# Geographic ranges and β-diversity: Discovering how many tree species there are where

RICHARD CONDIT, ROLANDO PÉREZ, SUZANNE LAO, SALOMON AGUILAR AND AGUSTIN SOMOZA

Diverse tropical communities are poorly known. How many species live where is fundamental to conservation, but complete species lists for non-vertebrate groups are seldom available from any but the smallest sampling units, and ecological theory offers only crude guidance in how to estimate species richness and species ranges from incomplete samples. Our goal is to develop methods for extrapolating tree species diversity and turnover from scattered tree plots. We tested several methods using data from 88 tree inventories in central Panama, combined with information from local floras and checklists. Of 1142 tree and shrub species recorded in plots and transects, 43% were restricted to the wetter climatic zone of the Caribbean slope and 27% to dry and intermediate climatic zones near the Pacific. We predict on theoretical grounds, however, that these figures are inflated because species-turnover is overestimated in incomplete samples. In two well-studied regions, the fraction of shared species increased linearly with the total number of species sampled, as predicted by theory. On this basis, we extrapolated the total number of species shared, concluding that forests at opposite ends of the Panama Canal share 70% of their species, even though single hectares from the two sites shared only 15%. For more remote and poorly sampled areas, our estimates of species richness and turnover remain hypothetical, underscoring the need for advancing theories and methods for evaluating species distribution patterns.

*Richard Condit, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948 USA. E-mail: condit@ctfs.si.edu*

*Rolando Pérez, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948 USA. E-mail: perezro@tivoli.si.edu*

*Suzanne Lao, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948 USA. E-mail: laoz@tivoli.si.edu*

*Salomón Aguilar, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948 USA. E-mail: ctfs@tivoli.si.edu*

*Agustín Somoza, Autoridad Nacional del Ambiente, Albrook, Panama. E-mail: ctfs@tivoli.si.edu*

## Introduction

Many or most neotropical tree species have poorly known geographic ranges. Large areas of forest in Central America and especially South America remain unexplored, and even where tree inventories have been carried out, they are seldom exhaustive. The large number of rare tree species in the tropics ensures that complete inventories are exceedingly time-consuming. Sparse and incomplete samples make it difficult to even hypothesize about factors limiting geographic ranges. Are most species limited by climate or soils? Or do dispersal and history prevent species from reaching areas where they could grow? Do tropical tree species have narrow ranges and high rates of endemism, or are most widespread generalists (Pitman *et al.* 1999, 2001). Species turnover (β-diversity) is an important element of conservation theory, since it determines how many species there are in total (γ-diversity) and how national park systems should be arranged to capture diversity.

Central Panama encompasses a range of climate and topography, and a variety of evidence suggests that climate limits the ranges of many tree species (Pyke *et al.* 2001). For one, observations at Barro Colorado Island on mortality during unusually severe drought suggest that the duration of the dry season is critical in limiting where some species can grow (Condit *et al.* 1995, 1996). Moreover, Panama tree communities show higher β-diversity than regions where climate is less variable (Condit *et al.* 2002). Here we begin with the assumption that some tree species are limited by dry season duration, and move forward to the question of how many species there are and how many are restricted in range to different positions along the rainfall gradient.

We invoke two approaches for extrapolating richness and ranges from scattered small samples. First, we estimate how many species there are in regions of fairly uniform climate and terrain by extrapolating species-area curves developed from precisely mapped inventories. We have a good basis for doing so: there is ample ecological literature on species-area relationships (Rosenzweig 1995) and some theory to back it up (Hubbell 2001). Beyond richness at single sites, we need to know how many species are shared between two different regions – this shows how many species have limited ranges. Far less has been published on this topic (see Cao *et al.* 1997), so we develop a simple theory here to address specifically what we need: how to extrapolate species-overlap between two regions when neither has been completely sampled.

## Study site

In central Panama, extensive forests line nearly the entire length of the Panama Canal, and a second large block of forest lies to the east, in Chagres National Park (Condit *et al.* 2001, Ibáñez *et al.* 2002). Toward the northern and eastern fringes of Chagres National Park, ridges and small peaks reach 600-1000 m ASL; southwest of the Canal there is a small patch of montane forest, also reaching 1000 m in elevation (Fig. 1, and see maps in Ibáñez *et al.* 2002). The rest of the area is below 300 m, but topographically variable with a complex mixture of geological formations (Stewart *et al.* 1980, Pyke *et al.* 2001). Forests of Chagres National Park are mostly old-growth, little disturbed for several centuries or more, while much of the forest along the Canal – which we call the Canal corridor – has been disturbed since the 19th century. Some sites near the Canal, though, such as Barro Colorado Island, have old-growth (Ibáñez *et al.* 2002).

Precipitation comes from the Caribbean, and the mountain chain running through Central America creates a rainshadow. As a result, the Caribbean coast is substantially moister

than the Pacific, with a shorter and less intense dry season (Condit *et al.* 2000). Annual precipitation near the Pacific entrance to the Canal is 1700 mm, and the dry season lasts an average of 129 days. At the Atlantic mouth of the Canal, rainfall is 3000 mm and the dry season lasts 102 days (Condit *et al.* 2000). The northeast part of the watershed gets 3500 mm of rain or more, with a dry season as short as 75 days.

## Materials and methods

### Censuses

We have carried out inventories of trees at 88 forested sites in the Panama Canal area (Fig. 1). Sites were chosen to evaluate different geological formations, and we have attempted to leave no major geographic gaps (Fig. 1). The samples includes 49 tree plots where every individual was marked, measured, identified, and precisely located in a carefully measured area: one 50-ha, one 40-ha, one 5.96-ha, one 4-ha, 37 1-ha, and nine 0.32-ha. In three large plots, every individual tree ≥ 1 cm diameter-breast-height (dbh) was censused. In most other plots, every tree ≥ 10 or ≥ 20 cm dbh was censused, and trees ≥ 1 cm dbh were censused in subquadrats; in the 0.32-ha plots, however, nothing < 10 cm was censused (see Condit 1998, Pyke *et al.* 2001, and Condit *et al.* 2004 for details). At 18 more sites, a square area of either 40 x 40 m (14 sites) or 100 x 100 m (4 sites) was marked and then thoroughly searched for tree species present. Finally, we censused 21 transects, recording tree species present along trails within an area < 1 km². Transects were not designed to cover precise areas nor to include every species.

Three inventories east of the Canal (Fig. 1) had elevation 600-850 m ASL and were considered the eastern-montane region. Three other sites to the southwest, around Cerro Campana, were also 600-850 m ASL (Fig. 1). All other inventories were < 420 m elevation.

All individuals encountered were placed in morphospecies categories. Most of these are fully identified, and we have collected many specimens and published species lists (Condit *et al.* 1996, 2004). Still, about 25% of the morphospecies remain unidentified, but in the analyses presented here, all are included, identified or not.

Ideally, we would compare inventories based on identical effort, but this would greatly hinder our ability to expand the dataset. Two considerations suggest that the enormous variation in census effort and methods can be ignored. First, what really matters is not whether the effort is identical everywhere, but whether the number of species sampled at each site is comparable, and species accumulate in a convenient way. In small samples, the number of new species encountered rises rapidly, but with additional sampling, this rate diminishes. As a result, inventories based on enormously different efforts include reasonably similar species counts. In fact, 78 of 88 inventories had between 70 and 180 species, even though sampling effort varied 1000-fold (200 individuals in 0.32-ha plots to 200,000 in the 50-ha plot). Second, most comparisons we make are based on regional groups of inventories, and this further overcomes differences in effort among individual censuses.

### Regions

To evaluate similarity of tree communities across the isthmus, we aggregated inventories into 11 geographic regions (Fig. 1). In one analyses, we compared two of these – the Sherman and Pacific regions, including the Sherman and Cocoli plots mentioned above plus several smaller inventories (Fig. 1). We then made broader comparisons by further aggregating the 11 regions, first into climatic zones: wet (Santa Rita, Chagres, Montane East, Campana), intermediate (Sherman, Barro Colorado, Pipeline, Laguna), and dry (Gamboa,
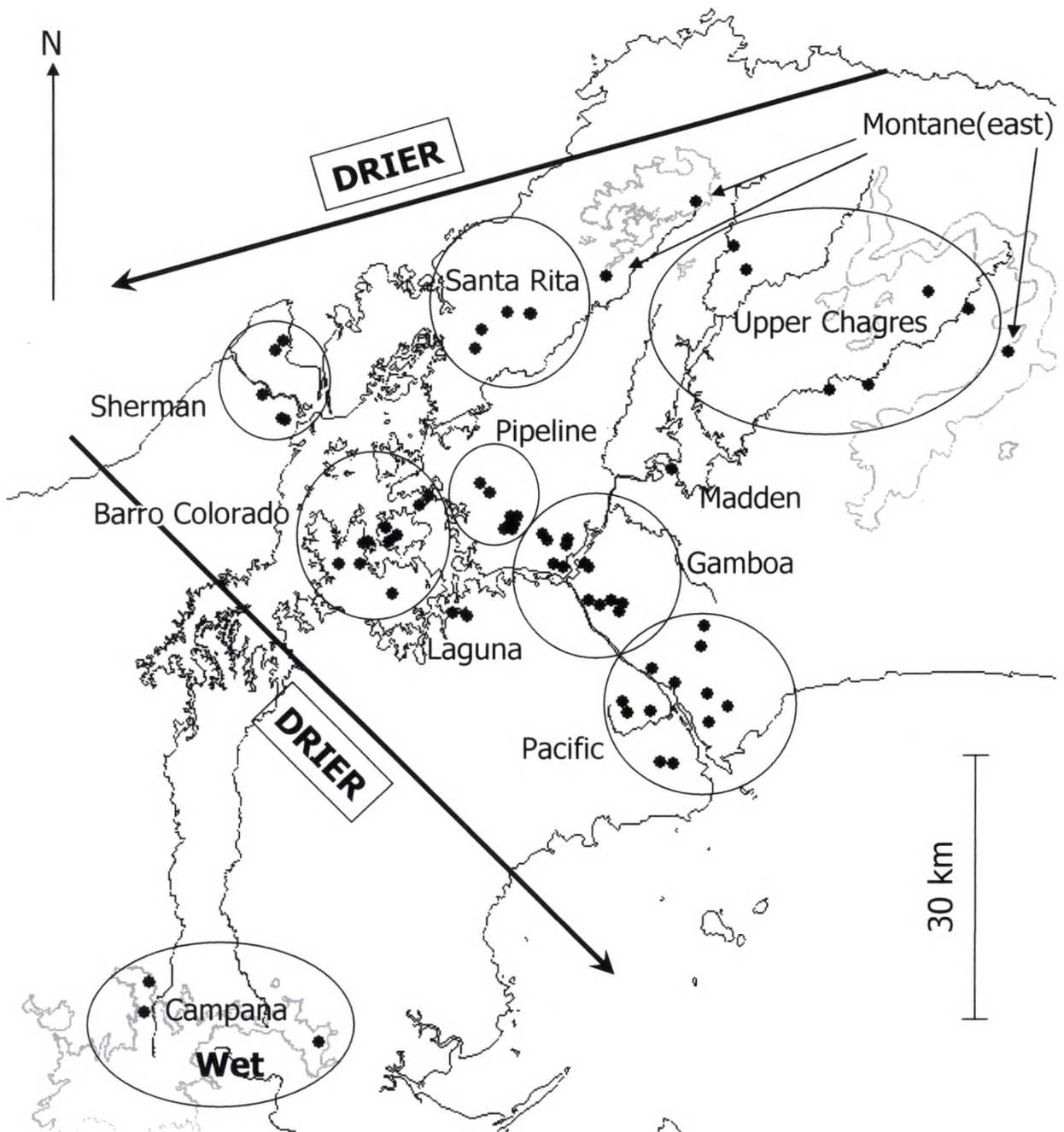
**Fig. 1**. Tree inventories in the Panama Canal area. Sites of inventories are marked with dots (in several places, nearby inventories appear as one dot). Regions described in text are indicated, and larger ones are circled. The climatic gradient is indicated with arrows.

Madden, and Pacific). We then compared the eastern wet forests with the main Canal corridor by aggregating the three regions east of the Canal (Santa Rita, Upper Chagres, Montane east) and comparing this to the aggregation of seven regions near the Canal (all remaining regions except Campana to the west).

### Species richness benchmarks

Our samples of species are incomplete. Our best evidence for this comes from published and unpublished checklists of the flora of Panama. Croat's (1978) flora of Barro Colorado Island includes 436 species which have free-standing stems at least 2 m tall (and thus approximately 1 cm dbh), but the 50-ha plot on the island plus two additional 1-ha plots had just 321 species. In addition, M. Correa (unpublished data) has updated D'Arcy's (1987) published checklist of Panama; Correa's dataset lists 863 tree or shrub species in the Canal Zone. All our inventories in the Canal corridor (roughly the same as the Canal Zone used by D'Arcy) include 650 species. The Correa checklist likewise can be used to count tree and shrub species in each of Panama's provinces and the entire country – 2880 species in 76,000 km² for the latter.

### Extrapolating species richness

From plots where individuals and species were counted, the diversity index Fisher's α was estimated using the procedure given in Condit *et al.* (1998). Fisher's α from one plot can be used to extrapolate species richness in larger samples, since the definition of the index includes a simple relationship between $S$ (species number) and $N$ (individuals):

$$S = \alpha \ln(1 + \frac{N}{\alpha}).$$

Fisher *et al.* (1943) devised the index, and Hubbell (2001) demonstrated that in a theoretical community on a homogeneous substrate (so every species can grow everywhere), this formula describes the relationship between $S$ and $N$.

We also extrapolate species counts in large areas using the power-function species-area-relationship, or SAR: $S = bA^z$, where $A$ = area. Assuming tree density is constant, this can also be given as $S = b'N^z$, with the same $z$ in both cases but a different $b$. Except at small scales, $S$ rises more rapidly with the power function than with Fisher's α, that is, the power function will extrapolate to higher species richness than will Fisher's α.

### Measuring similarity

To evaluate the similarity of tree communities on opposite sides of the isthmus of Panama, we counted the number of species shared between samples of various sizes. To do so, we sub-sampled the 5.96-ha plot near the Caribbean, at a site called Fort Sherman, and the 4-ha plot near the Pacific (called Cocoli). In each plot, species lists were taken from small, square, randomly-placed subquadrats of varying size. We express similarity between two subquadrats with the Sørensen similarity index, defined as the number of shared species divided by the mean of the species total in each of the two subquadrats.

### Similarity theory

All our estimates of similarity are based on incomplete tree species lists. Fisher's α and the SAR allow us to extrapolate species richness, but not the number of species shared. To do so, we present a simple theory of similarity.

Imagine two communities whose total species counts are unknown, and which share an unknown number of species. We sample a fraction of species $f_i$ at each ($i$ = 1,2). $T_i$ is the total number of species in community $i$, and $S_i = f_i T_i$ is the number of species sampled in each. Call the total number of species shared $T_c$ and the observed number shared $S_c$. Now make a

simple assumption: that the average probability of encountering shared species in a sample is equal to the average probability of encountering non-shared species. This assumption holds if species abundances at the two site are uncorrelated, that is, if abundant species at site 1 are absent, rare, or abundant at site 2 just as often as rare species at site 1. We examine this assumption in more detail below, but first use it to estimate $T_c$.

Given uncorrelated abundances, the number of species found shared between samples $S_1$ and $S_2$ is $S_c = f_1 f_2 T_c$. This is because the probability of finding one of the species in sample 1 is $f_1$, and likewise $f_2$ for sample 2, and under the assumption stated above, the two probabilities are independent, so the joint probability is their product. This equation can be rewritten

$$\frac{S_c}{S_1 S_2} = \frac{T_c}{T_1 T_2} = k,$$

because $S_i = f_i T_i$. The left side of this equation is the number of species shared between samples 1 and 2 divided by the product of the species count in each sample; the right side is the number of total shared species divided by the product of total species counts at the two sites. Call this index $k$. It can be considered a similarity index, differing from Jaccard and Sørensen indices in having a product in the denominator rather than a sum (the Sørensen index is

$$\frac{S_c}{0.5 (S_1 + S_2)}).$$

The equation above demonstrates that $k$ has a useful property as a similarity index: it is sample-size independent (but only under the assumption of independent probabilities). It can thus be used to estimate $T_c$, the total number of species shared, given that extrapolated estimates of species-richness are available ($T_1$ and $T_2$). Here we apply this method – after first checking that abundances in two communities are uncorrelated – to estimate the total number of species shared between Caribbean wet forests and Pacific coast forests in central Panama, and then more broadly, between the Canal corridor and eastern wet forests.

### Correlation in abundance

The lack of correlation in species abundance between two sites is crucial to the extrapolation described above. Consider two distinct aspects of this correlation. First, there is the correlation between abundances when species are present at both sites. Lack of correlation then means that the abundance of a species at one site does not predict abundance at the second site. The second aspect covers species present at one site but absent at the second. Here, a lack of correlation means that the abundance of species at site A does not depend on whether they are present or absent at site B.

It is easy in a qualitative sense to understand why the abundance correlation is crucial to similarity estimates from small samples. Initial sampling will uncover abundant species first and rare species later. If two sites have the same abundant species – a positive correlation in abundance – then early sampling tends to encounter shared species more rapidly than the equations above predict. Conversely, if there is a negative correlation in abundance, then small samples over represent non-shared species. Likewise, if species abundant at site A tend to be shared with site B, whereas rare species at site A tend to be absent at B, then small samples at site A are biased, since they disproportionately encounter shared species.

Conceptually, the correlation in abundance is easy to state, but judging whether species abundances at two sites are correlated in real datasets can be problematic. There are three groups of species which must be considered. First are those observed at both sites. We are certain about their distributions, and we can test directly whether their abundances are correlated. The only caveat is that small samples

hinder the estimate. The second group includes those observed at only one site. We do not know whether they are present at the second site, however, we can still include them in the question about abundance correlation. Does abundance at site A differ between those species observed to be absent at site B and those observed to be present at site B? The final group is comprised of species not observed at either site. We know nothing about them. Our only hope is that inferences based on cases one and two apply. If not – if unobserved species behave differently from observed species in terms of similarity – we are sunk, even if we can estimate exactly how many unobserved species there are.

We tested for abundance correlations after log-transforming population densities in plots, since densities were nearly log-normally distributed across species. First, simple regressions were done using those species present at two sites. Next, we compared mean log(abundance) at one site for those species present at a second site versus those absent at the second site. Confidence limits were estimated on log(abundance) using the variance across species and a standard $t$-test. To report abundances, we back-transformed mean and confidence limits, giving density in individuals ha$^{-1}$. We tested for correlation in abundance between the 4-ha Cocoli plot and the 5.96-ha Sherman plot, and then again between 34 small plots in the Canal corridor and 11 small plots in the eastern wet forest (Santa Rita and east-montane regions). In the latter case, abundance for each species was averaged across all plots in a region.

## Results and discussion

### *Range in Central Panama*
A total of 1142 tree and shrub morphospecies were tallied in the 88 inventories; 825 of these have been identified fully to species. The latter

**Table 1.** Local ranges of tree species recorded in inventories near the Panama Canal. Species were categorized with respect to presence-absence in three broad geographic zones (see text, Fig. 1). Widespread means occurring in all three zones.j

| Species range | no. species | % of species |
|---|---|---|
| Widespread | 197 | 17.3 |
| Wet forest only | 492 | 43.1 |
| Wet-intermediate | 113 | 9.9 |
| Intermediate only | 113 | 9.9 |
| Wet-dry only | 10 | 0.9 |
| Dry-intermediate | 130 | 11.4 |
| Dry side only | 87 | 7.6 |
| TOTAL | 1142 | 100 |

includes 12 subspecific taxa: six species have two separable subspecies. It does not include 20 cultivated species which occasionally appear in forest. Hereafter, we refer to the 1142 morphospecies or subspecies as 'species'.

By far the largest number of species was confined to the wet climatic zone – Santa Rita, Upper Chagres, and the higher elevations (Fig. 1, Table 1). The next largest group was widespread, occurring in all three zones. A smaller number of species was confined to the dry zone, or dry plus intermediate (Table 1). Sample distribution maps are shown in Figure 2. The overall trend suggests that species are limited climatically – if species were confined to zones due to dispersal or chance alone, then the number of species confined to different zones ought to be roughly equal. Three additional lines of evidence argue against dispersal limitation being the major cause of limited ranges: 1) the montane-east region is more similar to the distant montane region at Campana than it is to the intervening Pacific region, sharing 29% of its species with the former and just 7% with the latter; 2) a single limestone site in the Sherman region is more like the Pacific sites than nearby Sherman sites; and 3) the Laguna region is more like Upper

**Table 2**. Estimates of species richness in various regions, extrapolating with Fisher's α or an SAR (Fig. 3).

|  | Barro Colorado | Sherman | Pacific (Cocoli) | Canal corridor |
|---|---|---|---|---|
| large plot area (ha) | 50 | 5.96 | 4.0 | – |
| species in large plot | 301 | 227 | 169 | – |
| mean α in all plots | 34.3 | 35.6 | 28.4 | 33.7 |
| species in all inventories | 321 | 336 | 311 | 650 |
| area of entire region (ha) | 1567 | 15,000 | 11,000 | 62,000 |
| estimated individuals in entire region (x $10^6$) | 7.2 | 53.1 | 22.5 | ~2500 |
| estimated richness from α | 420 | 506 | 386 | 532 |
| estimated richness from SAR | 505 | – | – | 1000 |
| known species (plots plus flora/checklists) | 453 | – | – | 946 |

Chagres wet forests than the nearby Gamboa sites. Pyke *et al.* (2001) and Condit *et al.* (2002) mentioned the latter two examples.

However, it remains plausible that dispersal limitation is a factor in limiting some species distributions. There are many species (151 of 214, to be precise) occurring in montane forest east of the Canal but not at Campana, and the edaphically unusual Laguna site is missing many wet forest species. A caveat to conclusions about species absence, though, is that samples are incomplete, and it is always possible that missing species really are present in a region but have not been sampled.

### Extrapolated species richness

The total species count of trees and shrubs on Barro Colorado is 453, taking Croat's (1978) flora plus 17 new species we identified in the 50-ha plot. We estimate the uncertainty to be from 424 to 475, because species probably still remain to be discovered and because the lower

limit of 1 cm dbh produces a gray area regarding which species to include. In addition, there are some species Croat includes for which their is doubt, such as cases where no specimen nor living plant is known.

Fisher's α in the 50-ha plot is consistently close to 34 for samples larger than 1000 individuals (Condit *et al.* 1998). Given α = 34.3, the entire island (with 1567 ha and 4581 individuals per ha) would have 420 species (Table 2, Fig. 3). Alternatively, the power function $S = 173* A^{.1455}$ fits data from 1 ha to 50 ha within the large plot; extrapolating this to 1567 ha predicts 505 species (Table 2, Fig. 3). Thus, Fisher's α gives a slightly better prediction at local scale, as Hubbell's (2001) theory predicts, and provides a means for estimating total species richness in the Sherman region and the Pacific region (Fig. 1).

In the Pacific region, three plots have a mean Fisher's α of 28.4, leading to an estimate of 386 species in 11,000 ha of forest near the

**Fig. 2**. Distribution maps of species in tree inventories in the Panama Canal area. Presence is indicated by filled circles,  →
absence by empty circles. A) *Virola koschnyi* Warb. (Myristicaceae); B) *Tetragastris panamensis* (Engl.) Kunze (Burseraceae);
C) *Prockia crucis* P. Browne ex L. (Flacourtiaceae), D) *Hybanthus prunifolius* Schulze-Menz (Violaceae). Light gray lines indicate the 600-m contour; black lines show coasts and rivers.

A) *Virola koschnya*
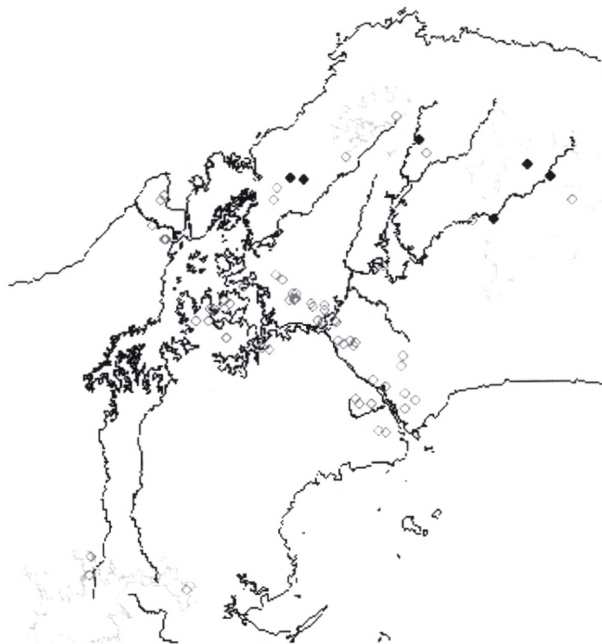
B) *Prockia crucis*
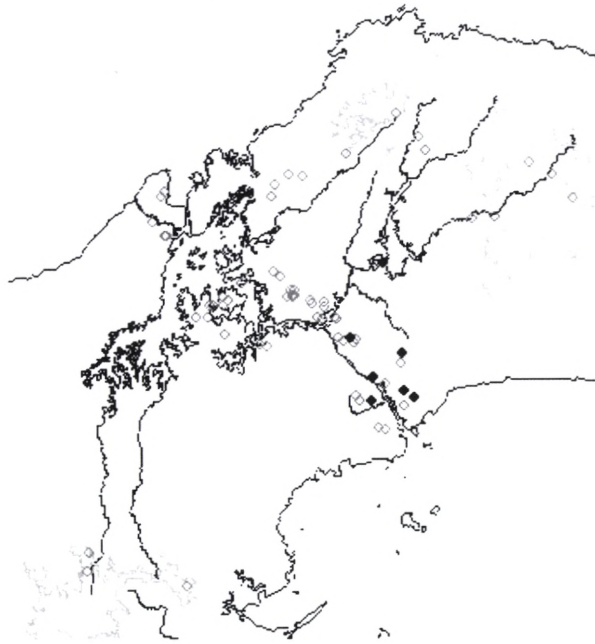
C) *Tetragastris panamensis*
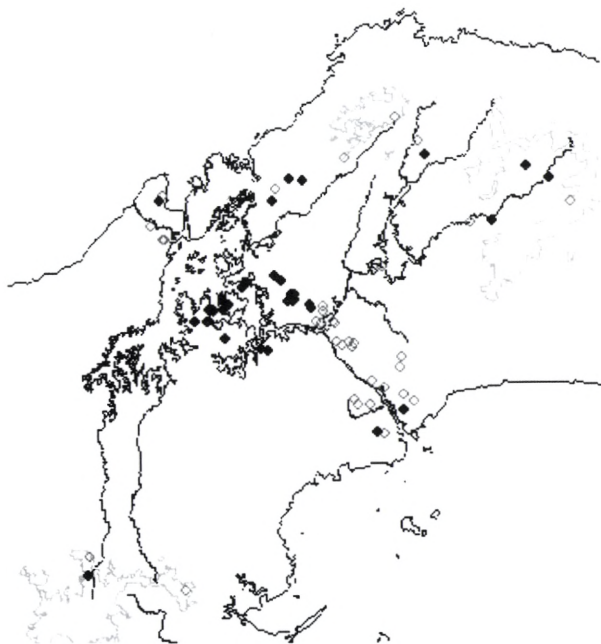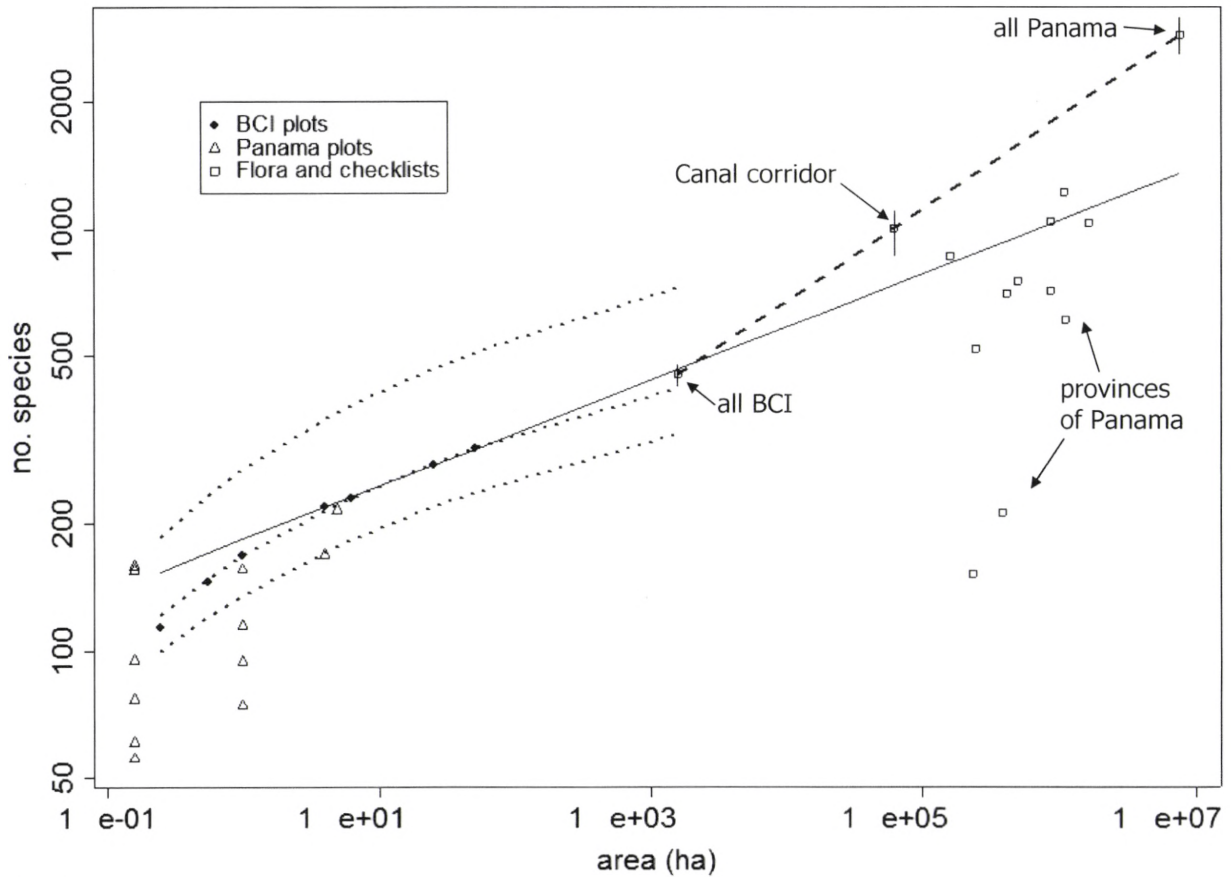
D) *Hybanthus prunifolius*
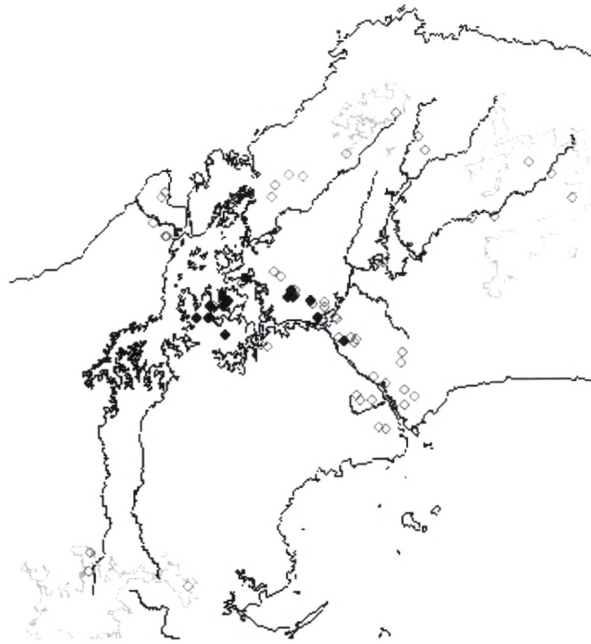
**Fig. 3**. Species-area relationship for trees and shrubs. The solid line gives the best fit power-function SAR for plots from 1 ha to 50 ha at Barro Colorado ($S=173\,area^{0.146}$). The dashed line is the fit to the three points indicated with arrows ($S=91\,area^{0.218}$). The dotted lines give predictions from Fisher's $\alpha$, with $\alpha = 26$, 35, and 60, representing mean values for a from the Pacific region, Barro Colorado, and eastern wet forests.

Pacific coast (using the mean density there of 2046 trees per ha $\geq 1$ cm dbh). The three plots have 191 species in total. We have carried out 15 additional inventories in the region, and the pooled species total is now 311 (Table 2), and it is not difficult to believe that further work would uncover another 75 species to reach the predicted total of 386. The same approach for the Sherman region ($\alpha = 35.6$, 3539 individuals per ha, 336 species observed) produces an estimate of 506 species in 15,000 ha (Table 2).

In the entire Canal corridor, the Panama checklist includes 863 species, and our plots have 73 more, leading to a total count of 946 tree and shrub species. Assuming additional species remain to be discovered, we estimate 1000 as the total, and this falls almost exactly on the power-function SAR between Barro Colorado and the entire nation of Panama: $S = 91 * A^{.218}$ (Fig. 3). On the other hand, with a mean Fisher's $\alpha$ per plot of 33.7 throughout the Canal corridor, extrapolation with $\alpha$ leads to an estimate of 532 species, well below the 650 we have recorded in plots and inventories

(Table 2, Fig. 3). At larger scale, where there is substantial climatic or soil heterogeneity, α substantially under-predicts species richness.

There are 107,000 ha of wet forests east of the Canal, covering a great deal of elevational heterogeneity. Based on how poorly Fisher's α extrapolated to the Canal corridor, we see no justification for its use at this larger scale, and we are left only with the SAR. Using $S = 91* A^{.218}$ (the dashed line in Fig. 3) leads to an estimated 1136 species in 107,000 ha; fitting a curve that includes the 50-ha plot, or the 50-ha plot plus additional 1-ha plots, changes this estimate to 1173 or 1180 species. But we do not place much confidence in these estimates. Other province counts fall far below the fitted SAR (Fig. 3), although this is probably due to lack of collecting. On the other hand, wet forests east of the Canal are locally very diverse. At Barro Colorado and Sherman, there were < 100 species per 0.16 ha plot, but Santa Rita and the Upper Chagres had > 150 species in similar plots (Table 3). We suggest that wet forests near the Canal have 1100-1300 species of trees and shrubs, but this is an hypothesis more than an estimate.

### Similarity of samples across the isthmus
Small plots in the Pacific and in Sherman regions shared few species. For instance, the mean 0.25-ha quadrat at the 4-ha plot at Cocoli

**Table 3.** Mean species richness in different regions of the Canal area from all plots in which every species was recorded; N is the number of plots.

| region | 40 x 40 m ≥ 1 cm dbh no. species | N | 1 ha ≥ 10 cm dbh no. species | N |
|---|---|---|---|---|
| Santa Rita | 174 | 2 | 162 | 2 |
| Upper Chagres | 155 | 4 | – | 0 |
| Sherman | 94 | 35 | 77 | 9 |
| Barro Colorado | 93 | 35 | 90 | 60 |
| Pipeline | 100 | 10 | 87 | 10 |
| Gamboa | 71 | 12 | 66 | 6 |
| Pacific | 57 | 25 | 57 | 6 |

(Pacific) shared 8.4 species with the mean 0.25-ha in the Sherman plot, out of 67 species in the Cocoli quadrat and 112 in the Sherman quadrat. The Sørensen similarity index from this sample would be 0.09. Increasing sub-quadrat size led to an increase in the Sørensen index that was linear with the number of species sampled at either site (Fig. 4). The linear increase is a prediction of the similarity theory, providing species abundances are uncorrelated at the two sites.

### Abundance correlation
The regression of the log-abundance in the Cocoli plot vs. log-abundance in the Sherman plot, for those 52 species found in both plots, had $r^2 = 0.01$. However, those species present at just one of the two sites indicated some correlation, although in a complicated sense. At Cocoli, mean abundance of all species was 2.3 ha$^{-1}$: 3.2 ha$^{-1}$ for those shared with Sherman and 2.0 ha$^{-1}$ for those unshared (based on log-transformation, averaging, then back-transformation). This difference was not statistically significant. For Sherman abundances, the pattern reversed: abundance of shared species was lower (1.9 ha$^{-1}$) than for unshared species (4.2 ha$^{-1}$), and this difference was significant. Despite this indication of correlated abundances, similarity rose in a linear fashion (Fig. 4). We believe it did so because of the reversal between sites – Sherman abundance was negatively associated with presence at Cocoli, but Cocoli abundance was positively associated with presence at Sherman. The conflicting patterns combine to produce an overall lack of correlation.

Comparing the Canal corridor and the wet forests east of the Canal, though, there was evidence for a positive correlation in abundance. For 194 species found in the small plots in both regions, the regression of log(abundance) between sites was significantly positive ($r^2 = 0.07$, $p < .001$). Moreover, the mean abundance
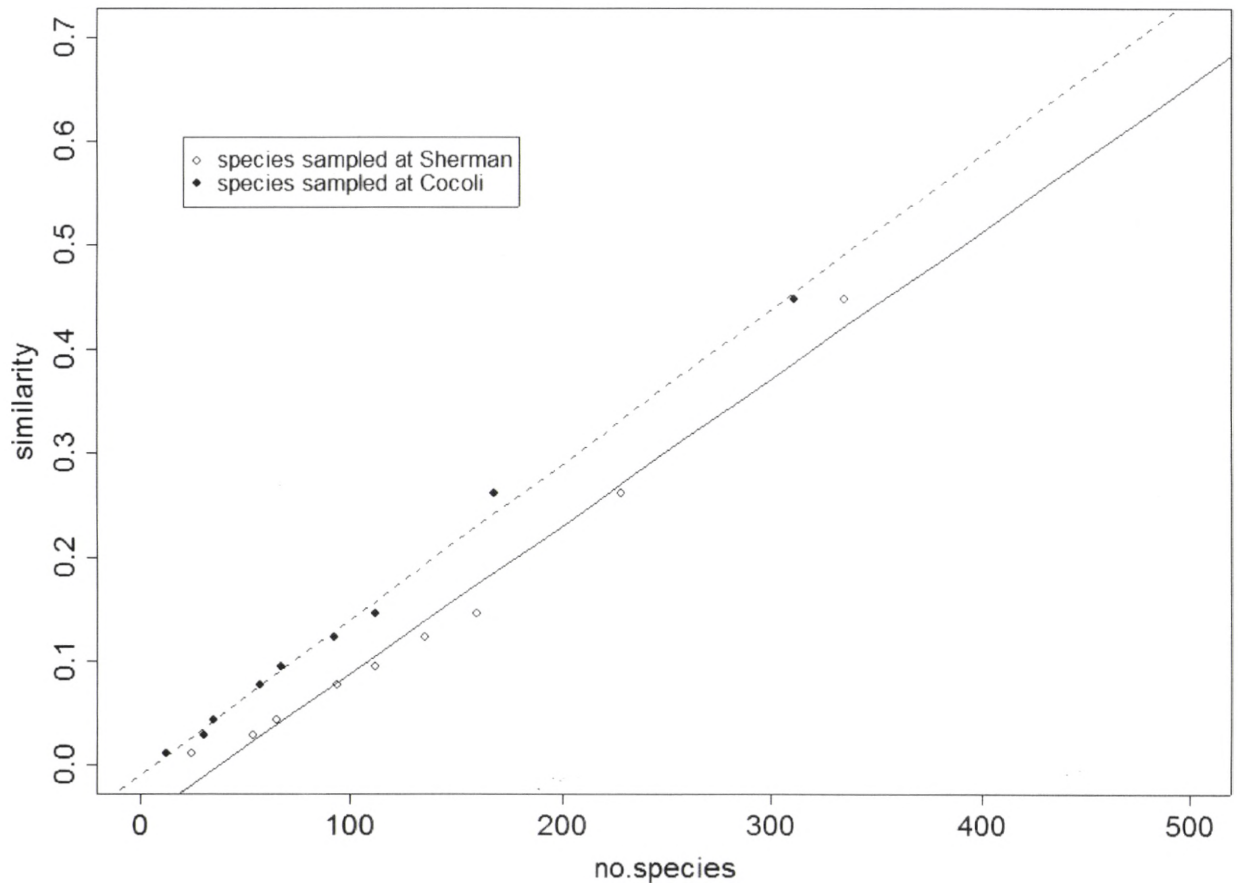
**Fig. 4**. Relationship between the Sørensen similarity index and sample size in quadrats from Sherman (Caribbean) and Cocoli (Pacific) plots. Solid circles show the Sørensen index graphed as a function of the number of species found in a sample at Sherman; the open circles show the index graphed as a function of species number at Cocoli. For example, the leftmost points give the index calculated between randomly drawn quadrats of 10 x 10 m in size, one in the Sherman plot, one in the Cocoli plot. The mean similarity index between all such pairs was 0.011, but the Sherman quadrats had a mean of 24 species and the Cocoli quadrats 12. Subsequent points show similarity for 20 x 20, 25 x 25, 50 x 50, 75 x 75, 100 x 100 m quadrats, the entire plots (4 ha at Cocoli, 5.96 at Sherman), and finally for all inventories in the two regions (see Fig. 1).

of those 194 species in the Canal corridor plots was 3.3 ha$^{-1}$, while the mean for 279 species in those same plots which did not occur in the wet forests was just 0.9 ha$^{-1}$, a statistically significant difference. For species occurring in wet forest plots, though, there was no significant difference in abundance between those shared with the Canal corridor and those not shared.

***Extrapolated similarity***

Assuming a lack of correlation in abundance, the theory presented above allows us to extrapolate the total number of species shared between Sherman and Pacific regions, given extrapolated species richness. In the two large plots, 52 species were shared, whereas 145 were shared in the complete set of inventories in

both zones. According to theory, the statistic $k$ should be sample-size independent: $k$ based on the two plots is

$$\frac{52}{169 \times 227} = 0.00136,$$

while based on all inventories,

$$\frac{145}{311 \times 336} = 0.00139,$$

It thus appears that the theory works quite well at these two sites. With it, we estimate that 271 species are shared between the two zones (Table 4). But there are two layers of uncertainty in this estimate, since it depends both on extrapolated species richness and extrapolated similarity; we do not pretend this is anything but a preliminary approximation. It is, however, the only estimate we have.

It seems certain that the total number of species shared between Pacific and Caribbean zones is higher than we would have concluded based on just the two moderately-sized plots. Contrary to our initial conclusions about how sharply the two forests differ in species composition, the fraction of species occurring in the Pacific zone that also occur near the Caribbean is 70% (271 of 386). This conclusion does not change the fact, though, that abundant species at the two sites are very different – not a single one of the top-ten ranking species in basal area at Sherman rank among the top-ten at Cocoli (Condit *et al.* 2004). A substantial number of species that are common in one region must occur rarely in the other, perhaps in locally moist (or dry) microsites.

A similar exercise can be applied to a comparison of the entire Canal corridor and the eastern wet forest. The observed number of species shared was 290, out of 650 in Canal corridor inventories and 694 in the wet forest. Assuming no correlation in abundance, we would extrapolate the true number shared to be 707-836, out of 1000 species in the corridor and 1100-1300 in the wet forest. However, there was a positive correlation in abundance between the two zones, and this invalidates the extrapolation. The positive correlation suggests that those species unobserved in either zone would tend to be absent in the second zone (that is, relative to those sampled).

If so, the estimate of 700-800 species shared is high. On the other hand, we are certain that the true number shared is higher than 290, the observed number shared. It would be useful to evaluate the severity of the overestimate when the assumption of no correlation in abundance is violated, but we have not tried to do so yet. For the sake of drawing a conclusion, we consider the lower figure of 700 (Table 4) as our best guess for the number of species shared. If so, there would be 1400-1600 in the two zones combined (Table 4). Based on the dashed SAR (Fig. 3), we would estimate 1281 species among the two regions (169,000 ha), and this varies little whether we add BCI 1- and 50-ha samples

**Table 4.** Estimates of total species shared between Sherman and Pacific regions as calculated from similarity theory. Area was estimated as the amount of forested area (Ibáñez *et al.* 2002) within the regions marked on Fig. 1. See text for explanation of estimates.

| region (area) | observed spp $S_i$ | observed shared $S_c$ | estimated spp $T_i$ | estimated shared $S_c$ |
|---|---|---|---|---|
| Sherman (150 km²) | 336 | 145 | 506 | 271 |
| Pacific (112 km²) | 311 | 145 | 386 | 271 |
| Total | 502 | – | 621 | – |

to the SAR. This means 800 must be shared, not 700, and suggests that the extrapolation based on similarity theory was close to the mark. Clearly, though, these estimates are fairly uncertain; recall the important caveat that applying the similarity theory requires that we make assumptions about unobserved (and thus rare) species based on the observed (common) species.

## Conclusions

Here is what we would like to conclude: that there are 1300-1400 tree species in central Panama, that 500-600 of these are restricted to wet forests of the Caribbean ridges and other mountain tops, 200 are restricted to the drier Canal corridor and Pacific slope, and 700-800 are widespread. The broad pattern – that there are more species restricted to the wet zone than the dry zone – seems defensible, but the actual species counts are hypothetical.

Scanty inventories and incomplete sampling should generally produce substantial underestimates of similarity between different forests, and we believe there has been a tendency to underestimate the distribution of tropical trees as a result. Existing methods for estimating species richness in poorly known areas, or the number of species with restricted ranges, are crude (Pitman & Jørgensen 2002). We agree with Pitman *et al.* (1999, 2001) that a large number of species are widespread, but we cannot properly judge just how many until there are better methods for extrapolating from small samples. We have tried to present a framework for doing so, and it is somewhat encouraging that in some circumstances, our theory accurately predicted changes in similarity. Whether the quantitative predictions of the theory will bear out in other regions, and whether its extrapolations are accurate, remain to be seen.

We close by encouraging thorough invento-

ries of tree species at more sites. Crucial to extrapolations are benchmarks where all species present are known with reasonable confidence, such as Croat's (1978) treatment of Barro Colorado Island, the 50-ha plot on the island, and M. Correa's growing and well-maintained checklist of the flora of Panama. More complete checklists of these sorts will be needed to advance theories on extrapolating species richness and ranges.

## Acknowledgments

## Literature cited

Cao, Y., Williams, W.P. & Bark, A.W. 1997. Effects of sample size (replicate number) on similarity measures in river benthic Aufwuchs community analysis. *Water Environment Research* **69**: 107-114.
Condit, R. 1998. *Tropical Forest Census Plots.* Springer-Verlag

and R. G. Landes Company, Berlin, Germany and Georgetown, Texas.

Condit, R., Aguilar, S., Hernández, A., Pérez, R., Angehr, G., Lao, S., Hubbell, S.P. & Foster, R.B. 2004. Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. *J. Trop. Ecol.* **20**: 1-22.

Condit, R., Foster, R.B., Hubbell, S.P., Sukumar, R., Leigh, E.G., Manokaran, N. & Lao, S. 1998. Assessing forest diversity on small plots: calibration using species-individual curves from 50 ha plots. *In*: Dallmeier, F. & Comiskey, J.A. (eds.), *Forest Biodiversity Diversity Research, Monitoring, and Modeling.* UNESCO, the Parthenon Publishing Group, Paris. Pp. 247-268.

Condit, R., Hubbell, S.P. & Foster, R.B. 1995. Mortality rates of 205 neotropical tree and shrub species and the impact of a severe drought. *Ecol. Monogr.* **65**: 419-439.

Condit, R., Hubbell, S.P. & Foster, R.B. 1996. Changes in a tropical forest with a shifting climate: results from a 50 ha permanent census plot in Panama. *J. Trop. Ecol.* **12**: 231-256.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Núñez, P.V., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H., Losos, E. & Hubbell, S.P. 2002. Beta-diversity in tropical forest trees. *Science* **295**: 666-669.

Condit, R., Robinson, W.D., Ibáñez, R., Aguilar, S., Sanjur, A., Martínez, R., Stallard, R., García, T., Angehr, G., Petit, L., Wright, S.J., Robinson, T.R. & Heckadon, S. 2001. Maintaining the canal while conserving biodiversity around it: a challenge for economic development in Panama in the 21st century. *Bioscience* **51**: 135-144.

Condit, R., Watts, K., Bohlman, S.A., Pérez, R., Hubbell, S.P. & Foster, R.B. 2000. Quantifying the deciduousness of tropical forest canopies under varying climates. *J. Veg. Sci.* **11**: 649-658.

Croat, T.R. 1978. *Flora of Barro Colorado Island.* Stanford University Press, Stanford, California.

D'Arcy, W.G. 1987. *Flora of Panama. Part I. Introduction and Checklist.* Missouri Botanical Garden, St. Louis, Missouri.

Fisher, R.A., Corbet, A.S. & Williams, C.B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12**: 42-58.

Hubbell, S.P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton University Press, Princeton, NJ.

Ibáñez, R., Condit, R., Angehr, G., Aguilar, S., Garcia, T., Martínez, R., Sanjur, A., Stallard, R., Wright, S.J., Rand, A.S. & Heckadon, S. 2002. An ecosystem report on the Panama Canal: monitoring the status of the forest communities and the watershed. *Environmental Monitoring and Assessment* **80**: 65-95.

Pitman, N.C.A. & Jørgensen, P.M. 2002. Estimating the size of the world's threatened flora. *Science* **298**: 989.

Pitman, N.C.A., Terborgh, J., Silman, M.R. & Núñez V., P. 1999. Tree species distributions in an upper Amazonian forest. *Ecology* **80**: 2651-2661.

Pitman, N.C.A., Terborgh, J., Silman, M.R., Núñez V., P., Neill, D.A., Palacios, W.A., Aulestia, M. & Céron, C.E. 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. *Ecology* **82**: 2101-2117.

Pyke, C.R., Condit, R., Aguilar, S. & Lao, S. 2001. Floristic composition across a climatic gradient in a neotropical lowland forest. *J. Veg. Sci.* **12**: 553-566.

Rosenzweig, M.L. 1995. *Species Diversity in Space and Time.* Cambridge University Press.

Stewart, R.H., Stewart, J.L. & Woodring, W.P. 1980. *Geologic map of the Panama Canal and vicinity.* United States Geological Survey, Reston, Virginia, USA.