



OPEN

Revisiting the hyperdominance of Neotropical tree species under a taxonomic, functional and evolutionary perspective

Gabriel Damasco^{1,2✉}, Christopher Baraloto³, Alberto Vicentini⁴, Douglas C. Daly⁵, Bruce G. Baldwin¹ & Paul V. A. Fine¹

Recent studies have leveraged large datasets from plot-inventory networks to report a phenomenon of hyperdominance in Amazonian tree communities, concluding that few species are common and many are rare. However, taxonomic hypotheses may not be consistent across these large plot networks, potentially masking cryptic diversity and threatened rare taxa. In the current study, we have reviewed one of the most abundant putatively hyperdominant taxa, *Protium heptaphyllum* (Aubl.) Marchand (Burseraceae), long considered to be a taxonomically difficult species complex. Using morphological, genomic, and functional data, we present evidence that *P. heptaphyllum sensu lato* may represent eight separately evolving lineages, each warranting species status. Most of these lineages are geographically restricted, and few if any of them could be considered hyperdominant on their own. In addition, functional trait data are consistent with the hypothesis that trees from each lineage are adapted to distinct soil and climate conditions. Moreover, some of the newly discovered species are rare, with habitats currently experiencing rapid deforestation. We highlight an urgent need to improve sampling and methods for species discovery in order to avoid oversimplified assumptions regarding diversity and rarity in the tropics and the implications for ecosystem functioning and conservation.

Accurately defining species and their distributions is critical to understanding the processes responsible for both the generation and the maintenance of biodiversity. The increasing availability and analysis of species distribution data through forest inventory networks (e.g., GBIF, ATDN, RAINFOR, PPBio, DRYFLOR, 2ndFOR) has resulted in important conclusions about species diversity (e.g.,^{1–3}) and related ecosystem processes (e.g.,^{4–6}). Yet, despite the scientific advances made possible by the analysis of large dataset networks, the diversity of tropical species remains critically understudied, even for well-studied groups such as Amazonian trees^{7,8}.

Estimates regarding the total known species richness of the Amazonian tree flora have been recently debated. The numbers vary from 6727⁹ (based principally on records from regional floras) to 10,071 species¹⁰ (based on plot-based checklists screened for taxonomic validity), but the actual richness is expected to reach over 16,000 species¹¹, given the large number of unsampled areas and the high diversity of extremely rare trees¹. Discrepancies between empirical and estimated values are usually attributed to at least one of three explanations: (i) undersampled regions; (ii) undescribed or undetermined taxa, collected but yet to be identified; and (iii) the presence of species complexes or cryptic lineages for which subtle morphological distinctions may have been recognized but no genetic data have been included to establish whether putative taxa are distinct. Although 1068 new species have been added since ter Steege and collaborators published the first ATDN (Amazon Tree Diversity Network) checklist¹⁰, there have been no systematic efforts to review the taxonomy of hyperdominant taxa (defined as the 227 most common species accounting for more than half of all trees in the Amazon Basin¹), especially those considered taxonomically difficult groups, morphospecies, or species complexes.

Potential species complexes are characterized by at least one of the following criteria (reviewed in¹²): (i) morphologically cryptic variation, (ii) recent evolutionary divergence and genetic introgression, (iii) broad

¹Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA. ²Departamento de Botânica e Zoologia, Universidade Federal do Rio Grande do Norte, Natal, RN 59072-970, Brazil. ³Institute of Environment, Florida International University, Miami, FL 33133, USA. ⁴Programa de Pós-Graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM 69080-971, Brazil. ⁵Institute of Systematic Botany, The New York Botanical Garden, Bronx, NY 10458, USA. ✉email: gabriel.damasco@ufrn.br

geographic range with incomplete sampling, and (iv) taxonomic redundancy (i.e., numerous synonyms). Plant species complexes are fairly common in the Amazon region (e.g.,^{13–15}) and based on the criteria above, we speculate that at least half of the 50 most hyperdominant Amazonian trees¹ may represent species complexes. If some or all of these hyperdominant species are actually species complexes that warrant treatment as multiple species or separately evolving taxa, numerous important consequences would arise. For example, resolution of such complexes would refine predictions about Amazonian species diversity and identify previously unknown and potentially threatened taxa and key habitats as new conservation priorities. In addition, it would enhance the understanding of ecosystem processes and global change if newly described taxa exhibit differential functional responses to variation in abiotic and/or biotic conditions.

Here, we review a putatively dominant and widespread plant species, *Protium heptaphyllum*, by conducting extensive population-level sampling across its geographic range. As circumscribed to date, *P. heptaphyllum* was listed as the 12th most common tree species in Amazonia¹, occurring also in the Cerrado and Atlantic forest of Brazil. *Protium* is one of the taxonomically best-studied genera in the Neotropics (e.g.,^{16–19}) and it was recently listed the second most abundant tree genus in the Amazon Basin¹. We combined morphological, ddRADseq (double digest Restriction site Associate DNA sequencing), and physiological data to answer the following questions: (i) Does this hyperdominant tree represent a single taxon or multiple taxa? (ii) What processes appear to have generated and maintained evolutionary divergence of lineages within *P. heptaphyllum*? (iii) To what extent do functional strategies vary across lineages? The answers to these questions are discussed here in detail as they play an important role in understanding the diversity, ecosystem processes and conservation of Amazonian forests.

Results

Morphological and molecular species delimitation. We find that *Protium heptaphyllum s.l.* represents at least eight different taxa based on both morphological and molecular evidence (Fig. 1, Figure S2, Table S2). The STRUCTURE analysis of genomic data recognized three major clusters that represent a broad pattern of genetic differentiation among major biogeographic regions in South America: (i) Central-Western Amazonia, (ii) Northeastern Amazonia, and (iii) Cerrado and Atlantic Forest domains (Fig. 1). However, a second STRUCTURE run within these major clusters revealed a finer pattern of population and lineage subdivision and identified a total of nine groups that were strongly supported according to the ΔK method and $\log P(X|K)$. RAxML and ExaBayes species trees were compatible with the STRUCTURE results, and all nine genetic clusters are monophyletic. SVDquartet and BPP (A01 analysis) also strongly support the monophyly of nine genetic clusters (posterior probability > 0.99). In addition, the BPP species delimitation analysis supported a nine-species model with posterior probabilities between 0.72–0.75 and an eight-species model with posterior probabilities around 0.17–0.20 regardless of the prior settings used.

Morphological analysis suggests that at least six distinct groups can be resolved based on normal mixture models (NMM) assuming 1 to 10 distinct morphological groups. The NMM results indicated a best-fit model with six distinct morphological groups (Figure S1). Nonetheless, models set for seven and eight species had similar but slightly lower ΔBIC -values. These groups can be distinguished based on several key criteria that have been recognized by the taxonomic specialists in *Protium* to delimit species (Table S1). The lineages clustered by STRUCTURE can be distinguished by both vegetative traits (e.g., plant habit, number of juga, leaflet size and shape, venation, leaflet thickness) and reproductive traits (inflorescence shape, flower density, petal shape and size, petal coloration, anthers and pistil size (See Table S2 and Figure S2 for morphological details and photos).

Based on the morphological disparity and the genetic clustering evidence, we conclude that *P. heptaphyllum s.l.* represents eight evolutionary lineages that warrant treatment as separate species. Although lineages from the Central Brazil Cerrado (light blue, Fig. 1) and the Atlantic Coast (dark blue) are morphologically distinct by key vegetative and reproductive traits, there is evidence for partial genetic admixture among these clades according to the STRUCTURE analysis. In addition, the BPP species delimitation inferred an eight-species scenario with lower but extant posterior probability mean of 0.18. Therefore, we opted for a conservative delimitation decision in which the Cerrado and Atlantic Forest groups are classified as a single taxon due to evidence of admixture and potential geneflow among lineages. Many of these putatively new lineages are geographically restricted and habitat specialists. When overlaying the geographic ranges of these lineages onto the density of *Protium heptaphyllum s.l.* from the Amazon Tree Diversity Network (ATDN, <http://atdn.myspecies.info/>; Fig. 1), it is probable that trees listed as *Protium heptaphyllum* in different areas are likely to be assigned to many of the putative new taxa described in Figure S2 and Table S2.

Demographic history and genetic differentiation. G-PhoCS estimates of divergence time and effective population sizes suggest that *Protium heptaphyllum s.l.* have experienced several distinct trajectories of demographic expansion. *P. heptaphyllum* populations diverged from the outgroup clade ca. 5 Ma after a relative increase in population size (Fig. 2). According to divergence time results, the earliest-diverging *P. heptaphyllum* lineages evolved in the Amazon basin and currently occur in white-sand forests and seasonally flooded habitats. On the other hand, late-diverging lineages diversified very recently, ca. 0.6 Ma, throughout the tropical rain forests and seasonally dry habitats from Central Brazil and the Atlantic Forest. Besides representing multiple distinct species in Amazonia, *P. heptaphyllum s.l.* shows contrasting values of current effective population sizes. In terms of genetic differentiation, F_{ST} and D_{xy} metrics indicate high similarity and potential gene flow among relatively recently diverged lineages from the Cerrado and Atlantic Forest (dark and light blue lineages, Fig. 3). As lineages become less phylogenetically related and spatially distant, F_{ST} and D_{xy} values increase, suggesting high genetic differentiation between Amazonian populations and more recently-diverged lineages from Central Brazil and the Atlantic coast.

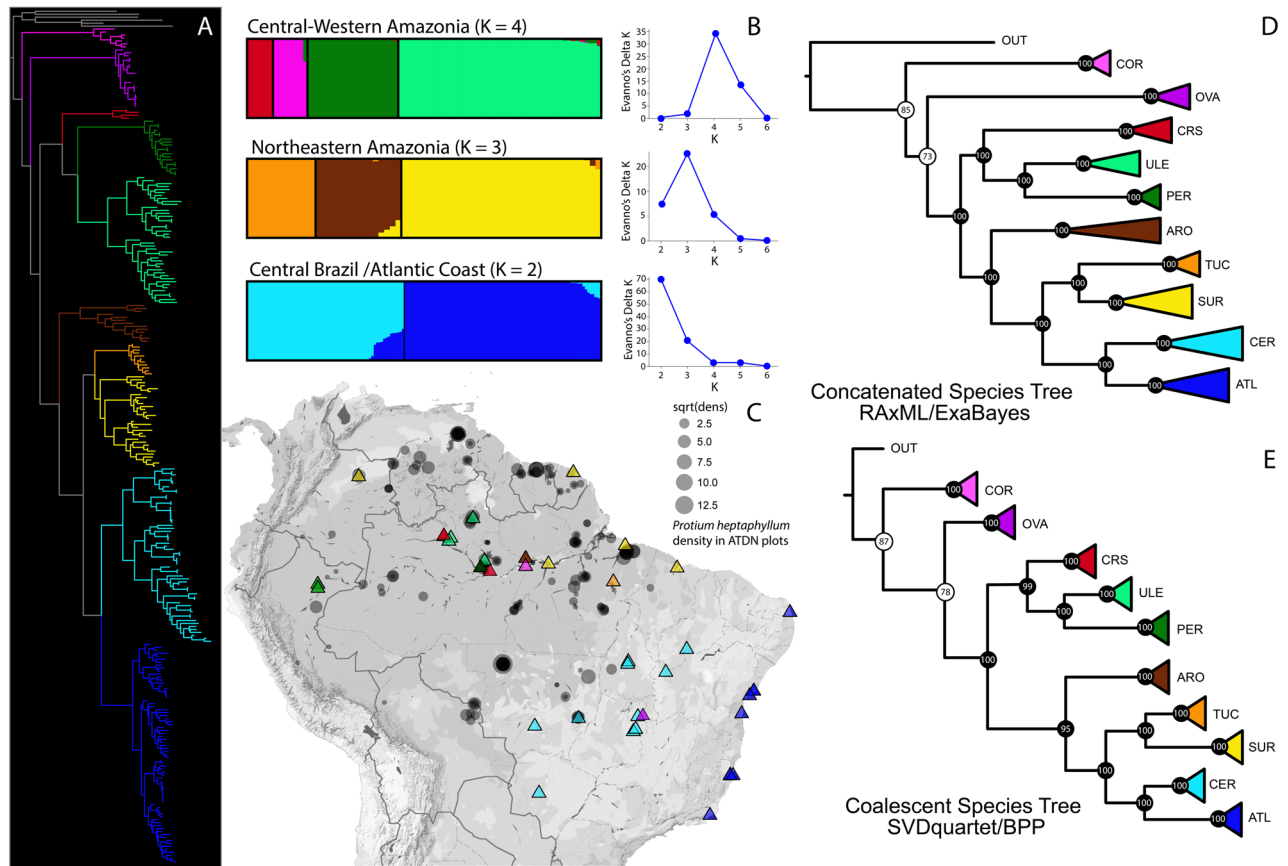


Figure 1. Phylogenetic reconstruction and population genetic structure of *Protium heptaphyllum* s.l. (Aubl.) Marchand. (A) Maximum likelihood (RAxML) phylogeny of *P. heptaphyllum* inferred with 1000 bootstrap replicates using the GTRGAMMA substitution model. The dataset consists of 285 samples and 6,234 loci. Trees were inferred from concatenated data sets using RAxML version 8.2.1060 (<https://github.com/stamatak/standard-RAxML>) and consensus tree visualized using the R package ape version 5.3 (<http://ape-package.ird.fr/>). (B) STRUCTURE analysis identified three main clusters that encompasses the Central/West and North/East of Amazonia, and the Central and Atlantic Coast of Brazil. The bar graphs represent a subsequent STRUCTURE run within each geographic group based on a combination of 10 replicate runs. Colors indicate the assignment of each individual to a particular genetically similar cluster. The number of clusters within each geographic region were identified based on the ΔK statistics. Analysis and figure visualization were performed in STRU CTURE version 2.3.4 (<http://web.stanford.edu/group/pritchardlab/structure.html>). (C) The location of each sampled population is displayed on the map (triangle symbols). Colors indicate the population assignment to a particular genetic cluster. Circle symbols corresponds to the plot location where *P. heptaphyllum* s.l. was sampled by the Amazon Tree Diversity Network (ATDN). Circle sizes are equivalent to the square root of density values per hectare in each of the ATDN plots. Figure generated using the R package ggmap version 3.0.0 (<https://www.rdocumentation.org/packages/ggmap/versions/3.0.0>). (D,E) Species tree inference based on concatenated data matrices ran in RAxML and ExaBayes software and based on the multispecies coalescent model ran in SVDquartet and BPP software, respectively. Node values represent bootstrap support and posterior probabilities for each clade. White circles correspond to support values that are not identical between RAxML bootstrapping and ExaBayes posterior probabilities. In this case, the lower support value was displayed in the figure. Both species trees were visualized using FigTree version 1.1 (<http://tree.bio.ed.ac.uk/software/figtree/>). See Figure S2 and Table S2 for more information about the putative new taxa.

Functional variation and environmental response. We found that *P. heptaphyllum* s.l. exhibited a complex and variable profile of functional trait characters with clear functional differences in mean trait values and trait breadth among lineages identified by the morphological and molecular analyses (Fig. 4). Lineages from the Amazon region occupying white-sand ecosystems and flooded *igapó* forests had very narrow trait variation. For example, the TUC (orange) and CRS (red) populations exhibit very low within-lineage variation in most functional traits (Fig. 4). In contrast, populations from Central Brazil Cerrado (light blue) and Atlantic Forest (dark blue) often showed greater intra-population trait variation corresponding to the different habitats the trees inhabit. For instance, *P. heptaphyllum* populations from the Atlantic Forest were commonly sampled across forest-savanna ecotones (e.g. rainforest and *restinga* transect gradients, populations with code MUS and PON, respectively) and leaf and habit traits in these populations exhibited phenotypic plasticity, corresponding to the local soil fertility and water availability (i.e. stunted trees or shrubs less than 2 m tall inhabiting dry and poor

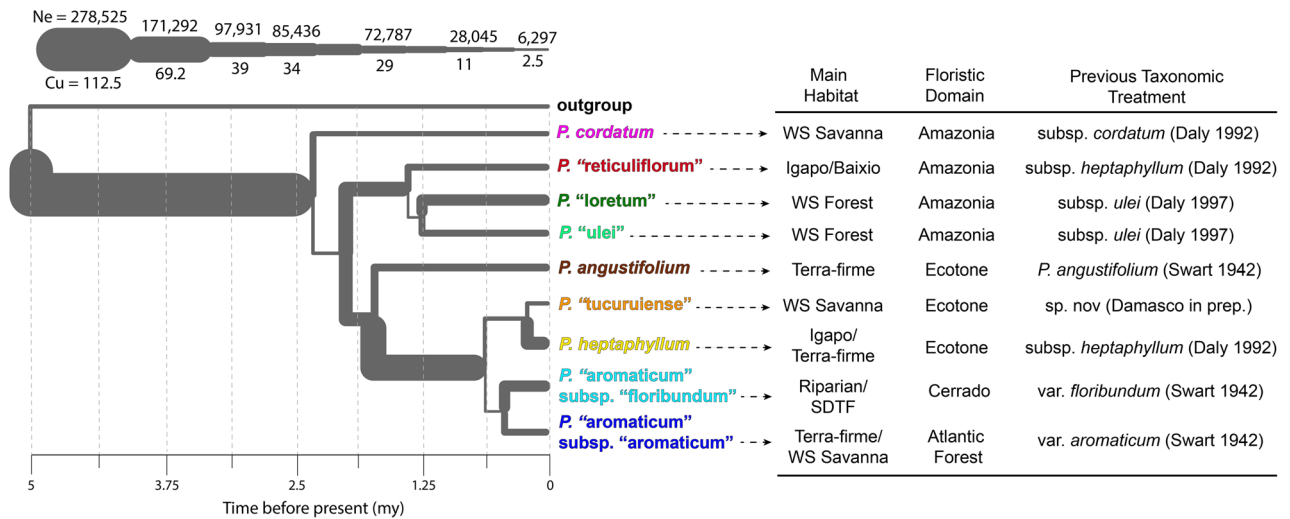


Figure 2. Historical demography and divergence estimate of *Protium heptaphyllum s.l.* populations. Effective population sizes (N_e) based on coalescent units (Cu), and divergence times estimated using the Generalized Phylogenetic Coalescent Sampler (G-PhoCS). The branch ranges (or width) correspond to 95% Bayesian credible intervals aggregated across three runs. Tip labels and colors represent the genetic population assignment based on STRUCTURE analysis. The plant names in quotes are informal, and they will be published in a separate taxonomic review of *P. heptaphyllum s.l.*. On the right, the table displays the habitat type, floristic domain, and the previous taxonomic treatment for each lineage. The floristic domain “Ecotone” corresponds to populations located in transitional areas between the Amazonian floristic domains and neighboring biome areas (e.g. Dry Forests and Grasslands in Colombia). Tree topology generated in FigTree version 1.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) and branch thickness varies accordingly to N_e and Cu values.

soils known as coastal *restingas*, and large 30 m trees occupying nutrient-rich and moister soils in the neighboring forest habitat).

The functional differences we observed among lineages are consistent with differential responses across heterogeneous environments. Overall, functional trait variability among different populations was significantly correlated with local climatic and soil conditions (Table 1). Leaf traits were mostly influenced by the precipitation in the wettest period, and wood traits were well correlated with precipitation deficit in the dry season. In addition, leaf and wood functional traits appear to be directly influenced by local soil conditions, including base saturation, cation exchange capacity and clay content (Table 1).

Discussion

Based on extensive geographic sampling of morphological, ddRADseq and functional trait data, we have identified eight distinct evolutionary lineages within the putative hyperdominant taxon *P. heptaphyllum*. We are using a “traditional” species concept—a typological or morphological species concept where we expect different species to have discrete morphological differences but also represent monophyletic groups with limited genetic admixture. Morphological traits overlap at varying levels, but genetic data suggest that this widespread group represents distinct, independently adapted lineages that diverged over the past million years. We refute the hypothesis that *P. heptaphyllum s.l.* is a single species and instead find evidence for recognizing eight independently evolving species (or lineages, *sensu*²⁰). Updates regarding the taxonomic treatment within *P. heptaphyllum s.l.* are in progress (e.g. *Protium cordatum* Huber *sensu*²¹) and detailed descriptions of new species are in preparation as part of a taxonomic revision (Table S2, Figure S2).

Even though a large number of new species have been recently described in the Neotropics^{9,10,22}, very few studies have attempted to resolve morphologically challenging species complexes (e.g.,^{13,15}). In the sections below, we discuss the implications of our results in the context of: (i) revisiting the concept of hyperdominance for Amazonian trees, (ii) improving richness and diversity estimates, (iii) understanding diversification within dominant tropical lineages, and (iv) refining predictions for ecosystem function.

Implications for the hyperdominance phenomenon. The fact that communities often harbor a small group of demographically abundant species in addition to a much larger number of rare species is not a recent discovery (e.g.,²³ as cited in³). This pattern, also called a species oligarchy and hyperdominance, was first reported for western Amazonian forests in the early 2000s^{24,25} and again on the pan-Amazonian scale based on data from the ATDN, one of the largest tree community datasets ever compiled in the tropics¹. Hyperdominant species have captured the imagination of many tropical ecologists. First, they have been thought to be more likely to be correctly identified than rare species²², allowing for people to use them as proxies for ecosystem-wide function. For example, Fauset et al.⁴ emphasized that hyperdominant species were responsible for half of carbon storage and productivity in the Amazon. Second, ecologists have proposed that hyperdominants have important shared demographic properties—they often have large geographic ranges but are only dominant in one or two regions and are often

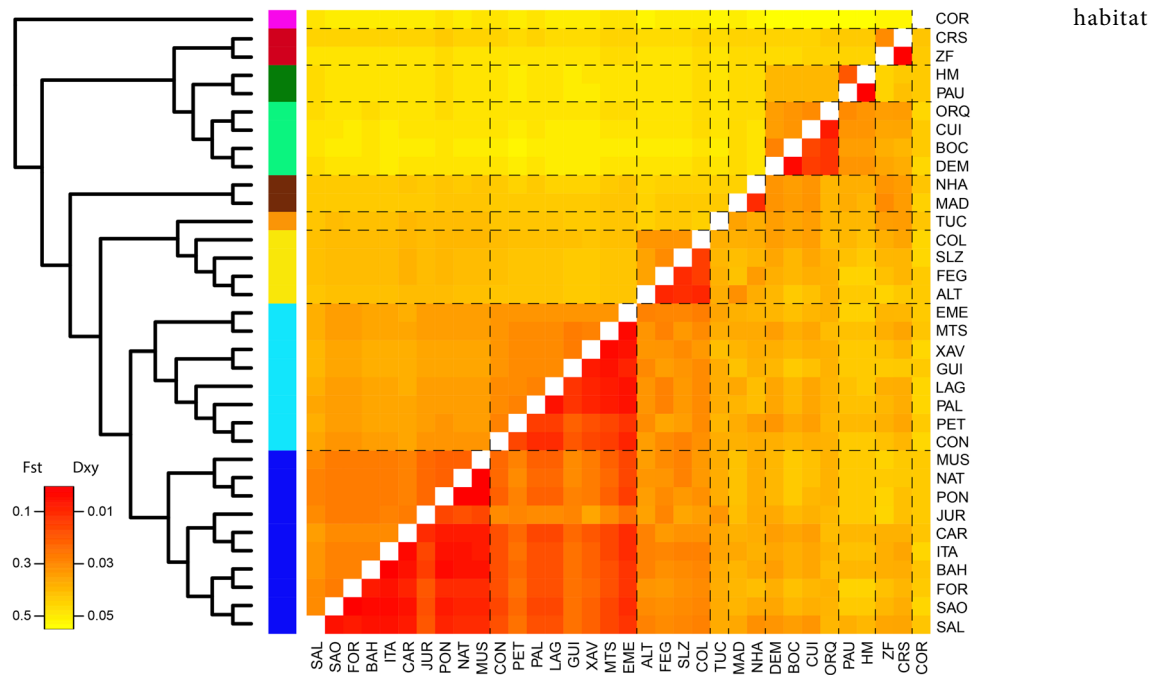


Figure 3. Genetic differentiation among populations of *Protium heptaphyllum s.l.* Pair-wise genetic distances among *P. heptaphyllum s.l.* populations based on two metrics: F_{ST} (lower matrix), a relative measure of genetic differentiation within and between pair-wise populations, and D_{xy} (upper matrix), an absolute measure of genetic divergence between pair-wise populations. Colored bars correspond to populations assignment based on the STRUCTURE analysis. Heatmap generated using heatmap function in RStudio version 1.1.447 (<https://rstudio.com/products/rstudio/>).

specialists¹. In contrast, our results suggest that the hyperdominant taxon *P. heptaphyllum s.l.* actually consists of several lineages warranting recognition as new species that have very different functional traits, that occupy distinct geographic ranges, and that can be rare or threatened. We postulate that similar conclusions could be reached with many of the other hyperdominant species, which are also thought to be members of species complexes (e.g., *Iriartea deltoidea* Ruiz and Pav.²⁶; *Eschweilera coriacea* (DC.) S.A. Mori²⁷). Further study is warranted prioritizing the study of these putative species complexes to refine our understanding of dominance across tropical regions²⁸. If fewer species are found to be true hyperdominants, this will render conservation efforts and ecosystem modeling exercises more complicated than has been discussed to date.

Taxonomic relevance. *Protium* is one of the best studied plant groups in the Neotropics (e.g.,^{17–19}). Currently, the genus consists of approximately 200 species, and their systematics has been studied by a collaborative team of taxonomists and evolutionary biologists. *Protium* has a wide geographic range of specimen sampling, and genomic data are available for a number of species^{18,29,30}. Also, current species descriptions in *Protium* are consistently founded on both morphological and molecular phylogenetic evidence¹⁹. Furthermore, the broad intraspecific sampling of key lineages/taxa within *Protium heptaphyllum s.l.* gives us high confidence in our results, in contrast to other plant groups that have not yet been subjected to similar intensive systematic study but often include diverse and abundant trees in the Amazon Basin. Here, we showed that a multidisciplinary effort and relatively short time investment on a hyperdominant species complex (3–5 years) has yielded the discovery of eight new lineages warranting species status.

Multiple other tree families are relatively well studied in the Neotropics (e.g., Annonaceae, Sapotaceae, Lecythidaceae, Rubiaceae, Chrysobalanaceae, Fabaceae, Melastomataceae), but are also frequently cited as containing a few species complexes with incomplete genetic divergence and requiring detailed revision (e.g., *Eschweilera*²⁷, *Pouteria*³¹). We acknowledge that the field of taxonomy is dynamic (ter Steege et al. 2019) in the sense that classification and species names are likely to change as new studies are performed. Besides that, the methods applied to describe, re-establish, or invalidate taxonomic entities are frequently inconsistent among taxonomists and experts, and can vary according to several species concepts and definitions^{32,33}. Inconsistency in taxonomic methods has been the focus of recent debates about tree species richness in Amazonia^{9,10}. While the authors of these studies disagreed about the number of Amazonian tree species, both sides agree that there is still much work to be done in order to improve taxonomy and to increase the pace of species discovery²². We propose that the establishment of a consensus prioritized list of other species complexes from putative hyperdominant taxa such as *P. heptaphyllum, s.l.* would represent a significant step to contribute to efficient progress of new species discovery. With the advent of cheaper and more rapid techniques for both molecular systematics and functional ecology, we expect that interdisciplinary approaches combined with an extensive populational sampling like

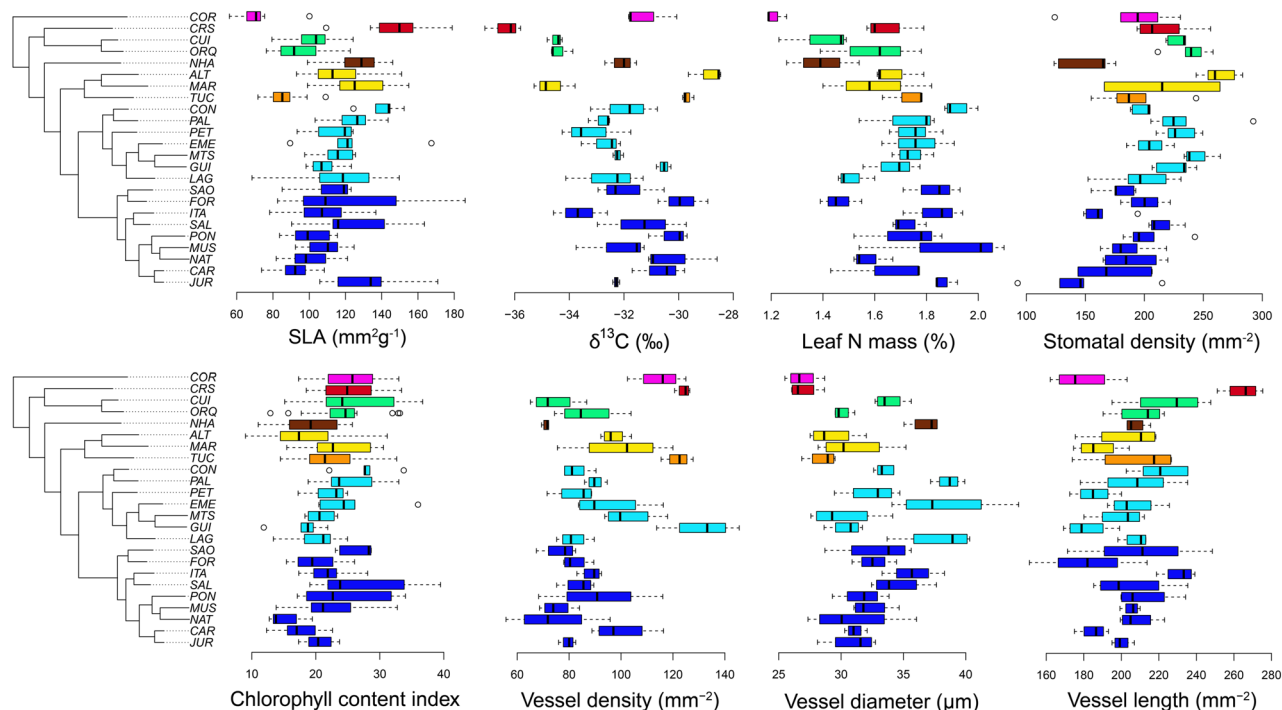


Figure 4. High trait variability among different populations of *Protium heptaphyllum s.l.* (A) Leaf and wood trait variation for multiple populations of *P. heptaphyllum s.l.* and color-coded by the genetic clusters identified by STRUCTURE analysis. The results confirm high variability within each population and among putatively discovered lineages. Functional traits were mapped onto a collapsed phylogeny including 24 geographically widespread populations. See Table S3 for complementary details about individual samples and Fig. 1 to access their geographic location. Boxplot graphs generated in RStudio version 1.1.447 (<https://rstudio.com/products/rstudio/>).

	SLA (mm ² g ⁻¹)	δ ¹³ C (‰)	Leaf N (%)	CCI	Stomatal Dens (mm ⁻²)	Vessel Diam (µm)	Vessel Dens (mm ⁻²)	Vessel length (µm)
Bioclimatic variables								
Mean diurnal range (max. temp.—min. temp)	0.156	0.017	0.212	-0.191	0.150	0.083	0.008	-0.139
Precipitation of driest quarter	-0.206	-0.187	0.000	0.120	0.271	-0.006	-0.174	0.485***
Precipitation coefficient of variation	0.201	-0.029	-0.107	-0.140	-0.063	0.049	0.078	-0.451**
Annual precipitation	-0.300**	-0.369**	-0.319**	-0.028	0.306	0.105	0.146	0.255
Precipitation of driest month	-0.152	-0.211	0.057	0.111	0.261	0.004	-0.209	0.523***
Precipitation of wettest month	-0.286	-0.220	-0.582****	-0.229	0.113	0.011	0.155	-0.135
Soil variables								
Calcium (mmol dm ⁻³)	-0.129	0.364*	-0.136	-0.232	-0.233	0.371*	-0.129	0.204
Magnesium (mmol dm ⁻³)	-0.302*	0.239	0.001	-0.027	0.021	0.304*	0.282	0.273
Carbon (%)	0.252	-0.226	-0.247	-0.469***	-0.492**	0.554****	0.099	0.210
Base saturation (%)	-0.287	0.385**	0.063	-0.127	0.001	0.193	0.084	0.208
Cation exchange capacity (mmol dm ⁻³)	0.197	-0.107	-0.319*	-0.295***	-0.490*	0.229	0.158	0.062
Clay content (%)	-0.201	-0.232	0.137	0.095	0.026	0.611****	0.065	0.373**

Table 1. Leaf and wood functional traits are influenced by climatic and soil conditions. Linear correlation among phylogenetically independent contrasts of leaf and wood functional traits and their respective bioclimatic and soil variables. Functional traits were measured in 24 geographically widespread populations. Bioclimatic values were extracted from the Worldclim dataset and soil features were obtained from chemical analysis. Significant correlations bolded (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$). SLA specific leaf area, δ¹³C carbon stable isotopes, N nitrogen, CCI chlorophyll content index, Dens. density, Diam. diameter.

we have done will be highly beneficial for the study of hyperdominant species complexes and to advance the estimates of Amazonian tree species diversity.

Refining the understanding of trait variation and functional response. The seven distinct lineages for the Central/West Amazon Basin and Northeast Amazonian regions showed substantial variation in leaf and wood traits (Fig. 4). Overall, more than half of the breadth of variation in these functional traits that has been observed across 2600 Amazonian species (e.g.,^{34–36}) can be found within this single species complex.

Moreover, the breadth of functional trait variation also varied markedly within lineages. The low within-lineage trait variation in the Amazonian TUC (orange) population (Fig. 4) is consistent with adaptation to dry savanna environments; that is, this lineage exhibits consistent high-water use efficiency (less negative 13C) and small, dense xylem vessels. In contrast, the CRS (red) population, which also exhibits low within-lineage functional trait variation, displays values of traits consistent with functional responses to flooding in Amazonian *igapó* habitats where it is found, and thus has low water use efficiency and large xylem vessels. Together, these lineages contribute to the broad variation observed in the species complex, while at the same time other lineages from Central Brazil and the Atlantic Coast (light and dark blue) show high within-lineage functional trait variation and broad distribution across habitats. Taken together across the phylogeny of the entire species complex, these functional traits can be considered to be highly labile, with some lineages adapting to contrasting extreme values whereas other sister lineages retain broad functional trait variation (Fig. 4).

Importantly, this variation is consistent with the evolutionary histories of each distinct lineage within the species complex. For example, the *P. heptaphyllum*, *P. “tucuruense”* and *P. “aromaticum”* (yellow, orange and blue) clades experienced a relatively recent population bottleneck and then expanded, likely with traits that permitted them to radiate into the drier savanna environments where they were then able to expand (Fig. 2). Some of these lineages have become adapted to distinct habitats and are responding very differently to current environmental conditions. Moreover, we suggest that these lineages will respond very differently to future changes in climatic conditions across the region, with some of them poorly equipped for the predicted increasing frequency and intensity of droughts³⁷.

Our results have important implications for ecosystem function. Since the advent of functional trait network studies (e.g., TRY Plant Trait Database³⁸), understanding how plant species behave in terms of their physiological performance over large scales has led to important predictions of the consequences of future scenarios of climate and land-use change^{39–41}. The importance of the Amazon region for the global climate and carbon cycle^{42,43} highlights the need to devote substantial effort to investigating the taxonomy of hyperdominant plant taxa. For instance, if some or most of the hyperdominant taxa actually represent multiple hidden evolutionary entities, as in *P. heptaphyllum s.l.*, a larger fraction of Amazonian tree species would contribute proportionally more to carbon storage and cycling than described by Fauset et al.⁴, rendering modeling exercises and management much more nuanced. In addition, a comprehensive taxonomic review of dominant tree lineages with concomitant screening of functional traits as we conducted for this species complex, would greatly improve the understanding of climate-induced functional shifts, such as described by Esquivel-Muelbert et al.⁴¹, and have potentially important consequences for the conservation of putative rare taxa that are nested within hyperdominant species complexes.

Understanding diversification in dominant lineages. The scenario of some hyperdominant or oligarchic taxa representing multiple diverged lineages is intriguing and relevant for understanding the processes of diversification in the Amazonian flora. Based on the demographic history of *P. heptaphyllum s.l.*, older lineages tend to be habitat specialists and less morphologically and functionally variable. In contrast, recently evolved lineages have colonized new areas and frequently experience gene flow across very large geographic distances, and they are morphologically and functionally more variable. We hypothesize that large population sizes may be associated with higher diversification rates due to the process of population expansion followed by radiation into different habitats. Therefore, dominant lineages would have better chances to speciate via habitat specialization and generate large clades than non-dominant lineages. Habitat specialization is considered to have evolved in many tropical plant groups^{44,45} and many different tree genera have become specialized in contrasting environments⁴⁶.

According to our results, *P. heptaphyllum s.l.* diverged from its common ancestors around five million years ago and diversified first in the Amazon region followed by an increase in population size. Many lineages subsequently became specialized to white-sand habitats, seasonally flooded forests in the Rio Negro Basin (*igapó* forests), and floodplain forests (*várzea* forests). However, more recently, lineages dispersed into neighboring floristic domains (e.g. Cerrado and Atlantic Forest) and ecotone areas, colonizing regions where the annual precipitation is currently lower and seasonal. These populations found in Central and Coastal Brazil are genetically very similar to each other despite the large geographical distances among them. Moreover, these populations are functionally more variable than the early-diverging lineages in the Amazon. In contrast, older lineages from the Amazon basin were found to be less morphologically plastic and more isolated in terms of gene flow. White-sand ecosystems in Amazonia are characterized by a patchy and geographically disjunct distribution⁴⁷ that could have inhibited dispersal of habitat-specialist populations, resulting in repeated speciation in white-sand forests in different regions.

Conservation implications. We show that at least four newly discovered lineages, including two “resurrected” species, are geographically restricted, demographically rare and endemic to white-sand vegetation in the Amazon (e.g. *P. cordatum* sensu Damasco et al. 2019a, *P. “tucuruense,”* *P. angustifolium* Swart, and *P. “reticuliflorum”*). Future studies aiming to review potential species complexes among hyperdominant species could indicate other threatened lineages warranting taxonomic recognition as well as new geographic areas for

conservation priority. While many studies have relied on datasets compiled by plot-inventory networks, more effort should be focused on increasing the pace of taxonomic research in the Neotropics^{22,28}. Deforestation rates in Amazonia are likely to increase in the next few years; and yet one third of the total number of tree species are likely still undescribed or undiscovered¹.

We believe that reports that roughly half of all biomass and carbon storage in one of the most diverse areas on Earth is stored by a very small group of “hyperdominant” species may be misinforming policy makers and stakeholders. Based on our integrative review of *P. heptaphyllum* s.l., we showed that genetic diversity and functional responses to environmental gradients are much greater than expected by the hyperdominance principle. Our results revealed that a single “hyperdominant” taxon contains several lineages warranting recognition as distinct species that include great functional diversity, including at least three that are relatively rare and potentially threatened. Metapopulations of a hyperdominant taxon may be interpreted as much more resilient to future global changes than relatively rare lineages within a species complex. Thus, our findings suggest not only critical issues with species level conservation for lineages that should now be considered threatened taxa and not hyper- nor dominant, but also that some of these putatively new lineages might be at risk of extinction due to future climate change and increasing deforestation. Although the distinction between species and population delimitation may not be of consequence for calculating current carbon storage, we argue that multiple distinct lineages with limited gene flow such as we describe here will likely respond very differently to current and future global changes than a larger interbreeding metapopulation of a single lineage. We therefore caution against simplistic assumptions about biogeochemical processes and ecosystem services reliant on the attractive simplicity of a putative hyperdominance phenomenon.

Methods

Field data collection. We conducted multiple field sampling across the geographic range of *Protium heptaphyllum* s.l. in South America (Fig. 1; Table S3). In total, 39 populations of *P. heptaphyllum* were sampled, in addition to ten closely related outgroup species selected based on a molecular phylogeny of the Protieae tribe¹⁸, i.e., *P. unifoliolatum*, *P. trifoliolatum*, *P. kruckhoffii*, *P. pillosum*, *P. widgrenii*, *P. icicariba*, *P. kleinii*, *P. brasiliense*, *P. ovatum*, and *P. dawsonii* (Table S3). Outgroup taxa were sampled in the field when possible; otherwise, herbarium specimens were used. For each sample, we recorded latitude, longitude, and elevation with a GPS device. Leaf material for DNA extraction was dried right after collection using silica gel and subsequently stored in a -20°C freezer. In addition, we performed leaf morphological measurements, extracted DNA, and collected functional trait data on the same plant specimens. Soil texture and fertility was analyzed from each population site. Morphological, ddRAD sequences, and functional trait data are provided in the supporting information. Fertile voucher specimens are in preparation to be deposited at the New York Botanical Garden (NY), the University Herbarium (UC) and *Instituto Nacional de Pesquisas da Amazonia* (INPA). Experimental research on plants, including the collection of plant material, comply with the relevant institutional, national, and international guidelines and legislation. Collections permits were provided by SISBIO (Biodiversity Authorization and Information System) in Brazil (document numbers 37726-3, 44650-1 and 48657-1). Formal identification of the plant material used in this study was performed by Douglas C. Daly, Gabriel Damasco, Paul V. A. Fine and Ricardo O. Perdiz.

Morphological measurements. We generated a character matrix with 59 continuous and 98 discrete traits (supplemental Table S1) to investigate the variability of morphological characters in a multidimensional space. Non-informative characters and missing data were excluded. We used the R package *clustvarsel* v.2.3.3⁴⁸ to reduce the dimensionality of the data by selecting the set of principal components most useful for discrimination without a priori grouping information. Vegetative and reproductive traits were measured on 104 specimens of *P. heptaphyllum* s.l.. To test the hypothesis that *P. heptaphyllum* s.l. represents multiple lineages worthy of taxonomic recognition, we fit the number of morphological clusters using the normal mixture models (NMMs) implemented in the R package *mclust* v.5.0⁴⁹. The Bayesian information criterion (BIC⁵⁰) was used to evaluate the best-fit number of morphological groups according to each NMM⁵¹.

DNA extraction and ddRAD library preparation. We extracted high-quality genomic DNA from 385 samples of *P. heptaphyllum* s.l. widely distributed throughout the Amazon, Atlantic Forest, and Cerrado, and six outgroup species (see Table S3 for details about specimens). DNA was extracted from ca. 100 mg of leaf tissue preserved in silica or from herbarium specimens when silica-dried leaves were not available. Extractions followed an updated version of the DNEasy Plant mini kit protocol (Qiagen, Crawley, U.K.). Double-digest RAD-seq libraries (paired-end length of 180 bp) were prepared for high-throughput sequencing following Peterson et al.⁵² and DNA was digested with *SphI*-HF and *EcoRI*-HF enzymes. DNA libraries were sequenced on five lanes of an Illumina HiSeq 4000 at the University of Berkeley QB3 facility.

ddRAD data analysis. We assembled the ddRAD-seq reads using the software *ipyrad* version 0.6.17⁵³, in order to generate two de novo assemblies with different levels of sample sizes and missing data. First, we used a clustering threshold of sequence similarity set to 0.85 and retained a data set including all loci shared by at least four samples. All other *ipyrad* parameters were set to default. In total, 23,922 loci and 39,770 SNPs were recovered from 385 samples (dataset 1). Although the average sample coverage was high (385 loci per sample), 56% of samples had high levels of missing data. We generated another assembly using a higher cluster threshold of sequence similarity (0.90) and including all loci shared by at least 10 samples. The low-missing-data assembly resulted in 6234 total filtered loci and 25,027 SNPs from 285 samples, and 21% missing data (dataset 2).

Genetic differentiation. To estimate the number of genetic clusters within *P. heptaphyllum s.l.* without a priori assumptions of individual assignments, we used the Bayesian clustering algorithm STRUCTURE⁵⁴ on dataset 1. We ran 10 replicates at each value of K for 500,000 generations with a burn-in of 50,000. $K=3$ was preferred with the ΔK method and had the highest $\log P(X|K)$. Those genetic clusters corresponded with the populations located in three major biogeographic regions: (i) the Amazon biome, (ii) ecotone zones between the Amazon and neighboring biomes, and (iii) a large geographic region encompassing the Cerrado and Atlantic Forest biomes (Fig. 1). In order to identify finer patterns of genetic structure, we ran a separate analysis within the three identified groups using similar parameter settings. We used Structure Harvester version 0.6.94 to compare alternative values of K based on the log probability ($\log P(X|K)$) and the ΔK statistic⁵⁵. Pairwise F_{ST} and D_{XY} were calculated among populations using dataset 1. F_{ST} was implemented based on the Weir and Cockerham equations⁵⁶ found in the `WCfst` function from the `hierfstat` R package⁵⁷. D_{XY} was calculated as described by Nei⁵⁸, with a custom R script slightly modified from the `genet.dist` function in `hierfstat`. Among different lineages, F_{ST} and D_{XY} values ranged from 0.1 to 0.43 and 0.028 to 0.055, respectively. In contrast, within same lineages, F_{ST} and D_{XY} values ranged from 0.08 to 0.19 and 0.021 to 0.032, respectively. D_{XY} measures absolute genetic distance, while F_{ST} measures genetic differentiation among populations relative to within populations⁵⁸. F_{ST} values under 0.19 (or 0.20) are usually interpreted as lacking significant genetic structuring or population subdivision. At the same time, F_{ST} values above 0.3 are seen as moderate population structuring. F_{ST} values above 0.5 are normally considered to be associated with strong population subdivision.

Divergence time and population size estimates. The demographic parameters were implemented in the Generalized Phylogenetic Coalescent Sampler (G-PhoCS version 1.2.3⁵⁹), a Bayesian method for inferring divergence times and effective population sizes from genome sequences. First, we generated a de novo assembly with 1971 loci and 43 samples selected from populations within the 10 different clusters defined by STRUCTURE and a collapsed clade of closely related outgroups. In the MCMC runs, we used the gamma distribution with $\alpha=1.0$ and $\beta=10,000$ for the mutation-scaled population sizes and divergence times, and a gamma distribution with $\alpha=0.002$ and $\beta=0.00001$ for the mutation-scaled migration rates. Each Markov Chain was run for 100,000 burn-in iterations, after which parameter values were sampled for 500,000 iterations every 50 iterations. Convergence was inspected manually for each run. Parameter calibration in the probabilistic model of G-PhoCS is scaled by mutation rate μ . We estimated an average mutation rate of $\mu=7.2 \times 10^{-5}$ mutations per site per generation given by $\tau=T\mu/g$, where τ is the parameter tau, g is the average generation time (in years), and T is the absolute divergence time (in years). We used a divergence time of 5 Ma (95% CI: 3.8–11.1 Ma) for the clade including *P. heptaphyllum* and close relative outgroups based on Fine et al.¹⁸ and the generation time of ten years was based on an estimation of the average reproductive age of many individuals observed in the field. Effective population sizes are given by $\theta=4N_e\mu$, where θ is parameter theta and N_e is the absolute effective population size, and the migration rates are given by $M=m/\mu$, where m is the probability of migration across two given populations.

Phylogenetic reconstruction. To reconstruct phylogenetic relationships among individuals we took multiple approaches. Trees were inferred from concatenated data sets using RAxML version 8.2.10⁶⁰ and ExaBayes version 1.5⁶¹ on dataset 2. For RAxML, maximum likelihood phylogenies were inferred with 1000 bootstraps using the GTRGAMMA substitution model. For ExaBayes analysis, we ran two Metropolis-coupling replicates with four coupled-chains (each with three heated chains) for 1×10^6 MCMC generations, sampled every 500 generations. Estimated sample size (ESS) for all parameters and branch lengths were summarized with the `post-ProcParam` tool and confirmed as sufficiently sampled after analyzing in Tracer v.1.6⁶¹. Finally, we generated a majority-rule consensus tree with the Exabayes 'consense' tool after a 25% burn-in. Each tree was rooted by the clade composed of *P. unifoliolatum*, *P. trifoliolatum*, *P. krukoffii*, *P. brasiliense*, *P. icariba*, and *P. widgrenii*.

Species tree inference and species delimitation. We inferred species trees on dataset 2 using the coalescent-based methods SVDquartet⁶² conducted in PAUP 4a157 (<http://phylosolutions.com/paup-test/>), and BPP version 3.4⁶³. The data were scored with different populations set as partitions using the A01 analysis⁶⁴. For SVDquartet, we used default settings except for the species tree option and conducted 1000 bootstrap replicates. The posterior distribution was independent of different starting species trees and the topology was consistent when comparing results of multiple runs. We used the A10 analysis in BPP version 3.4 to test species delimitation within the *P. heptaphyllum s.l.* group using a fixed species tree previously inferred by SVDquartet and BPP. We assigned equal prior probabilities to all species delimitation models (1 to 10 species) and all runs were sampled every 50 generations for 10,000 samples with a burn-in of 2,000. We evaluated convergence by comparing results of replicate runs. In BPP, we assigned equal probabilities for the rooted trees and set the inverse-gamma priors $\theta \sim \text{IG}(2, 50)$ for all θ s and $\tau \sim \text{IG}(4, 50)$ for the age of the root (τ_0) according to the demographic inference previously performed by G-PhoCS analysis.

Functional traits. In order to understand the physiological responses of distinct *P. heptaphyllum s.l.* populations across different climatic and environmental conditions, we measured several functional characters including leaf, wood, and chemical traits. The leaf traits measured were specific leaf area (SLA), leaf nitrogen content, leaf carbon content, leaf stable carbon isotope (^{13}C) composition, leaf chlorophyll content, leaf stomatal density. The anatomical wood traits were vessel diameter, vessel density, and vessel length. The phylogenetic independent contrast (PIC) was calculated for each functional trait using the R package `ape` version 5.3. Linear correlation between phylogenetically independent contrasts of functional traits were performed to test the relationship with local climatic and soil features. Details about each functional trait is available in Table S4.

Received: 13 July 2020; Accepted: 30 March 2021

Published online: 05 May 2021

References

1. ter Steege H. *et al.* Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092, <https://doi.org/10.1126/science.1243092> (2013).
2. Banda-R, K. *et al.* Plant diversity patterns in neotropical dry forests and their conservation implications. *Science* **353**, 1383–1387 (2016).
3. Draper, F. C. *et al.* Dominant tree species drive beta diversity patterns in Western Amazonia. *Ecology* **100**(4), e02636. <https://doi.org/10.1002/ecy.2636> (2019).
4. Fauset, S. *et al.* Hyperdominance in Amazonian forest carbon cycling. *Nat. Commun.* **6**, 6857.
5. Poorter, L. *et al.* Biomass resilience of Neotropical secondary forests. *Nature* **530**(7589), 211–214 (2016).
6. Rozendaal, D. M. *et al.* Biodiversity recovery of Neotropical secondary forests. *Sci. Adv.* **5**(3), eaa03114, <https://doi.org/10.1126/sciadv.aau3114> (2019).
7. Feeley, K.J. & Silman, M.R.. The data void in modeling current and future distributions of tropical species. *Global Change Biol.* **17**(1), 626–630 (2015).
8. Hopkins, M. J. Modelling the known and unknown plant biodiversity of the Amazon Basin. *J. Biogeogr.* **34**(8), 1400–1411 (2007).
9. Cardoso, D. *et al.* Amazon plant diversity revealed by a taxonomically verified species list. *Proc. Natl. Acad. Sci.* **114**, 10695–10700 (2017).
10. ter Steege H. *et al.* Towards a dynamic list of Amazonian tree species. *Sci. Rep.* **9**, 3501, <https://doi.org/10.1038/s41598-019-40101-y> (2019).
11. ter Steege H. *et al.* The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Sci. Rep.* **6**, 29549, <https://doi.org/10.1038/srep29549> (2016).
12. Bickford, D. *et al.* Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* **22**, 148–155 (2007).
13. Maas, P. J. M. *et al.* Confronting a morphological nightmare: Revision of the Neotropical genus *Guatteria* (Annonaceae). *Blumea-Biodivers. Evolut. Biogeogr. Plants* **60**(1–2), 1–219 (2015).
14. Pennington, R. T. & Lavin, M. The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytol.* **210**(1), 25–37 (2016).
15. Prata, E. M. *et al.* Towards integrative taxonomy in Neotropical botany: Disentangling the *Pagamea guianensis* species complex (Rubiaceae). *Bot. J. Linn. Soc.* **188**(2), 213–231 (2018).
16. Daly, D. C. *et al.* Burseraceae. In *Flowering Plants. Eudicots*. 76–104 (Springer, 2010).
17. Daly, D. C., Fine, P. V. A. & Martínez-Habibe, M. C. Burseraceae: A model for studying the Amazon flora. *Rodriguésia* **63**, 021–030 (2012).
18. Fine, P. V. A., Zapata, F. & Daly, D. C. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). *Evolution* **68**(7), 1988–2004 (2014).
19. Daly, D. C. & Fine, P. V. A. Generic limits re-visited and an updated sectional classification for Protium (tribe Protieae). *Studies in Neotropical Burseraceae XXV. Brittonia* **70**, 418–426 (2018).
20. de Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* **56**(6), 879–886 (2007).
21. Damasco, G., Daly, D. C., Vicentini, A. & Fine, P. V. A. Reestablishment of *Protium cordatum* (Burseraceae) based on integrative taxonomy. *Taxon* **68**(1), 34–46 (2019).
22. Baker, T. R. *et al.* Maximising synergy among tropical plant systematists, ecologists, and evolutionary biologists. *Trends Ecol. Evol.* **32**, 258–267 (2017).
23. Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**(3), 254–283 (1948).
24. Pitman, N. C. *et al.* Dominance and distribution of tree species in upper Amazonian terra firme forests. *Ecology* **82**(8), 2101–2117 (2001).
25. Macía, M. J. & Svenning, J. C. Oligarchic dominance in western Amazonian plant communities. *J. Trop. Ecol.* **21**(6), 613–626 (2005).
26. Henderson, A., Bernal, R., Galeano-Garces, G. *Field Guide to the Palms of the Americas* (Princeton University Press, 1997).
27. Mori, S. A. *et al.* Observations on the phytoecography of the Lecythidaceae clade (Brazil nut family). *Phytoneuron* **30**, 1–85 (2017).
28. Draper, F. C. *et al.* Quantifying tropical plant diversity requires an integrated technological approach. *Trends Ecol. Evol.* **35**(12), 1100–1109 (2020).
29. Afonso, L. F. *et al.* First draft genome of a Brazilian Atlantic rainforest Burseraceae reveals commercially-promising genes involved in terpenic oleoresins synthesis. *BioRxiv*: 467720, <https://doi.org/10.1101/467720> (2018).
30. Damasco, G., Shivakumar, V. S., Misciewicz, T. M., Daly, D. C. & Fine, P. V. A. Leaf transcriptome assembly of *Protium copal* (Burseraceae) and annotation of terpene biosynthetic genes. *Genes* **10**(5), 392 (2019).
31. Vivas, C. V. *et al.* DNA barcoding in Atlantic Forest plants: What is the best marker for Sapotaceae species identification?. *Genet. Mol. Biol.* **37**(4), 662–670 (2014).
32. Zachos, F.E. *Species Concepts in Biology* 801 (Springer, 2016).
33. Garnett, S. T. & Christidis, L. Taxonomy anarchy hampers conservation. *Nat. News* **546**(7656), 25 (2017).
34. Baraloto, C. *et al.* Functional trait variation and sampling strategies in species-rich plant communities. *Funct. Ecol.* **24**(1), 208–216 (2010).
35. Fortunel, C. *et al.* Wood specific gravity and anatomy of branches and roots in 113 Amazonian rainforest tree species across environmental gradients. *New Phytol.* **202**(1), 79–94 (2014).
36. Fortunel, C., Fine, P. V. A. & Baraloto, C. Leaf, stem and root tissue strategies across 758 Neotropical tree species. *Funct. Ecol.* **26**(5), 1153–1161 (2012).
37. Brando, P. M. *et al.* Droughts, wildfires, and forest carbon cycling: A pantropical synthesis. *Annu. Rev. Earth Planet. Sci.* **47**, 555–581 (2019).
38. Kattge, J. *et al.* TRY—A global database of plant traits. *Glob. Change Biol.* **17**(9), 2905–2935 (2011).
39. Laurance, W. F. *et al.* An Amazonian Forest and Its Fragments as a Laboratory of Global Change in Interactions Between Biosphere, Atmosphere and Human Land Use in the Amazon Basin 407–440 (Springer, 2016).
40. Ewers, R. M. *et al.* Predicted trajectories of tree community change in Amazonian rainforest fragments. *Ecography* **40**(1), 26–35 (2017).
41. Esquivel-Muelbert, A. *et al.* Compositional response of Amazon forests to climate change. *Glob. Change Biol.* **25**(1), 39–56 (2019).
42. Saleska, S. R. *et al.* Carbon in Amazon forests: Unexpected seasonal fluxes and disturbance-induced losses. *Science* **302**(5650), 1554–1557 (2003).
43. Malhi, Y. *et al.* Climate change, deforestation, and the fate of the Amazon. *Science* **319**(5860), 169–172 (2008).
44. Fine, P. V. A., Mesones, I. & Coley, P. D. Herbivores promote habitat specialization by trees in Amazonian forests. *Science* **305**(5684), 663–665 (2004).

45. Fine, P. V. A. *et al.* The growth–defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology* **87**(7), S150–S162 (2006).
46. Fine, P. V. A. & Baraloto, C. Habitat endemism in white-sand forests: Insights into the mechanisms of lineage diversification and community assembly of the Neotropical flora. *Biotropica* **48**(1), 24–33 (2016).
47. Adeney, J. M. *et al.* White-sand ecosystems in Amazonia. *Biotropica* **48**, 7–23 (2016).
48. Scrucca, L. & Raftery, A. E. *clustvarsel: A Package Implementing Variable Selection for Model-Based Clustering in R*. arXiv preprint arXiv:1411.0606 (2014).
49. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8**(1), 289 (2016).
50. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
51. Cadena, C. D., Zapata, F. & Jimenez, I. Issues and perspectives in species delimitation using phenotypic data: Atlantean evolution in Darwin's finches. *Syst. Biol.* **67**, 181–194 (2018).
52. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**(5), e37135. <https://doi.org/10.1371/journal.pone.0037135> (2012).
53. Eaton, D. A. R. & Overcast, I. Ipyrad: interactive assembly and analysis of RADseq data sets. *Bioinformatics* **36**(8), 2592–2594 (2016).
54. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**(2), 945–959 (2000).
55. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**(8), 2611–2620 (2005).
56. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
57. Goudet, J., Jombart, T. & Goudet, M. J. Package 'hierfstat'. *R Package Version 0.04-22*. <http://www.r-project.org>, <http://github.com/jgx65/hierfstat> (2015).
58. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.* **76**(10), 5269–5273 (1979).
59. Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**(10), 1031 (2011).
60. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9), 1312–1313 (2014).
61. Aberer, A. J., Kobert, K., Stamatakis, A. ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Mol. Biol. Evolut.* **31**, 2553–2556 (2014).
62. Chifman, J. & Kubatko, L. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**, 3317–3324 (2014).
63. Flouri, T., Jiao, X., Rannala, B. & Yang, Z. Species tree inference with bpp using genomic sequences and the multispecies coalescent. *Mol. Biol. Evol.* **35**(10), 2585–2593 (2018).
64. Yang, Z. The BPP program for species tree estimation and species delimitation. *Curr. Zool.* **61**, 854–865 (2015).

Author contributions

G.D., C.B., P.V.A.F. led the manuscript writing. G.D. performed fieldwork collection, molecular laboratory work and data analysis. B.G.B and D.C.D. mentored the manuscript writing. P.V.A.F., A.V. and D.C.D. provided funding and institutional support. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-88417-y>.

Correspondence and requests for materials should be addressed to G.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021