Instituto Nacional de Pesquisas da Amazônia – INPA
Programa de Pós-Graduação em Ecologia

PADRÕES BIOGEOGRÁFICOS, FUNCIONAIS E EVOLUTIVOS SOB O CONTROLE DE FILTROS GEOQUÍMICOS E CLIMÁTICOS NA AMAZÔNIA

Fernando Oliveira Gouvêa de Figueiredo

Manaus, Amazonas
Maio, 2017

Fernando Oliveira Gouvêa de Figueiredo

PADRÕES BIOGEOGRÁFICOS, FUNCIONAIS E EVOLUTIVOS SOB O CONTROLE DE FILTROS GEOQUÍMICOS E CLIMÁTICOS NA AMAZÔNIA

Orientadora: Dra. Flávia Regina Capellotto Costa (INPA)
Co-Orientador: Dr. Henrik Balslev (Aarhus University)
Co-Orientador: Dr. Finn Borchsenius (Aarhus University)

Tese apresentada ao Instituto Nacional de Pesquisas da Amazônia como parte dos requisitos para obtenção do título de Doutor em Biologia (Ecologia).

Manaus, Amazonas

Maio, 2017

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES

ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA DO INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA

Aos 14 dias do mês de junho do ano de 2017, às 14:00 horas, na Sala de Aula do PPG - ECO, Campus III, INPA/V8. Reuniu-se a Comissão Examinadora de Defesa Pública, composta pelos seguintes membros: o(a) Prof(a). Dr(a). **Fernanda de Pinho Werneck**, do Instituto Nacional de Pesquisas da Amazônia - INPA, o(a) Prof(a). Dr(a). **Bruce Walker Nelson**, do Instituto Nacional de Pesquisas da Amazônia - INPA, o(a) Prof(a). Dr(a). **Carlos Alberto Nobre Quesada**, do Instituto Nacional de Pesquisas da Amazônia - INPA, o(a) Prof(a). Dr(a). **Alberto Vicentini**, do Instituto Nacional de Pesquisas da Amazônia – INPA, o(a) Prof(a). Dr(a). **Ana Luisa Mangabeira Albernaz**, do Museu Paraense Emilio Goeldi – MPEG, tendo como suplentes o(a) Prof(a). Dr(a). Camila Cherem Ribas, do Instituto Nacional de Pesquisas da Amazônia - INPA, e o(a) Prof(a). Dr(a). Juliana Schietti de Almeida, do Instituto Nacional de Pesquisas da Amazônia - INPA, sob a presidência do(a) primeiro(a), a fim de proceder a argüição pública do trabalho de **TESE DE DOUTORADO** de **FERNANDO OLIVEIRA GOUVEA DE FIGUEIREDO**, intitulado **"PADRÕES BIOGEOGRÁFICOS, FUNCIONAIS E EVOLUTIVOS SOB O CONTROLE DE FILTROS GEOQUÍMICOS E CLIMÁTICOS NA AMAZÔNIA"**, orientado pelo(a) Prof(a). Dr(a). Flávia Regina Capellotto Costa, do Instituto Nacional de Pesquisas da Amazônia – INPA, coorientado pelo(a) Prof(a). Dr(a). Finn Borchsenius, da Aarhus Universitet e pelo(a) Prof(a). Dr(a). Henrik Balslev, da Aarhus Universitet.

Após a exposição, o(a) discente foi argüido(a) oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final:

☒ APROVADO(A) ☐ REPROVADO(A)

☒ POR UNANIMIDADE ☐ POR MAIORIA

Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof(a).Dr(a). Fernanda de Pinho Werneck	
Prof(a).Dr(a). Bruce Walker Nelson	
Prof(a).Dr(a). Carlos Alberto Nobre Quesada	Carlos A. Quesada
Prof(a).Dr(a). Alberto Vicentini	Alberto Vicentini
Prof(a).Dr(a). Ana Luisa Mangabeira Albernaz	Ana Luisa M Albernaz
Prof(a).Dr(a). Camila Cherem Ribas	
Prof(a).Dr(a). Juliana Schietti de Almeida	

Albertina Pimentel Lima

Coordenação PPG-ECO/INPA

INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA - INPA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA – PPG ECO
Av. André Araújo, nº 2936, Bairro – Petrópolis, Manaus-AM, CEP: 69.067-375
Site: http://pg.inpa.gov.br e-mail: pgecologia@gmail.com

F475 Figueiredo , Fernando Oliveira Gouvêa de

Padrões biogeográficos, funcionais e evolutivos sob o controle de filtros geoquímicos e climáticos na Amazônia / Fernando Oliveira Gouvêa de Figueiredo . --- Manaus: [s.n.], 2017.
155 f.: il.

Tese (Doutorado) --- INPA, Manaus, 2016.
Orientador: Flávia Regina Capellotto Costa
Coorientador: Henrik Balslev
Área de concentração: Ecologia

1. Clima . 2. Solo . 3. Biodiversidade Amazônica . I. Título.

CDD 551.6

Sinopse

Nesta tese investigamos como fatores edáficos e climáticos determinam padrões biogeográficos, funcionais e evolutivos de plantas amazônicas. Foram usados modelos de distribuição para prever distribuição de espécies, dados de comunidades para entender respostas funcionais da vegetação em função de gradientes edáficos, topográfico e climáticos, e análises filogenéticas para propor modelos evolutivos basaeados em estratégias funcionais, filtros geoquímicos e eventos geológicos e climáticos do passado.

Palavras-chave: 1. Solos. 2. modelos de nicho ecológico. 3. traços funcionais. 4. geologia. 5. florestas tropicais. 6. topografia. 7. mudanças climáticas. 8. diversificação.

Dedico esta tese aos meus queridos pais e irmãos que sempre me apoiaram nesta minha jornada aos trópicos.

AGRADECIMENTOS

Agradeço ao INPA e ao Programa de Pós-Graduação em Ecologia (PG-ECO) por todo apoio oferecido durante meus onze anos em Manaus

Ao CNPq pela concessão da minha bolsa e outros recursos fundamentais para o desenvolvimento do projeto de Doutorado.

A minha orientadora, Flávia Costa, pela amizade e por ter me orientado e apoiado, principalmente nos momentos mais difíceis.

Aos queridos amigos e companheiros da sala 9, Pancho, Flavinha, Cassiano e Isa, pela ótima convivência, amizade e pela troca de ideias.

A todos os colegas da família vegetal, novos e antigos, pela amizade e trocas de ideias científicas. Aos que se desgarraram da família, Thaíse, Cadu, Ká e Pri, pela amizade e apoio nos momentos difíceis.

Aos meus co-orientadores na Dinamarca – Henrik Balslev e Fin Borchsenius – por não terem poupado esforços para me receber na Universidade de Aarhus para realização de meu estágio internacional, e terem feito contribuições importantes nos capítulos desta tese.

Aos amigos Meethe, Wubing, Mide, Timo, Anne, Blas, Vicent, Bob, Peter, Dennis, Peder e todos os colegas do laboratório Ecoinformatics and Biodiversity da Universidade de Aarhus.

Aos amigos latinos americanos que conheci em Aarhus, Mauro Galetti e família, Carol Bello pelas conversas na nossa língua natal e Karina Galleti por elaborar deliciosos jantares.

Aos grandes amigos brasucas, Gabi Zuqim e Gabriel Moulatlet e os amigos finlandeses Hanna e Kalle que me receberam de braços abertos em Turku na Finlândia e pela colaboração que foram essenciais para o desenvolvimento desta tese.

Ao pessoal do Laboratório Temático de Biologia Molecular (LTBM) do INPA que me deram todo apoio para a realização das atividades laboratoriais.

Ao Mario Terra, por sua grande amizade e por tem me orientado pacientemente no LTBM e ter tornado minha vida entre pipetas e reagentes menos complicada.

A Mariana Naomi Saka e Thiago André pela parceria científica e amizade.

Aos pesquisadores Bruce Nelson, Mike Hoppkins, Alberto Vicentini, Marina Anciães a Laszlo Nagy pelas contribuições feitas na aula de qualificação.

Ao Eduardo Prata e Fernanda Werneck que forneceram reagentes essenciais para o desenvolvimento das atividades no LTBM.

Ao professor Bill Magnusson, por descomplicar o ensinamento em estatística e ter me ajudado a evoluir nesse universo mágico dos números.

A coordenadora do curso e amiga Albertina Lima, pelos conselhos e ajuda para enfrentar as questões burocráticas da Pós-Graduação.

Ao Marcos Silveira, Domingos Rodrigues e equipe, por fornecer prontamente os dados dos sítios PPBIO do Acre e Mato Grosso.

A toda equipe do PPBio-Cenban, Cida, Rafa e toda a turma, que sempre foram muito prestativos e que fazem um ótimo trabalho a frente desse importante instituto com foco no conhecimento e valorização da biodiversidade Amazônica.

A Andresa Mello e Giselle Viana, por ajudar a lidar com as questões administrativas do INPA, projetos e doutorado.

Aos grandes amigos Pardal e Saci, pela amizade e apoio nos momentos difíceis, agradeço a toda a turma da pelada, pelos momentos de descontração, a o Nory e seu fiel amigo Nestor, pela amizade e companheirismo durante nossa convivência na casinha do Petros.

A Ju Vieira, pelo seu amor, carinho, paciência e companheirismos nestes anos de convivência, pelos ensinamentos do mundo molecular e da vida cotidiana, obrigado por fazer parte dessa história S2.

Aos meus pais Saulo e Iara, irmãos Lu e Ju, minha cunhada e sobrinhas, Lívia e Maris', que me deram todo apoio nessa minha jornada.

Por fim, agradeço todas as forças evolutivas, determinísticas e estocásticas, que nos trouxeram até aqui.

"HOJE VEJO QUE A CANOA

- ÁGUAS ONDE NAVEGUEI-

É UM BARCO FEITO DE LUZ

DOS SONHOS QUE JÁ SONHEI"

(Celdo Braga & Célio Cruz)"

RESUMO

O objetivo desta tese foi revelar a importância de filtros geoquímicos edáficos e condições climáticas para entender a biogeografia, o funcionamento e a história evolutiva da biodiversidade Amazônica em um contexto de mudanças climáticas. No **primeiro capítulo** comparamos a importância relativa de dados espaciais de solo e dados climáticos, para prever a macro distribuição de plantas na Amazônia. Na larga escala geográfica, assume-se que as restrições eco fisiológicas determinadas pelo clima representam o principal controle dos limites de distribuição das espécies. No entanto, dados de solo representaram os principais descritores da distribuição de espécies no macro escala. Conclui-se que em um cenário de mudanças climáticas, barreiras edáficas deverão restringir a dispersão das espécies e dificultar o rastreamento de condições climáticas favoráveis. No **segundo capítulo**, usamos uma assembleia de plantas herbáceas de sub-bosque florestal (Zingiberales) para testarmos se existe convergência e/ou divergência nos traços funcionais (área foliar específica, altura e tamanho das sementes) das comunidades ao longo de gradientes regionais de solo, topografia e clima. No geral, a composição funcional das comunidades tende a convergir para o que é esperado pela teoria clássica custo-benefício das estratégias funcionais ao longo de gradientes de recursos: habitats com maior disponibilidade de recursos (nutrientes e água) selecionam traços funcionais indicadores de estratégias de crescimento rápido e ciclo de vida curto. No entanto, a composição funcional das comunidades divergiu fortemente ao longo do gradiente climático, sendo impossível prever a composição funcional em regiões secas. Esses resultados sugerem que o funcionamento das florestas amazônicas está intrinsicamente relacionado com características do solo e topografia, mas ainda existem muitas incertezas sobre como as mudanças climáticas irão afetar sua estrutura e funcionamento. No **terceiro capítulo**, testamos um modelo evolutivo baseado em características funcionais intrínsecas de linhagens, relações de nicho e transformações da paisagem amazônica em decorrência do soerguimento dos Andes. As linhagens com estratégia funcional de crescimento rápida apresentaram maiores taxas de especiação, tiveram maior riqueza, originaram-se mais recentemente, estiveram mais associadas com ambientes produtivos, e sua origem e mudanças temporais nas taxas de diversificação estiveram associadas principalmente com eventos geológicos do Mioceno. As linhagens com estratégia funcional de crescimento lento apresentaram o padrão oposto. Estes resultados revelam a importância de características funcionais, relações de nicho e eventos geológicos para compreender a história evolutiva da biota amazônica. Esta tese demonstra claramente que padrões biogeográficos, funcionais e evolutivos na Amazônia estão intimamente relacionados com filtros geoquímicos edáficos. Ignorar os efeitos destes filtros em modelos de mudanças climáticas pode levar a sérios erros de predição. Compreender onde, como e quando estes filtros atuam é um aspecto essencial para a conservação de uma Amazônia em constante mudança.

BIOGROGRAPHIC, FUNCTIONAL AND EVOLUTIONARY PATTERNS UNDER CONTROL OF GEOCHEMICHAL AND CLIMATIC FILTERS IN AMAZONIA.

ABSTRACT

The goal of this study was to reveal the importance of edaphic-geochemical filters and of climate conditions to understand the biogeographic, functional and evolutionary patterns of the Amazonian biodiversity in a scenario of climate change. In the **first chapter** we used species distribution models to compare the relative importance of soil and climate data to predict species ranges of Amazonian plants. Climate has generally been recognized as the main driver of species distribution at large scales. However, we found that soil attributes were the main predictors of large-scale species distribution. The strong control of species ranges by edaphic features might reduce species' abilities to track suitable conditions under climate change. In the **second chapter,** we used an assemblage of understory herbaceous plants (Zingiberales) to predict convergent and divergent functional trait response (specific leaf area, height and seed size) at community level along soil, hydro-topographic and climate gradients. Overall, the functional composition of communities tended to converge following the expectation of classical theory of plant strategy based on fast-slow grow trade-off: highly productive habitats (rich nutrient soil and bottomlands) selected functional traits indicating fast growth strategies. However, the functional composition of communities diverged widely along the broader climatic gradient, being impossible to predict which functional traits prevail in dry regions. These results suggest that functioning of Amazonian forests is tightly linked with geochemical conditions, but many uncertainties remain regarding how climate change will affect the functioning of tropical forests. In the **third chapter**, we tested an evolutionary model based on intrinsic clade functional strategy, niche-based relationships and past geochemical transformations of Amazonian landscapes driven by Andean uplift. Clades with fast growth strategy had higher diversification rates, higher species richness, arose more recently, were more associated with highly productive habitats, and their origin and diversification dynamics were associated with main geological events of the Miocene. Clades with slow growth strategies had the opposite patterns. These results reveal that, to understand the evolutionary history of Amazonian biodiversity, it is essential to take account of niche and functional aspects of species/clades and past geochemical transformations of landscapes driven by paleo-geological events. . In summary, we clearly demonstrate in this thesis that biogeographic, functional and evolutionary patterns in Amazonia are strongly controled by edaphic-geochemical filters. Overlooking the role of theses filters in models of climate change effects on tropical biodiversity may lead to unsatisfactory predictions. Understanding where, how and when these filters act is also essential for designing conservations strategies for an Amazonia under constant change.

Sumário

ARTIGO 1

ARTIGO 2

ARTIGO 3

LISTA DE TABELAS

Artigo 1

Table 1. List of the 42 Amazonian plant species, geographical regions of occurrence and the number of occurrence records (*N*) after removing outliers and applying a spatial filtering procedure with the indicated radius distance (*d*) to reduce sample bias. The geographical regions were classified as Amazonian rain forest (AM), Central American rainforest (CA), Cerrado biome (CB) and Atlantic rain forest (AT).

Table S1. Species occurrence data sources.

Artigo 2

Table 1: Summary results of mixed-effect GAMLSS models for CWD and CWM. The CWD models were fit using the zero adjusted Gamma distribution family and included only association between the mean (μ) parameter and predictor variables. The CWM models were fit using the three parameters *t* family distribution for LMA, the four parameters Box-Cox *t* for height and seed size, and included the association between mean and/or variance (σ) with predictor variables. Best models appear in bold and models where predictor variables were not significant ($P > 0.05$) are not shown. Null models are the first entry for each response variable. Values in brackets denote the scaled β coefficients.

Table S1: List of the 192 species in the ginger order of flowering plants and their functional trait values. LMA denotes leaf mass per unit of area.

LISTA DE FIGURAS

Artigo 1

Figure 1. Performance of CLIM+SOIL, CLIM and SOIL species distribution models of 42 plant species in Amazonia. The ranking of model performance is based on ΔAIC scores, and the number of species within each ranking class is shown in parentheses. The "=" indicates a difference smaller than 2 (no evidence that the model with lower ΔAIC is better) and "<" indicates a difference greater than 2 (evidence that model with lower ΔAIC is better). ΔAIC values for each species separately are presented in Appendix S3.

Figure 2. Contribution to explained variance (EV) of the unique and joint fractions of the soil and climate models for each of 42 Amazonian plant species (species of different life-forms are shown with different symbols). EV_{Joint} represents the EV that is jointly explained by soil and climate variables. Horizontal axis represents the difference between the unique contributions of the soil model (EV_{SOIL}) and climate model (EV_{CLIM}). Species for which the unique contribution of the soil model was higher than the unique contribution of the climate model are found to the right of the vertical midline.

Figure 3. Individual variables ranked by their importance in SDMs of 42 plant species in Amazonia. Importance is defined as the median value across all species of the explained variance (EV) in single-variable models, expressed as a percentage of the EV in the corresponding best model. Variables are coloured according to whether they were used in the CLIM models or in the SOIL models.

Figure 4. Spatial projection of the differences in suitability scores between the SDMs, mapped only in areas defined as unsuitable to species occurrence (averages across 42 plant species). Suitability scores range from 0 (low probability of species occurrence) to 100 (high probability of occurrence). a) The difference in suitability between the best and CLIM models ($\mathrm{S_{best}} - \mathrm{S_{clim}}$). b). The difference in suitability between the the best and SOIL models ($\mathrm{S_{best}} - \mathrm{S_{soil}}$). c) The Difference in suitability between the CLIM and SOIL models ($\mathrm{S_{clim}} - \mathrm{S_{soil}}$). Low values in panels a) and b) denote regions where the CLIM and SOIL models over-predict the suitable conditions relative to the best models, respectively. In panel c) values close to zero indicate areas where the CLIM and SOIL models jointly predict low suitable condition; negative values denote areas where the CLIM models predict low suitable condition, but the SOIL models over-predict and the positive values means the opposite. Note that for 42 species, the best models included both climatic and soil variables, and for the remaining two species, the best models included only soil variables. Maps are in Mollweide equal-area projection and the dashed line denotes the Equator.

Figure S1.1 Map of the 44 environmental layers used in SDMs with boundaries of Amazonian delineated in red. (C) and (S) are variables used in the CLIM and SOIL model, respectively. D, M and Q denote abbreviations for diurnal, month and quarter, respectively.

Figure S2.2 The *Linhom(r)* function computed for for different filtering (fd) distances for *Ischnosiphon martianus*. For this species, the selected filtering distance was 30 km, resulting in 30 presence records (n) free of spatial clustering. (*r*) is the radius distance setting by the algorithm at which *Linhom(r)* was estimated. The black line is the estimated *Linhom(r)* for the observed point process; grey bands represent the 95% bootstrap confidence interval for the observed *Linhom(r)*; the dotted red line is the estimated theoretical *Linhom(r)* assuming an inhomogeneous Poisson process.

Figure S3.3 Spatial projections of models for the 42 Amazonian plant species. From left to right, the first, second and third column panels denotes the spatial predictions of the relative occurrence probabilities (cumulative ROP of average predictions from the 10 fold cross-validation) of CLIM+SOIL model (CS), CLIM model (C) and SOIL (S) model, respectively. High values of cumulative ROP represent high probability of species occurrence (high suitability scores) and low values denote low probability of occurrence (low suitability scores). Model performance based on ΔAIC and median AUC are shown above the panels. The forth column panel shows the predicted distributional areas (orange areas) obtained by the best model (the CLIM+SOIL for the most of species) and areas in blue denotes the unsuitable climatic or/and soil conditions for species occurrence. Black crosses are the filtered presence records used to calibrate the SDM. The classification of distributional areas was based on setting the threshold on cumulative ROP that maximizes the sum of sensitivity plus specificity. Values above the forth column panels are the selected threshold (Th), omission rate (OR) and the fraction of predicted area (FPA) resulting from predictions' classification. The geographical background (area of predictions) was set in accordance to criteria defined in Appendix S2. Species are grouped by life-forms: trees, palms, lianas, monocot herbs and ferns.

Figure S3.4 Predicted species response curves (relative occurrence probabilities, ROPs) along the 44 environmental variables. Species response curves were built using the contrasts plots from R package 'visreg' (http://myweb.uiowa.edu /pbreheny/publications/visreg.pdf). ΔROP was obtained shifiting the *Xj* values by a reference value, $\bar{x}_j$, setting all species' ROP into a comparable scale (between 0 – 1). Lines in each of the 44 plots represent the response curves of each of the 42 species. Dashed black lines indicate models that were worse than the null model, based on ΔAIC.

Artigo 2

Figure 1. Life-forms in some genera of Zingiberales.

Figure 2. The study area and location of 451 plots where ginger community and environmental data were obtained. Background map represents the dry season length. Legend denotes the maximum number of consecutive months with less than 100 mm of precipitation along the 1998 -2004 period. Black triangles represent the RAPELD plots (250 m x 2m) and blue triangles represent plots with size of 100 m x 5m located along the Juruá river.

Figure 3. Correlation (Person's *r*) between functional traits based on species level (A – C) and community level (D – I). Correlation between traits at species level were obtained applying a logarithm transformation and vertical plot's axis are presented using that scale.

Figure 4. Changes in functional composition (CWM) of 451 herbaceous communities along environmental gradients. Continuous and dashed lines represent the mean (μ) and variance (σ) response of CWM, respectively. Predicted mean is the 0.5 quantile and predicted variance denotes the interval between the 0.25 and 0.75 quantiles of the *t* family distribution (A - C) and Box-Cox *t* distribution (D - L). For each variable the predicted mean mean and variance were obtained keeping all other variables constant at their mean values. The μ (mean) and/or σ (variance) displayed in the upper-right corner of panels indicates significant association between parameters and predictor variables and 'ns' denotes non significant association. Vertical axis from E to L are in logarithm scale. Dots denote the partial residual and light grey crosses represent the original values.

Artigo 3

Figure 1. Time-calibrated phylogeny base on nuclear ITS marker (A) and intrinsic clade factors across the seven Amazonian lineages of Marantaceae (B). Clades in (A) are coloured according with their diversifications rates (r) exhibited in B and grey represent non Amazonian lineages lineages, which some of them are collapsed in a unique node (see the complete phylogeny in Fig. S2). In (B) are the intrinsic clade LMA (mean values in gm^{-2}), diversification rates (*r,* in events per Mya)*,* stem ages (A_{st}, in Mya), crown ages (A_{cr}, in Mya), clade richness (R) and sampling fraction (sf) that means the proportion of extant taxa represented in the phylogeny. The diversification rates were estimated fitting a pure birth process, which was the best model compared with models assuming birth-death process (see detail in Methods).

Figure 2: Relationships between diversification rates, LMA, clade species richness and clade age. In A is the simple relation between diversification rates and LMA. In B – C are the partial regression of the model including LMA and stem age as predictor variables, and in D – E denote the partial regression the model including diversification rates and LMA as predictors. The fitted line in B – E was obtained using conditional plots implemented in *visreg* R-package (25). The *P* values represent the statistical significance of relations, R^2 the good-of-fitness of models including all variables and R^2_{out} de good-of-fitness of models without that variable.

Figure 3: Habitat association of the seven Marantacea clades. Vertical axis denotes the occurrence probabilities scaled to 0 – 1 using contrasting plot implemented in *visreg* R-package (25). Horizontal axis denotes the logarithm of sum of bases Ca, Mg plus K (SB), logarithm of vertical distance from drainage (VDD) and the dry season length (DSL). From top to bottom panels, clades are order based on their LMA values and colours are based on their stem age, which red represent the youngest blue the oldest clade. The statistical significance of clade-habitat association is given by the *P* values extracted from a GLM model using all the tree variables.

Figure 4. Diversification dynamics of the fast-growth and slow-growth clades. Lines represent the posterior mean and polygons the 95% credible interval for diversification, speciation and extinction rates. Fast-growth clades include Breviscapus, Comosae and Microcephalum and slow-growth clades include Ischnosiphon, Monotagma, Ornata and Straminea. Geological events (red triangles and grey polygons) were obtained in Hoorn *et al.* (5). Paleo climate spatial trends is based on data from Zachos *et al.* (26), assessed in ftp://ftp.ncdc.noaa.gov/pub/data/paleo/. ΔT °C is the approximate difference from present-day temperature scaled according to concentration of isotopes of oxygen (δ O^{18}) (see details in figure 2 of Zachos *et al.* (26)). Paleo climate points are coloured representing the gradient running from warm conditions (red) to cooler conditons (blue).

Figure S1. Bayesian phylogenetic tree for 151 taxa in Marantaceae family based on *matK* sequences. The five major clades are according to Prince & Kress (44) and four clades in grey are those with centre of diversity outside Amazonia. Grey bars denote the 95% confidence interval of divergence times.

Figure. S2. Bayesian phylogenetic tree for155 taxa in Calathea clade baed on ITS sequences. Clades in grey are those with centre of diversity outside lowlands Amazonian rain forest. Grey bars denote the 95% confidence interval of divergence times.

Figure S3. The study area and location of 451 plots where ginger community and environmental data were obtained. Background map represents the dry season length. Legend denotes the maximum number of consecutive months with less than 100 mm of precipitation along the 1998 - 2004 period. Black triangles represent the clusters of RAPELD plots (250 m x 2m) and blue triangles represent plots with size of 100 m x 5m located along the Juruá River. Clusters may have 5 to 57 plots spaced by at least 1 km.

Fig S4. Density distribution of LMA values for the seven Amazoninan clades in Marantaceae. Red lines donotes the fitted distribution of the three parameter *t* family distribution adjusted to the mean LMA values for each species.

INTRODUÇÃO GERAL

As florestas tropicais abrigam a maior parte da biodiversidade terrestre do planeta e desempenham um importante papel na regulação climática e manutenção dos serviços ecossistêmicos. Dentre as florestas tropicais encontradas nos diferentes continentes, a floresta amazônica representa a região com a maior cobertura florestal e ainda com um bom status de conservação. No entanto, esse imenso tapete verde e sua biodiversidade vêm sendo ameaçadas pelo avanço do desmatamento, da fronteira agrícola, pela instalação de grandes empreendimentos hidro energéticos e por mudanças climáticas (Nepstad *et al.* 2002; Fearnside 2006; Malhi *et al.* 2008; Nobre 2014). De maneira geral, essas pressões afetam sua intricada e complexa rede de interações ecológicas, responsáveis por manter o próprio sistema em funcionamento, assim como afetam sua capacidade de continuar gerando processos vitais e serviços ecossistêmicos que vão além dos seus limites geográficos (Lewis *et al.* 2011; Nobre 2014). Portanto, é urgente e necessário entender quais os fatores que controlam a distribuição de sua biodiversidade, regulam seu funcionamento e como este sistema diverso e complexo evoluiu.

Esse imenso tapete verde, cobiçado pelos conquistadores e admirado pelos naturalistas europeus que por aqui chegaram à era pós-colombiana, tem se revelado cada vez menos homogêneo. Talvez, uma das primeiras propostas de bioregionalização da Amazônia seja a feita por Wallace (1852) após suas expedições pelos principais rios amazônicos no século XIX. Segundo Wallace, certas espécies de plantas e animais só ocorriam em um dos lados dos grandes rios e, portanto, eles definiriam os limites de regiões biogeográficas distintas. Mais tarde, essa e outras observações feitas pelo naturalista ao redor do globo sustentaram a base primordial do conceito de especiação alopatria por vicariância, ou seja, barreiras físicas teriam isolado espacialmente as populações e o fluxo gênico entre elas, dando origem a novas espécies (Wallace, 1876).

Embora o efeito de rios como barreiras tem se verificado em padrões de distribuição e processos evolutivos em tetrápodes (Hayes & Sewlal 2004; Ribas *et al.* 2012), esse modelo parece não se aplicar no caso de plantas (Hughes *et al.* 2013). Na década de 70, Fittkau et al. (1975) propuseram a subdivisão ecológica da Amazônia baseada em macrorregiões definidas por suas características geológicas, que por sua vez, determinavam tipos de solo e a produtividade dos sistemas terrestres e aquáticos. Para os autores, os padrões de distribuição de espécies de plantas estavam condicionados às diferentes macrorregiões geoedáficas e, portanto, os grandes domínios vegetacionais estariam sendo definidos pelas características químicas e físicas do substrato amazônico. Como o clima (precipitação e temperatura) são relativamente homogêneos na maior parte do seu território, a geoquímica dos solos representaria a grande força ecológica seletiva ao longo das terras baixas amazônicas (Fittkau *et al.* 1975). Essas idéias influenciaram novas e mais complexas subdivisões ecológicas da Amazônia (Sombroek 2000) e a extensiva coleta de dados bióticos e abióticos que vem acontecendo nas últimas décadas têm confirmado a existência de uma grande heterogeneidade geológica e edáfica (Brasil 1978; Quesada *et al.* 2010) e reforçado a importância de filtros geoquímicos para a estruturação espacial da diversidade e composição de plantas na Amazônia (Tuomisto, Ruokolainen & Yli-Halla 2003; Tuomisto *et al.* 2016; ter Steege *et al.* 2006; Higgins *et al.* 2011; Figueiredo *et al.* 2014; Zuquim *et al.* 2014).

A noção de homogeneidade climática na Amazônia e seu desprezível fator seletivo proposto por Fitttkau *et al.* (1975) representam de certa forma uma visão simplista e subjetiva. Na região oeste, central e nordeste do bioma predominam climas com alta pluviosidade anual e com curtas estações secas enquanto que no Leste e nas periferias sul e norte predominam climas mais secos (Fig S1.1 no Material Suplementar – Cap. 1). A riqueza de espécies tende a aumentar em áreas mais úmidas (Gentry 1988; Clinebell *et al.* 1995; Steege *et al.* 2003; Stropp, Ter Steege & Malhi 2009), refletindo o fato de que muitas espécies são sensíveis à seca (Esquivel-Muelbert *et al.* 2016) e a composição florística varia de forma significativa em regiões onde o gradiente climático muda mais rapidamente (ter Steege *et al.* 2006; Toledo *et al.* 2011b). Estes resultados demonstram uma clara influência do clima determinando padrões de diversidade e composição na Amazônia.

Essa relação entre clima e diversidade sugere que a Amazônia não esteve e provavelmente não estará imune a oscilações climáticas do passado e do futuro. De fato, mudanças climáticas recentes já estão afetando as taxas de mortalidade de árvores e o ciclo do carbono na região (Lewis *et al.* 2011). Estudos paleo ecológicos mostram evidências de mudança de vegetação em resposta a oscilações climáticas passadas, como o aumento de diversificação em épocas mais quentes (Jaramillo & Cárdenas 2013) e aumento de extinção em épocas mais secas (Van Der Hammen & Hooghiemstra 2000). Simulações climáticas projetam uma intensificação de condições secas, principalmente no sul da Amazônia (Boisier *et al.* 2015) e consequentemente uma redução na diversidade de espécies (Feeley *et al.* 2012). Diante de um cenário de mudanças climáticas, três efeitos são esperados: (1) as espécies se adaptam às novas condições e sobrevivem; (2) as espécies migram rastreando condições climáticas adequadas ou (3) as espécies nem migram nem se adaptam e, portanto, se extinguem. Assumindo um cenário de baixa adaptabilidade, a capacidade de migração das espécies depende de processos de dispersão, que por sua vez depende de características das espécies (modos de dispersão), e características espaciais do habitat, como conectividade e grau de resistência à locomoção dos indivíduos (Opdam & Wascher 2004). Portanto, condições climáticas e não climáticas devem ser levadas em conta em modelos de projeção de impactos de mudanças climáticas sobre a distribuição espécies.

Modelos espaciais correlativos tem se revelado uma importante ferramenta para explorar potenciais efeitos de mudanças climáticas nas distribuições de espécies (Pearson & Dawson 2003). No entanto, para que estes modelos sejam efetivos, dois pressupostos básicos são necessários: (1) que as espécies estejam em equilíbrio com condições climáticas, ou seja, seus pontos de ocorrência e ausência estejam em áreas climáticas favoráveis e desfavoráveis, respectivamente, permitindo estimativas mais realistas de suas tolerâncias climáticas (nicho climático); (2) que as espécies sejam capazes de dispersar e rastrear condições climáticas favoráveis (Araujo & Peterson 2015). Sob estas condições, consegue-se projetar suas distribuições em cenários de mudanças climáticas. No entanto, como já discutido anteriormente, os limites de distribuição de espécies não são apenas determinados por condições climáticas, sendo talvez os filtros edáficos os principais limitantes das distribuições atuais das espécies de plantas amazônicas. Variáveis edáficas são raramente utilizadas em modelos espaciais de larga escala, pois se assume que seus efeitos são restritos a escalas locais (1 – 10 km) (Pearson & Dawson 2003). **No primeiro capítulo desta tese**, utilizamos modelos espaciais correlativos para avaliar quais seriam os melhores modelos – climáticos, edáficos ou mistos – para mapear a distribuição de 42 espécies de plantas amazônicas (incluindo árvores, palmeiras, lianas, herbáceas e samambaias). Através da sobreposição dos modelos espaciais das 42 espécies, mapeamos macrorregiões onde clima ou solo, ou ambos os fatores restringem os limites de distribuição de espécies.

Os filtros ambientais, como solo e clima, determinam a distribuição da diversidade e sua evolução através de restrições na funcionalidade de indivíduos, populações, espécies, comunidades e ecossistemas. A abundância ou escassez de nutrientes no solo influenciam diretamente nos processos fisiológicos dos indivíduos, que se refletem nos processos demográficos (taxas de nascimento, mortalidade, crescimento e dispersão), na estrutura funcional de comunidades e em última instância nas propriedades dos ecossistemas (ciclos energéticos e químicos) (Quesada *et al.* 2012; Reich 2014). Os filtros geoquímicos e climáticos tendem a selecionar certas expressões fenotípicas dos indivíduos, denominados de traços funcionais, gerando convergência funcional no nível de comunidades (Fortunel *et al.* 2014; Muscarella & Uriarte 2016). Por outro lado, processos locais, como a competição, as dinâmicas estocásticas de recursos e as dinâmicas populacionais de fonte-sumidouro atuam de forma oposta, gerando divergência funcional dentro das comunidades (Kraft, Valencia & Ackerly 2008) e até mesmo entre comunidades. Em ambientes ricos em recursos, como nutrientes e água, predominam espécies de rápido crescimento e ciclo de vida curto (estratégias aquisitivas) (Reich 2014). Consequentemente, as taxas de mortalidade de árvores tendem a ser mais altas nestes ambientes, gerando maior frequência de perturbação por queda e formação de clareiras (Quesada *et al.* 2012), o que em florestas pode afetar os processos demográficos e padrões funcionais no sub-bosque. Os efeitos podem ser diretos, afetando a mortalidade dos indivíduos, ou indiretos, gerando flutuações temporais nos níveis de luz que alcançam o sub-bosque. Na Amazônia, regiões mais chuvosas tendem a ter também maior frequência de tempestades, o que influencia a dinâmica de distúrbios e a formação de clareiras na floresta (Espírito-Santo *et al.* 2010; Negrón-Juárez *et al.* 2017). Variações topográficas também interferem nas taxas de mortalidade de árvores, provavelmente devido à instabilidade no terreno e a disponibilidade diferenciada de água (Ferry *et al.* 2010; Toledo *et al.* 2011a). Dessa forma, os filtros externos (solo, clima e topografia) poderiam, em tese, gerar tanto convergência quanto divergência funcional. Em um cenário de predominância de respostas funcionais divergentes sobre convergentes ao longo de gradientes climáticos, por exemplo, indicaria uma ausência de previsibilidade de como as comunidades ecológicas responderiam às mudanças climáticas. **No segundo capítulo** avaliamos se solo, topografia e clima afetam os padrões de convergência e divergência funcional de 451 comunidades de plantas herbáceas de sub-bosque (Zingiberales) ao longo das florestas de terras baixas da Amazônia brasileira.

Os padrões de diversidade que observamos no presente foram moldados por processos complexos e históricos que se desenvolveram ao longo de milhões de anos. Entender quais fatores e como eles afetaram a história evolutiva dos sistemas ecológicos pode trazer informações sobre como eles devem responder aos efeitos de mudanças ambientais no futuro (Jaramillo & Cárdenas 2013). Embora se tenha demonstrado que diferentes processos estejam envolvidos na formação das espécies tropicais (Bush 1994; Fine *et al.* 2005; Antonelli *et al.* 2009; Ribas *et al.* 2012), ainda temos a carência de um modelo de diversificação, ou seja, que proponha quais mecanismos permitiram a acumulação de espécies de maneira diferenciada entre linhagens, regiões e habitats. O soerguimento dos Andes tem sido reconhecido como um importante evento impulsionador da diversificação na Amazônia. Sugere-se que seus impactos na diversidade tenham ocorrido devido à formação de barreiras vicariantes, transformações geoquímicas e morfológicas da paisagem e alterações dos regimes climáticos (Hoorn *et al.* 2010). Por outro lado, estudos recentes têm sugerido que não é necessário invocar o soerguimento dos Andes e suas decorrentes transformações para explicar padrões evolutivos na Amazônia (Smith *et al.* 2014; Dexter *et al.* 2017), sendo a acumulação diferenciada de espécies entre linhagens determinada primeiramente pelo tempo que as espécies tiveram para se dispersar e especiar ao longo da paisagem amazônica (Smith *et al.* 2014). Uma previsão simples deste modelo é que a riqueza de espécie dentro de um clado aumenta proporcionalmente conforme aumenta a idade de origem dos clados. Outros modelos, no entanto, têm focado em processos funcionais e ecológicos como geradores de diversidade (Fine *et al.* 2005; Baker *et al.* 2014). Espécies da família Burseraceae proximamente relacionadas tendem a ocupar habitats edáficos distintos, sugerindo a importância da radiação adaptativa ao longo de gradientes ambientais como formadores de espécies. Um outro estudo, conduzido com 21 gêneros de árvores, demonstrou uma relação entre taxas de diversificação e estratégias funcionais (Baker *et al.* 2014). Gêneros que possuem estratégias funcionais aquisitivas (indivíduos de rápido crescimento e ciclo de vida curto) apresentaram taxas de diversificação maiores que gêneros com estratégias opostas. Isso seria explicado pela relação negativa entre o tempo do ciclo de vida dos organismos e a taxa de evolução molecular (Smith & Donoghue 2008). Como estratégias funcionais aquisitivas estão relacionadas a ambientes ricos em recursos minerais na Amazônia (Quesada *et al.* 2012), os resultados de Baker *et al.* (2014) sugerem que as transformações geoquímicas passadas ocorridas principalmente no Oeste da bacia, teriam contribuído para o aumento da diversificação de certos gêneros, e consequentemente para o aumento de riqueza em certas linhagens e regiões da Amazônia. Portanto, **no terceiro e último capítulo desta tese** usamos sete linhagens dentro da família Marantaceae (ervas monocotiledôneas) para testar se linhagens com traços funcionais característicos de estratégias de rápido crescimento tendem a ter altas taxas de diversificação e riqueza de espécies, contrastando com previsões do modelo de diversificação baseado no tempo de origem das linhagens, onde é esperada uma maior riqueza em clados antigos. Utilizamos ainda modelos de afinidade de habitats e análise de dinâmica de diversificação para revelar quais fatores ecológicos e paleoeventos (geológicos e climáticos) estariam associados com os padrões evolutivos na Amazônia.

OBJETIVOS

O principal objetivo desta tese foi revelar a importância de filtros geoquímicos e climáticos para o entendimento de padrões biogeográficos, funcionais e evolutivos de plantas na Amazônia.

Capítulo 1: Avaliar a importância relativa de dados espaciais de solo e clima para prever a distribuição de espécies amazônicas.

Capítulo 2: Testar se filtros geoquímicos, hidro topográficos e climáticos promovem convergência e/ou divergência funcional em comunidades vegetais amazônicas.

Capítulo 3: Demonstrar como e quando transformações geoquímicas e morfológicas da paisagem decorrentes de eventos históricos e oscilações climáticas influenciaram nos processos evolutivos de plantas amazônicas.

ARTIGO 1

Beyond climate control on species range: the importance of soil data to predict distribution of amazonian plant species

FIGUEIREDO, F.O.G, ZUQUIM, G., TUOMISTO, H., MOULATLET, G.M., BALSLEV & COSTA, F.R.C.
Manuscrito aceito na *Journal of Biogeograpy*

ABSTRACT

Aim To evaluate the relative importance of climatic vs. soil data when predicting species distributions for Amazonian plants and to gain understanding of potential range shifts under climate change.

Location Amazon rain forest.

Methods We produced Species Distribution Models (SDM) at 5 km spatial resolution for 42 plant species (trees, palms, lianas, monocot herbs and ferns) using species occurrence data from herbarium records and plot-based inventories. We modelled species distribution with Bayesian logistic regression using either climate data only, soil data only or climate and soil data together to estimate their relative predictive powers. For areas defined as unsuitable to species occurrence, we mapped the difference between the suitability predictions obtained with climate-only vs. soil-only models to identify regions where climate and soil might restrict species ranges independently or jointly.

Results For 40 out of the 42 species, the best models included both climate and soil predictors. The models including only soil predictors performed better than the models including only climate predictors, but we still detected a drought-sensitive response for most of the species. Edaphic conditions were predicted to restrict species occurrence in in the centre, the northwest and in the northeast of the Amazonia, while the climatic conditions were identified as the restricting factor in the eastern Amazonia, at the border of Roraima and Venezuela and in the Andean foothills.

Main conclusions Our results revealed that soil data, relative to climate, represents the most important predictor of plant species range in the Amazonia. The strong control of species ranges by edaphic features might reduce species' abilities to track suitable climate conditions under a drought-increase scenario. Future challenges are to improve the quality of soil data and couple them with process-based models to better predict species range dynamics under climate change.

Keywords Amazon rain forest**;** climate change; Bayesian Logistic Regression; Cation Exchange Capacity; Ecological Niche Models; SoilGrids; tropical soils; soil factors; species range; Species Distribution Models

INTRODUCTION

Climate is known to be an important determinant of species ranges at broad spatial scales, so an understanding of how climate change will affect ecological systems is urgently needed. The most obvious effect that climate change can be expected to have on biodiversity is shift in species' geographical distributions (Parmesan & Yohe, 2003; Thuiller, 2007). Species must either track suitable climate or adapt to novel climatic conditions scaping from extinction. In turn, some species can be favoured by climate changes and expand their ranges. However, range shifts may be decoupled from pure climatic tracking by other factors that also restrict species distribution, such as non-environmental spatial constraints, land-use, biotic interactions and soil conditions, (Hayes & Sewlal, 2004; Marage & Gégout, 2009; Blach-Overgaard *et al.*, 2010; Eiserhardt *et al.* 2011; Wisz *et al.*, 2013). Identification of both climatic and non-climatic constraints on broad-scale species distributions is therefore necessary to better predict how specie will respond to climate change.

Recent projections of global climate models predict significant climate change in Amazonia by the end of the 21st century (Boisier *et al.*, 2015), which would affect the plant diversity patterns in important ways (Olivares *et al.*, 2015). It is already known that soil conditions affect species occurrence patterns in Amazonia, and therefore they deserve special attention in the modelling of plant distributions. Edaphic heterogeneity is tightly related to geological features, such as soil age and mineralogical composition of the parent material (Irion, 1978; Quesada *et al.*, 2010). Whereas climate is relatively uniform over large parts of the Amazon basin, geological formations differ widely in age and geochemistry. Many studies suggest that soils exert the main control on broad-scale floristic patterns in Amazonia (Fittkau *et al.*, 1975; Tuomisto & Poulsen, 1996; Sombroek, 2000; Higgins *et al.*, 2011). However, studies focused on the southern and northern extremes of Amazonia have found a significant turnover of species along precipitation and temperature gradients (ter Steege *et al.*, 2006; Toledo *et al.*, 2011). This suggests that climatic conditions may be important in constraining species ranges at the borders of the biome. Although both climate and soil are considered important determinants of broad-scale floristic patterns in Amazonia, a rigorous evaluation of their relative roles for predicting plant species ranges has been overlooked.

Several studies have used spatial correlative models to predict current species distributions and potential range shifts under climate change for Amazonian species (Buermann *et al.*, 2008; Feeley *et al.*, 2012; Thomas *et al.*, 2012; Vedel-Sørensen *et al.*, 2013; Thomas *et al.*, 2014). However, either these studies relied on climatic and remote sensing variables, or used proxies, such as coarse resolution polygon-based maps of Ecoregions or soil types, to represent other environmental factors than climate (Feeley *et al.*, 2012; Vedel-Sørensen *et al.*, 2013; Thomas *et al*., 2014). In general, soil variables are often neglected in SDM because accurate and ecologically relevant data covering large extents are scarce. Recent efforts to develop digital soil maps, such as SoilGrids, have increased the availability of high-resolution grid-based soil information (Hengl *et al.*, 2014). The available soil maps include chemical and physical properties of the soil, as well as occurrence probabilities of soil classes, which can be expected to predict plant species range better than the previously used coarse-resolution proxies.

Here, we present the first broad-scale evaluation of how soil and climate variables interplay to predict plant species ranges in Amazonia using SDMs and a diverse set of climate and soil data layers. We focused on 42 plant species from phylogenetically distant groups that represent different growth forms: trees, palms, lianas, ferns and monocot herbs. We aimed to: (i) test whether the inclusion of soil variables improves the performance of climate-based SDMs; (ii) quantify the unique and joint contributions of climate and soil variables in explaining species distributions; (iii) quantify the relative importance of individual variables in predicting species distributions; and (iv) predict areas where soil and climatic conditions, independently and jointly, restrict species ranges. We expected that inclusion of a physiologically relevant set of edaphic predictors in SDM in combination with climate variables would produce more accurate predictions of species distributions and shed light on how soil and climate changes may interact to drive shifts in species ranges in Amazonia.

METHODS

Species and Environmental data

Species data. We focused on 42 plant species (Table 1) to investigate the predictive performance of climate and soil variables for mapping species distributions. The species were selected to cover a wide spectrum of plant sizes and life history strategies (trees, palms, lianas, monocot herbs and ferns) as well as different range sizes (some restricted to Amazonian rain forests and others also found in other biomes). Species occurrence records were obtained from two sources: plot-based inventories and herbarium records. To ensure data consistency, species that are easy to identify in the field were preferably considered, and all species in the final set had to be well represented in the species occurrence data. Details about the data sources can be found in Table S1, Appendix S1 in Supporting Information.

Environmental data. As climate predictors, we used both WorldClim variables and remote sensing data. Out of the available WorldClim data (http://www.worldclim.org), we used the 19 bioclimatic variables, which express 11 temperature and eight precipitation metrics at about 1 km resolution (WorldClim version 1.4; Hijmans *et al*., 2005). The WorldClim variables were produced by spatial interpolation between meteorological stations, which creates high uncertainty in the estimated climate data in many parts of Amazonia, where stations are sparse. Therefore, we also included precipitation data as estimated by the Tropical Rainfall Measuring Mission (TRMM) satellite (Kummerow *et al*., 1998; http://disc.sci.gsfc.nasa.gov). We converted monthly data from 1998 to 2004 of the TRMM product 3B43 V6 at a 0.25° resolution (about 28 km at the equator) into two climatic variables: annual precipitation and dry season length, defined as the maximum consecutive number of months with less than 100 mm of precipitation.

We used four remote sensing variables that describe terrain and forest structure properties: elevation, percentage tree cover, percentage evergreen tree cover and canopy height. Elevation was obtained from the Shuttle Radar Topography Mission (http://worldgrids.org). Tree cover and evergreen tree cover were derived from images of the NOAA's Advanced Very High Resolution Radiometer acquired in 1992–93 (DeFries *et al*., 2000; http://www.landcover.org). Canopy height was derived from the spaceborne Geoscience Laser Altimeter System (Simard *et al*., 2011; http://daac.ornl.gov). All remote sensing variables were originally provided at about 1 km resolution. We included the terrain and forest structure data in the pool of climate variables because these variables can be expected to represent ecophysiological constraints on species distribution via complex interactions between temperature, precipitation and solar radiation (Grubb, 1977; Kempes *et al*., 2011). Although soil and topography also affect forest structure, these relationships must be detected specially at fine-scale resolution.

Soil variables were obtained from the SoilGrids 250m database (https://www.soilgrids.org). We used 19 soil variables that provide predicted values for the surface soil layer (0 – 15 cm depth). These included: three soil texture variables (percent of clay, sand and silt), which are related to soil drainage; two chemical variables related to nutrient availability, the cation exchange capacity (CEC, measured in $\text{cmol}_c\ \text{kg}^{-1}$ at pH 7) and pH (determined in KCl); and other 14 variables representing different soil classes (following the classification of FAO's World Reference Base), which are related to soil physical conditions and nutrient availability (Quesada *et al.*, 2010). The soil classes were selected based on their moderate to high occurrence probabilities in the Neotropics. We used each soil class layer as an independent continuous predictor, where the value for a given pixel represents the probability of occurrence of that soil class. In this way, our SDMs take into account the uncertainty associated with the spatial limits of soil classes. This differs from the common approach of treating polygon-based soil classes as categorical variables in statistical models, so that each grid cell belongs to exactly one soil class. Extreme environmental conditions (bare soils areas, sparsely vegetated areas and water bodies) recognized as improbable to be occupied by our species were excluded by combining the spatial mask present in all aforementioned environmental data and also the spatial mask employed in SoilGrids 1km (Hengl *et al.*, 2014; product SMKISR accessed in http://worldgrids.org). The latter excludes several non-vegetated areas that are included in SoilGrids 250m. We rescaled all environmental data to a 0.05° spatial resolution (about 5 km at the equator) and both environmental and species data were transformed to the Mollweide equal-area projection before analysis. The maps of all 44 environmental variables are presented in Fig. S1.1.

Modelling framework

To evaluate the role of climate and soil factors in predicting species ranges we built a set of three SDMs for each species: (i) CLIM models based on WorldClim and remote sensing variables; (ii) SOIL models based on SoilGrids variables; (iii) and CLIM+SOIL models based on both sets of variables together. Our modelling approach includes six steps, as detailed below:

Step 1: Pre-modelling: detecting outliers, reducing sample bias and setting geographical background.

To improve SDM quality, we first coarsened the spatial resolution of the occurrence data. We retained only one presence record per 5x5 km grid cell and removed geographical and environmental outliers in the presence records. Then, we applied a spatial filter procedure to reduce sample bias (spatial aggregation) of occurrence records (Fig. S2.2). Finally, we constrained the area for model calibration, evaluation and prediction separately for each species to avoid predicting species presence in areas far beyond their potential ranges. Details of the pre-modelling procedures are given in Appendix S2.

Step 2: Predicting relative species occurrence probabilities using Bayesian Logistic Regression

To predict the relative occurrence probability of a species in each 5x5 km grid cell we applied logistic regression using presence and pseudo-absence points. Logistic regression using presence and pseudo-absence points is similar to methods widely used in ecology (e.g., MAXENT and resource selection functions). All of these methods are closely related to the Poisson point process model recently viewed as an appropriate way to analyse presence-only data in a regression framework (Warton & Shepherd, 2010; Aarts *et al.*, 2012; Renner & Warton, 2013; Renner *et al.*, 2015). To approximate our SDM to the Poisson point process model, one has to generate a sufficient number of pseudo-absences (background points) in a regular or uniform random design (Warton & Shepherd, 2010; Renner *et al.*, 2015). For each species, we generated 10,000 random background points over the geographical background area defined in step 1. We used Bayesian logistic regression instead of a conventional maximum likelihood logistic model because the maximum likelihood algorithm tends to overestimate regression coefficients when the number of presence records is very small in relation to absence or pseudo-absence points, or when the distribution of a species along an environmental gradient is very narrow (Hefley & Hooten, 2015). Bayesian logistic regression circumvents these problems by combining the information of the likelihood functions and a weakly informative default prior based on the Cauchy distribution to estimate model parameters (Gelman *et al.*, 2008).

We performed stepwise model selection (backward-forward) using Akaike Information Criterion (AIC) to find the most parsimonious model for each species separately. We allowed high correlated variables belonging to different source groups (WorldClim, TRMM climatic data, remote sensing data or SoilGrids) to compete in the same model selection, but we excluded variables that were too highly correlated with another variable in the same source group. Before the model selection procedure, we found pairs of variables that had correlation greater than 0.7 (absolute value of Pearson correlation) and excluded one of them which had greater variance inflection factor. We stopped the process when all pairwise correlation was less than 0.7. Only linear and second-order polynomial effects (Gaussian-shaped curve) were used. We started the backward-forward procedure including linear effects only. Second-order polynomials were allowed to enter the models at any step. We repeated model selection three times for each species to obtain the best CLIM+SOIL, CLIM and SOIL models.

Step 3: Evaluating model performance

To evaluate the performance of CLIM+SOIL, CLIM, and SOIL models, we employed two criteria. First, the goodness-of-fit of the models was compared using ΔAIC, which represents the difference in AIC between the model being evaluated and the best model (Burnham & Anderson, 2002). ΔAIC of the best model for a given species is therefore zero, and the higher ΔAIC value for a given model, the poorer the empirical support for it. As a rule of thumb, ΔAIC values smaller than 2 indicate models with good support. To evaluate if the best models are better than a null model, we also computed ΔAIC values for models including only the intercept term. Second, the spatial accuracy of models was evaluated by AUC, defined as the area under the receiver operating characteristic curve (Bradley, 1997). In a presence-background design, the false positive rate of the ROC plot is replaced by the fraction of predicted area. AUC values range from 0 to 1, with AUC values higher than 0.7 indicating models with good performance and values lower than 0.5 indicating that a model is worse than the null model (Araújo *et al.*, 2005). Before computing AUC, the relative occurrence probabilities (ROP) were transformed into cumulative values, which assigns a pixel the sum of all ROP values equal to or lower than the ROP values for that pixel, multiplied by 100 to give a percentage (Phillips *et al.*, 2006). Cumulative ROP (hereafter defined as suitability index) ranges from 0 to 100 and is better than the raw ROP output to visualize species range boundaries (Merow *et al.*, 2013). For the best models returned by the model selection procedure, we applied a 10-fold cross-validation and computed the AUC test of each replicate. The final prediction of each model was produced by averaging the 10 replicates returned by cross-validation.

For each species, we applied a two-tailed non-parametric Wilcoxon test to check if spatial predictions measured by the AUC were better than the null model. For those species with best models better than null, the difference in AUC scores between CLIM+SOIL, CLIM and SOIL models was tested using Generalized Mixed Effect Models (GLMM) (Pinheiro & Bates, 2000). Model type was used as the fixed term and species were set as the random term. Since AUC range from 0 to 1, we fitted the GLME assuming that AUC scores follow a beta distribution (*logit* transformation).

Step 4: Quantifying the unique and joint effects of climate and soil variables on species distribution models

Since Bayesian logistic regression can be viewed as a Generalized Linear Model with a constrained approach to estimate model parameters, we employed the same procedure of variance partitioning used in linear modelling (Borcard *et al.*, 1992) to quantify how much of the variances explained by the best CLIM+SOIL, CLIM and SOIL models were explained uniquely by climate or soil variables and jointly by both sets of variables. Given that we applied the model selection procedure to find the best models in each of the three model settings, the SOIL and CLIM models do not necessarily include the same subset of variables that were retained in the best CLIM+SOIL model. Hence, our approach is not a true variance partitioning, but an approximation of it. Percentage of total explained variance (EV) in logistic models, also known as pseudo-R^2, was given by:

$$EV = \frac{null\ deviance - residual\ deviance}{null\ deviance} * 100$$

as suggested by Zuur et al. (2009). The unique contribution of climate was computed as $EV_{clim+soil} - EV_{soil}$, and that of soil as $EV_{clim+soil} - EV_{clim}$. The joint contribution of climate and soil was obtained as $EV_{soil} + EV_{clim} - EV_{clim+soil}$. We report the unique and joint contributions of CLIM and SOIL model as percentages of EV rather than as percentages of total variance. This is done because pseudo-absence data may be generated also for pixels containing presence records, so the total variance will never be fully explained.

Step 5: Quantifying the relative importance of individual variables.

For each species, we ran Bayesian logistic regressions separately with each variable to quantify their relative importance. For each of 44 predictors, we ran a single-variable SDM and selected the best shape of the species response curve (no effect, linear, or second order polynomial) using ΔAIC scores. The EV of the single-predictor model was then expressed as a percentage of the EV of the best model identified in step 3. Finally, we ranked variables based on their median EV values over all species.

Step 6: *Mapping the difference between predictions of CLIM and SOIL models.*

Since our goal is to identify which factors (climatic, soil or both) potentially restrict species distribution, we mapped the difference between the models′ predictions only in areas defined as unsuitable, *i.e,* in areas that species are likely to be absent. To define the unsuitable areas, we applied a threshold on each species suitability scores that maximizes the sum of sensitivity (1 - omission error) and specificity (the fraction of predicted area) (Manel *et al.*, 2001). First, we mapped the difference in suitability between the best model (usually CLIM+SOIL) and each of the models being tested, the CLIM or SOIL models. In both case, high values (over-predictions) indicate areas that variables evaluated are unlikely to be the restricting factor for species occurrence. Second, to identify areas where both predictors may jointly or independently restrict species range, we mapped the difference between predictions of the CLIM and SOIL model. Values close to zero indicate areas where we hypothesised that both soil and climate restrict species range. Positive values indicate regions where the CLIM model over-predicted suitable conditions relative to SOIL model and, hence, we hypothesised that species range was mainly restricted by soil in such areas. Conversely, negative values indicate regions where climate may be the main restricting factor. Finally, we averaged the maps of all species across Amazonia *sensu latissimo* (Eva & Huber, 2005) to produce an overall predictive map of potential climate and soil constraint on species ranges.

All analyses were conducted using the R platform (v3.3.2; R Development Core Team, 2015). Manipulation of raster and vector data was done using ‘raster’ (Hijmans, 2006) and ‘sp’ (Pebesma & Bivand, 2005) packages. Data from GBIF and iDigBio were downloaded using the function *occ* from package ‘spocc’ (Chamberlain *et al.*, 2016). Geographical outliers were detected and removed using the function *rjack* of ‘biogeo’ package (Robertson, 2016). We executed the spatial filtering procedure using functions *lohboot* and *remove.duplicates* of the ‘spatstat’ (Baddley & Turner, 2005) and ‘dismo’ package (Hijmans *et al.*, 2017), respectively; exclusion of correlated variables with *vifcor* function of the ‘usdm’ package (Naimi, 2015); Bayesian logistic regression with function *bayesglm* of ‘arm’ package with default options (Gelman & Su, 2016); computation of AUC scores using function *auc* from package ‘MESS’ (Ekstrøm, 2016); evaluation of model significance with the *wilcox.test* function of the ‘stats’ package (R Development Core Team, 2015). The tests of differences in model performance were done fitting a GLMM using function *gamlss* of the ‘gamlss’ package (Rigby & Stasinopoulos, 2005). We implemented our own R function to perform the model selection procedure and the R scripts are available upon request.

RESULTS

Our results clearly indicate that soil data are important for the species distribution models developed in this study (Fig. 1). For 39 of the 42 species, the models combining climatic and soil data (CLIM+SOIL) were supported as the best models, and for the three remaining species, either the relative ranking of the CLIM+SOIL and SOIL models was uncertain, or the SOIL model had more support. The CLIM model was not supported as the best model for any of the species. Models including soil predictors only had better performance than models including climate predictors only. Across all species, and controlling for random effects, the AUC values for the SOIL models were significantly lower than those for CLIM+SOIL models ($P<0.001$) but higher than those for CLIM models ($P<0.005$). For every species, the best models showed fair to good prediction accuracy. All median AUC of the best models were greater than 0.7 (Fig. S3.3) and significantly greater than the null model (all $P < 0.005$ based on Wilcoxon test of AUC). All model statistics (ΔAIC, AUC, omission rates, fractions of predicted area and thresholds used to define unsuitable areas) are provided in Fig S3.3.

For 25 species (60%), more variance was explained uniquely by SOIL variables than uniquely by CLIM variables, and for 30 species (71 %), more than 50% of the explained variance was due to the joint contribution of CLIM and SOIL variables (Fig. 2). Annual precipitation, dry season length, canopy height, tree cover and Haplic Xanthic Ferralsols were the variables with the highest explained variance, on average, in the single-variable SDMs (Fig. 3). However, single-variable models never explained more than 15% of the variance explained by the corresponding best model, and the average was less than 8% for each one of the 44 variables. Species response curves along the most important CLIM variables revealed that most species tend to avoid dry conditions (drought-sensitive) and occur preferentially in tall closed forest (Fig. S3.4). The response curves of the 42 species along each of the 44 variables are shown in Appendix S3.

Visual assessment of the SDMs for each of the 42 species showed that SOIL models predicted more irregular and more patchy distributions than CLIM models (Fig. S3.3). CLIM models over-predicted suitable conditions in central Amazonia, upper Rio Negro, northern Guianas and in some areas of southern Amazonia, which reveals that it is more likely that species are absent in these areas due to soil restriction, not due to climate restriction (Fig. 4). The opposite was true in areas where the SOIL model over-predicted suitable conditions, *i.e,* in the Andean foothills, the border zone between Peru and Acre state, eastern Amazonia and the savannas of Roraima and Venezuela (Fig. 4). Both models predicted low suitability in southern Amazonia, which indicates that soil and climate jointly restrict species occurrence in this area (Fig. 4c).

DISCUSSION

Using SDM and a diverse set of climatic and edaphic predictors, we demonstrated that the soil-only models provide more accurate predictions than the climate-only models. This can be seen as contradicting the classical assumption of biogeography that climate is the main driver of broad-scale species distribution patterns (Gaston, 2003; Pearson & Dawson, 2003). The higher importance of soil relative to climate may reflect the fact that the variation in rainfall and temperature is narrow across large parts of Amazonia. On the other hand, the edaphic differences among regions can be so large that they act as filters preventing species from occurring over the entire biome. Indeed, the combined soil and climate models had good predictive power for most species, which is consistent with earlier findings suggesting that both climate and soil properties are important as drivers of plant diversity and community composition across Amazonia (Tuomisto *et al.*, 2003b; ter Steege *et al.*, 2006, Zuquim *et al.*, 2014).

Differences between the predictions made by the best models, climate-only and soil-only models are illustrative, because they indicate areas where species distributions might be constrained jointly by both groups of variables, or predominantly by one of them. ter Steege *et al*. (2006) found that the main axis of floristic variation across Amazonia was associated with a southwest-northeast soil fertility gradient and the second axis with a northwest-southeast climatic gradient. Our predictive maps reveal a more complex and slightly different spatial pattern of soil and climate control on species distribution than that. In particular, our results suggest that that soil might be more decisive than climate in restricting species occurrence particularly in the upper Rio Negro area, central Amazonia and the northern part of the Guiana region, which are covered predominantly by very acid and poor nutrient soils, or even by sandy soils (Sombroek, 2000; Quesada *et al.*, 2010). Our findings, therefore, are in accordance with previous studies that predict the absence of rich-soil species in large areas of Amazonia, particularly in the centre of biome (Fittkau *et al.* ,1975; Zuquim *et al.,* 2014; Tuomisto *et al.* 2016). On the other hand, where climate conditions vary markedly over short distances (in the Andean foothills, in the border between Peru and Acre state, eastern

Amazonia and the savannas of Roraima and Venezuela), climate may become more restrictive to species occurrence, especially if soil conditions are relatively uniform. The southern and north-western borders of Amazonia were predicted as unsuitable by both climate-only and soil-only models, suggesting that the northwest-southeast floristic gradient that was interpreted to be climatic by ter Steege *et al.* (2006) may reflect soil-related factors as well.

When the environmental variables were used separately in single-predictor SDMs, the ten most important predictors included precipitation-related variables (annual rainfall and dry season length), forest structural variables (canopy height and percentage of evergreen tree cover) and soil variables (Haplic Xanthic Ferralsols, pH and Haplic Acrisols). However, any variable alone explained less than 15% of the variance explained by the corresponding best model, indicating that species ranges are controlled by multiple factors. Many of the predictor variables are strongly correlated, and for most species the proportion of deviance explained jointly by climatic and edaphic variables was more than 50% of total explained variance. Such relationships can be causal, as temperature, precipitation and topography are among the main factors that affect soil formation (Jenny, 1994). Nevertheless, climate conditions seem to be less important than age and quality of the parent material in determining soil properties in Amazonia (Irion, 1978; Quesada *et al.*, 2010). Soils in most of lowland Amazonia are derived from sedimentary deposits, and different kinds of sediments have been deposited on top of each other during different time periods (Räsänen *et al.*, 1987; Hoorn et al. 2010). The relationship between topography, soil properties and climate is somewhat complex. In Amazonian lowlands, local and regional altitudinal differences can be related to soil properties (texture and nutrient content) (Tuomisto *et al.*, 2003a; Costa *et al.*, 2005; Higgins *et al.*, 2011), but in mountains there are also marked changes in temperature and solar radiation (Grubb 1977). In addition, forests structure can be affected by climate, especially the amount and seasonality of rainfall, and hydrological soil properties, which are largely defined by soil texture and topography (Levine et al. 2016). More detailed studies are still needed to disentangle the complex relationships and interactions among climate, soil, topography, forest structure and other variables that can affect species distributions.

Current hydrological trends and recent projections of global climate models predict increasing dry season length for the southern part of Amazonia (Boisier *et al.*, 2015). Increasingly dry conditions may cause forests to become shorter and more open (Hutyra *et al.*, 2005; Olivares *et al.*, 2015; Levine *et al.*, 2016). The species response curves along annual precipitation, dry season length and forest structure gradients suggest that most of our focal species avoid dry conditions and prefer tall and closed forests. This is consistent with the prevailing view that many Amazonian plant species are sensitive to drought (Nepstad *et al.*, 2007; Phillips *et al.*, 2010) and that their ranges may contract if dry seasons become more severe (Feeley *et al.*, 2012; Olivares *et al.*, 2015). Temperature showed weak predictive power in our SDMs, suggesting that changes in temperature may be less important for range shifts than changes in water availability.

Massive species loss in Amazonia is expected under climate change if species are not able to either track suitable climate by migration or adapt to the new conditions (Feeley *et al.*, 2012). If species distributions are strongly controlled by soils, species migrations may be effectively prevented by barriers created by large areas of unsuitable soil (Tuomisto et al. 2016). In our analyses, climatic variables alone explained species range limits in fewer areas than soil variables alone did, which suggests that species may indeed face problems in finding suitable routes to track climate change. Since soil properties change much slower than climate does, and climate does not seem to be an important driver of soil features in Amazonian lowlands (Quesada *et al.* 2010), the presence of edaphic constraints on species ranges can decouple potential range shifts from pure climate tracking.

Dispersal process and habitat heterogeneity are important aspect to understand species' range shift under climate change (Opdam & Wascher, 2004; Schurr *et al.*, 2012). Species range dynamics are affected by a combination of processes operating at different spatial scales. Regional climate changes force range shifts (expansion or contraction) while landscape heterogeneity and metapopulation dynamics control the velocity at which a species' range actually changes. For instance, where suitable soil conditions are restricted to isolated patches, dispersal rates within metapopulation are low and it becomes more difficult for species both to track suitable climate to new areas and to persist in areas that are becoming climatically unsuitable (Opdam & Wascher, 2004). Disregarding soil variables in fine-scale dynamic models has led to inaccurate predictions of potential distributional areas of species under climate change in Europe (Bertrand *et al.*, 2012). Although the fine spatial configuration of soil properties is not reflected in the coarse spatial resolution of our analyses, our results suggest that species distributions are more irregular and patchy than the rather continuous areas that get predicted by climate-only models. The inclusion of an edaphic dimension, therefore, in the forthcoming process-based dynamic models should guide useful conservation and management strategies to protect Amazonian plant species under climate change.

Our analyses followed the standard procedure of using environmental data that are publicly available. However, WorldClim data have been criticised as inaccurate (Soria-Auza *et al.*, 2010). Although the quality of the SoilGrids data layers has not yet been properly assessed, inaccuracies must be expected simply because the density of field sampling in Amazonia is very low. In addition, our analyses were carried out at a 5 km resolution, which reduced our ability to capture fine-scale soil variation. A probable consequence of this is that the models predict edaphic niches to be wider than they are in reality, which leads to spatial over-prediction of suitable conditions. Despite these limitations, the contribution of soil data to the modelled species geographical ranges was significant, and we expect that better soil data will further reinforce this finding.

Another potential problem is that the available soil variables may not be ecologically optimal for species distribution modelling. For example, the soil data available in SoilGrids do not include the summed concentration of base cations (Ca, Mg, K, Na), which has been shown to be an important determinant of floristic patterns in several studies (Tuomisto *et al.*, 2003b; Higgins *et al.*, 2011; Figueiredo *et al.*, 2014; Zuquim *et al.*, 2014, Tuomisto et al., 2016, Cámara-Leret *et al.* 2017). Cation exchange capacity (CEC) is available, but CEC is affected both by base cations and by aluminium, which is not a plant nutrient and becomes toxic to some species at high levels (Delhaize & Ryan, 1995). Indeed, CEC did not come out as an important variable in our SDMs, probably because it is a mixed variable that does not properly represent either the soil fertility gradient or the soil toxicity gradient. Soil pH did emerge as important in our SDMs, even though studies based on plot-based inventories have often found pH to be less important than the sum of base cations for explaining species occurrence patterns in Amazonia (Ruokolainen *et al.*, 2007; Tuomisto et al., 2016). It is beyond the scope of the present study to assess the extent of such problems, although they are likely to lead to underestimating the importance of soil factors in SDMs.

The relative importance of climate as opposed to edaphic drivers of species distribution has traditionally been considered a matter of scale of analysis. Climate is expected to be useful at broad extents (e.g. global, continental and regional) and soil is expected to be relevant only at small extents and in high-resolution studies (e.g., landscape and local scales) (Pearson & Dawson, 2003; Eiserhardt *et al.*, 2011). In our analyses, soil variables emerged as the most important even at the continental extent. This is no doubt related to relative homogeneity of the climate within Amazonia, at least when compared to subtropical, temperate and montane areas where climatic differences are much larger and, therefore, more restrictive to plant distributions. However, even at global-scale, edaphic variables can better explain ecological patterns (*e.g.*, the variation in some leaf functional traits) than do climatic variables (Maire *et al.*, 2015). In Amazonia, soils mediate several ecological patterns and processes, such as large-scale floristic turnover (ter Steege *et at.,* 2006; Tuomisto *et al.*, 2016), biotic interactions (Fine *et al.*, 2004), species diversification (Fine *et al.*, 2005), forest dynamics and resilience (Quesada *et al.*, 2012, Levine *et al.*, 2016). We believe that it is time to scale-up the importance of soil in ecology and biogeography. The inclusion of the soil component in macroecological models will also provide a better framework toward understanding the complex impacts of climate change on ecological systems.

ACKNOWLEDGEMENTS

We thank Mike Hopkins for the Zingiberales data from Amazonian herbaria, Robert Magill for data of the MBG herbarium, and Eric Anderson for the English language review. Funding from the Brazilian Program for Biodiversity Research - (PPBio), CNPq (#47890820/2-1, led by FC); CNPq (fellowship to F.F); CNPq-PPG7 (#556988/2005-1), FINEP/Projeto Integrado MCT-EMBRAPA (led by Ana L. K. M. Albernaz); Biological Dynamics of Forest Fragments (BDFFP, grant to F.F.); Hidroveg Project – FAPEAM/ FAPESP (#1428/2010 led by FC and Javier Tomasella), Academy of Finland (grants to H.T.) and Danish National Research Council (grant to H.B.). This is paper number XXX from BDFFP series.

REFERENCES

Aarts, G., Fieberg, J. & Matthiopoulos, J. (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, **3**, 177–187.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species – climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

Baddeley, A. & Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software* **12**, 1-42. Available at http://www.jstatsoft.org/v12/i06/.

Bertrand, R., Perez, V. & Gégout, J.-C. (2012) Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of Quercus pubescens in France. *Global Change Biology*, **18**, 2648–2660.

Blach-Overgaard, A., Svenning, J.C., Dransfield, J., Greve, M. & Balslev, H. (2010) Determinants of palm species distributions across Africa: the relative roles of climate, non-climatic environmental factors, and spatial constraints. *Ecography*, **33**, 380–391.

Boisier, J.P., Ciais, P., Ducharne, A. & Guimberteau, M. (2015) Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nature Climate Changes*, **5**, 656–660.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

Bradley, A.P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, **30**, 1145–1159.

Buermann, W., Saatchi S., Smith T., Zutta B., Chaves J., Milá B. & Graham C. (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, **35**, 1160-1176.

Burnham, K.K.P. & Anderson, D.R.D. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. Springer-Verlag, New York.

Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H. & Kristiansen, S.M. (2017) Modelling responses of western Amazonian palms to soil nutrients. *Journal of Ecology*, **105**, 367–381.

Chamberlain, S., Ram, K. & Hart, T. (2016) spocc: Interface to species occurrence data sources. *R package version 0.5.0.* Available at https://CRAN.R-project.org/package=spocc

Costa, R.C.F., Magnusson, W.E. & Luizão, R.C. (2005) Mesoscale distribution patterns of Amazonian understorey herbs in relation to topography, soil and watersheds. *Journal of Ecology*, **93**, 863–878.

DeFries, R.S., Hansen, M.C., Townshend, J.R.G., Janetos, A.C. & Loveland, T.R. (2000) A new global 1 km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, **6**, 247–254.

Delhaize, E. & Ryan P. (1995) Aluminium toxicity and tolerance in plants. *Plant Physiology*, **107**, 315-321.

Ekstrøm, C. (2016). MESS: Miscellaneous Esoteric Statistical Scripts. *R package version 0.4-3*. Available at https://CRAN.R-project.org/package=MESS

Eiserhardt, W., Svenning J., Kissling W. & Balslev H. (2011) Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Annals of Botany*, **108**, 1391-1416.

Eva, H.D. & Huber, O. (2005) *A proposal for defining the geographical boundaries of Amazonia*, Office for Official Publications of the European Communities, Luxemburg.

Feeley, K.J. (2012) Distributional migrations, expansions, and contractions of tropical plant species as revealed in dated herbarium records. *Global Change Biology,* **18**, 1335–1341.

Feeley, K.J., Malhi, Y., Zelazowski, P. & Silman, M.R. (2012) The relative importance of deforestation, precipitation change, and temperature sensitivity in determining the future distributions and diversity of Amazonian plant species. *Global Change Biology*, **18**, 2636–2647.

Figueiredo, F.O.G., Costa, F.R.C., Nelson, B.W. & Pimentel, T.P. (2014) Validating forest types based on geological and land-form features in central Amazonia. *Journal of Vegetation Science*, **25**, 198–212.

Fine, P., Mesones, I. & Coley, P.D. (2004) Herbivores promote habitat specialization by trees in Amazonian forests. *Science*, **305**, 663-665.

Fine, P., Daly D., Muñoz G., Mesones I. & Cameron K. (2005) The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae in the western Amazon. *Evolution*, **59**, 1464–1478.

Fittkau, E.J., Junk, W., Klinge, H. & Sioli, H. (1975) Substrate and vegetation in the Amazon region. *Berichte der Internationalen Symposien der Internationalen Vereinigung für Vegetationskunde Herausgegeben von Reinhold Tüxen.* (ed. by J. Cramer), pp. 73–90. J. Cramer, Vaduz.

Gaston, K.J. (2003) *The Structure and Dynamics of Geographical Ranges*, Oxford University Press, Oxford.

Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y. (2008) Weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistic*, **2**, 1360–1383.

Gelman, A & Su, Y-S. (2016) arm: Data analysis using regression and multilevel hierarchical models. *R package version 1.9-1.* Available at https://CRAN.R-project.org/package=arm

Grubb P.T. (1977) Control of forest growth and distribution on wet tropical mountains: with special reference to mineral nutrition. *Annual Review of Ecology and Systematics*, **8**, 83–107.

Hayes, F.E. & Sewlal, J.A.N. (2004) The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. *Journal of Biogeography*, **31**, 1809–1818.

Hefley, T.J. & Hooten, M.B. (2015) On the existence of maximum likelihood estimates for presence-only data. *Methods in Ecology and Evolution*, **6**, 648–655.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G. & Gonzalez, M.R. (2014) SoilGrids1km —Global soil information based on automated mapping. *PLoS ONE*, **9**, e105992.

Higgins, M., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R. & Räsänen, M. (2011) Geological control of floristic composition in Amazonian forests. *Journal of Biogeography*, **38**, 2136–2149.

Hijmans, J.R. (2016) raster: Geographic data analysis and modeling. *R package version 2.5-8*. Avaliable at https://CRAN.R-project.org/pakcage=raster.

Hijmans, R.J, Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **1978**, 1965–1978.

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2017). dismo: Species Distribution Modeling. *R package version 1.1-4*. Available at https://CRAN.R-project.org/package=dismo

Hoorn, C., Wesselingh, F.P., Steege, H. ter, Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T. & Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.

Hutyra, L.R., Munger, J.W., Nobre, C.A., Saleska, S.R., Vieira, S.A. & Wofsy, S.C. (2005) Climatic variability and vegetation vulnerability in Amazônia. *Geophysical Research Letters*, **32**, doi:10.1029/2005GL024981.

Irion, G. (1978) Soil infertility in the Amazonian rain forest. *Naturwissenschaften*, **10**, 515–519.

Jenny, H. (1994) *Factors of soil formation: a system of quantitative pedology*. Dover books on Earth sciences. Dover Publications.

Kempes, C.P., West, G.B., Crowell, K., & Girvan, M. (2011) Predicting maximum tree heights and other traits from allometric scaling and resource limitations. *PLoS ONE*, **6**, e20551.

Levine, N.M., Zhang, K., Longo, M., Baccini, A., Phillips, O.L., Lewis, S.L., Alvarez-Dávila, E., Segalin de Andrade, A.C., Brienen, R.J.W., Erwin, T.L., Feldpausch, T.R., Monteagudo Mendoza, A.L., Nuñez Vargas, P., Prieto, A., Silva - Espejo, J.E., Malhi, Y. & Moorcroft, P.R. (2016). Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. *Proceedings of the National Academy of Sciences*, **113,** 793–797.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. & Simpson, J. (1998) The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.

Maire, V., Wright, I.J., Prentice, I.C., Batjes, N.H., Bhaskar, R., van Bodegom, P.M., Cornwell, W.K., Ellsworth, D., Niinemets, Ü., Ordonez, A., Reich, P.B. & Santiago, L.S. (2015) Global effects of soil and climate on leaf photosynthetic traits and rates. *Global Ecology and Biogeography*, **24**, 706–717.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence absence models in ecology; the need to count for prevalence. *Journal of Applied Ecology*, **38**, 921–931.

Marage, D. & Gégout, J.-C. (2009) Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology and Biogeography*, **18**, 88–97.

Merow, C., Smith, M.J. & Silander, J. a. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Miles, L., Grainger, A. & Phillips, O. (2004) The impact of global climate change on tropical forest biodiversity in Amazonia. *Global Ecology and Biogeography*, **13**, 553–565.

Naimi, B. (2015) usdm: Uncertainty analysis for species distribution models. *R package version 1.1-15.* Available at https://CRAN.R-project.org/package=usdm.

Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P. & Cardinot, G. (2007). Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology*, **88**, 2259–2269.

Olivares, I., Svenning, J-C., van Bodegom, P.M., & Balslev, H. (2015) Effects of warming and drought on the vegetation and plant diversity in the Amazon Basin. *The Botanical Review*, **81**, 42–69.

Opdam, P. & Wascher, D. (2004) Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation*, **117,** 285-297.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of speces: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

Pebesma, E.J & Bvand, R.S. (2005) Classes and methods for spatial data in R. *R News 5(2).* Available at https://cran.r-project.org/doc/Rnews/

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, O. L., van der Heijden, G., Lewis, S. L., *et al.* (2010). Drought-mortality relationships for tropical forests. *The New Phytologist*, **187**, 631–46.

Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*, Springer Verlag, New York.

Quesada, C.A., Phillips, O.L., Schwarz, M., *et al.* (2012) Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences*, **9**, 2203–2246.

Quesada, C.A., Lloyd, J., Schwarz, M., *et al.* (2010) Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences*, **7**, 1515–1541.

R Developement Core Team (2015) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Avaliable at https://www.R-project.org .

Räsänen, M.E., Salo, J.S. & Kalliola, R.J. (1987) Fluvial perturbance in the western Amazon basin: regulation by long-term sub-Andean tectonics. *Science*, **238**, 1398–1401.

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S., Popovic, G. & Warton, D.I. (2015) Point process models for presence-only analysis – a review. *Methods in Ecology and Evolution*, **6**, 366–379.

Renner, I.W. & Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274-281.

Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape. *Applied Statistics,* **54**, 507-554.

Robertson, M. (2016) biogeo: Point data quality assessment and coordinate conversion. *R package version 1.0,* Available at https://CRAN.R-project.org/package=biogeo

Ruokolainen, K., Tuomisto, H., Macía, M.J., Higgins, M.A. & Yli-Halla, M. (2007) Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *Journal of Tropical Ecology*, **23**, 13.

Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B., Hartig, F., Kissling, W.D., Linder, H.P., Midgley, G.F., Schröder, B., Singer, A. & Zimmermann, N.E. (2012) How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography*, **12,** 2146–2162.

Simard, M., Pinto, N., Fisher, J.B. & Baccini, A. (2011) Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, **116**, doi:10.1029/2011JG001708.

Sombroek, W. (2000) Amazon landforms and soils in relation to biological diversity. *Acta Amazonica*, **30**, 81–100.

Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K. & Böhner, J. (2010) Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling*, **221**, 1221–1229.

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

Thomas, E., Caicefo, C.A., Loo, J. & Roeland, K. (2014) The distribution of the Brazil nut (*Bertholletia excelsa*) through time: from range contraction in glacial refugia, over human-mediated expansion, to anthropogenic climate change. *Boletim do Museu Paraense Emílio Goeldi. Ciências Naturais*, **9**, 267–291.

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G. & van Etten, J. (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the Neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal. *PLoS ONE*, **7**, e47676.

Thuiller, W. (2007) Climate change and the ecologist. *Nature*, **448**, 550–552.

Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Chuvinã, J., Leaño, C., Licona, J.C., ter Steege, H. & Bongers, F. (2011) Patterns and determinants of floristic variation across lowland forests of Bolivia. **43**, 405–413.

Tuomisto, H. & Poulsen, A.D. (1996) Influence of edaphic specialization on pteridophyte distribution in neotropical rain forests. *Journal of Biogeography*, **23**, 283–293.

Tuomisto, H., Ruokolainen, K., Aguilar, M. & Sarmiento, A. (2003a) Floristic patterns along a 43-km long transect in an Amazonian rain forest. Journal of Ecology, **91**, 743–756.

Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003b) Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, **299**, 241–244.

Tuomisto. H., Moulatlet, G., Balslev, H., Emilio, T., Figueiredo, F.O.G., Pedersen, D. & Ruokolainen, K. (2016) A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *Journal of Biogeography*, **12,** 2400–2411.

Vedel-Sørensen, M., Tovaranonte, J., Bøcher, P., Balslev, H. & Barfod, A. (2013) Spatial distribution and environmental preferences of 10 economically important forest palms in western South America. *Forest Ecology and Management*, **307**, 284-292.

Warton, D.I. & Shepherd, L.C. (2010) Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Annals of Applied Statistics*, **4**, 1383–1402.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D. a., Aastrup, P. & Svenning, J.C. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Zuquim, G., Tuomisto, H., Jones, M.M., Prado, J., Figueiredo, F.O.G., Moulatlet, G.M., Costa, F.R.C., Quesada, C.A. & Emilio, T. (2014) Predicting environmental gradients with fern species composition in Brazilian Amazonia. *Journal of Vegetation Science*, **25**, 1195–1207.

Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R.* Springer, New York.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Species occurrence data sources (Table S1) and maps of the 44 environmental layers used in SDMs (Fig. S1).

Appendix S2. Description of the pre-modelling procedures: detection of outliers, reductions of sample bias (Fig. S2) and definitions of geographical background.

Appendix S3. Spatial projections of SDM for the 42 Amazonian plant species (Fig. S3.3) and the predicted species response curves along each of the environmental layers (Fig. S3.4)

BIOSKETCH

All authors share interest in understanding the ecological drivers of Amazonian biodiversity distribution.

Author contributions: F.F and F.C. conceived the idea and led the writing; F.F. analysed the data; all authors collected data and helped with the writing.

Editor: Steven Higgins

TABLES

Table 1. List of the 42 Amazonian plant species, geographical regions of occurrence and the number of occurrence records (*N*) after removing outliers and applying a spatial filtering procedure with the indicated radius distance (*d*) to reduce sample bias. The geographical regions were classified as Amazonian rain forest (AM), Central American rainforest (CA), Cerrado biome (CB) and Atlantic rain forest (AT).

Plant group	Species	Family	Distribution	*N*	*d* (km)
Tree	*Caryocar glabrum* (Aubl.) Pers.	Caryocaraceae	AM	375	50
Tree	*Caryocar microcarpum* Ducke	Caryocaraceae	AM	106	25
Tree	*Couepia dolichopoda* (Prance) Sothers & Prance	Chrysobalanaceae	AM	24	5
Tree	*Eperua falcata* Aubl.	Fabaceae	AM	72	50
Tree	*Henriquezia nitida Spruce ex Benth.*	Rubiaceae	AM	43	5
Tree	*Inga alba* (Sw.) Willd.	Fabaceae	AM,CA,CB,AT	534	50
Tree	*Jacaranda copaia* (Aubl.) D. Don	Bignoniaceae	AM,CA	513	50
Tree	*Jacaratia spinosa* (Aubl.) A. DC.	Caricaceae	AM,CA,CB,AT	238	50
Tree	*Manilkara huberi* (Ducke) A. Chev.	Sapotaceae	AM	178	50
Tree	*Mezilaurus itauba* (Meisn.) Taub. ex Mez	Lauraceae	AM	318	50
Tree	*Minquartia guianensis* Aubl.	Olacaceae	AM,CB	183	50
Tree	*Nectandra turbacensis* (Kunth) Nees	Lauraceae	AM,CA,CB	89	50
Tree	*Poecilanthe effusus* Huber	Fabaceae	AM	154	50
Tree	*Simarouba amara* Aubl.	Simaroubaceae	AM,CA,CB,AT	576	50
Tree	*Siparuna decipiens* (Tul.) A. DC.	Siparunaceae	AM	411	50
Palm	*Astrocaryum aculeatum* G. Mey.	Arecaceae	AM	96	50
Palm	*Bactris acanthocarpa* Mart.	Arecaceae	AM,AF	108	50
Palm	*Iriartea deltoidea* Ruiz & Pav.	Arecaceae	AM.CA	133	50

Palm	*Leopoldinia pulchra* Mart.	Arecaceae	AM	45	15
Palm	*Oenocarpus bataua* Mart.	Arecaceae	AM	106	50
Palm	*Socratea exorrhiza* (Mart.) H. Wendl.	Arecaceae	AM,CA	210	50
Liana	*Callichlamys latifolia* (Rich.) K. Schum.	Bignoniaceae	AM,CA,CB,AT	231	50
Liana	*Machaerium amplum* Benth.	Fabaceae	AM,CB,AF	83	20
Liana	*Machaerium ferox* (Mart. ex Benth.) Ducke	Fabaceae	AM	62	5
Liana	*Machaerium multifoliolatum* Ducke	Fabaceae	AM	41	15
Liana	*Martinella obovata* (Kunth) Bureau & K. Schum.	Bignoniaceae	AM,CA,CB	162	50
Liana	*Pleonotoma jasminifolia* (Kunth) Miers	Bignoniaceae	AM,CB	83	15
Liana	*Rourea cuspidata* Benth. ex Baker	Connaraceae	AM,CB	74	30
Monocot herb	*Goeppertia fragilis* (Gleason) Borchs. & S. Suárez	Marantaceae	AM	20	5
Monocot herb	*Goeppertia loeseneri* (J.F. Macbr.) Borchs. & S. Suárez	Marantaceae	AM	33	5
Monocot herb	*Goeppertia zingiberina* (Körn.) Borchs. & S. Suárez	Marantaceae	AM	69	20
Monocot herb	*Heliconia schumanniana* Loes.	Heliconiaceae	AM	93	10
Monocot herb	*Hylaeanthe hexantha* (Poepp. & Endl.) A.M.E. Jonker & Jonker	Marantaceae	AM	45	5
Monocot herb	*Ischnosiphon martianus* Eichler ex Petersen	Marantaceae	AM	28	30
Monocot herb	*Monotagma ulei* K. Schum.	Marantaceae	AM	30	5
Monocot herb	*Renealmia breviscapa* Poepp. & Endl.	Zingiberaceae	AM	80	50
Fern	*Adiantum pulverulentum* L.	Pteridaceae	AM,CA,CB,AT	257	50
Fern	*Adiantum tomentosum* Klotzsch	Pteridaceae	AM	120	40
Fern	*Cyathea pungens* (Willd.) Domin	Cyatheaceae	AM,CA,CB,AT	168	50
Fern	*Cyclopeltis semicordata* (Sw.) J. Sm.	Lomariopsidaceae	AM,CA	190	50
Fern	*Lindsaea guianensis* (Aubl.) Dryand.	Lindsaeaceae	AM,CA,CB,AT	100	50
Fern	*Schizaea elegans* (Vahl) Sw.	Schizaeaceae	AM,CA,CB,AT	199	50

FIGURES

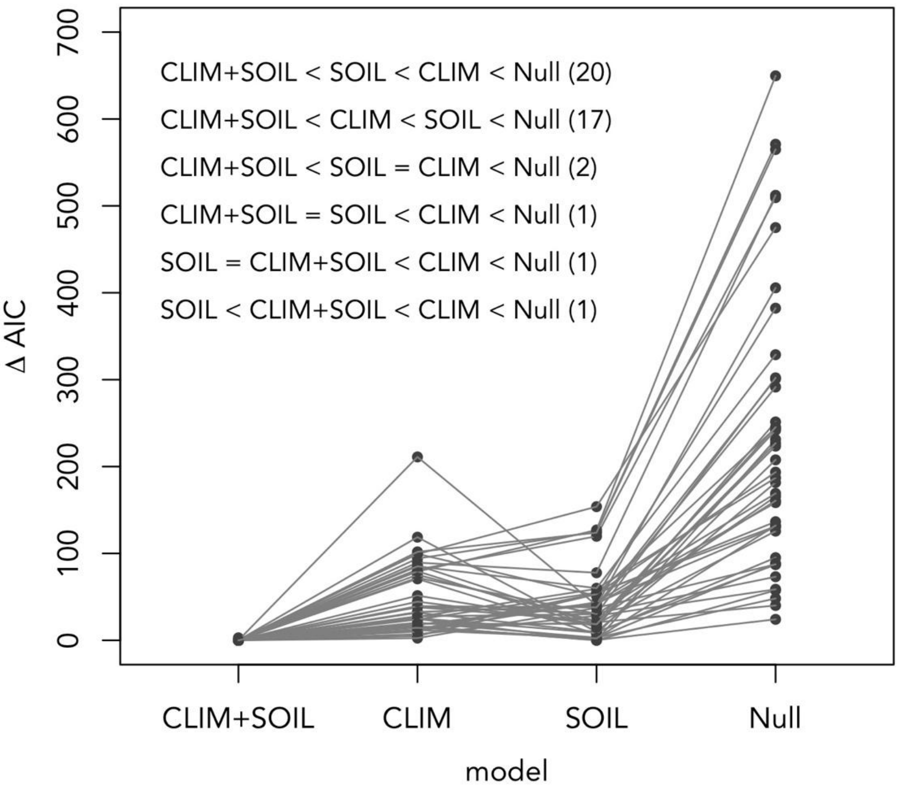


Fig 1. Performance of CLIM+SOIL, CLIM and SOIL species distribution models of 42 plant species in Amazonia. The ranking of model performance is based on ΔAIC scores, and the number of species within each ranking class is shown in parentheses. The "=" indicates a difference smaller than 2 (no evidence that the model with lower ΔAIC is better) and "<" indicates a difference greater than 2 (evidence that model with lower ΔAIC is better). ΔAIC values for each species separately are presented in Appendix S3.

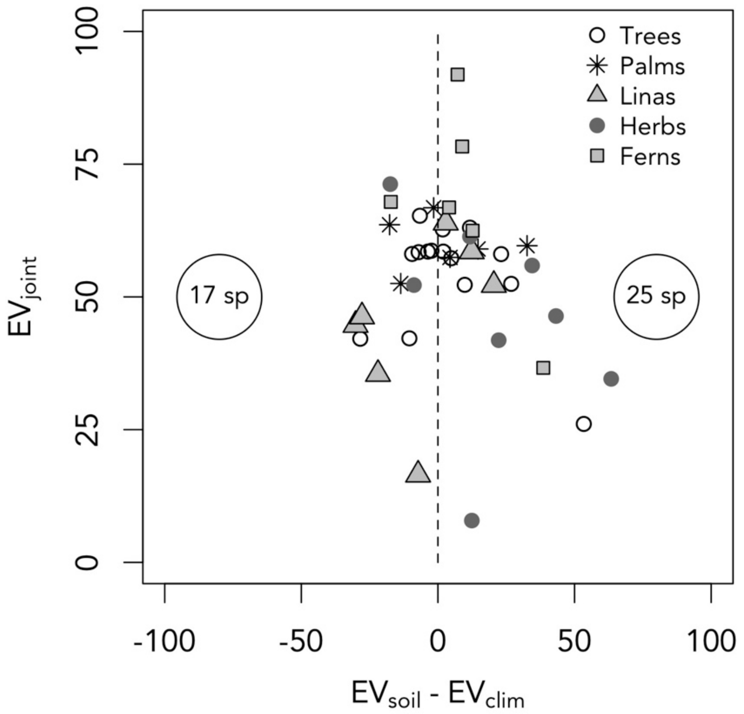


Fig. 2. Contribution to explained variance (EV) of the unique and joint fractions of the soil and climate models for each of 42 Amazonian plant species (species of different life-forms are shown with different symbols). EV_{Joint} represents the *EV* that is jointly explained by soil and climate variables. Horizontal axis represents the difference between the unique contributions of the soil model (EV_{SOIL}) and climate model (EV_{CLIM}). Species for which the unique contribution of the soil model was higher than the unique contribution of the climate model are found to the right of the vertical midline.

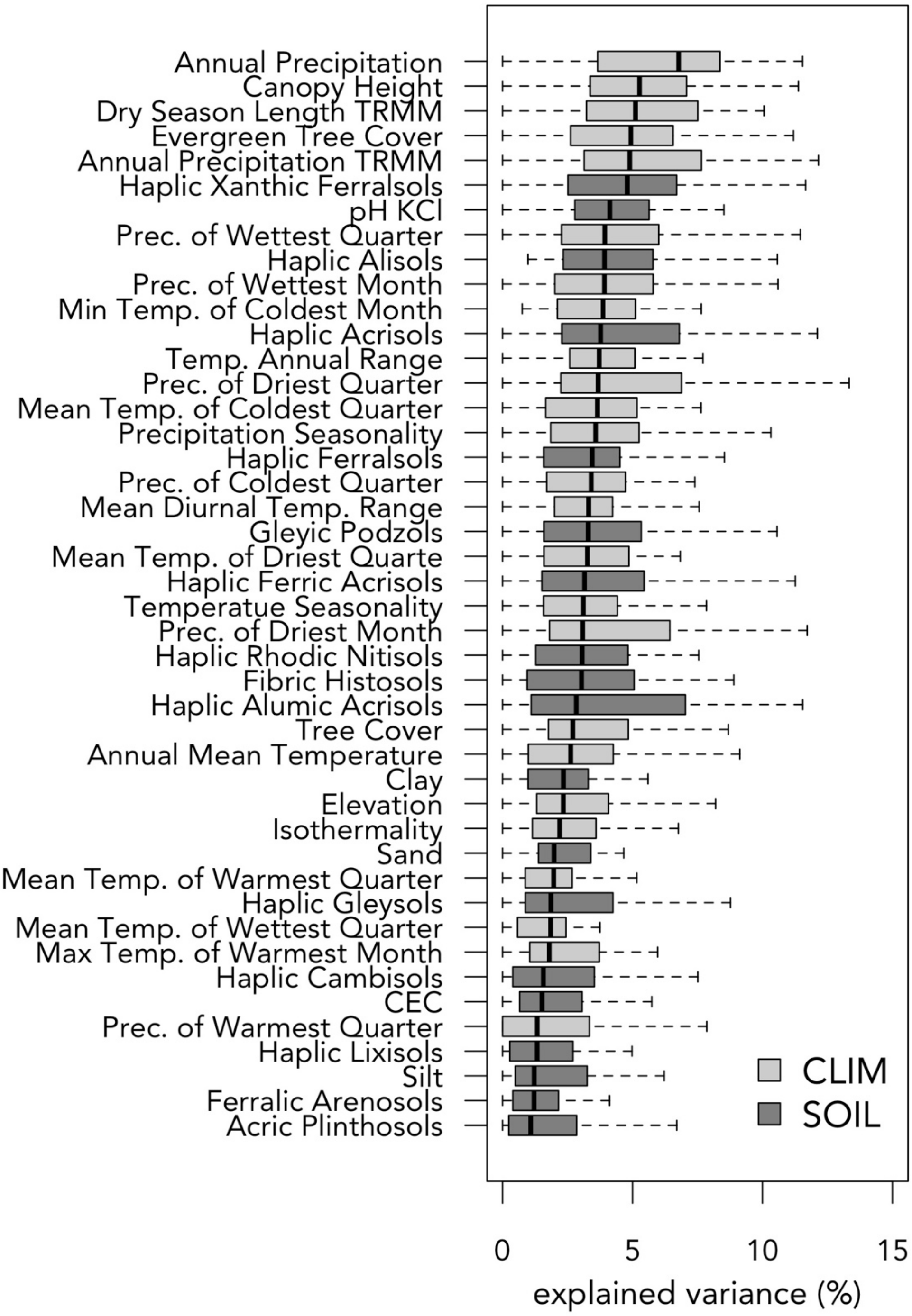


Fig. 3 Individual variables ranked by their importance in SDMs of 42 plant species in Amazonia. Importance is defined as the median value across all species of the explained variance (EV) in single-variable models, expressed as a percentage of the EV in the corresponding best model. Variables are coloured according to whether they were used in the CLIM models or in the SOIL models.

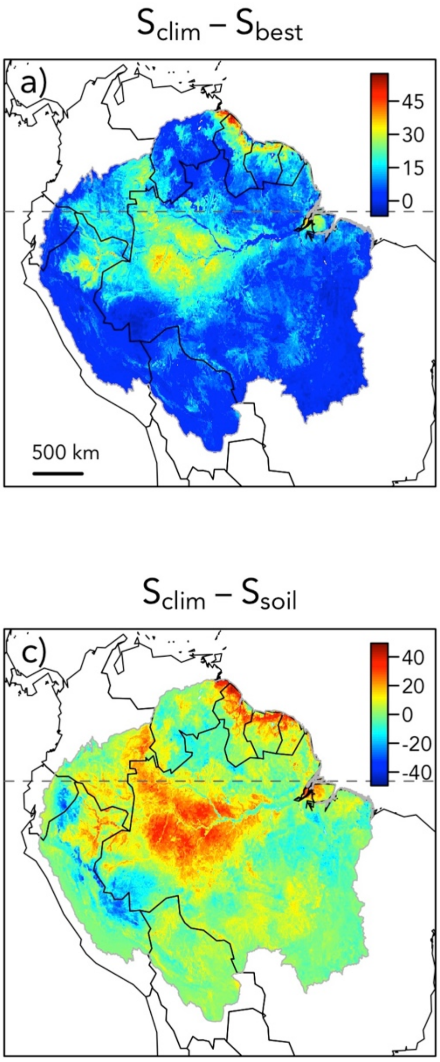


Fig. 4 Spatial projection of the differences in suitability scores between the SDMs, mapped only in areas defined as unsuitable to species occurrence (averages across 42 plant species). Suitability scores range from 0 (low probability of species occurrence) to 100 (high probability of occurrence). a) The difference in suitability between the best and CLIM models ($S_{clim} - S_{best}$). b). The difference in suitability between the the best and SOIL models ($S_{soil} - S_{best}$). c) The Difference in suitability between the CLIM and SOIL models ($S_{clim} - S_{soil}$). High values in panels a) and b) denote regions where the CLIM and SOIL models over-predict the suitable conditions relative to the best models, respectively. In panel c) values close to zero (green) indicate areas where the CLIM and SOIL models jointly predict low suitable condition; negative values (blue) denote areas where the CLIM models predict low suitable condition, but the SOIL models over-predict and the positive values (red) means the opposite, so blue indicates where climate is the main restricting factor and red indicates where soil is

the main restricting factor. Note that for 42 species, the best models included both climatic and soil variables, and for the remaining two species, the best models included only soil variables. Maps are in Mollweide equal-area projection and the dashed line denotes the Equator.

ARTIGO 1 – MATERIAL SUPLEMENTAR

Journal of Biogeography

SUPPORTING INFORMATION

Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species

Fernando O. G. Figueiredo[1,2*], Gabriela Zuquim[3], Hanna Tuomisto[3], Gabriel M. Moulatlet[3],
Henrik Balslev[4], and Flávia R. C. Costa[2]

Appendix S1. Species occurrence data sources (Table S1) and maps of the 44 environmental layers used in SDMs (Fig. S1.1).

Species occurrence data sources

Species occurrence records were obtained from two complementary sources: plot-based inventories and herbarium records, for which most of the data is available on the web. Details about plot-based sampling can be found in: Emilio *et al.* 2010 (trees); Tuomisto *et al.* 2016 (palms); Costa *et al.* 2006, Figueiredo *et al.* 2014 and Moulatlet *et al.* 2014 (monocot herbs); Tuomisto *et al.* 2014, Zuquim *et al.* 2014, and Tuomisto *et al.* 2016 (ferns). Reference are provided in end of Appendix S1. The local of data assessed are given in the Table S1.

Table S1. Species occurrence data sources.

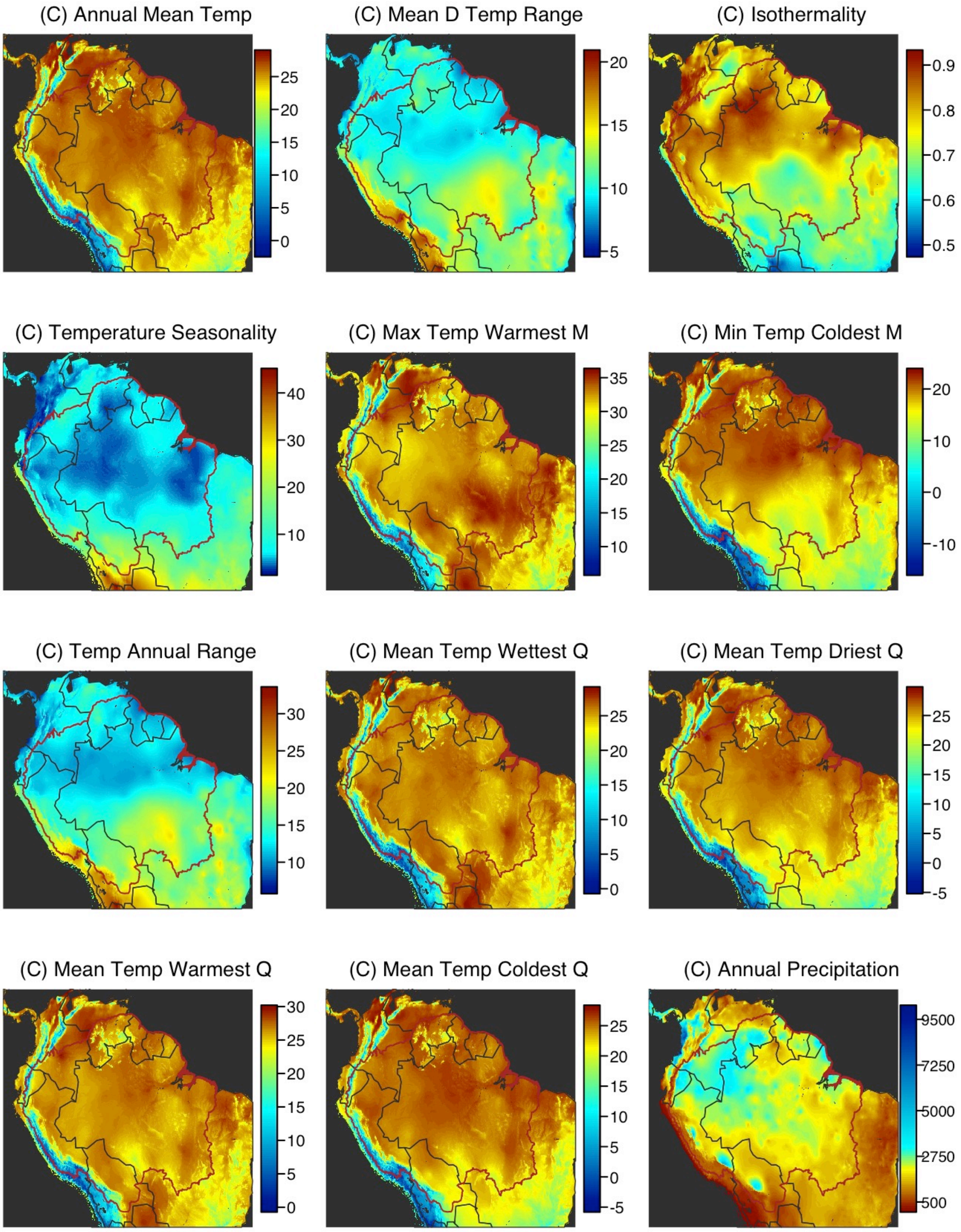
Data type	Data source	Local accessed	Plant group
Herbarium records	Global Biodiversity Information Facility (GBIF)	www.gbif.org	trees, palms[*], lianas, monocot herbs, ferns
	Integrated Digitized Biocollections (iDigBio)	https://www.idigbio.org	trees, palms[*], lianas, ferns
	SpeciesLink	http://splink.cria.org.br	monocot herbs
	Herbario do Instituto Nacional de Pesquisas da Amazônia (INPA)	http://herbaria.plants.ox.ac.uk/bol/brahms	monocot herbs
	Herbario IAN – Embrapa Amazônia Oriental (IAN)		
	Herbario do Museu Paraense Emílio Goeldi (MG)		
	New York Botanical Garden's Virtual Herbarium (NY)	http://sciweb.nybg.org/science2	monocot herbs
	Missouri Botanical Garden (MBG)	www.tropicos.org	monocot herbs[†]
Plot-based inventory	Projeto RADAM Brasil	https://archive.org/download/ftp-geoftp.ibge.gov.br 201401	trees
	Brazilian Program for Biodiversity Research (PPBio)	https://ppbio.inpa.gov.br/	monocot herbs[§], ferns
	Amazon Research Team – University of Turku	http://www.utu.fi/en/sites/amazon (not avaliable on web)	ferns

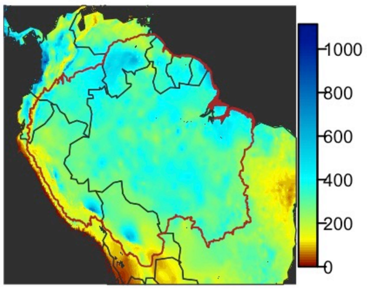
* include also occurrence data recorded in systematic plot-based inventories.

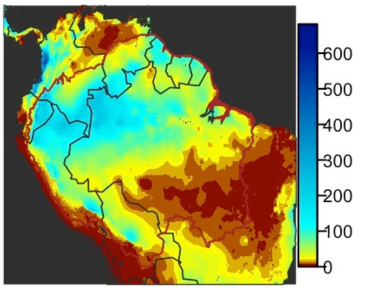
§ include published and non-published data.

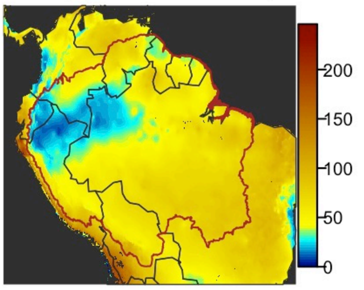
† provided directly by the institution due to large amount of data.

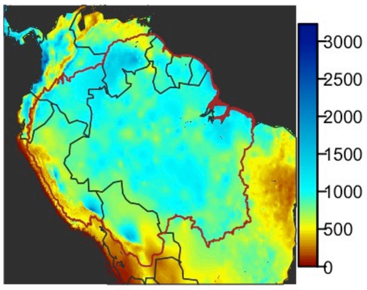
FS1.1

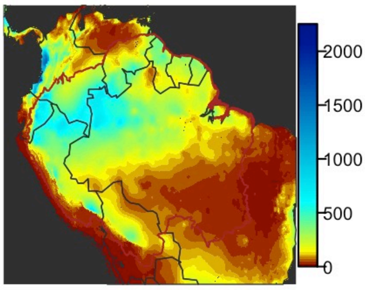

(C) Annual Mean Temp
25
20
15
10
5
0
(C) Mean D Temp Range
20
15
10
5
(C) Isothermality
0.9
0.8
0.7
0.6
0.5
(C) Temperature Seasonality
40
30
20
10
(C) Max Temp Warmest M
35
30
25
20
15
10
(C) Min Temp Coldest M
20
10
0
-10
(C) Temp Annual Range
30
25
20
15
10
(C) Mean Temp Wettest Q
25
20
15
10
5
0
(C) Mean Temp Driest Q
25
20
15
10
5
0
-5
(C) Mean Temp Warmest Q
30
25
20
15
10
5
0
(C) Mean Temp Coldest Q
25
20
15
10
5
0
-5
(C) Annual Precipitation
9500
7250
5000
2750
500

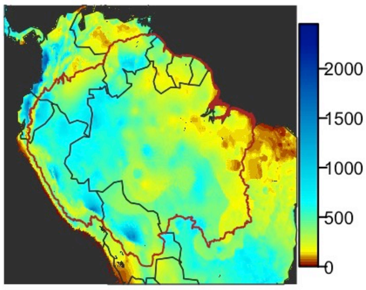

(C) Prec Wettest M

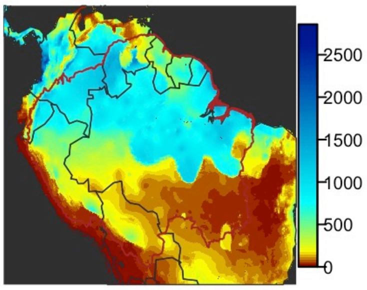

(C) Prec Driest M

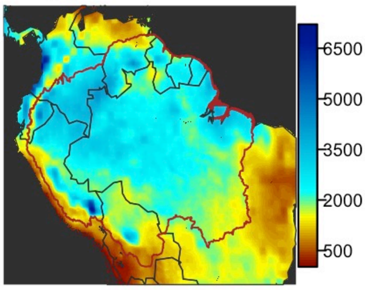

(C) Prec Seasonality

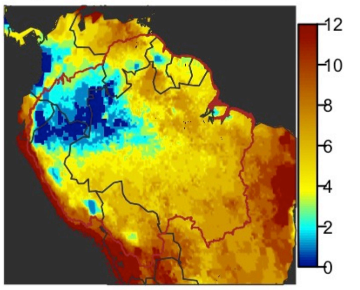

(C) Prec Wettest Q

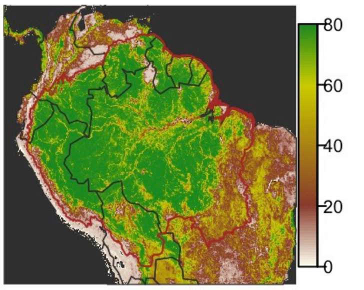

(C) Prec Driest Q

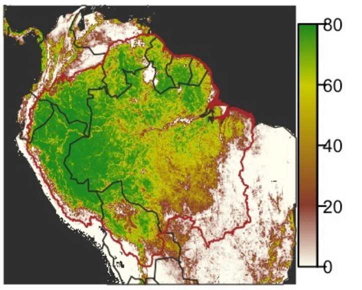

(C) Prec Warmest Q

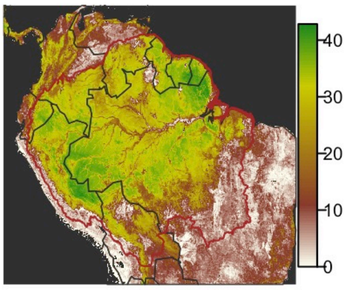

(C) Prec Coldest Q

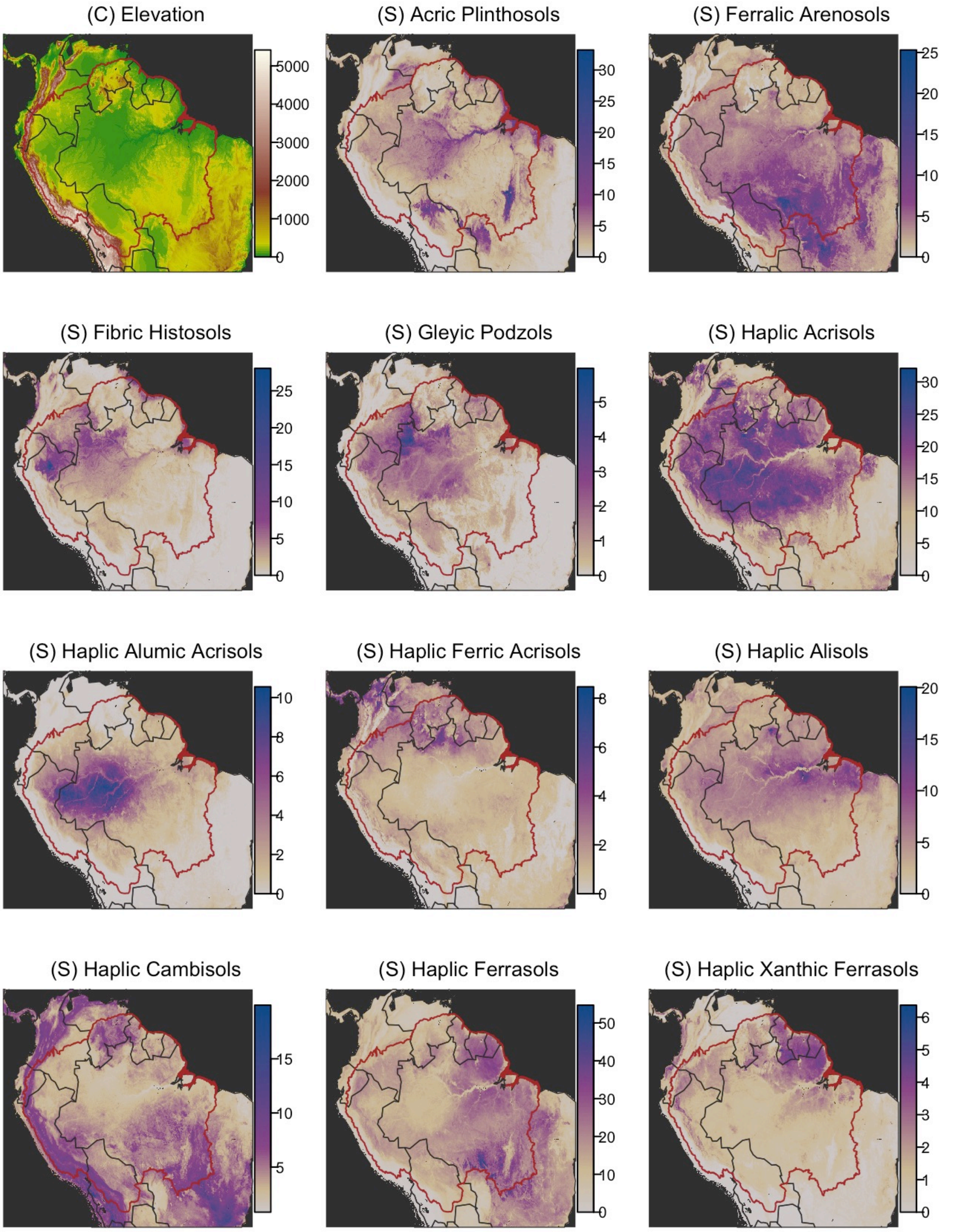

(C) Annual Prec TRMM


(C) Dry Season Length TRMM


(C) Tree Cover


(C) Evergreen Tree Cover


(C) Canopy Height


(C) Elevation
5000
4000
3000
2000
1000
0
(S) Acric Plinthosols
30
25
20
15
10
5
0
(S) Ferralic Arenosols
25
20
15
10
5
0
(S) Fibric Histosols
25
20
15
10
5
0
(S) Gleyic Podzols
5
4
3
2
1
0
(S) Haplic Acrisols
30
25
20
15
10
5
0
(S) Haplic Alumic Acrisols
10
8
6
4
2
0
(S) Haplic Ferric Acrisols
8
6
4
2
0
(S) Haplic Alisols
20
15
10
5
0
(S) Haplic Cambisols
15
10
5
(S) Haplic Ferrasols
50
40
30
20
10
0
(S) Haplic Xanthic Ferrasols
6
5
4
3
2
1
0

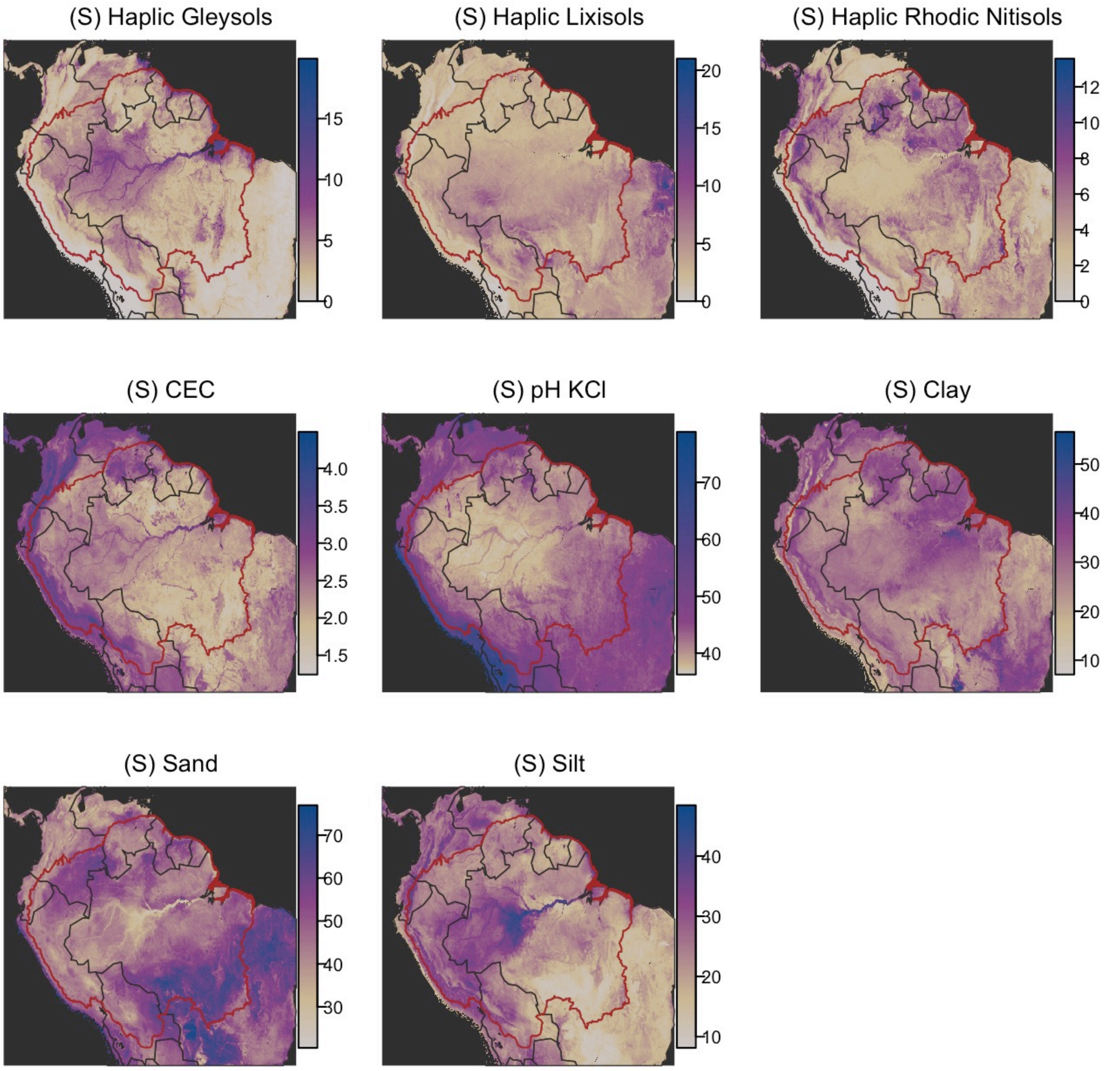


Figure S1.1 Map of the 44 environmental layers used in SDMs with boundaries of Amazonian delineated in red. (C) and (S) are variables used in the CLIM and SOIL model, respectively. D, M and Q denote abbreviations for diurnal, month and quarter, respectively.

References

Costa F.R.C., Magnusson W.E., & Luizao R.C. (2005) Mesoscale distribution patterns of Amazonian understorey herbs in relation to topography, soil and watersheds. *Journal of Ecology*, **93**, 863–878.

Emilio T., Nelson B.W., Schietti J., Desmoulière S.J.-M., Espírito Santo H.M. V., & Costa F.R.C. (2010) Assessing the relationship between forest types and canopy tree beta diversity in Amazonia. *Ecography*, **33**, 738–747.

Figueiredo F.O.G., Costa F.R.C., Nelson B.W., & Pimentel T.P. (2014) Validating forest types based on geological and land-form features in central Amazonia. *Journal of Vegetation Science*, **25**, 198–212.

Moulatlet G.M., Costa F.R.C., Rennó C.D., Emilio T., & Schietti J. (2014) Local hydrological conditions explain floristic composition in lowland Amazonian forests. *Biotropica*, **46**, 395–403.

Tuomisto H., Zuquim G., & Cárdenas G. (2014) Species richness and diversity along edaphic and climatic gradients in Amazonia. *Ecography*, 1–13.

Tuomisto. H., Moulatlet, G., Balslev, H., Emilio, T., Figueiredo, O.G.F., Pedersen, D. & Ruokolainen, K. (2016) A compositional turnover zone of biogeographical

Appendix S2. Description of the pre-modelling procedures: detection of outliers, reductions of sample bias (Fig. S2.2) and definitions of geographical background.

Detecting environmental and geographical outliers

For each species dataset, we reduced the multiple number of occurrences to one record per 5 km^2, i.e., the spatial resolution of our study. Since our SDM only includes lowland-forest species we used SRTM and Tree Cover layer to remove records that were located in high altitudes and non-forest habitats. Environmental outliers were detected and removed using the reverse jackknife algorithm (Chapman, 2005) implemented in Biogeo R package (Robertson *et al.*, 2016). For all plant groups, geographical outliers were visually identified comparing the species range maps published in Monographs (when available) and the distribution of our presence points. Just for the palm *Socratea exorrhiza* some occurrence records were identified as outliers (presence records located in Atlantic rainforest biome), and then were manually removed.

Spatial filtering to reduce sample bias in presence records

We employed a spatial filter procedure to reduce spatial bias in presence records. Spatial filtering consists in removing presence records that are within a predefined spatial radius, hereafter called filtering distance (Boria *et al.*, 2014). Recent studies advocated that presence-only records of species occurrence should be viewed as a particular realization of a spatial point process, the inhomogeneous Poisson process (Renner *et al.*, 2015). The *K* Ripley's function and its modified version, *L* function, provide a useful statistic to describe inter-point spatial dependence and clustering. Here we employed the inhomogeneous version of the *L* function (Baddeley *et al.*, 2000) to objectively choose filtering distance. To choose filtering distance for each species we set different filtering radius starting at 5 km and varying by 5 km until to 50 km. For each filtering distance we estimated the *L* function *Linhom(r)* of the observed point process. Observed *Linhom(r)* values greater than the estimated theoretical *Linhom(r)* are indicative of spatial clustering. We set the filtering distance when 95% of the estimated theoretical *Linhom(r)* values were higher than low band of the 95% bootstrap confidence interval of observed spatial point process. Below (Fig. S1) we present the *Linhom(r)* for each filtering distance for *Ischnosiphon martianus*. All the procedures were performed in R program. Duplicated points within the filtering distance were removed using the function *remove.duplicates* of *dismo* package; confidence bands and theoretical *Linhom(r)* were computed using the function *lohboot* of the *spatstat* package. In the function *lohboot* we set 'best' for the argument *correction* and 'TRUE' for the argument *global*.

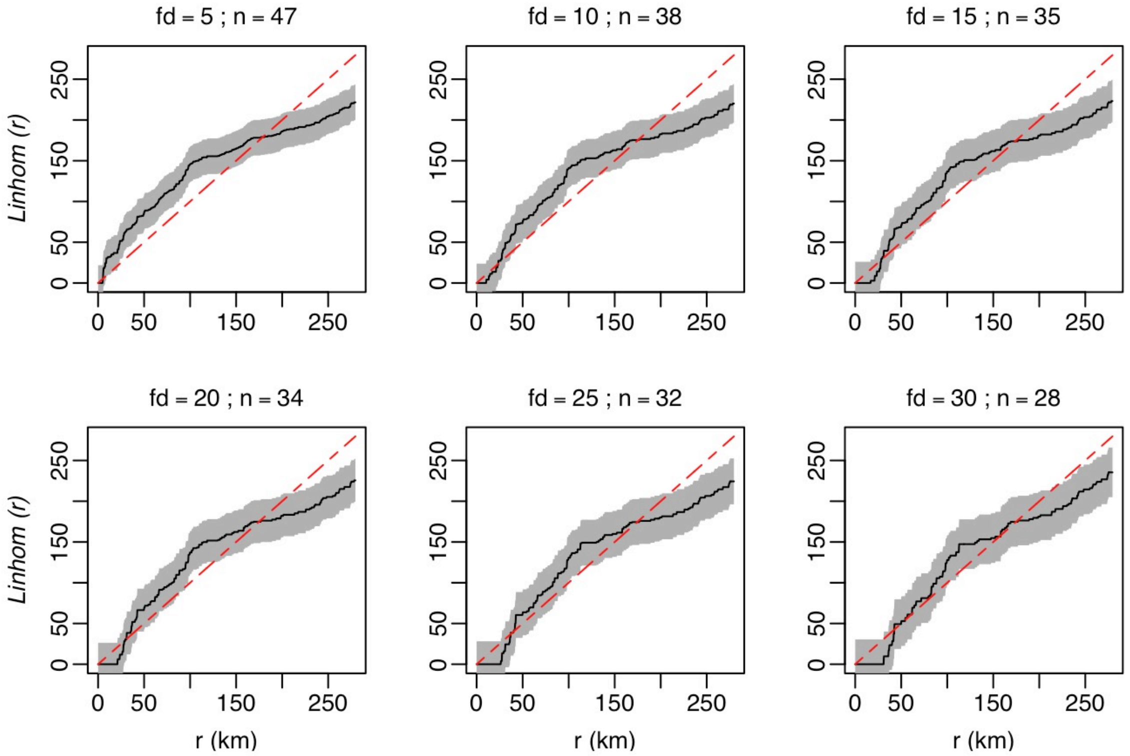


Figure S2.2 The *Linhom(r)* function computed for for different filtering (fd) distances for *Ischnosiphon martianus*. For this species, the selected filtering distance was 30 km, resulting in 30 presence records (n) free of spatial clustering. (*r*) is the radius distance setting by the algorithm at which *Linhom(r)* was estimated. The black line is the estimated *Linhom(r)* for the observed point process; grey bands represent the 95% bootstrap confidence interval for the observed *Linhom(r)*; the dotted red line is the estimated theoretical *Linhom(r)* assuming an inhomogeneous Poisson process.

Delimiting the area for model calibration, evaluation and prediction

To avoid over-predictions in species range, i.e., predict species to be present in areas far beyond their actual occurrence, we spatially restricted the area for model calibration, evaluation and prediction. For each species we used the median geographical distance of filtered occurrence records as the radius buffer and set the geographic background as being the smallest convex polygon around buffer geometry. By doing that, we adjusted the size of background area with the size of the species range. Radius buffer was allowed to vary between a minimum and maximum distance of 500 km and 1000 km, respectively. See in Appendix S3 the background areas set for each species.

References

Chapman, A. D. (2005) Uses of primary species-occurrence data, version 1.0. – Report for the Global Biodiversity Information Facility, Copenhagen.

Baddeley, A.J., Møller, J. & Waagepetersen, R. (2000) Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, **54**, 329–350.

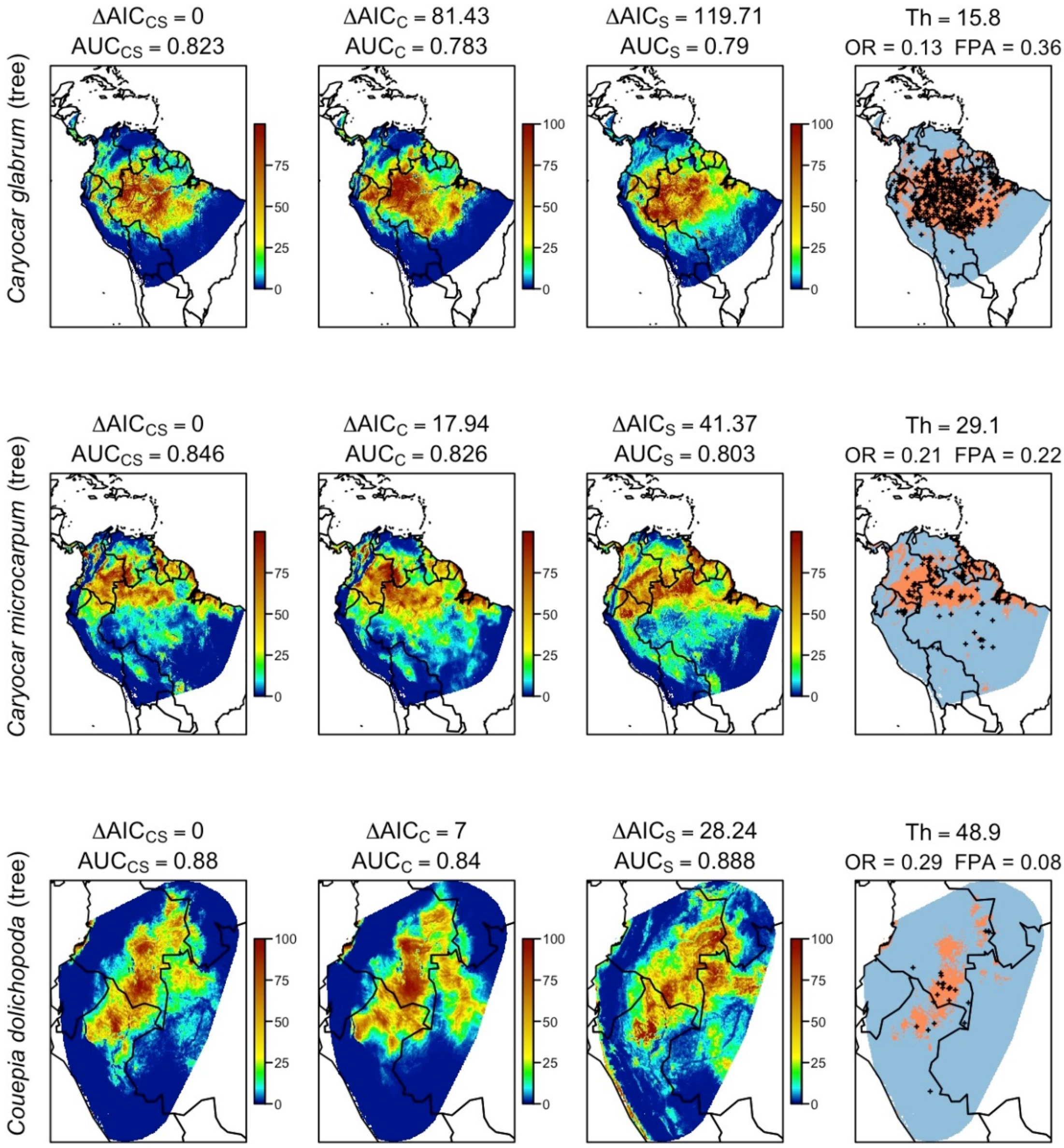
Boria, R.A., Olson, L.E., Goodman, S.M. & Anderson, R.P. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, **275**, 73–77.

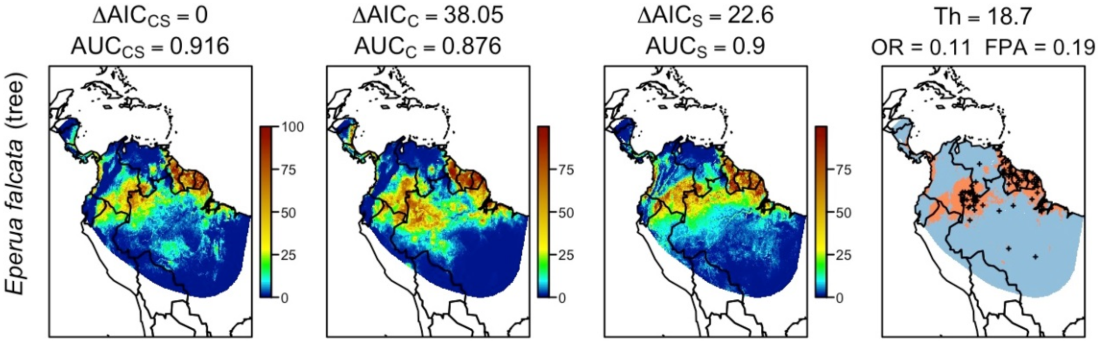
Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S., Popovic, G. & Warton, D.I. (2015) Point process models for presence-only analysis – a review. *Methods in Ecology and Evolution*, **6**, 366–379.

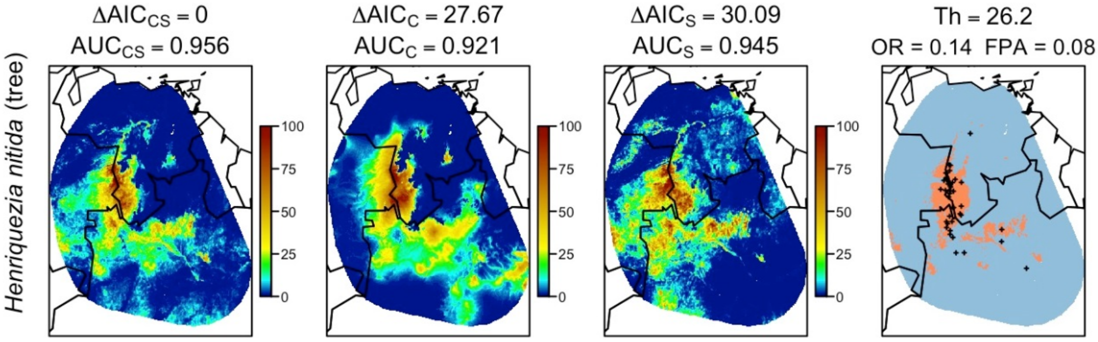
Robertson M.P., Visser V., & Hui C. (2016) Biogeo : an R package for assessing and improving data quality of occurrence record datasets. *Ecography*, **39**, 394–401.

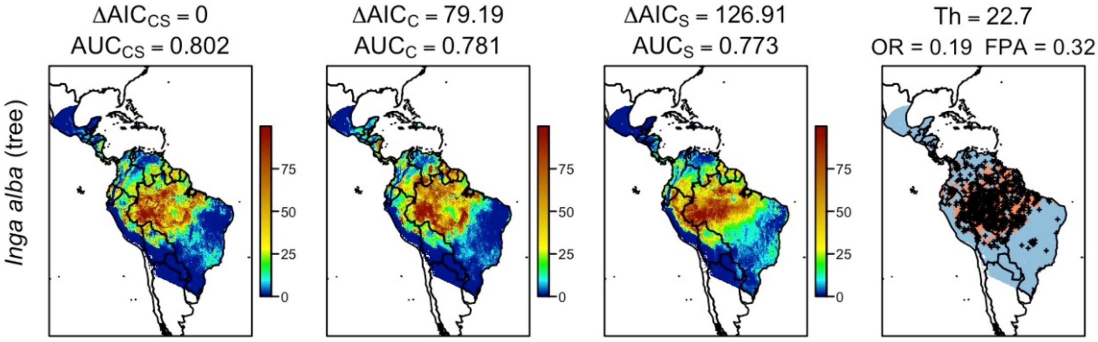
Appendix S3.3 Spatial projections of SDM for the 42 Amazonian plant species (Fig. S3.3) and the predicted species response curves along each of the environmental layers (Fig. S3.4)

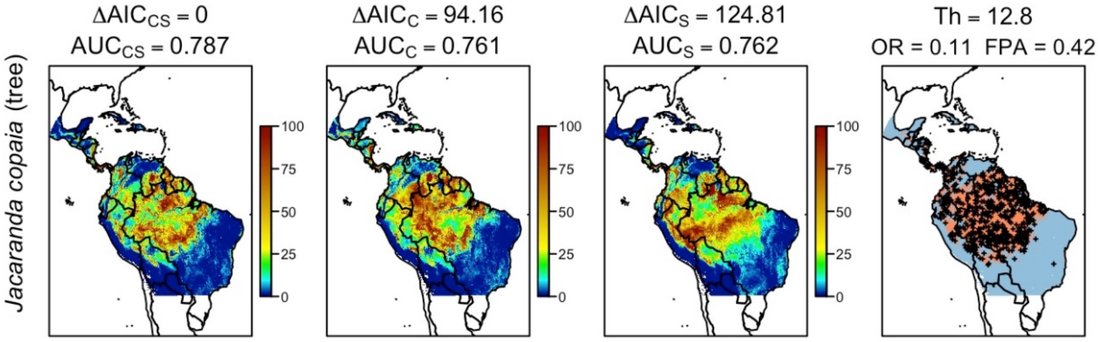
Fig. S3.3

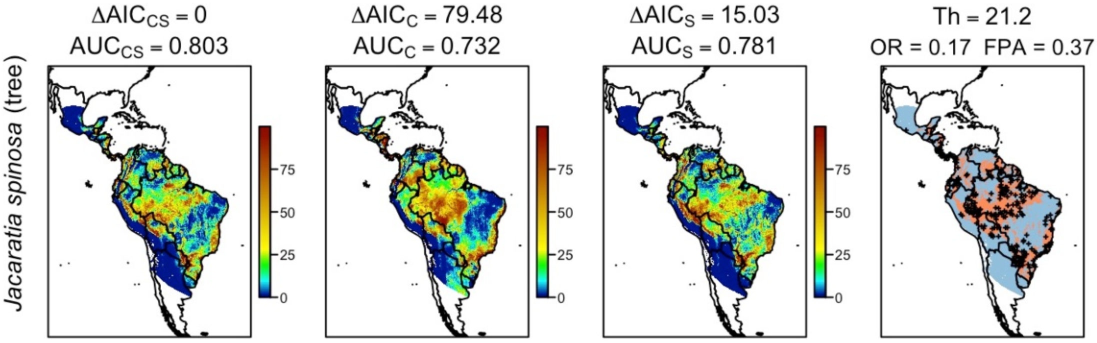


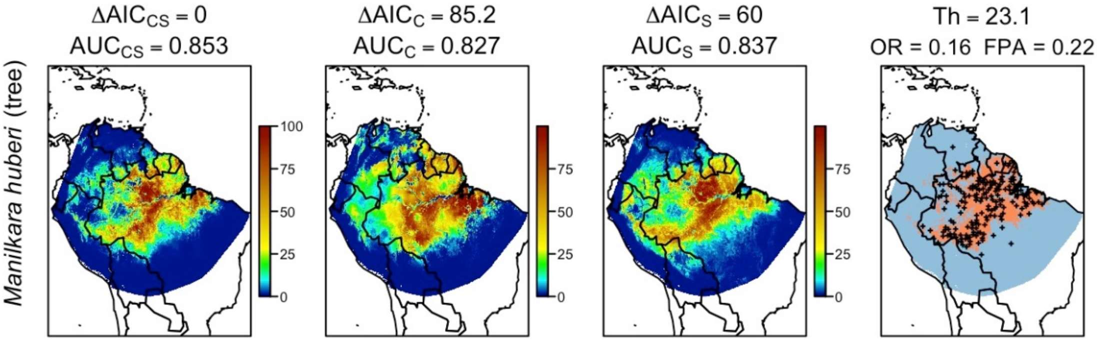

Eperua falcata (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.916$
$\Delta AIC_{C} = 38.05$
$AUC_{C} = 0.876$
$\Delta AIC_{S} = 22.6$
$AUC_{S} = 0.9$
Th = 18.7
OR = 0.11 FPA = 0.19
100
75
50
25
0

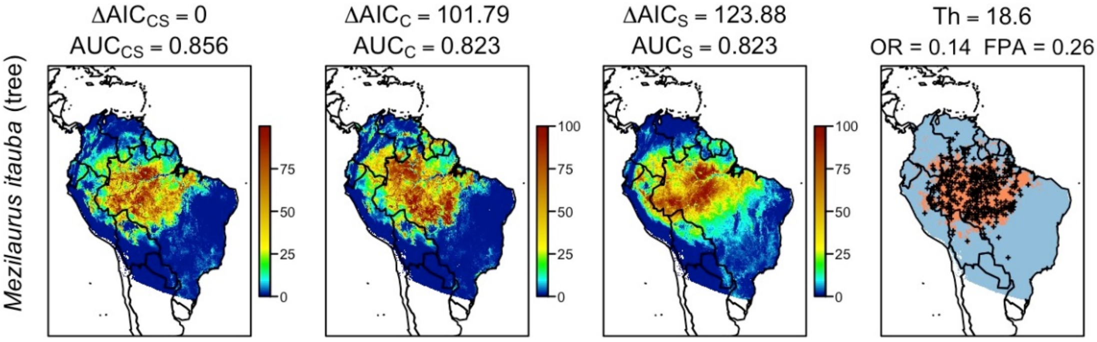

Henriquezia nitida (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.956$
$\Delta AIC_{C} = 27.67$
$AUC_{C} = 0.921$
$\Delta AIC_{S} = 30.09$
$AUC_{S} = 0.945$
Th = 26.2
OR = 0.14 FPA = 0.08
100
75
50
25
0

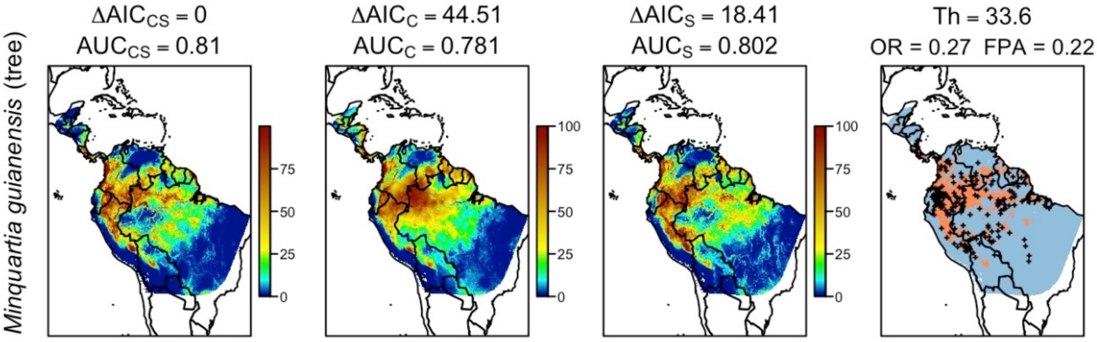

Inga alba (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.802$
$\Delta AIC_{C} = 79.19$
$AUC_{C} = 0.781$
$\Delta AIC_{S} = 126.91$
$AUC_{S} = 0.773$
Th = 22.7
OR = 0.19 FPA = 0.32
75
50
25
0

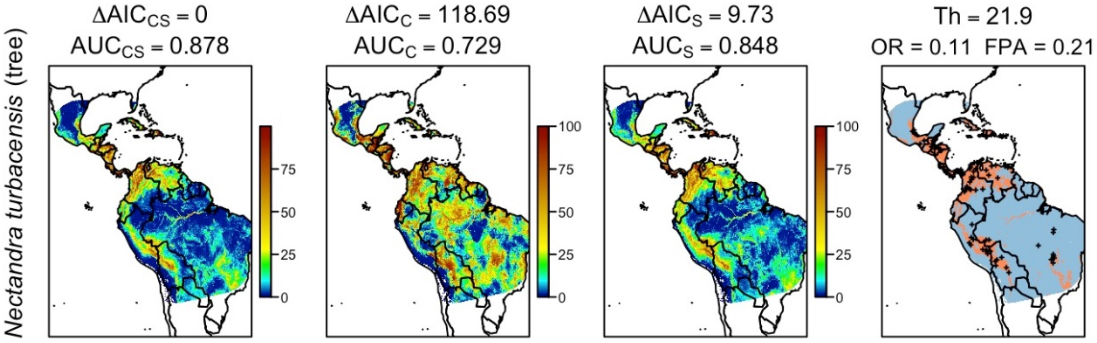

Jacaranda copaia (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.787$
$\Delta AIC_{C} = 94.16$
$AUC_{C} = 0.761$
$\Delta AIC_{S} = 124.81$
$AUC_{S} = 0.762$
Th = 12.8
OR = 0.11 FPA = 0.42
100
75
50
25
0

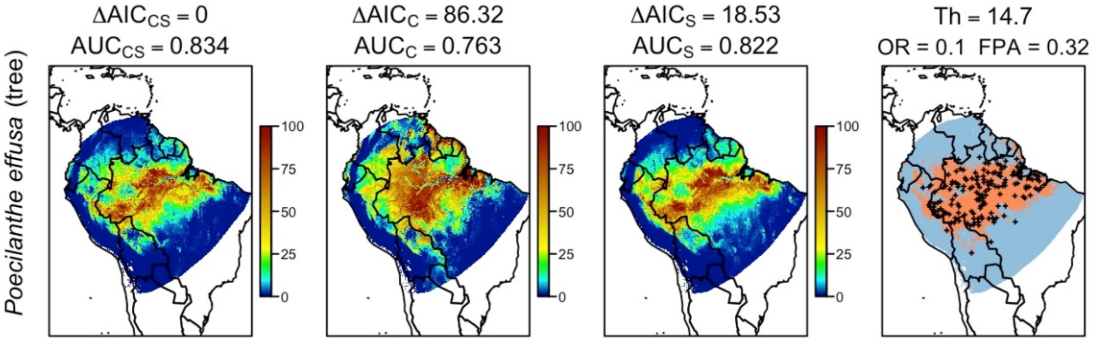

Jacaratia spinosa (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.803$
$\Delta AIC_{C} = 79.48$
$AUC_{C} = 0.732$
$\Delta AIC_{S} = 15.03$
$AUC_{S} = 0.781$
Th = 21.2
OR = 0.17 FPA = 0.37
75
50
25
0

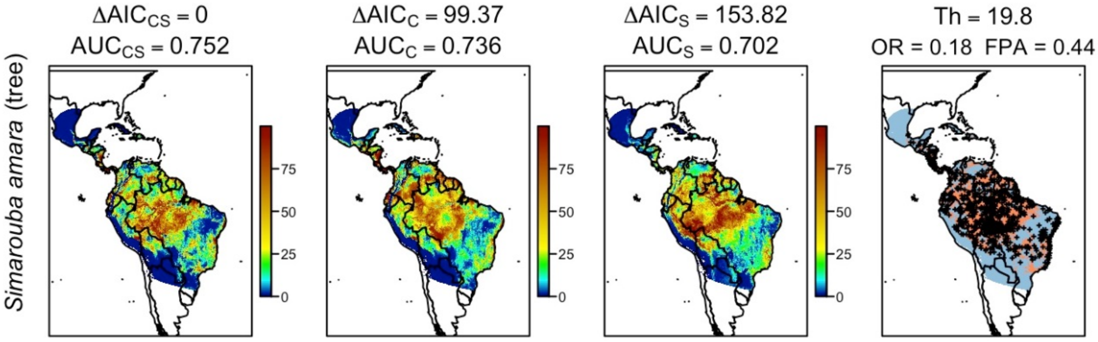

Manilkara huberi (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.853$
$\Delta AIC_{C} = 85.2$
$AUC_{C} = 0.827$
$\Delta AIC_{S} = 60$
$AUC_{S} = 0.837$
Th = 23.1
OR = 0.16 FPA = 0.22
100
75
50
25
0

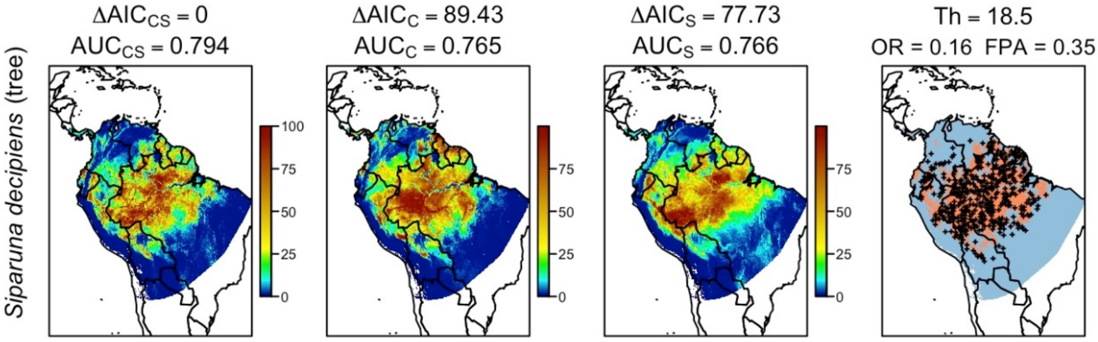

Mezilaurus itauba (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.856$
$\Delta AIC_{C} = 101.79$
$AUC_{C} = 0.823$
$\Delta AIC_{S} = 123.88$
$AUC_{S} = 0.823$
Th = 18.6
OR = 0.14 FPA = 0.26
100
75
50
25
0

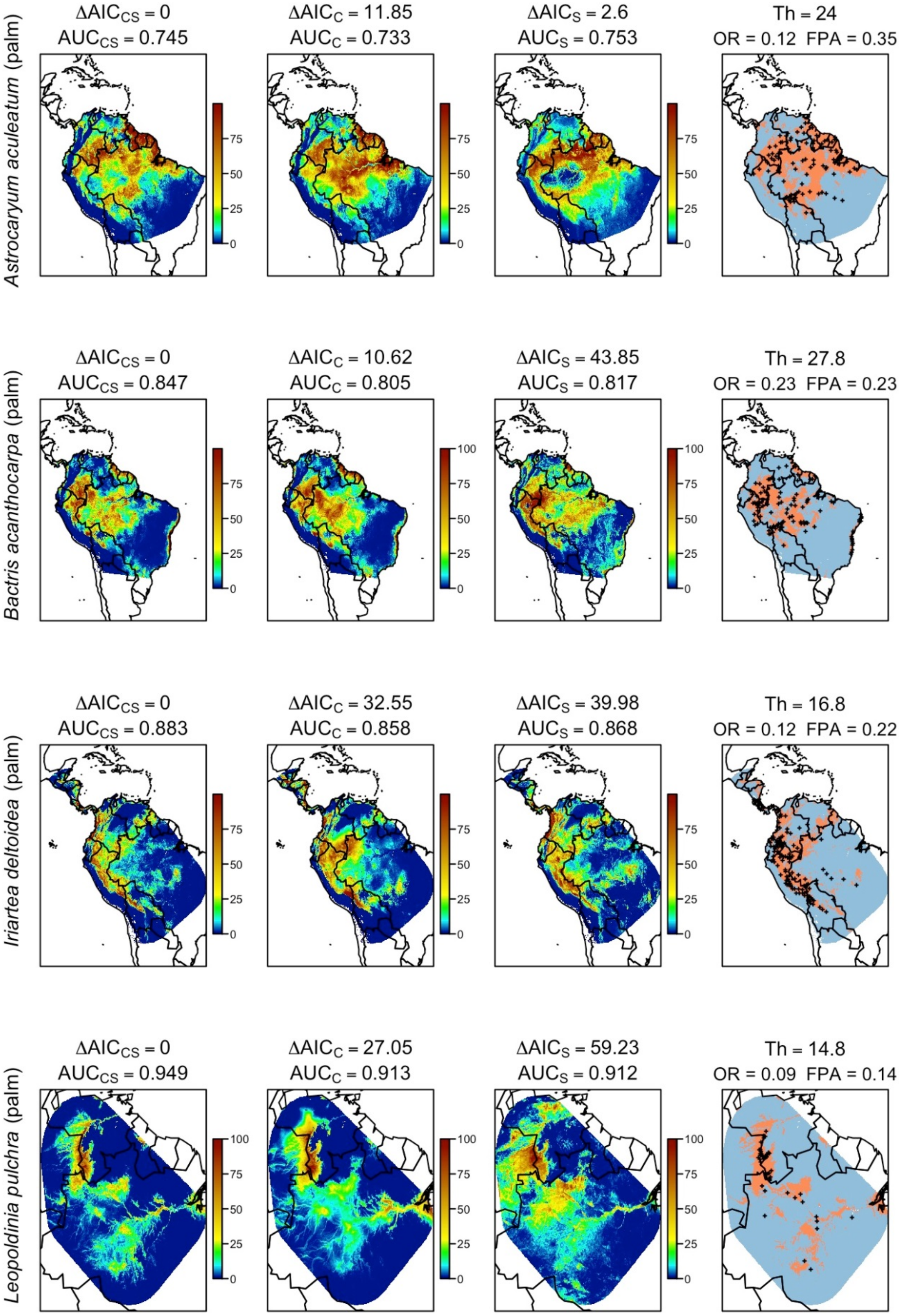

Minquartia guianensis (tree)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.81$
$\Delta AIC_{C} = 44.51$
$AUC_{C} = 0.781$
$\Delta AIC_{S} = 18.41$
$AUC_{S} = 0.802$
Th = 33.6
OR = 0.27 FPA = 0.22
100
75
50
25
0

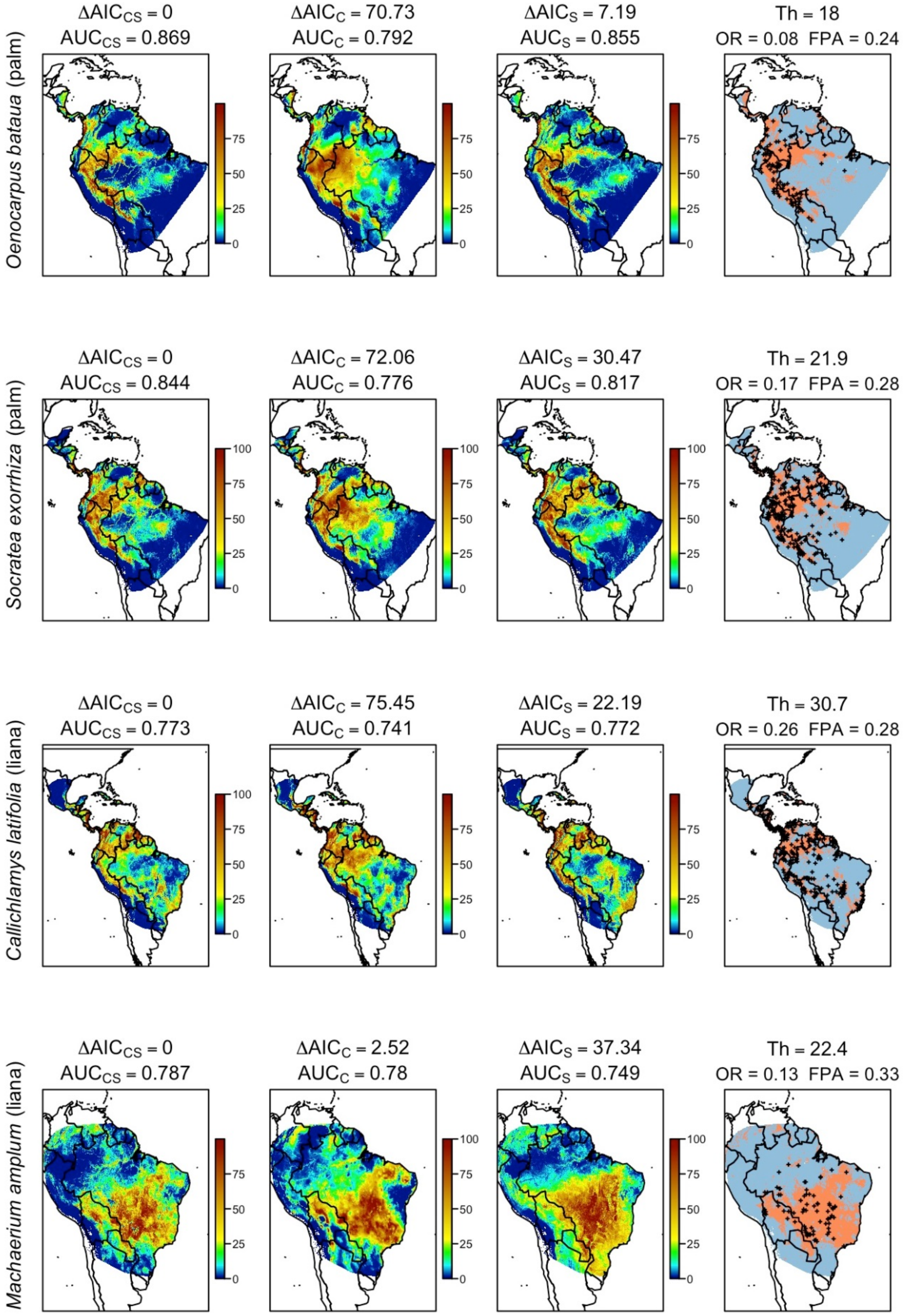

Nectandra turbacensis (tree)
ΔAIC$_{CS}$ = 0
AUC$_{CS}$ = 0.878
75
50
25
0
ΔAIC$_{C}$ = 118.69
AUC$_{C}$ = 0.729
100
75
50
25
0
ΔAIC$_{S}$ = 9.73
AUC$_{S}$ = 0.848
100
75
50
25
0
Th = 21.9
OR = 0.11 FPA = 0.21

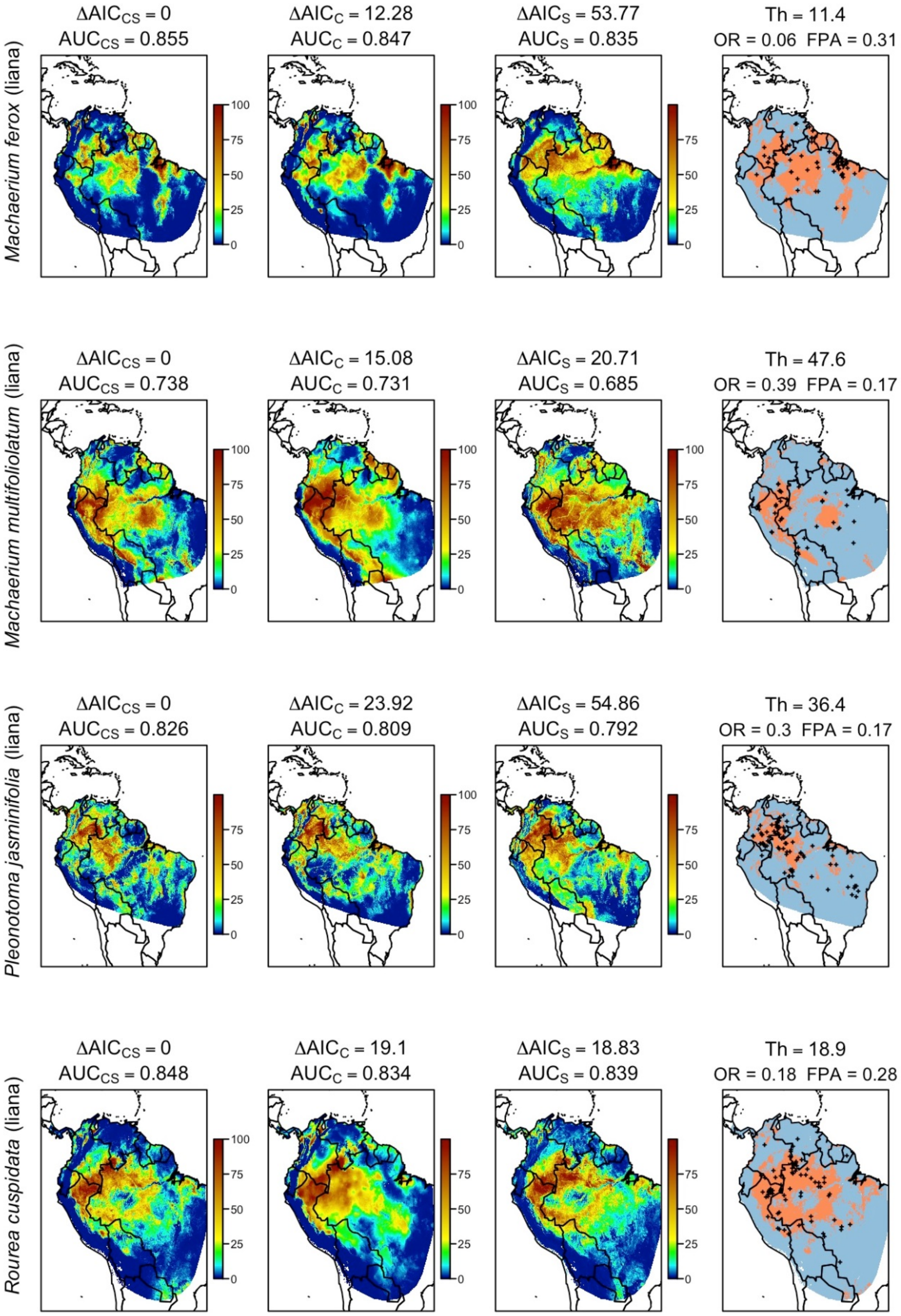

Poecilanthe effusa (tree)
ΔAIC$_{CS}$ = 0
AUC$_{CS}$ = 0.834
100
75
50
25
0
ΔAIC$_{C}$ = 86.32
AUC$_{C}$ = 0.763
100
75
50
25
0
ΔAIC$_{S}$ = 18.53
AUC$_{S}$ = 0.822
100
75
50
25
0
Th = 14.7
OR = 0.1 FPA = 0.32


Simarouba amara (tree)
ΔAIC$_{CS}$ = 0
AUC$_{CS}$ = 0.752
75
50
25
0
ΔAIC$_{C}$ = 99.37
AUC$_{C}$ = 0.736
75
50
25
0
ΔAIC$_{S}$ = 153.82
AUC$_{S}$ = 0.702
75
50
25
0
Th = 19.8
OR = 0.18 FPA = 0.44

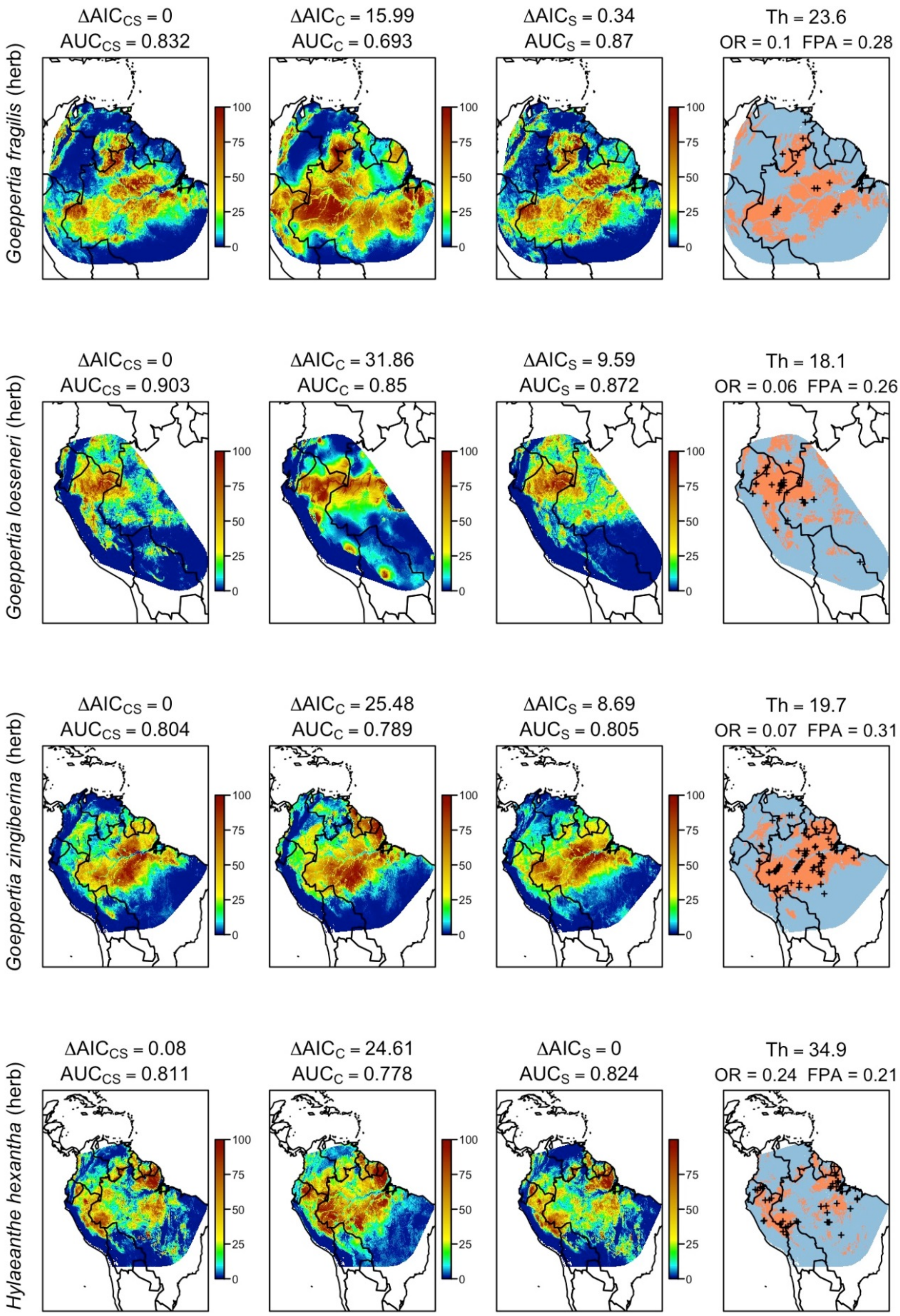

Siparuna decipiens (tree)
ΔAIC$_{CS}$ = 0
AUC$_{CS}$ = 0.794
100
75
50
25
0
ΔAIC$_{C}$ = 89.43
AUC$_{C}$ = 0.765
75
50
25
0
ΔAIC$_{S}$ = 77.73
AUC$_{S}$ = 0.766
75
50
25
0
Th = 18.5
OR = 0.16 FPA = 0.35

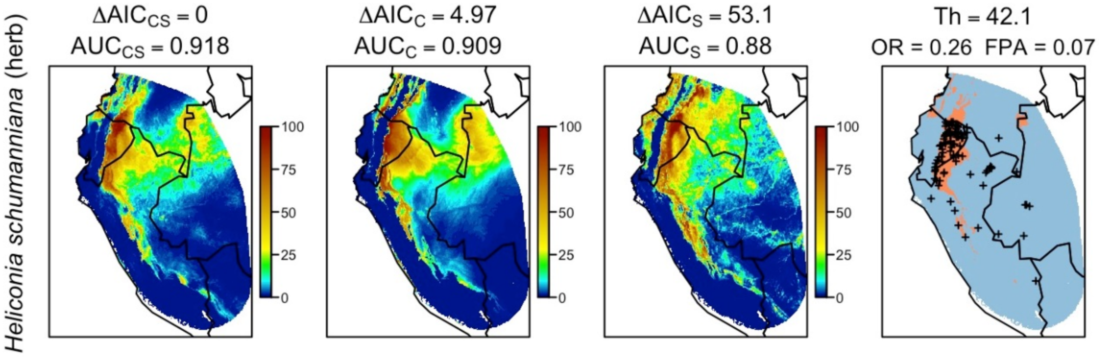

Astrocaryum aculeatum (palm)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.745$
$\Delta AIC_{C} = 11.85$
$AUC_{C} = 0.733$
$\Delta AIC_{S} = 2.6$
$AUC_{S} = 0.753$
Th = 24
OR = 0.12 FPA = 0.35
Bactris acanthocarpa (palm)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.847$
$\Delta AIC_{C} = 10.62$
$AUC_{C} = 0.805$
$\Delta AIC_{S} = 43.85$
$AUC_{S} = 0.817$
Th = 27.8
OR = 0.23 FPA = 0.23
Iriartea deltoidea (palm)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.883$
$\Delta AIC_{C} = 32.55$
$AUC_{C} = 0.858$
$\Delta AIC_{S} = 39.98$
$AUC_{S} = 0.868$
Th = 16.8
OR = 0.12 FPA = 0.22
Leopoldinia pulchra (palm)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.949$
$\Delta AIC_{C} = 27.05$
$AUC_{C} = 0.913$
$\Delta AIC_{S} = 59.23$
$AUC_{S} = 0.912$
Th = 14.8
OR = 0.09 FPA = 0.14
0
25
50
75
100

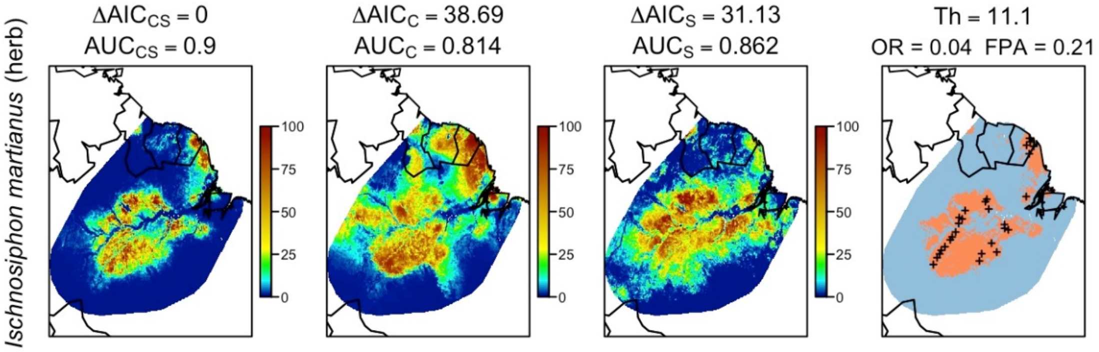

Oenocarpus bataua (palm)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.869$
$\Delta AIC_{C} = 70.73$
$AUC_{C} = 0.792$
$\Delta AIC_{S} = 7.19$
$AUC_{S} = 0.855$
Th = 18
OR = 0.08 FPA = 0.24
0
25
50
75
Socratea exorrhiza (palm)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.844$
$\Delta AIC_{C} = 72.06$
$AUC_{C} = 0.776$
$\Delta AIC_{S} = 30.47$
$AUC_{S} = 0.817$
Th = 21.9
OR = 0.17 FPA = 0.28
0
25
50
75
100
Callichlamys latifolia (liana)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.773$
$\Delta AIC_{C} = 75.45$
$AUC_{C} = 0.741$
$\Delta AIC_{S} = 22.19$
$AUC_{S} = 0.772$
Th = 30.7
OR = 0.26 FPA = 0.28
0
25
50
75
100
Machaerium amplum (liana)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.787$
$\Delta AIC_{C} = 2.52$
$AUC_{C} = 0.78$
$\Delta AIC_{S} = 37.34$
$AUC_{S} = 0.749$
Th = 22.4
OR = 0.13 FPA = 0.33
0
25
50
75
100

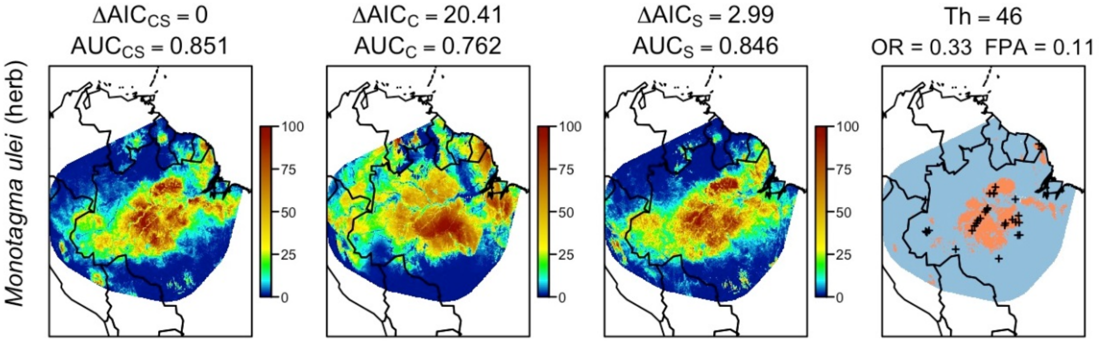


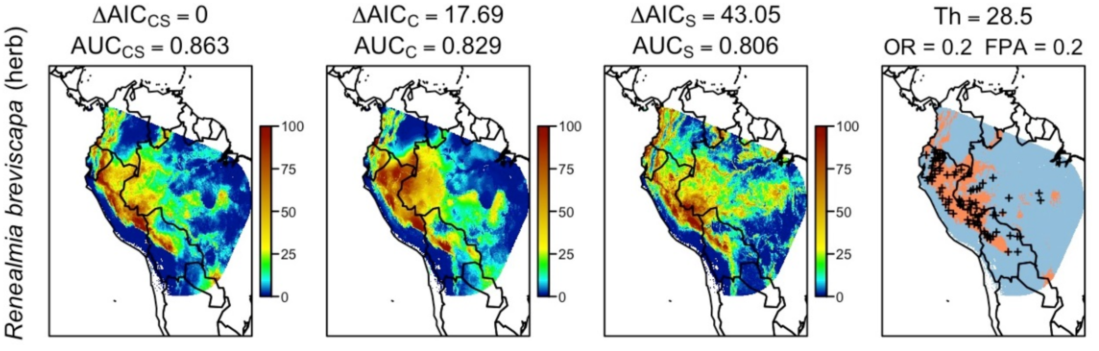
Machaerium ferox (liana)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.855$
$\Delta AIC_{C} = 12.28$
$AUC_{C} = 0.847$
$\Delta AIC_{S} = 53.77$
$AUC_{S} = 0.835$
Th = 11.4
OR = 0.06 FPA = 0.31
Machaerium multifoliolatum (liana)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.738$
$\Delta AIC_{C} = 15.08$
$AUC_{C} = 0.731$
$\Delta AIC_{S} = 20.71$
$AUC_{S} = 0.685$
Th = 47.6
OR = 0.39 FPA = 0.17
Pleonotoma jasminifolia (liana)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.826$
$\Delta AIC_{C} = 23.92$
$AUC_{C} = 0.809$
$\Delta AIC_{S} = 54.86$
$AUC_{S} = 0.792$
Th = 36.4
OR = 0.3 FPA = 0.17
Rourea cuspidata (liana)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.848$
$\Delta AIC_{C} = 19.1$
$AUC_{C} = 0.834$
$\Delta AIC_{S} = 18.83$
$AUC_{S} = 0.839$
Th = 18.9
OR = 0.18 FPA = 0.28
100
75
50
25
0

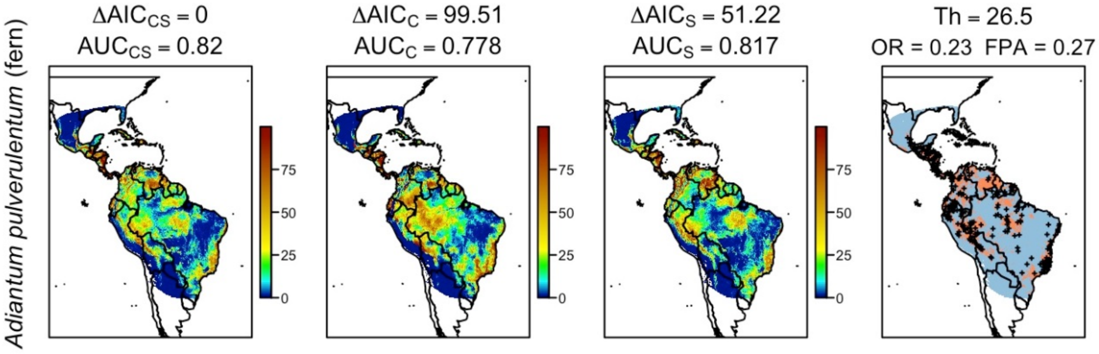

Goeppertia fragilis (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.832$
$\Delta AIC_{C} = 15.99$
$AUC_{C} = 0.693$
$\Delta AIC_{S} = 0.34$
$AUC_{S} = 0.87$
Th = 23.6
OR = 0.1 FPA = 0.28
Goeppertia loeseneri (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.903$
$\Delta AIC_{C} = 31.86$
$AUC_{C} = 0.85$
$\Delta AIC_{S} = 9.59$
$AUC_{S} = 0.872$
Th = 18.1
OR = 0.06 FPA = 0.26
Goeppertia zingiberina (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.804$
$\Delta AIC_{C} = 25.48$
$AUC_{C} = 0.789$
$\Delta AIC_{S} = 8.69$
$AUC_{S} = 0.805$
Th = 19.7
OR = 0.07 FPA = 0.31
Hylaeanthe hexantha (herb)
$\Delta AIC_{CS} = 0.08$
$AUC_{CS} = 0.811$
$\Delta AIC_{C} = 24.61$
$AUC_{C} = 0.778$
$\Delta AIC_{S} = 0$
$AUC_{S} = 0.824$
Th = 34.9
OR = 0.24 FPA = 0.21
100
75
50
25
0

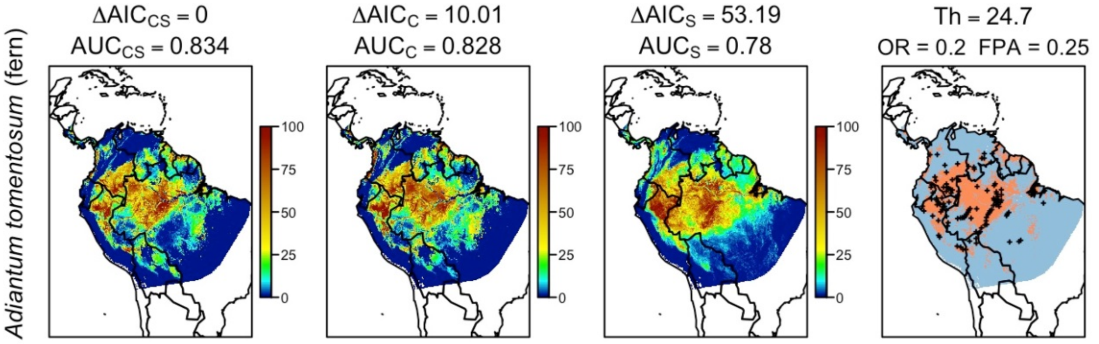

Heliconia schumanniana (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.918$
$\Delta AIC_{C} = 4.97$
$AUC_{C} = 0.909$
$\Delta AIC_{S} = 53.1$
$AUC_{S} = 0.88$
Th = 42.1
OR = 0.26 FPA = 0.07
100
75
50
25
0

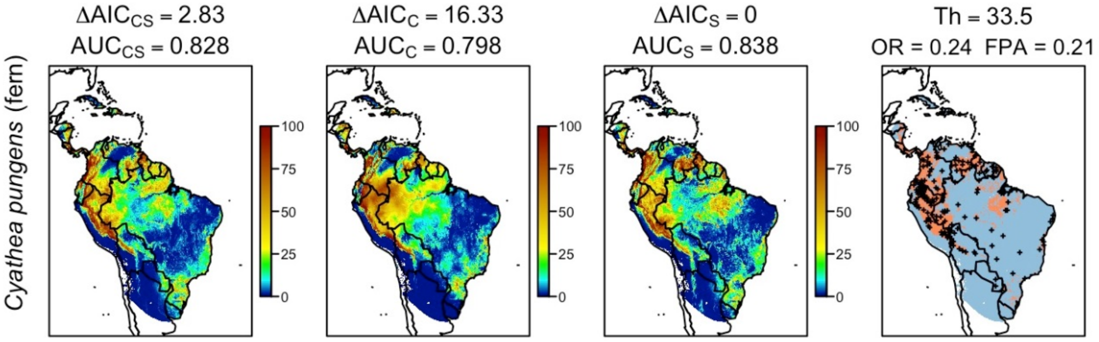

Ischnosiphon martianus (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.9$
$\Delta AIC_{C} = 38.69$
$AUC_{C} = 0.814$
$\Delta AIC_{S} = 31.13$
$AUC_{S} = 0.862$
Th = 11.1
OR = 0.04 FPA = 0.21
100
75
50
25
0

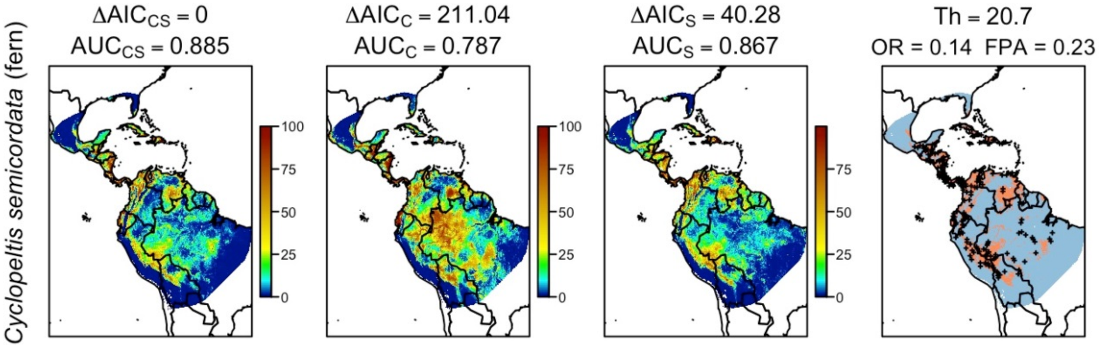

Monotagma ulei (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.851$
$\Delta AIC_{C} = 20.41$
$AUC_{C} = 0.762$
$\Delta AIC_{S} = 2.99$
$AUC_{S} = 0.846$
Th = 46
OR = 0.33 FPA = 0.11
100
75
50
25
0


Renealmia breviscapa (herb)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.863$
$\Delta AIC_{C} = 17.69$
$AUC_{C} = 0.829$
$\Delta AIC_{S} = 43.05$
$AUC_{S} = 0.806$
Th = 28.5
OR = 0.2 FPA = 0.2
100
75
50
25
0


Adiantum pulverulentum (fern)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.82$
$\Delta AIC_{C} = 99.51$
$AUC_{C} = 0.778$
$\Delta AIC_{S} = 51.22$
$AUC_{S} = 0.817$
Th = 26.5
OR = 0.23 FPA = 0.27
75
50
25
0


Adiantum tomentosum (fern)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.834$
$\Delta AIC_{C} = 10.01$
$AUC_{C} = 0.828$
$\Delta AIC_{S} = 53.19$
$AUC_{S} = 0.78$
Th = 24.7
OR = 0.2 FPA = 0.25
100
75
50
25
0


Cyathea pungens (fern)
$\Delta AIC_{CS} = 2.83$
$AUC_{CS} = 0.828$
$\Delta AIC_{C} = 16.33$
$AUC_{C} = 0.798$
$\Delta AIC_{S} = 0$
$AUC_{S} = 0.838$
Th = 33.5
OR = 0.24 FPA = 0.21
100
75
50
25
0


Cyclopeltis semicordata (fern)
$\Delta AIC_{CS} = 0$
$AUC_{CS} = 0.885$
$\Delta AIC_{C} = 211.04$
$AUC_{C} = 0.787$
$\Delta AIC_{S} = 40.28$
$AUC_{S} = 0.867$
Th = 20.7
OR = 0.14 FPA = 0.23
100
75
50
25
0

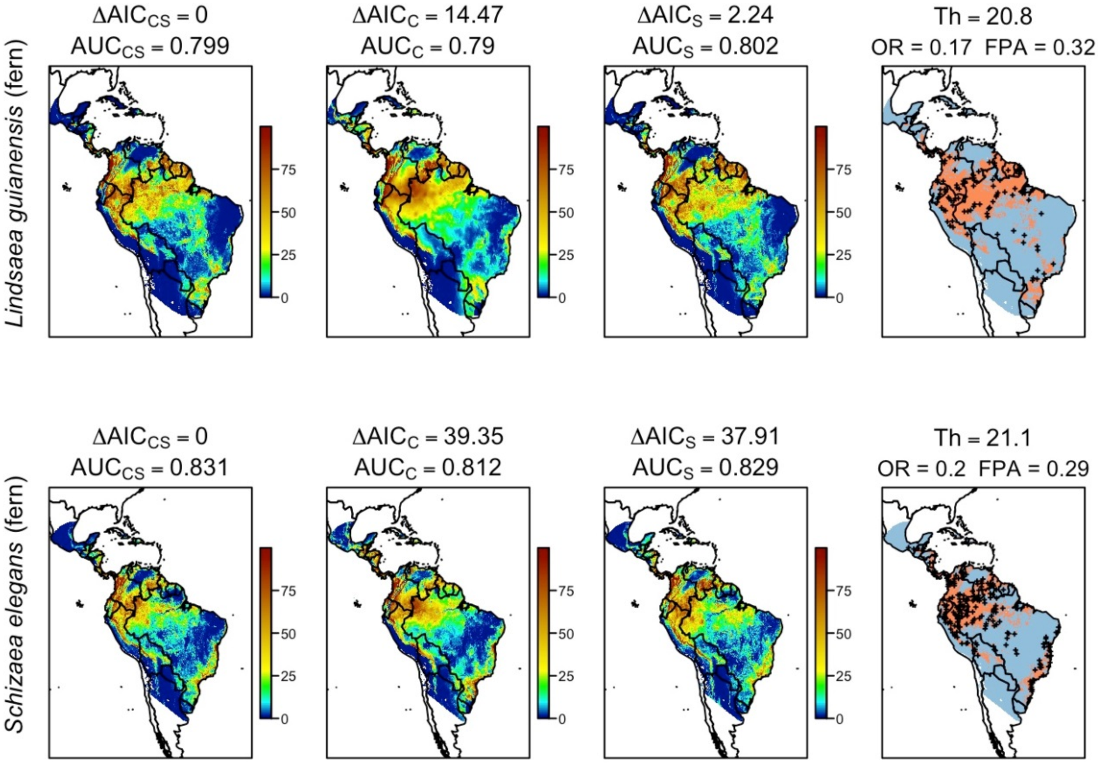
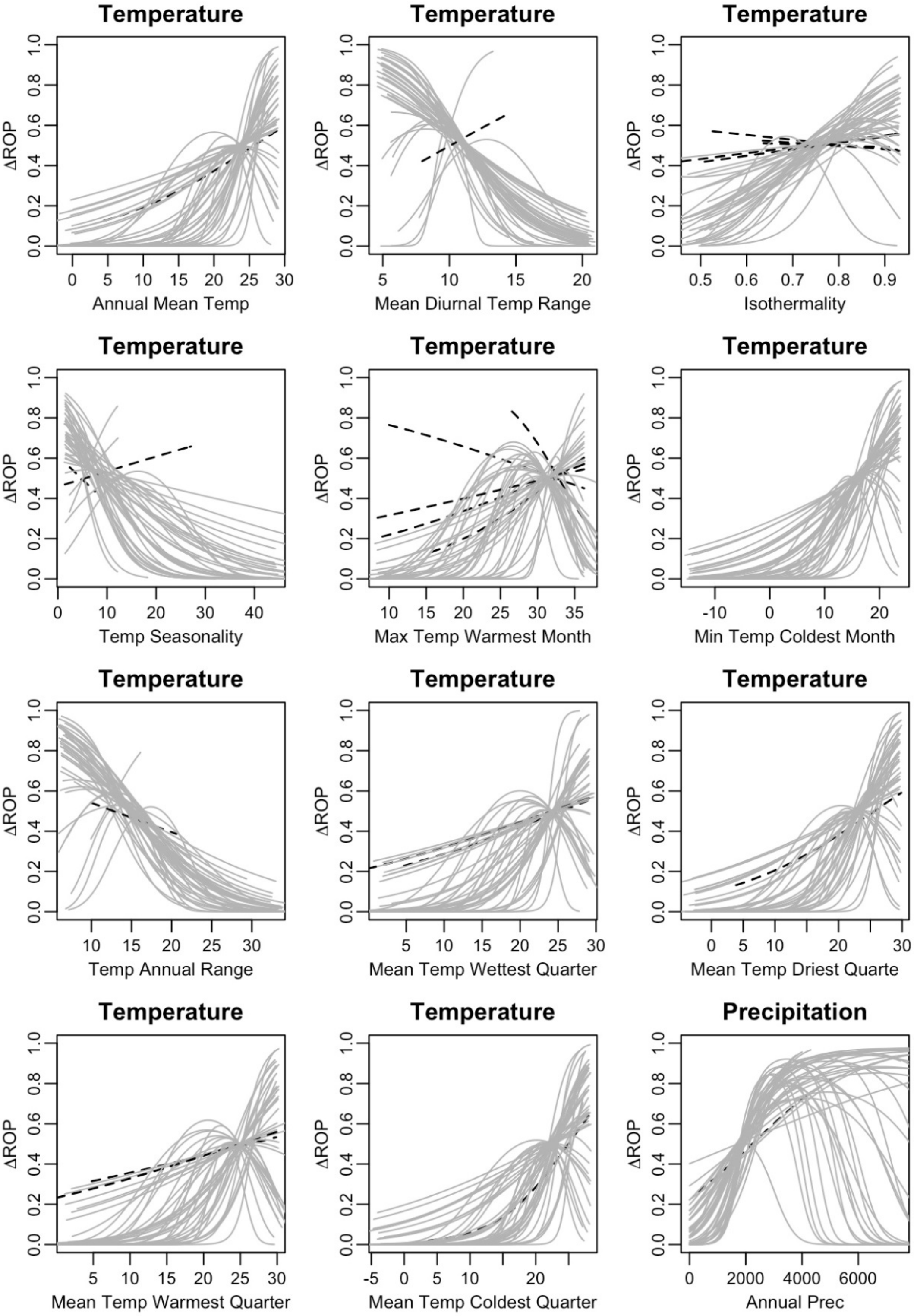
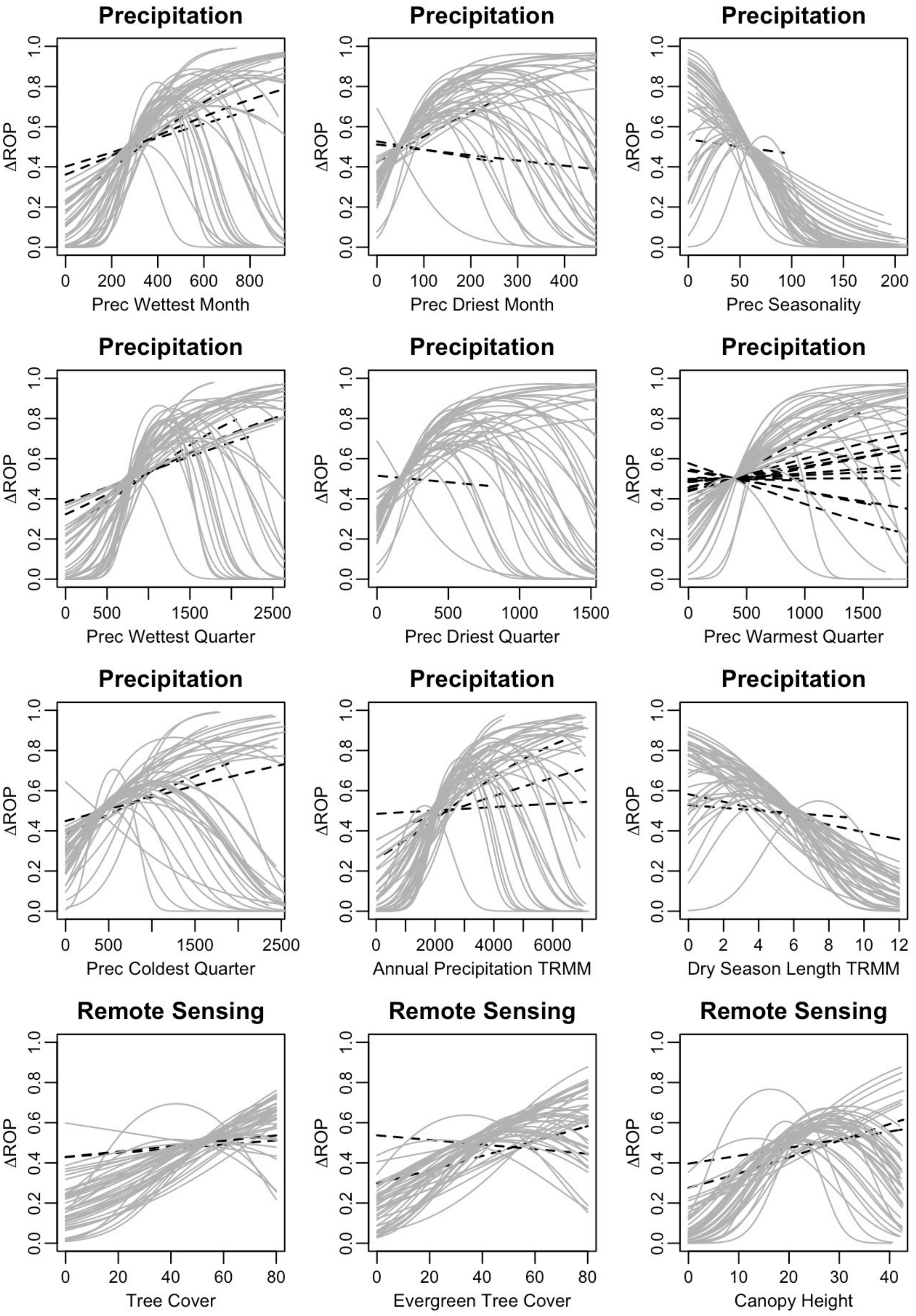


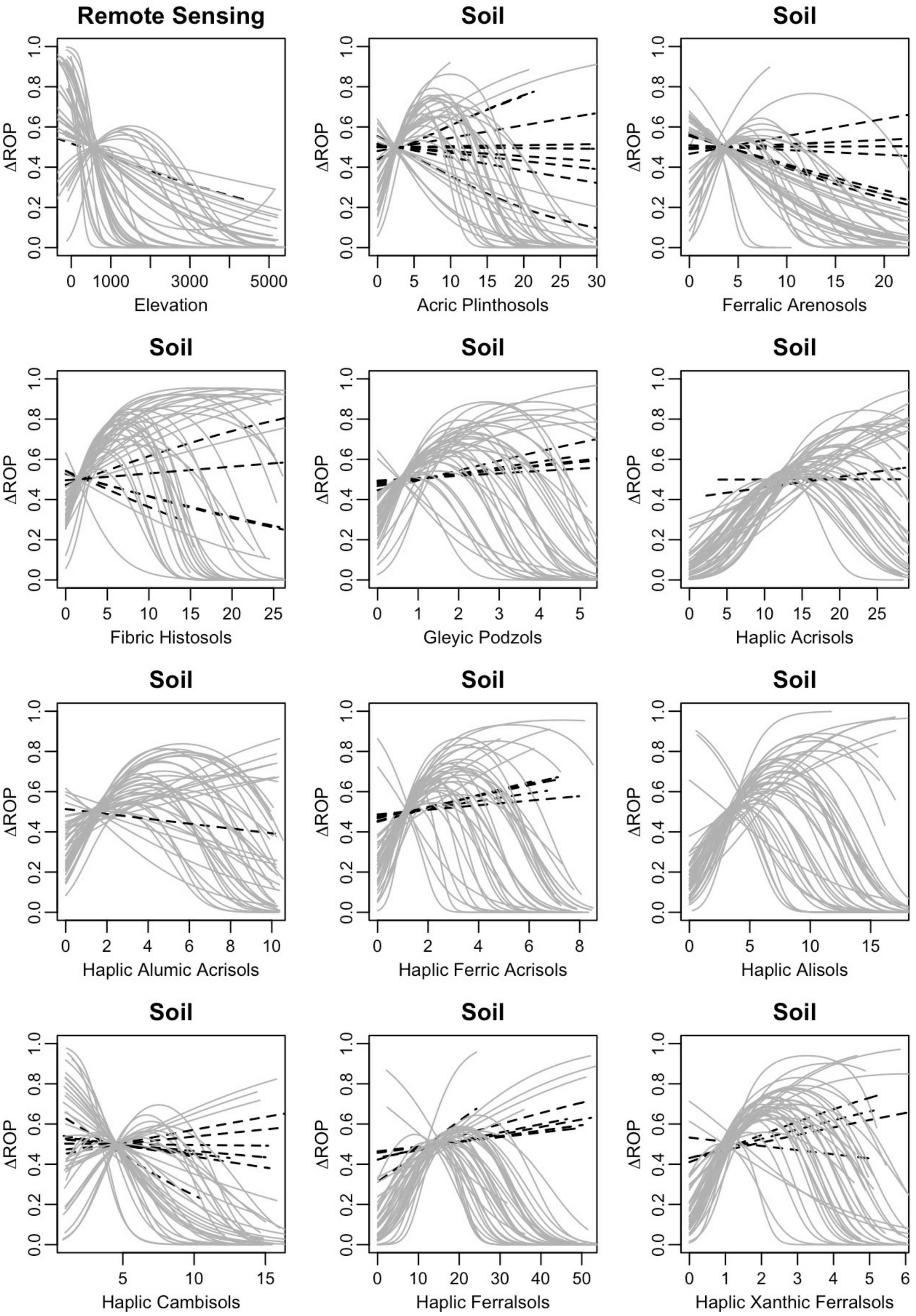
Figure S3.3 Spatial projections of models for the 42 Amazonian plant species. From left to right, the first, second and third column panels denotes the spatial predictions of the relative occurrence probabilities (cumulative ROP of average predictions from the 10 fold cross-validation) of CLIM+SOIL model (CS), CLIM model (C) and SOIL (S) model, respectively. High values of cumulative ROP represent high probability of species occurrence (high suitability scores) and low values denote low probability of occurrence (low suitability scores). Model performance based on ΔAIC and median AUC are shown above the panels. The forth column panel shows the predict distributional areas (orange areas) obtained by the best model (the CLIM+SOIL for the most of species) and areas in blue denotes the unsuitable climatic or/and soil conditions for species occurrence. Black crosses are the filtered presence records used to calibrate the SDM. The classification of distributional areas was based on setting the threshold on cumulative ROP that maximizes the sum of sensitivity plus specificity. Values above the forth column panels are the selected threshold (Th), omission rate (OR) and the fraction of predicted area (FPA) resulting from predictions' classification. The geographical background (area of predictions) were set in accordance to criteria defined in Appendix S2. Species are grouped by life-forms: trees, palms, lianas, monocot herbs and ferns.

Fig S3.4



Temperature
ΔROP
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
Annual Mean Temp
Temperature
ΔROP
5
10
15
20
Mean Diurnal Temp Range
Temperature
ΔROP
0.5
0.6
0.7
0.8
0.9
Isothermality
Temperature
ΔROP
0
10
20
30
40
Temp Seasonality
Temperature
ΔROP
10
15
20
25
30
35
Max Temp Warmest Month
Temperature
ΔROP
-10
0
10
20
Min Temp Coldest Month
Temperature
ΔROP
10
15
20
25
30
Temp Annual Range
Temperature
ΔROP
5
10
15
20
25
30
Mean Temp Wettest Quarter
Temperature
ΔROP
0
5
10
20
30
Mean Temp Driest Quarte
Temperature
ΔROP
5
10
15
20
25
30
Mean Temp Warmest Quarter
Temperature
ΔROP
-5
0
5
10
20
Mean Temp Coldest Quarter
Precipitation
ΔROP
0
2000
4000
6000
Annual Prec


Precipitation
ΔROP
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
Prec Wettest Month
Precipitation
0
100
200
300
400
Prec Driest Month
Precipitation
0
50
100
150
200
Prec Seasonality
Precipitation
0
500
1500
2500
Prec Wettest Quarter
Precipitation
0
500
1000
1500
Prec Driest Quarter
Precipitation
0
500
1000
1500
Prec Warmest Quarter
Precipitation
0
500
1500
2500
Prec Coldest Quarter
Precipitation
0
2000
4000
6000
Annual Precipitation TRMM
Precipitation
0
2
4
6
8
10
12
Dry Season Length TRMM
Remote Sensing
0
20
40
60
80
Tree Cover
Remote Sensing
0
20
40
60
80
Evergreen Tree Cover
Remote Sensing
0
10
20
30
40
Canopy Height


Remote Sensing
ΔROP
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
3000
5000
Elevation
Soil
Acric Plinthosols
0
5
10
15
20
25
30
Soil
Ferralic Arenosols
Soil
Fibric Histosols
Soil
Gleyic Podzols
Soil
Haplic Acrisols
Soil
Haplic Alumic Acrisols
Soil
Haplic Ferric Acrisols
Soil
Haplic Alisols
Soil
Haplic Cambisols
Soil
Haplic Ferralsols
Soil
Haplic Xanthic Ferralsols

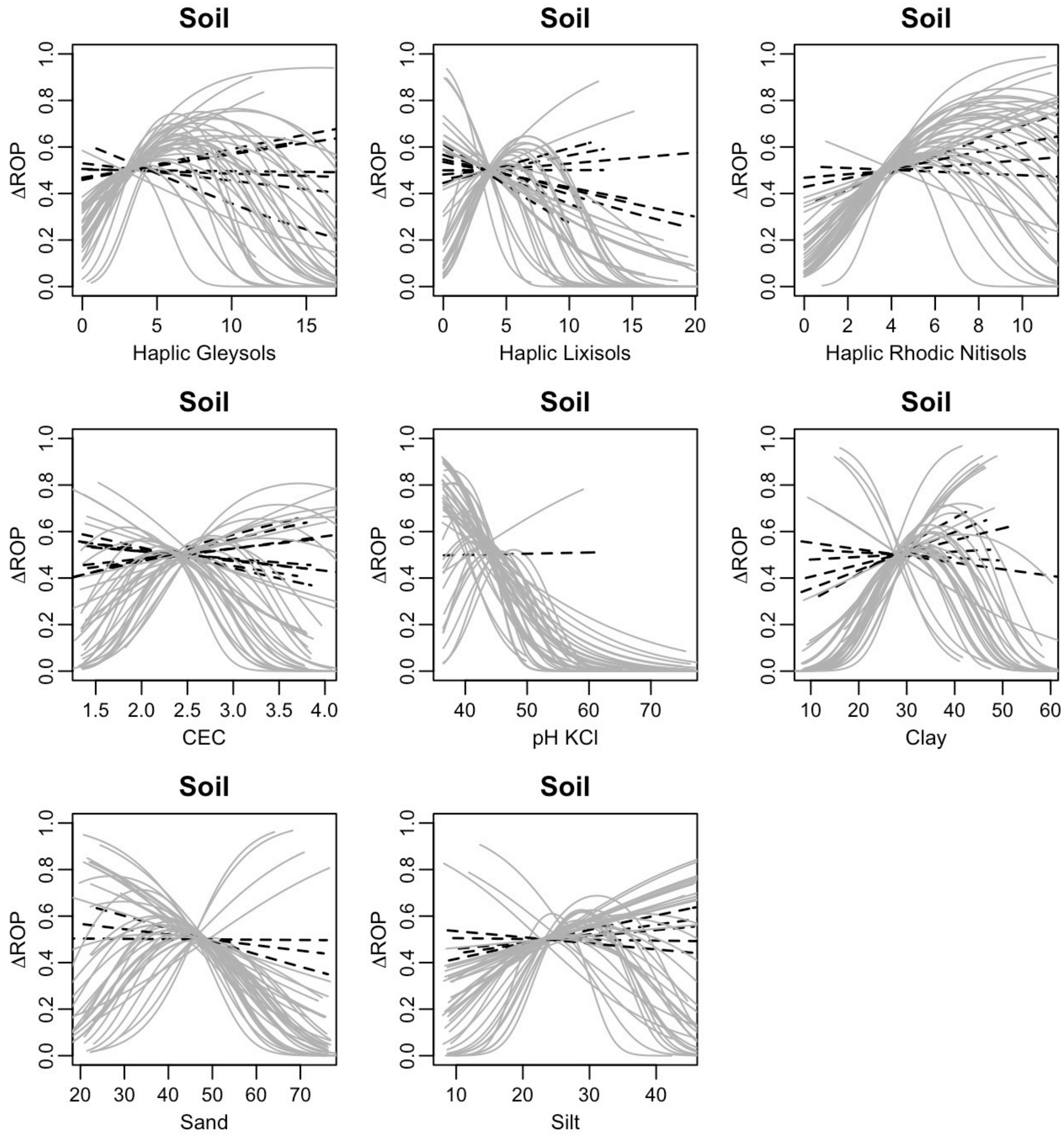


Figure S3.4 Predicted species response curves (relative occurrence probabilities, ROPs) along the 44 environmental variables. Species response curves were built using the contrasts plots from R package ‘visreg’ (http://myweb.uiowa.edu /pbreheny/publications/visreg.pdf). ΔROP was obtained shifiting the Xj values by a reference value, $\bar{x}_j$, setting all species’ ROP into a comparable scale (between 0 – 1). Lines in each of the 44 plots represent the response curves of each of the 42 species. Dashed black lines indicate models that were worse than the null model, based on ΔAIC.

ARTIGO 2

Terrain-based resources and climate drive convergent and divergent changes in the functional composition of herbaceous communities in tropical forest.

FIGUEIREDO, F.O.G, MOULATLET, G., ZUQUIM, G., EMÍLIO, T., TUOMISTO, H., SILVEIRA, M., DOMINGOS, J.R. & COSTA, F. R. C, Manuscrito em preparação para *Journal of Ecology.*

Abstract

1. External forces, such as soil, topography and climate tend to promote convergence in functional composition among communities. However, these variables are also related to disturbance regimes in tropical forest, which can promote divergence in functional composition within and even among communities. Divergent effects have not been addressed in regional studies, yet they can decrease our ability to predict the functional response of communities under environmental changes.
2. Using 451 inventory plots of understory herbaceous ginger assemblages, distributed across the Brazilian Amazon lowland rain forest, we measured the leaf mass per area (LMA), height and seed size for 192 species.We examined the roles of soil, hydro-topography and climate gradients as predictors of: (1) functional trait variability within local communities (2) convergent and divergent changes in functional composition among communities.
3. At the local within-community scale, variability in seed size and height tend to increase in bottomlands and in poor soils, a sign of divergence. At the larger Amazonian scale, functional composition converged more than it diverged along environmental gradients, except for height. Convergence followed a classical fast versus slow growth trade-off, with richer soils having fast-growth traits: lower means for LMA, height and seed size. The filtering effect of soil on LMA was the strongest among all traits and environmental variables. Functional composition diverged along climatic gradients, with communities in dry regions having widely distinct functional compositions. Among traits, LMA was under higher convergent pressure, while height and seed size were under higher divergent pressure.
4. *Synthesis* The functional composition among herbaceous communities in Amazonian is under both convergent and divergent pressures. The strength and direction of these forces depend on which environmental gradient and which trait is analysed. While all traits converged along soil fertility gradients, traits related to light acquisition and dispersal tend to diverge along topographic and climate gradients. These results suggest that natural stochastic gap-disturbance and non-equilibrium population dynamics partially control the functional composition of understory herbaceous communities and indicate there is still much uncertainty in how climate change will affect the structure and functioning of Amazonian rain forest.

Introduction

Convergence and divergence have been seen as opposite forces defining the functional structure of communities. Convergence is driven by environmental filtering and tends to restrict the functional variation within communities, while divergent forces tend to promote local functional trait variation (Grime 2006; Kraft, Valencia & Ackerly 2008; Cornwell & Ackerly 2009). So far, studies of large scale plant functional patterns in tropical forests have focused on how environmental forces drive the variation of optimum functional traits and strategies among communities, revealing a trait-strategy convergence governed by climate and edaphic conditions (ter Steege *et al.* 2006; Fyllas *et al.* 2009; Fortunel *et al.* 2014; Muscarella & Uriarte 2016). On the other hand, local scale studies have explicitly evaluated how traits diverge within communities, in a search for signals of interspecific competition controlling local species coexistence (Kraft *et al.* 2008; Sterck *et al.* 2011). Several processes beyond interspecific competition may constitute divergent forces, including fine scale spatial or temporal resource heterogeneity (*e.g.* light gap dynamics), source-sink population dynamics, and other stochastic natural disturbances. If regional environmental filters are not strong, divergent forces may overcome their selection pressure, increasing the trait diversity within communities as well as the variance among trait communities′ optimum. This perspective is lacking on studies of regional trait structure between communities, which mostly view variance as random errors. Recognizing the specific factors and processes that can generate predictable variance in traits at large scales can improve understanding and modelling of community and ecosystem dynamics.

Functional traits represent the phenotypic expression of individuals and relate to population vital rates (growth, recruitment, mortality) and plant life strategies. Direct measurements of the life cycle of organisms or of plant tissues are not straightforward, especially in high diversity ecosystems and in large scale applications, so plant functional ecologists have been using a set of traits as proxies of plant growth strategy. These include height, leaf mass per area (LMA, in g m^{-2}) and seed size (Westoby 1998; Reich 2014; Díaz *et al.* 2016). Height of plants is directly related to organism longevity and to light niche position, especially in forests. Taller species have more access to light, but with high cost to produce and sustain the stem. Seed size summarizes the reproduction and establishment strategy. Large-seeded species produce few seeds with high supply of reserves, enabling them to establish under scarce-resource conditions, while small seeds produced in large quantities enable species to colonize micro-sites that are rare in space and time (*e.g.* light-gaps) (Dalling, Hubbell & Silvera 1998; Kitajima 2002). LMA strongly summarizes the global spectrum of leaf economics, running from fast to slow return of investment in leaf tissues and nutrients (Wright *et al.* 2004). Ecological theory predicts that at low level of resources (nutrients, light, water), species grow slowly because they often invest in highly durable tissues at either leaf, wood and root levels (Reich 2014). Under these conditions species have low mortality rates, high longevity, and expressing phenotypes marked by stature, seed size and LMA. The opposite strategy is expected under high availability of resources.

Soil and climate represent key environmental filters of functional trait strategies of trees across Amazonian forests. Tree species with slow growth strategies (conservative traits) tend to successfully establish and thrive in nutrient poor forests. Fast growth species (acquisitive traits) dominate in nutrient rich forests (ter Steege *et al.* 2006; Fyllas *et al.* 2009; Quesada *et al.* 2012). Climate is expected to modulate leaf traits and plant size at the global scale, but often a weak relation to climate has been found for leaf traits (Wright *et al.*; Maire *et al.* 2015). In tropical forests, patterns of tree functional change along rainfall gradients have been contradictory. Fyllas et al. (2009) found a positive relationship between LMA and annual precipitation in Amazonian forests, whereas Muscarella & Uriarte (2016) found a negative relationship across a rainfall gradient in Central America. However, organisms do not access the total amount of water provided by rainfall. Water availability depends also on the local topo-edaphic conditions, such as the soil capacity to retain water and the proximity to water table. Topography strongly influences the variation of hydraulic traits of species. It may be as stronger a filter of functional strategies as climate (Cosme *et al.* 2017) but has not yet been included in studies of large scale functional patterns.

Natural disturbances are expected to interact asymmetrically with soil, climate and topography in tropical forests, promoting trait diversity and generating what may be recognized as noise in the observed trait community optimum (mean trait values) among communities. It has been documented that fast growth and corresponding high rates of tree mortality lead to a higher frequency of disturbance in rich soils or soils with physical limitations (Quesada *et al.* 2012), which can potentially increase local spatial heterogeneity of light. Frequency of windstorms should be higher in areas with high levels of precipitation and may also increase the dynamics of forest gap-disturbance (Espírito-Santo *et al.* 2010; Schietti *et al.* 2016; Negrón-Juárez *et al.* 2017). Tree mortality rates tend also to be higher in bottomlands than on hilltops and plateaus (Ferry *et al.* 2010; Toledo *et al.* 2011). All these connexions between regional environmental factors and disturbance regimes can change the expectations from classical community-trait trade-off theory by generating trait divergence, more so in those traits related to dispersal and light acquisition.

Trait-based community studies in the tropics have mostly targeted canopy trees and neglected understory species. Whether environmental gradients affect the functional structure of both these groups in the same way is an open question. A reversal of the dicot strategy along light gradients has been proposed for understory monocot herbs (Swenson 2009). Light demanding species showed higher density tissues (leaf and culm) than shade tolerant ones, the opposite pattern observed in trees, suggesting a biomechanical constraint on species trait variation. Height of understory herbs is directly related to their light interception capacity and may define the light niche position of species. The higher intensity and frequency of gap-disturbance regimes associated with rich soils, bottomland and wet regions, as expected above, may thus favour the predominance of light demanding species in such habitats. According to the understorey monocot strategy proposed by Swenson (2009), it would be possible that communities in these habitats would have predominantly tall species with high LMA, i.e., the opposite expectation of classical acquisitive-conservative trade-off for trees.

Here, we investigate the variation of functional trait patterns of 420 herbaceous communities across the Amazon basin under a trait divergence-convergence perspective. Our objectives are two-fold. First we test whether regional environmental gradients (soil, topography and climate) affect the within community trait variation (Community Weighted Deviation, CWD) of LMA, height and seed size. Our expectation is that the within community trait variation will increase towards nutrient-rich soils, bottomlands and wet areas in response to the higher disturbance levels associated with these conditions. The second objective is to test whether regional environmental gradients promote trait convergence (evaluated by the mean response of Community Weighted Means, CWM) or mean trait divergence (evaluated by the variance response of CWM) among communities. Following the classical fast-slow growth trade-off, we expect that LMA, height and seed size CWM should decrease as soil nutrients and water availability increase. However, according to understory monocot functional strategy along light-gradients, LMA and height should increase in more disturbed areas - the rich soils, bottomlands and wet regions - an opposite pattern to that expected by the classical functional trade-off. In terms of the variance response of CWM, we also expected that variance of mean trait values will increase in rich soils, bottomlands and wet conditions, given the high level of disturbance associated to them. By joint modelling the mean and variance parameter of the community trait means we can estimate the relative importance of convergence and divergence forces that structure the functional changes across herbaceous communities in tropical forests.

METHODS

Taxonomic group: Our study was based on functional trait patterns of the Zingiberales (hereafter gingers), a rich pantropical order within monocots, with about 2,000 species (Kress *et al.* 2001). Ginger species are rhizomatous herbs, varying in habit from small rosulate herbs (10 cm height) to tall palm-like or scandent plants (5m tall) (Fig. 1). In Amazonian lowland rain forest, ginger species greatly contribute to the abundance and diversity of understory herbaceous plants (Costa 2004; Costa, Magnusson & Luizao 2005) and floristic variations of ginger assemblages are in accordance with other taxonomic groups (Figueiredo *et al.* 2014; Tuomisto *et al.* 2016)

Study area and plots design:

This study was conducted in the Brazilian Amazonia. Our sample design included only non-inundated lowland rain forest, covering a wide range of soil and topography conditions. Based on time series of Tropical Rainfall Measuring Mission (TRMM) orbital sensor data (Kummerow *et al.* 1998) annual precipitation ranges from 1,900 to 3,200 mm across the study area.

Biotic and abiotic data were obtained in 451 plots of 500 m^2 established in non-inundated forest (Fig Sx). Data came from two plot designs: (1) 412 plots of 250 x 2 m set up on topographical isoclines, following the RAPELD protocol (Costa & Magnusson 2010); (2) 39 plots of 100 x 5 m sub-sampled from 500 x 5 m transects oriented to include representative local topographic variation (Tuomisto *et al.* 2016). Altitudinal values at 5 m intervals allong the 500 m long transect were extracted from a digital elevation model to guide the selection of a 100m long continuous section having internal topographical variation as low as possible, in order to make the two sample designs comparable. RAPELD sample units are aggregated in spatial modules, but with minimum geographic distance between plots of 1 km.

Data collection

Species data - We identified and counted every clump with a height >5 cm rooted inside the plot. Clumps were defined as groups of stems or leaves arising from the soil less than 20 cm from each other, or based on our field experience with each species. We identified the species based on specialized literature and sent fertile material to specialists Helen Kennedy and Paul Maas, to confirm identifications. Voucher collections are deposited in the INPA Herbarium in Manaus, BR; the University of British Columbia Herbarium, CA; and the National Herbarium of the Netherlands at Utrecht University, NL.

Functional Trait data - Our analyses were based on three functional traits: LMA, height and seed size. LMA is the ratio between the dry leaf mass and fresh (*i.e*, completely expanded) leaf area. Here, we adapted the original method to estimate LMA, using small pieces of the lamina of dried material from herbarium collections. For each individual, a maximum of 4 pieces (3 cm^2 average) were excised from different parts of the lamina (at the middle and close to the base and apex), excluding the main vein. We carefully chose leaf pieces where the tissue had no or very little shrinkage and avoided young leaves. LMA for each species was calculated as the average of all pieces obtained from a maximum of six individuals and minimum of one individual, when the species was rare and collected only once. Height and seed size were obtained by searching monographs, field guides and our field observations. Seed size was obtained using the formula for area of an ellipse based on the two largest dimensions of the seed reported in the literature. We were unable to obtain measures of LMA, height and seed size for 23, 23 and 109 species respectively. Trait values in these cases were assigned based on the mean trait value of their genus.

Environmental data - To represent soil fertility, topographic and climatic gradients we used the concentration of exchangeable base cations (Ca, Mg e K) defined as sum of bases, the vertical distance from the nearest drainage (VDD) and the dry season length in months, respectively. The sum of exchangeable bases is a good proxy for soil fertility in Amazonia since it is well correlated with phosphorus concentration, a limiting soil nutrient (Quesada *et al.* 2010), and is strongly related to diversity patterns across the basin (Higgins *et al.* 2011; Figueiredo *et al.* 2014; Tuomisto *et al.* 2016). Fertility was determined from soil samples taken at each plot. The protocols for soil sampling and laboratory analyses are described in previous studies (Figueiredo *et al.* 2014; Moulatlet *et al.* 2014; Zuquim *et al.* 2014; Tuomisto *et al.* 2016).

The topographical descriptor was based on the vertical distance from drainage (VDD), extracted from SRTM digital elevation models (SRTM-DEM). VDD have been successfully used to map local hydrological conditions and to explain floristic patterns in Amazonia (Rennó *et al.* 2008; Schietti *et al.* 2013; Moulatlet *et al.* 2014). Plots with low VDD values indicate that individuals are closer to bottomlands and consequently closer to water table and with better access to water than individuals in plots with high VDD values, located often on hill tops and plateaus, distant from water table. VDD values were obtained from three similar algorithms. For 198 plots, located in the interfluve Purus-Madeira and in the state of Pará, VDD was derived from the Height Above Nearest Drainage (HAND) algorithm using the 90m resolution SRTM-DEM (Rennó *et al.* 2008; Moulatlet *et al.* 2014; Pinheiro, Rennó & Escada 2015). For 193 plots located north of the Amazon River in central Amazonia, VDD was obtained from the Elevation Over Channel Network algorithm using the hydrological tools of the SEXTANTE spatial data analysis library (http://www.sextante- gis.com/), coupled with the open-source gvSIG v 1.10 software (http://www.gvsig.org/web/) (see details in Figueiredo *et al.* 2014). For the remaining plots, VDD was obtained from 30m SRTM-DEM using the Vertical Distance to Channel Network (Bock & Köthe 2008) algorithm, implemented in QGis software through the SAGA-toolbox.

Dry season length was estimated from data of the TRMM satellites (Kummerow *et al.* 1998) available at http://disc.sci.gsfc.nasa.gov. We converted monthly data from 1998 to 2004 of the TRMM product 3B43 V6 at a 0.25° resolution (about 28 km at the equator) into the dry season length variable, defined here as the maximum number of consecutive months with less than 100 mm of precipitation along that period. For each plot, the dry season length value was extracted after rescaling the data to a 0.05° (~ 5 km) spatial resolution. We used the *raster* package in R to manipulate and process TRMM variable (R development core team 2015; Hijmans *et al.* 2017).

Data analysis

In order to evaluate the effect of environmental gradients on functional trait patterns within and among communities, we first estimated two trait community values for each plot. We estimated the community weighted mean (CWM) to characterize the optimum trait strategy for each community (Díaz *et al.* 2007; Violle *et al.* 2007) and the community weighted deviance (CWD) to characterize the within community trait variability. The CWM and CWD are given by the following formulas:

$$CWM = \sum_{k=1}^{n_j} w_{k,j} \times T_{k,j}$$

$$CWD = \sqrt{\frac{\sum_{k=1}^{n_j} w_{k,j}(T_{k,i} - CWM)}{\sum_{k=1}^{n_j} w_{k,j} - 1}}$$

where *w* is the relative abundance of the species *j* in the community (plot) *k*, and *T* is the trait value of the species *j* in the community *k*.

To evaluate the convergence-divergence patterns of trait response among and within communities, we performed generalized additive models for location, scale and shape (GAMLSS) (Rigby & Stasinopoulos 2007) to jointly estimate the effect of environmental gradients on the mean (μ) and variance (σ) of community-trait response. The general idea is to fit the model assuming that response variables (CWM or CWD of each trait) can be represented by a general distribution family, which include not only the location parameter (mean), but also a scale parameter (variance), and shape parameters (skewness and kurtosis). The approach consists in modelling all parameters as linear and/or non-linear responses of predictor variables using specific link functions for each of the parameters. For CWD, we only fitted a mean response model as a function of environmental variables because our expectations are only based on that parameter. The CWD model was fitted using a zero adjusted Gamma distribution for the response variable, since we recorded only one species in some plots and thus CWD can assume values equal to zero.

For CWM, we fitted a model based on the mean and variance response of each functional trait. Another feature of GAMLSS is that we can test several distribution families that best fit the distribution of response variables. For height and seed size, we tested several continuous distribution assuming y>0, which run from simple family distributions, *e.g.* exponential with one parameter, to more complex ones, *e.g.* Box-Cox *t* with four parameters. For LMA, we included also the family distributions that do not have lower bounds, *e.g.* Normal distribution. The chosen family distribution was that with low values of the generalized Akaike information criteria (GAIC) obtained fitting the available distributions on observed values of CWM (without environmental predictors). However, the best family sometimes lead to problems in the convergence optimization of model parameters when the predictor variables are included in the model. Thus, the second best, third best and so on were selected until the full model (with environmental predictors) have reached good convergence properties. For height and seed size, CWM models were fit using a Box-Cox *t* with four parameters assuming a log link for mean, variance and kurtosis and identity link for the skewness parameter. For LMA, the CWM model was fit using a *t* family distribution with 3 parameters using identity link for mean and log link for variance and skewness parameters.

After setting the distribution that better fits our response variables, we performed a model selection procedure using again the GAIC to set which independent variables better explain the variation of each parameter of our response variables. For CWD, we only evaluated the mean response, because our ecological expectations are based on this parameter, and thus the intercept of the variance parameter of a zero adjusted Gamma distribution was estimated. For CWM, we tested mean and variance responses as functions of environmental predictors. The intercept of skewness and kurtosis parameters were also estimated, but their values were not allowed to vary in response to environmental predictors. Model selection procedure started with forward approach applied first in the mean model and subsequently in the variance model. The procedure finishes running the backward selection on the contrary direction, from variance to mean. Only linear relationships between response and predictor variables were allowed.

We performed GAMLSS models using the mixed effect design (Zuur *et al.* 2009). The spatial modules of RAPELD protocol and plots in the Juruá River (Tuomisto *et al.* 2016) were set as the random term in the GAMLSS model, for two reasons. First, to estimate model parameters and the significance statistics controlling for the spatial autocorrelation present in our sampling design, due to spatial aggregation of plots within localities. And second, to control the potential bias in the estimation of model parameters generated by differences in protocols for soil and topographic data. The sum of bases and vertical distance from drainage were logarithmically transformed before the analysis to fit the assumptions of normality required in linear models.

To compare the relative importance of convergent and divergent effects of predictors on the CWM, we fitted three kinds of models: one allowing only the environmental effect on the mean parameter (convergent effect); the second allowing only the environmental effect on the variance parameter (divergent effect); and a third model allowing the effect of environment in the mean and variance parameters. Models' performance was evaluated using the ΔGAIC and generalized R^2 (Nagelkerke 1991) implemented for GAMLSS models (Rigby & Stasinopoulos 2007).

GAMLSS models were run using the *gamlss* R-package and model selection was based on the function *stepGAICAll.A* from the same package (Rigby & Stasinopoulos 2005).

RESULTS

General results

We recorded 47,129 herb individuals belonging to 192 species. Five plots had only one species and three or less individuals and were removed from the analyses. The minimum, median and maximum abundance per plot was was 4, 66 and 897 individuals, respectively. The median of species richness per plot was 8, varying from a minimum of 1 to a maximum of 23 species. Functional traits were often moderately or weakly inter-correlated at species and community level. At the species level, a high and positive correlation was observed between LMA and height (Fig. 3a). Communities with high local variance (CWD) in height were also those with high variance in seed sizes (Fig. 3f) and communities with dominance (CWM) of taller species tended also to have a dominance of species with larger seeds (Fig. 3i). The complete list of species and their functional trait values is presented in Table S1 of Supporting Information.

Environmental effects on within community trait variation (CWD)

We found a weak effect of environmental gradients on local, within community trait variability (Table 1). The variation of seed size within communities was slightly higher in poor nutrient soils and in bottomlands (Table 1). The variation of height followed the same pattern as seed size, but the best height-environment model had very week support when compared with a null model (ΔAIC < 2).

Environmental effects on functional composition among communities (CWM)

We found significant convergent and divergent effects of environmental gradients on functional trait composition among communities. The best model (low ΔAIC and high R^2) included significant effects of the environmental variables on mean and variance response of

CWM, except for LMA, whose mean response was associated with environmental gradients (Table 1). Models that included only the mean response of CWM had better performance than models that included only the variance response, except for height (Table 1).

The rate of change in trait values among communities varied strongly among distinct traits and environmental predictors. As expected by the classical fast-slow growth trade-off, the mean LMA was lower in nutrient rich soils and bottomlands, but contrary to our expectation, there was no change in LMA along the dry season length gradient (Fig 4 a-c). The standardized rate of change of mean LMA values along the soil fertility gradient (*b*=-2.52) was the highest comparing all models, and was twice bigger than the rate of change along the topography gradient (*b*=1.18).

The patterns of change of mean height and seed size along environmental gradients were very similar (Fig. 4d-i), probably due to the high correlation between these two traits at community level (Fig. 4f and 4i). As for LMA, height and seed size also decreased as soil fertility increases, but with smaller rates of change than LMA (Table 1, Fig 4a). Mean seed size slightly decreased in bottomlands, but with a smaller rate of change than the rate at which the variance increased (Table 1).

Contrary to the observed for LMA, we found a significant effect of topography and climate on the among-community variance of height and seed size. As expected, the mean community values of height and seed size diverged more than in bottomlands than in uplands (Fig. 4e and f). In bottomlands we found wide variation in terms of optimum traits given by CWM, while in uplands there was a predominance of herbs with intermediate stature and medium seed size. Contrary to our expectation, the variance of CWM height and seed size increased in drier conditions, while in wet areas the dominant species were those with medium height and seed size (Fig. 4g and h).

Discussion

Our findings reveal that regional environmental gradients affect both the within-community and among-community functional trait patterns of the ginger's assemblage in Amazonian forests. At local communities, the within-community variability (CWD) in seed size and height tend to increase in bottomlands and in poor soils. At the larger Amazonian scale, functional composition mostly converged towards higher LMA, height and seed size in nutrient-rich soils, in accordance with classical fast-slow growth trade-off. Height and seed size tended to diverge among communities across topographic and climatic gradients. The filtering effect of soil in LMA was the strongest among all traits and environmental variables, revealing that soil is a key factor driving regional changes of functional composition across the Amazonian basin. Bottomlands had low values of LMA and seed size and no predictable functional compositional changes were observed along the climatic gradient. The functional composition diverged more among communities in bottomlands and dry regions than in uplands and wet regions, but only in respect to height and seed size, suggesting that local disturbance events and disequilibrium population dynamics may be relevant factors to explain changes in functional composition at regional scales in Amazonian forests.

Within-community variation in seed size and height tend to increase in bottomlands and in poor soils

Our expectation that local trait variance would be associated to environmental conditions that promote higher frequency of disturbances was partially valid. The within community variabilities of seed size and height were slightly greater in bottomlands. Seed size is related to dispersal and colonization strategies, and height of understory plants may be directly linked to their light niche position. Bottomlands are more prone to natural disturbance than uplands, given terrain instability and physical constraints of their soil, leading to higher rates of tree mortality (Ferry *et al.* 2010; Toledo *et al.* 2011; Quesada *et al.* 2012). The higher the tree mortality rates, the higher the spatial and temporal heterogeneity of light in the understory, which would explain the presence of species with more variable seed sizes and heights co-occurring in local communities within bottomlands.

Tree mortality tends also to be higher in rich soils (Quesada *et al.* 2012), but contrary to our expectation, local trait variance of height and seed size slightly increased as soil fertility decreased. A potential explanation is that light may not be limiting in rich soils due to high intensity and frequency of tree mortality events, which makes the canopy uniformly more open and light more widely available. Therefore, the strategy of being tall may not be necessary or advantageous in this environment. At the same time, high productivity habitats also favour the fast-growth strategy (Reich 2014) which may limit the potential of species to achieve high stature, as this would potentially divert resources that could be better invested in faster reproduction.

Local trait variability was not associated with climatic gradients. We expected that wet regions would have higher local trait variability assuming that storms are more frequent in such regions. In central Amazonia, the frequency of storms have been indexed by the number of days per year with precipitation higher than 20 mm of rain (Espírito-Santo *et al.* 2010), which may not be directly associated with the size of dry season length, as used here. Further investigations may be done in order to check if this storm-index is valid for the whole Amazon basin, and if it useful to predict patterns of trait variability in plant communities.

Soil-resource and topographic gradients drive convergence in functional composition among communities

Functional composition (CWM) among herbaceous communities converged more than diverged along environmental gradients, with soil gradients having the strongest filtering effect. Trait values observed along the soil gradient are in perfect accordance with the classical fast-slow growth trade-off, which predicts that acquisitive traits (low LMA, low stature and small seed size) dominate in highly productive habitats, while conservative traits predominate in low resource conditions (Reich 2014). Indeed, the rate of change in community trait means was higher along the soil gradient than along topographic or climatic gradients, reinforcing that soil-fertility is the key driver of regional changes in functional compositional across Amazonian lowland rain forest, as previously described for trees (Fyllas *et al.* 2009; Quesada *et al.* 2012).

The functional composition of ginger´s assemblage also converged along topographical gradients following the expectation of classical acquisitive-conservative strategy trade-off along water-resource gradients (Hoffmann *et al.* 2005; Reich 2014). Uplands tend to suffer more with water deficit than bottomlands, favouring the predominance of conservative strategies (low LMA and larger seed size). A similar trend was observed at a local scale for trees, species associated with plateaus tend also to have higher LMA than species associated with valleys in central Amazonia (Cosme *et al.* 2017). Soil on plateaus in central Amazonia are slightly more nutrient rich than valleys, indicating that increase of LMA in uplands is more probably associated with water availability than soil nutrients in this local scale (Cosme *et al.* 2017). Topographic variation has been disregarded in regional scale studies of community patterns, and probably these investigations fail to capture tight links between changes in functional composition and water deficit, which depends not only on climate, but the interaction of it with topography.

Changes in functional composition of ginger assemblage were not associated to the climatic gradient. In tropical forests, functional changes among tree communities along rainfall gradients have been contradictory, with some studies showing that LMA is higher in wet conditions (Fyllas *et al.* 2009) and others showing the opposite pattern (Muscarella & Uriarte 2016). We did not find any tendency of increasing or decreasing of trait values along the climatic gradient, probably because most trait variation was explained by soil-resources and water-topography gradients. The aforementioned studies only evaluated direct relations between community traits and climatic gradients, which can lead spurious correlation if terrain-based environmental filters were not taken into account, although the sampling design in Muscarella & Uriate (2016) included a single soil type. The absence of climate effect on LMA could also arise by the presence of deciduous species in dry regions. Deciduous species are often associated with dry regions and tend to have lower LMA than evergreen ones (Poorter *et al.* 2009) which could annulated the trend of increase LMA (conservative strategy) in directions to dry regions. Although we have not directly evaluated the deciduous pattern in our dataset, our field observations suggest that some short species of *Goeppertia* and *Chamaecostus* can exhibit this behaviour. However, as LMA is under the strong control of soil conditions, it is expected that deciduous may be conditioned jointly by the dry conditions and high concentrations of nutrients in soils.

Our findings are in perfect agreement with the classical fast-slow growth trade-off. Species with acquisitive traits dominate in habitats with high levels of resources (rich soils and bottomlands). Another expectation however, according to the monocot functional strategy along light-gradients (Swenson 2009), was that the high levels of light reaching the understory of rich soils and bottomlands, as consequence of high rates of tree mortality (Ferry *et al.* 2010; Quesada *et al.* 2012) would favour the predominance of tall species with high LMA in these habitats. This reversal of the dicot strategy was previously described at the species level. Here we did not observe this pattern at community level and regional scale, where external environmental filters drive the community optimum functional strategy. However, about half of variation in the mean trait values of communities was not explained by our external environmental filters, which may partially represent the noise generated by the monocot strategy along light-gradients.

Climatic and topographic gradients promote divergence in functional composition among communities

The functional composition of understory herbs diverged along topographic and climatic gradients, but not along the soil nutrients gradient. We found that bottomlands and dry regions supported communities with very distinct dominant functional strategies in terms of height and seed size. Height of understory herbs seems to be directly linked to their capacity to intercept light, in which often, tall species may be the light-demanding and short species the shade-tolerant. Small-seeded species are better adapted to colonize gaps (Dalling *et al.* 1998; Westoby *et al.* 2010) and thus are expected to predominate in forests with high gap-disturbance. Thus, stochastic gap-disturbance events may be driving changes in light conditions above the canopies and thus promoting the high variation in mean trait values across bottomlands. The dominant trait, however, will be driven by the intensity and frequency of gap-disturbance events.

The functional composition also diverged more in dry than in wet regions. As climate becomes drier, the forest becomes shorter, more open and consequently more heterogeneous. According to the metapopulation model, colonization rates are expected to decrease and local extinction rates to increase in highly heterogeneous habitats (Levins 1969). In addition, the dry regions in Amazonia found in its border and in the east, may have suffered some degree of forests fragmentations and local extinctions driven by long-term cycles of high-low precipitations regimes in the last 2 Mya (Van Der Hammen & Hooghiemstra 2000; Cheng *et al.* 2013). Metapopulation dynamics and long-term climate oscillation can interact to set the vegetation into disequilibrium, *i.e*, local communities will never contain all species climatically adapted due to dispersal limitation (Svenning & Sandel 2013). Thus, population disequilibrium dynamics may be driving divergent functional compositional patterns of understory herbs in Amazonian dry regions.

Conclusions

Our study shows that external environmental filters, especially soil and topography, are the major determinants of functional composition among understory herbaceous communities across Amazonian forests. Local trait variation is also associated to external filters, partially due to the correlation between terrain instability and disturbance and partially by constraints imposed by high levels of soil nutrients. The changes in functional composition among communities are in agreement with expectations based on the classical fast-slow growth trade-off: acquisitive traits (low LMA, low stature and small seeds) predominate in resource-rich habitats (rich soil and bottomlands), while conservative traits predominate in resource-poor habitats, which is exactly the pattern observed for tropical forest tree communities. Mean functional composition was not associated with climate, on the contrary, functional composition widely diverged along the climatic gradient, potentially due to a long-term effect of past climate oscilation on population dynamics in dry regions. Recently, regional climate simulations have been confirmed a potential increase of dry conditions in southern Amazonia (Boisier *et al.* 2015) and thus, there is an urgent need in understanding how climate change will affect the structure and functioning of Amazonian forest (Olivares *et al.* 2015). Our findings revealed that terrain features (soil and topography) are crucial to predict composition functional across Amazonian rainforest but there is still much uncertainty in predicting which functional patterns will prevail in specific future climatic conditions.

Table 1: Summary results of mixed-effect GAMLSS models for CWD and CWM. The CWD models were fit using the zero adjusted Gamma distribution family and included only association between the mean (μ) parameter and predictor variables. The CWM models were fit using the three parameters *t* family distribution for LMA, the four parameters Box-Cox *t* for height and seed size, and included the association between mean and/or variance (σ) with predictor variables. Best models appear in bold and models where predictor variables were not significant ($P > 0.05$) are not shown. Null models are the first entry for each response variable. Values in brackets denote the scaled β coefficients.

Trait	Response variable	Model	ΔAIC	R^2
LMA	CWD	**(μ) ~ 1|random(sites)**	**0**	**0.35**
	CWM	(μ) ~ 1|random(sites)	63.28	0.38
		(μ) ~ (-2.52)SB + (1.18)HAND + 1|random(sites) **(σ) ~ 1**	**0**	**0.469**
		(μ) ~ 1 (σ) ~ (0.14)SB	58.59	0.392
Height	CWD	(μ) ~ 1|random(sites)	1.88	0.381
		(μ) ~ (-0.04)SB + (-0.06)HAND + 1|random(sites)	**0**	**0.386**
	CWM	(μ) ~ 1|random(sites)	45.83	0.398
		(μ) ~ (-0.04)SB + (-0.03)HAND + 1|random(sites)	41.73	0.407
		(μ) ~ 1|random(sites) (σ) ~ (-0.26)HAND + (0.23)DSL	3.57	0.457
		(μ) ~ (-0.05)SB + 1|random(sites) **(σ) ~ (-0.25)HAND + (0.23)DSL**	**0**	**0.463**
Seed size	CWD	(μ) ~ 1|random(sites)	11.73	0.469
		(μ) ~ (-0.04)SB + (-0.06)HAND + 1|random(sites)	**0**	**0.487**
	CWM	(μ) ~ 1|random(sites)	75.73	0.503
		(μ) ~ (-0.08)SB + (0.05)HAND + 1|random(sites)	37.54	0.548
		(μ) ~ 1|random(sites) (σ) ~ (-0.36)HAND + (0.23)DSL	49.44	0.536
		(μ) ~ (-0.07)SB + (0.08)HAND + 1|random(sites) **(σ) ~ (-0.58)HAND + (0.29)DSL**	**0**	**0.587**

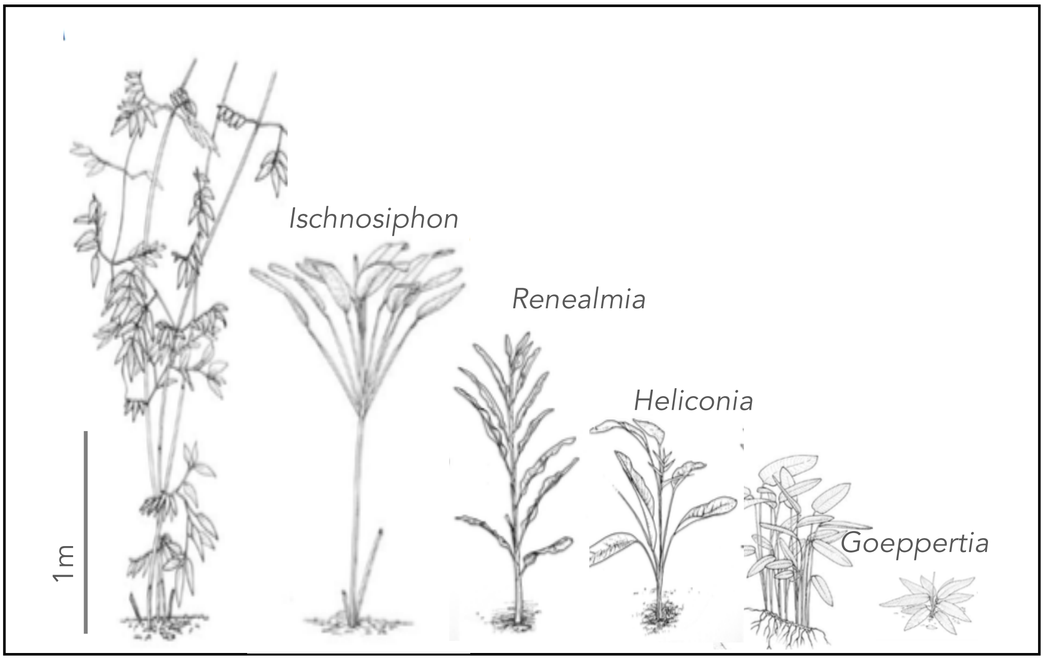
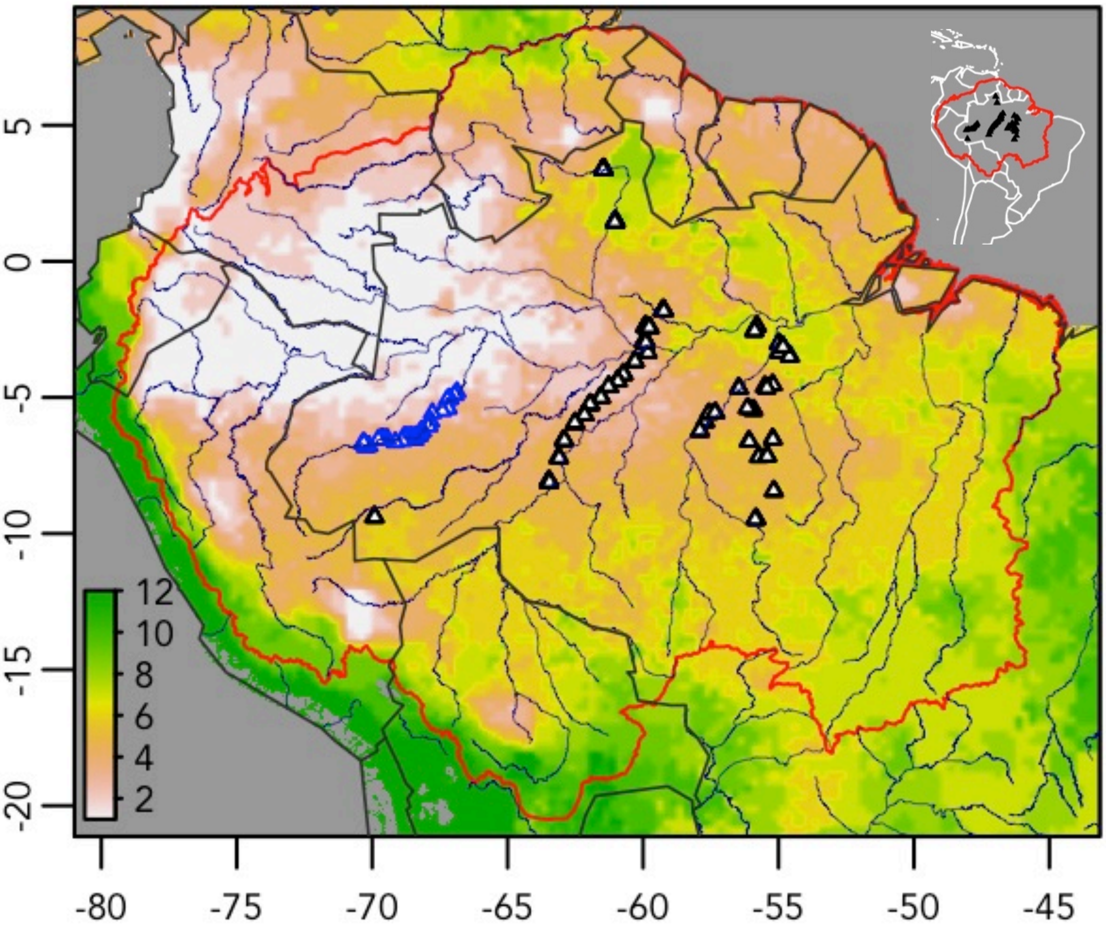


Figure 1: Life-forms in some genera of Zingiberales.

Figure 2. The study area and location of 451 plots where ginger community and environmental data were obtained. Background map represents the dry season length. Legend denotes the maximum number of consecutive months with less than 100 mm of precipitation along the 1998 - 2004 period. Black triangles represent the clusters of RAPELD plots (250 m x 2m) and blue triangles represent plots with size of 100 m x 5m located along the Juruá River. Clusters may have 5 to 57 plots spaced by at least 1 km.



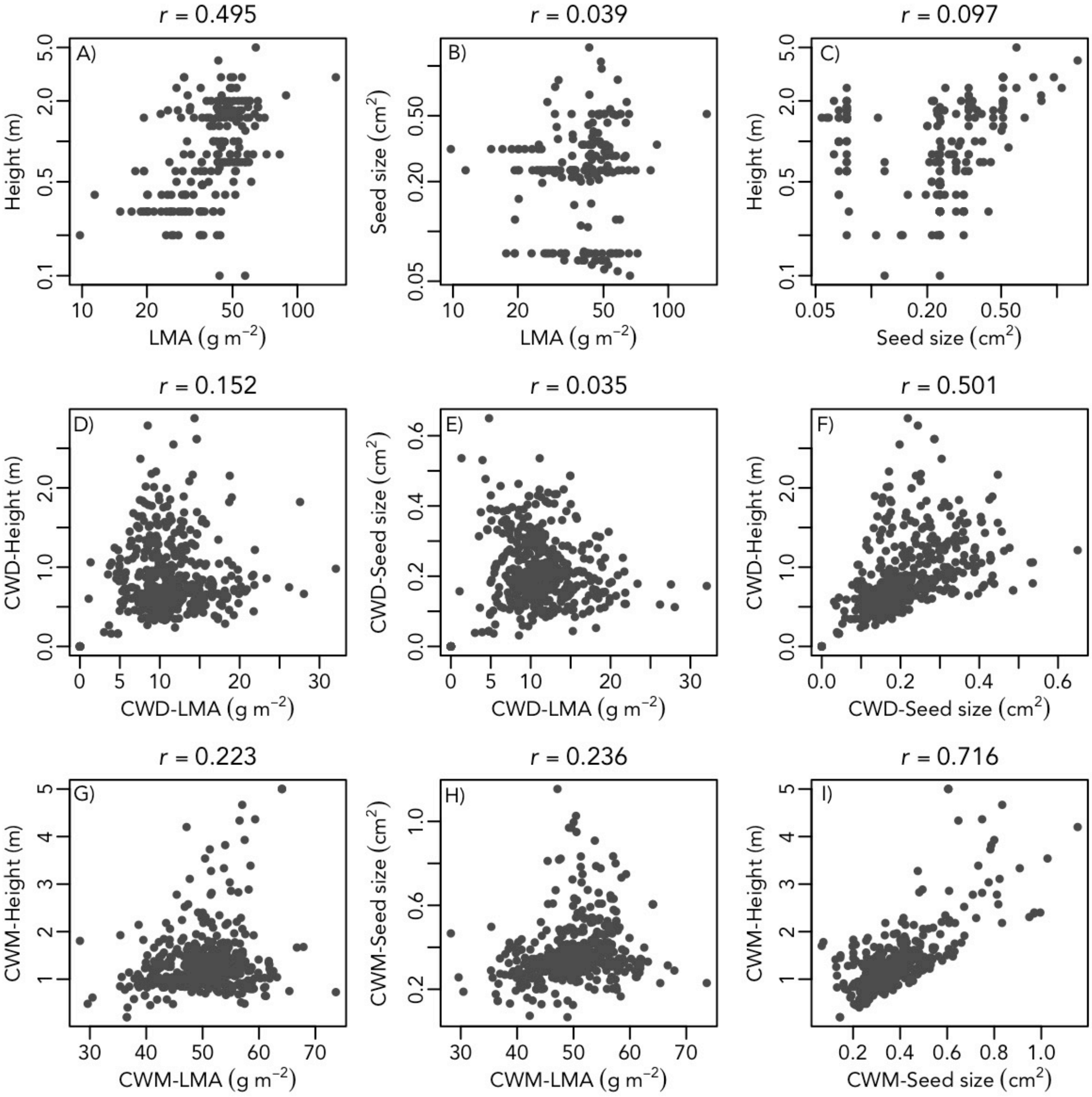


Fig 3. Correlations (Person's *r*) between functional traits at the species (A – C) and community levels (D – I). Correlations between traits at species level were obtained applying a logarithm transformation and vertical plot's axes are presented using that scale.

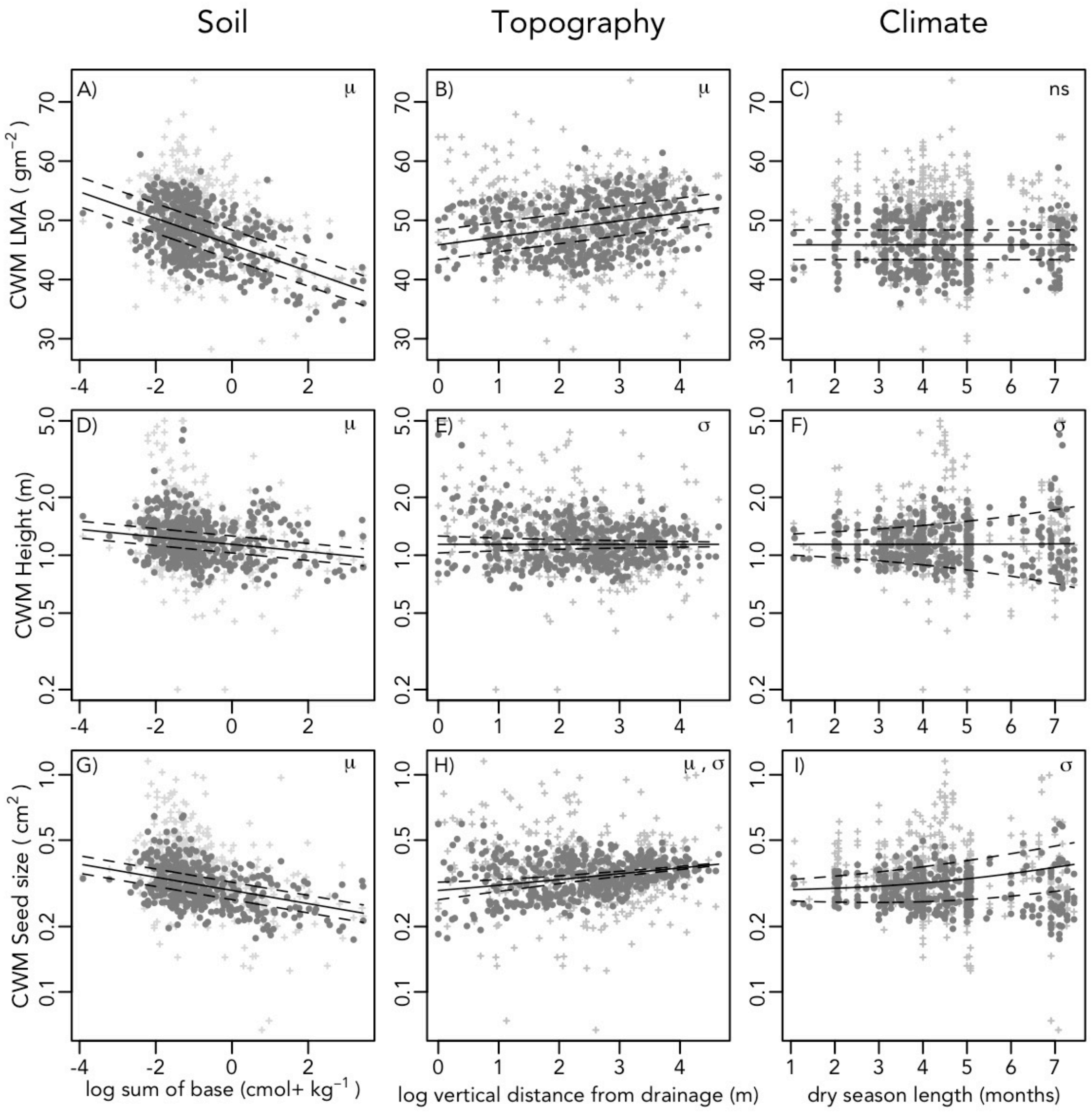


Fig 4. Changes in functional composition (CWM) of 451 herbaceous communities along environmental gradients. Continuous and dashed lines represent the mean (μ) and variance (σ) response of CWM, respectively. Predicted mean is the 0.5 quantile and predicted variance denotes the interval between the 0.25 and 0.75 quantiles of the *t* family distribution (A - C) and Box-Cox *t* distribution (D - L). For each variable the predicted mean and variance were obtained keeping all other variables constant at their mean values. The μ (mean) and/or σ (variance) symbols displayed in the upper-right corner of panels indicates which of these parameters had a significant association with predictor variables and ‘ns’ denotes non-significant association. Vertical axes from E to L are in logarithm scale. Dots denote the partial residual and light grey crosses represent the original values.

References

Bock, M. & Köthe, R. (2008) Predicting the depth of hydromorphic soil characteristics influenced by groud water. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, **19**, 13–22.

Boisier, J.P., Ciais, P., Ducharne, A. & Guimberteau, M. (2015) Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nature Climate Change*, **5**, 656–660.

Cheng, H., Sinha, A., Cruz, F.W., Wang, X., Edwards, R.L., D'Horta, F.M., Ribas, C.C., Vuille, M., Stott, L.D. & Auler, A.S. (2013) Climate change patterns in Amazonia and biodiversity. *Nature communications*, **4**, 1–6.

Cornwell, W.K. & Ackerly, D.D. (2009) Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, **79**, 109–126.

Cosme, L.H.M., Schietti, J., Costa, F.R.C. & Oliveira, R.S. (2017) The importance of hydraulic architecture to the distribution patterns of trees in a central Amazonian forest. *New Phytologist*.

Costa, F.R.C. (2004) Structure and composition of the ground-herb community in a terra-firme Central Amazonian forest. *Acta Amazonica*, **34**, 53–59.

Costa, F.R.C. & Magnusson, W.E. (2010) The need for large-scale, integrated studies of biodiversity – the experience of the Program for Biodiversity Research in Brazilian Amazonia. *Natureza & Conservação*, **8**, 3–12.

Costa, F.R.C., Magnusson, W.E. & Luizao, R.C. (2005) Mesoscale distribution patterns of Amazonian understorey herbs in relation to topography, soil and watersheds. *Journal of Ecology*, **93**, 863–878.

Dalling, J.W., Hubbell, S.P. & Silvera, K. (1998) Seed dispersal, seedling establishment and gap partitioning among tropical pioneer trees. *Journal of Ecology*, **86**, 674–689.

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I.C., Garnier, E., Bönisch, G., Westoby, M., Poorter, H., Reich, P.B., Moles, A.T., Dickie, J., Gillison, A.N., Zanne, A.E., Pierce, S., Shipley, B., Kirkup, D., Casanoves, F., Joswig, J.S., Günther, A. & Falczuk, V. (2016) The global spectrum of plant form and function. *Nature*, **529**, 167–171.

Díaz, S., Lavorel, S., de Bello, F., Quétier, F., Grigulis, K. & Robson, T.M. (2007)

Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences*, **104**, 20684–9.

Espírito-Santo, F.D.B., Keller, M., Braswell, B., Nelson, B.W., Frolking, S. & Vicente, G. (2010) Storm intensity and old-growth forest disturbances in the Amazon region. *Geophysical Research Letters*, **37**.

Ferry, B., Morneau, F., Bontemps, J.D., Blanc, L. & Freycon, V. (2010) Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest. *Journal of Ecology*, **98**, 106–116.

Figueiredo, F.O.G., Costa, F.R.C., Nelson, B.W. & Pimentel, T.P. (2014) Validating forest types based on geological and land-form features in central Amazonia. *Journal of Vegetation Science*, **25**, 198–212.

Fortunel, C., Paine, C.E.T., Fine, P.V.A., Kraft, N.J.B. & Baraloto, C. (2014) Environmental factors predict community functional composition in Amazonian forests. , 145–155.

Fyllas, N.M., Patiño, S., Baker, T.R., Nardoto, G.B., Martinelli, L.A., Quesada, C.A., Paiva, R., Schwartz, M., Horna, V., Mercado, L.M., Santos, A., Arroyo, L., Jimenez, E.M., Luizão, F., Neill, D.A., Silva, N., Prieto, A., Rudas, A., Silveira, M., Vieira, I.C.G., Lopez-Gonzalez, G., Malhi, Y., Phillips, O.L. & Lloyd, J. (2009) Basin-wide variations in foliar properties of Amazonian forest : phylogeny , soils and climate. *Biogeosciences*, **6**, 2677–2708.

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

Van Der Hammen, T. & Hooghiemstra, H. (2000) Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Science Reviews*, **19**, 725–742.

Higgins, M., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R. & Räsänen, M. (2011) Geological control of floristic composition in Amazonian forests. *Journal of biogeography*, **38**, 2136–2149.

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2017). dismo: Species Distribution Modeling. R package version 1.1-4. Available at https://CRAN.R- project.org/package=dismo

Hoffmann, W.A., Franco, A.C., Moreira, M.Z. & Haridasan, M. (2005) Specific leaf area explains differences in leaf traits between congeneric savanna and forest trees. *Functional Ecology*, **19**, 932–940.

Kitajima, K. (2002) Do shade-tolerant tropical tree seedling depend longer on seed reserves?

Functional growth analysis of three Bignoniaceae species. *Functional Ecology*, **16**, 433–444.

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

Kress, W.J., Prince, L.M., Hahn, W.J. & Zimmer, E.A. (2001) Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. *Systematic biology*, **50**, 926–944.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. & Simpson, J. (1998) The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.

Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, **15**, 237–240.

Maire, V., Wright, I.J., Prentice, I.C., Batjes, N.H., Bhaskar, R., van Bodegom, P.M., Cornwell, W.K., Ellsworth, D., Niinemets, Ü., Ordonez, A., Reich, P.B. & Santiago, L.S. (2015) Global effects of soil and climate on leaf photosynthetic traits and rates. *Global Ecology and Biogeography*, **24**, 706–717.

Moulatlet, G.M., Costa, F.R.C., Rennó, C.D., Emilio, T. & Schietti, J. (2014) Local hydrological conditions explain floristic composition in lowland Amazonian forests. *Biotropica*, **46**, 395–403.

Muscarella, R. & Uriarte, M. (2016) Do community-weighted mean functional traits reflect optimal strategies? *Proceedings of the Royal Society B*, **283**, 20152434.

Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.

Negrón-Juárez, R., Jenkins, H., Raupp, C., Riley, W., Kueppers, L., Magnabosco Marra, D., Ribeiro, G., Monteiro, M., Candido, L., Chambers, J. & Higuchi, N. (2017) Windthrow variability in Central Amazonia. *Atmosphere*, **8**, 28.

Olivares, I., Svenning, J.C., van Bodegom, P.M. & Balslev, H. (2015) Effects of warming and drought on the vegetation and plant Ddversity in the Amazon basin. *Botanical Review*, **81**, 42–69.

Pinheiro, T.F., Rennó, C.D. & Escada, M. isabel S. (2015) HAND terrain descriptor for mapping regional scale eco-hydrological units of Amazon terra firme. *Revista Brasileira de Cartografia*, **67**, 691–700.

Poorter, H., Niinemets, Ü., Poorter, L., Wright, I.J., Villar, R., Niinemets, U., Poorter, L., Wright, I.J. & Villar, R. (2009) Causes and consequences of variation in leaf mass per area (LMA):a meta-analysis. *New phytologist*, **182**, 565–588.

Quesada, C.A., Lloyd, J., Schwarz, M., Patiño, S., Baker, T.R., Czimczik, C., Fyllas, N.M., Martinelli, L., Nardoto, G.B., Schmerler, J., Santos, A.J.B., Hodnett, M.G., Herrera, R., Luizão, F.J., Arneth, A., Lloyd, G., Dezzeo, N., Hilke, I., Kuhlmann, I., Raessler, M., Brand, W.A., Geilmann, H., Moraes Filho, J.O., Carvalho, F.P., Araujo Filho, R.N., Chaves, J.E., Cruz Junior, O.F., Pimentel, T.P. & Paiva, R. (2010) Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences*, **7**, 1515–1541.

Quesada, C.A., Phillips, O.L., Schwarz, M., Czimczik, C.I., Baker, T.R., Patiño, S., Fyllas, N.M., Hodnett, M.G., Herrera, R., Almeida, S., Alvarez Dávila, E., Arneth, A., Arroyo, L., Chao, K.J., Dezzeo, N., Erwin, T., Di Fiore, A., Higuchi, N., Honorio Coronado, E., Jimenez, E.M., Killeen, T., Lezama, A.T., Lloyd, G., López-González, G., Luizão, F.J., Malhi, Y., Monteagudo, A., Neill, D.A., Núñez Vargas, P., Paiva, R., Peacock, J., Peñuela, M.C., Peña Cruz, A., Pitman, N., Priante Filho, N., Prieto, A., Ramírez, H., Rudas, A., Salomão, R., Santos, A.J.B., Schmerler, J., Silva, N., Silveira, M., Vásquez, R., Vieira, I., Terborgh, J. & Lloyd, J. (2012) Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences*, **9**, 2203–2246.

R Developement Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Avaliable at https://www.R-project.org .

Reich, P.B. (2014) The world-wide “fast-slow” plant economics spectrum: A traits manifesto. *Journal of Ecology*, **102**, 275–301.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J. & Waterloo, M.J. (2008) HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, **112**, 3469–3481.

Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, **54**, 507–554.

Rigby, R.A. & Stasinopoulos, D.M. (2007) Generalized additive models for location, scale and shape (GAMLSS) in R. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, **54**, 507–554.

Schietti, J., Emilio, T., Rennó, C.D., Drucker, D.P., Costa, F.R.C., Nogueira, A., Baccaro, F.B., Figueiredo, F., Castilho, C. V, Kinupp, V., Guillaumet, J.-L., Garcia, A.R.M., Lima, A.P. & Magnusson, W.E. (2013) Vertical distance from drainage drives floristic composition changes in an Amazonian rainforest. *Plant Ecology & Diversity*, **7**, 1–13.

Schietti, J., Martins, D., Emilio, T., Souza, P.F., Levis, C., Baccaro, F.B., Pinto, J.L.P. da V., Moulatlet, G.M., Stark, S.C., Sarmento, K., de Araújo, R.N.O., Costa, F.R.C., Schöngart, J., Quesada, C.A., Saleska, S.R., Tomasella, J., Magnusson, W.E. & Nardoto, G.B. (2016) Forest structure along a 600 km transect of natural disturbances and seasonality gradients in central-southern Amazonia. *Journal of Ecology*, **104**, 1335–1346.

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

Sterck, F., Markesteijn, L., Schieving, F. & Poorter, L. (2011) Functional traits determine trade-offs and niches in a tropical forest community. *Proceedings of the National Academy of Sciences*, **108**, 20,627-20,632.

Svenning, J.C. & Sandel, B. (2013) Disequilibrium vegetation dynamics under future climate change. *American journal of botany*, **100**, 1266–1286.

Swenson, N.G. (2009) Herbaceous monocot plant form and function along a tropical rain-forest light gradient: a reversal of dicot strategy – CORRIGENDUM. *Journal of Tropical Ecology*, **25**, 569.

Toledo, J.J., Magnusson, W.E., Castilho, C. V. & Nascimento, H.E.M. (2011) How much variation in tree mortality is predicted by soil and topography in Central Amazonia? *Forest Ecology and Management*, **262**, 331–338.

Tuomisto, H., Moulatlet, G.M., Balslev, H., Emilio, T., Figueiredo, F.O.G., Pedersen, D. & Ruokolainen, K. (2016) A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *Journal of Biogeography*, **43**, 2400–2411.

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

Westoby, M. (1998) A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, **199**, 213–227.

Westoby, M., Leishman, M., Lord, J., Poorter, H. & Schoen, D.J. (2010) Comparative ecology of seed size and dispersal. , **351**, 1309–1318.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K. & Gulias, J. The worldwide leaf economics spectrum. , **12**, 821–827.

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

Zuquim, G., Tuomisto, H., Jones, M.M., Prado, J., Figueiredo, F.O.G., Moulatlet, G.M., Costa, F.R.C., Quesada, C.A. & Emilio, T. (2014) Predicting environmental gradients with fern species composition in Brazilian Amazonia. *Journal of Vegetation Science*, **25**, 1195–1207.

ARTIGO 2 – MATERIAL SUPLEMENTAR

Table S1: List of the 192 species in the ginger order of flowering plants and their functional trait values. LMA denotes leaf mass per unit of area.

Family	Genus	Species	LMA (gm^{-2})	Height (m)	Seed size (cm^2)
Costaceae	Chamaecostus	*Chamaecostus acaulis*	26.59	0.2	0.074
Costaceae	Chamaecostus	*Chamaecostus congestiflorus*	71.93	0.8	0.074
Costaceae	Chamaecostus	*Chamaecostus fragilis*	33.46	0.6	0.074
Costaceae	Chamaecostus	*Chamaecostus fusiformis*	45.89	0.7	0.074
Costaceae	Chamaecostus	*Chamaecostus sp1*	19.34	0.6	0.074
Costaceae	Chamaecostus	*Chamaecostus sp1 pec*	17.67	0.6	0.074
Costaceae	Chamaecostus	*Chamaecostus sp2*	29.26	0.6	0.074
Costaceae	Costus	*Chamaecosuts fragilis*	27.18	0.6	0.074
Costaceae	Costus	*Costus acreanus*	53.52	1.7	0.074
Costaceae	Costus	*Costus amazonicus*	59.98	2.0	0.074
Costaceae	Costus	*Costus arabicus*	42.23	1.8	0.074
Costaceae	Costus	*Costus fusiformis*	57.44	0.7	0.074
Costaceae	Costus	*Costus guanaiensis var guanaiensis*	57.09	2.0	0.074
Costaceae	Costus	*Costus laisus*	42.11	1.0	0.074
Costaceae	Costus	*Costus lasius*	40.67	1.0	0.074
Costaceae	Costus	*Costus scaber*	54.11	1.5	0.074
Costaceae	Costus	*Costus sp1 acre*	43.45	1.5	0.074
Costaceae	Costus	*Costus sp1 cunia*	42.29	1.5	0.074
Costaceae	Costus	*Costus sp1 m01*	40.58	1.5	0.074
Costaceae	Costus	*Costus sp1 pec*	19.36	1.5	0.074
Costaceae	Costus	*Costus sp1 VIRMAR*	46.11	1.5	0.074
Costaceae	Costus	*Costus sp2*	27.85	1.7	0.074
Costaceae	Costus	*Costus sp2 m01*	40.58	1.5	0.074

Costaceae	Costus	*Costus sp3 pec*	33.07	1.5	0.074
Costaceae	Costus	*Costus sp4*	23.16	1.6	0.074
Costaceae	Costus	*Costus sp5*	31.79	1.8	0.074
Costaceae	Costus	*Costus sp6*	31.61	1.7	0.074
Costaceae	Costus	*Costus sp7*	40.58	1.5	0.074
Costaceae	Costus	*Costus sp7 pec*	59.03	1.8	0.074
Costaceae	Costus	*Costus sp8*	54.98	1.7	0.074
Costaceae	Costus	*Costus spiralis*	65.52	2.0	0.074
Costaceae	Costus	*Costus sprucei*	40.58	1.5	0.074
Costaceae	Costus	*Costus zingiberoides*	28.81	1.5	0.074
Costaceae	Dimerocostus	*Dimerocostus strobilaceus*	49.30	2.5	0.074
Heliconiaceae	Heliconia	*Heliconia 46*	39.79	2.0	0.334
Heliconiaceae	Heliconia	*Heliconia acuminata*	43.78	0.9	0.296
Heliconiaceae	Heliconia	*Heliconia aemygdiana*	30.92	2.2	0.821
Heliconiaceae	Heliconia	*Heliconia bihai*	48.09	2.0	0.350
Heliconiaceae	Heliconia	*Heliconia brachyantha*	52.14	1.4	0.236
Heliconiaceae	Heliconia	*Heliconia chartacea*	44.46	2.2	0.334
Heliconiaceae	Heliconia	*Heliconia den angustifolia*	36.10	1.5	0.361
Heliconiaceae	Heliconia	*Heliconia den densiflora*	30.73	1.0	0.361
Heliconiaceae	Heliconia	*Heliconia hirsuta*	25.05	1.7	0.334
Heliconiaceae	Heliconia	*Heliconia juliani*	53.20	2.0	0.304
Heliconiaceae	Heliconia	*Heliconia juruana*	88.56	2.2	0.334
Heliconiaceae	Heliconia	*Heliconia lasiorachis*	42.89	1.7	0.247
Heliconiaceae	Heliconia	*Heliconia lourtegiae*	51.42	0.6	0.334
Heliconiaceae	Heliconia	*Heliconia psittacorum*	40.71	1.3	0.334
Heliconiaceae	Heliconia	*Heliconia rostrata*	49.69	2.5	0.334

Heliconiaceae	Heliconia	*Heliconia schumanniana*	52.14	1.6	0.227
Heliconiaceae	Heliconia	*Heliconia sp1 chand*	43.03	0.8	0.334
Heliconiaceae	Heliconia	*Heliconia sp1 pec*	44.09	1.5	0.334
Heliconiaceae	Heliconia	*Heliconia spathocircinata*	42.73	1.8	0.334
Heliconiaceae	Heliconia	*Heliconia stricta*	44.27	1.7	0.371
Heliconiaceae	Heliconia	*Heliconia tenebrosa*	48.72	0.8	0.334
Heliconiaceae	Heliconia	*Heliconia velutina*	46.80	1.8	0.350
Marantaceae	Calathea	*Calathea crotalifera*	45.10	1.7	0.225
Marantaceae	Calathea	*Calathea lutea*	52.20	2.5	0.250
Marantaceae	Calathea	*Calathea striata*	44.49	0.9	0.237
Marantaceae	Ctenanthe	*Ctenanthe eriacae*	37.86	2.0	0.212
Marantaceae	Ctenanthe	*Ctenanthe sp1 acre*	35.66	0.2	0.212
Marantaceae	Ctenanthe	*Ctenanthe sp1 pime*	32.03	0.5	0.212
Marantaceae	Goeppertia	*Goeppertia altissima*	63.53	1.3	0.454
Marantaceae	Goeppertia	*Goeppertia attenuata*	53.23	1.0	0.234
Marantaceae	Goeppertia	*Goeppertia brev m10*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia cannoides*	70.44	1.5	0.234
Marantaceae	Goeppertia	*Goeppertia capitata*	42.48	2.0	0.217
Marantaceae	Goeppertia	*Goeppertia comosa*	46.39	2.0	0.217
Marantaceae	Goeppertia	*Goeppertia curaraya*	35.19	0.2	0.234
Marantaceae	Goeppertia	*Goeppertia elliptica*	50.49	0.7	0.234
Marantaceae	Goeppertia	*Goeppertia exscapa*	29.13	0.4	0.234
Marantaceae	Goeppertia	*Goeppertia fragilis*	50.65	1.7	0.234
Marantaceae	Goeppertia	*Goeppertia fucata*	42.34	0.2	0.106
Marantaceae	Goeppertia	*Goeppertia hopkinsii*	25.45	0.7	0.234
Marantaceae	Goeppertia	*Goeppertia loeseneri*	37.89	0.8	0.212

Marantaceae	Goeppertia	*Goeppertia maassiorum*	50.50	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia metallica*	25.97	0.4	0.196
Marantaceae	Goeppertia	*Goeppertia micans*	36.56	0.2	0.144
Marantaceae	Goeppertia	*Goeppertia microcephala*	43.84	0.2	0.147
Marantaceae	Goeppertia	*Goeppertia neblinensis*	42.38	0.8	0.234
Marantaceae	Goeppertia	*Goeppertia ovata*	27.44	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia panamensis*	19.98	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia picturata*	40.80	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia polytricha*	29.03	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp m2*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp m3*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp m4*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp m5*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp m9*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp1 acre*	30.34	0.4	0.234
Marantaceae	Goeppertia	*Goeppertia sp1 brev uat*	35.93	0.7	0.234
Marantaceae	Goeppertia	*Goeppertia sp1 pec*	24.65	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp11 altissima*	52.17	1.5	0.454
Marantaceae	Goeppertia	*Goeppertia sp15*	20.26	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp16*	30.22	0.4	0.234
Marantaceae	Goeppertia	*Goeppertia sp17*	28.02	0.2	0.234
Marantaceae	Goeppertia	*Goeppertia sp18*	19.33	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp19*	37.08	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp2 orna vir*	65.39	1.8	0.234
Marantaceae	Goeppertia	*Goeppertia sp20*	11.44	0.4	0.234
Marantaceae	Goeppertia	*Goeppertia sp21*	35.31	0.3	0.234

Marantaceae	Goeppertia	*Goeppertia sp22*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp24*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp25*	25.69	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp26*	35.59	0.6	0.234
Marantaceae	Goeppertia	*Goeppertia sp28*	36.09	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp29*	61.06	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp3*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp3 acre*	26.56	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp30*	37.00	0.6	0.234
Marantaceae	Goeppertia	*Goeppertia sp31*	23.40	0.4	0.234
Marantaceae	Goeppertia	*Goeppertia sp32*	31.53	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp33*	43.49	0.1	0.234
Marantaceae	Goeppertia	*Goeppertia sp35*	30.01	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp36*	24.56	0.2	0.234
Marantaceae	Goeppertia	*Goeppertia sp38*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp39*	36.17	0.5	0.234
Marantaceae	Goeppertia	*Goeppertia sp4*	48.43	0.7	0.234
Marantaceae	Goeppertia	*Goeppertia sp40*	28.12	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia sp41*	44.18	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia straminea*	82.77	0.8	0.234
Marantaceae	Goeppertia	*Goeppertia taeniosa*	35.12	0.3	0.234
Marantaceae	Goeppertia	*Goeppertia variegata*	44.55	1.7	0.454
Marantaceae	Goeppertia	*Goeppertia zingiberina*	62.30	0.8	0.279
Marantaceae	Hylaeanthe	*Hylaeanthe hexantha*	20.19	0.4	0.157
Marantaceae	Hylaeanthe	*Hylaeanthe unilateralis*	19.41	0.6	0.118
Marantaceae	Ischnosiphon	*Ischnosiphon arouma*	58.24	2.0	0.825

Marantaceae	Ischnosiphon	*Ischnosiphon cannoideus*	43.40	1.3	0.200
Marantaceae	Ischnosiphon	*Ischnosiphon canoideus*	39.28	1.3	0.200
Marantaceae	Ischnosiphon	*Ischnosiphon cerotus*	43.01	1.5	0.670
Marantaceae	Ischnosiphon	*Ischnosiphon crassispicus*	150.87	3.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon gracilis*	47.50	1.7	0.503
Marantaceae	Ischnosiphon	*Ischnosiphon grandibracteatus*	49.00	3.0	0.962
Marantaceae	Ischnosiphon	*Ischnosiphon hirsutus*	32.14	0.8	0.261
Marantaceae	Ischnosiphon	*Ischnosiphon killipii*	29.82	3.0	0.746
Marantaceae	Ischnosiphon	*Ischnosiphon lasiocoleus*	47.20	1.6	0.393
Marantaceae	Ischnosiphon	*Ischnosiphon leucophaeus*	58.62	0.8	0.234
Marantaceae	Ischnosiphon	*Ischnosiphon longiflorus*	42.92	4.0	1.292
Marantaceae	Ischnosiphon	*Ischnosiphon martianus*	57.32	1.2	0.503
Marantaceae	Ischnosiphon	*Ischnosiphon obliquus*	27.44	2.5	0.605
Marantaceae	Ischnosiphon	*Ischnosiphon paryrizinho*	60.17	2.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon petiolatus*	53.69	0.9	0.548
Marantaceae	Ischnosiphon	*Ischnosiphon pub puberulus*	35.55	2.5	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon pub verruculosus*	48.63	2.5	1.060
Marantaceae	Ischnosiphon	*Ischnosiphon sp1*	29.74	3.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp10*	48.24	1.9	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp2*	55.22	3.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp4*	50.29	3.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp5*	59.72	1.5	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp6*	48.24	1.9	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp7*	65.71	1.8	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp8*	50.65	2.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon sp9*	44.16	3.0	0.512

Marantaceae	Ischnosiphon	*Ischnosiphon surumuensis*	55.39	2.0	0.512
Marantaceae	Ischnosiphon	*Ischnosiphon ursinus*	47.18	1.3	0.512
Marantaceae	Maranta	*Maranta ciclophylla*	16.97	0.3	0.314
Marantaceae	Maranta	*Maranta humilis*	20.07	0.4	0.314
Marantaceae	Maranta	*Maranta sp1 acre*	23.29	0.3	0.314
Marantaceae	Maranta	*Maranta sp1 jur*	21.55	0.3	0.314
Marantaceae	Maranta	*Maranta sp1 pec*	20.22	0.3	0.314
Marantaceae	Maranta	*Maranta sp1 pime*	25.78	0.3	0.314
Marantaceae	Maranta	*Maranta sp2 acre*	9.77	0.2	0.314
Marantaceae	Maranta	*Maranta sp2 jur*	24.94	0.4	0.314
Marantaceae	Maranta	*Maranta sp2 pime*	18.75	0.3	0.314
Marantaceae	Maranta	*Maranta sp3 pime*	15.04	0.3	0.314
Marantaceae	Monophyllanthe	*Monophyllanthe araracuarensis*	59.57	0.7	0.118
Marantaceae	Monophyllanthe	*Monophyllanthe oligophylla*	57.17	0.1	0.118
Marantaceae	Monotagma	*Monotagma angustissimum*	44.98	1.0	0.396
Marantaceae	Monotagma	*Monotagma aurantispathum*	29.64	0.8	0.227
Marantaceae	Monotagma	*Monotagma breviscapum*	35.96	0.7	0.385
Marantaceae	Monotagma	*Monotagma contractum*	47.59	0.7	0.205
Marantaceae	Monotagma	*Monotagma contrariosum*	44.15	1.0	0.271
Marantaceae	Monotagma	*Monotagma densiflorum*	55.37	1.3	0.227
Marantaceae	Monotagma	*Monotagma exile*	41.66	0.4	0.248
Marantaceae	Monotagma	*Monotagma flavicomum*	30.30	0.3	0.428
Marantaceae	Monotagma	*Monotagma floribundum*	63.34	1.6	0.286
Marantaceae	Monotagma	*Monotagma humile*	29.77	0.4	0.225
Marantaceae	Monotagma	*Monotagma juruanum*	42.57	0.7	0.440
Marantaceae	Monotagma	*Monotagma laxum*	50.52	1.0	0.302

Marantaceae	Monotagma	*Monotagma lilacinum*	46.24	0.7	0.221
Marantaceae	Monotagma	*Monotagma plurispicatum*	48.76	0.8	0.271
Marantaceae	Monotagma	*Monotagma secundum*	53.25	0.7	0.324
Marantaceae	Monotagma	*Monotagma sp1*	57.18	0.7	0.286
Marantaceae	Monotagma	*Monotagma sp1 acre*	54.93	0.7	0.286
Marantaceae	Monotagma	*Monotagma sp2 pime*	51.91	0.7	0.286
Marantaceae	Monotagma	*Monotagma sp3 m10*	35.78	0.3	0.286
Marantaceae	Monotagma	*Monotagma sp3 pime*	46.19	0.4	0.286
Marantaceae	Monotagma	*Monotagma sp4 pime*	42.75	0.6	0.286
Marantaceae	Monotagma	*Monotagma sp5 cunia*	43.50	0.6	0.286
Marantaceae	Monotagma	*Monotagma sp5 jur*	42.99	0.7	0.286
Marantaceae	Monotagma	*Monotagma tomentosum*	51.75	0.6	0.319
Marantaceae	Monotagma	*Monotagma tuberosum*	26.04	0.2	0.226
Marantaceae	Monotagma	*Monotagma ulei*	38.73	0.5	0.211
Marantaceae	Monotagma	*Monotagma vaginatum*	44.93	0.7	0.402
Strelitziaceae	Phenakospermum	*Phenakospermum guyannense*	64.07	5.0	0.605
Zingiberaceae	Renealmia	*Renealmia acreana*	50.38	1.5	0.059
Zingiberaceae	Renealmia	*Renealmia alpinia*	44.14	2.0	0.063
Zingiberaceae	Renealmia	*Renealmia aromatica*	48.93	1.7	0.067
Zingiberaceae	Renealmia	*Renealmia breviscapa*	52.62	0.8	0.063
Zingiberaceae	Renealmia	*Renealmia cernua*	58.16	1.5	0.057
Zingiberaceae	Renealmia	*Renealmia floribunda*	66.37	1.5	0.054
Zingiberaceae	Renealmia	*Renealmia monosperma*	40.56	0.3	0.076
Zingiberaceae	Renealmia	*Renealmia sp1*	38.09	1.5	0.067
Zingiberaceae	Renealmia	*Renealmia sp1 br319*	45.85	1.0	0.067
Zingiberaceae	Renealmia	*Renealmia sp1 pec*	45.85	1.0	0.067

Zingiberaceae	Renealmia	*Renealmia sp1 pime*	45.85	1.0	0.067
Zingiberaceae	Renealmia	*Renealmia sp10*	49.30	1.5	0.067
Zingiberaceae	Renealmia	*Renealmia sp4 jur*	32.88	0.4	0.067
Zingiberaceae	Renealmia	*Renealmia sp5 jur*	51.48	0.4	0.067
Zingiberaceae	Renealmia	*Renealmia sp7 jur*	39.76	0.6	0.067
Zingiberaceae	Renealmia	*Renealmia sp8*	40.76	1.0	0.067
Zingiberaceae	Renealmia	*Renealmia sp9 jur*	45.85	1.0	0.067
Zingiberaceae	Renealmia	*Renealmia thyrsoidea*	39.24	1.5	0.108

ARTIGO 3

Landscape evolution, functional trade-offs and boosted diversification in Neotropical forests

FIGUEIREDO, F.O.G, TERRA, M.H.A, SAKA, M., CAMPOS, J.V., ANDRÉ, T., MOULATLET, G., ZUQUIM, G., EMÍLIO, T., TUOMISTO, H., BORCHSENIUS, F., BALSLEV, H., SILVEIRA, M., DOMINGOS, J.R.& COSTA, F. R. C. Manuscrito em preparação para *PNAS.*

ABSTRACT

We propose a diversification model for Amazonia based on intrinsic clade functional traits, habitat association and past geological events. Using seven clades of Marantaceae, a species rich Neotropical family of monocots, we demonstrate that short versus long generation time, corresponding respectively to fast growth rate on fertile substrate versus slow growth rate on infertile substrate, is the major determinant of widely distinct diversification rates observed across the Marantaceae. Species richness in each lineage is well predicted by these functional constraints and not by clade age. Clades with acquisitive strategies were often associated with highly productive habitats – nutrient-rich soils and bottomlands – and their origin and diversification dynamics match past geological events mediated by Andean uplift. We further show that a pure allopatric-dispersal model disconnected from past geological and ecological forces is inadequate to explain floristic evolutionary and diversity patterns in Amazonian lowlands. A major geochemical and physical transformation of the Amazon landscape in the Miocene, driven by Andean uplift, increased soil fertility over large portions of Amazonia, providing the opportunity for the evolution of clades with fast growth and short generation times, boosting rates of molecular evolution and of diversification in Amazonian forests.

INTRODUCTION

Amazonia harbors the greatest diversity of plants and animal in the world. Such spectacular species accumulation through time has been linked with several historical events and complex processes (1–3), but a mechanistic explanation, linking past events to diversification dynamics and diversity patterns, remains elusive. Historical events include climate oscillations and geological-landscape transformations (4, 5). The refuge theory based on Pleistocene climate oscillations, popular in the 1970s, was gradually replaced by alternative models as its predictions did not find support in the spatial and temporal evidence (1, 6, 7). With the advance of paleo-ecological and molecular studies, geological events coupled with landscape transformation have emerged as key elements to explain biological diversification patterns in Amazonia (5, 8, 9).

The intense physical landscape change caused by Andean uplift during the Neogene is postulated to have promoted speciation in tetrapodes via an allopatric-vicariance process (8). In this model, the emergence of mountains and rivers isolated widespread populations, interrupting their gene flow and forming new species. Recently, the allopatric-vicariance model has been criticized because spatial and temporal genetic patterns between closely related lineages vary widely and do not match the timing of geological events (10). In contrast to the vicariance model, the dispersal model assumes that allopatric speciation is driven by dispersal events and the speciation outcome from organism-specific abilities to persist and disperse in the landscape (10). According to this model, diversity is expected to be higher in old clades as they have had more time to accumulate species and also higher in clades with poor dispersal abilities becuase gene flow is more restricted (10). The allopatric-vicariance model also does not find support in plant phylogenies (3). Molecular analysis of four species rich and abundant clades of Neotropical trees showed that local forest communities in Amazonia are comprised of lineages widely randomly distributed across the phylogenetic tree (11). It therefore seems unlike that vicariance events have pooled common evolutionary histories into distinct and isolated geographic regions. These recent findings refute the idea of a deep imprint of geological events and their associated barriers on the structure of diversification and of evolutionary patterns in lowland rain forest of Amazonia.

The transformation caused by the Andean uplift, however, goes beyond the emergence of mountain and river barriers. The intense geological activity promoted an intense input of nutrients mainly in Western Amazonia and a gradual transformation of lowlands substrates that extends to Central and Eastern Amazonia (5, 12). From the late Oligocene onward, the predominant uplands habitats with low productivity of the Amazonia Craton were replaced by a high productivity bottomland landscape (13). These geo-physical gradients are strongly related to several spatial diversity and evolutionary patterns in Amazonian plants. Floristic differences are well documented along bottomland to uplands gradients of non-inundated forest (14, 15). Soil derived from the Miocene mega-wetlands are markedly richer in nutrients, harbor a distinct floristic composition and have higher species richness compared to forest growing on soils derived from nutrient-poor substrates(16–18). Species-habitat associations mapped within plant phylogenies has revealed that (1) closely related species have different soil affinities, (2) multiple events of speciation probably occurred by parapatric speciation across a heterogeneous edaphic landscape and (3) the origin of species associated with clay soils coincided with sedimentation history in the Miocene (9). All these findings suggest that the Andean uplift may have left a deep imprint in diversification patterns through the chemical and physical transformation of the broader Amazon landscape, favoring species divergence along this ecological dimension. However, it remains elusive how these past geological events can be linked to explain large differences in diversity and speciation rates across tropical plant lineages.

Availability of resources is tightly related to plant functional strategies. The topographic and soil resource gradients of tropical forests have been shown to be associated with these strategies (19–21). Species with acquisitive functional traits (*e.g.* low woody density, low stature and low leaf mass per unit of area), having high turnover rates (low longevity), predominate in nutrient rich habitats and bottomlands, while conservative functional strategies are more common in nutrient poor habitats and uplands. An interesting study combining phylogenies across 51 tree genera and plot-based demographic data in Amazonia showed that clades of species with fast growth strategy (and short generation times) showed markedly high rates of diversification and tended to be richer than clades with a slow growth strategy (22). These findings inspired us to build and test a mechanistic model framework linking geological history, habitat heterogeneity, intrinsic functional traits of lineages and the diversification patterns of Amazonian flora. The key point is to include past geological events and the appearance of resource rich soils as a driver of evolution of species with short generation times and high speciation rates.

Here, we use phylogenetic data and a functional trait proxy for fast versus slow growth strategies for a set of 100 species from seven clades of a species-rich understory herbaceous family (Marantaceae, ~230 recognized species in Amazonian), found in 400 standard plots, each having environmental data (soil and topography), to present an ecological model of diversification based on the history of geochemical and physical transformation of Amazonian substrate. We hypothesize that, when low productivity upland habitats were replaced in the Western Amazon by higher productivity bottomland habitats, due to the intensification of Andean uplift at the Oligocene-Miocene boundary (~23 Mya), the origin of new clades with acquisitive functional traits was favoured, and that shorter life cycles of these fast-growing taxa lead to higher diversification rates in these new environments. By contrast, old clades are expected to have conservative functional traits (longer life cycles), low diversification rates and to be associated with habitats resembling the Amazonian landscape that predominated prior to the early stages of the Andean uplift (upland habitats with nutrient poor soil). In terms of current diversity patterns, we expect that variation in species richness across clades will be better predicted by clade functional traits and by diversification rates than by clade age because the increase in diversification rates of fast-growt and young clades must overcome the species acuumulation in old clades.Finally, we explore the diversification dynamics of clades with distinct functional strategies, searching for temporal congruences between shifts in diversification rates and the major paleo geological and climatic events.

RESULTS

Clade functional traits, and not age, explain diversification rates and diversity patterns

The clade functional strategy here is represented by the leaf mass per area (LMA, g m^{-2}). LMA describes a spectrum of leaf economics, running from fast to slow return on investments of resources and of dry mass in leaves, and is related with organisms' fitness and population demographic rates (23, 24). Species with low LMA often have the acquisitive strategy: fast growth, high mortality rates and short generation times.The opposite is generally true for species with high LMA (conservative strategy). For the diversification rates, six models were tested, running from simple (includes only births with constant speciation rates) to more complex (include births deaths with speciation varying exponentially over time)

Species richness, diversification rates (*r)*, ages and mean LMA values vary markedly across the seven lowland rain forests lineages of Marantaceae (Fig. 1). For all seven clades, diversification rates were best estimated by a pure birth model process, i.e., with no extinctions and constant speciation rates. The three clades with the highest diversification rates (Breviscapus, Comosae and Microcephalum) were also those with the lowest LMA, *i.e*., clades with acquisitive functional strategy, and in average were younger than the four remaining clades (Fig. 1). In fact, the intrinsic diversification rates were strongly dependent on the clade functional trait (Fig. 2a), and both diversification rate and LMA predicted relatively well the species richness across the seven lineages (Fig 2b –d). Clades with an acquisitive strategy had higher diversification rates and were richer than those with conservative strategies (Fig, 2b-d). Contrary to predictions of the age-dispersal model, species richness was not associated with clade age, (Fig, 2c and 2e).

Clade habitat-association related to functional traits and time of origin

The habitat association of the seven clades was also clearly related to the diversification rates, functional strategies and time of origin. Clades that originated before 30 Mya, with conservative traits and low diversification rates (Straminea, *Monotagma*, *Ischnosiphon* and Ornata) tended to be associated to soils with low to medium levels of nutrients (Fig 3). Straminea and Ornata clades were also associated with uplands, while *Monotagma* and

Ischnosiphon did not show any topographic preferences (Fig 3). On the other hand, the three lineages that originated between 25 – 20 Mya (Microcephalum, Comosae and Breviscapus), those with the highest diversification rates and acquisitive functional strategies, had better associated to rich soils. The Breviscapus clade was more variable in terms of association to the soil gradient, with high preferences for both poor or rich soils, and avoiding soils with intermediate levels of nutrients. All clades had preferences to wet conditions and tended to avoid dry conditions (Fig 3).

Diversification dynamics related to paleo events.

Until now, we assumed that diversifications rates can only be constant, or vary linearly or exponentially through time. The analysis employed in this section relaxes this assumption allowing speciation and extinction to freely vary in any time. We revealed first, an impressive congruence between diversification dynamics and paleo geological and climatic events, and second, markedly distinct behaviour dynamics between clades with different functional strategies (Fig 4). The diversification dynamics of conservative clades started around 45 Mya in the middle Eocene and the speciation and extinction rates kept low and constant until the late Oligocene – early Miocene (~20 Mya). This period marks a phase of intensification of Andean uplift and the beginning of the fluvial-wetland system (~23 Mya) and match with a slight increase in the speciation and diversification rates. After that, a new constant regime in the speciation and extinction rates of these clades predominates until the present. By contrast, the diversification dynamics of acquisitive clades started later and was more dynamic than the diversification history of conservative clades (Fig 4). The most recent common ancestor of acquisitive clades (Breviscapus, Comosae e Microcephalum) dated from 23.7 (19.7 – 29.8, 95% confidence interval) and matched again with the early stages of the western fluvial-wetlands system. The speciation rates kept constant for the following 10 Mya. Interestingly, the extinction rates of this group slightly increase until the middle Miocene (~15 Mya), making indistinguishable the differences between diversification rates of acquisitive and conservative clades. The speciation rates of acquisitive clades only started to speed up slightly before to the end of the Pebas phase and the beginning of the Acre system (~10 Mya), the same period where extinction rates started to decline resulting in a rapid increase in diversification rates. Speciation and diversification rates of fast growth clades reached a peak around the end of Pliocene, followed by an abrupt decline in the beginning of Pleistocene (~3 Mya) which stabilized only in the last million years.

DISCUSSION

Intrinsic clade functional trait-strategy drives diversification rates and diversity patterns.

Our results clearly establish a link between an intrinsic clade functional trait (LMA), past geological events, diversification dynamics and current species diversity. We found a strong functional-evolutionary trade-off structuring the phylogenetic patterns of seven major clades of Marantaceae in Amazonia lowland rain forest. Clades with an acquisitive functional strategy (low LMA) tended to be younger, had high diversification rates and higher species richness than clades with a conservative strategy (high LMA). The acquisitive strategy is associated with short generation times, which speed up the rate of molecular evolution resulting in an increase of speciation events (Smith & Donoghue 2008), independent of the mechanism involved in speciation – vicariance, isolation by distance or ecological divergence. By contrast, species with high LMA invest in high durable tissues and consequently, tend to have long generation times, low rates of molecular evolution and low rates of speciation. Our findings are in perfect accordance with a previous study in Amazonia which showed that tree genera with high turnover rates (high rates of mortality) tend to have also high diversification rates (22).

Assuming a neutral process, where speciation and extinction rates are randomly distributed in the phylogeny, old clades would have more species merely due to the larger time to accumulate species than new clades (25). The age-diversity relationship is invoked to partially explain differences in diversity between regions (*e.g.* latitudinal gradient in species richness) and across clades within biomes (26–28). Within the Neotropical, the temporal constraining effect on clade's species accumulation was found to be a major determinant of variation in species richness across bird lineages (10). Our findings for Amazonian Marantaceae clades do not support the age-diversity model. On the contrary, recent clades of Marantaceae often exhibited higher species richness than old clades because the speed up in diversification rates of young clades driven by the origin of fast-growth strategy was big enough to counterattack the temporal restriction for species accumulation. Although the age-diversity model has not been directly contrasted with functional-diversification model in the previous tropical studies, it seems plausible that functional-diversification trade-off controls also the diversity pattern across tropical tree clades, instead of the age-diversity relationship. The *Inga* genus is emblematic here, since it is the richest genus and have the highest diversification rates among Amazonian trees, most of species are fast-growth and its origin (6 Mya) and radiation (4 Mya) are very recent (22, 29).

Deep imprint of paleo-geochemical transformation on evolutionary patterns in Marantaceae

The functional fast-slow trade-off observed across our seven lineages correlates with their preferences along the resource-substrate gradients. This relationship agrees with previous studies demonstrating the soil-fertility gradient as the main driver of the fast-slow growth trade-off in tropical forests (20, 30). However, the role of the soil-resource gradient as a driver of ecological patterns transcends ecological time scales and goes deep into the evolutionary/geological scale as we have shown here. The slow-growth clades - associated to soils with low or intermediate levels of nutrients - evolved early in the Eocene, while the emergence of fast-growth clades – associated with nutrient-rich soils – dated from the late Oligocene. Since the fast-growth strategy is strongly conditioned by resource availability (31), this suggests that the birth of these clades was only made possible by the emergence of highly productive environments in Amazonia. Paleo-geological and ecological evidences strongly support the replacement of a low productivity Craton-derived landscape by a high productivity fluvial system and an incipient wetlands landscape in Western Amazonia in late Oligocene-Early Miocene (~23 Mya), driven by an intense phase of the Andean uplift (5, 13). Occasional marine incursions observed in this period (32) may have also contributed significant levels of nutrients, especially calcium, to Amazonian ecosystems. Still today, it is possible to detect the resource-soil gradients related to sediments from the early Miocene (Pebas formation) structuring ecological and biogeographical patterns in Western forests, including for the herbaceous groups analysed here (18). The gap-disturbance regime tends also to be higher in rich soils, due to short generation times and high mortality rates of trees (20), allowing higher penetration of light to the understory, which is another important resource structuring the functional fast-slow strategy trade-off (31). Our results suggest therefore that the emergence of high productivity and dynamic habitats were essential to the evolution the evolution of a fast growth strategy and consequently enhancement of the diversification rates.

The imprint of the long-term topographic transformation along Amazonian history on functional and evolutionary patterns across Marantaceae clades is less evident than that by geochemical transformation, because most clades showed no preference along topographic gradients. However, association exclusively with uplands were found in clade with high LMA and associations exclusively with bottomlands were observed for the Breviscapus clade, that with the highest diversification rates and LMA. The establishment of the modern Amazonian drainage system started only in the late Miocene (~9 Mya) and is relatively more recent than geochemical transformation that took place in Amazonia in the early-middle Miocene (13, 33), which may favour recent speciation events driven by adaptive radiation along topographic gradients. In central Amazonia, for instance, it is common to find member of sisters species pair associated to valleys and the other to plateaus (21) suggesting that speciation across topographical gradients may be relatively recent and may have occurred several times within clades. This evolutionary pattern would lead to the absence of a general topographical association at any higher than species level. Further studies can elucidate this question mapping the topographic affinities within clades through a similar approach as that used by Finne *et al*. (9). Contrasting he absence of topographical association of most clades, all of them were associated with wet climatic conditions, revealing that dry conditions are somewhat detrimental to diversification of all clades analysed here. The inexistence of a deep evolutionary divergence associated to large-scale water conditions however, indicate that past climate oscillations have not driven the emergence of new clades and their differences in intrinsic diversification rates, functional strategies and diversity patterns.

Distinct diversification dynamics between acquisitive/conservative clades and the role of past geological and climatic events.

The diversification dynamics signature of two groups with distinct functional strategies was quite distinct in terms of response to past geological and climatic events. The diversification rates of conservative clades were in an almost pure steady-state along their 44 Mya of existence, with a gradual reduction in extinction rates and a short increase in the speciation rates at the early Miocene (~20 Mya). The intense transformation of the landscape driven by intensification of Andean uplift seem to have slightly impacted the speciation rates of conservative clades, since some clades in this group have intermediate values of LMA that must have been favoured by the increment of resources. While the dynamics of conservative clades can be described as constant and stable process, we can not say the same for of acquisitive clades.

According to our molecular tree and niche analysis, the evolution of Breviscapus, Comosae and Microcephalmum clades begins in late Oligocene (~23 Mya), probably due to the emergence of high productive habitats in Western Amazonia that favoured the rapid evolution of fast-growth functional strategy. Paleo-environmental reconstruction suggests that fluvial environments, some marine incursions and an incipient wetland established in Western Amazonia at early Miocene (~ 24 to 16 Mya) and evolved into a huge and complex mega-wetland, formed by lacustrine, swamp and riverine systems in the middle-early Miocene (~16 a 11 Mya, Pebas phase), covering an area represented today by the Western lowlands (5, 13). Although acquisitive clades have been already established in the Pebas phase, the steady-state of speciation rates during this period and the slight increase in extinction rates lead to a reduction in net diversification to levels indistinguishable from diversification rates of conservative clades. This scenario suggests that diversification of western lineages associated with non-inundated forest was not favoured by the expansion of mega-wetlands, probably due to reduction of suitable area. Evidences of the inhibiting effect on speciation process caused by Pebas system has been also found in evolutionary history of Neotropical palms (34). The boosted diversification of acquisitive clades was only observed at the end of Pebas phase when a new stage of intensity of Andean uplift drove the gradual replacement of mega-wetlands by fluvio-tidal wetlands and the early Amazon river system (~11 to 7 Mya, Acre phase) (13). The expansion of terrestrial habitats over the wetlands, supported by predominance of continental vertebrates in fossil records (35), may have favoured dispersal events across the landscape and speciation events could have emerged by the combination of long dispersal events, geographical isolation of populations and process of adaptive divergence (36) with species radiating across a highly heterogeneous topo-edaphic landscape.

The net diversification rates continued speeding up and reach the peak at the late Pliocene and abruptly declined at early Pleistocene (~3 Mya). The Plio-Pleistocene period mark a transition from warm to cold and dry climatic conditions (1, 37). Evidences from fossil pollen records also indicate a general decline in plant diversification along Pleistocene by the increase of extinction rates (38), but in the case of Marantaceae clades these climate changes may have affected diversification by a slow down effect on speciation rates. The negative effect of cooling-dry conditions observed only in acquisitive clades and not in conservative clades suggests that fast-growth tropical plants may depend on relatively high temperatures and humidity, possibly to sustain their higher metabolism rates. Due to biophysical constraints, a really fast-growth strategy is only possible if all vital resources (*e.g,* nutrients, carbon, water and radiations) are provided in enough quantities to keep elevated the metabolic rates (31). By contrast, current evidences have shown that slow-growth species are less sensitive to drought (39), supporting the distinct response between acquisitive and conservative clades to paleo-climate changes.

The dependence of acquisitive clades on warm-wet conditions also suggests that the rapid raising temperature at late Oligocene could have also favoured the emergence of the acquisitive clades. The increase of palynological diversity in Neotropics associated with warming periods from the Paleocene to the early Miocene (65 to 20 Mya) (40) have been attributed to a speed up in biochemical kinetics, rate of molecular evolution and speciation rates driven by the increase of temperatures (41). However, diversification rates in Marantaceae did not increase in the Miocene Climatic Optimum warming period (17 to 15 Mya), and a strong increase in diversification rates happened at middle – late Miocene (~10 Mya) when the temperature conditions resembled those before the late Oligocene warming period. The paleo-climate cycles only, therefore, were unlike to have favoured the birth of acquisitive clades and seeped up the diversification rates in Marantaceae, leading us to conclude that Andean uplift and the landscape transformation have played a decisive role in this process.

CONCLUDING REMARKS

By combining plant growth trade-off and long-term diversification dynamics we provided clear evidences that landscape transformation boosted by Andean uplift was the paramount driver of the diversification history of Marantaceae clades in Amazonian lowland rainforest. This process may have involved two main mechanisms, each prevailing at a different stage. The intense Andean uplift (~23 Mya) drove the emergence of highly dynamic and productive environments and promoted evolution of lineages with fast demographic strategies to live in such abundant resource conditions via trait-niche divergence. Fast growth strategies associated to short generation times, led to fast molecular evolution, speeding up the speciation rates (42). This process, however, stayed latent for the next 10 Mya due to low availability of vast continental areas. As the mega-wetlands system gave place to terrestrial environments due to the second phase of intense mountain raising (~10 Mya), new extensive areas of terrestrial habitats became available. Over this vast and heterogeneous topo-edaphic landscape diversification speeded up by the combination of allopatric-dispersal and adaptive divergence (9, 10, 36). Diversification slowed down in the cooling phase during Pleistocene, but until there, most of the species known today have already established. The higher the speciation rates, the higher was species' accumulation, resulting in high number of extant species. Since more than 50% of species in the Marantaceae-Calathea clade are found in the three fast-growth clades, it is reasonable to assume that such tumultuous past events caused by Andean uplift in Amazonia significantly contributed to current diversity of its lowlands rain forest. We agree with previous studies that is not necessary to invoke the allopatric vicariance-geological model to explain evolutionary history and diversity patterns in Amazonia (10, 11). However, a pure dispersal model, disconnected from past geological events, ecological forces and focused in recent speciation events do not seem to be the best alternative (10, 11). We propose that coupling functional trait-niche framework to the long-term diversification dynamics must provide plentiful insights into the complex evolutionary history of tropical forests and potential mechanisms underlying the origin and evolution of its spectacular biological diversity. Geological and ecological forces shaped the evolutionary history and diversity patterns in Amazonian forest.

REFERENCES

1. Bush MB (1994) Amazonian speciation: a necessarily complex model. *J Biogeogr* 21(1):5–17.

2. Rull V (2011) Neotropical biodiversity: Timing and potential drivers. *Trends Ecol Evol* 26(10):508–513.

3. Hughes CE, Pennington RT, Antonelli A (2013) Neotropical Plant Evolution: Assembling the Big Picture. *Bot J Linn Soc* 171(1):1–18.

4. Haffer J (1969) Speciation in Amazonian Forest Birds. *Science (80-)* 165(3889):131–137.

5. Hoorn C, et al. (2010) Amazonia through time : Andean uplifit, climate change, landscape evolution and biodiversity. *Science (80-)* 927:927–931.

6. Antonelli A, et al. (2009) Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. *Amazonia: Landscape and Species Evolution. A Look into the Past*, pp 386–404.

7. Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centers, refugia and botanical collection density in Brazilian Amazonia. *Nature* 345:714–716.

8. Ribas CC, et al. (2012) A palaeobiogeographic model for biotic diversification within

Amazonia over the past three million years. (July 2011):681–689.

9. Fine PVA, Daly DC, Muñoz G V., Mesones I, Cameron KM (2005) The contribution of edaphic heterogeneity to the evolution and diversity of burseraceae trees in the Western Amazon. *Evolution (N Y)* 59(7):1464–1478.

10. Smith BT, et al. (2014) The drivers of tropical speciation. *Nature* 515:406–409.

11. Dexter KG, et al. (2017) Dispersal assembly of rain forest tree communities across the Amazon basin. *Proc Natl Acad Sci* I(1). doi:10.1073/pnas.1613655114.

12. Wanderley-Filho JR, Eiras JF, Da Cruz Cunha PR, Van Der Ven PH (2010) The Paleozoic Solimões and Amazonas Basins and the Acre Foreland Basin of Brazil. *Amazonia: Landscape and Species Evolution: A Look into the Past* (Wiley-Blackwell, Oxford), pp 29–37.

13. Hoorn C, Wesselingh FP, Hovikoski J, Guerrero J (2010) The development of the Amazonian mega-wetland (Miocene; Brazil, Colombia, Peru, Bolivia). *Amazonia, Landscape and Species Evolution: A Look into the Past*, pp 123–142.

14. Valencia R, et al. (2004) Tree species distributions and local habitat variation in the Amazon : large forest plot in eastern Ecuador. 214–229.

15. Moulatlet GM, Costa FRC, Rennó CD, Emilio T, Schietti J (2014) Local hydrological conditions explain floristic composition in lowland Amazonian forests. *Biotropica* 46(4):395–403.

16. Tuomisto H, Zuquim G, Cárdenas G (2014) Species richness and diversity along edaphic and climatic gradients in Amazonia. *Ecography (Cop)* (May):1–13.

17. Higgins M, et al. (2011) Geological control of floristic composition in Amazonian forests. *J Biogeogr* 38(11):2136–2149.

18. Tuomisto H, et al. (2016) A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *J Biogeogr* 43(12):2400–2411.

19. Fyllas NM, et al. (2009) Basin-wide variations in foliar properties of Amazonian forest : phylogeny , soils and climate. *Biogeosciences* 6:2677–2708.

20. Quesada CA, et al. (2012) Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences* 9(6):2203–2246.

21. Cosme LHM, Schietti J, Costa FRC, Oliveira RS (2017) The importance of hydraulic architecture to the distribution patterns of trees in a central Amazonian forest. *New Phytol*. doi:10.1111/nph.14508.

22. Baker TR, et al. (2014) Fast demographic traits promote high diversification rates of Amazonian trees. *Ecol Lett* 17(5):527–536.

23. Wright IJ, et al. (2004) The worldwide leaf economics spectrum. *Nature* 428:821–827.

24. Poorter L, et al. (2008) Are functional traits good predictors of demographic rates? Evidence from five neotropical forests. *Ecology* 89(7):1908–1920.

25. Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (Princiton University Press, New Jersey).

26. Antonelli A, et al. (2015) An engine for global plant diversity: highest evolutionary turnover and emigration in the American tropics. *Front Genet* 6(April):130.

27. Couvreur TLP, Forest F, Baker WJ (2011) Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biol* 9(1):44.

28. Wiens JJ (2011) The causes of species richness patterns across space, time, and clades and the role of "ecological limits." *Q Rev Biol* 86(2):75–96.

29. Richardson J, Pennington RT, Pennington TD, Hollingsworth PM (2001) Rapid diversification of a species-rich genus of Neotropical rain forest trees. *Science (80-)* 293:2242–2245.

30. Russo SE, Brown P, Tan S, Davies SJ (2008) Interspecific demographic trade-offs and soil-related habitat associations of tree species along resource gradients. *J Ecol* 96(1):192–203.

31. Reich PB (2014) The world-wide "fast-slow" plant economics spectrum: A traits manifesto. *J Ecol* 102(2):275–301.

32. Marine influecne in Amazonia: evidence from geological record (2010). *Amazonia: Landscape and Species Evolution. A Look into the Past* (Wiley-Blackwell, Oxford), p 447.

33. Hoorn C, et al. (2017) The Amazon at sea: Onset and stages of the Amazon River from a marine record with reference to vegetation turnover in the Plio-Pleistocene. *Glob Planet Change*. doi:10.1016/j.gloplacha.2017.02.005.

34. Roncal J, Kahn F, Millan B, Couvreur TLP, Pintaud JC (2013) Cenozoic colonization and diversification patterns of tropical American palms: Evidence from Astrocaryum (Arecaceae). *Bot J Linn Soc* 171(1):120–139.

35. Latrubesse EM, da Silva SAF, Cozzuol M, Absy ML (2007) Late Miocene continental sedimentation in southwestern Amazonia and its regional significance: biotic and geological evidence. *J South Am Earth Sci* 23(1):61–80.

36. Kawecki TJ (2008) Adaptation to marginal habitats. *Annu Rev Ecol Evol Syst* 39(1):321–342.

37. Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Source Sci New Ser* 292(5517):686–693.

38. Van Der Hammen T, Hooghiemstra H (2000) Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quat Sci Rev* 19(8):725–742.

39. Ouédraogo DY, Mortier F, Gourlet-Fleury S, Freycon V, Picard N (2013) Slow-growing species cope best with drought: Evidence from long-term measurements in a tropical semi-deciduous moist forest of Central Africa. *J Ecol* 101(6):1459–1470.

40. Jaramillo C, Rueda MJ, Mora GG (2006) Cenozoic plant diversity in the Neotropics. *Science (80-)* 311(March):1893–1896.

41. Jaramillo C, Cárdenas A (2013) Global warming and neotropical rainforests: a historical perspective. *Annu Rev Earth Planet Sci* 41(1):741–766.

42. Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science (80-)* 322(5898):86–89.

43. Breheny P, Burchett W (2013) Visualization of regression models using visreg. *http://myweb.uiowa.edu/pbreheny/publications/visreg.pdf*. Available at: http://cran.r-project.org/package=visreg.

44. Prince LM, Kress WJ (2006) Phylogenetic Relationships and Classification in Marantaceae: Insights from Plastid DNA Sequence Data. *Taxon* 55(2):281.

45. Borchsenius F, Suárez LSS, Prince LM (2012) Molecular phylogeny and redefined generic limits of Calathea (Marantaceae). *Syst Bot* 37(3):620–635.

46. Saka MN (2016) Revisão taxonômica e análise filogenética das espécies extra-amazônicas de Goeppertia Ness, clado Breviscapus (Marantacaea). Dissertation (Universidade Estadual Paulista Júlio de Mesquita Filho - Rio Claro).

47. Costa FRC, Espinelli FP, Figueiredo FOG (2011) *Guide to the Zingiberales of PPBio sites in Brazilian Western Amazonia*. (Áattema, Manaus).

48. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.

49. Janssens SB, et al. (2016) Evolutionary dynamics and biogeography of Musaceae reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *New Phytol* 210(4):1453–1465.

50. Costa FRC, Magnusson WE (2010) The need for large-scale, integrated studies of biodiversity – the experience of the Program for Biodiversity Research in Brazilian Amazonia. *Nat Conserv* 8:3–12.

51. Figueiredo FOG, Costa FRC, Nelson BW, Pimentel TP (2014) Validating forest types based on geological and land-form features in central Amazonia. *J Veg Sci* 25(1):198–212.

52. Quesada CA, et al. (2010) Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* 7(5):1515–1541.

53. Zuquim G, et al. (2014) Predicting environmental gradients with fern species composition in Brazilian Amazonia. *J Veg Sci* 25(5):1195–1207.

54. Rennó CD, et al. (2008) HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sens Environ* 112(9):3469–3481.

55. Schietti J, et al. (2013) Vertical distance from drainage drives floristic composition changes in an Amazonian rainforest. *Plant Ecol Divers* 7(1–2):1–13.

56. Pinheiro TF, Rennó CD, Escada M isabel S (2015) HAND terrain descriptor for mapping regional scale eco-hydrological units of Amazon terra firme. *Rev Bras Cartogr* 67:691–700.

57. Bock M, Köthe R (2008) Predicting the depth of hydromorphic soil characteristics influenced by groud water. *Hamburg Beiträge zur Phys Geogr und Landschaftsökologie* 19:13–22.

58. Kummerow C, Barnes W, Kozu T, Shiue J, Simpson J (1998) The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J Atmos Ocean Technol* 15(3):809–817.

59. Morlon H, et al. (2016) RPANDA: An R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol Evol* 7(5):589–597.

60. Rigby RA, Stasinopoulos DM (2007) Generalized additive models for location, scale and shape (GAMLSS) in R. *J R Stat Soc Ser C Appl Stat* 54(3):507–554.

61. Höhna S, May MR, Moore BR (2015) TESS: An R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. *Bioinformatics* 32(5):789–791.

FIGURES

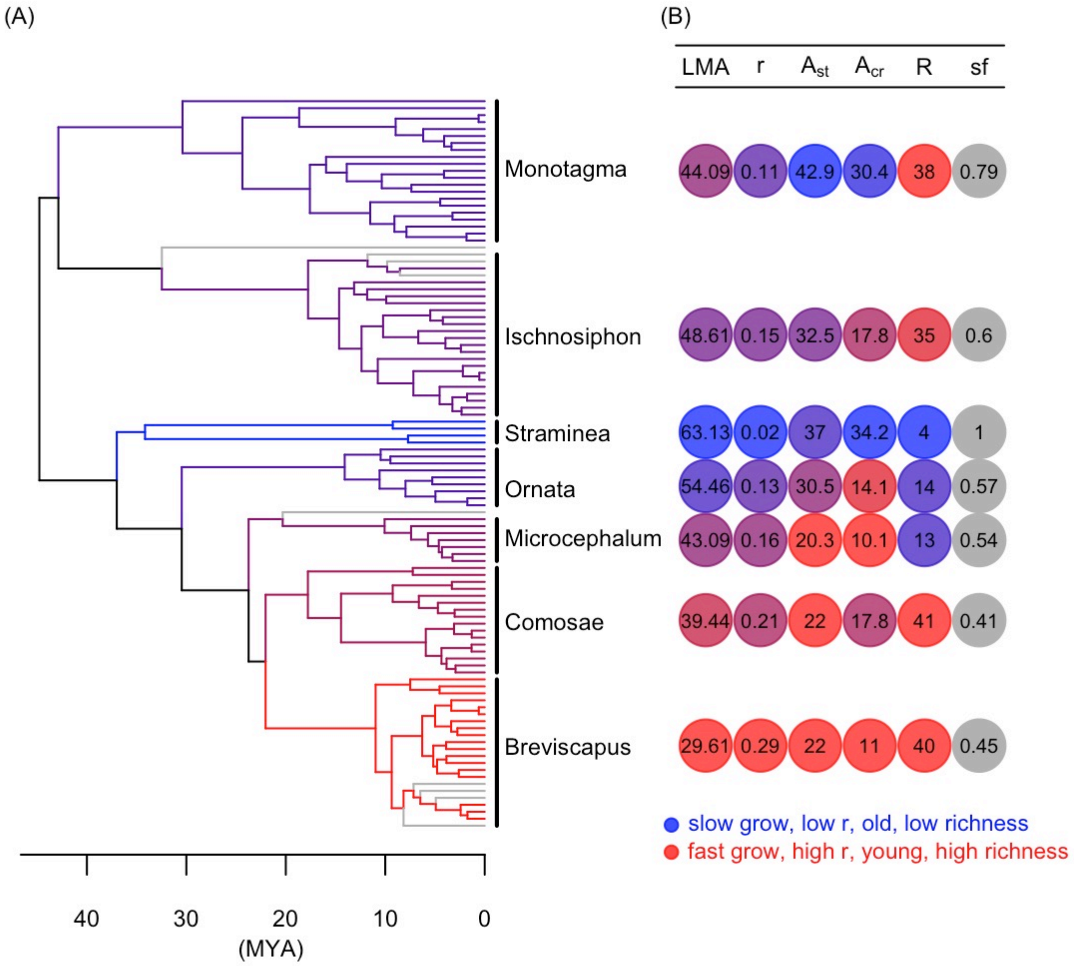


Figure 1. Time-calibrated phylogeny based on the nuclear ITS marker (A) and intrinsic clade factors across the seven Amazonian lineages of Marantaceae (B). Clades in (A) are coloured according with their diversifications rates (r) exhibited in B and grey represent non Amazonian lineages, with some of them collapsed in a unique node (see the complete phylogeny in Fig. S2). In (B) are the intrinsic clade LMA (mean values in gm^{-2}), diversification rates (*r,* in events per Mya), stem ages (A_{st}, in Mya), crown ages (A_{cr}, in Mya), clade richness (R) and sampling fraction (sf) that means the proportion of extant taxa represented in the inferred phylogeny. Diversification rates were estimated fitting a pure birth process, which was the best model compared with models assuming birth-death process (see details in Methods and Results).

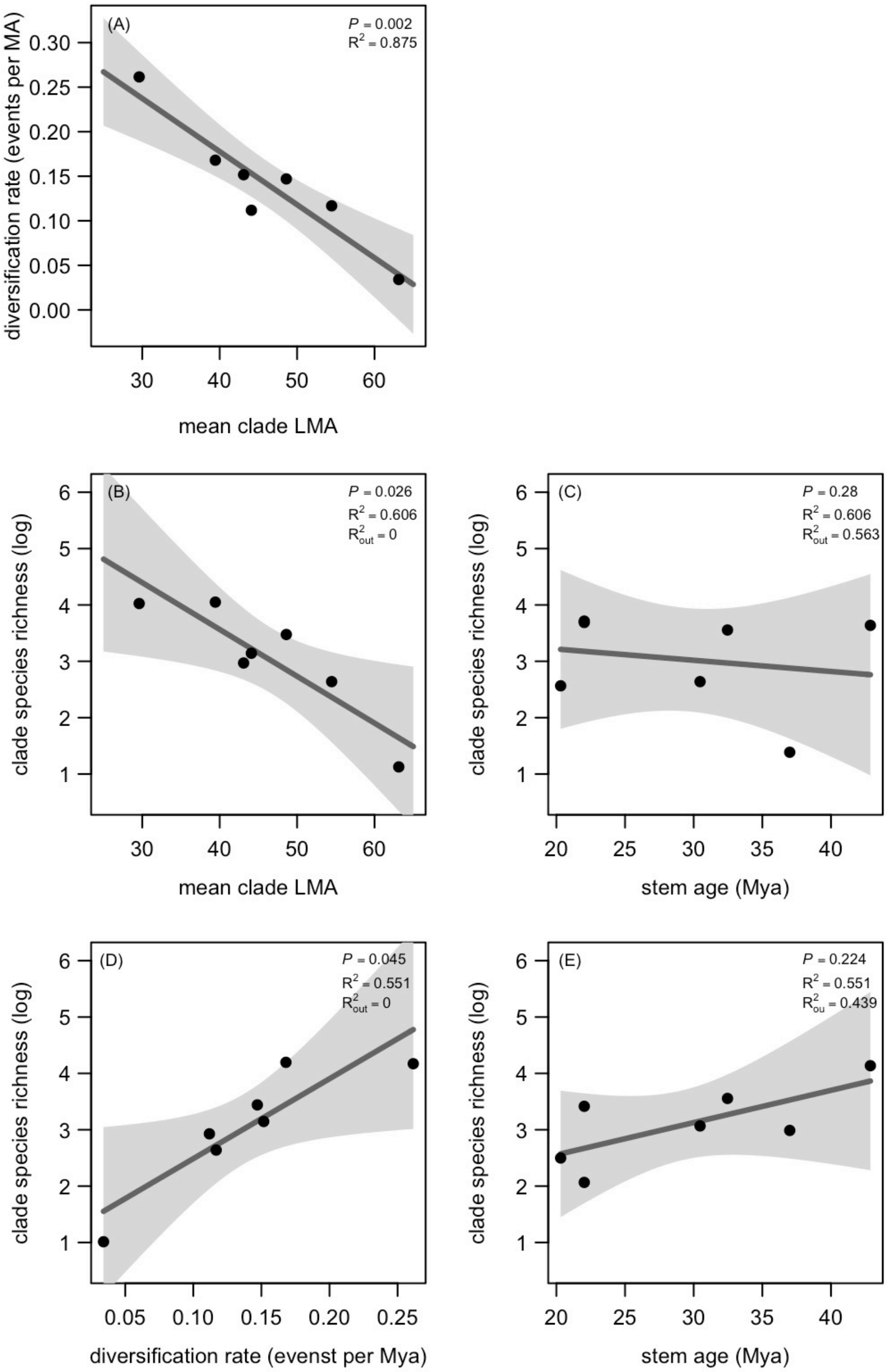


Figure 2: Relationships between diversification rates, LMA, clade species richness and clade age. In A is the simple relation between diversification rates and LMA. In B – C are the partial regression of the model including LMA and stem age as predictor variables, and in D – E denote the partial regression of the model including diversification rates and LMA as predictors. The fitted line in B – E was obtained using conditional plots implemented in *visreg* R-package (43). The *P* values represent the statistical significance of relations, R^2 the good-of-fitness of models including all variables and R^2_{out} de good-of-fitness of models without that variable.

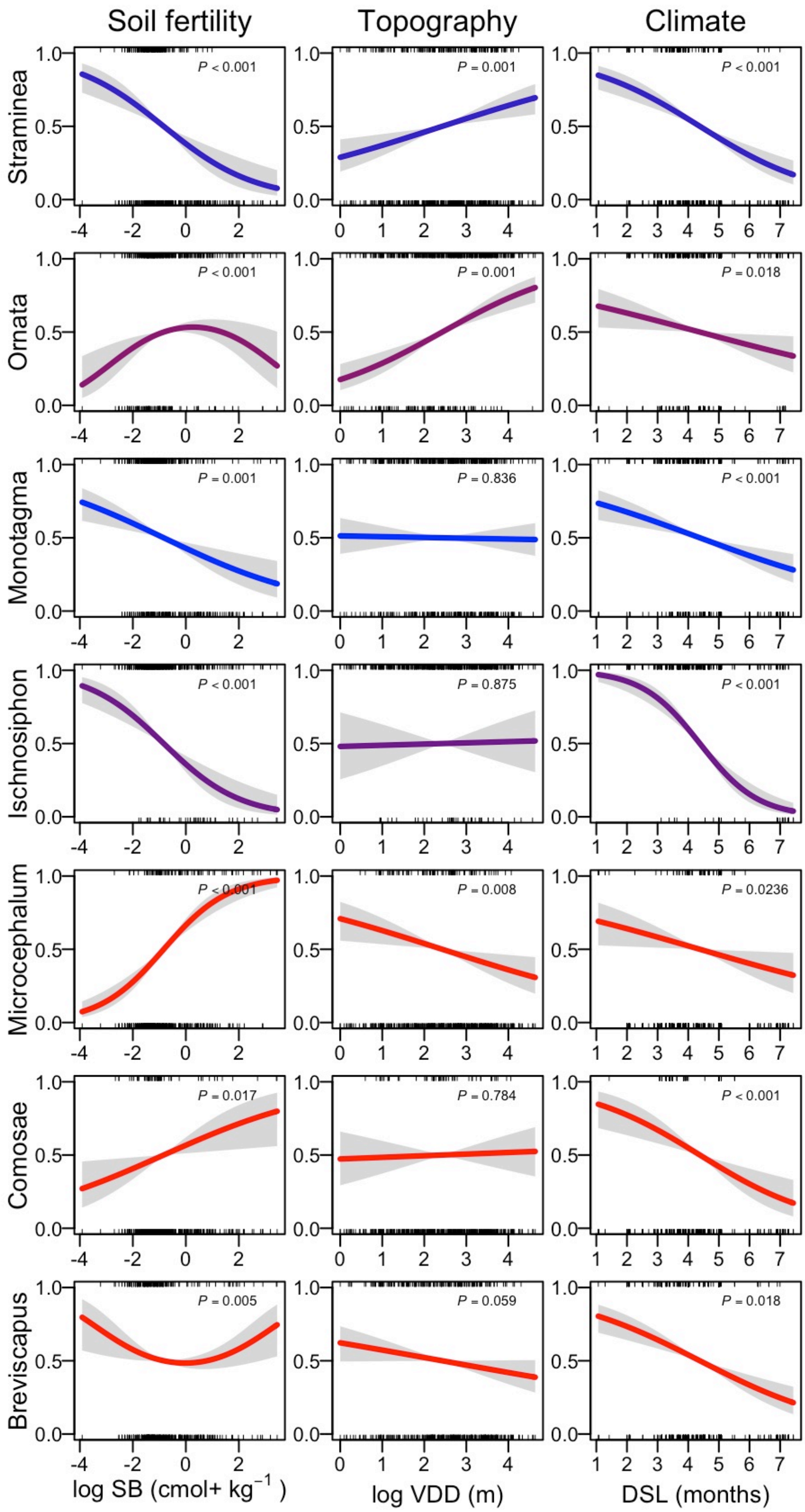

Soil fertility
Topography
Climate
Straminea
Ornata
Monotagma
Ischnosiphon
Microcephalum
Comosae
Breviscapus
1.0
0.5
0.0
-4
-2
0
2
0
1
2
3
4
1
2
3
4
5
6
7
P < 0.001
P = 0.001
P < 0.001
P < 0.001
P = 0.001
P = 0.018
P = 0.001
P = 0.836
P < 0.001
P < 0.001
P = 0.875
P < 0.001
P < 0.001
P = 0.008
P = 0.0236
P = 0.017
P = 0.784
P < 0.001
P = 0.005
P = 0.059
P = 0.018
log SB (cmol+ kg^{-1})
log VDD (m)
DSL (months)

Figure 3: Habitat association of the seven Marantaceae clades. Vertical axis denotes the occurrence probabilities scaled to 0 – 1 using contrasting plot implemented in *visreg* R-package (43). Horizontal axis denotes the logarithm of sum of bases Ca, Mg plus K (SB), logarithm of vertical distance from drainage (VDD) and the dry season length (DSL). From top to bottom panels, clades are order based on their LMA values and colours are based on their stem age, which red represent the youngest blue the oldest clade. The statistical significance of clade-habitat association is given by the *P* values extracted from a GLM model using all the tree variables.

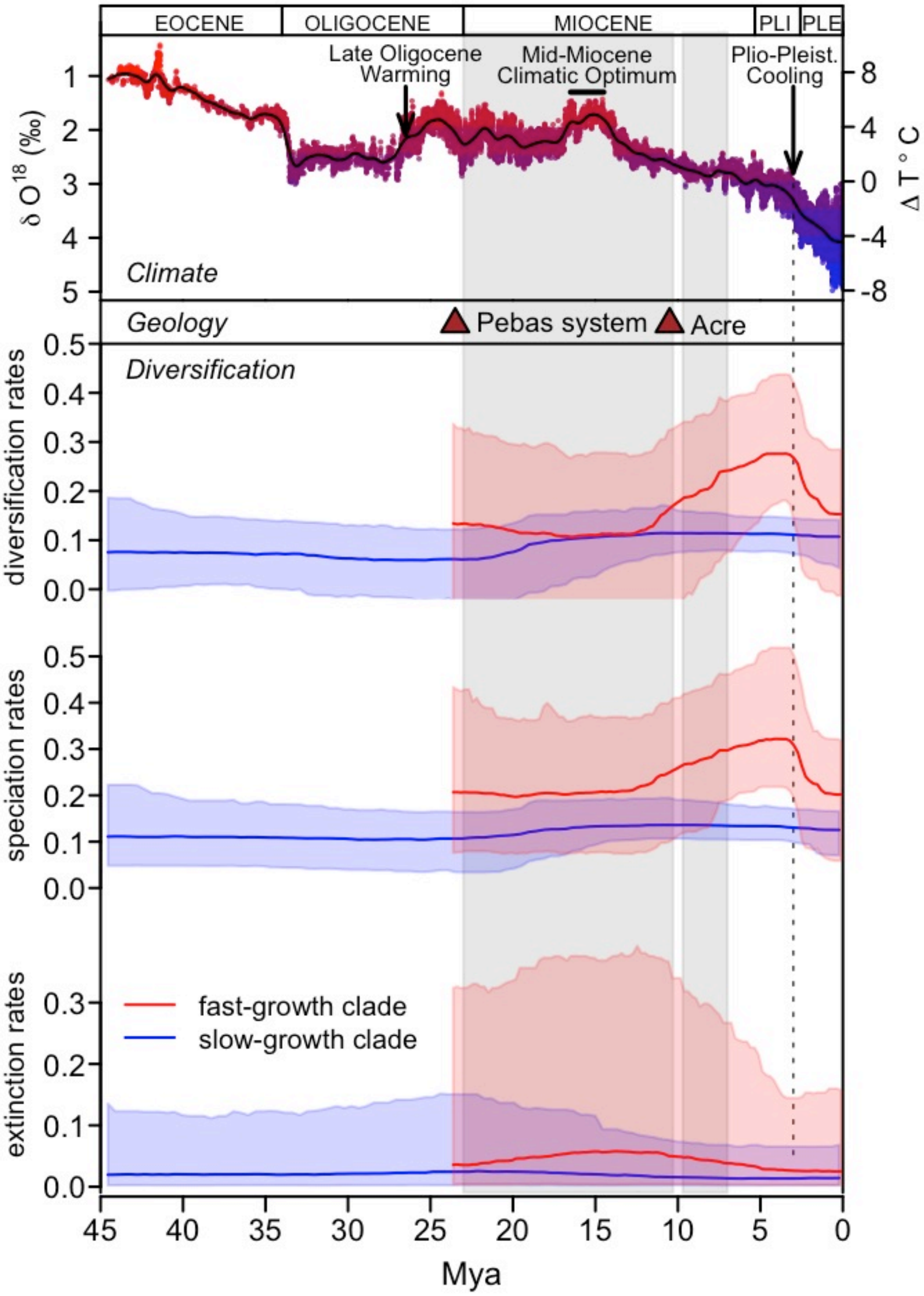


Figure 4. Diversification dynamics of the fast-growth and slow-growth clades. Lines represent the posterior mean and polygons the 95% credible interval for diversification, speciation and extinction rates. Fast-growth clades include Breviscapus, Comosae and Microcephalum and slow-growth clades include Ischnosiphon, Monotagma, Ornata and Straminea. Geological events (red triangles and grey polygons) were obtained in Hoorn *et al.* (5). Paleo climate spatial trends is based on data from Zachos *et al.* (37), assessed in ftp://ftp.ncdc.noaa.gov/pub/data/paleo/. ΔT °C is the approximate difference from present-day temperature scaled according to concentration of isotopes of oxygen (δ O^{18}) based on benthic foraminifera record (see details in figure 2 of Zachos *et al.* (37)). Paleo climate points are coloured representing the gradient running from warm conditions (red) to cooler conditons (blue).

METHODS

Target taxa - Marantaceae is species rich pantropical Monocot family within the order Zingiberales (gingers). Most species are rhizomatous herbs, varying in habit from 10 cm tall small rosulate to 4 m tall palm-like or scandent. Nearly 550 species are recognized world-wide but most (~450 spp) are found in the Neotropics (44). Here, we focus on the Neotropical Calathea clade, a lineage of four genera (*Calathea, Goeppertia, Ischnosiphon and Monotagma*) with 360 species, occurring in Amazonia, Atlantic Coastal Forest and Central America. *Goeppertia* is the richest genus within Marantaceae (~250 species). It is split into six well-supported clades defined in previous phylogenetics studies (45, 46): Breviscapus, Comosae, Microcephalum, Ornata, Scapifolia and Straminea. Clades with diversity centers outside the lowland rain forest of the Amazon basin (genera Calathea, Scapifolia and a sub-clade of Breviscapus) were removed from our main analyses, but were included to build a time-calibrated phylogenetic tree at the family level (see details in next section). In summary, our main analyses were based on seven Amazonian clades: five clades in *Goeppertia* (Breviscapus, Comosae, Microcephalum, Ornata e Straminea) and the two genera *Ischnosiphon* and *Monotagma*. Images and taxonomic details of some of species included here can be found in (47).

Molecular data - Our molecular data included sequences of 170 Marantaceae taxa overall. For 70 taxa, DNA samples were collected by our team in Amazonia and Atlantic forest sites. The remainder were obtained from GenBank, accessed at https://www.ncbi.nlm.nih.gov/genbank/. To calibrate the tree of species in seven Amazonian clades, we used GenBank data for all available sequences of all Marantaceae clades. DNA was extracted from leaf tissues using the CTAB protocol and the DNeasy Plant minikit

(Qiagen, Valencia, California). Molecular analyses were based on sequences of nuclear ribosomal internal transcribed spacer regions (ITS 1 and 2; 850 bp), the plastid matK gene, and flanking 30 trnK intron, as described in Borchsenious et al. (45). PCR fragments of ITS were amplified with annealing temperatures of 54–58°C and matk, following the protocol in Borchsenious et al. (45) and Saka (46). Sequencing was conducted at the Laboratório de Temático de Biologia Molecular at INPA, Manaus, Brazil and at Macrogen Inc., Seoul, South Korea. We aligned sequences using the MAFT algorithm and then made fine-tune adjustments manually using Geneious software.

Phylogenetic inference and temporal reconstruction – To estimate our time-calibrated topology we performed a Bayesian analysis using BEAST v1.8.3 (48). This was done in two stages. (1) We first estimated a family-wide, time-calibrated topology using 151 *matk* sequences. The root age of our plastid tree was based on diversification times estimated from an Order-level topology calibrated with three well-identified fossils of gingers (49). We set the root age of the matK tree using a normal distribution prior with mean of 60 Mya and 2 standard deviations (sd). The analysis was run using a GTR substitution model (Gamma site heterogeneity model with four categories), an uncorrelated relax clock with lognormal relaxed distribution (ucld) and a Birth-Death model of speciation. Exponential priors for ucld.mean and ucld.sd were set to 0.001 and 0.5 events per million years, respectively. (2) Based on the node ages estimated in the first step, we then estimated the ITS tree containing 150 sequences of taxa from the Calathea clade only, setting the root age to 45 Mya (normal distribution with 5 sd). We applied the same settings used for the matK tree, but with ucld.mean and ucld.sd set to 0.003 and 0.6, respectively. This gave some low effective sample size (ESS) values for some parameters so we ran the analysis again using the simpler HKY substitution model and the Yule process model of speciation. For each topology, the plastid tree and nuclear tree, the tree's parameters were obtained after a run of $2x10^7$ generations of Markov Chain Monte Carlo (MCMC), sampling each 20,000 generations. With these modifications the ESS values exceeded 200 for both trees, which indicate good convergence in the parameter estimations. We burned-in 10% of the 1,000 Bayesian trees and obtained the maximum clade credibility tree (MCC) using TreeAnnotator software. The MCC trees of plastid and nuclear markers are shown in Figures S1 and S2, respectivelly.

Trait-based data - The leaf mass area (LMA in m g^{-1}) was used to summarize the functional strategy of clades and was obtained from 190 individuals of 118 species sampled in our field plot inventories (see plot-based data description below). LMA is the ratio between the dry leaf mass and fresh (*i.e*, completely expanded) leaf area. Here, we adapted the original method to estimate LMA, using small pieces of the lamina of dried material from herbarium collections. For each individual, a maximum of 4 pieces (3 cm^2 in average) was cut in different parts of the lamina (at middle and close to the base and apex), excluding the main vein. We choose carefully for leaf pieces where the tissue had no or very little shrinkage and avoided young leaves. LMA for each species was calculated as the average of values obtained in maximum of six individuals and minimum of one individual, when the species was rare and collected only once.

Plot-based data - Presence-absence data of species in Calathea clade were obtained in 451 plots of 500 m^2 established in non-inundated forest (Fig S3). Data came from two plot design: (1) 409 plots with 250m x 2m set up on the topographical isoclines following the RAPELD protocol (50); (2) 39 plots with 100m x 5 m sub-sampled from 500m x 5m transects oriented to include representative local topographic variation (18). Altitudinal values for each 5m points within the 500m long transect were extracted from digital elevation model to guide the selection of the continuous 5m sub-units with low internal topographical variation as possible in order to make the two sample design comparable. Sample units of RAPELD design are aggregated in spatial modules with minimum geographic distance between plots of 1 km. The details about field sampling can be found in Figueiredo et al. (51) and Tuomisto et al. 2016 (18).

Environmental data – To represent soil fertility, topographic and climatic gradients we used the concentration of exchangeable base cations (Ca, Mg e K) defined as sum of base, the vertical distance from drainage (VDD) and the dry season length, respectively. The sum of base is a good proxy of soil fertility in Amazonia since it is well correlated with Phosphorus concentration (52) and is strongly related to diversity patterns across the basin (17, 18, 51). The protocol of soil sample and the laboratory analysis are described in previous published studies (15, 18, 51, 53)

The topographical descriptor was based on the vertical distance from drainage (VDD), extracted from SRTM digital elevation models (SRTM-DEM). VDD have been successfully used to map local hydrological conditions and to explain floristic patterns in Amazonia (15, 54, 55). Plots with low VDD values indicate that individuals are closer to bottomlands and consequently closer to water table and with better access to water than individuals in plots with high VDD values, located often in hill tops and plateaus, distant from water table. VDD were obtained from three similar algorithms. For 198 plots, located in the interfluve Purus-Madeira and in the Pará state, VDD was derived from the Height Above Nearest Drainage (HAND) algorithm using the 90m SRTM-DEM (15, 54, 56). For 193 plots located at north of Amazonas river in central Amazonia, VDD were obtained from the Elevation Over Channel Network algorithm using the hydrological tools of the SEXTANTE spatial data analysis library (http://www.sextante- gis.com/), coupled with the open-source gvSIG v 1.10 software (http://www.gvsig.org/web/) (see details in Figueiredo *et al.* 2014). For the remained plots, VDD was obtained from 30m SRTM-DEM using the Vertical Distance do Channel Network (57) algorithm, implemented in the QGis software trough SAGA-toolbox.

Climatic variable was estimated by the TRMM satellite (58) available at http://disc.sci.gsfc.nasa.gov. We converted monthly data from 1998 to 2004 of the TRMM product 3B43 V6 at a 0.25° resolution (about 28 km at the equator) into the dry season length variable, defined here as the maximum number of consecutive months with less than 100 mm of precipitation along that period. For each plot the dry season length value were extract after rescale the data to a 0.05° (~ 5km) spatial resolution. We used the *raster* package in R to manipulate and process TRMM variable (R development core team 2015; Hijmans *et al.* 2017).

Intrinsic clade diversification rate and relation to trait functional, clade age and species richness – The diversifications rates for each of the seven Amazonian clades were estimated using maximum likelihood inference, implemented in R RPANDA package (59). For each phylogeny, we fitted six different models. Three pure birth models (no extinction) with and different parameterization for speciation rates for each model: time-constancy, time-linear variation and time-exponential variation. And three bird-death model allowing extinction with constant, combined with same parameterization of speciation rates used before. Diversification rates were defined here as the maximum rate observed thought time. For instance, in the case of models with ascending diversification rates, rates would be defined at present and for descending-rate models, at clade origin (initial diversification rates). For model with constant rate, no time adjustment is needed. Essentially, model parameters are estimated given three information a priori: a time-calibrated phylogeny tree, the total time which process happened (crow age or stem age) and the number of extant species. A fourth optional parameter is the fraction of extant species represent in the phylogeny. This is essential for us since we do not have a complete phylogeny and the sampling effort is unbalanced across clades. We extracted accepted taxa names by genus in the World Checklist of Selected Plant families, accessed at http://apps.kew.org/wcsp/home.do. Based on ancillary literature and our taxonomic knowledge about the group we removed no valid taxa and add other not included in the list. Taxa were also classified based their occurrence in the four major regions in the Neotropics: Amazonia, Atlantic forest, Cerrado biome and Central America). We assumed that taxa not identified at species level included in our phylogeny are one of those in the checklist, and thus the sampling fraction of each clade was defined as the ratio of tips in the phylogeny and the total accepted Amazonian taxa belonging that clade. Diversification rates were estimated using both, crown ages and stem ages. The best model was assessed comparing Akaike information criteria with corrected sample size (AICc).

In order to test if clade diversification were related to clade functional strategy we applied a simple linear model using the mean LMA taken for clades as independent variable. Since for all clades the time-constant rate model was set the best model, no temporal adjustment was need to set a single rate value for each clade. We estimated the mean LMA values for each clade fitting a three parameter *t* family distribution (mean, variance and skewness) over the distribution of species trait values using the *gamlss* R-package (60) in order to reduce the effect of outlier in the mean estimates (Fig. S4). To evaluate whether functional strategy, diversification rates and clade age predict clade species richness we performed multiple linear model. Two models were fitted separately, one using LMA and clade age as independent variables, and other using diversification rates and clade age. This was done since we presume that diversification rates and LMA were correlated and because we do not have statistical degree of freedoms enough to fit complexes model (*e.g.* with three variable). The relative importance of each variable was obtained estimating the reduction in the adjusted R^2 after removing each one variable from the model. Linear models were run using the *stat* R package.

Clade habitat association - We estimated clade habitat affinities fitting a Generalized Linear Model with a binomial family distribution (logistic regression). The presence-absence data for each clade obtained in the 448 plots were regressed against logarithm of sum of base, logarithm of vertical distance from drainage and dry season length. Linear and second order polynomial response curves were allowed and best fitted models were chosen using AIC values. All species recorded in the plot inventories were included, even they were not represented in the phylogeny.

Diversification dynamics and relation to paleo events– To evaluate changes in the trajectory of diversification dynamics between clades with different functional strategies and their association with past geological and climatic events we used a compound Poisson process on Mass-Extinction Times (CoMET) algorithm implemented in *TESS* R-package (61). The advantage of CoMET is that it does not assumes *a priori* a unique and independent branching-process that generated the branching-pattern of our tree, *e.g.* a pure birth or birth-death process with exponential varying rates. Instead, it estimates rates (speciation, extinction and mass extinction) in any point of time using compound Poisson process (CPP) model and reversible-jump MCMC to average over a vast set of possible birth-death process (61). We merged our seven clade into two groups, one consisting of the three clades with low LMA values – acquisitive clades (Breviscapus, Comosae and Microcephalum) and the conservative clades with the four remained lineages, and performed the CoMET analysis separately for each of them. We did not consider the mass extinction events in our model and thus we disabled this option before running the algorithm. The prior for speciation and extinction rates were defined using the automatic empirical hyperprior procedure already implemented in CoMET. To ensure sufficient converge in reversible-jump MCMC we used the auto-stopping rules setting high number of maximum iteration (10^8) and the minimum ESS equal to 500. Thus, the the MCMC stops after this threshold has been reached. The main geological events related to Andean uplift (5) and paleo climate reconstruction (37) were plotted in conjunct with diversification dynamics profiles to identify potential association between past events and shifts in the diversification rates.

SUPPORTING INFORMATION

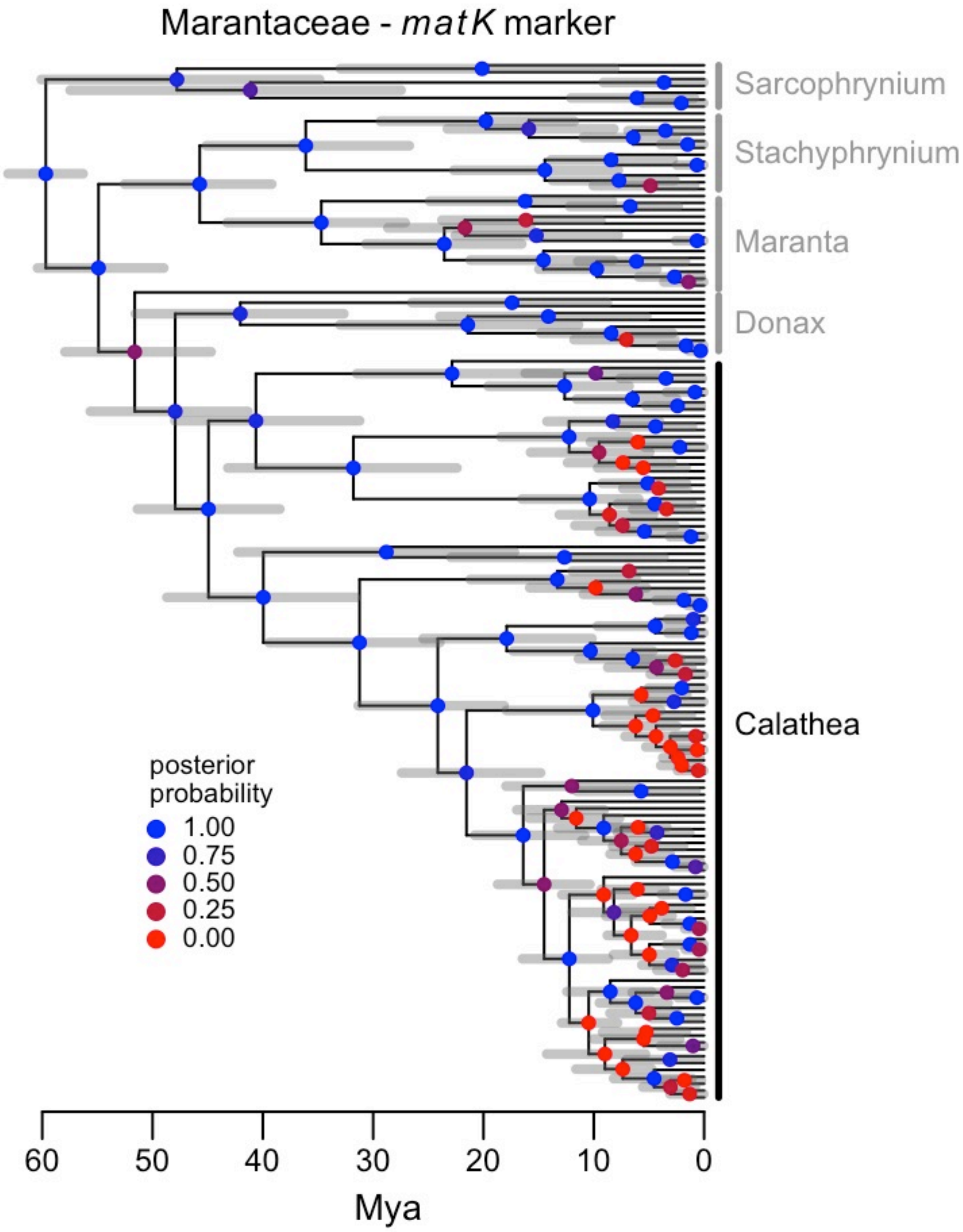


Fig. S1: Bayesian phylogenetic tree for 151 taxa in Marantaceae family based on *matK* sequences. The five major clades are according to Prince & Kress (44) and four clades in grey are those with centre of diversity outside Amazonia. Grey bars denote the 95% confidence interval of divergence times.

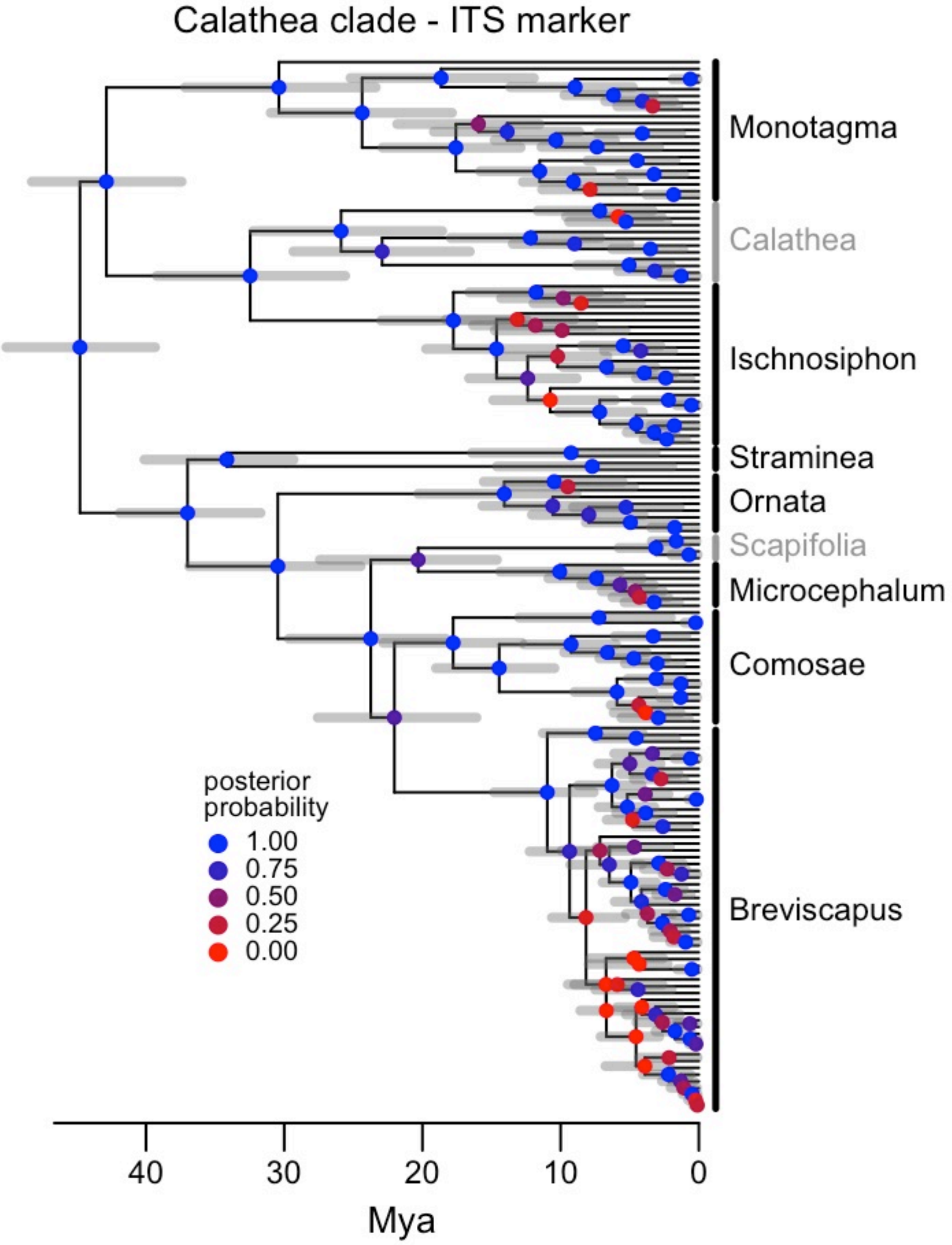


Fig. S2: Bayesian phylogenetic tree for155 taxa in Calathea clade baed on ITS sequences. Clades in grey are those with centre of diversity outside lowlands Amazonian rain forest. Grey bars denote the 95% confidence interval of divergence times.

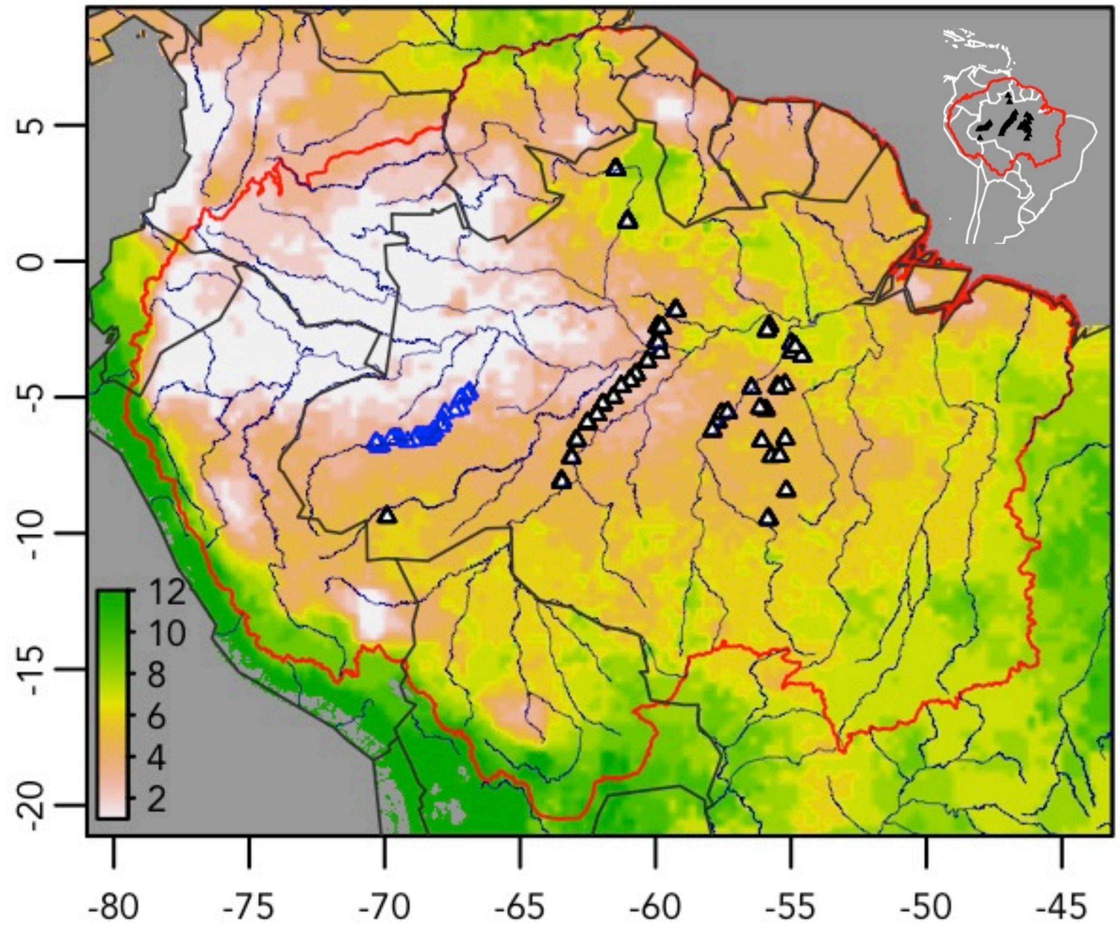


Fig S3. The study area and location of 451 plots where ginger community and environmental data were obtained. Background map represents the dry season length. Legend denotes the maximum number of consecutive months with less than 100 mm of precipitation along the 1998 - 2004 period. Black triangles represent the clusters of RAPELD plots (250 m x 2m) and blue triangles represent plots with size of 100 m x 5m located along the Juruá River. Clusters may have 5 to 57 plots spaced by at least 1 km.

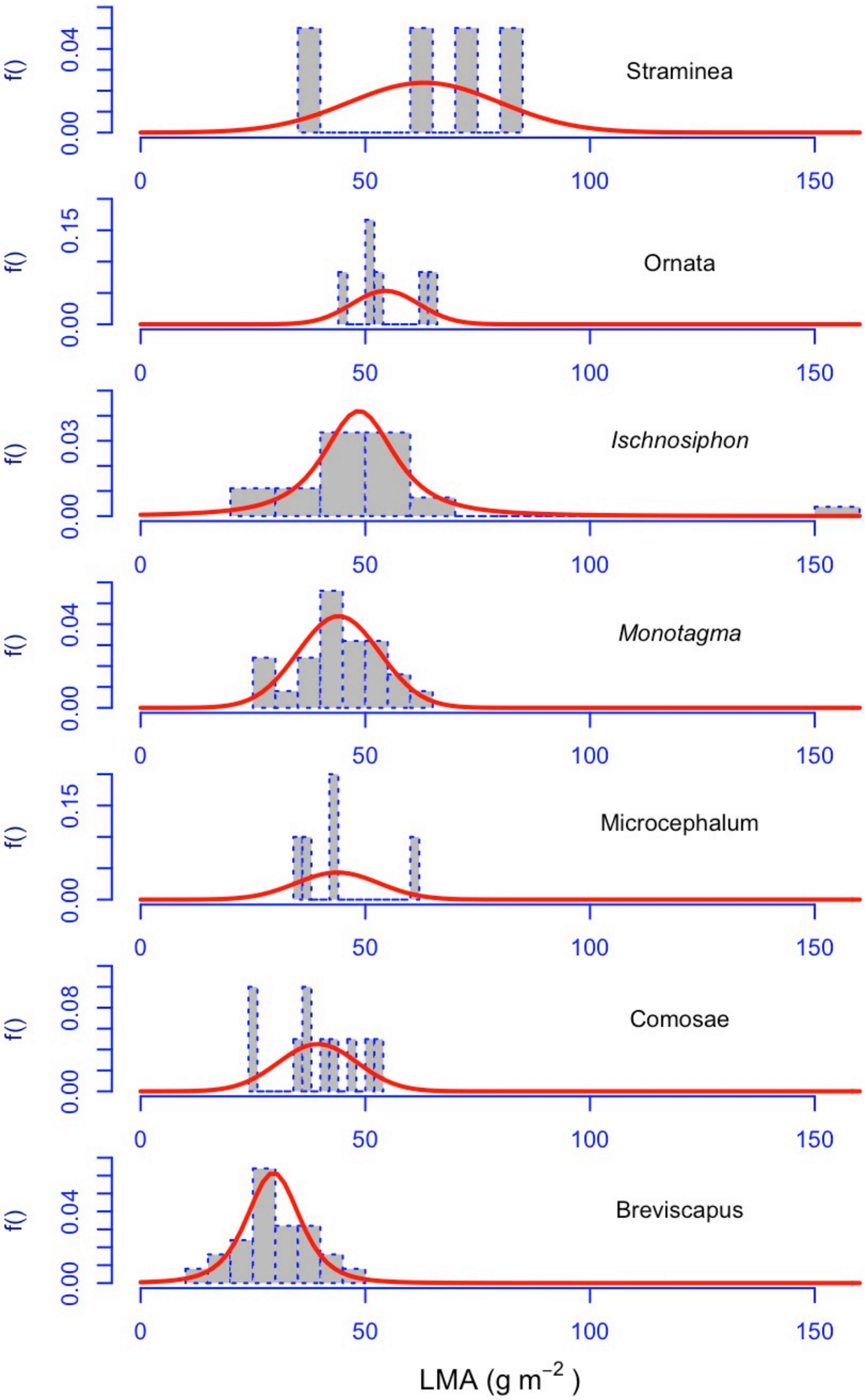


Fig S4. Density distribution of LMA values for the seven Amazoninan clades in Marantaceae. Red lines donotes the fitted distribution of the three parameter *t* family distribution adjusted to the mean LMA values for each species whint that clade.

SINTESE

Demonstramos nesta tese que os filtros biogeoquímicos edáficos representam o principal fator determinante de macro padrões espaciais e evolutivos de plantas nas florestas de terras baixas da Amazônia, tendo o clima um papel coadjuvante nesta história.

No primeiro capítulo mostramos que os limites de distribuição das espécies são controlados tanto por variáveis edáficas quanto por variáveis climáticas. Porém os modelos baseados em variáveis de solo tiveram um desempenho melhor que os modelos puramente climáticos indicando que os filtros edáficos são os principais controladores da distribuição de espécies de plantas na Amazônia. Segundo nossas projeções espaciais, barreiras edáficas estão limitando a ocorrência de espécies na região central, noroeste e nordeste da Amazônia enquanto que barreiras climáticas estão limitando ocorrências na região sudoeste, leste e norte. A região sul da Amazônia apresentou tanto clima como solos inadequados para ocorrência de certas espécies. Mostramos, portanto, que as distribuições das espécies de plantas na Amazônia não estão em perfeito equilíbrio com condições climáticas e que é essencial a inclusão de variáveis edáficas em modelos de projeção de impactos climáticos sobre a diversidade de plantas na Amazônia. A baixa resolução espacial dos dados de solo disponíveis para Amazônia limita o uso destes dados para aplicações de alta resolução, como por exemplo, definição de potenciais rotas de migração num cenário de mudanças climáticas e mapeamento de potenciais áreas de refúgios. A produção de mapas de solo mais acurados para Amazônia é essencial e urgente.

No segundo capítulo mostramos que as comunidades de plantas herbácea de sub-bosque apresentaram convergência funcional ao longo de gradientes edáficos e hidro topográficos. Espécies com traços funcionais característicos de estratégias de rápido crescimento predominaram nos solos ricos em nutrientes e nas áreas próximas dos cursos d'água, mostrando a importância seletiva dos filtros geoquímicos e hidrológicos. No entanto, as relações mais fortes foram encontradas em nutrientes do solo e massa foliar específica. As áreas próximas aos igarapés também apresentaram maior divergência funcional que as áreas altas, tanto dentro como entre comunidades, principalmente para traços funcionais ligados a dispersão e captação de luz. Este resultado sugere a influência de processos estocásticos, como a dinâmica de perturbação e clareiras, estruturando uma parte da diversidade funcional de plantas herbáceas em florestas tropicais. Não constatamos nenhum efeito seletivo do gradiente de seca nas respostas funcionais entre as comunidades. Ao contrário do que esperávamos encontramos divergência funcional significativa nas áreas secas e não nas áreas úmidas. Essa divergência funcional pode estar sendo gerada por desequilíbrios entre taxas de colonização e extinção local causados pela heterogeneidade do habitats e oscilações climáticas decorridas no Pleistoceno. Estes resultados mostram que ainda existem muitas incertezas sobre como as florestas Amazônicas responderão funcionalmente às alterações no regime de chuvas futuras, mas que a compreensão do seu funcionamento passa necessariamente pelo controle de filtros biogeoquímicos edáficos e filtros hidrológicos ligados às propriedades do terreno.

No terceiro e último capítulo demonstramos a existência de uma clara relação entre características funcionais das linhagens, taxas de diversificação, riqueza de espécies e transformações geoquímicas da paisagem em decorrência do soerguimento dos Andes. Primeiro, demonstramos que existe uma relação linear positiva fortíssima entre taxa de diversificação e traços funcionais. Linhagens com traços indicadores de estratégias de rápido crescimento e ciclo de vida curto (estratégias aquisitivas) tiveram taxas de diversificação superiores que os clados com estratégias conservativas, sugerindo um efeito de tempo de geração dos organismos sobre taxas de evolução molecular e taxas de especiação na Amazônia. Os clados com estratégia aquisitiva foram no geral os clados que surgiram em épocas mais recentes e mesmo assim tenderam a apresentar maior riqueza de espécies, rejeitando a hipótese que o tempo de origem determina o número de espécies que se acumulam nas linhagens. Os clados com estratégia aquisitiva também estiveram mais associados com ambientes mais ricos em nutrientes e áreas próximas aos cursos de água, sugerindo que o surgimento destas linhagens esteve associado com o aparecimento de ambientes mais produtivos na Amazônia. De fato, os dados moleculares mostram que essas linhagens surgiram no fim do Oligoceno e início do Mioceno (~20 Ma), época que marca o início da primeira fase de transformações intensas na paisagem Amazônia, onde sistemas menos produtivos deram lugar a sistemas mais produtivos na parte oeste da bacia. As análises de dinâmica mostraram, no entanto que apenas há 10 Ma, houve uma aceleração nas taxas de diversificação dos clados com estratégia aquisitiva. Essa época marca a segunda fase de processos intensos de soerguimento dos Andes, onde grandes complexos de áreas alagáveis deram lugares a paisagens terrestres no Oeste, sugerindo que a disponibilidade de hábitats foi fundamental para disparar as taxas de especiação das linhagens com estratégias aquisitivas. Detectamos também uma queda brusca nas taxas de especiação dos clados aquisitivos ao longo da transição do Plio-Pleistoceno, época que marca a transição de climas globais mais quentes e úmidos para climas mais frios e secos. Estes resultados mostram que transformações bioquímicas da passagem e oscilação climáticas tiveram um efeito contrário na diversidade de Marantaceae na Amazônia: o primeiro, acelerando as taxas de especiação, e o segundo freiando-as. Demonstramos, portanto, que o soerguimento dos Andes e suas decorrentes transformações geoquímicas e físicas impulsionaram a diversificação e contribuíram com a formação da diversidade Amazônica.

Os resultados obtidos nos três capítulos desta tese mostraram que os filtros geoquímicos são o principal estruturador de padrões macroespaciais, funcionais e evolutivos de plantas na Amazônia. Embora as relações entre padrões eco evolutivos e condições climáticas não tenham sido tão previsíveis, não implica, no entanto, que a floresta Amazônia está imune aos efeitos das mudanças climáticas. Os nossos resultados sugerem que o entendimento dos efeitos das mudanças ambientais sobre a Amazônia passa por reconhecer e incluir os filtros biogeoquímicos em quaisquer modelos ecológicos e evolutivos que objetivem compreender os mecanismos geradores e mantenedores de sua biodiversidade.

REFERÊNCIAS

Antonelli, A., Quijada-mascareñas, A., Crawford, A.J., John, M., Velazco, P.M. & Wüster, W. (2009) Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. *Amazonia: landscape and species evolution. A look into the Past*, pp. 386–404.

Araujo, M.B. & Peterson, A.T. (2015) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527–1539.

Baker, T.R., Pennington, R.T., Magallon, S., Gloor, E., Laurance, W.F., Alexiades, M., Alvarez, E., Araujo, A., Arets, E.J.M.M., Aymard, G., de Oliveira, A.A., Amaral, I., Arroyo, L., Bonal, D., Brienen, R.J.W., Chave, J., Dexter, K.G., Di Fiore, A., Eler, E., Feldpausch, T.R., Ferreira, L., Lopez-Gonzalez, G., van der Heijden, G., Higuchi, N., Honorio, E., Huamantupa, I., Killeen, T.J., Laurance, S., Leaño, C., Lewis, S.L., Malhi, Y., Marimon, B.S., Marimon Junior, B.H., Monteagudo Mendoza, A., Neill, D., Peñuela-Mora, M.C., Pitman, N., Prieto, A., Quesada, C.A., Ramírez, F., Ramírez Angulo, H., Rudas, A., Ruschel, A.R., Salomão, R.P., de Andrade, A.S., Silva, J.N.M., Silveira, M., Simon, M.F., Spironello, W., Steege, H. ter, Terborgh, J., Toledo, M., Torres-Lezama, A., Vasquez, R., Vieira, I.C.G., Vilanova, E., Vos, V.A. & Phillips, O.L. (2014) Fast demographic traits promote high diversification rates of Amazonian trees. *Ecology Letters*, **17**, 527–536.

Boisier, J.P., Ciais, P., Ducharne, A. & Guimberteau, M. (2015) Projected strengthening of Amazonian dry season by constrained climate model simulations. *Nature Climate Change*, **5**, 656–660.

Brasil 1978. Projeto RADAMBRASIL - Geologia, geomorfologia, pedologia,vegetação e uso potencial da terra. Depto Nacional de Produção Mineral. **1-34.**

Bush, M.B. (1994) Amazonian speciation: a necessarily complex model. *Journal of Biogeography*, **21**, 5–17.

Clinebell, R.R., Phillips, O.L., Gentry, A.H., Stark, N. & Zuuring, H. (1995) Prediction of neotropical tree and liana species richness from soil and climatic data. *Biodiversity and Conservation*, **4**, 56–90.

Dexter, K.G., Lavin, M., Torke, B., Twyford, A.D., Kursar, T.A., Coley, P.D., Drake, C., Hollands, R. & Pennington, R.T. (2017) Dispersal assembly of rain forest tree communities across the Amazon basin. *Proceedings of the National Academy of Sciences*, **I**.

Espírito-Santo, F.D.B., Keller, M., Braswell, B., Nelson, B.W., Frolking, S. & Vicente, G. (2010) Storm intensity and old-growth forest disturbances in the Amazon region. *Geophysical Research Letters*, **37**.

Esquivel-Muelbert, A., Baker, T.R., Dexter, K.G., Lewis, S.L., ter Steege, H., Lopez-Gonzalez, G., Monteagudo Mendoza, A., Brienen, R., Feldpausch, T.R., Pitman, N., Alonso, A., van der Heijden, G., Peña-Claros, M., Ahuite, M., Alexiaides, M., Álvarez Dávila, E., Murakami, A.A., Arroyo, L., Aulestia, M., Balslev, H., Barroso, J., Boot, R., Cano, A., Chama Moscoso, V., Comiskey, J.A., Cornejo, F., Dallmeier, F., Daly, D.C., Dávila, N., Duivenvoorden, J.F., Duque Montoya, A.J., Erwin, T., Di Fiore, A., Fredericksen, T., Fuentes, A., García-Villacorta, R., Gonzales, T., Guevara Andino, J.E., Honorio Coronado, E.N., Huamantupa-Chuquimaco, I., Killeen, T.J., Malhi, Y., Mendoza, C., Mogollón, H., Jørgensen, P.M., Montero, J.C., Mostacedo, B., Nauray, W., Neill, D., Vargas, P.N., Palacios, S., Palacios Cuenca, W., Pallqui Camacho, N.C., Peacock, J., Phillips, J.F., Pickavance, G., Quesada, C.A., Ramírez-Angulo, H., Restrepo, Z., Reynel Rodriguez, C., Paredes, M.R., Sierra, R., Silveira, M., Stevenson, P., Stropp, J., Terborgh, J., Tirado, M., Toledo, M., Torres-Lezama, A., Umaña, M.N., Urrego, L.E., Vasquez Martinez, R., Gamarra, L.V., Vela, C.I.A., Vilanova Torre, E., Vos, V., von Hildebrand, P., Vriesendorp, C., Wang, O., Young, K.R., Zartman, C.E. & Phillips, O.L. (2016) Seasonal drought limits tree species across the Neotropics. *Ecography*, 1–12.

Fearnside, P.M. (2006) Dams in the Amazon: Belo Monte and Brazil's hydroelectric development of the Xingu River Basin. *Environmental Management*, **38**, 16–27.

Feeley, K.J., Malhi, Y., Zelazowski, P. & Silman, M.R. (2012) The relative importance of deforestation, precipitation change, and temperature sensitivity in determining the future distributions and diversity of Amazonian plant species. *Global Change Biology*, **18**, 2636–2647.

Ferry, B., Morneau, F., Bontemps, J.D., Blanc, L. & Freycon, V. (2010) Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest. *Journal of Ecology*, **98**, 106–116.

Figueiredo, F.O.G., Costa, F.R.C., Nelson, B.W. & Pimentel, T.P. (2014) Validating forest types based on geological and land-form features in central Amazonia. *Journal of Vegetation Science*, **25**, 198–212.

Fine, P.V.A., Daly, D.C., Muñoz, G. V., Mesones, I. & Cameron, K.M. (2005) The contribution of edaphic heterogeneity to the evolution and diversity of burseraceae trees in the Western Amazon. *Evolution*, **59**, 1464–1478.

Fittkau, E.J., Junk, W., Klinge, H. & Sioli, H. (1975) Substrate and vegetation in the Amazon region. *Berichte der Internationalen Symposien der Internationalen Vereinigung für Vegetationskunde Herausgegeben von Reinhold Tüxen.* (ed J. Cramer), pp. 73–90. J. Cramer, Vaduz.

Fortunel, C., Paine, C.E.T., Fine, P.V.A., Kraft, N.J.B. & Baraloto, C. (2014) Environmental factors predict community functional composition in Amazonian forests. , 145–155.

Gentry, A. (1988) Changes in plant community diversity and floristic compostion on environmental and geographical gradients. *Annals of the Missouri Botanical Garden*, **75**, 1–34.

Van Der Hammen, T. & Hooghiemstra, H. (2000) Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Science Reviews*, **19**, 725–742.

Hayes, F.E. & Sewlal, J.A.N. (2004) The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. *Journal of Biogeography*, **31**, 1809–1818.

Higgins, M., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R. & Räsänen, M. (2011) Geological control of floristic composition in Amazonian forests. *Journal of biogeography*, **38**, 2136–2149.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartin, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundenberg, J., Stadler, T., Sarkinen, T. & Antonelli, A. (2010) Amazonia through time : Andean uplifit, climate change, landscape evolution and biodiversity. *Science*, **927**, 927–931.

Hughes, C.E., Pennington, R.T. & Antonelli, A. (2013) Neotropical Plant Evolution: Assembling the Big Picture. *Botanical Journal of the Linnean Society*, **171**, 1–18.

Jaramillo, C. & Cárdenas, A. (2013) Global warming and neotropical rainforests: a historical perspective. *Annual Review of Earth and Planetary Sciences*, **41**, 741–766.

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

Lewis, S.L., Brando, P.M., Phillips, O.L., van der Heijden, G.M.F. & Nepstad, D. (2011) The 2010 Amazon drought. *Science (New York, N.Y.)*, **331**, 554.

Malhi, Y., Roberts, J.T., Betts, R.A., Killeen, T.J., Li, W. & Nobre, C.A. (2008) Climate change, deforestation, and the fate of the Amazon. *Science*, **319**, 169–72.

Muscarella, R. & Uriarte, M. (2016) Do community-weighted mean functional traits reflect optimal strategies? *Proceedings of the Royal Society B*, **283**, 20152434.

Negrón-Juárez, R., Jenkins, H., Raupp, C., Riley, W., Kueppers, L., Magnabosco Marra, D., Ribeiro, G., Monteiro, M., Candido, L., Chambers, J. & Higuchi, N. (2017) Windthrow variability in Central Amazonia. *Atmosphere*, **8**, 28.

Nepstad, D., Mcgrath, D., Alencar, A., Barros, A.C., Carvalho, G., Santilli, M. & Diaz, M.C.V. (2002) Frontier Governance in Amazonia. *Science*, **295**, 629–631.

Nobre, A.D. (2014) *O Futuro Climático Da Amazônia*. São José dos Campos.

Opdam, P. & Wascher, D. (2004) Climate change meets habitat fragmentation: Linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation*, **117**, 285–297.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of speces: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

Poorter, L., Wright, S.J., Paz, H., Ackerly, D.D., Condit, R., Ibarra-Manríquez, G., Harms, K.E., Licona, J.C., Martínez-Ramos, M., Mazer, S.J., Muller-Landau, H.C., Peña-Claros, M., Webb, C.O. & Wright, I.J. (2008) Are functional traits good predictors of demographic rates? Evidence from five neotropical forests. *Ecology*, **89**, 1908–1920.

Quesada, C.A., Phillips, O.L., Schwarz, M., Czimczik, C.I., Baker, T.R., Patiño, S., Fyllas, N.M., Hodnett, M.G., Herrera, R., Almeida, S., Alvarez Dávila, E., Arneth, A., Arroyo, L., Chao, K.J., Dezzeo, N., Erwin, T., Di Fiore, A., Higuchi, N., Honorio Coronado, E., Jimenez, E.M., Killeen, T., Lezama, A.T., Lloyd, G., López-González, G., Luizão, F.J., Malhi, Y., Monteagudo, A., Neill, D.A., Núñez Vargas, P., Paiva, R., Peacock, J., Peñuela, M.C., Peña Cruz, A., Pitman, N., Priante Filho, N., Prieto, A., Ramírez, H., Rudas, A., Salomão, R., Santos, A.J.B., Schmerler, J., Silva, N., Silveira, M., Vásquez, R., Vieira, I., Terborgh, J. & Lloyd, J. (2012) Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences*, **9**, 2203–2246.

Reich, P.B. (2014) The world-wide "fast-slow" plant economics spectrum: A traits manifesto. *Journal of Ecology*, **102**, 275–301.

Ribas, C.C., Aleixo, A., Nogueira, A.C.R., Miyaki, C.Y., Cracraft, J. & Andre, A. (2012) A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. , 681–689.

Smith, S.A. & Donoghue, M.J. (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–89.

Smith, B.T., Mccormack, J.E., Cuervo, A.M., Hickerson, M.J., Aleixo, A., Cadena, C.D., Pérez-Emán, J., Burney, C.W., Xie, X., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Derryberry, E.P., Prejean, J., Fields, S. & Brumfield, R.T. (2014) The drivers of tropical speciation. *Nature*, **515**, 406–409.

Sombroek, W. (2000) Amazon landforms and soils in relation to biological diversity. *Acta Amazonica*, **30**, 81–100.

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

Steege, H.T.E.R., Pitman, N., Sabatier, D., Hout, P.V.A.N.D.E.R. & Daly, D.C. (2003) A spatial model of tree α -diversity and -density for the Amazon. , 1–19.

Stropp, J., Ter Steege, H. & Malhi, Y. (2009) Disentangling regional and local tree diversity in the Amazon. *Ecography*, **32**, 46–54.

Toledo, J.J., Magnusson, W.E., Castilho, C. V. & Nascimento, H.E.M. (2011a) How much variation in tree mortality is predicted by soil and topography in Central Amazonia? *Forest Ecology and Management*, **262**, 331–338.

Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Chuvinã, J., Leaño, C., Licona, J.C., ter Steege, H. & Bongers, F. (2011b) Patterns and determinants of floristic variation across lowland forests of Bolivia. , **43**, 405–413.

Tuomisto, H., Moulatlet, G.M., Balslev, H., Emilio, T., Figueiredo, F.O.G., Pedersen, D. &

Ruokolainen, K. (2016) A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *Journal of Biogeography*, **43**, 2400–2411.

Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, **299**, 241–244.

Wallace, A.R. (1852) On the monkeys of the Amazon. *Proceedings of the Zoological Society of London*, **20**, 107–110.

Wallace, A.R. (1876) *The geographical distribution of animals*. Harper, New York, **1**. 635p.

Zuquim, G., Tuomisto, H., Jones, M.M., Prado, J., Figueiredo, F.O.G., Moulatlet, G.M., Costa, F.R.C., Quesada, C.A. & Emilio, T. (2014) Predicting environmental gradients with fern species composition in Brazilian Amazonia. *Journal of Vegetation Science*, **25**, 1195–1207.