We found that at small sample sizes of 5 to 10 individuals, estimates of unbiased heterozygosity outperformed those based on observed heterozygosity or allelic diversity for both low- and high-diversity populations. We also found that when comparing across populations in which different numbers of individuals were sampled, rarefaction provided a useful way to compare estimates of allelic diversity. We recommend that standard errors should be reported for all diversity estimators, especially when sample sizes are small. We also recommend that at least 20 to 30 individuals be sampled in microsatellite studies that assess genetic diversity when working in a population that has an unknown level of diversity. However, research on critically endangered populations (where large sample sizes are impossible or extremely difficult to obtain) should include measures of genetic diversity even if sample sizes are less than ideal. These estimates can be useful in assessing the genetic diversity of the population.

Conservation geneticists often use microsatellite loci as genetic tools to estimate population statistics such as heterozygosity and allelic diversity and to determine whether a population was bottlenecked (Beaumont and Bruford 2000, Frankham et al. 2002). Large sample sizes or large numbers of loci (Nei 1978, Gregorius 1980, Carvalho and Hauser 1994, Baverstock and Moritz 1996, Ruzzante 1998, Toro et al. 2002, Kalinowski 2005) are recommended for the estimation of genetic statistics using diploid markers such as microsatellites. However, in threatened and endangered vertebrate populations it can be difficult to obtain large sample sizes to accurately estimate genetic diversity, because individuals are often dispersed and difficult to locate (e.g., large home ranges, limited habitat availability) or are found in remote areas that are expensive to reach and difficult to access (e.g., island populations). In addition, in non-model systems large numbers of microsatellite loci are often not available. Although modeling is a useful approach to determine the efficacy of sampling methods, empirical studies are necessary to confirm theoretical predictions.

Theory predicts that some diversity estimators perform better than others at small sample sizes. Unbiased estimates of expected heterozygosity (H_e) are expected to have less bias, more precision, and greater accuracy than estimates based on observed heterozygosities (H_o; Nei 1978, 1987). Also, estimates of alleles per locus at small sample sizes can be greatly biased, especially when compared to populations from which a larger sample size is obtained (Petit et al. 1998). One way to alleviate this problem is through rarefaction to the smallest sample size (Petit et al. 1998, Leberg 2002, Kalinowski 2004). Based on the genotypes of 100 song sparrows *Melospiza melodia* from Attu and adjacent islands, we sought to empirically test these theories and to determine how much confidence researchers should have when estimates are derived from limited numbers of samples in a genetically depauperate population. In addition, we sought to provide guidance to researchers on whether or not it is necessary to sample more extensively in an area when sampling goals are not met. We also examined the effect of small sample size on genetic estimates in a genetically diverse dataset. We then compared these estimates with those from the low-diversity population to see whether or not estimates differed between the datasets. We tested to see whether these datasets fit theoretical predictions of higher variances in more diverse populations (Nei 1978). Our goal was to glean some general sampling guidelines from these comparisons.

Materials and methods

Whole genomic DNA from the tissues of 200 song sparrows from Attu Island (n = 84) and nearby Shemya and Nizki islands (n = 16), Hyder (n = 18), Alexander Archipelago (n = 30), and Copper River Delta (n = 28) in Alaska, and Queen Charlotte Islands, British Columbia (n = 24), were extracted following Glenn (1997). Samples from Attu, Shemya, and Nizki Islands were grouped into the low genetic diversity population, and samples from Hyder, Alexander Archipelago, Copper River Delta, and Queen Charlotte Islands were grouped into the high genetic diversity population. We recognize that the high-diversity group does not represent a single population (Pruett and Winker 2005), but it was necessary to group several locations to obtain a sample size equivalent to that of the low-diversity population. Samples were obtained between April and November over several years. Known or suspected parent-offspring or sibling relatives were excluded. Eight microsatellite loci were amplified for all individuals using fluorescent dye-labeled primers developed for song sparrows (Mme1, 2, 3, 7, 8, 12; Jeffery et al. 2001) and for two other bird species (Escu1; Hanotte et al. 1994, GF05; Petren 1998), and were then genotyped using an ABI 373A or 3100 automated sequencer. These loci are all highly polymorphic in other song sparrow populations (Keller et al. 2001, Chan and Arcese 2002, Pruett and Winker 2005). Five to ten loci are often used in the estimation of conservation genetic parameters of natural populations that have not been studied extensively as model systems (e.g., Edwards et al. 2004, Mock et al. 2004, Schwartz et al. 2005).

For both populations, we calculated commonly used diversity statistics in conservation genetics, including the number of alleles per locus, average observed heterozygosity (H_o), and unbiased average expected heterozygosity (H_e). Methods for calculating the number of alleles per locus and average H_o are provided in Frankham et al. (2002). Average unbiased H_e was calculated using equation 8.4 in Nei (1987). These values calculated from the full sample of 100 individuals for each population were treated as the actual or known value for that group. We then created datasets of different sample sizes (5, 10, 20, 30, 40, and 50) by randomly sampling (without replacement) individuals from each population dataset (low and high diversity) of 100 song sparrows using Microsoft Excel 2002 (Microsoft Co., Seattle). 10,000 random datasets were created for each sample size.

To determine the relative bias and precision for each sample size, we calculated the mean of the 10,000 simulations for each dataset. We calculated standard errors for each simulated dataset and then averaged these values across all simulations. This error measurement is commonly reported in conservation genetics literature (Ciofi et al. 1999, Akst et al. 2002, Oyler-McCance et al. 2005). Unbiased standard errors were determined by calculating the variance of the mean heterozygosity (Nei 1987, equations 8.1–8.8).

We also examined the accuracy (how close the estimator is to the true value) of estimates by calculating the scaled root mean square error (SRMSE) for each sample size for both populations (Walther and Moore 2005). We tested how well rarefaction of alleles from the empirical 100-individual datasets to each sample size (5, 10, 20, 30, 40, 50) corresponded to the mean of the 10,000 random datasets (Petit et al. 1998). Rarefaction is commonly used to compare allelic diversity across unequal sample sizes (Petit et al. 1998, Leberg 2002).

Results

For the low-diversity population, actual values of genetic diversity estimates based on 100 individuals were substantially lower than those in the high-diversity population (Table 1). The average number of alleles per locus for the low-diversity population was 3.50 and for the high-diversity population 15.38. Heterozygosity values for the high-diversity population ($H_e = 0.792$, $H_o = 0.793$) were much higher than for the low diversity population ($H_e = 0.223$, $H_o = 0.189$).

For all of the sample sizes, the standard error range encompassed the actual value for high- and low-diversity populations for unbiased heterozygosity estimates and for

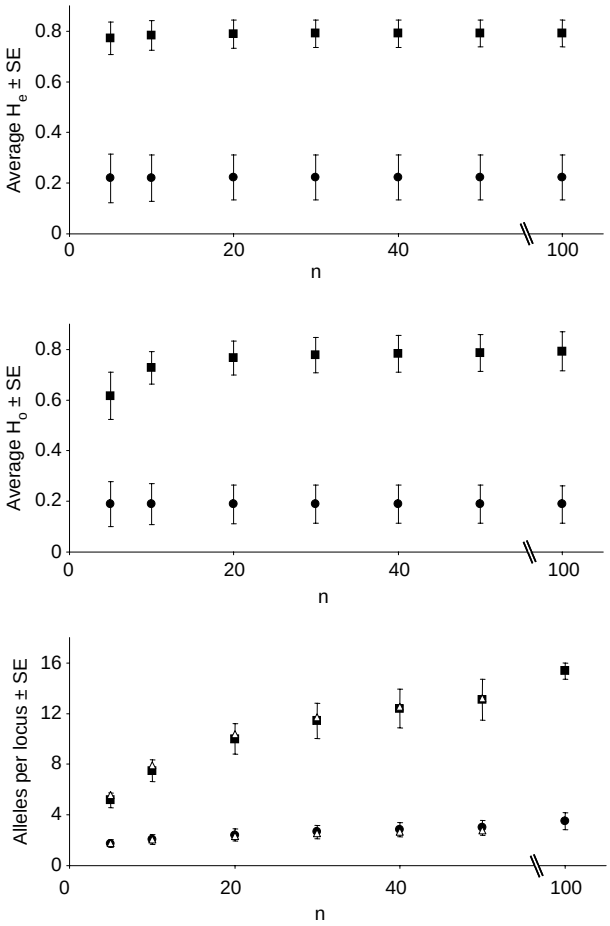

Fig. 1. Mean and standard error (SE) for estimates of average unbiased expected heterozygosity (H_e), average observed heterozygosity (H_o), and alleles per locus for low- (dark circles) and high-diversity (dark squares) populations of song sparrows. Values are based on 10,000 random datasets for various sample sizes. Estimates of allelic richness (white triangles), based on rarefaction from a sample of 100 for each sample size, are also provided in comparison with the alleles per locus means.

Table 1. Allele frequencies for eight microsatellite loci from low- and high-genetic diversity song sparrow populations.

Locus/Alleles	Low	High	Locus/Alleles	Low	High	Locus/Alleles	Low	High
Mme1			Mme3			Mme12		
132	0.000	0.039	160	0.010	0.000	182	0.010	0.005
134	0.085	0.000	162	0.000	0.104	188	0.900	0.629
136	0.040	0.005	164	0.000	0.004	200	0.000	0.033
138	0.220	0.010	170	0.000	0.071	206	0.005	0.028
144	0.000	0.164	172	0.990	0.659	212	0.080	0.061
146	0.645	0.198	174	0.000	0.004	218	0.000	0.094
148	0.010	0.080	176	0.000	0.028	224	0.005	0.074
150	0.000	0.011	178	0.000	0.063	230	0.000	0.036
152	0.000	0.255	180	0.000	0.052	236	0.000	0.040
154	0.000	0.011	182	0.000	0.011	Escu1		
156	0.000	0.005	188	0.000	0.004	128	0.000	0.004
158	0.000	0.208	Mme7			132	0.000	0.024
160	0.000	0.004	100	0.000	0.035	134	0.000	0.232
164	0.000	0.013	114	0.000	0.070	138	0.515	0.107
Mme2			116	0.000	0.007	140	0.000	0.008
120	0.000	0.004	118	0.000	0.023	142	0.000	0.000
124	0.000	0.025	120	0.000	0.065	144	0.010	0.096
126	0.000	0.004	122	1.000	0.137	146	0.000	0.132
128	0.000	0.004	124	0.000	0.033	148	0.395	0.175
136	0.000	0.026	126	0.000	0.186	150	0.080	0.032
138	0.000	0.300	128	0.000	0.093	152	0.000	0.140
140	0.000	0.040	130	0.000	0.181	154	0.000	0.043
142	0.730	0.122	132	0.000	0.064	156	0.000	0.010
144	0.030	0.015	134	0.000	0.044	GF05		
146	0.000	0.033	136	0.000	0.050	184	0.000	0.004
148	0.000	0.021	138	0.000	0.011	186	0.000	0.025
150	0.010	0.101	140	0.000	0.004	192	0.000	0.004
152	0.000	0.057	Mme8			194	0.000	0.057
154	0.000	0.030	201	0.000	0.014	196	0.000	0.033
156	0.000	0.025	205	0.000	0.008	198	0.000	0.188
158	0.000	0.004	207	0.000	0.004	200	0.000	0.023
160	0.000	0.011	208	0.000	0.021	202	0.000	0.158
162	0.000	0.087	210	0.000	0.119	206	0.000	0.024
164	0.000	0.064	211	0.000	0.007	208	0.000	0.021
166	0.000	0.007	213	1.000	0.162	210	0.000	0.043
172	0.000	0.021	215	0.000	0.103	212	0.960	0.057
212	0.230	0.000	217	0.000	0.256	214	0.000	0.025
			218	0.000	0.027	216	0.010	0.049
			219	0.000	0.107	218	0.000	0.074
			220	0.000	0.004	220	0.000	0.030
			221	0.000	0.009	222	0.000	0.046
			223	0.000	0.030	224	0.000	0.048
			224	0.000	0.000	226	0.000	0.045
			225	0.000	0.021	228	0.000	0.020
			227	0.000	0.026	232	0.030	0.021
			228	0.000	0.004	234	0.000	0.004
			229	0.000	0.069	238	0.000	0.004
			234	0.000	0.011			

the H_o estimates of the low-diversity population (Fig. 1). At $n = 5$ for the high-diversity population, the standard error did not overlap the actual value (Fig. 1), but it did overlap for the larger sample sizes. Standard error values for the alleles per locus did not include the actual value until sample sizes reached 40 to 50 individuals (Fig. 1). However, rarefied estimates of alleles overlapped the simulated values for all sample sizes (Fig. 1).

As expected from theory (Nei and Roychoudhury 1974, Nei 1978), deviations from the actual value for all genetic diversity parameters were largest at small sample sizes ($n = 5$–10 individuals; Fig. 2). The accuracy of H_e and H_o estimates at small sample sizes was better in higher than lower diversity populations. Deviations from the true value of alleles per locus were similar for high- and low-diversity populations across all sample sizes, with estimates from the low-diversity population performing slightly better (Fig. 2). In all instances, low- and high-diversity populations had estimates that began to approach one another at sample sizes > 20 (Fig. 2).

Discussion

The song sparrows of Attu Island and nearby Nizki and Shemya islands provide the opportunity for a real-world assessment of how differences in sampling schemes can affect the performance of genetic diversity estimates in genetically depauperate vertebrate populations. Our empirical results corroborate theory. Unbiased estimators perform better at low sample sizes, and rarefaction provides a useful way to compare unequal sample sizes, even when

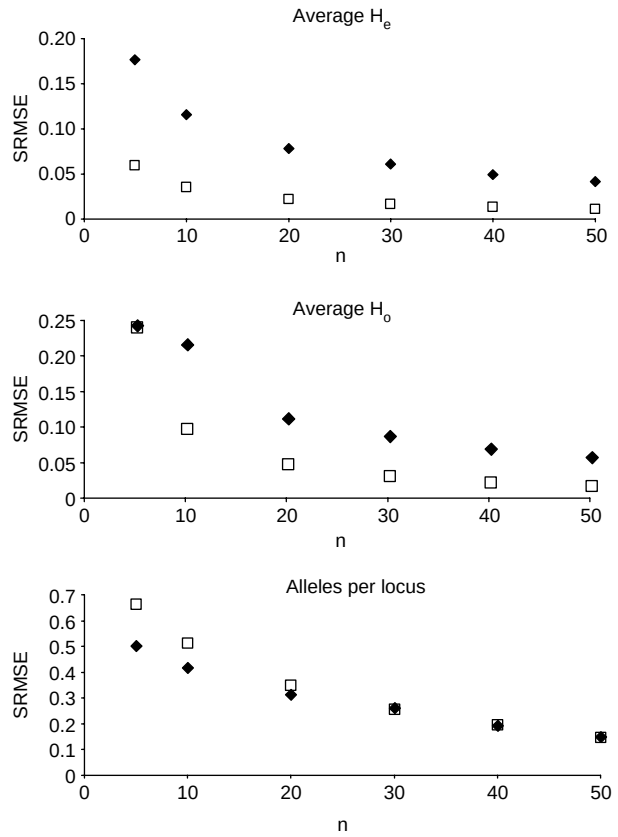

Fig. 2. Accuracy of average unbiased expected heterozygosity (H_e), average observed heterozygosity (H_o), and alleles per locus estimates for low- (black diamonds) and high-diversity (white squares) populations of song sparrows measured using scaled root mean square errors (SRMSE). Values are based on 10,000 random datasets for various sample sizes (n).

one sample is substantially smaller than another. Error at small sample sizes when unbiased average H_e and average H_o are estimated is primarily caused by a lack of accuracy in both low- and high-diversity populations. The mean and standard error encompass the true value for average H_e and H_o estimates in most instances (except for $n = 5$ for H_o in the high-diversity population). We suggest that researchers should report standard errors along with average heterozygosities. It is likely that the error range will overlap the true value even at smaller sample sizes (5 to 10), especially when unbiased estimates of H_e are used. Thus, unbiased average H_e and the standard error of these values should be useful in conservation assessments, even when sampling goals are not achieved.

Allelic diversity is thought to reflect the long-term evolutionary potential of a population better than heterozygosity (Allendorf 1986, Petit et al. 1998, Leberg 2002). However, at samples of less than 40, the number of alleles per locus is a poor measure of diversity. These findings correspond well with theory (Sjögren and Wynoni 1994, Petit et al. 1998, Haavie et al. 2000). Error is caused by sampling bias, a lack of precision, and inaccuracy. This leads one to wonder whether or not allelic-based estimators can provide useful measures of diversity. We recommend the use of average unbiased H_e instead of number of alleles per locus when measurements of a single population are determined. However, when comparing across samples with differing numbers of individuals we support using a rarefaction method (Leberg 2002, Kalinowski 2004). When we rarified from 100 individuals to smaller sample sizes, we found that this method provided almost an exact replica of the simulated values (Fig. 1). Even when sample sizes from populations were as different as 5 and 100, they were readily comparable. Through rarefaction it is possible to determine whether there is a difference in diversity with very unequal sample sizes.

Comparisons between low- and high-diversity populations showed differing patterns. When unbiased average H_e and average H_o estimates were compared, the high-diversity population deviated less from the true value at small sample sizes than the low-diversity population. However, the opposite pattern was found for the allelic diversity estimates. Thus, some estimators perform better in high-diversity populations (unbiased H_e) and others in low-diversity populations (number of alleles per locus) at small sample sizes. A general pattern found in our simulations is that values based on high and low diversity populations begin to converge when they approach a sample size of 20 to 30. We recommend that when the nuclear genetic diversity of a population is unknown researchers should strive for a sample size of at least 20 and preferably 30 individuals. Estimates based on these sample sizes should provide a useful measure of genetic diversity in all populations whether they are genetically depauperate or diverse.

Acknowledgements – We thank D. Rocque, E. Saillant, E. Humphries, T. Glenn, T. Roberts, M. Patten, and two anonymous reviewers for comments on the manuscript.

References

Akst, E. P., Boersma, P. D. and Fleischer, R. C. 2002. A comparison of genetic diversity between the Galápagos Penguin and the Magellanic Penguin. – Conserv. Genetic. 3: 375–383.

Allendorf, F. W. 1986. Genetic drift and loss of alleles versus heterozygosity. – Zoo Biol. 5: 181–190.

Beaumont, M. A. and Bruford, M. W. 2000. Microsatellites in conservation genetics. – In: Goldstein, D. B. and Schlotterer, C. (eds). Microsatellites: evolution and applications. Oxford University Press, Oxford, pp. 165–182.

Baverstock, P. R. and Moritz, C. 1996. Project design. – In: Hillis, D. M., Moritz, C. and Mable, B. K. (eds). Molecular systematics. Sinauer, pp. 17–47.

Carvalho, G. R. and Hauser, L. 1994. Molecular genetics and the stock concept in fisheries. – Rev. Fish. Biol. Fish. 4: 326–350.

Chan, Y. and Arcese, P. 2002. Subspecific differentiation and conservation of song sparrows *Melospiza melodia* in the San Francisco bay region inferred by microsatellite loci analysis. – Auk 119: 641–657.

Ciofi, C., Beaumont, M. A., Swingland, I. R. and Bruford, M. W. 1999. Genetic divergence and units for conservation in the Komodo dragon *Varanus komodoensis*. – Proc. R. Soc. B 266: 2269–2274.

Edwards, T., Schwalbe, C. R., Swann, D. E. and Goldberg, C. S. 2004. Implications of anthropogenic landscape change on inter-population movements of the desert tortoise *Gopherus agassizii*. – Conserv. Genet. 5: 485–499.

Frankham, R., Ballou, J. D. and Briscoe, D. A. 2002. Introduction to conservation genetics. – Cambridge University Press, Cambridge, UK.
Glenn, T. 1997. Genetic bottlenecks in long-lived vertebrates: mitochondrial and microsatellite DNA variation in American alligators and whooping cranes. – PhD thesis, University of Maryland, College Park.
Gregorius, H. 1980. The probability of losing an allele when diploid genotypes are sampled. – Biometrics 36: 643–652.
Haavie, J., Saetre, G. P. and Moum, T. 2000. Discrepancies in population differentiation at microsatellites, mitochondrial DNA and plumage colour in the pied flycatcher – inferring evolutionary processes. – Mol. Ecol. 9: 1137–1148.
Hanotte, O., Zanon, C., Pugh, A., Grieg, C., Dixon, A. and Burke, T. 1994. Isolation and characterization of microsatellite loci in a passerine bird: the reed bunting *Emberiza schoeniclus*. – Mol. Ecol. 3: 529–530.
Jeffery, K. J., Keller, L. F., Arcese, P. and Bruford, M. W. 2001. The development of microsatellite loci in the song sparrow, *Melospiza melodia*, and genotyping errors associated with good quality DNA. – Mol. Ecol. Notes 1: 11–13.
Kalinowski, S. T. 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. – Conserv. Genet. 5: 539–543.
Kalinowski, S. T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? – Heredity 94: 33–36.
Keller, L. F., Jeffery, K. J., Arcese, P., Beaumont, M. A., Hochachka, W. M., Smith, J. N. and Bruford, M. W. 2001. Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. – Proc. R. Soc. B 268: 1387–1394.
Leberg, P. L. 2002. Estimating allelic richness: effects of sample size and bottlenecks. – Mol. Ecol. 11: 2445–2449.
Mock, K. E., Latch, E. K. and Rhodes, O. E. 2004. Assessing losses of genetic diversity due to translocation: long-term case histories in Merriam's turkey *Meleagris gallopavo merriami*. – Conserv. Genet. 5: 631–645.
Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. – Genetics 89: 583–590.
Nei, M. 1987. Molecular evolutionary genetics. – Columbia University Press, New York.
Nei, M. and Roychoudhury, A. K. 1974. Sampling variances of heterozygosity and genetic distance. – Genetics 76: 370–390.
Oyler-McCance, S. J., St. John, J., Taylor, S. E., Apa, A. D. and Quinn., T. W. 2005. Population genetics of Gunnison sage-grouse: implications for management. – J. Wildl. Manage. 69: 630–637.
Petit, R., El Mousadik, A. and Pons, O. 1998. Identifying populations for conservation on the basis of genetic markers. – Conserv. Biol. 12: 844–855.
Petren, K. 1998. Microsatellite primers from *Geospiza fortis* and cross species amplification in Darwin's finches. – Mol. Ecol. 7: 1782–1784.
Pruett, C. L. and Winker, K. 2005. Northwestern song sparrow populations show genetic effects of sequential colonization. – Mol. Ecol. 14: 1421–1434.
Ruzzante, D. E. 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. – Can. J. Fish. Aquat. Sci. 55: 1–14.
Schwartz, M. K., Ralls, K., Williams, D. F., Cypher, B. L., Pilgrim, K. L. and Fleischer, R. C. 2005. Gene flow among San Joaquin kit fox populations in a severely changed ecosystem. – Conserv. Genet. 6: 25–37.
Sjögren, P. and Wynoni, P. I. 1994. Conservation genetics and detection of rare alleles in finite populations. – Conserv. Biol. 8: 267–270.
Toro, M., Barragán, C., Óvilo, C., Rodrigañez, J., Rodriguez, C. and Silió, L. 2002. Estimation of coancestry in Iberian pigs using molecular markers. – Conserv. Genet. 3: 309–320.
Walther, B. A. and Moore, J. L. 2005. The concepts of bias, precision, and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. – Ecography 28: 815–829.