TITLE

Optimal strategies for sampling functional traits in species-rich forests

AUTHORS AND AFFILIATIONS

C. E. Timothy Paine (Corresponding author), c.e.t.paine@stir.ac.uk

- Biological and Environmental Sciences, University of Stirling, Stirling, FK9 4LA, UK, phone: +44 (0)1786 467785, fax: +44 (0) 1786 467843

Christopher Baraloto, Chris.Baraloto@ecofog.gf

- INRA, UMR "Ecologie des Forêts de Guyane", 97387, Kourou Cedex, French Guiana
- International Center of Tropical Botany, Department of Biology, Florida International University, Miami, Florida, 33199 USA

Sandra Diaz, sdiaz@efn.uncor.edu

- Instituto Multidisciplinario de Biología Vegetal (IMBIV-CONICET) and Departamento de Diversidad Biológica y Ecología, FCEFyN, Universidad Nacional de Córdoba, CC 495, 5000 Córdoba, Argentina.

AUTHOR CONTRIBUTIONS
All authors conceived the question, CETP and CB collected data with the BRIDGE team, CETP performed the simulations and wrote the first paper with contributions from all authors.

ABSTRACT

1. Functional traits provide insight into a variety of ecological questions, yet the optimal sampling method to estimate the community-level distribution of plant functional trait values remains a subject of debate, especially in species-rich forests.
2. We present a simulation analysis of the trait distribution of a set of nine completely sampled permanent plots in the lowland rain forests of French Guiana.
3. Increased sampling intensity consistently improved accuracy in estimating community-weighted means and variances of functional trait values, whereas there was substantial variation among functional traits, and minor differences among sampling strategies.
4. Thus, investment in intensified sampling yields a greater improvement in the accuracy of estimation than does an equivalent investment in sampling design complication.
5. Notably, 'taxon-free' strategies frequently had greater accuracy than did abundance-based strategies, which had the additional cost of requiring botanical surveys.
6. We conclude that there is no substitute for extensive field sampling to accurately characterize the distribution of functional trait values in species-rich forests.

INTRODUCTION

Functional traits impact plant fitness via their effects on recruitment, growth, reproduction and survival (Lavorel & Garnier 2002; Violle et al. 2007). Scaling from functional traits measured on individuals up to community-level distributions of trait values has provided insight into a variety of issues surrounding community assembly and ecosystem processes (McGill et al. 2006). Traits have been particularly useful in species-rich forests where the rarity of many species has stymied efforts to characterize their niche requirements (Baraloto, Paine, Poorter, et al. 2010; Fortunel et al. 2013). Standardized protocols for the measurement of functional traits and intensive efforts to compile trait data have fuelled explosive growth of functional trait databases (Kattge et al. 2011; Pérez-Harguindeguy et al. 2013), which in turn, support the development of trait-based dynamic global vegetation models (Scheiter, Langan, & Higgins 2013).

An investigator's choice of sampling strategy can have a major influence on the inferences he or she makes regarding the strength and extent of ecological processes. For example, common sampling

designs for coring of trees can systematically bias estimations of forest growth and productivity (Nehrbass-Ahles et al. 2014). The increasing use of functional traits in plant ecology spurred Baraloto et al. (2010a) to assess trait-sampling strategies in species-rich forests. After measuring the traits on every individual tree in nine 1-ha plots of lowland tropical rain forest, they simulated performing four sampling strategies over a gradient of sampling intensity. Their study, however, suffered from a number of defects, which we remedy in the current contribution. First, their analysis did not assess interactions between strategy and intensity. Second, Baraloto et al. (2010a) classified the results of their simulations into successes and failures using an arbitrary criterion, then analysed these binomally-distributed variables as though they followed Gaussian distributions. Third, their analysis did not take into account variation among sites. Finally, Baraloto et al. (2010a) examined sampling strategies based on species abundance only in passing, although such strategies are widely used in functional trait studies in extra-tropical ecosystems (Conti & Díaz 2013; Fisichelli, Frelich, & Reich 2014; Lavorel et al. 2008; Pakeman & Quested 2007). This paper updates Baraloto et al. (2010a) with an expanded and refined response to the question: What sampling strategy yields the most accurate characterization of community-level trait distributions in species-rich forest communities?

We expand upon Baraloto *et al.* (2010a) by evaluating the interacting effects of sampling strategy, sampling intensity, and functional trait identity on the probability of accurately characterizing the community-level distribution of functional trait values. We characterize the distribution of trait values in terms of its mean and variance (Díaz et al. 2007; Violle et al. 2007). First, we obtain the true community-level distributions by assessing the trait values of every individual tree (Baraloto *et al.* 2010). Then, in simulation, we estimate the mean and variance of functional traits that result from applying each sampling strategy over a range of sampling intensities. We compare these estimates with the true values to determine the accuracy of estimation for every combination of strategy, intensity and trait. We include sampling strategies in which individuals are selected based upon their regional or local stem density or basal area, as measures of abundance, and we analyse their accuracy using mixed-effect models.

MATERIALS AND METHODS

We sampled functional traits on 4672 individual trees representing 668 species in nine 1-ha plots of lowland rain forest in French Guiana between November 2007 and September 2009 (Baraloto, Paine, Patiño, et al. 2010). In each plot, all trees >10 cm diameter at breast height (d.b.h.) were mapped, measured for d.b.h. and climbed to obtain a branch for leaf samples. For the current study, we used the same 10 functional traits as analysed by Baraloto *et al.* (2010a). The eight leaf traits are correlated with the global leaf economics spectrum (Díaz et al. 2004; Wright et al. 2004): chlorophyll content, concentrations of $\delta^{13}C$, N and $\delta^{15}N$, C:N ratio, thickness, toughness and specific leaf area (SLA), one is associated with the global wood economics spectrum (Chave et al. 2009): bole density, and one is of uncertain affinities (Baraloto, Paine, Poorter, et al. 2010): laminar surface

area. The methods of trait sampling in the field and subsequent laboratory analysis are detailed in Baraloto *et al.* (2010a).

We investigated the accuracy of 12 plot-based strategies. Three are based on abundance estimated as basal area, three are based on abundance estimated as number of stems, three are 'taxon-free' (sensu Lavorel et al. 2008), and the final three have varying motivations. The strategies vary widely in the amount of field data required prior to trait sampling, and each emphasizes different aspects of the tree community (Table 1). Most strategies were included because they have recently been used in published studies. The BA REGION and BA LOCAL strategies sample individuals on the basis of the basal area of their conspecifics in the region or local plot, respectively. BA INDIVIDUAL, on the other hand, samples individuals on the basis of their own basal area. It, therefore, tends to sample very large individuals. As such, BA REGION requires botanical inventories and d.b.h. measurements from multiple plots, whereas BA LOCAL and BA INDIVIDUAL require botanical inventories and d.b.h. measurements from a single local plot. Variants of these strategies have been applied in non-tropical communities (Conti & Díaz 2013; Fisichelli et al. 2014; Lavorel et al. 2008; Pakeman & Quested 2007). The commonness strategies (COMMON REGION and COMMON LOCAL) are similar, but are based on stem density in the region or in the local plot, and require either regional or local botanical inventories, respectively. The SPECIES strategy flips COMMON LOCAL on its head, and weights species inversely by their local abundance, thus emphasizing relatively rare species.

The three taxon-free strategies require no botanical determinations. In CLUMP, three random spatial locations within the plot are selected, and trees were selected based on their proximity to these foci. A variant of CLUMP has been used by the RAINFOR project (Patiño et al. 2009). The SIZE strategy stratifies the tree community into five equally spaced size-classes based on d.b.h, from which individuals were drawn at random (Poorter, Bongers, & Bongers 2006). In RANDOM, individual trees are randomly selected from the plot. Although we have not seen this strategy used in the literature, it provides a basis for comparison with other more complicated strategies.

The final three strategies are more idiosyncratic. The ONE PER SP strategy samples a single individual from each species in each plot. Trait values for all individuals of each species in the plot are then set equal to the conspecifics of the sampled individuals (Kraft, Valencia, & Ackerly 2008). This strategy requires a full botanical inventory prior to sampling. The DATABASE strategy samples one individual per species from the region and applies that individual's trait values to all conspecifics (Paine et al. 2012; ter Steege et al. 2006). Note that our implementation of the DATABASE strategy may overestimate its accuracy, because the trait values we use come from a regional data base including only the plots under consideration, rather than from other regions, and because species–trait combinations for which no data were available were left as missing data, rather than being replaced with genus- or family level means (Baraloto, Paine, Patiño, et al. 2010). Finally, the basis of comparison for all sampling strategies was the BRIDGE strategy, in which every individual in the plot is sampled (Baraloto et al. 2012).

We simulated each strategy 500 times in each of the nine plots in the BRIDGE network. Sampling strategies, except those for which sampling intensity does not vary (BRIDGE, DATABASE and ONE PER SP), were executed over a gradient of sampling intensity (1, 2, 3, 5, 10, 20, 30, 40 and 50% of individuals sampled per plot). Sampling intensities greater than 50% are impractical as they require essentially as much work as does 100% (BRIDGE) sampling. In contrast to Baraloto et al. (2010a), the sampling of individuals was probabilistic. Thus, in RANDOM, all individuals in a plot had an equivalent probability of being sampled (equal weights for all individuals), whereas unequal weights were used in all other strategies. In BA REGION, for example, an individual's probability of being sampled was proportional to the BA of its conspecifics in the region (Table 1).

We assessed the performance of each sampling strategy, sampling intensity and trait as the accuracy in recovering the true mean and variance of trait values. Baraloto et al. (2010), considered resampling events as successes (or failures) with respect to an arbitrary threshold. We, in contrast, define accuracy as the absolute value of the per cent difference between the trait values estimated in each re-sampling event to those derived from BRIDGE sampling. We analysed the performance of each of the nine strategies for which intensity was varied using a linear mixed-effect model. Accuracy values were log-transformed for analysis to improve the normality of residuals, and back-transformed for presentation. Tests of significance are irrelevant in the analysis of simulation output. Rather, our primary interest was to assess the relative importance of variation among sampling strategies, functional traits, and sampling intensity in determining the accuracy of estimation of the mean and variance in trait values. Thus, we modelled the log-transformed accuracy of estimation as a normally distributed variable based on the three-way interaction of sampling intensity, strategy and trait. Plots were included as a random effect. We calculated the percentage of variance in the accuracy of estimations explained by sampling strategy, intensity and trait identity using the method of Nakagawa and Schielzeth (2013). Simulations and mixed-effect models were implemented in R 3.1.1 (R Core Development Team 2014) using the lme4 package (Bates et al. 2014). The R script used to conduct the simulations is available as Appendix S1 in Supporting Information.

RESULTS

Sampling intensity, sampling strategy, and the trait under consideration affected the probability of accurately estimating the true mean and variance in functional trait values (Fig. 1). Executed at low sampling intensities, all strategies yielded inaccurate estimates of trait means and variances, but increased intensity consistently increased the accuracy of estimation. Over a broad range of sampling intensity, the taxon-free strategies of CLUMP, SIZE and RANDOM had consistently superior performance in estimating both the mean and variance of most traits. Contrastingly, the strategies based upon basal area or commonness, whether assessed locally or regionally, performed less well. In general, performing botanical inventories prior to trait sampling did not improve the accuracy of estimating the mean or, especially, the variance in trait values. The error in estimates of the trait variances exceeded that of trait means by an order of magnitude, on average.

Despite sampling of every species, and approximately one-third of the individuals in a plot, the ONE PER SP strategy performed no better than sampling an equivalent number of randomly selected individuals. Notably, the DATABASE strategy estimated means and variances approximately as well as field-based sampling strategies that were executed at 10% sampling intensity. There was considerable variation among traits in the probability of correctly estimating their mean and variance, as certain traits were recalcitrant to intensified sampling. Accurate estimates of mean leaf toughness and $\delta^{15}N$ were achieved only rarely, regardless of the sampling intensity or strategy employed.

We partitioned the variance in the accuracy of estimating means and variances of functional trait values to determine the relative importance of sampling strategy, sampling intensity and trait identity (Fig. 2). The most important determinants of accurately estimating the mean and variance in trait values were the trait under investigation and sampling intensity, respectively. Sampling strategy, in contrast, was far less important in determining the accuracy of estimation. Interactions among sampling intensity, strategy, and trait were weak, implying that increasing sampling intensity increases the probability of accurate estimation for all traits and strategies, not only some. Similarly, sampling strategies performed equivalently well (or poorly) in estimating mean and variance for all traits. Finally, sampling strategy, intensity and trait together jointly explained more variance in the accuracy of estimation for means, whereas the percentage of variance that remained unexplained was greater for trait variance.

DISCUSSION

Investment in sampling strategy and sampling intensity

By assessing the interactions between sampling strategy, sampling intensity and trait identity, we dissected the relative contributions of each to the accurate estimation of trait distributions. Overall, sampling intensity was the single best predictor of accuracy in the estimation of the mean and variance of trait values (Fig. 2). Statistical interactions with trait identity and sampling strategy were weak, indicating that more-intense sampling consistently yielded more accurate estimations. Moreover, sampling intensity explained far more variance in the probability of accurate estimation than did sampling strategy, indicating that an investment in sampling intensity would yield a greater improvement in the accuracy of estimation than would an equivalent investment in a complicated sampling design. Sampling intensity was an especially strong predictor of the variance in trait values (Fig. 2). In other words, intense sampling is essential for accurately estimating trait variance. It is not possible to advocate a minimal intensity of trait sampling for all trait-based studies, however, as their tolerance of inaccuracy in functional trait estimation will vary, depending on their objectives. Nevertheless, sampling just 10% of individuals is likely to yield sufficiently small estimation errors in most cases.

Traits varied widely in terms of the accuracy at which their distributions could be estimated. Means and variances for some, such as foliar thickness and $\delta^{13}C$, were accurately estimated by sampling

relatively few individuals. Others, including foliar toughness and $\delta^{15}N$, defied accurate estimation. Accordingly, trait identity affected the accuracy of estimation of trait means (Fig. 2). Interestingly, among-trait variation was much less in the estimation of trait variance. Similar variation among traits has been reported for temperate herbaceous communities (Pakeman & Quested 2007, Lavorel et al. 2008). The reasons behind this variation among traits remain unclear. One possibility is that leaf toughness and $\delta^{15}N$ are more sensitive than other traits we analysed to extrinsic sources of error, which affect their measurement, and thereby complicate efforts to estimate their distributions. Variation in soil isotopic signatures can cause substantial variation in $\delta^{15}N$ (Nardoto et al. 2014), whereas leaf toughness, in this study, was prone to measurement error arising from variation in the distance of the punch test to veins in the leaf blade (Baraloto et al. 2010a). Though reducing sources of extrinsic variation may increase the accuracy of estimating trait values, we do not know why the sampling necessary to obtain accurate estimates varies among traits.

Taxon-free strategies have superior performance

Our simulations indicate that simple, taxon-free, sampling strategies can often characterise community-level functional trait distributions more accurately than can more-intricate strategies. Notably, strategies based on stem density or basal area of species, whether assessed locally or regionally, performed less well than did taxon-free sampling strategies for a given level of sampling intensity (Fig. 1). Abundance-based sampling also implies a greater commitment of resources to field sampling, because it must be preceded by local or regional botanical inventories (Table 1). The SIZE strategy, which entailed the random sampling of size-stratified individuals, yielded the least-biased estimates of community-level trait distributions, with CLUMP and RANDOM running closely behind. These results are consistent with Baraloto *et al.* (2010a), though that study examined only one, slightly different, abundance-based strategy. They also agree with a recent examination of sampling designs for dendrochronology studies, in which random sampling yielded the least-biased estimates of tree growth and productivity (Nehrbass-Ahles et al. 2014).

Therefore, we advocate the use of taxon-free sampling techniques for estimating community-level trait distributions in species-rich forests, though they have relatively poor performance in Alpine grasslands (Lavorel et al. 2008). As a practical matter, concentrating field sampling at a few spatial foci, as in CLUMP, is likely to simplify logistics, accounts for spatial variation in species composition, and may provide a minor improvement in accuracy over random sampling.

What explains the poor performance of abundance-based strategies? In diverse ecosystems, such as those studied here, even the most common species represent a small proportion of individuals. For example, the most common tree species in the forests we studied, *Eschweilera coriacea* (Lecythidaceae), represented only 3.4% of all observed individuals. Across our nine forests, the ten most-common species together represent only 20% of individuals. Contrastingly, at the scale of the Amazon basin, *E. coriacea* and 226 other 'hyperdominants' (1.4% of the approximately 16,000

species of Neotropical trees), represent 50% of all individual trees. Their broad-scale dominance is attributable to their wide spatial ranges, however, rather than to numerical abundance at local sites (ter Steege et al. 2013). In less species-rich plant communities, or those with a less-equitable distribution of individuals among species, abundance-based sampling strategies may be more appropriate. The accuracy of abundance-based sampling strategies would also be reduced if the most-common species did not have the most-typical trait values. This situation should be uncommon, because environmental filtering, which leads to the sorting of species along environmental gradients (Grime 2006), is strong in Neotropical forests (Kraft et al. 2008; Paine et al. 2011). In a recent study of French Guiana trees, species abundance was negatively correlated with the distinctiveness of trait values, as predicted by environmental filtering (Mouillot et al. 2013). Mouillot et al. found similar relationships for herbaceous plants in the Alps, and fish on coral reefs in the Pacific. Abundance-based trait sampling strategies therefore yield poor results in diverse ecosystems not because common species have atypical traits, but rather because even the common species in such species-rich ecosystems are represented by few individuals. They should be more accurate in ecosystems in which more individuals are concentrated among fewer species.

We reiterate the caution sounded by Baraloto *et al*. (2010a) regarding the use of database-derived trait values to estimate community-level trait distributions. They are tempting because the cost associated with extracting trait data from a database is minuscule, compared to that incurred through even low-intensity field work (Baraloto et al. 2010a). Moreover, global trait databases such as TRY (Kattge et al. 2011) are essential for broad-scale comparative studies. At the relatively small scale of this study, however, the DATABASE strategy rarely yielded accurate estimations of means or variances. Sampling as few as 10% of individuals was typically sufficient to yield more accurate estimates that those obtained through DATABASE lookups. The ONE PER SP strategy (Kraft et al. 2008) yielded good performance in almost all cases (Fig. 1). In diverse ecosystems, however, it could require sampling of half of all individuals (Pitman et al. 2001). We conclude that, despite the commitment of time and money required, there is no substitute for extensive field sampling to accurately characterize the distribution of trait values in species-rich forests.

LITERATURE CITED

Albert, C.H., Thuiller, W., Yoccoz, N.G., Douzet, R., Aubert, S. & Lavorel, S. (2010) A multi-trait approach reveals the structure and the relative importance of intra-vs. interspecific variability in plant traits. *Functional Ecology*, **24**, 1192–1201.

Baraloto, C., Hardy, O.J.O., Paine, C.E.T., Dexter, K.G., Cruaud, C., Dunning, L.T., Gonzalez, M.-A.A., Molino, J.-F., Sabatier, D., Savolainen, V. & Chave, J. (2012) Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *Journal of Ecology*, **100**, 690–701.

Baraloto, C., Paine, C.E.T., Patiño, S., Bonal, D., Hérault, B. & Chave, J. (2010) Functional trait variation and sampling strategies in species-rich plant communities. *Functional Ecology*, **24**, 208–216.

Baraloto, C., Paine, C.E.T., Poorter, L., Beauchene, J., Bonal, D., Domenach, A.-M., Hérault, B., Patiño, S., Roggy, J.-C. & Chave, J. (2010) Decoupled leaf and stem economics in rain forest trees. *Ecology Letters*, **13**, 1338–1347.

Bates, D.M., Maechler, M., Bolker, B.M. & Walker, S. (2014) lme4: Linear mixed-effects models using Eigen and S4.

Chave, J., Coomes, D.A., Jansen, S., Lewis, S.L., Swenson, N.G. & Zanne, A.E. (2009) Towards a worldwide wood economics spectrum. *Ecology letters*, **12**, 351–66.

Conti, G. & Díaz, S. (2013) Plant functional diversity and carbon storage - an empirical test in semi-arid forest ecosystems (S Lavorel, Ed.). *Journal of Ecology*, **101**, 18–28.

Díaz, S., Hodgson, J.G., Thompson, K., Cabido, M., Cornelissen, J.H.C., Jalili, A., Montserrat-Martí, G., Grime, J., Zarrinkamar, F., Asri, Y., Band, S.R., Basconcelo, S., Castro-Díez, P., Funes, G., Hamzehee, B., Khoshnevi, M., Pérez-Harguindeguy, N., Pérez-Rontomé, M.C., Shirvany, A., Vendramini, F., Yazdani, S., Abbas-Azimi, R., Bogaard, A., Boustani, S., Charles, M., Dehghan, M., de Torres-Espuny, L., Falczuk, V., Guerrero-Campo, J., Hynd, A., Jones, G., Kowsary, E., Kazemi-Saeed, F., Maestro-Martínez, M., Romo-Díez, A., Shaw, S., Siavash, B., Villar-Salvador, P. & Zak, M.R. (2004) The plant traits that drive ecosystems: Evidence from three continents. *Journal of Vegetation Science*, **15**, 295.

Díaz, S., Lavorel, S., Bello, F. de, Quétier, F., Grigulis, K. & Robson, T.M. (2007) Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 20684–9.

Fisichelli, N. A., Frelich, L.E. & Reich, P.B. (2014) Temperate tree expansion into adjacent boreal forest patches facilitated by warmer temperatures. *Ecography*, **37**, 152–161.

Fortunel, C., Paine, C.E.T., Fine, P.V. A., Kraft, N.J.B. & Baraloto, C. (2013) Environmental factors predict community functional composition in Amazonian forests. *Journal of Ecology*, **102**, 145–155.

Grime, J. (2006) Trait convergence and trait divergence in herbaceous plant communities: mechanisms and consequences. *Journal of Vegetation Science*, **17**, 255–260.

Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B. & Wright, I.J. (2011) TRY--a global database of plant traits. *Global Change Biology*, **17**, 2905–2935.

Kraft, N.J.B., Valencia, R. & Ackerly, D.D. (2008) Functional traits and niche-based tree community assembly in an Amazonian forest. *Science*, **322**, 580–582.

Lavorel, S. & Garnier, E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, **16**, 545–556.

Lavorel, S., Grigulis, K., McIntyre, S., Williams, N.S.G., Garden, D., Dorrough, J., Berman, S., Quétier, F., Thébault, A. & Bonis, A. (2008) Assessing functional diversity in the field – methodology matters! *Functional Ecology*, **22**, 134–147.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, **21**, 178–185.

Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S.S., Lavorel, S., Mouquet, N., Paine, C.E.T., Renaud, J. & Thuiller, W. (2013) Rare species support vulnerable functions in high-diversity ecosystems (GM Mace, Ed.). *PLoS Biology*, **11**, e1001569.

Nardoto, G.B., Quesada, C.A., Patiño, S., Saiz, G., Baker, T.R., Schwarz, M., Schrodt, F., Feldpausch, T.R., Domingues, T.F., Marimon, B.S. et al. (2014) Basin-wide variations in Amazon forest nitrogen-cycling characteristics as inferred from plant and soil 15N:14N measurements. *Plant Ecology & Diversity*, **7**, 173–187.

Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.

Nehrbass-Ahles, C., Babst, F., Klesse, S., Nötzli, M., Bouriaud, O., Neukom, R., Dobbertin, M. & Frank, D. (2014) The influence of sampling design on tree-ring based quantification of forest growth. *Global Change Biology*. **20**, 2867-2885.

Paine, C.E.T., Baraloto, C., Chave, J. & Hérault, B. (2011) Functional traits of individual trees reveal ecological constraints on community assembly in tropical rain forests. *Oikos*, **120**, 720–727.

Paine CET, Baraloto C, Diaz S (2015) Selected leaf and stem traits from the BRIDGE database. Data from: Optimal strategies for sampling functional traits in species-rich forests. TRY Downloadable Files Archive https://www.try-db.org/TryWeb/Data.php#7

Paine, C.E.T., Norden, N., Chave, J., Forget, P.-M., Fortunel, C., Dexter, K.G. & Baraloto, C. (2012) Phylogenetic density dependence and environmental filtering predict seedling mortality in a tropical forest. *Ecology Letters*, **15**, 34–41.

Pakeman, R.J. & Quested, H.M. (2007) Sampling plant functional traits: What proportion of the species need to be measured? *Applied Vegetation Science*, **10**, 91–96.

Patiño, S., Lloyd, J., Paiva, R., Baker, T.R., Quesada, C.A., Mercado, L.M., Schmerler, J., Schwarz, M., Santos, A. J.B., Aguilar, A., Czimczik, C.I., Gallo, J., Horna, V., Hoyos, E.J., Jiménez, E., Palomino, W., Peacock, J., Peña Cruz, A., Sarmiento, C., Sota, A., Turriago, J.D., Villanueva, B., Vitzthum, P., Alvarez Dávila, E., Arroyo, L.L., Baraloto, C., Bonal, D., Chave, J., Costa, A.C.L., Herrera, R., Higuchi, N., Killeen, T.J., Leal, E., Luizão, F.J., Meir, P., Monteagudo, A., Neill, D.A., Núñez Vargas, P., Peñuela, M.C., Pitman, N.C.A., Priante Filho, N., Prieto, A., Panfil, S.N., Rudas, A., Salamão, R., Silva, N., Silveira, M., Soares deAlmeida, S., Torres-Lezama, A., Vásquez-Martínez, R., Vieira, I.C.G., Malhi, Y. & Phillips, O.L. (2009) Branch xylem density variations across the Amazon Basin. *Biogeosciences*, **6**, 545–568.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S. & Poorter, H. (2013) New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, **61**, 167–234.

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

Poorter, L., Bongers, L. & Bongers, F. (2006) Architecture of 54 moist-forest tree species: traits, trade-offs, and functional groups. *Ecology*, **87**, 1289–301.

R Core Development Team. (2014) R: A Language and Environment for Statistical Computing.

Scheiter, S., Langan, L. & Higgins, S. (2013) Next-generation dynamic global vegetation models: learning from community ecology. *New Phytologist*, **198**, 957–969.

ter Steege, H., Pitman, N.C.A., Phillips, O.L., Chave, J., Sabatier, D., Duque, A., Molino, J.-F., Prévost, M.-F., Spichiger, R., Castellanos, H., von Hildebrand, P., Vásquez, R. & Vásquez-Martínez, R. (2006) Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature*, **443**, 444–447.

ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., et al. (2013) Hyperdominance in the Amazonian Tree Flora. *Science*, **342**, 1243092–1243092.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. & Garnier, E. (2007) Let the concept of trait be functional! *Oikos*, **116**, 882–892.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C.H., Midgley, J.J., Navas, M.-L., Niinemets, Ü., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L.D., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J. & Villar, R. (2004) The worldwide leaf economics spectrum. *Nature*, **428**, 821–827.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

Appendix S1 *R script for all analyses*

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

TABLES

Table 1 Trait sampling strategies investigated in this study.

Type	Name	How are individuals sampled?	Weighting*	Requirement*	Intensity
Abundance-based	BA REGION	Proportional to species basal area in the region	Overweights species that are regionally common, widespread and large-statured	Botanical inventories and d.b.h. from the region	1-50%
	BA LOCAL	Proportional to species basal area in the local plot	Overweights species that are locally common and large-statured	Botanical inventory and d.b.h. from local plot	1-50%

	BA INDIVIDUAL	Proportional to individual basal area	Overweights species with large-statured individuals, regardless of abundance	Botanical inventory and d.b.h. from local plot	1-50%
	COMMON REGION	Proportional to regional abundance	Overweights species that are common and widespread, regardless of stature	Botanical inventories from the region	1-50%
	COMMON LOCAL	Proportional to local abundance	Overweights species that are locally common, regardless of stature or regional abundance	Botanical inventory of local plot	1-50%
	SPECIES	Inversely proportional to local abundance	Overweights species that are locally rare	Botanical inventory of local plot	1-50%
Taxon-free	RANDOM	Randomly	Accurately reflects local species composition	None	1-50%
	SIZE	Stratify individuals by stature. Draw individuals randomly from strata	Overweights species with large individuals	d.b.h. from local plot, but no botanical inventory	1-50%
	CLUMP	Rank individuals by distance to three randomly located foci. Draw individuals based on their proximity to nearest focus.	Accurately reflects local relative abundance. Accounts for geographic variation in species composition	None	1-50%
Special	ONE PER SP	Draw one individual from the plot to represent each species	Overweights species that are locally rare	Botanical inventory of local plot	~33%
	DATABASE	Draw one individual from the region to represent each species	Overweights species that are locally rare	Regional checklist of species	100%
	BRIDGE	Select all individuals	Accurately reflects local species composition	Botanical inventory of local plot	100%

Notes:

Weighting is defined with respect to drawing individuals at random.

REQUIREMENT INDICATES FIELD WORK REQUIRED PRIOR TO TRAIT SAMPLING.

FIGURES

Figure 1

The probability of correctly estimating the community mean or variance in functional trait values increases with increasing sampling intensity and varies among sampling strategies and the trait under consideration. Lines show predictions from linear mixed-effect models, applied to simulation output. Larger values represent less-accurate estimates of the true mean and variance. Note that the X-axis is log-transformed. Results from the ONE PER SP, DATABASE and BRIDGE strategies are shown as points because they are only defined at a single sampling intensity (Table 1). Vertical arrows indicate that the estimation error of DATABASE and ONE PER SP exceed the plotted region for Leaf $\delta^{15}N$.

Figure 2

The per cent variance in the probability of accurately estimating community mean and variance in trait values explained by sampling intensity (I), sampling strategy (S), the sampled trait (T) and their interactions. Note that the per cent variance explained by some interactions is so slight that their bars are scarcely visible.

Figure 1

This article is protected by copyright. All rights reserved.

Figure 2