



## Performance of nonparametric species richness estimators in a high diversity plant community

A. CHIARUCCI<sup>1</sup>, N. J. ENRIGHT<sup>2\*</sup>, G. L. W. PERRY<sup>3</sup>, B. P. MILLER<sup>2</sup> and B. B. LAMONT<sup>4</sup> <sup>1</sup>Dipartimento di Scienze Ambientali 'G. Sarfatti', Università di Siena, Via P.A. Mattioli 4, 53100 Siena, Italy, <sup>2</sup>School of Anthropology, Geography & Environmental Studies, University of Melbourne, Australia 3010, <sup>3</sup>Department of Geography, King's College London, Strand, London WC2R 2LS U.K. and <sup>4</sup>School of Environmental Biology, Curtin University of Technology, GPO Box UL1987 Perth, Australia 6001

**Abstract.** The efficiency of four nonparametric species richness estimators — first-order Jackknife, second-order Jackknife, Chao2 and Bootstrap — was tested using simulated quadrat sampling of two field data sets (a sandy 'Dune' and adjacent 'Swale') in high diversity shrublands (kwongan) in south-western Australia. The data sets each comprised > 100 perennial plant species and > 10 000 individuals, and the explicit (*x-y* coordinate) location of every individual. We applied two simulated sampling strategies to these data sets based on sampling quadrats of unit sizes 1/400th and 1/100th of total plot area. For each site and sampling strategy we obtained 250 independent sample curves, of 250 quadrats each, and compared the estimators' performances by using three indices of bias and precision: MRE (mean relative error), MSRE (mean squared relative error) and OVER (percentage overestimation). The analysis presented here is unique in providing sample estimates derived from a complete, field-based population census for a high diversity plant community. In general the true reference value was approached faster for a comparable area sampled for the smaller quadrat size and for the swale field data set, which was characterized by smaller plant size and

higher plant density. Nevertheless, at least 15–30% of the total area needed to be sampled before reasonable estimates of  $S_i$  (total species richness) were obtained. In most field surveys, typically less than 1% of the total study domain is likely to be sampled, and at this sampling intensity underestimation is a problem. Results showed that the second-order Jackknife approached the actual value of  $S_i$  more quickly than the other estimators. All four estimators were better than  $S_{obs}$  (observed number of species). However, the behaviour of the tested estimators was not as good as expected, and even with large sample size (number of quadrats sampled) all of them failed to provide reliable estimates. First- and second-order Jackknives were positively biased whereas Chao2 and Bootstrap were negatively biased. The observed limitations in the estimators' performance suggests that there is still scope for new tools to be developed by statisticians to assist in the estimation of species richness from sample data, especially in communities with high species richness.

**Key words.** Bootstrap, Chao2, Jackknife, sample-based accumulation curve, species richness estimation.

### INTRODUCTION

Species richness is the most fundamental component of species diversity (Colwell & Coddington,

1994), and its estimation is one of the most common measures used in ecological research (Wilson, 1988; Rosenzweig, 1995; Purvis and Hector, 2000). It is often preferred to diversity due to the greater ease with which it can be determined for any sample, and given the lack of agreement as

\* Corresponding author.

to which of the myriad of diversity indices most adequately integrates both the numbers and abundances of species. However, sample species richness invariably underestimates the total richness of the plant communities or areas sampled. Complete enumeration of species richness within an extensive study area is generally not feasible, and can prove difficult for sampling units that exceed even a few hundreds of square meters in size (Palmer, 1995; Palmer *et al.*, 2002). Consequently, a number of different methods have been devised for estimating total species richness from samples. Among these, the nonparametric estimators are considered the most promising since they do not require any assumptions about community structure (Palmer, 1990, 1991; Colwell and Coddington, 1994; Chazdon *et al.*, 1998; Zelmer and Esch, 1999).

The most widely used nonparametric estimators are those based on the Bootstrap and Jackknife procedures developed by Efron and Tibshirani (1976), Heltshe and Forrester (1983) and Smith and van Belle (1984), and those developed by Chao (1984). These estimators were developed to estimate the number of classes in a statistical population from samples, such as species richness in communities (Heltshe and Forrester, 1983; Chao, 1984; Smith and van Belle, 1984), or the number of individuals in a population based on capture–recapture data (Burnham and Overton, 1978, 1979; Chao, 1987). One common assumption of the nonparametric estimators is that although the species have different probabilities of being collected, these probabilities are temporally and spatially constant throughout the study (Burnham and Overton, 1979). They require no assumptions about within-quadrat species interactions, whereas abundance models typically assume independence among species (Smith and van Belle, 1984). In addition, they have the important advantage that they consider quadrats as random samples of space and not as random samples of individuals (Heltshe and Forrester, 1983). Nonparametric estimators that use abundance data may be difficult to use in plant communities containing clonal plants since ramets and genets are not readily separable. Estimators based on incidence data (presence/absence) are thus particularly attractive for estimating the number of species in plant communities. According to Gotelli and Colwell (2001) an ideal estimator

should; (i) reach its own asymptote much sooner than the sample-based rarefaction curve levels off, and (ii) approximate the empirical asymptote in an unbiased way, when tested over many benchmark datasets.

The efficacy of the various richness estimators has been tested and compared mostly using simulated datasets (Burnham and Overton, 1979; Heltshe and Forrester, 1983; Pollock and Otto, 1983; Baltanás, 1992; Mingoti and Meeden, 1992; Norris and Pollock, 1996; Polulin, 1998; Walther and Morand, 1998; Zelmer and Esch, 1999). In plant communities, relatively few tests have been performed with real field data (Palmer, 1990, 1991; Chazdon *et al.*, 1998; Skov and Lawesson, 2000; Chiarucci *et al.*, 2001). All of these field-based studies analysed the performance of the species richness estimators by using data collected from only a limited fraction of the total area for which the estimation was run. This approach conforms to the reality of field sampling procedures of necessity used in community ecological studies, where conclusions about large areas must be based on estimates derived from samples.

A more complete test of species richness estimator performance might be afforded by applying these estimators to a data set in which the identity and location is known for every individual plant within the reference area and is not limited to a sample of locations (quadrats) only. A test along these lines, using field data, was attempted by Hellmann and Fowler (1999), who sampled five 0.4 ha forest plots with low to moderate species richness in Michigan (from 14 to 56 species), dividing the plots by a grid and enumerating all species within the grid (but without recording the explicit spatial position of each plant). Gimaret-Carpentier *et al.* (1998) performed a test on a complete census of trees = 1 cm d.b.h. identified and mapped in a 25-ha plot in the Pasoh Forest Reserve, Malaysia, and concluded that Chao1 and Chao2 estimators performed better than Jackknife estimators.

The availability of two data sets from high diversity Mediterranean-type shrublands in SW Australia in which the position (stem base) of each individual plant was recorded for more than 10 000 individuals per site offers the opportunity to test the current group of favoured nonparametric estimators of species richness by sampling

from a whole, known and species rich, population of plants within the plots and using sampling units (quadrats) of any chosen size. The aim of the present study is to test the performance of the most widely used nonparametric species richness estimators — First- and Second-order Jackknife, Bootstrap and Chao2 — using incidence data and simulating different sampling strategies and intensities for two species-rich sites ( $n > 100$  species in each case).

## MATERIALS AND METHODS

### Data sources and community structure

Field data were collected from two sites in the South Eneabba Flora Reserve, a part of the northern sandplain shrublands (kwongan), 270 km north of Perth, Western Australia. These shrublands are fire-prone with a mean fire return period of approximately 15 years (Enright *et al.*, 1996). Many species show adaptations to cope with fire, including ability to regrow vegetatively, and seed-based adaptations including fire stimulated germination of soil-stored seeds and release of canopy-stored (serotinous) seed crops (Enright and Lamont, 1989). Hnatiuk and Hopkins (1981) reported species richness of up to 130 species per 100 m<sup>2</sup> for sample quadrat data collected from the reserve for numerical classification of vegetation, and environmental correlation. They described clear differences in species composition for low sandy dunes and adjacent flatlands, based primarily on difference in depth of unconsolidated acid sands over less permeable lateritic substrate that affected water availability. We collected data from one site on a sandy dune crest and another in the adjacent flatlands (swale) to provide two high diversity data sets with contrasting community structure and composition. Both sites were last burned about 15 years ago and represented mature phase vegetation.

The sample areas were 40 × 40 m at the crest and 30 × 30 m at the swale, respectively. At each site the sample area was surveyed by dumpy level and permanent metal posts were installed on a 5 × 5 m grid. The identity of every plant was recorded within each 5 × 5 m subplot and the spatial position where plant stems emerged from the ground was recorded as the distance (in cm) from each of two known subplot corners. These values were later converted to *x-y* co-ordinates. For clonal plants, such as some rhizomatous sedges and rushes, spatial position was recorded as the centre of the patch of ramets. While this is a simplification, it is analogous to using stem position as the location for plants with canopies held above ground. In all cases, canopy width in two directions was also recorded so that alternative approaches to definition of species co-occurrence are possible (but are not explored here). A small number of plant species was excluded from the field survey; those smaller than 5 cm in both height and canopy diameter, and species that may be visible at some times of the year, but not others (examples included Droseraceae, Orchidaceae). The analysis therefore focuses on plant species with perennial above-ground parts.

A total of 10 550 and 12 844 individuals were enumerated at the crest and the swale sites, respectively, and more than 100 species were recorded in each site (Table 1). The sites were both dominated numerically by individuals from the three predominantly southern hemisphere families Myrtaceae, Proteaceae and Restionaceae, with the Cyperaceae, Dilleniaceae and Epacridaceae also prominent. Taking into account density data, no species occurred with relative abundance higher than 10% in either site, while 30 and 22 species were represented by fewer than 10 individuals (i.e. a relative abundance of < 0.01%) in the crest and swale sites, respectively. Species-abundance distribution from the swale site closely fitted the log-normal distribution,

**Table 1** General properties of the crest and swale vegetation sample sites, South Eneabba Flora Reserve, Western Australia. Area is in m<sup>2</sup>, *N* is total number of individuals sampled, Inds/m<sup>2</sup> is mean number of individuals per m<sup>2</sup>, *H'* is sample Shannon-Weiner Diversity based on density values, *J'* is sample Evenness

Site	Area	<i>N</i>	Inds/m <sup>2</sup>	Species	Genera	Families	<i>H'</i>	<i>J'</i>
Crest	1600	10 550	6.59	107	62	21	3.76	0.79
Swale	900	12 844	14.27	104	58	22	3.94	0.85

while that from the crest approximated the geometric series. In both cases, genus- and family-abundance relationships were consistent with the species-abundance distributions. Shannon-Weiner diversity was high and evenness was moderate for both sites, and the main difference between sites was the higher density of individuals per unit area in the swale (Table 1).

### Simulated sampling design

The data were tabulated as a list of individuals by species and *x-y* co-ordinate location (stem base). A computer program (Excel Add-in routine) was written by GLWP to allow either quadrat-based or individual-based sampling of these plants with and without replacement for any sample size (quadrat size or number of individuals). In the present paper we explore the use of quadrat-based sampling, which is the most common approach to field sampling, and to the comparison of nonparametric estimators reported by other workers.

Instead of choosing a fixed quadrat size for both sites, which differed in total area, we selected quadrat sizes that represented a constant proportion of the whole plot. For each site we applied two sampling strategies that used different quadrat sizes, representing, respectively, 1/400th (small quadrats) and 1/100th (large quadrats) of the plot area. In the first case, the quadrats were 2 × 2 m and 1.5 × 1.5 m for the crest and swale plot, respectively, and in the second, quadrats were 4 × 4 m and 3 × 3 m, respectively. For each plot we sampled 250 accumulation curves, using 250 quadrats for each curve. With the small quadrat size, the maximum potential area sampled per curve was 62.5% of the total plot area, while for the large quadrat size it was 250% of the total plot area. Our sampling strategy was based on sampling with replacement, as formally required for unbiased estimation using these nonparametric methods (Heltshe and Forrester, 1983; Smith and van Belle, 1984), so that the actual area explored by the sample quadrats may be smaller than the corresponding maximum. The use of a very high proportion of area sampled, by the large quadrat size, allows a complete statistical evaluation of the performance of the estimators in order to find the point where the estimates stabilize.

### Species richness estimators

The species richness estimators tested in the present survey were the mean species accumulation curve, which reports the mean number of species found with a given number of quadrats, and corresponds to the sample-based rarefaction curve (Gotelli and Colwell, 2001), and the most widely used nonparametric estimators using incidence data: First-order Jackknife, Second-order Jackknife, Bootstrap and Chao2. All these estimators use presence/absence data in sampling units, and are thus based on the spatial dispersion of species more than on rigorous measures of abundance.

The First-order Jackknife is based on the presence of 'unique species' in sampling units, defined as the species recorded in only one sampling unit (Heltshe and Forrester, 1983; Colwell and Coddington, 1994) and is expressed by the formula:

$$S_{Jack1} = S_{obs} + \left( \frac{n-1}{n} \right) \cdot k$$

where  $S_{obs}$  is the number of species observed in the sample,  $n$  is the sample size (number of quadrats) and  $k$  is the number of unique species.

The Second-order Jackknife (Burnham and Overton, 1978, 1979; Colwell and Coddington, 1994) takes into account the number of 'unique species',  $k$ , as well as the number of species that occur in exactly two samples  $m$ , named 'duplicates' (Colwell and Coddington, 1994) such that:

$$S_{Jack2} = S_{obs} + \left[ \frac{k \cdot (2n-3)}{n} - \frac{m(m-2)^2}{n \cdot (n-1)} \right]$$

The bootstrap estimator (Efron, 1979; Smith and van Belle, 1984; Colwell and Coddington, 1994) is based on the frequency distribution of the species found in the sample, i.e. the proportion of quadrats containing each species, according to the following formula:

$$S_{Boot} = S_{obs} + \sum_{j=1}^{S_{obs}} (1 - p_j)^n$$

where  $S_{obs}$  is the number of species observed in the pooled sample, and  $p_j$  is the proportion of plots containing the species  $j$ .

Chao (1984) developed another group of nonparametric estimators. These estimators were originally designed for an individual-based sampling

approach, taking into account the species recorded just once or twice ('singletons' and 'doubletons', respectively, Colwell and Coddington, 1994). A later modification (Chao, 1987) allowed estimation for quadrat-based data using species found in just one or two samples, without any reference to the true species abundance ('uniques' and 'duplicates'). The formula for this estimator was, however, undefined for zero duplicates. In his new version of the program EstimateS, Robert Colwell (<http://viceroi.eeb.uconn.edu/EstimateS>) provides a bias-corrected version supplied by Anne Chao. The bias-corrected formula, which was used in this study, is:

$$S_{Chao2} = S_{obs} + \frac{k^2}{2 \cdot (m + 1)} - \frac{k \cdot m}{2 \cdot (m + 1)^2}$$

### Comparison of estimators' performance

For each site and sampling strategy we obtained 250 independent sample curves, of 250 quadrats each, and then used these curves to obtain mean curves and empirical measures of variability. Relative Error (RE) and Square Relative Error (SRE) were used to evaluate the performance of each estimator. RE is a relative measure of bias between the estimated value and the true value of species richness and was calculated by the formula:

$$RE = \frac{S_{est} - S_i}{S_i}$$

where  $S_{est}$  is the estimated value and  $S_i$  is the true value of species richness observed in the plot ( $S_i = 107$  and  $104$  for the crest and swale plots, respectively). We also calculate the square of RE (SRE) which expresses the closeness of the estimated value to the true value, without considering the sign of the deviation; in addition, SRE weights estimates further from  $S_i$  more heavily.

The bias of the different estimators over the 250 independent sample curves was assessed by the mean value of RE (MRE). The precision of the estimators was assessed by the mean value of SRE (MSRE) and by a measure of overestimation (OVER), an additional indicator expressing the percentage of estimates higher than the true reference value  $S_i$ .

MRE is equivalent to the 'mean deviation' of Palmer (1990) and to the 'bias' of Hellmann and Fowler (1999), except that it is scaled by the true

value, being equivalent to the 'bias' used by Walther and Morand (1998). MSRE is equivalent to the 'mean square proportional deviation' of Palmer (1990) and the 'deviation' of Walther and Morand (1998). If all RE values have the same sign, MRE and MSRD are related. However, if an estimator has a range of estimations above and below the true value, it may have a low MRE but a high MSRE. A perfectly unbiased estimator should have MRE and MSRD values of zero and a value of OVER of 50% (Palmer, 1990; Walther and Morand, 1998).

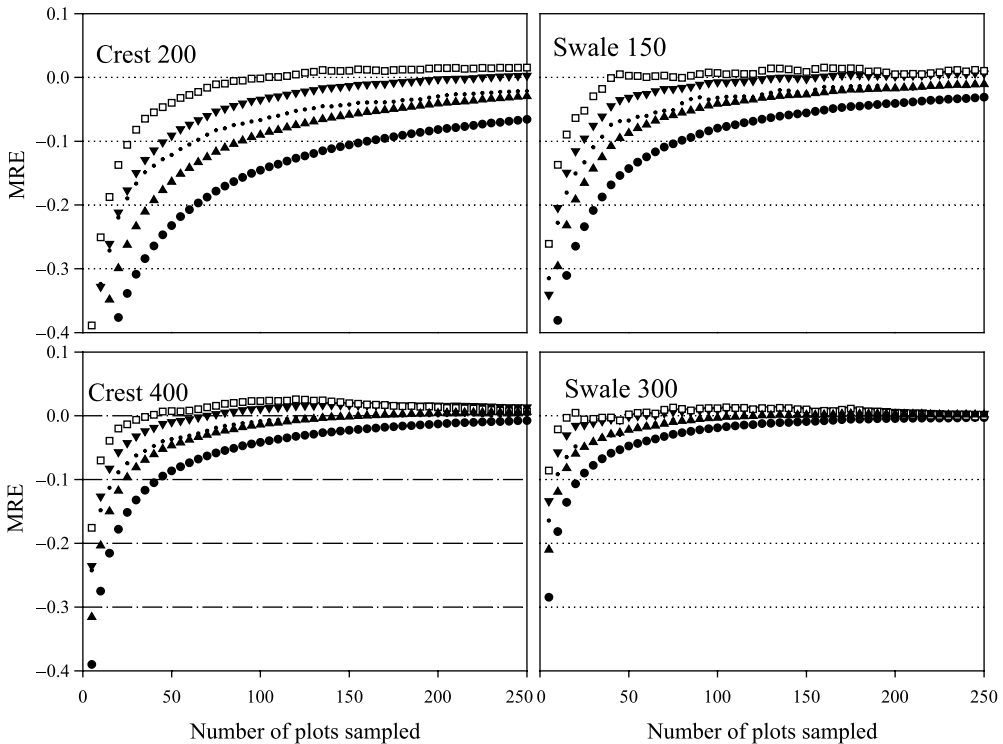
### RESULTS

The patterns of MRE values show that the performance of the nonparametric estimators progressively improves as the number of quadrats sampled increases (Table 2, Fig. 1). Estimates from the swale were consistently about 50% less biased than those for the crest, apart from some estimates for the Second-order Jackknife using the small sample quadrats. Estimates within sites were also consistently better for the small quadrat size for a given total area sampled, in terms of reducing the MRE. To sample the same amount of area with the small quadrats requires four times the number of quadrats with respect to the large quadrats, and so would likely provide a better spatial coverage of the total sample space relative to the distribution of large sample quadrats. We did not consider the possibility of compensational effects of reduced amount of sampling effort needed in the field where large quadrat size is used.

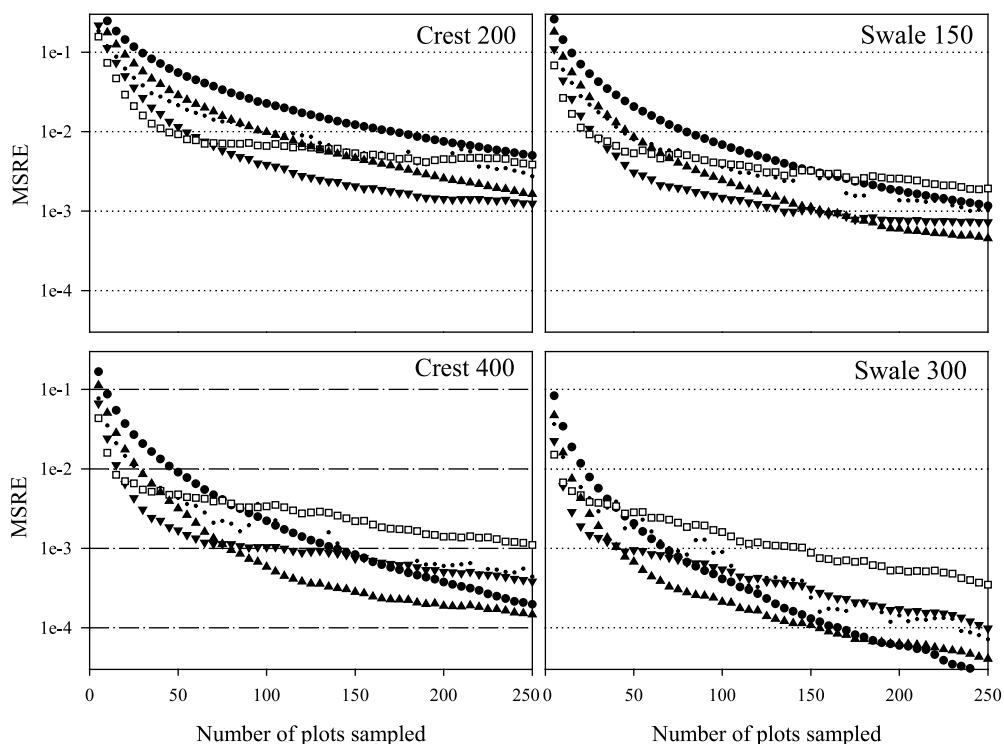
Among the different estimators,  $S_{obs}$  (observed species richness) was consistently the most biased, and even with a sample size corresponding to 50% of the area of the whole plot its estimates were biased by 4.1%–8.7% (Table 2). The estimator with the best MRE was Second-order Jackknife, followed by First-order Jackknife, Chao2 and Bootstrap, respectively. This pattern was almost the same for every sample size (number of sample quadrats), site and quadrat size (Fig. 1). However, even the estimates provided by the best performing estimator, Second-order Jackknife, remained negatively biased until the sample size reached a number of quadrats corresponding to 15–30% of the whole plot area. The number of quadrats required to obtain this proportion of area

**Table 2** Results of the MRE (mean relative error) analysis for a given percentage of the total area of plot sampled by the two different sizes of sample quadrats in the crest and swale sites

Quadrat size	Percentage of plot sampled							
	5%		10%		20%		50%	
	small	large	small	large	small	large	small	large
<b>Crest</b>								
$S_{obs}$	-0.377	-0.390	-0.264	-0.275	-0.171	-0.178	-0.082	-0.087
Chao2	-0.220	-0.243	-0.138	-0.148	-0.08	-0.089	-0.030	-0.037
Jack1	-0.212	-0.235	-0.114	-0.127	-0.049	-0.057	-0.003	-0.010
Jack2	-0.137	-0.176	-0.055	-0.07	-0.01	-0.02	0.015	0.007
Boot	-0.299	-0.316	-0.193	-0.203	-0.11	-0.118	-0.041	-0.047
<b>Swale</b>								
$S_{obs}$	-0.265	-0.285	-0.169	-0.182	-0.099	-0.107	-0.041	-0.048
Chao2	-0.151	-0.164	-0.074	-0.092	-0.040	-0.048	-0.016	-0.022
Jack1	-0.115	-0.134	-0.045	-0.057	-0.016	-0.016	0.002	-0.002
Jack2	-0.064	-0.086	-0.001	-0.021	0.000	0.005	0.005	0.002
Boot	-0.192	-0.210	-0.108	-0.119	-0.055	-0.060	-0.017	-0.022



**Fig. 1** MRE (mean relative error) in relation to number of quadrats sampled, by site (Crest, Swale) and quadrat size (150 × 150, 200 × 200, 300 × 300, 400 × 400 cm), for the different richness estimators tested: observed number of species (black dots), Chao2 (small dots), Jackknife 1 (down triangles), Jackknife 2 (open quadrats) and Bootstrap (up triangles).



**Fig. 2** MSRE (mean square relative error) in relation to number of quadrats sampled, by site (Crest, Swale) and quadrat size ( $150 \times 150$ ,  $200 \times 200$ ,  $300 \times 300$ ,  $400 \times 400$  cm), for the different richness estimators tested: observed number of species (black dots), Chao2 (small dots), Jackknife 1 (down triangles), Jackknife 2 (open quadrats) and Bootstrap (up triangles).

sampled was 60–120 for the small quadrat size and 15–30 for the large quadrat size. The other estimators needed a much higher sampling intensity to obtain comparable results. After reaching the true reference value ( $S_i$ ), the First and Second order Jackknives showed a positive bias that continued even when sample size exceeded 100% of the whole plot area (Fig. 1). Chao2 and Bootstrap estimators had a lower bias after this point, showing a more stable behaviour. None of the estimators provided reliable richness estimates for proportion of total area sampled less than 15%.

The MSRE (Fig. 2, Table 3) values showed a similar trend to the MRE values described above; in particular, with small sample size Second-order Jackknife showed the lowest MSRE, followed by First-order Jackknife, Chao2, Bootstrap and  $S_{obs}$ . However, Second-order Jackknife did not maintain the best MSRE performance across the full range of sample sizes, with the First-

order Jackknife and Bootstrap becoming the most precise (lower MSRE) at large sample sizes (Fig. 2). Again, the small quadrats provided consistently better MSRE for the same proportion of total area sampled (Table 3). The OVER analyses indicated that only the Second-order Jackknife, and to a lesser extent the First-order Jackknife, stabilized around the optimum percentage overestimation of 50% (Table 4, Fig. 3). However, such values were reached only when at least 20%–30% of the whole plot area had been sampled, representing a large number of quadrats. When a very large number of quadrats was added to the curve, representing a sample area of about 200% (for the large quadrat size) of the whole plot area, Second-order Jackknife, First-order Jackknife and Chao2 started to show a decrease of OVER, while Bootstrap continued to increase (Fig. 3). By definition, the overestimation by  $S_{obs}$  was constantly 0.

**Table 3** Results of the MSRE (mean squared relative error) analysis for a given percentage of the total area of plot sampled by the two different sizes of sample quadrats in the crest and swale sites

Quadrat size	Percentage of plot sampled							
	5%		10%		20%		50%	
	<i>small</i>	<i>large</i>	<i>small</i>	<i>large</i>	<i>small</i>	<i>large</i>	<i>small</i>	<i>large</i>
<b>Crest</b>								
<i>S<sub>obs</sub></i>	0.144	0.167	0.072	0.087	0.031	0.037	0.007	0.009
<i>Chao2</i>	0.062	0.077	0.028	0.035	0.013	0.015	0.005	0.004
<i>Jack1</i>	0.050	0.067	0.017	0.024	0.005	0.006	0.001	0.002
<i>Jack2</i>	0.029	0.044	0.011	0.016	0.007	0.007	0.005	0.005
<i>Boot</i>	0.093	0.112	0.040	0.051	0.014	0.017	0.003	0.003
<b>Swale</b>								
<i>S<sub>obs</sub></i>	0.070	0.083	0.029	0.034	0.010	0.012	0.002	0.002
<i>Chao2</i>	0.028	0.037	0.012	0.014	0.005	0.006	0.001	0.002
<i>Jack1</i>	0.016	0.022	0.005	0.006	0.002	0.002	0.001	0.001
<i>Jack2</i>	0.011	0.015	0.007	0.007	0.005	0.005	0.003	0.003
<i>Boot</i>	0.038	0.047	0.013	0.016	0.004	0.004	0.001	0.001

**Table 4** Results of OVER (percentage overestimation) analysis for a given percentage of the total area of plot sampled by the two different sizes of sample quadrats in the crest and swale sites

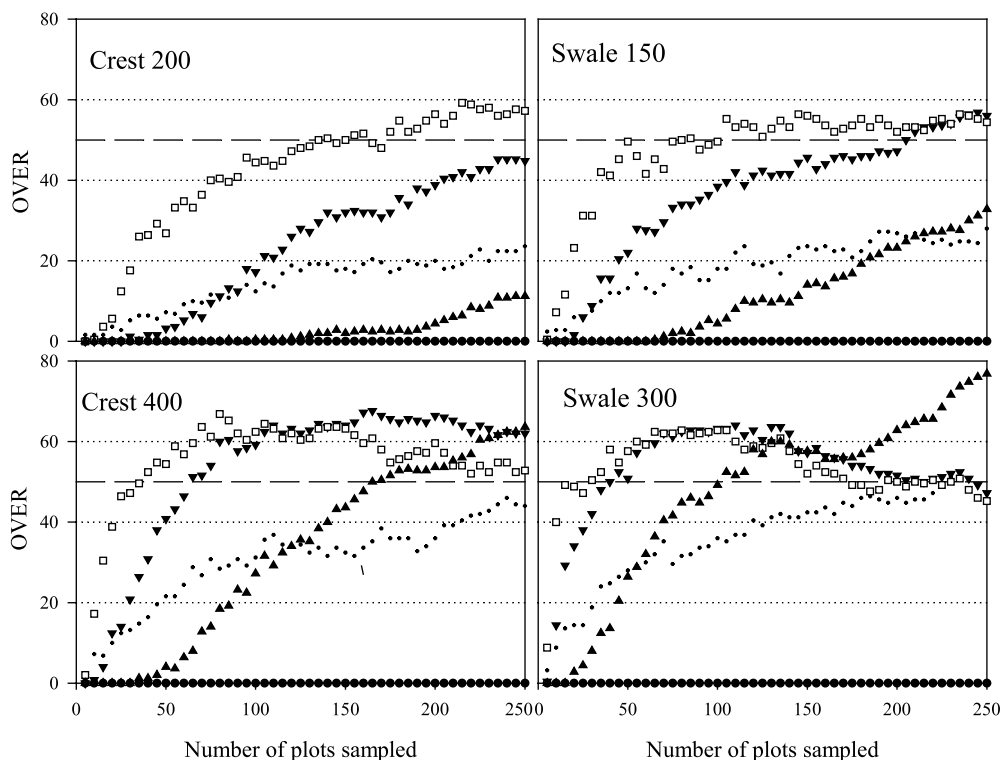
Quadrat size	Percentage of plot sampled							
	5%		10%		20%		50%	
	<i>small</i>	<i>large</i>	<i>small</i>	<i>large</i>	<i>small</i>	<i>large</i>	<i>small</i>	<i>large</i>
<b>Crest</b>								
<i>S<sub>obs</sub></i>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Chao2</i>	3.6	2.4	6.4	7.2	11.2	10.0	20.0	21.6
<i>Jack1</i>	0.0	0.0	1.6	0.8	11.2	12.4	38.8	40.8
<i>Jack2</i>	5.6	2.0	26.4	17.2	40.4	38.8	56.4	54.4
<i>Boot</i>	0.0	0.0	0.0	0.0	0.0	0.0	4.4	4.0
<b>Swale</b>								
<i>S<sub>obs</sub></i>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<i>Chao2</i>	6.0	3.2	12.0	8.8	16.8	14.4	26.8	28.0
<i>Jack1</i>	1.6	0.4	15.6	14.4	34.0	34.0	47.2	50.8
<i>Jack2</i>	23.2	8.8	41.2	40.0	50.0	48.8	52.0	57.6
<i>Boot</i>	0.0	0.0	0.0	0.0	2.4	2.8	23.2	26.4

## DISCUSSION

The present study allowed a unique test of the efficiency of several well-known nonparametric estimators of species richness for two field data sets where all the individual plants were enumerated and located. To our knowledge, this is one of the first studies where simulated sampling has

been applied to such a complete field data set for individual plants to estimate the performance of the most commonly used nonparametric estimators. Gimaret-Carpentier *et al.* (1998) tested the nonparametric estimators of species richness on a complete census of trees exceeding a chosen threshold size within their study area, in the Pasoh Forest Reserve, Malaysia. However they





**Fig. 3** OVER (percentage overestimation) in relation to number of quadrats sampled, by site (Crest, Swale) and quadrat size ( $150 \times 150$ ,  $200 \times 200$ ,  $300 \times 300$ ,  $400 \times 400$  cm), for the different richness estimators tested: observed number of species (black dots), Chao2 (small dots), Jackknife 1 (down triangles), Jackknife 2 (open quadrats) and Bootstrap (up triangles).

used clusters of a fixed number of individuals as sampling units, rather than quadrats, and limited their analysis to tree species only. Their data set comprised 13 678 trees  $> 1$  cm d.b.h. within a study area of 25 ha, while we used data on 10 550 and 12 844 individual plants, identified and mapped within two much smaller study plots. These authors found that First- and Second-order Jackknives overestimated the true number of species for samples sizes up to 9000 individuals. While Chao1 and Chao2 estimators also overestimated the number of species, they were closer to the true species richness.

Many authors have tested the performance of the same estimators by using simulated sampling on contrived data sets (Burnham and Overton, 1979; Heltshe and Forrester, 1983; Pollock and Otto, 1983; Baltanás, 1992; Mingoti and Meeden, 1992; Norris and Pollock, 1996; Polulin, 1998; Walther and Morand, 1998; Zelmer and Esch,

1999). However, while contrived data may reflect theoretical species-abundance distributions for a hypothetical plant community, it is unlikely that these data will necessarily reflect the spatial distribution and aggregation properties of real communities. Our results clearly show that the swale site, where abundances fitted a log-normal distribution and plant density was much higher, yielded more accurate estimates, and at lower sample intensity, than did the crest site, where abundances followed a geometric distribution and plant densities were lower.

On the other hand, authors working with real data sets (Palmer, 1990, 1991; Chazdon *et al.*, 1998; Keating and Quinn, 1998; Skov and Lawesson, 2000; Chiarucci *et al.*, 2001) have usually lacked data on the whole community and so could not simulate different sampling designs. Chazdon *et al.* (1998) found that the Incidence-based

Coverage Estimator (ICE), an index based on the concept of sample coverage (Chao and Lee, 1992), was the best performing estimator. However, this estimator requires an arbitrary threshold to define 'infrequent species' and can produce therefore different results from the same data set. In addition, Longino *et al.* (2002) found that the behaviour of ICE was not stable with different data sets of ant communities at La Selva. Chazdon *et al.* (1998) found Chao2 to be the second best estimator, quickly producing stable estimates as number of sampled plots increases, with First-order Jackknife also considered to be relatively stable at small sample size, whereas, Second-order Jackknife was found to exceed the true value, and then fall to a value closer to the true one. They observed that the Chao2 estimator provided good approximations, being relatively insensitive to number of quadrats, and approaching the true value quickly despite the patchiness of species distribution. In our test Chao2 was outperformed by the First-order Jackknife and never reached the performance level of Second-order Jackknife at small sample sizes.

Palmer (1990, 1991) tested the Bootstrap and First- and Second-order Jackknife estimators using  $40 \times 2 \text{ m}^2$  quadrats randomly located in each of  $30 \times 0.1 \text{ ha}$  field plots. All of these were found useful relative to the number of observed species in reducing the bias, but overall, the Second-order Jackknife gave the best performance. The area covered by the sample quadrats in the study by Palmer (1990) was 8%, which is a relatively high proportion compared to that which might be feasible in a field study of a large area. In our test this proportion was reached by sampling, respectively, 32 small or 8 large sample quadrats. At this sample size the Second-order Jackknife outperformed the other estimators. In the swale site, overestimation for this index had already reached approximately 40% (theoretical optimum = 50%) at this sample size; a value otherwise exceeded only by First-order Jackknife among the other estimators even when up to 60% of the total plot area was sampled.

Skov and Lawesson (2000) used 325 nested plots, from  $2 \text{ m}^2$  to  $200 \text{ m}^2$  each, systematically placed to survey a managed forest of 325 ha in Denmark, and tested Second-order Jackknife and Chao2 estimators. They found that both the estimators were useful in reducing bias relative

to the observed number of species, and that the Second-order Jackknife constantly outperformed Chao2. Chiarucci *et al.* (2001) tested all the non-parametric estimators we tested in the present survey on data collected in 50 plots of  $50 \text{ m}^2$  randomly located in a nature reserve of 431 ha. First- and Second-order Jackknives were found to give the best overall performance: Second-order Jackknife gave the best estimate of total species richness, but was slightly less precise than the First-order Jackknife when applied to life-form data sets which showed different abundance distributions. First-order Jackknife had a higher, and consistently negative, bias, but slightly better precision. The sample used by Chiarucci *et al.* (2001) represented less than 0.1% of the total area for which the estimates were performed, while in the test done by Skov and Lawesson (2000) the area sampled represented 0.02%–2%. According to our tests, the Second-order Jackknife provides the best estimates within this range of area sampled. It was the most precise estimator (lower MSRE, and OVER closer to 50%) for small sample size (less than 10–15% of the total area sampled), but with larger sample sizes (~30–70% of the total area sampled) the First-order Jackknife was slightly better. Bootstrap was the best estimator once the sample approximated the total plot area.

An approach more comparable to ours was used by Hellmann and Fowler (1999), who recorded the presence of all species in five 0.4 ha plots divided into a series of smaller ( $5 \times 5 \text{ m}$ ) quadrats within which species presence/absence was recorded. These subplots were then sampled. They found that for sample size lower than ~25% of the whole plot the least biased estimator was the Second-order Jackknife, followed by the First-order Jackknife and Bootstrap. With increases in sample size, these estimators became, in the same order, positively biased. Hellmann and Fowler (1999) also observed how the Second-order Jackknife provided the most variable estimates, followed by the First-order Jackknife and the Bootstrap. However, their results were biased because of the sampling strategy adopted, which was based on sampling without replacement, rather than with replacement as required for unbiased estimation (Heltshe and Forrester, 1983; Smith and van Belle, 1984). We also observed an overestimation by the same estimators, and in the

same order, as reported by Hellmann and Fowler (1999) but the magnitude was much smaller.

Although nonparametric estimators have been considered promising, the results obtained up to now using field data sets have been far from conclusive (Palmer, 1990, 1991; Gimaret-Carpentier *et al.*, 1998; Hellmann and Fowler, 1999; Skov and Lawesson, 2000; Chiarucci *et al.*, 2001). This is in part due to the fact that almost all previous analyses of nonparametric estimators using field data have been based on samples — so that  $S_i$  was not really known, and have used a variety of different sampling strategies and scales of resolution — making comparison between studies difficult. Because  $S_i$  is known for the two field data sets investigated here, and because they incorporate the real complexities of spatial aggregation and species-abundance patterns within high diversity plant communities, the findings of the present work allow a critical evaluation of these estimators.

Overall, the results obtained here are not encouraging. With a low number of quadrats sampled, the Jackknife estimators provided the most accurate estimates of  $S_i$ , as also reported by Smith and van Belle (1984). However, even the best estimator, Second-order Jackknife, approached  $S_i$  too slowly as area sampled increased, requiring a larger sample (at least 15–30% coverage of the survey domain) than could be expected in most field surveys. Estimated  $S_i$  continued to increase until the area sampled was close to 100% of the total area, then exceeded the actual  $S_i$  before declining again for sample areas > 150% of the total plot area, i.e. it failed to reach an asymptote. This was also true of the First-order Jackknife. The Chao2 and Bootstrap estimators were consistently more negatively biased than the Jackknife estimators at all sample sizes up to 100% of the total area and had lower percentage over-estimation with respect to the optimal value of 50%. D'Alessandro and Fattorini (2002) have recently demonstrated that when many species with a very low probability of detectability are present in a community, the Bootstrap and Jackknife estimators are inadequate to estimate its total species richness.

In many cases, plot-based sampling is essential to the objectives of field surveys and estimation of species richness is only one of a number of desired outputs. Here, it is still reasonable to seek to identify and use the best available estimator,

albeit understanding its limitations. However the best performing estimators can vary from data set to data set (Colwell and Coddington, 1994; D'Alessandro and Fattorini, 2002; Longino *et al.*, 2002). When interest is in estimating the number of species in a very large area it is clear that the whole area cannot be completely surveyed. For example, in the studies by Skov and Lawesson (2000) and Chiarucci *et al.* (2001) the total area covered by sample quadrats was as little as 0.1–2%. In such large areas many species are localized and their detection probability is low, arguably rendering the nonparametric estimators ineffective (D'Alessandro and Fattorini, 2002).

While nonparametric estimators certainly provide a reduction in bias relative to observed species number, and our results suggest that the Second-order Jackknife is best, the estimates obtained can hardly be expected to be accurate and are not likely to be easy to interpret. Gotelli and Colwell (2001; p. 389) note that a common weakness among existing estimators when applied to hyperdiverse taxa is that '[ ... ] they often fail to reach any asymptote at all, rising more or less in parallel with the still-steep sample-based rarefaction curve'. Our results suggest that this assertion will prove true for any real data set comprising a small sampled area from a high diversity community. Estimates for communities with lower richness will almost always show better accuracy at lower sample size due to, on average, more equitable spatial distribution of species among samples. This can be at least partly illustrated by comparing observed species richness using random, individual-based vs. quadrat-based samples. Random, individual-based sampling removes any effects of aggregation, but not species-abundance patterns, on the probability that a species will be recorded. Thus, the extent to which quadrat-based richness falls below individual-based richness, for samples with equal number of individuals, represents a measure of the overall impact of aggregation on the likely efficiency of richness estimators (Gotelli and Colwell, 2001). Table 5 summarizes results of such a comparison for the crest and swale plots across samples ranging from the equivalent of 1% to 50% of the whole survey area, and it is clear that quadrat-based species richness lies consistently below individual-based estimates. The underestimation impacts of aggregation are

**Table 5** A comparison of mean species richness ( $S_{obs}$ ) in the crest and swale plots calculated using a random individual-based sampling procedure vs. quadrat (small and large quadrat sizes) sampling, with equal  $n$ . Values in parentheses are the percentages by which the quadrat-based estimates fall below the individual-based estimates of  $S_{obs}$

$n$	% sample	Crest			Swale		
		$S_{obs}$ Individual	Small Q	Large Q	$S_{obs}$ Individual	Small Q	Large Q
100	1	40	37 (7.5)	33 (17.5)	44	41 (6.8)	42 (4.5)
200	2	54	49 (9.3)	45 (16.7)	59	55 (6.8)	54 (8.5)
500	5	69	66 (4.3)	64 (7.2)	75	72 (4.0)	71 (5.3)
1000	10	79	78 (1.3)	76 (3.8)	85	83 (2.4)	82 (3.5)
2000	20	88	88 (0.0)	86 (2.3)	93	92 (1.1)	92 (1.1)
5000	50	98	97 (1.0)	97 (1.0)	99	99 (0.0)	98 (1.0)

most apparent for small sample sizes, for large sampling units (quadrat size), and for the crest (where plants are larger and densities are lower) relative to the swale.

While there is a continuing need for further comparative studies of the performance of these richness estimators on different empirical and theoretically derived data sets (Colwell and Coddington, 1994; Gotelli and Colwell, 2001), our results are likely to diminish the expectations from such work. Clearly, there is scope for new or modified estimators to be developed by statisticians to assist field ecologists in the estimation of species richness from sample data, especially for communities with high species richness and differing species-abundance and spatial aggregation patterns, and where sampling can represent only a very small fraction of the total study area. There is also a need for consideration to be given to the processes or mechanisms that may underlie observed diversity patterns and thus influence the performance of the various indicators (M. Huston, pers. comm.). In the plots described here, future studies will investigate aspects of biotic (spatial patterns of extant species, soil seedbanks, recruits after fire) and abiotic (soil nutrients, soil moisture) spatial heterogeneity to address these issues.

## ACKNOWLEDGMENTS

The authors acknowledge the assistance of a number of colleagues and hired helpers who assisted in the field data collection exercise. The University of Melbourne facilitated field data

collection and collaboration with coauthors through financial assistance to NJE from its Study Leave and research grants (MRGS) programs. Curtin University assisted with field costs through an Infrastructure grant to BBL.

## REFERENCES

- Baltanás, A. (1992) On the use of some methods for the estimation of species richness. *Oikos* **65**, 484–492.
- Burnham, K.P. & Overton, W.S. (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 623–633.
- Burnham, K.P. & Overton, W.S. (1979) Robust estimation of the size of population size when capture probabilities vary among animals. *Ecology* **60**, 927–936.
- Chao, A. (1984) Non-parametric estimation of the classes in a population. *Scandinavian Journal of Statistics* **11**, 265–270.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791.
- Chao, A. & Lee, S.M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* **87**, 210–217.
- Chazdon, R.L., Colwell, R.K., Denslow, J.S. & Guariguata, M.R. (1998) Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. *Forest Biodiversity in North, Central and South America and the Caribbean: Research and Monitoring* (eds F. Dallmeier and J. Corniskey), pp. 285–309. Parthenon Publishing, Paris.
- Chiarucci, A., Maccherini, S. & De Dominicis, V. (2001) Evaluation and monitoring of the flora in

- a nature reserve by estimation methods. *Biological Conservation* **101**, 305–314.
- Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London, Series B* **345**, 101–118.
- D'Alessandro, L. & Fattorini, L. (2002) Resampling estimators of species richness from presence-absence data: why they don't work. *Metron* **61**, in press.
- Efron, B. & Thisted, R. (1976) Estimating the number of unseen species: how many words did Shakespeare know? *Biometrika* **63**, 35–41.
- Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.
- Enright, N.J. & Lamont, B.B. (1989) Seed banks, fire season, safe sites and seedling recruitment in five co-occurring *Banksia* species. *Journal of Ecology* **77**, 1111–1122.
- Enright, N.J., Lamont, B.B. & Marsula, R. (1996) Canopy seed bank dynamics and optimum fire regime for the highly serotinous shrub, *Banksia hookeriana*. *Journal of Ecology* **84**, 9–17.
- Gimaret-Carpentier, C., Pélissier, R., Pascal, J.P. & Houllier, F. (1998) Sampling strategies for the assessment of tree species diversity. *Journal of Vegetation Science* **9**, 161–172.
- Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**, 379–391.
- Hellmann, J.J. & Fowler, G.W. (1999) Bias, precision, and accuracy of four measures of species richness. *Ecological Applications* **9**, 824–834.
- Heltshe, J.F. & Forrester, N.E. (1983) Estimating species richness using the Jackknife procedure. *Biometrics* **39**, 1–11.
- Hnatiuk, R. & Hopkins, A. (1981) An ecological analysis of kwongan vegetation south of Eneabba, WA. *Australian Journal of Ecology* **6**, 423–438.
- Keating, K.A. & Quinn, J.F. (1998) Estimating species richness: the Michaelis-Menten model revisited. *Oikos* **81**, 411–416.
- Longino, J.T., Coddington, J. & Colwell, R.K. (2002) The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology* **83**, 689–702.
- Mingoti, S.A. & Meeden, G. (1992) Estimating the total number of distinct species using presence and absence data. *Biometrics* **48**, 863–875.
- Norris, J.L. III & Pollock, K.H. (1996) Non-parametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* **52**, 639–649.
- Palmer, M.W. (1990) The estimation of species richness by extrapolation. *Ecology* **71**, 1195–1198.
- Palmer, M.W. (1991) Estimating species richness: the second order jackknife reconsidered. *Ecology* **72**, 1512–1513.
- Palmer, M.W. (1995) How should one count species? *Natural Areas Journal* **15**, 124–135.
- Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S. & Wohlgemuth, T. (2002) Quantitative tools for perfecting species lists. *Environmetrics* **13**, 121–137.
- Pollock, K.H. & Otto, M.C. (1983) Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* **39**, 1035–1049.
- Polulin, R. (1998) Comparison of three estimators of species richness in parasite component communities. *Journal of Parasitology* **8**, 485–490.
- Purvis, A. & Hector, A. (2000) Getting the measure of biodiversity. *Nature* **405**, 212–219.
- Rosenzweig, M.L. (1995) *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.
- Skov, F. & Lawesson, J.E. (2000) Estimation of plant species richness from systematically placed plots in a managed forest ecosystem. *Nordic Journal of Botany* **20**, 477–483.
- Smith, E.P. & van Belle, G. (1984) Non-parametric estimation of species richness. *Biometrics* **40**, 119–129.
- Walther, B.A. & Morand, S. (1998) Comparative performance of species estimation methods. *Parasitology* **116**, 395–405.
- Wilson, E.O., ed. (1988) *Biodiversity*. National Academy Press, Washington, DC.
- Zelmer, D.A. & Esch, G.W. (1999) Robust estimation of parasite component community richness. *Journal of Parasitology* **85**, 592–594.