Syst. Biol. 51(2):303–316, 2002

Systematic Data in Biodiversity Studies: Use It or Lose It

V. A. FUNK[1] AND K. S. RICHARDSON[2,3]

[1] *U.S. National Herbarium, Smithsonian Institution, Washington, D.C. 20560-0166, USA; E-mail: funkv@nmnh.si.edu*
[2] *Department of Geography, McGill University, Montreal, Quebec, H3A 2K6, Canada*

Abstract.—Systematic data in the form of collections data are useful in biodiversity studies in many ways, most importantly because they serve as the only direct evidence of species distributions. However, collecting bias has been demonstrated for most areas of the world and has led some to propose methods that circumvent the need for collections data. New methods that model collections data in combination with abiotic data and predict potential total species distribution are examined using 25,111 records representing 5,123 species of plants and animals from Guyana; some methods use the reduced number of 320 species. These modeled species distributions are evaluated and potential high-priority biodiversity sites are selected based on the concept of irreplaceability, a measure of uniqueness. The major impediments to using collections data are the lack of data that are available in a useful format and the reluctance of most systematists to become involved in biodiversity and conservation research. [Biodiversity; collections data; conservation planning; Guyana; irreplaceability.]

Systematics can and should play an important role in biodiversity studies and conservation biology. From the identification and classification of the organisms involved to the final determination of sites for protected status, systematic data are crucial for conservation and biodiversity studies. Various kinds of systematic data can be used for conservation planning, including collections data, phylogenetic data, classifications, and observational data. In this study, we use collections data, defined here as species data from vouchered collections. These data are crucial because they are a permanent record of a species at a given location at a specific time and because they can be checked by experts for proper identification. There are different ways to approach the use of collections data in conservation issues, and careful consideration must be given to the planning and execution of studies. The underlying assumption of all conservation planning is that we use the best available data at any given time—but when using collections data, what constitutes the best data? Unfortunately, a database that describes the full distribution of all biodiversity is rarely, if ever, available. This leaves no choice but to use surrogates such as vegetation maps, numerical classification of environmental variables, species data for selected taxonomic groups, or genetic data to describe biodiversity (e.g., Mackey et al., 1988, 1989; Belbin, 1995; Pressey and Logan, 1995; Faith and Walker, 1996; Pressey, 1997; Moritz and Faith, 1998; Wessels et al., 1999; Margules and Pressey, 2000). However, none of these surrogates is robust enough on its own to be heralded as the best. The most convincing biodiversity surrogates may come from a combination of environmental variables and species data (Ferrier, 2002); however, an analysis of any country or region will always be limited by the availability of data.

The question often arises: Are collections data of any use for conservation decision-making? Some would argue they are not, that collections data are too flawed to be of any use. The limitations to using collections data are that they can be (1) geographically biased, favoring more easily accessed areas; (2) taxonomically incomplete, including only easy-to-study species, which gives undue weight to a few taxa; and (3) temporally biased, based on one survey, and that one usually not carried out during the wet season (Faith and Walker, 1996; Ferrier, 1997; Funk et al. 1999). Among the techniques developed to deal with these limitations are modeling known species records on biophysical data to create a more geographically complete map of known and modeled species distributions; using "indicator" taxonomic groups to represent other taxonomic groups; and using additional historic data from museums and herbaria, expert knowledge and sight records to supplement recent field surveys (Austin and Margules, 1986; Margules and Austin,

---

[3] Present address: Cooperative Research Centre for Rainforest Ecology and Management, Department of Zoology and Entomology, University of Queensland, St. Lucia, Queensland, Australia 4072.

1994; Ferrier, 1997; Howard et al., 1998; Van Jaarsveld et al., 1998). In addition, several techniques have been developed to circumvent the use of species data altogether. These techniques use abiotic surrogates of biodiversity, including land classifications, vegetation maps, numerical classification of environmental variables, and ordination of environmental variables (Mackey et al., 1988, 1989; Belbin, 1993, 1995; Pressey and Logan, 1995; Faith and Walker, 1996; Wessels et al., 1999; Faith et al., 2001). The problem with these latter techniques, however, is that they are not informed by the biological data (Ferrier et al., 1999). As a result, decisions on what areas to conserve are made without regard to what species may or may not be in those areas.

When one has decided to incorporate collections data into conservation planning, there exists a plethora of methodologies that can be used to select priority biodiversity sites. Most conservation planners accept that a network of conservation sites needs to be complementary in nature, that is, where each site complements the biodiversity of other sites within any given network (Vane-Wright et al., 1991; Pressey et al., 1993, 1994; Margules et al., 1994). Incorporating the concept of complementarity ensures that sties are selected to maximize the representation of different species. Recently the concept of complementarity has been enhanced by the introduction of the concept of irreplaceability (Pressey et al., 1993; Ferrier et al., 2000). Irreplaceability refers to a measure of uniqueness, such that the irreplaceability value of a site reflects the relative importance of that site for achieving an explicit conservation target (Ferrier et al., 2000). Although the political decision to gazette a site for conservation may depend on further analyses of economic, political, and other potential land uses, this complementarity/irreplaceability approach has been used successfully to select areas of high biodiversity priority in Australia, South Africa, and the USA (Pressey et al., 1993; Rebelo, 1994; Pressey, 1994; Lombard et al., 1997, 1999; Davis et al., 1999).

This study presents examples of how collections data can be used in conservation planning in Guyana. It does not present the definitive maps of species richness, restrictedness, or high-priority biodiversity sites for Guyana; rather, it demonstrates how conservation planning techniques can be applied to collections data. Collections data are used to explore the options and demonstrate the role these data can play in (1) identifying sampling gaps; (2) improving survey design; (3) reducing collecting biases; (4) building richness, restricted range (endemicity), and distribution maps; and (5) selecting priority biodiversity sites for possible conservation. We contend that not only are such data useful, they are vital if one wants reliable results. In this study, data taken from collections housed in museums and herbaria, as well as recent survey data all from Guyana, South America, were stored in a Geographic Information System (GIS) and were used to provide examples on how those collections data might be employed.

First, collections data are used to examine how well data from known localities held in the GIS database sample actual environmental and geographical space. The identification of sampling gaps provides information on how complete surveys are, where new surveys need to be carried out, and whether further studies should target sites that are geographically or environmentally different from previous collecting sites.

Second, collections data are used to generate species-richness maps for both "known locality" data and "modeled distribution" occurrences. The modeled distribution of a species can be estimated by using either a presence-only modeling technique or, if presence–absence data are available, more statistically sophisticated modeling techniques. In the case of Guyana, the majority of the data used here were collected in an ad hoc manner and hence are presence-only. The predicted distribution of a species can potentially reveal whether it is a widespread species, a rare species, or a species for which distribution is highly correlated with climatic, terrain, or substrate properties (e.g., a high-elevation species). Predicted distributions of species can be used in the same way as species data are used to derive maps showing species richness, restricted ranges/endemicity, or rarity. Known localities and modeled distribution maps, used in conjunction with expert knowledge, can serve as both a management tool (e.g., for endangered, rare, or threatened species) and a conservation-planning tool to locate important biodiversity sites. Here known locality data and modeled species occurrence are used to provide an example of how species

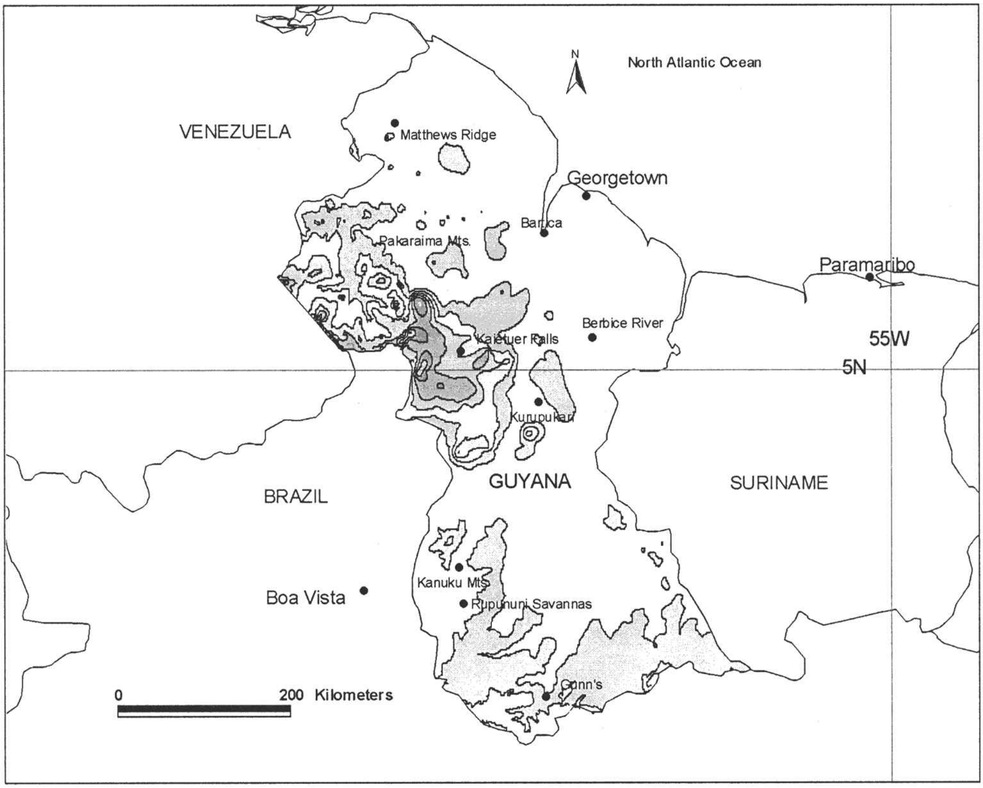

FIGURE 1. A map of Guyana showing location, 250-m contour lines, and localities of interest to this study.

richness and restrictedness vary when modeled data are used.

Lastly, known locality and modeled distribution data are combined and used to select priority biodiversity sites that maximize the complementarity and summed irreplaceability of sites using an interactive conservation-planning tool, C-Plan (New South Wales National Park Service [NSW NPWS], 1999). As each site is selected to meet the required conservation target, the "accumulation of species" is examined for plants and animals to demonstrate the response of different groups. Although different conservation targets can be applied to the same dataset, the example presented in this paper demonstrates how collections data can be used as a biodiversity surrogate in Guyana with a fixed target of representing each species in at least two high-priority biodiversity sites. A target of two was chosen as an example that would strive to build some redundancy into a network of selected sites; in a real-world situation at least one representation of a species would be a known locality from a vouchered specimen.

Why Guyana (Fig. 1)? Many of the problems facing other countries are not an issue in Guyana: It is not a large country (215,000 km^2), a large amount of its land is intact or only marginally damaged (ea. 70%), and it has a small human population concentrated along the coast (ea. 800,000 in 10% of its territory). In addition, although previously poorly known biologically, exploration in the past 18 years has generated a wealth of information for some organisms in some parts of the country. Most importantly, although there are few protected areas in Guyana, the government of Guyana is interested in developing a National Protected Area System. Based on this initiative, several preliminary studies on how existing data might be used in the development of a National Protected Areas System have been conducted (Ter Steege, 1998; Funk et al., 1999; Richardson and Funk, 1999; Ter Steege et al., 2000) but no final decisions have yet been made.

METHODS

The techniques presented in this paper use a set of collections data and environmental variables developed and modified for this study. For several of the analyses, only some of the environmental variables and/or a partial dataset of the collections data is used. In all cases, the methodologies for data collection and the creation of environmental layers are identical and are summarized below.

Collections Data

The collections data were provided by the Biological Diversity of the Guianas Program (BDG) of the Smithsonian Institution (www.mnh.si.edu/biodiversity/bdg). For the past 18 years the BDG has had an active field program in Guyana and has developed a database of many of the historical specimens housed at major institutions in Europe and the USA. Data were collected from both historical collections from Guyana housed at museums and herbaria around the world and from recent field collections. When using collections data to establish the biodiversity of an area, one assumes that the species collected are representative of the diversity, that the area is well-collected, that the identifications are correct, and that the species distributions themselves are a good indicator of biodiversity. Although some of these assumptions may not hold true in the case of Guyana, collections data provide the only direct evidence of the distribution of a species. Data from recent collections usually have more detailed information, may have more accurate identifications, and in the past 10 years usually have been georeferenced by using Global Positioning System (GPS) technology. Historical specimens are valuable because they have information on past distributions; however, habitat modification may have altered or eliminated the species at a particular location. Gathering data from historical specimens is a time-consuming task because whenever possible one must not only check the identification but also try to locate the collecting location for georeferencing. However, this process is still faster and cheaper than going into the field. One important feature of the BDG database is that all species records are "specimen-based" (no observational data are included); as a result, every record has a voucher for which the identification can be verified. Specimen-based data allows one to adjust for nomenclature changes and easily accommodates the splitting of species and the description of new species. The majority of the data for the historical collections come from seven institutions (in alphabetical order by country): the Royal Ontario Museum, Canada; the Royal Botanic Gardens at Kew and the Natural History Museum, England; the University of Guyana, Guyana; the University of Utrecht Herbarium, The Netherlands; and the Smithsonian Institution, American Museum of Natural History, and the New York Botanical Garden, USA.

Data were entered into a DBase IV database especially designed for the BDG program (Funk et al., 1999). Only species records that were geocoded in the database were included (31% of initial records were not used because of geocoding problems). In total, 5,123 species (25,111 records) are represented, including 4,482 species of plants and 641 species of animals.

Environmental Variables

Data on climatic variables and vegetation types were assembled from various sources. Temperature and rainfall data were taken primarily from the Hydrometeorological Service of Guyana records, which recorded weather data from 234 stations over 20 years in Guyana. These data were also compared and where necessary complemented with data collected by the Centre for Resource and Environmental Studies in Australia, which compiled the rainfall and temperature data for a Commonwealth-funded project that used meteorological information on Guyana. Average monthly rainfall data for 72 weather stations over 16 years were entered into a database. The 72 stations were chosen from the 234 countrywide stations on the basis of two criteria: uninterrupted rainfall data for at least 10 years and accurate knowledge of the station's location. Average minimum and maximum monthly temperatures for 45 weather stations over 14 years were also entered into a database. The stations for temperature data were selected according to the same criteria as for rainfall. A digital elevation model (DEM) of the country was obtained from the U.S. Geological Survey (1996) at a scale of 1:1,000,000, a grid size of approximately 1 km^2. Digitized streams and rivers and spot heights were used to correct drainage basins with ANUDEM (Hutchinson, 1989), a program that interpolates elevation data onto a regular grid by using a finite-difference method. The rivers and streams were digitized from the 1:100,000 topographical maps of Guyana with ARC/INFO (ESRI, 1998). Monthly mean climate data for temperature and rainfall were spatially interpolated for the entire area of Guyana by using ANUSPLIN (Hutchinson, 1993), a program that calculates climate surfaces from individual points for which the longitude, latitude, and elevation are known. The climatic surfaces were then fitted to the DEM by using ANUCLIM (Hutchinson, 1998) to produce regular grids of monthly mean climate at the same spatial resolution as the DEM (approximately 1 km^2). All environmental variables were created at a grid cell resolution of 1 km^2.

Sampling Gaps

The role of collections data in conservation planning can be greatly enhanced by using simple techniques that stratify collecting regimes in terms of environmental and geographical space. In Guyana, where research money to conduct more collecting expeditions is very difficult to acquire and where trips into the interior are expensive, each survey can potentially improve the conservation-decision process by sampling either a new locality or a new taxonomic group. The identification of sampling gaps is important because it allows one to make the most of research funds by maximizing the amount of information acquired during each new expedition. One way to identify sampling gaps is to look at the distribution of known collecting sites within the total environment space.

To illustrate how collections data can be used to examine sampling gaps in environmental space, two environmental variables were chosen from the fitted surfaces produced by ANUCLIM: mean annual rainfall and mean annual temperature. These variables were chosen as they are the most readily available environmental variables in most countries and are commonly used, however data for any other variables such as temperature of the hottest period, precipitation of the wettest month, ruggedness, lithology, and soil fertility can be used. The total environmental space for these two variables was determined by plotting the annual temperature and annual mean rainfall for every grid cell in Guyana. The annual mean temperature and annual mean rainfall were then plotted for each collecting site (Fig. 2).

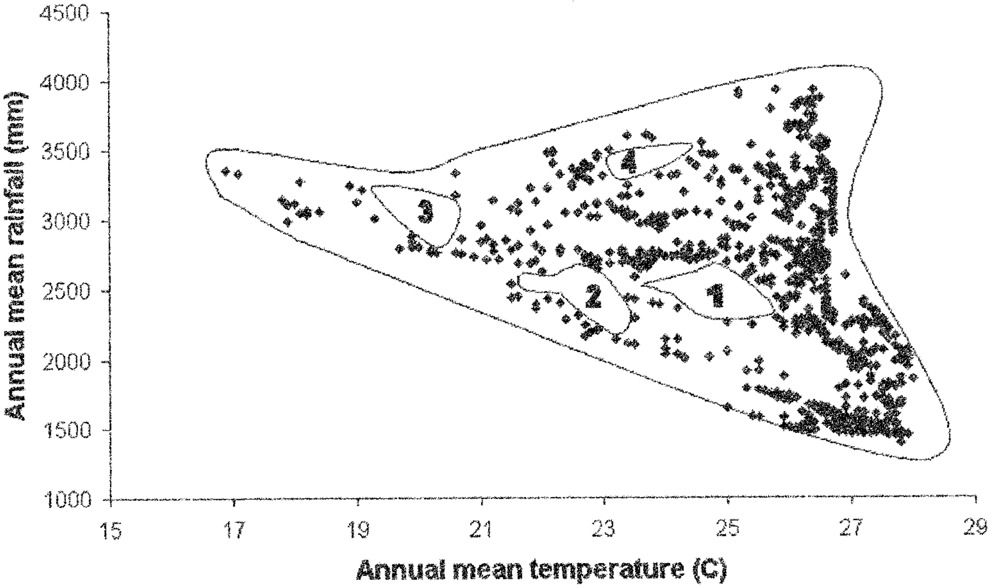

FIGURE 2. Plot of the total environmental space, based on annual mean temperature (C) and annual mean rainfall (mm) (polygon), and actual collecting sites (black diamonds). Polygons 1, 2, 3, and 4 are examples of the obvious gaps that exist, illustrating that not all of the environmental space available in Guyana has been sampled.

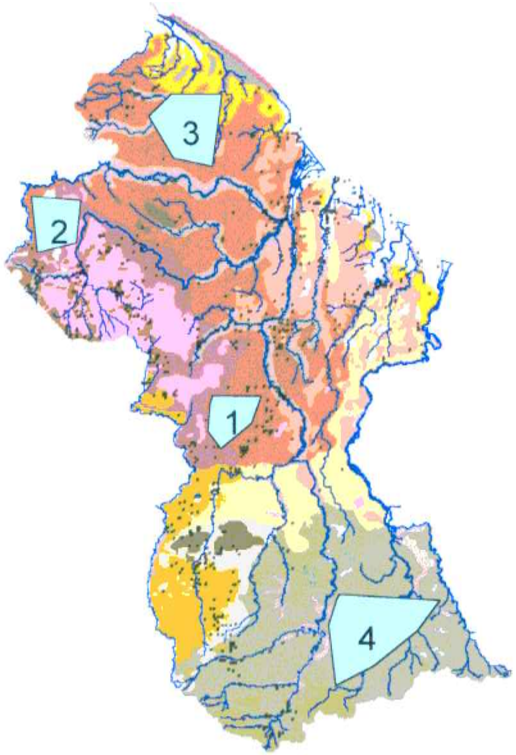

FIGURE 3. A map of known collecting sites and the four polygons representing the sampling gaps drawn in Figure 2 drawn over the vegetation classes (modified from Huber et al., 1995) and main rivers of Guyana. The different vegetation classes are described in more detail in Huber et al. (1995).

Four polygons were drawn to demonstrate some of the sampling gaps (Fig. 2). Although these polygons are arbitrary in their size and shape, they indicate real sampling gaps in the collections data. The four polygons were projected back onto the map of Guyana to illustrate where those gaps were located geographically and where the collecting sites were located in relation to the different vegetation types (modified from Huber et al., 1995) and the major rivers, which constitute the primary access routes into the interior of Guyana (Fig. 3). The plot of collecting sites in Guyana revealed gaps in the areas of middle to high rainfall and temperature ranges (Fig. 2). These gaps are most apparent in the Pakaraima Mountains, the central lowland areas, and the upper Berbice area. Notable gaps in sampling different vegetation classes in Guyana include areas in the following:

Polygon 1: Tall/medium, evergreen, non-flooded forest (rainforest) and low, evergreen seasonally flooded forest. These forests are commercially valuable forests and are primarily under the control of different concessionaires;

Polygon 2: Tall/medium, montane sclerophyllous forests, found in the Pakaraima Mountains, which are difficult to access;

Polygon 3: Tall/medium, evergreen, non-flooded forest (rainforest), commercially valuable, and tall/medium pre- and basimontane forests of the foothills of the Pakaraima Mountains; and

Polygon 4: Tall, evergreen hill-land forests found in the southeast of the country, currently inaccessible for logistical and political reasons.

When examined countrywide, a clear pattern of sampling along roads, rivers, and streams is apparent, and sampling gaps correspond to areas of high elevation and parts of the country that are difficult to access. Specifically, areas in the Pakaraimas and southeast are undersampled. Also undersampled are the lowland forests, where access is limited by forestry concessions. However, preliminary data from these areas indicate a great potential for high species diversity.

Modeled Distribution and Species Richness Maps

Sites with greater species richness have generally been considered more important for conservation than the sites deemed species-poor (Myers, 1988, 1990; Mittermeier and Werner, 1990). Given that complete inventories of species are impractical, particularly in species-rich tropical areas, the utility of species richness and other species-based approaches depends on the extent to which results from limited data sets can be generalized. In this study, we used known locality data and potential distribution data from modeled distributional maps to enhance species-richness maps. However, other studies have also examined how well certain taxonomic groups act as indicators for other taxonomic groups (Pearson and Cassola, 1992; Prendergast et al., 1993; Williams et al., 1996; Howard et al., 1998; Van Jaarsveld et al., 1998; Moritz et al., 2001). For computational reasons, a subset of the total species database was used to examine differences in species richness and restricted range by using known locality and modeled data. This reduced database, comprising 320 species from both animal and plant groups, was used because each had 10 or more locational points across Guyana and could be reasonably modeled using the presence-only modeling tool, DOMAIN (Carpenter et al., 1993). The threshold of 10 locational points or greater was chosen after examining the results of modeling both fewer and greater minimum numbers of locational points and determining at which number the error in predicting distributions stabilized. The accuracy of the predicted models of species that used 10 or more locational points was examined by removing some of the data for each species before modeling and then testing whether known sites were predicted. Expert opinion was used to verify the distributions. However, because of the "10 or more" selection criterion, this method cannot address questions involving rarity.

The potential distribution of each species was modeled by using DOMAIN, a presence-only data modeling technique (Carpenter et al., 1993). Modeling distributions of species assumes that differences in species composition and abundance at any given location can largely be explained by differences in environmental factors, such as temperature, moisture, nutrients, and evaporation (Nix, 1982; Austin et al., 1994; Busby, 1986; Margules et al., 1988; Belbin, 1995). In the case of Guyana, modeling was used to increase the geographical coverage of the likely distribution of a species and remove some of the sampling biases when species records were collected from opportunistic and easily accessible sites. However, the reliability of modeled species data is still partially a function of the degree of spatial biases present in the locality data (Margules and Pressey, 2000).

The steps taken to model the potential distributions of each species were as follows:

1. The digital elevation model, vegetation map, lithology map, and mean monthly rainfall of the driest month (October) were selected as the variables for modeling species distributions. These variables were verified by an expert as relevant to the distribution of plants and animals in Guyana.
2. Each species was modeled by the DOMAIN program using the selected variables. Figure 4 presents an example of the known locality data for a species and the modeled distribution from DOMAIN. Preliminary verification by an expert has confirmed the likelihood of the modeled distribution.
3. A similarity map was produced for each species, showing the likelihood of the species being present in a given area. The similarity maps of the modeled distribution of each species were reclassified in a GIS program (ArcView version 3.2 [ESRI], 2000) to show the known localities and modeled distributions having a similarity value of 95% or greater. A similarity value of 95% or greater was chosen as a conservative cutoff point for the modeled distribution of a species (Fig. 4).
4. The modeled distributions were then used to improve the species-richness map. The species-richness maps for the known

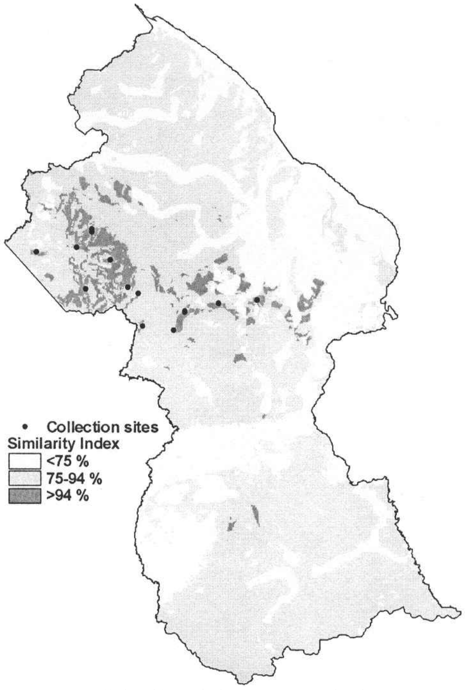

FIGURE 4. A map illustrating the known distribution of a plant species (Melastomatcecae; *Leandra purpurea* Gleason: black dots) and the modeled distribution of the same species (shaded gray areas) based on modeling the known localities with four abiotic factors by using DOMAIN. Three levels of similarity are shown; only a similarity index of 95% or more was used in the analysis.

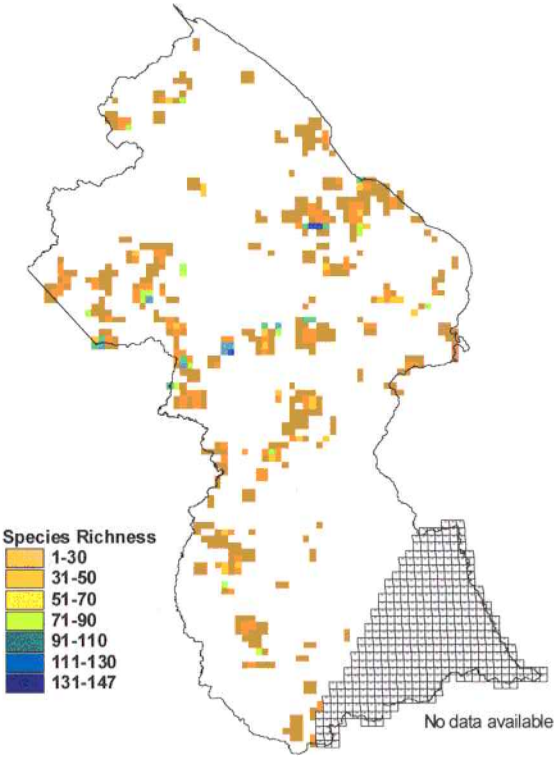

FIGURE 5. Species richness based on known locality data for the reduced dataset of 320 species.

locality data-only and for the known locality plus modeled distribution data were created by superimposing the distribution of each species on top of one another in a GIS (ArcView version 3.2; ESRI, 2000).

5. A map of the restricted range values, a measurement of endemism, was calculated in the same manner, by using an index of restricted range. The restricted range index was calculated by counting the number of grid cells in which each species occurred, taking its inverse, and summing the total for each grid cell, as follows:

$$\text{Restricted range index} = \sum_{i=1}^{S} Q^{-1}$$

where S is the total number of species, and Q is the total number of grid cells included in each species range. Thus, species with very restricted ranges had higher scores, the most restricted species (those occurring in only one grid cell) scoring 1.0 on the restricted range scale. Data were transformed using log $(x + 1)$ to normalize the variance. Both the known locality data and the modeled data were used to calculate restricted range values and to produce the resulting map of restricted range.

The species-richness maps of the known locality data-only and the known locality plus modeled distribution maps for 320 species reveal differences in the number of species in a given grid cell and the distribution of the sites with the greatest species richness (Figs. 5 and 6). The species richness maps of Guyana based on only the known locality data showed variations of from 1 to 147 species per grid cell (Fig. 5). Only 0.10% of the total grid cells had high species richness (131–147 species). Conversely, 39% had low species richness (1–30 species). Using the known locality and modeled distribution data yielded species richness values ranging from 10 to 185 species, and the number of grid cells that had high species richness increased to 8.5% (Fig. 6). Some of the minor differences between the richness maps result from the aggregation of cells when the data are modeled. These differences can be reduced if a smaller grid cell is used. Based on the species patterns of this map (Fig. 6), some of

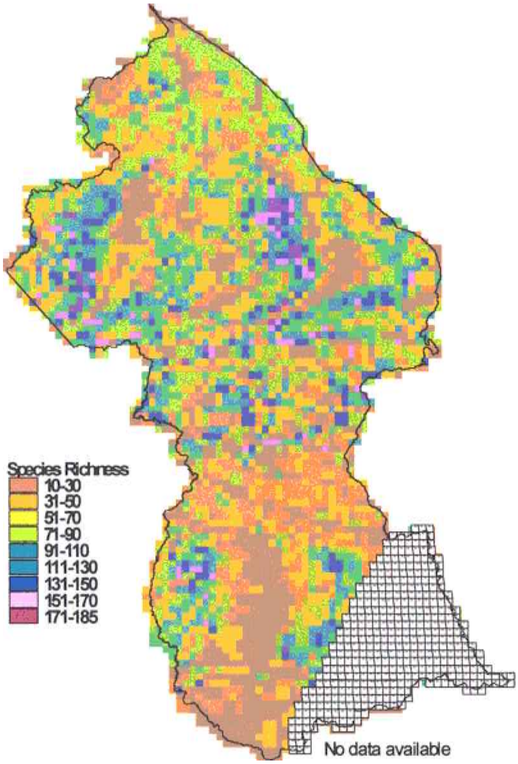

FIGURE 6. Estimated species richness based on known locality data and modeled distributions for the reduced dataset of 320 species.

the areas of estimated high species richness are Kaieteur Falls and the Potaro river gorge; Kurupukari and the central Essequibo River; Bartica and the lower Cuyuni, Mazaruni, and Essequibo Rivers; the Pakaraima Mountains, including Mts. Ayanganna, Roraima, and Wokomon; the Upper Cuyuni and Mazaruni rivers; the Kanuku Mountains and the Rewa River; the Upper Berbiece River; and a few scattered areas in the Rupununi Savannas. This disregards the sites in the far southeast of the country, however, where virtually no collecting has taken place and very few abiotic data are available. Areas with apparent low species richness (especially in the south) may turn out to be richer than shown when all the species data (5,123 species) are used rather than the reduced dataset of 320 species because poor collecting in the southern part of the country affects the modeled distributions, as the southern parts of the country have unique vegetation and rainfall/temperature patterns and therefore appropriately show a low species number when modeled.

The restricted range index varied for known locality data-only and known locality and modeled distribution data in a similar manner as for species richness (Fig. 7). Interestingly, the overlap between areas of high species richness and areas of very restricted species was quite high (71.4%). Perhaps certain species-rich areas in Guyana are also centers of endemism; moreover, these areas may have similar biogeographical features.

Location of Priority Biodiversity Sites

One of the main drawbacks with using only species richness or restrictedness data to select priority biodiversity sites for conservation is that such data do not ensure that the different species in an area are conserved. For instance, a grid cell might be relatively species-poor, but if that cell adds the most species not already represented in an existing network of conservation sites, then it may be the cell that is most important in terms of conservation (Flather et al., 1997). However, one can represent many more species in a network of sites if decisions are made to use species richness or restricted range values and the complementarity principle to select sites (in this case, an index of summed irreplaceability) (Vane-Wright et al., 1991; Pressey et al., 1993; Margules et al., 1994).

The known and modeled data of the 320 species were used to select priority sites for biodiversity conservation using a grid size of 8 km × 8 km (2,978 grid cells across the country, each being 64 km^2). This grid size was chosen for demonstrative purposes only. Although the size of the grid is arbitrarily set in most conservation planning exercises, the size may influence the quantity and location of priority biodiversity sites (Flather et al., 1997; Reid, 1998). Excluded from analyses were 418 of the grid cells from the bottom southeast corner of the country, an area from which very little information has been collected because of logistical and political problems. In addition, 157 grid cells representing urban areas and cultivated fields were removed from the analyses as little natural biodiversity is left in them.

Priority biodiversity sites were selected by using an interactive software package, C-Plan (NSW NPWS, 1999), which runs as an extension in ArcView (version 3.2; ESRI, 2000). Sites were selected to maximize the

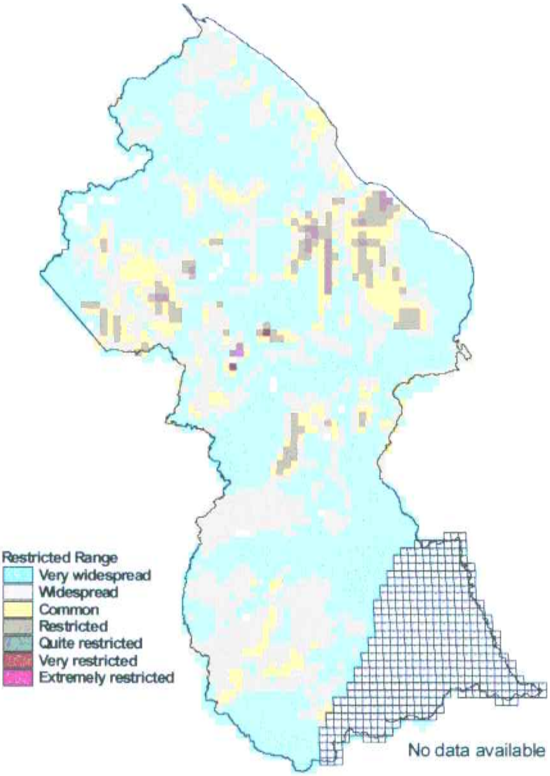

FIGURE 7. Restricted range values based on known locality data and modeled distribution data for the reduced dataset of 320 species.

rate of species accumulation by an iterative process based on estimated summed irreplaceability, defined as the sum for all species of the likelihood that a site would be required as part of a network of sites to achieve a set target, which in this instance was the representation of each species within at least two sites (Pressey et al., 1993; Ferrier et al., 2000). The minimum set of sites needed to satisfy the target was calculated using an interactive stepwise algorithm that selected sites based on their highest summed irreplaceability. Every time the algorithm selected a site, the potential contribution of all the other sites was recalculated and the next most appropriate site was selected. C-Plan allows the user to verify the species composition of each site selected and keeps a log of the order of selection.

For Guyana, in order to capture each species at least two times in a network of sites were required 27 grid cells out of a possible 2,978 (Fig. 8). These 27 sites were selected to maximize the complementarity of species between the sites and the relative irreplaceability value of each site. A few sites are in the northwest near the Venezuelan border and the coast, and several are in the Pakaraima Mountains, including in the vicinity of Kaieteur, Mt. Ayanganna, and the upper reaches of the Mazaruni River. Three sites are in the northeast corner; the Essequibo River between Bartica and Kurupukari has three sites; and the Berbice River has several. Below the 4th parallel are four sites: two in the Kanuku Mountains, one near the border with Surinam, and one in the far south just north of Gunn's. No doubt additional data from southern Guyana would change the results.

The rate of species accumulation for plants and animal shows that 80% of the plant species (95% of animal species) are represented in 19 grid cells (Fig. 9); however, an additional 8 grid cells are necessary to capture all species at least two times. Although species have been divided only into plants and animals, further work can be done to examine the efficiency of a taxonomic group to

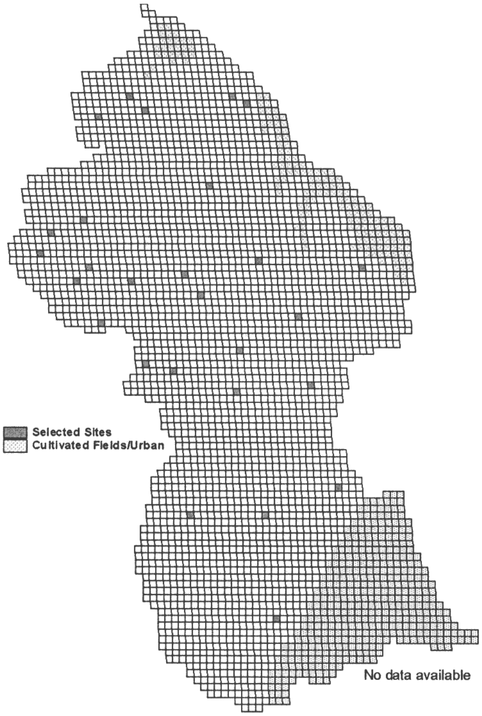

FIGURE 8. The location of high-priority biodiversity sites (dark squares) selected in C-Plan, given the options selected; other sites are probable given other options. Also indicated are cultivated fields and urban areas (shaded) and the area of the county with insufficient data (gray).

act as a surrogate for other taxonomic groups in selecting priority biodiversity sites.

CONCLUSIONS

Understanding patterns of biodiversity may be the key to conserving remaining species, especially in tropical areas such as Guyana. Although collections data may not be the perfect surrogate of biodiversity, they can assist in several ways in the prioritization of important biodiversity areas, as demonstrated in this paper. When sites were selected that maximized the summed irreplaceability values, 27 sites were required to capture each species at least two times in a network of sites. Although the number of sites needed to represent each species a given number of times may vary with the size of the grid, the rules of selection, and the amount of collection data used, at a minimum the list of sites acts as a starting point for conservation planners. For Guyana, a relatively poor country, with limited means to conserve its biodiversity, having a biologically meaningful start to the design of a National System of Protected Areas is a huge step forward.

As with most land use issues, conservation decision-making is largely a matter of real estate, and location is everything (Kiester et al., 1996). At least two factors influence the choice of locations for conservation. First, conservation areas have to compete with other land uses, particularly those that deliver short-term economic benefits to governments and are perceived to be necessary for economic development. Second, different species occupy different locations and in general, overall species diversity increases with area (Connor and McCoy, 1979). Thus, finding locations that are available, species-rich, and relatively large is one of the main challenges for conservation. Identifying locations that are appropriate for conservation is a multiple-step procedure. It involves collecting geographical, biological, political, and in some cases, social data; analyzing the data; and finally weighing potential trade-offs with other land uses. Because other land uses such as agriculture and forestry are often in competition for the same land, the decision to protect certain locations is usually done in the context of national land use planning. Future conservation sites in Guyana may differ from the high-priority biodiversity sites identified in this study by incorporating some assessment of the urgency or "vulnerability" with which priority biodiversity sites should be conserved. Sites that have a high biodiversity priority may have a low vulnerability index and therefore a low overall conservation priority. Such sites might include tops of mountains, steep slopes on mountains, and remote forests, which are all self-protecting to some degree. On the other hand, medium-priority biodiversity sites may be very threatened by human activities and be under immediate threat of losing their biodiversity. Those sites would be high-priority conservation sites. Other considerations not incorporated in this paper with regard to prioritizing sites include conserving sites for long-term persistence of evolutionary processes and genetic diversity (Moritz, 2002) and retention of variation within a species (Desmet et al., 2002).

Using collections data in biodiversity studies adds a dimension to the results that are used for conservation decisions. To ask the

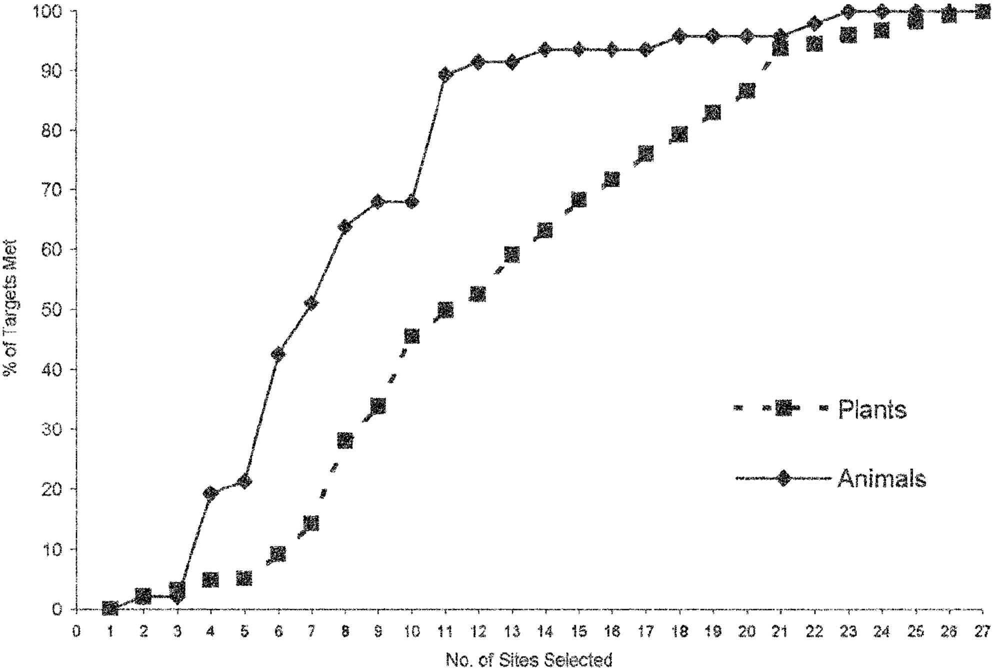

FIGURE 9. The species accumulation curve for plant and animal groups as each site is selected. The graph shows that plant and animal data cannot be used interchangeably and that all sites are required to capture the diversity of the 320 modeled species distributions.

question, "Where should conserved areas be placed?" without collections data means that the areas are selected on the basis of such factors as percentages of ecosystems (as determined by broad-scale vegetation maps) or modeling of abiotic factors such as rainfall and elevation into environmental domains. It is conceivable that decisions could be made to conserve areas where the species composition is more depauperate or when there is a relatively common species composition. Collections data add biotic influence by looking at the species composition for areas as well as the abiotic factors. Stork (1995) asked, "Why are actions affecting biodiversity and conservation based on inadequate information?" and proposed three reasons:

1. Data are unavailable, incomplete, or unreliable.
2. Data are not presented in format useful for policy-makers and managers.
3. Data are incorrectly interpreted.

Clearly, systematists are capable of helping solve these problems but do not usually get involved in biodiversity and conservation projects. Some of the reluctance to get involved in the science of conservation is attributable no doubt to time constraints but also perhaps there is a feeling that conservation is somehow a "softer" science. However, if we don't work with the conservation biologists to make use of collections data in biodiversity studies, then they will continue to move forward with options that utilize only abiotic data rather than including biotic data. Truly, we must use collections data or lose the opportunity to have these data play a role in conservation.

ACKNOWLEDGMENTS

We thank the Smithsonian Institution for (1) base funding provided to the Biological Diversity of the Guianas Program by the National Museum of Natural History (NMNH) Associate Director for Research and Collections; (2) support from the Office of Fellowships and Grants; (3) support from the NMNH Biodiversity Programs Office; and (4) help from the BDG staff, especially Carol Kelloff and Tom Hollowell. In Guyana, we gratefully acknowledge the assistance provided by (1) the Centre for the Study of Biological Diversity; (2) the Faculty of Natural Sciences, University of Guyana; (3) the USAID office; (4) the Government of Guyana; (5) our Amerindian colleagues; (6) and the World Bank–GEF for some funding in the beginning of the project and for funding two preliminary reports produced by Conservation International and the Smithsonian Institution on GIS-based data and historical collections data. We also thank Bob Pressey for allowing us to use C-Plan and Simon Ferrier for discussions and comments on analyses. George Schatz and Simon Ferrier provided many helpful comments on the manuscript. As usual, without the cooperation of herbaria and museums around the world we would not have been able to gather the data necessary for this study, and we therefore thank the following institutions: the University of Guyana, Guyana; the Royal Botanic Gardens at Kew and the Natural History Museum, England; Royal Ontario Museum, Canada; the University of Utrecht Herbarium, The Netherlands; and the Smithsonian Institution the American Museum of Natural History, and the New York Botanical Garden, USA. This is number 60 in the Smithsonian's Biological Diversity of the Guianas Program Publication Series.

REFERENCES

AUSTIN, M. P., AND C. R. MARGULES. 1986. Assessing representativeness. Pages 45–67 *in* Wildlife conservation evaluation (M. B. Usher, ed.). Chapman and Hall, London.

AUSTIN, M. P., A. O. NICHOLLS, M. D. DOHERTY, AND J. A. MEYERS. 1994. Determining species response functions to an environmental gradient by means of a b-function. J. Veg. Sci. 5:215–228.

BELBIN, L. 1993. Environmental representativeness: Regional partitioning and reserve selection. Biol. Conserv. 66:223–230.

BELBIN, L. 1995. A multivariate approach to the selection of biological reserves. Biodivers. Conserv. 4:951–963.

BUSBY, J. R. 1986. A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. Aust. J. Ecol. 11:1–7.

CARPENTER, G., A. N. GILLISON, AND J. WINTER. 1993. DOMAIN: A flexible modelling procedure for mapping potential distributions of plants and animals. Biodivers. Conserv. 2:667–680.

CONNOR, E. F., AND E. D. MCCOY. 1979. The statistics and biology of the species area relationship. Am. Nat. 113:791–833.

DAVIS, F. W., D. M. STOMS, AND S. ANDELMAN. 1999. Systematic reserve selection in the USA: An example from the Columbia Plateau ecoregion. Parks 9:31–41.

DESMET, P. G., R. M. COWLING, A. G. ELLIS, AND R. L. PRESSEY. 2002. Integrating biosystematic data into conservation planning: Perspectives from southern Africa's Succulent Karoo. Syst. Biol. 51:00–00 (this issue).

ESRI. 1998. PC ARC/INFO 3.5.2. Environmental Systems Research Institute, Redlands, CA. http://www.esri.com

ESRI. 2000. ArcView GIS version 3.2. Environmental Systems Research Institute, Redlands, CA. http://www.esri.com

FAITH, D. P., C. R. MARGULES, P. A. WALKER, J. STEIN, AND G. NATERA. 2001. Practical application of biodiversity surrogates and percentage targets for conservation in Papua, New Guinea. Pac. Con. Biol. 6:289–303.

FAITH, D. P., AND P. A. WALKER. 1996. Environmental diversity: On the best possible use of surrogate data for assessing the relative biodiversity of sets of areas. Biodivers. Conserv. 5:399–415.

FERRIER, S. 1997. Biodiversity data for reserve selection: Making best use of incomplete information. Pages 315–329 *in* National parks and protected areas: Selection, delimitation, and management (J. J. Pigram and R. C. Sundell, eds.). Centre for Water Policy Research, Armidale, Australia.

FERRIER, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? Syst. Biol. 51:00–000 (this issue).

FERRIER, S., M. R. GRAY, G. A. CASSIS, AND L. WILKIE. 1999. Spatial turnover in species composition of ground-dwelling arthropods, vertebrates and vascular plants in northeast New South Wales: Implications for selection of forest reserves. Pages 68–76 *in* The other 99%: The conservation and biodiversity of invertebrates (W. Ponder and D. Lunney, eds.). Transactions of the Royal Zoological Society of New South Wales, Mosman.

FERRIER, S., R. L. PRESSEY, AND T. W. BARRETT. 2000. A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for further refinement. Biol. Conserv. 93:303–325.

FLATHER, C. H., K. R. WILSON, D. J. DEAN, AND W. MCCOMB. 1997. Identifying gaps in conservation networks: Of indicators and uncertainty in geographic-based analyses. Ecol. Appl. 7:531–542.

FUNK, V. A., M. F. ZERMOGLIO, AND N. NASSIR. 1999. Testing the use of specimen based collecting data and GIS in biodiversity exploration and conservation decision-making in Guyana. Biodivers. Conserv. 8:727–751.

HOWARD, P., P. VISKANIC, T. R. B. DAVENPORT, F. W. KIGENYI, M. BALTZER, C. J. DICKINSON, J. S. LWANGA, R. A. MATTEWS, AND A. BALMFORD. 1998. Complementarity and the use of indicator groups for reserve selection in Uganda. Nature 394:472–475.

HUBER, O., G. GHARBARRAN, AND V. A. FUNK. 1995. Preliminary vegetation map of Guyana. Biological Diversity of the Guianas Program. Smithsonian Institution, Washington, D.C., USA.

HUTCHINSON, M. F. 1989. ANUDEM version 1.1. Centre for Environmental Studies, ANU, Canberra.

HUTCHINSON, M. F. 1993. ANUSPLIN version 2.1. Centre for Environmental Studies, ANU, Canberra.

HUTCHINSON, M. F. 1998. ANUCLIM version 1.8. Centre for Environmental Studies, ANU, Canberra.

KIESTER, A. R., J. M. SCOTT, B. CSUTI, R. F. NOSS, B. BUTTERFIELD, K. SAHR, AND D. WHITE. 1996. Conservation prioritization using GAP data. Conserv. Biol. 10:1332–1342.

LOMBARD, A. T., R. M. COWLING, R. L. PRESSEY, AND P. J. MUSTART. 1997. Reserve selection in a species-rich and fragmented landscape on the Agulhas Plain, South Africa. Conserv. Biol. 11:1101–1116.

LOMBARD, A. T., C. HILTON-TAYLOR, A. G. REBELO, R. L. PRESSEY, AND R. M. COWLING. 1999. Reserve selection in the Succulent Karoo, South Africa: Coping with high composition turnover. Plant Ecol. 142: 35–55.

MACKEY, B. G., H. A. NIX, M. F. HUTCHINSON, J. P. MACMAHON, AND P. M. FLEMING. 1988. Assessing the representativeness of places for conservation reservation and heritage listing. Environ. Manage. 12:501–514.

MACKEY, B. G., H. A. NIX, J. A. STEIN, AND S. E. CORK. 1989. Assessing the representativeness of the wet tropics of Queensland world heritage property. Biol. Conserv. 50:279–303.

MARGULES, C. R., AND M. P. AUSTIN. 1994. Biological models for monitoring species decline: The construction and use of data bases. Philos. Trans. R. Soc. London B 344:69–75.

MARGULES, C. R., I. D. CRESSWELL, AND A. O. NICHOLLS. 1994. A scientific basis for establishing networks of protected areas. Pages 327–350 *in* Systematics and conservation evaluation (P. L. Forey, C. J. Humphries, and R. I. Vane-Wright, eds.). Clarendon, Oxford, U.K.

MARGULES, C. R., A. O. NICHOLLS, AND R. L. PRESSEY. 1988. Selecting networks of reserves to maximize biological diversity. Biol. Conserv. 43:63–76.

MARGULES, C. R., AND R. L. PRESSEY. 2000. Systematic conservation planning. Nature 405:243–253.

MITTERMEIER, R. A., AND T. B. WERNER. 1990. Wealth of plants and animals unites "Megadiversity" countries. Tropicus 4:4–5.

MORITZ, C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Syst. Biol. 51:00–00.

MORITZ, C., AND D. P. FAITH. 1998. Comparative phylogeograpghy and the identification of genetically divergent areas for conservation. Mol. Ecol. 7:419–429.

MORITZ, C., K. S. RICHARDSON, S. FERRIER, G. B. MONTIETH, S. E. WILLIAMS, AND T. WHIFFEN. 2001. Biogeographic concordance and efficiency of taxon indicators for establishing conservation priority for a tropical rainforest biota. Proc. R. Soc. Biol. Ser. B 268:1875–1881.

MYERS, N. 1988. Threatened biotas: "Hotspots" in tropical forests. Environmentalist 8:1–20.

MYERS, N. 1990. The biodiversity challenge: Expanded hot-spot analysis. Environmentalist 10:243–256.

NEW SOUTH WALES, NATIONAL PARK SERVICE. 1999. C-Plan Module, version 3.03. http://www.ozemail.com/~cplan

NIX, H. A. 1982. Environmental determinants of biogeography and evolution in Terra Australis. Pages 47–66 *in* Evolution of the flora and fauna of Australia (W. R. Baker and P. J. M. Greenslade, eds.). Peacock Publishers, Adelaide.

PEARSON, D. L., AND F. CASSOLA. 1992. World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): Indicator taxon for biodiversity and conservation studies. Conserv. Biol. 6:376–391.

PRENDERGAST, J. R., R. M. QUINN, B. C. LAWTON, B. C. EVERSHAM, AND D. W. GIBBONS. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature 365: 335–337.

PRESSEY, R. L. 1994. Ad hoc reservations: Forward or backward steps in developing representative reserve systems? Conserv. Biol. 8:662–668.

PRESSEY, R. L. 1997. Priority conservation areas: Towards an operational definition for regional assessments. Pages 337–357 *in* National parks and protected areas: Selection, delimitation, and management (J. J. Pigram and R. C. Sundell, eds.). Centre for Water Policy Research, Armidale, Australia.

PRESSEY, R. L., C. J. HUMPHRIES, C. R. MARGULES, R. I. VANE-WRIGHT, AND P. H. WILLIAMS. 1993. Beyond opportunism: Key principles for systematic reserve selection. Trend. Ecol. Evol. 8:124–128.

PRESSEY, R. L., I. R. JOHNSON, AND P. D. WILSON. 1994. Shades of irreplaceability: Towards a measure of the contribution of sites to a reservation goal. Biodivers. Conserv. 3:242–262.

PRESSEY, R. L., AND V. S. LOGAN. 1995. Reserve coverage and requirements in relation to partitioning and generalization of land classes: Analyses for western New South Wales. Conserv. Biol. 9:1506–1517.

REBELO, A. G. 1994. Using the Proteaceae to design a nature reserve network and determine priorities for the Cape Floristic Region. Pages 375–396 *in* Systematics and conservation evaluation (P. L. Forey, C. J. Humphries, and R. I. Vane-Wright, eds.). Clarendon, Oxford, U.K.

REID, W. V. 1998. Biodiversity hotspots. Trend. Ecol. Evol. 13:275–280.

RICHARDSON, K. S., AND V. A. FUNK. 1999. An approach to designing a protected area system in Guyana. Parks 9:7–16.

STORK, N. E. 1995. Inventorying and monitoring. Pages 457–543 *in* Global biodiversity assessment (V. H. Heywood and R. T. Watson, eds.). Cambridge Univ. Press, Cambridge.

TER STEEGE, H. 1998. The use of forest inventory data for a national protected area strategy in Guyana. Biodivers. Conserv. 7:1457–1483.

TER STEEGE, H., M. J. JANSEN-JACOBS, AND V. K. DATADIN. 2000. Can botanical collections assist in a National Protected Area Strategy in Guyana? Biodivers. Conserv. 9:215–240.

U.S. GEOLOGICAL SURVEY. 1996. Digital elevation model of South America. USGS EROS Data Center, Sioux Falls, South Dakota.

VANE-WRIGHT, R. I., C. J. HUMPHRIES, AND P. H. WILLIAMS. 1991. What to protect? Systematics and the agony of choice. Biol. Conserv. 55:235–254.

VAN JAARSVELD, A. S., S. FREITAG, S. L. CHOWN, C. MULLER, S. KOCH, H. HULL, C. BELLAMY, M. KRUGER, S. ENDRODY-YOUNGA, M. W. MANSELL, AND C. H. SCHOLTZ. 1998. Biodiversity assessment and conservation strategies. Science 279:2106–2108.

WESSELS, K. J., S. FREITAG, AND A. S. VAN JAARSVELD. 1999. The use of land facets as biodiversity surrogates during reserve selection at a local scale. Biol. Conserv. 89:21–38.

WILLIAMS, P., D. GIBBONS, C. MARGULES, A. REBELO, C. HUMPHRIES, AND R. PRESSEY. 1996. A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds. Conserv. Biol. 10:155–174.

Received 1 March 2001; accepted 1 October 2001
Associate Editor: Ann K. Sakai