

# Co-occurrence based assessment of habitat generalists and specialists: a new approach for the measurement of niche width

JASON D. FRIDLEY\*§, DAVID B. VANDERMAST\*¶,

DANE M. KUPPINGER†, MICHAEL MANTHEY‡ and ROBERT K. PEET†‡

\*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599–3280, USA,

†Curriculum in Ecology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3275, USA,

‡Institute of Botany and Landscape Ecology, Ernst-Moritz-Arndt University, 17487 Greifswald, Germany

## Summary

**1** Much ecological theory is based on the characterization of ecological habits of species as ‘generalist’ or ‘specialist’, but standard measures for placing species along a generalist-specialist gradient do not exist.

**2** We introduce a method for quantifying habitat specialization (i.e. relative niche widths) using species co-occurrence data. Generalists should co-occur with many species, whereas specialists should co-occur with relatively few species, given equal plot occurrences. We quantify this concept using a generalist-specialist metric ( $\theta$ ) derived from a beta diversity statistic.

**3** We evaluate the ability of our generalist-specialist metric to correctly rank species according to simulated (known) niche widths. Our technique is generally robust to a wide variety of niche distribution structures and sampling designs, but surveys strongly biased toward certain habitats can undermine the ability of  $\theta$  to accurately describe niche widths for underrepresented species.

**4** We apply our technique to three spatially nested surveys of the large woody flora (> 1 cm d.b.h.) of the south-eastern USA. For each dataset we rank the generalist-specialist tendencies of all species of non-trivial occurrences, including 113 species across the Southeast, 71 species of southern Appalachian forests, and 44 species of the 6800-ha Joyce Kilmer-Slickrock Wilderness Area (NC and TN, USA).

**5** Rankings of species’  $\theta$ -values were generally consistent among datasets of different spatial extent. Generalist species (e.g. *Ilex opaca*, *Ulmus rubra*, *Morus rubra*, *Prunus serotina*, *Acer rubrum*) were often those with large geographical ranges, particularly for  $\theta$ -values from the largest dataset, and overall were more likely to be bird-dispersed, deciduous, and shade tolerant. South-eastern specialist species (e.g. *Taxodium* spp., *Abies fraseri*, *Quercus laevis*, *Pinus pungens*, *Pinus palustris*) were those associated with stressful or unusual conditions, such as a long duration of flooding, high fire frequency, or extreme cold or dry climates.

**6** Our study demonstrates that increasingly available, large-survey datasets can contribute niche-related species information in the absence of detailed environmental or habitat measurements. Applications include new assessments of relationships between species traits, ecological and environmental tolerances, and species packing in different assemblages.

*Key-words:* niche breadth, habitat selection, habitat diversity, south-east U.S.A., woody flora, southern Appalachians, Joyce Kilmer-Slickrock Wilderness, beta diversity, additive partitioning, ecoinformatics

*Journal of Ecology* (2007)

doi: 10.1111/j.1365-2745.2007.01236.x

## Introduction

A well known but poorly understood phenomenon in community ecology is that some species have wider niches than others. That some species occur in many habitats, and others in few, has important implications for our understanding of community-level processes and for species conservation issues. Despite this, the estimation of relative niche sizes has been a persistent problem in the long history of quantifying species niches in ecology (e.g. Hutchinson 1957; Horn 1966; Wuenscher 1969). The classic approach of niche delimitation for plants is to measure species response or abundance along one or a few environmental gradients (Whittaker 1956; Ellenberg 1974; Austin & Gaywood 1994; Bazzaz 1996), and interpret these response patterns using metrics of niche breadth (Levins 1968; Colwell & Futuyma 1971; Bazzaz 1991). However, for the vast majority of species the important environmental gradients that determine occurrence are not known, much less used to determine relative niche sizes. As a result, attempts at defining habitat specializations have been few and problematic (Burgman 1989; Witkowski & Lamont 1997). Because habitats are not discrete, a habitat-based estimate of species generalism is inaccurate and potentially misleading (Bazzaz 1991; cf. Thompson *et al.* 1998). A measure of niche breadth along one or a few gradients is also misleading because species that are generalists on one axis may be specialists on another (Crawley 1997).

To achieve a classification of species along a generalist–specialist gradient, an approach is needed that allows for robust estimation of whether a species occurs in many or few ‘habitats’, without the need for defining habitat or attempting to measure the N-dimensional hypervolume for each species. One way to do this is to let patterns of species co-occurrences themselves define habitat diversity. All else being equal, habitat generalists co-occur with many species across their range, while habitat specialists co-occur with relatively few species. Co-occurrence data thus offer an approach that is in effect a biological assay for ‘habitat diversity’ or ‘niche width’ that requires no assumptions about the definition of a habitat or the most critical environmental factors that control plant species distributions. This effectively avoids the need to measure the unknown (and potentially unknowable) factors that determine plant niches, and instead lets co-occurrence patterns alone suggest the width of plant niches.

In this paper we introduce a method for rank-ordering species along a gradient of niche width – or what we will refer to as habitat generalism vs. specialism – that only requires information on the relative degree to which focal species co-occur with other species within a given area. Our method takes into account several factors that could obscure the relationship of niche width and total co-occurrences, such as variation in local (alpha) diversity among habitats and survey designs that are biased toward collection of certain habitats or species.

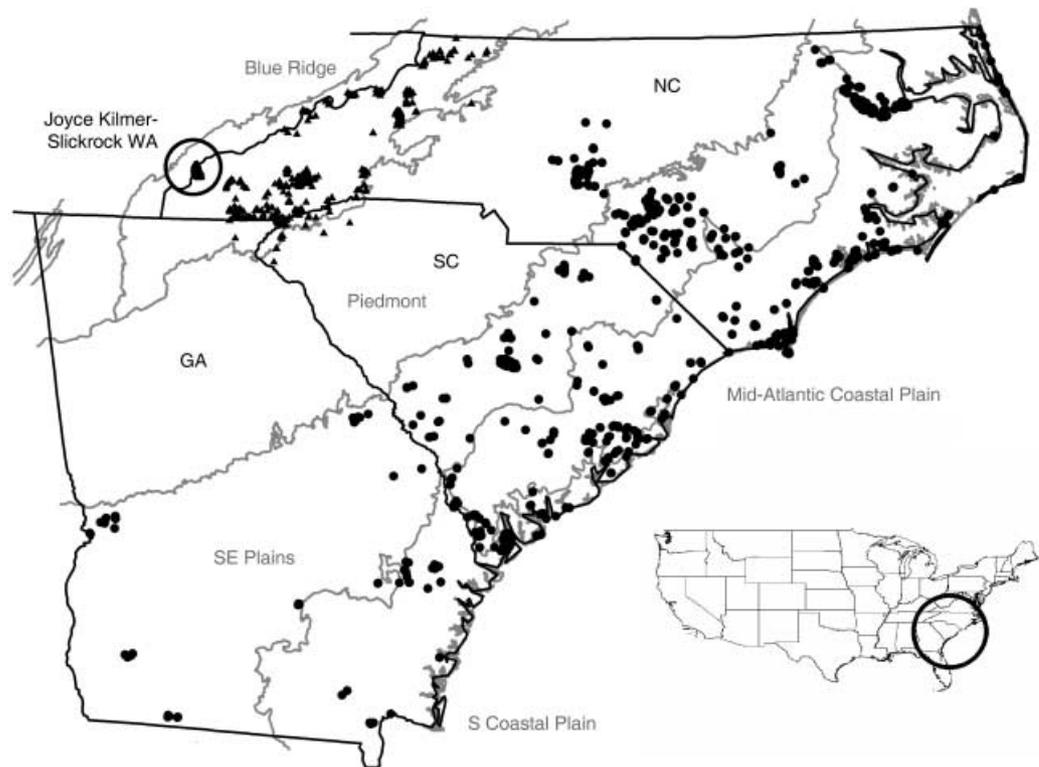
We demonstrate the effectiveness of our approach at removing these possible artefacts through simulation. We then use our method to quantify habitat generalism and specialism for large woody plant species of the south-eastern USA, using the three following nested spatial extents of a large vegetation data base (Fig. 1): (i) the Carolinas and parts of Georgia and Tennessee (NC, SC, GA and TN) including montane, piedmont, and coastal habitats, (ii) Southern Appalachian montane forests, and (iii) the Joyce Kilmer-Slickrock Wilderness area (NC) near the southern terminus of the Appalachians. Finally, we examine whether our measures of niche width are correlated with other species characteristics, including some simple life-history attributes (growth and reproductive traits) and selected environmental properties of species’ ranges.

## Methods

### θ: A GENERALIST-SPECIALIST METRIC BASED ON CO-OCCURRENCES

We assume that species that occur in many different habitats – generalists – will have a relatively high rate of species turnover among plots in which they occur. Specialist species, regardless of their frequency in the data set, will exhibit relatively low species turnover in the plots in which they occur because they consistently occur with the same species. Species turnover among plots, or differentiation diversity, was defined as  $\beta$  diversity by Whittaker (1960) and has been estimated using myriad metrics and statistical techniques (Koleff *et al.* 2003), which are profitably grouped into gradient and non-gradient approaches (Vellend 2001). For a co-occurrence measure free of assumptions about environment and distance we require the equivalent of a non-gradient index of beta diversity. We considered multiple versions of non-gradient  $\beta$  diversity (Koleff *et al.* 2003) and selected the technique of ‘additive partitioning’ of diversity components (Lande 1996; Loreau 2000; Veech *et al.* 2002),  $\beta = \gamma - \mu(\alpha)$ , where  $\gamma$  is the cumulative number of species among plots and  $\mu(\alpha)$  is mean plot species richness. We chose this form of  $\beta$  diversity because it controls for different levels of associated local (alpha) richness among species, which would spuriously inflate estimates of habitat generalism. For example, a species may co-occur with a relatively large number of species across its range, but consistently occur in the same species-rich community. Subtracting mean plot richness from total number of co-occurrences removes the effect of variation in habitat richness, which we demonstrate further below.

A second critical issue is that species differ widely in frequency of occurrence for any given dataset. Species frequently represented in a plot database will tend to co-occur with relatively many species (i.e. have a high associated  $\gamma$  diversity) regardless of the degree of species turnover among plots in which they occur. The representation of any one species in a plot database is the



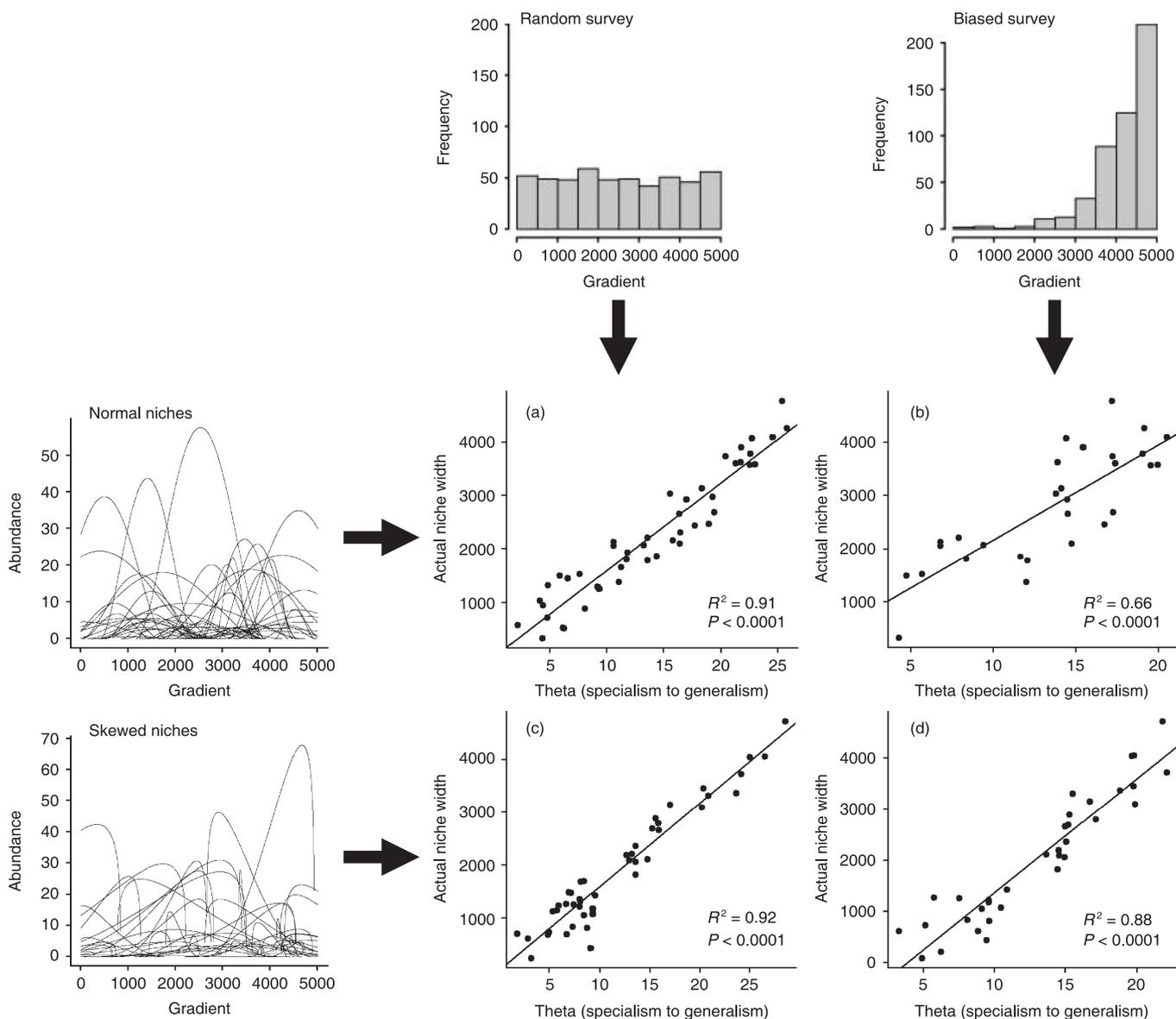
**Fig. 1** Location of 2480 vegetation plots of the Carolina Vegetation Survey (NC, SC and GA, USA) within the context of Level III Ecoregions (US EPA 2002). Montane upland forest plots are indicated with triangles, which include plots in the Joyce Kilmer-Slickrock Wilderness Area (circle).

result of both survey site selection and the species' habitat specialization; it is thus crucial that the effects of site selection be removed before calculating species turnover (see Burgman 1989). To remove the signature of overall abundance in the dataset we used a randomization technique whereby we randomly chose a fixed number of plots containing a focal species (50 plots for the largest dataset, 20 for the two smaller datasets) before calculating  $\beta$ , thereby keeping total plot frequency constant between species. For each species we applied this randomization 100 times and took the average  $\beta$ -value. To avoid confusion with various measures of community-level turnover, we term our metric for the generalist–specialist tendency of a species  $\theta$ . In this way, a species like *Acer rubrum*, which occurred in a large percentage of all plots, had the same plot representation as a less common species like *Tsuga caroliniana*. On the other hand, *A. rubrum* occurs in many plots partly because it occurs in many habitats; this aspect is retained in our calculation because a random selection of 50 plots containing *A. rubrum* will have considerable species turnover due to the large geographical and ecological separation of those plots. In contrast, the southern Appalachian endemic *T. caroliniana*, although also represented by 50 plots, will co-occur with many fewer species due to its small range and site restrictions. In practice, the choice of the size of random plot selection depends on the relative frequency of species in the dataset and the compositional variance of plots containing a given species; small sample sizes for a very common

species increase the likelihood of a non-typical subset of species co-occurrences, and thus contribute to large variance in the estimation of  $\theta$ . However, large sample sizes can exclude all but the most common species from the analysis, depending on the distribution of species frequencies in the dataset. Our choice of 50 plots for the largest dataset and 20 for the two smaller datasets represents a compromise between a higher number, which would reduce variance due to sample size effects, and a lower number, which allows inclusion of all but the rarest species. The specific algorithm for calculation of  $\theta$  is described in Appendix S1 (see Supplementary material) and our implementation in R 2.2.0 (R Development Core Team 2005) is provided as Appendix S2.

#### SIMULATION: TESTING THE EFFECTIVENESS OF $\theta$

Our proposed metric is at best a well-informed hypothesis of how best to characterize niche width using species occurrence data. To verify the efficacy and robustness of our technique, we ran simulations of randomly parameterized species distributions of known niche width along a single 'environmental' gradient (Fig. 2, left panels). Our approach was to calculate  $\theta$  using our simulated data and assess the ability of  $\theta$  to predict the known (one-dimensional) niche widths of simulated species, particularly in response to different scenarios of niche shape and survey designs that could potentially introduce error into the relationship of  $\theta$  and niche



**Fig. 2** Four simulation scenarios of different niche shapes along a single gradient (normal (symmetrical) vs. skewed) and different survey designs (random locations along the gradient vs. biased towards one end), and the corresponding relationship between a co-occurrence-based estimate of niche width ( $\theta$ ) and known niche widths of simulated species. (a) Species abundances are distributed normally along the gradient and surveyed with 500 gradient locations randomly chosen. (b) Species abundances as in (a) but surveyed in a biased manner whereby sampling intensity increases along the gradient. (c) Species abundances are skewed along the gradient (in either direction) but are surveyed randomly. (d) Species abundances as in (c) but surveyed in the biased design of (b). Each simulation is of 50 species. Niches were modelled using the Beta function with parameters described in Table 1. Regressions results describe the specific iteration displayed on the left.

width. Specifically, we focused on (i) whether species abundances are distributed normally along the gradient (with a central peak and symmetric decrease in abundance on either side) or skewed in either direction (such that peaks are located near a range limit), and (ii) whether allocation of plots in a survey dataset is randomly distributed along the gradient or is strongly biased to one end. Both issues relate to the ability of a survey to adequately capture the entirety of a species' niche, which in turn has obvious ramifications for the ability of an analysis of occurrences to effectively determine niche width. Strongly skewed niche distributions decrease the ability of a survey to capture large portions of the niche where a species is of low abundance;

similarly, highly biased survey designs describing only a limited range of available habitats will underestimate the occurrence of species inhabiting poorly surveyed habitats.

For each simulation we created niches of 50 species using the Beta function (Minchin 1987) with random parameterizations of niche width, amplitude and location of optimum. Parameters are described in Table 1. Normally distributed niches were created by fixing the shape parameters  $a$  and  $b$  at 1.99, while skewed niches were created by fixing either  $a$  or  $b$  at 1.99 and the other at 0.25. For each simulation of 50 niches along the gradient, we created a dataset of species occurrences in 500 'survey plots' by assigning a plot to a gradient location

**Table 1** Parameter values for niche simulations using the Beta function (after Minchin 1987)

Parameter	Description	Value
S	Total species along gradient	50
$A_0$	Niche amplitude	Lognormal RV of mean = 2 and S D = 1 (log scale)
m	Location of niche optimum	Uniform RV of gradient range 0–5000
r	Niche width	Uniform RV between 10 and 5000
a	Niche shape, left side	1.99 or 0.25
b	Niche shape, right side	1.99 or 0.25

(between 0 and 5000) and randomly selecting  $n$  individuals, where probability of species selection was weighted by the abundance of species at that location. Location assignment to plots was determined by either a uniform random variable of the range of the gradient (each site could be surveyed with equal probability; Fig. 2, top left panel) or in a biased manner where the probability of assignment increased exponentially along the gradient (Fig. 2, top right panel). In this way we created four scenarios corresponding to either normal or skewed niche distributions, and random or biased survey allocation (Fig. 2). For each of these scenarios we ran 100 simulations and for each simulation we evaluated the ability of  $\theta$  to predict the 50 known niche widths using the Pearson correlation coefficient ( $r$ ) with a threshold of 20 occurrences (under which a species was not included in the  $\theta$  calculation). R code used for these simulations is available as Appendix S3.

#### STUDY AREAS AND VEGETATION SURVEY DATA

We calculated  $\theta$  from field survey data using three nested spatial extents of vegetation survey plots from the Carolina Vegetation Survey (archived by the North Carolina Botanical Garden, Chapel Hill, NC, USA). Begun in 1988, the Carolina Vegetation Survey (CVS) archive includes vegetation plot data for the purposes of vascular plant inventory and classification of plant communities of the Southeastern USA. From this archive, we selected 2480 plots, most (see below) 0.1 ha (20 × 50 m) in size, containing a full inventory of individual tree composition and stem size data. Richness values of plots less than 0.1 ha were corrected using species–area transformations (see below). This dataset of 2480 plots covers the full spatial extent of the Carolinas and parts of Georgia and Tennessee, from coastal dune vegetation to the mountains of the southern Appalachians (Fig. 1), sampled between 1988 and 2000. The plots were generally collected as regional projects, with particular concentrations in montane forests of the southern Blue Ridge Mountains (*c.* 550 plots) and pine savannas of the Southeastern Coastal Plain (*c.* 450 plots). Within a project the landscape was typically subdivided by dominant vegetation types and plots were evenly distributed across types. In general, more undisturbed or mature vegetation is better represented in the data than recently disturbed areas. All plots were surveyed with the CVS protocol

described in Peet *et al.* (1998). Vascular plant nomenclature was standardized to follow Kartesz (1999), with the exception that we follow Weakley's (2005) recommendation of the specific epithet *Quercus montana* over its synonym *Q. prinus*.

For these 2480 vegetation plots, we chose to limit the species pool of our analysis to mid- to large-sized woody species, for the purposes of: (i) computational feasibility (reduction of several thousand taxa to less than 200); (ii) providing a more coherent assemblage of species that generally exist within the same vertical strata and functional group, and thus are expected to compete for the same resources; and (iii) using a group of taxa generally well distributed geographically within our study area, but representing a wide variety of perceived habitat affinities. Our total species pool, thus constrained, consisted of 179 species that could attain at least 1 cm diameter at breast height (d.b.h.), as evinced by at least one such occurrence in the CVS dataset. A species was considered present in a plot if it existed as a stem > 0 cm d.b.h.; thus, we did not include presences of seedlings or small saplings.

We selected two nested subregions of the CVS dataset to examine the sensitivity of our niche width metric to spatial scale. A dataset of 579 0.1-ha plots of montane upland forests (MUF) surveyed within the Blue Ridge province of the Southern Appalachians (Fig. 1) included a total of 105 woody species for use in our analysis and a maximum linear extent of *c.* 300 km. MUF plots include forested (closed-canopy) communities of many types, including spruce–fir, northern hardwoods, xeric and mesic deciduous and coniferous forests, etc., across an elevation range of 245–1902 m. Our dataset of smallest spatial extent was composed of 133 0.1-ha plots from Joyce Kilmer-Slickrock Wilderness Area (hereafter 'Slickrock') nested within the MUF dataset, including 72 woody species for analysis distributed over 6800 ha and an elevation range of 333–1618 m (Newell *et al.* 1997). Slickrock contains a wide variety of montane forest communities and those above *c.* 1030 m elevation are largely old-growth (Newell *et al.* 1997). Species pools of the MUF and Slickrock datasets are subsets of the 179-species pool of the larger CVS dataset, and the same criterion for species presence applies (stem > 0 cm d.b.h.).

All plots of the MUF and Slickrock datasets were 0.1 ha in size. However, 964 of the 2480 plots in the largest CVS dataset were less than 0.1 ha, some as small as 0.01 ha, for the surveying of communities that

**Table 2** Relationships between  $\theta$ -values for 113 woody species of the Carolina Vegetation Survey dataset and 27 other species characteristics, relating to life-history attributes and environmental properties of their distribution. Variables are described in text in more detail.  $n$  is the number of observations for each correlation,  $r$  is the correlation coefficient between the variable and  $\theta$ ,  $P$  is the unadjusted (single-test)  $P$ -value, and Adj.  $P$  is the  $P$ -value after adjustment for multiple testing, after Hochberg (1988). Bold indicates  $P < 0.05$

Variable	Source	$n$	$r$	$P$	Adj. $P$
Growing degree days range	Thompson <i>et al.</i> (1999)	70	0.68	<b>&lt;0.001</b>	<b>&lt;0.001</b>
Annual temperature range	Thompson <i>et al.</i> (1999)	70	0.55	<b>&lt;0.001</b>	<b>&lt;0.001</b>
Leaf duration	Burns & Honkala (1990)	109	–	<b>0.003</b>	0.081
Moisture index	Thompson <i>et al.</i> (1999)	70	–0.31	<b>0.009</b>	0.256
Dispersed by birds? (yes/no)	Burns & Honkala (1990)	102	–	<b>0.013</b>	0.344
Life span, maximum	Burns & Honkala (1990)	46	–0.32	<b>0.031</b>	0.783
Dispersal distance, maximum	Burns & Honkala (1990)	50	–0.30	<b>0.036</b>	0.867
Life span, typical	Burns & Honkala (1990)	54	–0.28	<b>0.041</b>	0.915
Height, maximum	Burns & Honkala (1990)	57	–0.27	<b>0.046</b>	0.915
Shade tolerance (1–3)	USDA & NRCS (2004)	74	–	0.060	0.915
Growth rate, mean	Prasad & Iverson (1999)	57	–0.23	0.089	0.915
Age at first seed production	Burns & Honkala (1990)	54	–0.19	0.159	0.915
Age at optimum seed production	Burns & Honkala (1990)	47	–0.19	0.196	0.915
pH, mean occurrence	Prasad & Iverson (1999)	58	0.16	0.240	0.915
Growth rate, maximum	Prasad & Iverson (1999)	57	–0.14	0.298	0.915
Adaptation to fire (1–4)	Prasad & Iverson (1999)	58	–	0.351	0.915
Age at decline of seed production	Burns & Honkala (1990)	45	–0.14	0.375	0.915
Canopy tree? (yes/no)	Burns & Honkala (1990)	95	–	0.394	0.915
Moisture index range	Thompson <i>et al.</i> (1999)	70	0.10	0.407	0.915
Seed mass	USDA & NRCS (2004)	74	–0.08	0.498	0.915
Flooding tolerance (1–3)	Prasad & Iverson (1999)	58	–	0.525	0.915
January temperature, median	Thompson <i>et al.</i> (1999)	70	–0.07	0.592	0.915
Temperature of coldest month, median	Thompson <i>et al.</i> (1999)	70	–0.06	0.594	0.915
Growing degree days, median	Thompson <i>et al.</i> (1999)	70	–0.06	0.638	0.915
Fire tolerance (1–3)	Prasad & Iverson (1999)	58	–	0.672	0.915
Potential evapotranspiration, mean	Prasad & Iverson (1999)	58	0.05	0.713	0.915
Annual temperature, median	Thompson <i>et al.</i> (1999)	70	0.01	0.915	0.915

were perceived to be smaller than 0.1 ha (e.g. rock outcrops, small wetlands and dune vegetation). Because the diversity of habitats is greatly enhanced by retaining these plots in our analysis, but local richness values are confounded by differences in plot sizes, we retained these plots but used corrected plot richness values based on observed species–area relationships in the other 1516 0.1-ha plots. Most plots in the CVS archive contain nested species richness data for  $10 \times 10$  m subquadrats (see Fridley *et al.* 2005), allowing estimation of 0.1-ha species richness based on extrapolation from values of smaller quadrat sizes. Using nonlinear curve fitting of the Arrhenius species–area equation ( $S = cA^Z$ ; Fridley *et al.* 2005) to species–area data from the other 1516 plots, we estimated 0.1 ha richness values for the remaining 964 plots. Although we use these corrected values of plot richness in our analysis, there was little influence of this procedure on our results, as (i) even correcting 0.01–0.1 ha plot sizes changed local richness by only about 10 species (corresponding to an estimated  $Z$ -value of 0.35), and (ii) as detailed below, differences in mean local richness among species had little influence on  $\theta$ -values.

#### SPECIES LIFE-HISTORY DATA

We compiled ecological and life-history attributes of species from several sources, including the USDA

PLANTS database (USDA & NRCS 2004), the USFS Silvics Manual (Burns & Honkala 1990), the *USGS Climate–Vegetation Atlas of North America* (Thompson *et al.* 1999), and the *USFS Climate Change Tree Atlas* (Prasad & Iverson 1999). The 27 life-history and environmental variables chosen for our analysis and their sources are listed in Table 2. We focused on attributes that could be obtained for at least a third of our 110 species, particularly regarding dispersal traits (mode, seed size and age of reproduction), growth characteristics (growth rate, mature height, lifespan and shade tolerance), leaf duration (deciduous/evergreen), tolerances to stress and disturbance (fire, flooding and low temperatures), and summaries of climate distribution. Environmental variables included median values and 10th–90th quantile ranges of distributions of climate values at a 25-km<sup>2</sup> grid delimited by the complete North American range of 70 of our selected species, as reported by Thompson *et al.* (1999), as well as summarized range values for potential evapotranspiration and pH compiled by Prasad & Iverson (1999). Prasad & Iverson (1999) also categorized species according to tolerances of flooding and fire, and used USDA-USFS Forest Inventory and Analysis (FIA) data to calculate species-specific growth rates. Most data on life-history attributes were obtained from Burns & Honkala (1990), except for seed mass and a simple index of shade tolerance, both obtained from USDA PLANTS (USDA & NRCS 2004).

We tested for significant correlations between calculated  $\theta$ -values from the full extent CVS database and the 27 variables listed in Table 2. Values of most variables were not available for all 113 selected species (Table 2). We report  $P$ -values from independent tests of linear regressions and adjusted  $P$ -values for conducting multiple significance tests using Hochberg's (1988) modification of the Bonferroni technique, as summarized in Legendre & Legendre (1998). We did not attempt similar correlations for the smaller two datasets due to restrictions on sample size and the overall similarity of  $\theta$  rankings for species among the three datasets.

## Results

### ROBUSTNESS OF $\theta$ IN SIMULATIONS

While good estimates of niche widths were provided by  $\theta$  regardless of niche shape or survey design scenario,  $\theta$  was a better predictor of niche width under some circumstances (Fig. 2). Niche widths of species with skewed niche distributions were predicted as well as those with normal niches (Fig. 2 (a) and (c); mean correlation ( $r$ ) of  $\theta$  and niche width for 100 simulations was 0.925 and 0.949 for normal and skewed niches, respectively). Biased surveys reduced the ability of  $\theta$  to accurately predict niche widths but correlations remained high (mean  $r$  of 0.875 and 0.901 for normal and skewed niches). This is in part because a biased survey produces fewer species able to meet minimum occurrence requirements to enter the  $\theta$  calculation; those species whose ranges would be most poorly represented are not evaluated by  $\theta$ . In addition, for a highly biased survey design, niche widths were on average better predicted by the total number of species co-occurrences of each species (mean  $r$  of 0.884 and 0.922 for normal and skewed niches) than  $\theta$ . This occurs because, for species with relatively large niche width but which are more concentrated in those habitats poorly represented in the data, a random selection of plots is biased away from finding the bulk of their co-occurrences, and thus they are over-penalized by  $\theta$ . This should be especially true for species that are just above the occurrence threshold.

### PATTERNS OF HABITAT SPECIALIZATION AMONG 113 WOODY SPECIES OF THE CAROLINA VEGETATION SURVEY

Values of  $\theta$  for the 113 species that met the 50-plot occurrence cut-off are listed in Table 3, from the most generalist species (*Ilex opaca* = 106) to the most specialist (*Taxodium ascendens* = 24). Given a random set of 50 plots from the CVS dataset containing *Ilex opaca*, there were on average 122 woody species from our species pool of 179 in these plots and an average plot richness of 16 woody species; subtracting the average plot richness value from the average total co-occurrence value yields  $\theta = 106$ . At the other extreme, 50 plots of *Taxodium ascendens* contained an average of 29 total species with

an average plot richness of  $c. 5$ , giving  $\theta = 24$ . Average plot richness,  $\mu(\alpha)$ , was only a small factor in the calculation of  $\theta$ , ranging from 4.6 to 19.5 with a mean of 14.4 (SD = 3.2), compared to 24–106 for  $\theta$ , with a mean of 66.1 (SD = 19.3). Species that had extremely low values of associated plot richness also yielded low  $\theta$ -values, but above a  $\mu(\alpha)$  threshold of  $c. 10$  species there was no relationship between mean plot richness and  $\theta$  ( $P > 0.1$ ).

The total number of occurrences of each species in the CVS dataset was significantly (ANOVA,  $P < 0.01$ ) but weakly ( $R^2 = 0.08$ ) correlated with  $\theta$  (Fig. 3a), with widespread species like *Acer rubrum* and *Nyssa sylvatica* more likely to be generalists than geographical- and habitat-restricted species like *Abies fraseri* and *Nyssa aquatica*. Several species were very well represented in the database but had very low  $\theta$ -values (*Pinus palustris* and *Quercus laevis*); other species were represented by few plots but were associated with high species turnover within those plots (*Ulmus rubra*, *Morus rubra* and *Ostrya virginica*). Species'  $\theta$ -values were closely predicted by total number of co-occurrences in the data, regardless of associated plot richness or plot frequency ( $R^2 = 0.86$ ,  $P < 0.001$ ).

### SENSITIVITY OF HABITAT SPECIALIZATION TO SPATIAL EXTENT OF OCCURRENCES

#### Montane upland forests

Values of  $\theta$  for 71 species that met the 20-plot occurrence cut-off for the montane upland forests dataset are listed in Table S1. Values of  $\theta$  range from 51.8 (*Carpinus caroliniana*) to 31 (*Abies fraseri* and *Viburnum lantanoides*), with a mean of 44.5 (SD = 4.4), and were well predicted by total number of co-occurrences in the MUF dataset ( $R^2 = 0.57$ ,  $P < 0.0001$ ). On average, there were seven more woody species per 0.1 ha in montane upland forests than for the full CVS dataset (mean richness = 21.1, SD = 2.2).  $\theta$  was not constrained by low values of total plot occurrences in the MUF dataset; high frequencies of occurrence guaranteed high  $\theta$ -values, but high  $\theta$  could also occur with very few plot occurrences, such as those of *Carpinus caroliniana* (Fig. 3b, Table S1). Despite these differences, there was a strong positive correlation between species'  $\theta$ -values calculated using the CVS and MUF datasets ( $P < 0.001$ , Fig. 4). Species that maintain high abundances in montane forests but have wide geographical distribution, such as *Acer rubrum* and *Prunus serotina*, have high  $\theta$ -values for both datasets, while species whose distributions are centred in the mountains, such as *Tsuga canadensis*, *Rhododendron maximum*, and *R. catawbiense*, have relatively higher  $\theta$ -values in montane forests.

#### Forests of the Joyce Kilmer-Slickrock Wilderness Area

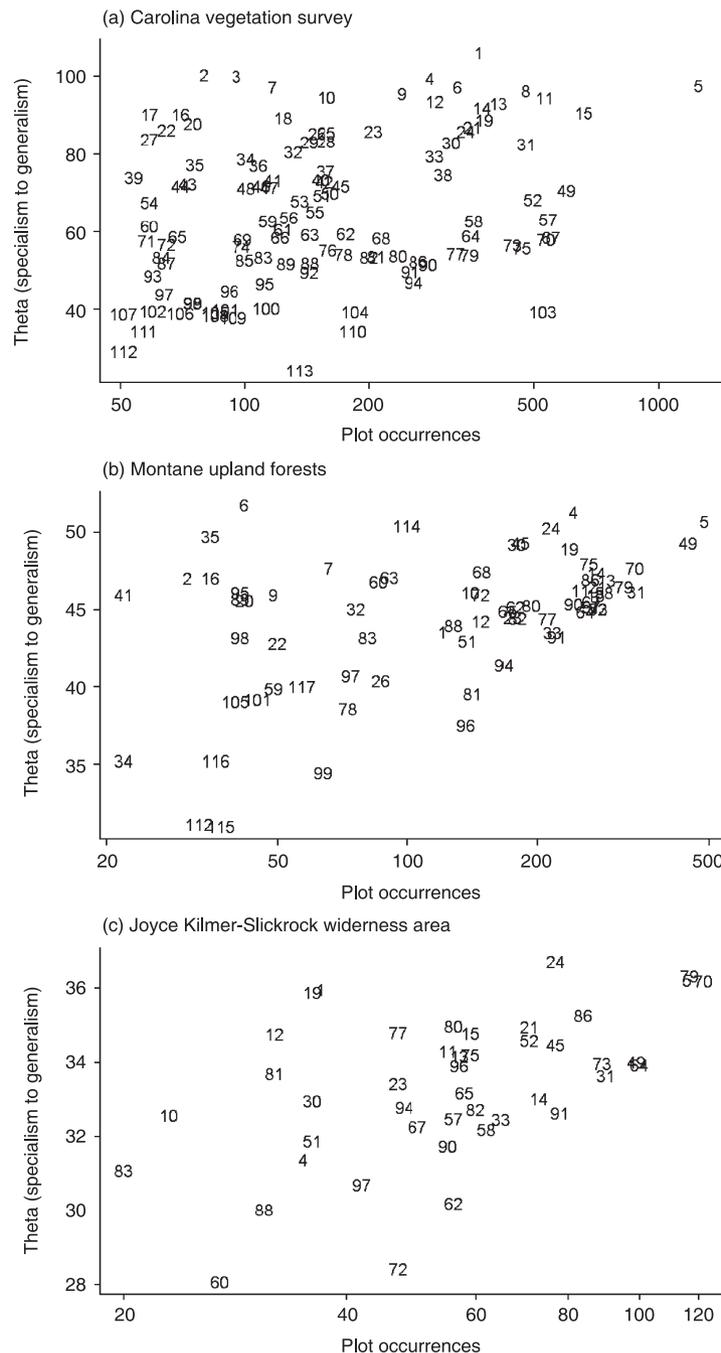
Values of  $\theta$  for 44 species that met the 20-plot occurrence cut-off for the Slickrock dataset are listed in Table S2. Values of  $\theta$  range from 36.7 (*Fagus grandifolia*) to 28.1

**Table 3** Habitat generalist–specialist tendencies for 113 common woody species of the Southeastern USA as represented by the Carolina Vegetation Survey, listed in descending order of  $\theta$  (generalists to specialists). See text for calculation of  $\theta$  (SD based on 100 permutations of selecting 50 plots). Mean 0.1-ha richness of plots containing given species is listed as  $\mu(\alpha)$ , followed by total number of co-occurrences in the data ( $\gamma$ ), total number of plot occurrences, and the mean number of co-occurrences in a selection of 50 plots containing the species

Code	Species	$\theta$	$\theta$ SD	$\mu(\alpha)$	Total $\gamma$	Total plots	$\mu(\gamma)$ , 50 plots
1	<i>Ilex opaca</i>	106.16	4.84	16.00	162	368	122.16
2	<i>Ulmus rubra</i>	100.31	4.05	17.22	131	80	117.53
3	<i>Morus rubra</i>	100.01	3.53	16.86	132	96	116.87
4	<i>Prunus serotina</i>	99.45	6.19	16.15	149	282	115.60
5	<i>Acer rubrum</i>	97.59	5.34	14.82	166	1256	112.41
6	<i>Carpinus caroliniana</i>	97.28	4.89	15.83	150	328	113.11
7	<i>Ostrya virginiana</i>	97.19	4.70	17.38	135	117	114.57
8	<i>Liquidambar styraciflua</i>	96.25	5.41	13.04	155	480	109.29
9	<i>Diospyros virginiana</i>	95.63	6.24	11.02	144	241	106.65
10	<i>Carya cordiformis</i>	94.72	5.10	16.74	136	158	111.46
11	<i>Cornus florida</i>	94.45	5.46	17.17	154	531	111.62
12	<i>Carya alba</i>	93.58	6.01	17.26	151	288	110.84
13	<i>Carya glabra</i>	93.05	5.58	17.24	150	411	110.29
14	<i>Fraxinus americana</i>	91.84	7.49	16.85	147	375	108.69
15	<i>Nyssa sylvatica</i>	90.60	6.28	15.75	159	661	106.35
16	<i>Juglans nigra</i>	90.28	3.20	16.60	115	70	106.88
17	<i>Cercis canadensis</i>	90.22	1.52	19.51	111	59	109.73
18	<i>Vaccinium arboreum</i>	89.35	4.73	12.96	122	124	102.31
19	<i>Quercus alba</i>	88.74	6.51	17.65	149	379	106.39
20	<i>Carya ovata</i>	87.85	3.10	18.19	112	75	106.04
21	<i>Sassafras albidum</i>	86.94	6.04	16.87	148	356	103.81
22	<i>Crataegus macrosperma</i>	86.15	3.93	16.22	108	65	102.37
23	<i>Quercus velutina</i>	85.86	7.27	18.27	138	205	104.13
24	<i>Fagus grandifolia</i>	85.79	4.87	15.86	139	343	101.65
25	<i>Ulmus alata</i>	85.53	5.27	14.82	127	158	100.35
26	<i>Symplocos tinctoria</i>	85.33	5.99	16.86	127	150	102.19
27	<i>Quercus phellos</i>	83.79	2.38	14.43	101	59	98.22
28	<i>Quercus nigra</i>	83.34	4.34	13.96	119	158	97.30
29	<i>Juniperus virginiana</i>	83.09	6.03	13.87	119	143	96.96
30	<i>Hamamelis virginiana</i>	82.87	6.37	17.46	144	316	100.33
31	<i>Liriodendron tulipifera</i>	82.45	6.91	17.03	147	480	99.48
32	<i>Lindera benzoin</i>	80.45	4.30	16.10	120	131	96.55
33	<i>Tilia americana</i>	79.33	7.20	16.21	134	288	95.54
34	<i>Quercus falcata</i>	78.66	4.82	12.85	108	101	91.51
35	<i>Viburnum nudum</i>	77.23	3.04	14.28	100	76	91.51
36	<i>Persea palustris</i>	76.99	3.72	12.49	100	108	89.48
37	<i>Morella cerifera</i>	75.53	4.81	11.72	107	157	87.25
38	<i>Pinus taeda</i>	74.56	4.97	10.68	118	303	85.24
39	<i>Pinus echinata</i>	73.90	2.02	14.01	89	54	87.91
40	<i>Fraxinus pennsylvanica</i>	73.34	6.80	12.98	115	153	86.32
41	<i>Asimina triloba</i>	73.22	3.92	13.97	101	117	87.19
42	<i>Quercus laurifolia</i>	72.79	5.57	13.29	108	156	86.08
43	<i>Vaccinium fuscatum</i>	72.27	3.53	11.82	92	73	84.09
44	<i>Quercus pagoda</i>	71.69	2.82	15.15	91	70	86.84
45	<i>Amelanchier arborea</i>	71.64	5.98	17.10	115	171	88.74
46	<i>Quercus stellata</i>	71.64	5.63	11.15	104	110	82.79
47	<i>Platanus occidentalis</i>	71.28	5.35	12.40	103	115	83.68
48	<i>Quercus michauxii</i>	71.07	3.10	15.70	98	101	86.77
49	<i>Quercus rubra</i>	70.58	4.99	16.49	123	600	87.07
50	<i>Ulmus americana</i>	69.94	5.19	13.15	100	161	83.09
51	<i>Vaccinium stamineum</i>	69.29	5.38	17.63	113	154	86.92
52	<i>Oxydendrum arboreum</i>	68.26	6.54	16.96	127	499	85.22
53	<i>Celtis laevigata</i>	67.79	4.10	12.97	95	136	80.76
54	<i>Carya pallida</i>	67.47	1.41	14.89	84	59	82.36
55	<i>Magnolia virginiana</i>	65.10	5.66	10.32	97	148	75.42
56	<i>Quercus hemisphaerica</i>	63.54	4.90	11.76	89	128	75.30
57	<i>Quercus montana</i>	63.01	4.64	13.83	108	542	79.16
58	<i>Robinia pseudoacacia</i>	62.73	4.30	16.48	109	359	79.21

Table 3 continued

Code	Species	$\theta$	$\theta$ SD	$\mu(\alpha)$	Total $\gamma$	Total plots	$\mu(\gamma)$ , 50 plots
59	<i>Pinus virginiana</i>	62.62	4.42	16.37	91	114	78.99
60	<i>Calycanthus floridus</i>	61.46	2.36	18.87	82	59	80.33
61	<i>Ilex decidua</i>	60.54	6.05	12.75	90	124	73.29
62	<i>Magnolia acuminata</i>	59.37	4.63	17.54	99	176	76.91
63	<i>Vaccinium corymbosum</i>	59.25	4.44	16.63	89	144	75.88
64	<i>Halesia tetraptera</i>	58.80	4.96	16.22	101	353	75.02
65	<i>Rubus canadensis</i>	58.63	2.05	16.27	78	69	74.90
66	<i>Ilex vomitoria</i>	58.57	4.62	12.49	84	122	71.06
67	<i>Kalmia latifolia</i>	58.50	4.56	16.60	111	551	75.10
68	<i>Amelanchier laevis</i>	58.27	4.33	16.12	91	214	74.39
69	<i>Quercus virginiana</i>	57.93	5.13	11.33	80	99	69.26
70	<i>Tsuga canadensis</i>	57.92	4.45	16.21	101	535	74.13
71	<i>Cyrilla racemiflora</i>	57.64	4.31	10.63	71	58	68.27
72	<i>Hydrangea arborescens</i>	56.62	2.28	17.73	77	65	74.35
73	<i>Betula lenta</i>	56.50	3.78	16.30	100	446	72.80
74	<i>Acer negundo</i>	56.04	4.55	12.82	80	98	68.86
75	<i>Rhododendron maximum</i>	55.68	3.77	16.15	97	469	71.83
76	<i>Nyssa biflora</i>	55.17	7.89	8.52	92	159	63.69
77	<i>Pinus strobus</i>	54.34	3.95	17.34	90	324	71.68
78	<i>Pinus rigida</i>	53.99	5.08	17.30	87	174	71.29
79	<i>Acer pensylvanicum</i>	53.93	3.56	15.61	86	350	69.54
80	<i>Aesculus flava</i>	53.77	3.17	14.14	88	235	67.91
81	<i>Quercus coccinea</i>	53.49	7.01	17.62	97	208	71.11
82	<i>Ilex montana</i>	53.38	2.99	15.98	81	200	69.36
83	<i>Vaccinium simulatum</i>	53.38	3.48	16.09	75	111	69.47
84	<i>Fraxinus caroliniana</i>	53.33	4.37	10.54	68	63	63.87
85	<i>Persea borbonia</i>	52.69	3.94	12.63	76	100	65.32
86	<i>Magnolia fraseri</i>	52.28	3.55	16.77	85	263	69.05
87	<i>Osmanthus americanus</i>	51.94	2.24	13.46	69	65	65.40
88	<i>Rhododendron calendulaceum</i>	51.85	2.41	17.92	78	144	69.77
89	<i>Rhododendron minus</i>	51.69	4.24	17.10	81	126	68.79
90	<i>Castanea dentata</i>	51.52	2.68	17.94	80	278	69.46
91	<i>Acer saccharum</i>	49.59	2.54	15.02	77	252	64.61
92	<i>Quercus marilandica</i>	49.41	4.86	8.17	73	143	57.58
93	<i>Ilex coriacea</i>	48.58	3.11	9.37	60	60	57.95
94	<i>Betula alleghaniensis</i>	46.98	3.13	13.28	72	257	60.26
95	<i>Rhododendron catawbiense</i>	46.59	2.89	15.06	66	112	61.65
96	<i>Gaylussacia ursina</i>	44.65	2.35	18.61	70	92	63.26
97	<i>Prunus pensylvanica</i>	43.74	1.99	13.99	59	64	57.73
98	<i>Tsuga caroliniana</i>	41.65	3.84	17.40	65	75	59.05
99	<i>Acer spicatum</i>	41.49	2.50	11.87	56	75	53.36
100	<i>Quercus incana</i>	40.16	3.95	7.35	57	113	47.51
101	<i>Picea rubens</i>	39.69	1.75	11.91	54	90	51.60
102	<i>Quercus lyrata</i>	39.39	1.60	11.47	52	60	50.86
103	<i>Pinus palustris</i>	39.21	5.03	4.99	78	528	44.20
104	<i>Pinus serotina</i>	39.19	7.78	6.03	68	185	45.22
105	<i>Leucothoe recurva</i>	39.05	2.28	18.45	61	85	57.50
106	<i>Taxodium distichum</i>	38.89	2.54	9.45	51	70	48.34
107	<i>Sabal palmetto</i>	38.59	0.75	11.14	49	51	49.73
108	<i>Pinus pungens</i>	38.36	2.99	16.76	60	85	55.12
109	<i>Quercus margarettiae</i>	37.71	2.48	8.23	53	94	45.94
110	<i>Quercus laevis</i>	34.26	2.93	5.92	55	183	40.18
111	<i>Nyssa aquatica</i>	34.22	1.60	9.02	44	57	43.24
112	<i>Abies fraseri</i>	29.19	0.29	10.74	39	51	39.93
113	<i>Taxodium ascendens</i>	24.20	3.77	4.58	39	136	28.78
	Other:						
114	<i>Rubus allegheniensis</i>						
115	<i>Viburnum lantanoides</i>						
116	<i>Sorbus americana</i>						
117	<i>Castanea pumila</i>						



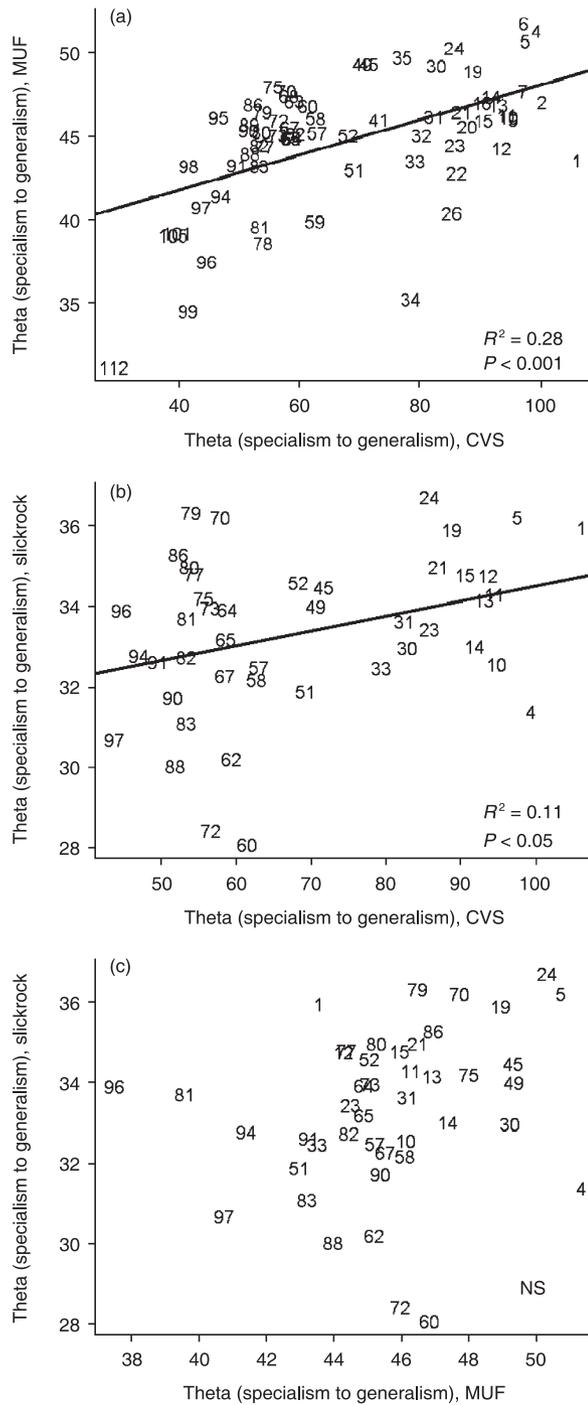
**Fig. 3** Co-occurrence-based estimate of niche width ( $\theta$ ) vs. the total number of plot occurrences for woody species inhabiting three nested areas of different spatial extent. Carolina Vegetation Survey spans *c.* 800 km of the Southeast USA, Montane Upland Forests comprise a nested linear extent of *c.* 300 km within the CVS area, and the Joyce-Kilmer Slickrock Wilderness Area comprises a 6800-ha area within the MUF region. Occurrences are log-transformed. Species codes are listed in Table 3.

(*Calycanthus floridus*), with a mean of 33.4 (SD = 2.0). Values of  $\theta$  were not related to values of mean plot richness ( $P > 0.5$ ), but were significantly associated with total number of plot occurrences ( $R^2 = 0.27$ ,  $P < 0.001$ ; Fig. 3c) and total number of species co-occurrences ( $R^2 = 0.64$ ,  $P < 0.001$ ) within Slickrock. As with the MUF dataset,  $\theta$ -values calculated with the CVS dataset significantly predicted  $\theta$ -values calculated with 133 Slickrock plots ( $P < 0.05$ , Fig. 4); however, MUF

and Slickrock  $\theta$ -values were not correlated ( $P > 0.1$ ; Fig. 4).

#### CORRELATION OF $\theta$ -VALUES WITH LIFE-HISTORY TRAITS

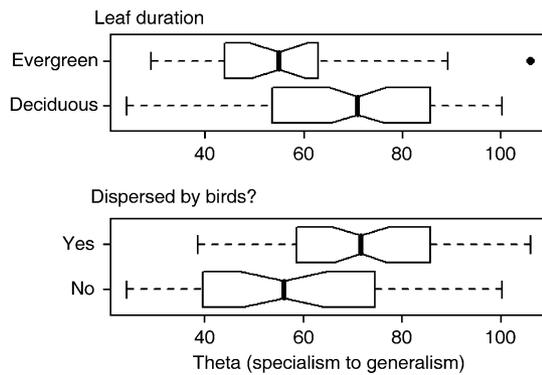
We used single contrasts and a multiple testing criterion to determine whether values of  $\theta$  derived from the full extent CVS dataset were related to a suite of life-history



**Fig. 4** Pairwise contrasts of co-occurrence-based estimates of niche width ( $\theta$ ) of each species using different nested datasets of differing spatial extent: (a) Montane upland forests vs. the Carolina Vegetation Survey (CVS) area, (b) Joyce Kilmer-Slickrock Wilderness Area vs. the CVS area, and (c) Joyce Kilmer-Slickrock vs. Montane upland forests. Statistics and mean regression line are from simple linear regressions.

traits and summary statistics for the environmental distributions of the 113 species in the CVS analysis. Table 2 lists the 27 variables, their correlation to  $\theta$ , single-test  $P$ -values, and  $P$ -values corrected for multiple testing. Nine of the 27 variables were significantly correlated with  $\theta$  in separate tests of significance ( $P < 0.05$ ), and two variables were retained after corrections for multiple testing: the total range of growing degree days and the total range annual temperature, of the entire North

American range of each species. In separate tests of significance, deciduous species and those species dispersed by birds had higher  $\theta$ -values than evergreens and non-bird dispersed species, respectively (Table 2; Fig. 5). In addition, species with a shorter life span (both typical and maximum), lower mean moisture index, shorter maximum dispersal distance, and shorter stature were associated with higher  $\theta$ -values, although none of these were significant after multiple test correction (Table 2).



**Fig. 5** Difference in co-occurrence-based estimates of niche width ( $\theta$ ) values among species of contrasted life history, including leaf duration (evergreen ( $n = 28$ ) or deciduous ( $n = 81$ )) and whether seeds are dispersed by birds (Yes,  $n = 61$ ; No,  $n = 41$ ). Box plots show median values as heavy vertical lines, 95% confidence intervals as notches, 25th to 75th quantiles within boxes, data range within whiskers, and extreme outliers as dots. The evergreen outlier is *Ilex opaca*.

## Discussion

### $\theta$ : A MEANINGFUL MEASURE OF NICHE WIDTH?

Our metric of niche breadth is novel in that it relies only on patterns of species co-occurrences to produce a relative index of habitat generalism vs. specialism. Measurement of species performance along gradients is not necessary; instead we allow patterns of non-gradient species turnover (Vellend 2001) within plots inhabited by a focal species to indicate the degree to which a species occurs in a variety of ‘habitats’. Our approach shares affinities with indirect ordination techniques (Legendre & Legendre 1998) in that it uses patterns of species co-occurrences to dictate species groupings, although we focus on species differences in compositional turnover rather than relative species positions along compositional gradients. Our metric of niche width makes no assumptions about niche shape, distribution and configuration of habitats, or how species are distributed along gradients.

In our approach we assume that niches are perfectly delimited by the species that inhabit them (Levins & Lewontin 1985), a perspective that is not shared by all researchers (see Herbold & Moyle 1986). For example, if some habitat types have been more historically common than others, their niches may be better filled owing to the time required for refined adaptations (Schamp *et al.* 2003). Indeed, our definition of a generalist derives from our assumption that environmental diversity is accurately reflected by the diversity of species that inhabit those environments. Belief in the occurrence of ‘empty niches’ will disconnect our metric from a true niche width. However, given myriad direct and indirect environmental factors, and potential interactions and nonlinearities in their effects on growth, fitness, and dispersal, we assert that an approach that assays species behaviours themselves is a more tractable option to describe niches than the daunting task of identifying

with certainty all the relevant environmental variables and measuring species response to them. Our procedure of combining a particular metric for species turnover (additive partitioning of diversity) with a permutation technique for equalizing plot occurrences among species has several desirable properties, but we advise caution in applying it without careful consideration of the structure of a given survey dataset. Although correlations of niche width and  $\theta$ , or niche width and total co-occurrences, were both very high ( $r > 0.85$ ) in our simulations, comparison of estimated niche widths of any two species will be more variable and subject to the frequency of a species in the survey, especially as frequency nears the chosen threshold of occurrences used in the  $\theta$  calculation. Nonetheless,  $\theta$  can provide valuable information on the generalist tendencies of species in a given area, which can then be the basis of further inquiry using trait or environmental data.

### GENERALIST VS. SPECIALIST SPECIES OF THE SOUTHEASTERN USA

Rankings of species along a generalist–specialist gradient from the CVS dataset exhibit a strong signature of geographical range size (Thompson *et al.* 1998). The top six generalist species (*Ilex opaca*, *Ulmus rubra*, *Morus rubra*, *Prunus serotina*, *Acer rubrum* and *Carpinus caroliniana*) are widespread and abundant in the Southeast in all major bioregions, from the coastal plain to relatively high elevations in the southern Appalachians. Unsurprisingly, tolerance to a wide variety of climate conditions is thus a central component to niche width on a large geographical scale. More unexpected was the relation of habitat generalism at this scale to deciduousness (Fig. 5). After *I. opaca*, the most generalist evergreen species is *Vaccinium arboreum*, 18th on the list, followed by *Juniperus virginiana* at 29th. That deciduous species in the CVS dataset had higher  $\theta$ -values is probably due to the particular habitat affinities of the relatively few evergreen species in the Southeastern Mountains and Piedmont. Several evergreen species in the CVS dataset are restricted to relatively stressful habitats of cold (*Abies fraseri* and *Picea rubens*) or low nutrient availability (several *Pinus* spp., such as *P. pungens*, *P. palustris* and *P. serotina*). Datasets of continental extent that cover a large range of latitudes may reveal a different relationship of niche breadth and deciduousness. This may be because evergreen habits are common to a wide variety of climatic conditions, and include most boreal and tropical species, while deciduousness is restricted to habitats with strongly seasonal distributions of temperature or water availability (Chabot & Hicks 1982).

Most of the top-ranked generalists have fleshy seeds or nuts that can be dispersed long distances by birds and other animals (*I. opaca*, *M. rubra*, *P. serotina*, *C. caroliniana*, *O. virginiana*, *D. virginiana*, *Carya* spp., *C. florida* and *N. sylvatica*). Whether this is characteristic of species with large ranges, or whether there is an additional role of seed dispersal in habitat generalism is

unclear. Species of wide climatic and edaphic tolerances should have the ability to disperse quickly and broadly to open sites for establishment (Ozinga *et al.* 2005). Better dispersers may also continually supply marginal habitats with seed rain, and thus their occurrence in some habitats may be more associated with mass effects (Shimda & Ellner 1984) than autonomous populations (Grime 1998), thus giving the appearance of a wide niche. It is also interesting that several of the top generalists are typically understorey species (*I. opaca* (ranked 1), *M. rubra* (3), *C. caroliniana* (6), *O. virginiana* (7), *D. virginiana* (9), *C. florida* (11) and *C. canadensis* (17)), and the other two in the top five are very shade tolerant (*U. rubra* and *A. rubrum*) and thus disproportionately represented in the understorey. These species may be 'specialists' for a universally distributed 'habitat' of the forest understorey – low irradiance and a buffered microclimate of higher humidity and lower diurnal and seasonal temperature changes (Geiger 1965).

Specialist species in the CVS dataset tended to be those that most frequently occur in relatively extreme environmental circumstances. Cypress (*Taxodium ascendens* and *T. distichum*) and *Nyssa aquatica* typically occur in habitats that are inundated nearly year-round. *Abies fraseri* and *Picea rubens* occupy the coldest environments in the Southeastern USA. Some of the hottest and driest habitats in our study area contain the oaks *Quercus laevis*, *Q. margarettiae* and *Q. incana*. *Pinus pungens*, a central-southern Appalachian endemic, inhabits dry ridges with infertile soils; it and several other pines (especially *P. palustris* and *P. serotina*) are favoured by a regular fire regime.

We did not apply a phylogenetic correction technique to assess whether significant relationships between traits and niche width could be overestimated due to the overrepresentation of certain clades in regions of the generalist-specialist gradient. However, genus- and family-level observations suggest such phylogenetic corrections would be minor. Species-rich genera in the Southeast, including *Quercus* and *Pinus*, as well as other multiple-species genera such as *Acer*, were generally well spread along the gradient (Table 3). One exception were species of the genus *Carya*. Our analysis has most *Carya* as generalists in the Southeastern dataset (Table 3), most likely due to the wide geographical range of most of these species across the Southeast and also the fact that the conspicuous specialist species (*C. aquatica*, *C. carolinae-septentrionalis*, *C. laciniosa* and *C. myristiciformis*) were unlikely to appear in > 50 plots in the dataset. Well represented families (Betulaceae, Ericaceae, Fagaceae and Pinaceae) were also generally found along the entire gradient, although pines and ericads were better represented in the specialist region (Table 3).

#### SCALE SENSITIVITY OF NICHE WIDTH

Estimated niche widths for the same species derived from nested datasets differing greatly in spatial extent

(from *c.* 15–800 km) were significantly and positively correlated (Fig. 4). Generalist species across the Southeast, such as *Acer rubrum*, *Ilex opaca*, *Fagus grandifolia* and *Quercus alba*, also tended to co-occur with relatively many species at both the regional (southern Appalachians) and landscape (Slickrock) scales. Specialist species, including *Abies fraseri*, *Acer spicatum* and *Prunus pensylvanica*, occurred with relatively few species across all scales. Our choice of nested locations may, in part, explain why such consistency should occur. Although much smaller in extent, both smaller datasets included a relatively large range of elevations. As such, those species of large longitudinal range (such as the above generalists) would be somewhat accommodated in habitat by the large elevation range in the smaller datasets, and consequently the reduction in total number of habitats may not have been as large as the reduction in area might imply.

Exceptions to the consistency of  $\theta$  rankings across scales are perhaps more interesting than the overall correlation. *Acer pensylvanicum*, for example, is geographically restricted to the cooler climates of Northeastern USA. It extends south only in the Appalachians, and thus has a relatively low  $\theta$ -value for the Southeast. Within the southern Appalachians, however, this species occurs in many habitat types across a range of elevations, and therefore has a high associated  $\theta$ -value when examined within strictly montane datasets (Fig. 4, species 79). *Acer pensylvanicum* is also a highly shade-tolerant species, and therefore fits the general pattern of understorey species as generalists, perhaps because they are 'specialists' for a ubiquitous understorey environment. A similar case could be made for *Tsuga canadensis*. Although in the absence of disturbance *Tsuga* ultimately becomes a principal canopy component where it occurs, it is perhaps the most shade-tolerant USA tree (Burns & Honkala 1990) and is a common understorey component of nearly all mesic forest types in the southern Appalachians. Like *A. pensylvanicum*, its presumed climatic adaptations restrict its occurrence largely to montane forests in the Southeast, but it is one of the most generalist species within these montane shaded environments (Fig. 4).

#### $\theta$ AND THEORIES OF NICHE WIDTH

There are remarkably few hypotheses concerning the relationship of niche width to species traits. Most discussions of niche width based on survey data either focus on the relationship between plot frequency (global occurrence) and local abundance (e.g. Brown 1984; Burgman 1989; Gaston 1996; Thompson *et al.* 1998) or consider species characteristics in relation to range size and latitude (e.g. Vazquez & Stevens 2004; Morin & Chuine 2006). Basic trade-off theory suggests that species tolerant of a wide variety of ambient conditions should be poorer competitors for specific habitats, but our analysis using  $\theta$  does not suggest that generalist species are generally poor competitors. To

the contrary, many of the most generalist species are highly tolerant of light competition, and species like *Acer rubrum* are increasing in abundance across the eastern USA (Abrams 1992). However, our analysis of niche breadth using co-occurrence data does suggest that unusually stressful habitats (e.g. chronically flooded, extremely xeric and oligotrophic, routinely burned, or climatically harsh) contain species that are specially adapted to such conditions at the expense of occurring elsewhere. At least part of this habitat restriction is due to competition – we note specialists such as *Taxodium* spp., the specialist species of *Quercus* and *Pinus*, and *Rhododendron* spp. will grow widely under cultivation in more benign habitats. Our analysis, in part, supports conceptual models in which competition plays a key role in restricting certain species to stressful habitats (Grime 1974; Austin & Smith 1989; Keddy 1990). Without knowledge of the fundamental environmental tolerances of the species in our region, however, we cannot assess the core aspects of such stress-based trade-offs, such as whether species of stressful habitats have wider fundamental niches (following Austin & Smith 1989).

Species inhabiting many habitat types, large areas, and a large range of environmental conditions might generally be thought to harbour higher levels of genetic diversity (cf. Stebbins 1942, 1980). Many of the generalist species across the Southeast described in this study are well known genetic complexes, composed of either relatively distinct varieties or clines of genetically based morphological or physiological variation. Generalist species thought to display significant clinal genetic variation as reported by Burns & Honkala (1990) include *Acer rubrum*, *Carpinus caroliniana* and *Juglans nigra*. Several generalist species have also been placed into geographically based varieties, including *Fagus grandifolia*, *Ostrya virginiana*, *Diospyros virginiana*, *Nyssa sylvatica*, *Carya glabra* and *Cercis canadensis*. Each of these species has a large geographical range, however, and thus it is difficult to determine whether genetic diversity contributes to niche width in a habitat sense, rather than simply in terms of population size. Our metric  $\theta$  could be used in species-level comparisons of genetic diversity and niche width, particularly for regional and landscape scales that are less influenced by the correlation of  $\theta$  and range size.

## Conclusions

We have used a plot-based survey to demonstrate the utility of our non-gradient, species co-occurrence approach for studies of relative niche width for the species pool of a particular area. Our approach is generally applicable to any dataset from which species co-occurrences can be derived, although we advise caution in the interpretation of  $\theta$  as a descriptor of niche width when using a survey strongly biased toward particular habitats. Analysis of relative niche widths, from habitat generalists to specialists, for woody plant species of the

Southeastern USA suggests niche width is related to certain species characteristics, such as understorey growth habit and mode of seed dispersal, which require more in-depth study. Determination of habitat generalism for the same species using nested datasets of different spatial extent reveals important scale-dependence in habitat generalism for several species, and relatively stable niche width values for others.

Our approach has important applications for theoretical and applied issues in community ecology. Future studies can incorporate calculation of  $\theta$  into analysis of species packing (does the distribution of  $\theta$  change for similar taxa in different regions; does one area have relatively more generalists in its flora than another?). Distributions of  $\theta$  can also be compared between different taxonomic or life-form groups of the same survey locations (are herbs more specialized than trees? Are more vagile animals more generalist?). Conservation strategies might focus on sites with a high number of both generalists (to maximize total surveyed diversity) and specialists (to protect species unlikely to be found elsewhere). The increasing availability of large species-location databases suggests that there is much to be gained from large-scale computational approaches to detecting patterns of species behaviour, and application of  $\theta$  can be a key component for testing behaviour- and trait-related hypotheses in the absence of detailed environmental information.

## Acknowledgements

We gratefully acknowledge the more than 600 individuals who have participated in collection of vegetation data in the Carolina Vegetation Survey archive, and particularly C. Newell, T. Wentworth, A. Weakley and M. Schafale. Colleagues in the UNC Plant Ecology Laboratory provided helpful discussion. Plot collection was made possible through challenge cost-share grants from the United States Forest Service. Data compilation and analysis was supported by NSF grants DBI-9905838 and DBI-0213794 to RKP. JDF was supported by the National Parks Ecological Research Fellowship program, a partnership between the National Park Service, the Ecological Society of America, and the National Park Foundation and funded through a generous grant from the Andrew W. Mellon Foundation. MM was supported by a grant from the German Academy of Natural Scientists Leopoldina (BMBF-LPD 9901/8–94).

## References

- Abrams, M.D. (1992) Fire and the development of oak forests. *Bioscience*, **42**, 346–353.
- Austin, M.P. & Gaywood, M.J. (1994) Current populations on environmental gradients and species response curves in relation to continuum theory. *Journal of Vegetation Science*, **5**, 473–487.
- Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35–47.
- Bazzaz, F.A. (1991) Habitat selection in plants. *American Naturalist*, **137**, S116–S130.

- Bazzaz, F.A. (1996) *Plants in Changing Environments*. Cambridge University Press, Cambridge.
- Brown, J.H. (1984) On the relationship between abundance and distribution of species. *American Naturalist*, **124**, 255–279.
- Burgman, M.A. (1989) The habitat volumes of scarce and ubiquitous plants: a test of the model of environmental control. *American Naturalist*, **133**, 228–239.
- Burns, R.M. & Honkala, B.H. (Technical Coordinators) (1990) *Silvics of North America: 1. Conifers; 2. Hardwoods*. Agriculture Handbook 654. US Department of Agriculture, Forest Service, Washington, DC. [http://www.na.fs.fed.us/spfo/pubs/silvics\\_manual](http://www.na.fs.fed.us/spfo/pubs/silvics_manual).
- Chabot, B.F. & Hicks, D.J. (1982) The ecology of leaf life spans. *Annual Review of Ecology and Systematics*, **13**, 229–259.
- Colwell, R.K. & Futuyma, D.J. (1971) On the measurement of niche breadth and overlap. *Ecology*, **52**, 567–576.
- Crawley, M.J. (1997) *The structure of plant communities*. *Plant Ecology*, 2nd edn. Blackwell, Oxford.
- Ellenberg, H. (1974) *Zeigwerte der Gefässpflanzen Mitteleuropas*. *Scripta Geobotanica*, **9**. Goeltze, Gottingen.
- Fridley, J.D., Peet, R.K., White, P.S. & Wentworth, T.R. (2005) Connecting fine- and broad-scale patterns of species diversity: species–area relationships of Southeastern US flora. *Ecology*, **86**, 1172–1177.
- Gaston, K.J. (1996) The multiple forms of the interspecific abundance–distribution relationship. *Oikos*, **75**, 211–220.
- Geiger, R. (1965) *The Climate Near the Ground*. Harvard University Press, Cambridge, MA, USA.
- Grime, J.P. (1974) Vegetation classification by reference to strategies. *Nature*, **250**, 26–31.
- Grime, J.P. (1998) Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *Journal of Ecology*, **86**, 902–910.
- Herbold, B. & Moyle, P.B. (1986) Introduced species and vacant niches. *American Naturalist*, **128**, 751–760.
- Hochberg, Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, 800–802.
- Horn, H.S. (1966) The measurement of ‘overlap’ in comparative ecological studies. *American Naturalist*, **100**, 419–424.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415–457.
- Kartesz, J.T. (1999) *A Synonymized Checklist and Atlas with Biological Attributes for the Vascular Flora of the United States, Canada, and Greenland*. 1st edn. Synthesis of the North American Flora, Version 1.0 (ed. by J.T. Kartesz & C.A. Meacham). North Carolina Botanical Garden, Chapel Hill, NC, USA.
- Keddy, P.A. (1990) Competitive hierarchies and centrifugal organization in plant communities. *Perspectives on Plant Competition* (ed. by J.B. Grace & D. Tilman), pp. 265–289. Academic Press, San Diego, USA.
- Koleff, P., Gaston, K.J. & Lennon, J.J. (2003) Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, **72**, 367–382.
- Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, **76**, 5–13.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*. Elsevier, New York.
- Levins, R. (1968) *Evolution in Changing Environments*. Princeton University Press, Princeton, NJ.
- Levins, R. & Lewontin, R.C. (1985) *The Dialectical Biologist*. Harvard University Press, Cambridge, MA.
- Loreau, M. (2000) Are communities saturated? On the relationship between alpha, beta and gamma diversity. *Ecology Letters*, **3**, 73–76.
- Minchin, P. (1987) Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio*, **71**, 145–156.
- Morin, X. & Chuine, I. (2006) Niche breadth, competitive strength and range size of tree species: a trade-off based framework to understand species distribution. *Ecology Letters*, **9**, 185–195.
- Newell, C.L., Peet, R.K. & Harrod, J.C. (1997) *Vegetation of Joyce Kilmer-Slickrock Wilderness, North Carolina*. Report to US Forest Service. Department of Agriculture, National Forests of North Carolina. Biology Department, Chapel Hill, NC, USA.
- Ozinga, W.A., Schaminee, J.H.J., Bekker, R.M., Bonn, S., Poschod, P., Tackenberg, O., Bakker, J. & van Groenendael, J.M. (2005) Predictability of plant species composition from environmental conditions is constrained by dispersal limitation. *Oikos*, **108**, 555–561.
- Peet, R.K., Wentworth, T.R. & White, P.S. (1998) A flexible, multipurpose method for recording vegetation composition and structure. *Castanea*, **63**, 262–274.
- Prasad, A.M. & Iverson, L.R. (1999–ongoing) *A Climate Change Atlas for 80 Forest Tree Species of the Eastern United States*. [Database]. Northeastern Research Station, USDA, Forest Service, Delaware. URL: <http://www.fs.fed.us/ne/delaware/atlas/index.html>.
- R Development Core Team (2005) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Schamp, B.S., Aarssen, L.W. & Lee, H. (2003) Local plant species richness increases with regional habitat commonness across a gradient of forest productivity. *Folia Geobotanica*, **38**, 273–280.
- Shimada, A. & Ellner, S. (1984) Coexistence of plant species with similar niches. *Vegetatio*, **58**, 29–55.
- Stebbins, G.L. (1942) The genetic approach to problems of rare and endemic species. *Madrono*, **6**, 241–258.
- Stebbins, G.L. (1980) Rarity of plant species: a synthetic viewpoint. *Rhodora*, **82**, 77–86.
- Thompson, R.S., Anderson, K.H. & Bartlein, P.J. (1999) *Distributions of Important Trees and Shrubs in North America*. US Geological Survey Professional Paper 1650 A&B, online, Version 1.0, 14 December 1999, <http://pubs.usgs.gov/pp/p1650-a/index.html>.
- Thompson, K., Hodgson, J.G. & Gaston, K.J. (1998) Abundance–range size relationships in the herbaceous flora of central England. *Journal of Ecology*, **86**, 439–448.
- US EPA (2002) *Level III Ecoregions of the United States, August 2002 revision*. National Health and Environmental Effects Research Laboratory, Corvallis, OR, USA.
- USDA & NRCS (2004) *The PLANTS Database, Version 3.5* (<http://plants.usda.gov>). National Plant Data Center, Baton Rouge, LA, USA.
- Vazquez, D.P. & Stevens, R.D. (2004) The latitudinal gradient in niche breadth: concepts and evidence. *American Naturalist*, **164**, E1–E19.
- Veech, J.A., Summerville, K.S., Crist, T.O. & Gering, J.C. (2002) The additive partitioning of species diversity: recent revival of an old idea. *Oikos*, **99**, 3–9.
- Vellend, M. (2001) Do commonly used indices of  $\beta$ -diversity measure species turnover? *Journal of Vegetation Science*, **12**, 545–552.
- Weakley, A.S. (2005) *Flora of the Carolinas, Virginia, and Georgia, Working Draft of March 2005*. North Carolina Botanical Garden. <http://www.herbarium.unc.edu/flora.htm>.
- Whittaker, R.H. (1956) Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1–80.
- Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 279–338.

J. D. Fridley,  
D. B. Vandermast,  
D. M. Kuppinger,  
M. Manthey &  
R. K. Peet,

Witkowski, E.T.F. & Lamont, B.B. (1997) Does the rare *Banksia goodii* have inferior vegetative, reproductive or ecological attributes compared with its widespread co-occurring relative *B. gardneri*? *Journal of Biogeography*, **24**, 469–482.  
Wuenschel, J.E. (1969) Niche specification and competition modeling. *Journal of Theoretical Biology*, **25**, 436–443.

Received 15 November 2006; revision accepted 16 February 2007

Handling Editor: Fernando Maestre

### Supplementary material

The following supplementary material is available for this article:

**Appendix S1** Algorithm for calculation of a co-occurrence based generalist–specialist metric.

**Appendix S2** R code implementing the generalist–specialist metric.

**Appendix S3** R code for niche width simulations.

**Table S1** Habitat generalist–specialist tendencies for 71 woody species of montane upland forests of the southern Appalachians (USA)

**Table S2** Habitat generalist–specialist tendencies for 44 woody species of Joyce Kilmer-Slickrock Wilderness Area, NC (USA)

This material is available as part of the online article from <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2745.2007.01236.x>

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.