DR. FREDERICK C DRAPER (Orcid ID : 0000-0001-7568-0838)

MS. XIMENA TAGLE (Orcid ID : 0000-0003-4152-2051)

Article type : Articles

Running head: Tree dominance and beta diversity

Dominant tree species drive beta diversity patterns in Western Amazonia

Frederick C Draper 1,2*, Gregory P. Asner1, Eurídice N. Honorio Coronado 3, Timothy R. Baker4, Roosevelt García-Villacorta 5, Nigel C. A. Pitman 6, Paul V.A. Fine7, Oliver L. Phillips 4, Ricardo Zárate Gómez 3, Carlos A. Amasifuén Guerra 8, Manuel Flores Arévalo 8, Rodolfo Vásquez Martínez 9, Roel J.W Brienen 4, Abel Monteagudo-Mendoza 9,10, Luis A. Torres Montenegro 8, Elvis Valderrama Sandoval 8, Katherine H. Roucoux 11, Fredy R. Ramírez Arévalo 12, Ítalo Mesones Acuy 7, Jhon Del Aguila Pasquel 3,13, Ximena Tagle Casapia 3, Gerardo Flores Llampazo 14, Massiel Corrales Medina 15, José Reyna Huaymacari 8 and Christopher Baraloto 2

*Correspondence: Email: freddie.draper@gmail.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ecy.02636

This article is protected by copyright. All rights reserved.
Abstract

The forests of Western Amazonia are among the most diverse tree communities on Earth, yet this exceptional diversity is distributed highly unevenly within- and among communities. In particular, a small number of dominant species account for the majority of individuals while the large majority of species are locally and regionally extremely scarce. By definition, dominant species contribute little to local species richness (alpha diversity), yet the importance of dominant species in structuring patterns of spatial floristic turnover (beta diversity) has not been investigated. Here, using a network of 207 forest inventory plots, we explore the role of dominant species in determining regional patterns of beta diversity (community-level floristic turnover and distance-decay relationships) across a range of habitat types in northern lowland Peru. Of the 2031 recorded species in our dataset, only 99
of them accounted for 50% of individuals. Using these 99 species it was possible to reconstruct the overall features of regional beta diversity patterns, including the location and dispersion of habitat types in multivariate space, and distance-decay relationships. In fact, our analysis demonstrated that regional patterns of beta diversity were better maintained by the 99 dominant species than by the 1932 others, whether quantified using species abundance data or species presence/absence data. Our results reveal that dominant species are normally common only in a single forest type. Therefore, dominant species play a key role in structuring Western Amazonian tree communities, which in turn has important implications, both practically for designing effective protected areas, and more generally for understanding the determinants of beta diversity patterns.

Key Words
Dominance, beta diversity, Western Amazonia, species turnover, tree species, common species, rare species, tropical forest communities, Loreto, habitat specificity

Introduction
A consistent pattern among all ecological communities is that few species are common and many are rare (Preston 1948, MacArthur 1957). Common species, by virtue of their high abundance, dominate key ecosystem functions such as primary productivity and biomass storage (Grime 1998), and therefore have a vital role in underpinning ecosystem services (Gaston and Fuller 2008, Gaston 2010). The hyper-diverse tree communities of Amazonia are no exception to this ‘few common and many rare’ rule, and a number of recent studies have emphasized that a small number (<5%) of tree species, termed ‘hyperdominants’ or ‘oligarchs’, account for the majority of individuals at large spatial scales (>100,000 km²)
Amazon hyperdominants also dominate some ecosystem functions, for example, ~1% of Amazonian tree species account for 50% of stored biomass and woody productivity (Fauset et al. 2015). Despite our increased understanding of the functional importance of dominant species, little is known of their role in driving spatial patterns of community level turnover in species composition.

It is evident that in highly diverse ecosystems, such as Amazonian forests, it is rare species rather than common species that contribute the most to patterns of local and regional species richness (i.e. alpha and gamma diversity) (Hubbell 2013). The contribution that common species make to patterns of species turnover through space (i.e. beta diversity), on the other hand, is far less clear. In some other ecological systems, such as boreal tree communities and global avian populations, common species have a disproportionately strong influence on patterns of beta diversity (Nekola and White 1999, Gaston et al. 2007). However, the ecological literature presents a conflicting narrative regarding the potential importance of common species in determining patterns of beta diversity in tropical forests. On one hand, species that are common in one habitat type are more likely to be common in other habitat types than rare species (Pitman et al. 2001, 2014), suggesting that dominant species have broader habitat requirements than the average species, and will therefore decrease turnover among habitat types and will be a poor proxy of whole community patterns of beta diversity.

On the other hand, patterns of species turnover in common species may be broadly representative of whole community patterns of species turnover at regional scales (Arellano et al. 2016), and the exclusion of (typically rare) unidentified morphospecies has almost no discernible effect on patterns of beta diversity in Amazonian forests (Pos et al. 2014).

Together these two observations imply that observed beta diversity may be driven more by common species than rare species. Furthermore, at a 50 ha scale, the distance-decay of...
floristic similarity in tree communities is determined primarily by the most common species, with the removal of the most common 10% of species from the analysis having a much greater effect than the removal of the rarest 80% of species (Morlon et al. 2008).

In this paper we aim to investigate the role of dominant species in determining regional-scale patterns of tree beta diversity across a range of habitat types in Loreto, a 37 million ha region covering the Northern Peruvian Amazon basin. As beta diversity has been used to refer to a range of concepts and measurements in ecology (Tuomisto 2010, Anderson et al. 2011), we investigate three specific types of beta diversity pattern: 1) Floristic similarity among plots belonging to different habitat types; 2) Floristic similarity among plots within the same habitat type; and 3) The distance-decay of floristic similarity among plots within the same habitat type, which provides a measure of the turnover in species composition as a function of geographic distance.

Loreto is an ideal region within which to explore these patterns, firstly because it is among the most diverse regions on Earth at both local and regional scales (Gentry 1988, ter Steege et al. 2003, Stropp et al. 2009). Secondly, Loreto harbours a range of habitat types that vary substantially in their composition and diversity (Gentry 1981, Tuomisto et al. 1995, Vormisto et al. 2000, Fine and Kembel 2011, Draper et al. 2018b), which arise as a result of a variable geology and dynamic geomorphology (Räsänen et al. 1990). We assembled a dataset of 207 forest inventory plots, spanning the five main habitat types found in Loreto to address the question: How much do dominant species contribute to patterns of floristic dissimilarity within and among habitat types at a regional scale in Loreto? We aim to answer this question by testing the following hypotheses:
H1a: Dominant species are poor predictors of beta diversity patterns among habitat types because they are common in several habitat types (i.e. habitat generalists).

H1b: Dominant species are good predictors of beta diversity patterns among habitat types when species abundance data are used but not when occurrence data are used, because although they occur in several habitat types they only achieve high abundance in a single habitat type.

H1c: Dominant species are always good predictors of beta diversity patterns among habitat types using both occurrence and abundance data because they occur most frequently in a single habitat type.

H2: Dominant species are poor predictors of distance-decay relationships within habitat types.

An additional important issue concerning this analysis is that the abundance of a species strongly determines its landscape-scale detectability, i.e. common species are likely to be well sampled while rare species are significantly under-sampled. Under-sampling is known to have large impacts on patterns of beta-diversity (Cardoso et al. 2009, Beck et al. 2013). To address this concern, we focus our analysis and interpretation on common species, which are less likely to suffer from sampling issues. However, we also explore the effect of under-sampling, and test the hypothesis (H3) that rare species are a poor predictor of beta diversity because they are so poorly sampled that it is not possible to interpret any spatial patterns.
Methods

Forest inventories

Forest plot inventory data were taken from several contributing sources and, as a result, varied in plot size and sampling protocol used (Table 1). The compiled dataset consists of 207 forest plots covering the major forest habitat types in Loreto: terra firme forest, white-sand forest, seasonally-flooded forest, peatland pole forest and palm swamp forest (Figure 1). A summary of plot sources can be found in Table 1, and further details in Appendix S2. Most data were downloaded from the ForestPlots.net online database (Lopez-Gonzalez et al. 2011).

Standardizing datasets

As plot data came from several different sources and determinations were made by numerous different botanists, it was not possible to standardize our taxonomy across the entire dataset. As a result, individuals that could not be identified to species level were removed from the dataset and excluded from all subsequent analysis. We expect this removal of unidentified morphospecies to have little effect on our analysis of beta diversity. A similar plot-based study has shown strong correlations ($R^2 > 0.98$) between estimates of beta diversity (ecological distances) including and excluding unidentified morphospecies (Pos et al. 2014). As plot size (and therefore the number of stems per plot) varied among our datasets we converted our abundance data into relative abundances (i.e. number of individuals per species/total individuals per plot).

Because some habitat types (e.g. white-sand forests) were over-represented in our plot dataset compared to their distribution in Loreto, it was necessary to standardise our dataset so that our list of dominant species was as representative as possible of the wider Loreto region.
rather than our environmentally biased dataset. To do this we estimated that the different habitat types account for the following proportions of forested land area in Loreto: *Terra firme* forest 60 %, seasonally-flooded forest 20 %, palm swamp 10%, white-sand forest 5 %, peatland pole forest 5 %, based on the best maps of habitat types available in Loreto (Josse et al. 2007, Draper et al. 2014, Palacios et al. 2016, Asner et al. 2017). We then created a correction factor (actual proportion of habitat type area in plot dataset / predicted proportion of habitat type area). The abundance of each species in each plot was then multiplied by the corresponding correction factor. This corrected dataset provided an estimate of population size for each species within Loreto, and was used for all subsequent analysis.

**Identifying dominant species**

Dominant species were defined simply as the minimum number of species that together account for 50 % of all individual trees across the corrected dataset following the definition of ter Steege *et al* (2013). Species that were not identified as dominant have been labelled rare throughout the rest of the manuscript. We did not attempt to distinguish between the two dimensions of dominance, local abundance and regional frequency (Arellano et al. 2014), nor did we attempt to ensure a representative number of dominant species from each habitat type. Our definition was purposefully coarse, as we sought to frame our definition of dominance in the simplest terms.

Our approach to identifying dominant species is dependent on our sample size as it assumes our data were broadly representative of the habitat types in Loreto, and sufficiently large enough to be a reasonable proxy for species composition. To assess how sample size may affect the robustness of our list of dominant species we used a resampling approach. This consisted of repeatedly (1000 times) recalculating the identity of dominant species whilst
excluding either 10 or 50% of plots at random from each habitat type in our dataset. We were then able to calculate the proportion of our list of dominant species that are found in these subsampled datasets.

*Analysing patterns of beta diversity*

In order to summarize patterns of beta diversity among plots both within and among habitat types, we used NMDS ordinations using the *vegan* package in R (Oksanen et al. 2013). NMDS ordinations were produced in two ways: Firstly, species abundance data were used to construct an ecological distance matrix using the Hellinger distance (Legendre and Legendre 2012), which was then used to construct the first NMDS ordination. This ordination provides a way of assessing how similar plots are to one another based on the abundance of tree species. Secondly, species presence/absence data was used to construct a different ecological distance matrix using the Jaccard index, this distance matrix then formed the basis for the second NMDS. This ordination provides a way of assessing how similar plots are to one another based on the occurrence of tree species. Both NMDS ordinations were optimized for two axes and run for 100 iterations or until a convergent solution was reached.

*Quantifying dissimilarity within and among habitat types*

To quantify floristic dissimilarity among habitat types we used the PERMANOVA method (Anderson 2001), which tests the significance of habitat types in determining plot locations in multivariate space. This method tests whether plots in the same habitat type are more floristically similar to one-another than would be expected if the same number of plots were drawn randomly from all plots across all habitat types. The PERMANOVA method was applied using the *adonis* function in the R package *vegan*.

This article is protected by copyright. All rights reserved.
To quantify floristic dissimilarity within habitat types we used the multivariate dispersion method (Anderson 2006) to assess the dispersion of plots within each habitat type in multivariate space. This method tests how floristically similar plots are to one another within each habitat type. The multivariate dispersion method was applied using the betadisper function in the R package vegan. Combined, these methods present a robust approach to identifying location vs dispersion effects in multivariate space (Anderson and Walsh 2013). We then compare how these statistics varied among the three datasets including: i) all species, ii) only dominant species and iii) only rare species.

To test whether the different minimum stem diameter cut-offs used across the plot dataset (ranging from 2 – 10 cm) had a significant impact on our analyses of beta-diversity, we repeated the distance-based ordination analyses using only those stems greater than 10 cm dbh. We found results from these repeat analyses to be remarkably consistent with the original analyses (Appendix S1: Fig S2, Table S2 and Table S3), and therefore these repeat analyses have not been discussed further.

Model based ordinations

The use of distance-based approaches to analyse multivariate abundance data has received substantial criticism, as these approaches are based on the incorrect assumption that there is no mean variance relationship in species abundance data (Warton et al. 2012, 2015b). As a result of this assumption, ordinations based on distance metrics may be confounding patterns of beta diversity with statistical artefacts. In order to test whether the patterns we observed in this dataset were properties of the dataset rather than artefacts of distance based measures, we also used a model-based approach to construct alternative ordination plots (Hui et al. 2015). Specifically, we fit a pure latent variable model (Warton et al. 2015a), using Bayesian
Markov Chain Monte Carlo (MCMC) estimation in the R package *boral* (Hui 2016). Default model parameters were used apart from the fourth hyper parameter, which affects the distribution of default priors, and was reduced from the default of 30 to 20 in order to reduce computational requirements and allow the model to run. Using the model output, posterior latent variable medians for each forest plot can then be plotted in an ordination, where the first two axes represents the two most important axes of floristic variation (Hui 2016).

*Distance-decay analysis*

To quantify the contribution of dominant species to the distance-decay of floristic similarity within forests types at a regional scale, we used binomial generalized linear models (GLMs). Geographic distance was the predictor variable and floristic similarity was the response variable, and a log link function was used following Millar *et al.* (2011). We conducted this analysis using both species abundance (standardised Hellinger distance) and presence-absence data (standardised Jaccard index) for all species, only dominant species and only rare species. The average model fits along with the 95% confidence intervals surrounding these fits have been plotted in order to visualise the distance-decay relationships. This approach provides a measure of how species composition varies within forest types as a function of geographic distance.

*Quantifying habitat specificity*

We used two approaches to assess the habitat specificity of species. Again, we completed this analysis for all species that occurred in five or more plots, judging that species that occurred in fewer than five plots were too rare for their habitat specificity to be assessed. Our first approach was simply to quantify the proportion of individuals that are restricted to a single
habitat type for each species in the habitat type where that species is most abundant. We also ran this analysis once all species abundance values had been normalised to one, i.e. reduced the data set to presence/absence. We did this to ensure that our results were not dominated by single plots where common species have achieved exceptionally high abundance. Our second approach was to use indicator species analysis (Dufrêne and Legendre 1997), to assess which species are significant indicators of a particular habitat type. Indicator species analysis was conducted using the R package labdsv (Roberts 2013).

**Quantifying the effect of under-sampling on patterns of beta diversity**

To estimate the effect of under-sampling on patterns of beta diversity, we used a resampling procedure to purposefully under-sample dominant species. This procedure consisted of sampling a set number (5, 10, 15, or 20) individuals from each dominant species. These individuals were drawn at random from the Loreto-wide population of each dominant species. We then generated a Hellinger distance matrix and NMDS ordinations in the same way as was done with the original species pool.

**Results**

**Identifying dominant species**

From our dataset of 60,000 individual trees made up of 2031 species, we found that just 99 species (~5%; Appendix S1: Table S1) accounted for 50% of individual trees. Because these relatively few species collectively represent half of all trees in our dataset, we call them ‘dominants’, equivalent to the term ‘hyperdominant’ as used by others in the context of Amazon trees (e.g., ter Steege et al. 2013). The dominant species in Loreto come from 29 families, with Arecaceae, Myristicaceae and Fabaceae accounting for more than a third (36) of the 99 species (13, 12, and 11 species respectively).
In our simulation analysis, we found this list of 99 dominant species to be reasonably robust when 10% of the plot data was excluded, although the most abundant dominant species were more stable than the least abundant. For example, all of the 30 most abundant dominant species in our list were found in > 85% of the simulated datasets, while only 24 of the 30 least abundant dominant species occurred in the majority of simulated datasets. The list was far less robust in the simulations where 50% of the plot data were excluded. In these simulations, all of the 30 most abundant dominant species were found in ~ 30% of the simulated datasets. Again, this became more pronounced with the less abundant dominant species; for example, only 15 of the 30 least abundant dominant species occurred in a majority of simulated datasets.

**Floristic similarity among habitat types**

The NMDS ordination figures constructed using all 2031 recorded species from the dataset show clear patterns, with plots clustering based on habitat type. This clustering by habitat type is significant (based on PERMANOVA tests) and consistent in ordinations constructed using both abundance and presence-absence data (Figure 2 panels A and B). When NMDS ordinations were replotted using only the 99 dominant species, we found broadly similar patterns to the ordinations constructed using all 2031 species (Figure 2 panels C and D). Nevertheless, ordinations of dominant species exhibit increased overlap between some habitat types. In particular, a number of seasonally-flooded forest plots appear more similar to *terra firme* plots once only dominant species are analysed. Again, the clustering was significant based on the PERMANOVA test, and habitat types were found to be most distinctive from one another when only common species are included in the analysis ($r^2 = 0.22$; Table 2), suggesting that any visual overlap was not significant.
The ordination plots created using rare species (1932 species) also show broadly similar clustering patterns to ordinations that were constructed using all 2031 species or using only 99 dominant species (Figure 2; panels E and F). However, removing dominant species results in substantial overlap among palm-swamp and seasonally-flooded forest plots in ordinations constructed with both abundance and presence-absence data. This overlap is reflected in the PERMANOVA test which demonstrated that habitat types are least distinct when only rare species are included in the analysis ($r^2 =0.08$, Table 2).

*Floristic similarity within habitat types*

Overall floristic similarity within habitat types increased when only dominant species were included and decreased when dominant species were excluded. The results from the multivariate dispersion analysis suggest that average dispersion of plots within habitat types was consistently lowest when only dominant species were included, and dispersion was consistently highest in the dataset that included only rare species (Table 3).

*Model-based ordinations*

Overall the latent variable model-based ordinations were broadly consistent with the distance based NMDS analyses (Figure 4). Again, the ordination that was created using only the 99 dominant species showed the strongest clustering, however, like in the NMDS analysis there is some overlap in similarity among forests types. In particular, seasonally-flooded forest appears to be floristically poorly defined and overlaps with both *terra firme* and palm swamp habitat types. The latent variable model-based ordination constructed using the 1932 rare species (Figure 4 panel B) showed far more overlap in similarity than in the ordination constructed with only dominant species, especially between seasonally-flooded forest and white-sand forest, as well as palm swamp and white-sand habitat types.
**Distance-decay analysis**

There are clear patterns in the distance-decay relationships within habitat types when all species are considered, both using abundance and presence/absence datasets (Figure 3 panels A and B). Peatland pole forests, palm swamp forests and seasonally-flooded forests all exhibit relatively steep declines in similarity with increasing distance between plots, whereas *terra firme* and white-sand forests are characterized by a much shallower distance-decay pattern. The distance-decay models based on the 99 dominant species are similar to the models based on all 2031 species; however, there is an overall increase in similarity when rare species are excluded (Figure 3; panels C and D). This increase in similarity is especially evident in peatland pole forest and palm swamp habitat types, and when models are constructed using presence/absence data. The distance-decay models based on the 1932 rare species were substantially different from models based on all species. Excluding dominant species led to lower overall similarity among plots within habitat types, as well as less variation in distance-decay rates among models of different habitat types (Figure 3; panels E and F). The gradient of all distance-decays models became more gradual as a result of excluding common species.

**Habitat specificity analysis**

Our analysis of habitat specificity demonstrated that most individuals of most species occurred in a single habitat type (Figure 5 panel B). Importantly, this was true of both dominant and rare species, and there was no significant difference between these groups. The mean proportion of individuals per species that were found in a single habitat type was 78% for dominant species and 83% for rare species. However, the overlapping 95% confidence interval notches on the boxplot indicate there is no significant difference between these two groups (Figure 5 panel B).

This article is protected by copyright. All rights reserved.
Once species abundances had been standardised to unity (i.e. converted to presence/absence), a similar pattern was found: Most species predominantly occur in a single habitat type, and again this was true for both dominant as well as for rare species (Figure 5 panel C). The mean proportion of per-species occurrences found in a single habitat type was 73% for dominant species and 78% for rare species. Again, the overlapping 95% confidence intervals on the boxplot indicate there is no significant difference between these two groups (Figure 5 panel C).

Finally, 317 species were identified as significant indicator species in our dataset ($P \leq 0.05$). Of these significant indicators, 80 were dominant species (Appendix S1: Table S1), which is equivalent to 81% of dominant species. The remaining 237 indicator species represented just 12% of rare species.

The effect of under-sampling

Under-sampling dominant species had a strong effect on observed patterns of beta diversity (Figure 6). Previously observed patterns of among habitat type turnover within dominant species were still detectable (but less distinct) when 20 individuals were sampled per species. However, patterns became increasingly less pronounced when fewer individuals were sampled. When five individuals were sampled per species, there was no evidence of clustering of plots into habitat types (Figure 6).

Discussion

Our results demonstrate that in some of the most diverse tropical forests on Earth, patterns of beta diversity in tree communities are more robust to excluding the rarest 95% of species than to excluding the 5% most common. The patterns of floristic dissimilarity among habitat types
that we observed are broadly consistent using either traditional distance-based multivariate approaches or a Bayesian model-based framework, suggesting that these results are not a product of using distance-based statistics. Furthermore, we found the same beta diversity patterns when analyses were conducted using either species abundance or species presence-absence data. This suggests that our results are not merely a product of dominant species controlling beta diversity patterns due to their exceptionally high abundance in individual plots, but rather that in general, dominant species are strongly associated with particular habitat types and localities.

How can so few species have such an important influence on regional scale patterns of beta diversity? Our analysis of habitat specificity demonstrates that most dominant species occur most frequently in a single habitat type, even though 88% of dominant species are found in more than one habitat type. Interestingly, dominant species in our dataset do not appear to be significantly less habitat specific than rare species, when considering either abundance or presence-absence of species. This pattern of similar habitat specificity persists despite the fact that we would expect rarer species to be more habitat specific by chance simply because they are likely to be under-sampled in our dataset. This finding therefore supports our third hypothesis that dominant species are good predictors of beta diversity patterns among habitat types because they occur most frequently in a single habitat type. These results appear to contrast with previous findings in nearby forests, which have suggested that species that are dominant in one habitat type are more likely to occur in other habitat types (Pitman et al. 2001, 2014). However, only eight species from these previous studies were found to be ‘super oligarchs’, i.e. common across several habitat types, which is consistent with our findings. Furthermore, at the even larger scale of Amazonia, most dominant species are significant indicator species of a single habitat type (ter Steege et al. 2013).
The results from the habitat specificity analysis suggest that dominant species are relatively specialized to a single habitat type, within which they may be functionally optimal. This within-habitat success may be explained by the growth-defense trade-off across edaphic gradients (Fine et al. 2006). This hypothesis states that the combination of impoverished soils and herbivory in low fertility habitats (e.g. white-sand forests) leads to higher investment in defense, while in high fertility habitat types (e.g. *terra firme* forest) species are more likely to invest more heavily in growth. Therefore, this mechanism would suggest that highly successful dominant species are likely to have optimal traits for one habitat, which in turn leads to a trade-off against success in other habitat types. This habitat specific dominance may also allow dominant species to serve as source populations, supplementing neighbouring populations in sub-optimal habitat types, consistent with the concept of source-sink dynamics (Pulliam 1988). Such an interpretation would be consistent with the original oligarchy hypothesis, which predicts that dominant species should only dominate in relatively homogeneous environmental settings (Pitman et al. 2013).

Our results also indicate that the reduced spatial patterns seen in rare species in our dataset results from under-sampling of these species. Our under-sampling simulations approach demonstrates that even when there is a strong underlying pattern, it can be masked if species are represented by fewer than 20 individuals (Figure 6). Given that approximately 75 % of species in our dataset are represented by fewer than 20 individuals, and half are represented by fewer than 6 individuals, under-sampling is likely masking patterns in rare species. Indeed, when we attempted to isolate and examine patterns of beta diversity in the rarest 50% of species, we found no patterns in the data (Appendix S1: Figure S1). Therefore, it is impossible to assign any ecological interpretation to the reduced beta-diversity patterns in rare species. To the best of our knowledge, there are currently few regional datasets in the lowland tropics that surpass ours in terms of number of plots across a range of forest types.

This article is protected by copyright. All rights reserved.
Therefore, it is reasonable to assume that it is not currently possible to describe the spatial patterns of abundance for most tropical forest tree species.

There is substantial overlap between our list of 99 dominant species and published lists of oligarchic species in Western Amazonia (Pitman et al. 2001, 2014) and hyperdominant species in the entire basin (ter Steege et al. 2013). Approximately 40% of species found on our list of Loreto dominants are also listed as oligarchs in upland terra firme or swamp forests (Appendix S1: Table S1) (Pitman et al. 2001, 2014). This overlap is somewhat surprising as Pitman’s studies were conducted in more fertile regions closer to the Andes, did not consider pole forest or white sand forest habitat types, and used different criteria for defining dominance which incorporated both total species abundance and frequency in a given habitat type. There is even greater overlap between our list of dominant species in Loreto and Amazonian hyperdominants, with 47% of species in our Loreto list also occurring on the list of Amazonian hyperdominants (Appendix S1: Table S1). Again, this comes as some surprise as Loreto represents 6% of the area of Amazonia, and many Amazonian hyperdominants do not occur in Loreto. Moreover, our study includes peatland pole forest which is not included in the basin-wide analysis. Peatland pole forests are known to have a distinct composition, and several species (e.g. *Tabebuia insignis* and *Cybianthus spicatus*) would not be dominant if this forest type were excluded.

**Practical implications**

The results presented in this study have important practical implications for understanding patterns of beta diversity in hyper-diverse Amazonian forests using both plot inventory and remote sensing approaches. Our results suggest that it should be possible to use dominant species as a proxy for community-level patterns of beta diversity because dominant species
appear to be broadly representative of overall patterns of beta diversity. This is important because identifying these 99 dominant species is a more straight-forward task than a full tree species inventory, and taxonomic identifications are more likely to be valid for dominant species than for rarer species (Baker et al. 2017), although some dominant species may represent species complexes (Damasco et al. in press). As dominant species are functionally important and include species from most major clades, it may be more informative to focus on these common species than on a taxonomically restrictive list, when resources are limited, as has been suggested elsewhere (Higgins and Ruokolainen 2004, Ruokolainen et al. 2007). Rapid inventories of dominant species may highlight areas that are outliers in multivariate space and therefore may have a particularly distinctive flora. Examples from our dataset are the two plots OLL-03 and OLL-04, which are clear floristic outliers. These plots represent peatland pole forests that receive occasional inundation and have an extremely stunted canopy (~5m) heavily dominated by *Cybianthus spicatus* (Primulaceae) and the swamp specialist *Tabebuia insignis* var. *monophylla* (Bignoniaceae). To the best of our knowledge, these plots have provided unique floristic records of this habitat type in Amazonia, and therefore warrant further investigation.

Our results also have important implications for understanding the spatial ecology of tropical forests through remote sensing technologies. Recent advances using airborne imaging spectroscopy have demonstrated that beta diversity can now be estimated at a landscape scale remotely (Féret and Asner 2014a, 2014b, Draper et al. 2018a). These approaches use unsupervised machine learning to identify c. 30-50 “spectral species”, which are subsequently used as a proxy for species composition. Whilst the results of these studies correlate strongly with plot-based measures of beta diversity, questions remain regarding the legitimacy of using 50 spectral species as proxies for the hundreds of biological species that exist within these landscapes. Our findings demonstrate that 99 features (species) in
multivariate space were sufficient to understand patterns of beta diversity at a regional scale (< 10^5 km²). Therefore, 50 multivariate features (e.g., spectral species) may be sufficient for understanding patterns of beta diversity at smaller landscape scales (> 100 km²) using imaging spectroscopy approaches. Furthermore, because dominant species make up a large fraction of the sunlit canopy, they play a key role in determining the identity of the species that constitute the forest functional classes (FFCs) that underpin national-scale functional beta-diversity (Asner et al. 2017). Therefore, our results suggest that dominant species are crucial to driving the turnover in large-scale functional composition in response to geologic substrate and elevation.

Limitations of this study

An important caveat of our approach is that in order to identify regionally dominant species, it is first necessary to assemble extensive regional floristic inventories of all species abundances across the full range of habitat types. Therefore, the list of dominant species presented in this study applies only to the studied region of Loreto. Key assumptions of this study are that we have adequately sampled the major habitat types in Loreto, and that we have included sufficient plots to identify the true dominant species. The consistency of our list of dominants when 10% of plots are excluded suggests that our list is robust; however, it is difficult to predict how our list of dominants would change if our dataset included a further 100 or 1000 plots. Moreover, our dataset may be missing some dominant species, because although we have comprehensive sampling of the major habitat types in Loreto, we have limited samples from unknown or poorly known habitat types. For example, a type of white-sand forest heavily dominated by the arboreal palm *Mauritia carana* (Arecaceae) has been reported to occupy significant areas of southern Loreto white-sand forests (Fine et al. 2006b, 2007).
Torres Montenegro et al. 2015). Yet *M. carana* is a very rare species in our dataset (8 individuals in total) and is therefore not included in our list of dominants.

**Conclusions**

In this study, we have highlighted the important role of dominant species in determining patterns of beta diversity within and among different habitat types in Western Amazonia. Dominant species are geographically widespread in Loreto, and while they may superficially appear to be widely occurring habitat generalists, our data set shows that they often predominantly occur in single habitat types. Despite the widespread distribution of dominant species, it is possible to understand spatial patterns of beta diversity such as distance-decay relationships, using a heavily restricted dataset, consisting of only dominant species.

We have focused on the wealth of information to be gained by understanding a few dominant species, however, we caution that our findings do not diminish the value of rare species to understanding regional diversity patterns. Instead, this study reveals how little we know of the spatial ecology of most Amazonian tree species. Moreover, in this study, we have focused exclusively on the taxonomic dimension of beta diversity opposed to functional beta diversity. Rare species are known to have distinct trait combinations, especially of vulnerable traits, and therefore contribute disproportionately to functional diversity (Mouillot et al. 2013, Leitão et al. 2016). The distinct functional composition of rare species may lead to rare species having an enhanced role in structuring functional beta diversity in Amazonian forests. Similarly, if two or more rare species occupy similar trait space, they may collectively have an important role in driving patterns of functional beta diversity despite their individually low abundance. We propose that understanding the functional trait profiles of both dominant and rare species should be a future research priority. This understanding will provide new insights
into why some species achieve dominance, as well as revealing the underlying mechanisms that determine species distributions, which in turn govern patterns of beta diversity.

Acknowledgments
This study was supported through a joint project between the Carnegie Institution for Science and the International Center for Tropical Botany at Florida International University. GPA and FCD were supported by a grant from the John D. and Catherine T. MacArthur Foundation and the Leonardo DiCaprio Foundation. Plot installations, fieldwork and botanical identification by the authors and colleagues has been supported by several grants including a NERC PhD studentship to FCD (NE/J50001X/1), an ‘Investissement d’avenir’ grant from the Agence Nationale de la Recherche (CEBA, ref. ANR-10- LABX-25-01), a Gordon and Betty Moore Foundation grant to RAINFOR, the EU’s Seventh Framework Programme (283080, ‘GEOCARBON’) and NERC Grants to OLP (Grants NER/A/S/2000/0053, NE/B503384/1, NE/F005806/1, and a NERC Postdoctoral Fellowship), and a National Geographic Society for supporting forest dynamics research in Amazonian Peru (grant #5472-95). OLP is supported by an ERC Advanced Grant and is a Royal Society-Wolfson Research Merit Award holder.
Literature cited


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


MacArthur, R. H. 1957. On the relative abundance of bird species. Proceedings of the


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


Table 1. Summary of plot inventory data used in this analysis

<table>
<thead>
<tr>
<th>Habitat type</th>
<th><em>Terra firme</em> forest</th>
<th>Seasonally-flooded forest</th>
<th>White-sand forest</th>
<th>Palm swamp forest</th>
<th>Peatland pole forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. plots</td>
<td>96</td>
<td>27</td>
<td>55</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Plot size</td>
<td>0.1-1</td>
<td>0.5-1</td>
<td>0.025-1</td>
<td>0.1-0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Min. diameter</td>
<td>2.5</td>
<td>2.5</td>
<td>2 or 10</td>
<td>10[2]</td>
<td>10[2]</td>
</tr>
<tr>
<td>Total area (ha)</td>
<td>42.75</td>
<td>12.56</td>
<td>18.47</td>
<td>7.76</td>
<td>5.5</td>
</tr>
<tr>
<td>Total stems</td>
<td>33,815</td>
<td>7,006</td>
<td>20,742</td>
<td>4,818</td>
<td>9,427</td>
</tr>
<tr>
<td>Total species</td>
<td>1749</td>
<td>579</td>
<td>603</td>
<td>258</td>
<td>99</td>
</tr>
<tr>
<td>No. dominant species</td>
<td>87</td>
<td>71</td>
<td>54</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>No. indicator species</td>
<td>143</td>
<td>95</td>
<td>41</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

**Reference**
- Pitman et al. 2008; Honorio Coronado et al. 2015; Baraloto et al. 2011
- Draper et al. 2018
Table 2. PERMANOVA test results indicating the significance of habitat type in explaining the location of plots in multivariate space based on either abundance data (Hellinger distance) or presence-absence data (Jaccard distance) for the entire dataset (all species), a subsampled data set using only dominant species and a subsampled dataset using only rare species.

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>$R^2$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abundance data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all species</td>
<td>9.34</td>
<td>0.16</td>
<td>0.001</td>
</tr>
<tr>
<td>dominant species</td>
<td>14.43</td>
<td>0.22</td>
<td>0.001</td>
</tr>
<tr>
<td>rare species</td>
<td>4.5</td>
<td>0.08</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Presence/absence data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all species</td>
<td>7.23</td>
<td>0.13</td>
<td>0.001</td>
</tr>
<tr>
<td>dominant species</td>
<td>15.46</td>
<td>0.23</td>
<td>0.001</td>
</tr>
<tr>
<td>rare species</td>
<td>4.85</td>
<td>0.09</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 3. Multivariate dispersal of plots within habitat types. Distances given are mean Euclidean distances from plots to the median spatial location of each habitat type in NMDS multivariate space.

<table>
<thead>
<tr>
<th></th>
<th>Pole forest</th>
<th>Palm swamp</th>
<th>Terra firme</th>
<th>Seasonally flooded</th>
<th>White sand</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abundance data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all species</td>
<td>0.46</td>
<td>0.46</td>
<td>0.63</td>
<td>0.63</td>
<td>0.59</td>
</tr>
<tr>
<td>dominant species</td>
<td>0.42</td>
<td>0.4</td>
<td>0.59</td>
<td>0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>rare species</td>
<td>0.64</td>
<td>0.65</td>
<td>0.65</td>
<td>0.66</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>Presence/absence data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all species</td>
<td>0.52</td>
<td>0.6</td>
<td>0.64</td>
<td>0.65</td>
<td>0.62</td>
</tr>
<tr>
<td>dominant species</td>
<td>0.41</td>
<td>0.5</td>
<td>0.58</td>
<td>0.56</td>
<td>0.54</td>
</tr>
<tr>
<td>rare species</td>
<td>0.6</td>
<td>0.65</td>
<td>0.65</td>
<td>0.67</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Figure 1. Map of field sites within the department of Loreto, Peru.

Figure 2. NMDS ordinations based on: data using all 2031 recorded species (panels A and B); data using only 99 dominant species (panels C and D); and data using all 1932 rare species (panels E and F). Ordinations in the first column (panels A, C and E) were constructed using species abundance data and a Hellinger distance matrix. Ordinations in the second column (panels B, D and F) were constructed using species presence/absence data and a Jaccard distance matrix. Colours and shapes correspond to different habitat types: Purple triangles (terra firme forest); green diamonds (seasonally-flooded forest); yellow inverted-triangles (white-sand forests); red circles (peatland pole forests) and blue squares (palm swamp forest).
Figure 3. Distance-decay curves among plots within five habitat types. Different rows correspond to different datasets as follows: Panels A and B were made using data from all 2031 recorded species, panels C and D were made using only the 99 dominant species, panels E and F were made using all 1932 rare species. In the left column similarity is the inverse Hellinger distance derived from species abundance data, while in the right column similarity is the Jaccard index derived from species presence/absence data. Lines represent mean GLM model fits, and shading represents 95% confidence intervals of model fits.

Figure 4. Latent variable model-based ordinations using posterior latent variable for each plot if: all species are included in the model (panel A), only 99 dominant species are included in the model (panel B) and only 1932 rare species are included in the model (panel C). Colours and shapes correspond to different habitat types: Purple triangles (terra firme forest); green diamonds (seasonally-flooded forest); yellow inverted-triangles (white-sand forests); red circles (peatland pole forests) and blue squares (palm swamp forest).

Figure 5. Distribution of habitat specificity for dominant and rare species. Habitat specificity corresponds to the proportion of individuals of each species that are restricted to the habitat type where that species is most abundant (panel B). Panel C shows the same measure of habitat specificity when all abundances have been normalised to one (i.e. species occurrences). Horizontal lines indicate median values, red points indicate mean values, boxes indicate the 25th and 75th percentiles, vertical lines indicate 1.5 x the inter-quartile range. Notches in boxes indicate approximate confidence intervals surrounding the median, therefore overlapping notches indicate no significant difference.

Figure 6. NMDS ordinations constructed with Hellinger distance matrices derived from under-sampled datasets of the 99 dominant species. Panels represent different levels of under-sampling: 20 individual per species (A); 15 individuals per species (B); 10 individuals per
species (C) and 5 individuals per species (D). Colours and shapes correspond to different habitat types: Purple triangles (*terra firme* forest); green diamonds (seasonally-flooded forest); yellow inverted-triangles (white-sand forests); red circles (peatland pole forests) and blue squares (palm swamp forest).
This article is protected by copyright. All rights reserved.