bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title: The Maximum Entropy Formalism of statistical mechanics in a biological application: a quantitative analysis of tropical forest ecology.

Authors: Edwin Pos*,1,2 Luiz de Souza Coelho,3 Diogenes de Andrade Lima Filho,3 Rafael P. Salomão,4,5 Iêda Leão Amaral,3 Francisca Dionízia de Almeida Matos,3 Carolina V. Castilho,6 Oliver L. Phillips,7 Juan Ernesto Guevara,8,9 Marcelo de Jesus Veiga Carim,10 Dairon Cárdenas López,11 William E. Magnusson,12 Florian Wittmann,13,14 Mariana Victória Irume,3 Maria Pires Martins,3 Daniel Sabatier,15 José Renan da Silva Guimarães,10 Jean-François Molino,15 Olaf S. Bánki,2 Maria Teresa Fernandez Piedade,16 Nigel C.A. Pitman,17 Abel Monteagudo Mendoza,18 José Ferreira Ramos,3 Joseph E. Hawes,19 Everton José Almeida,20 Luciane Ferreira Barbosa,20 Larissa Cavalheiro,20 Márcia Cléia Vilela dos Santos,20 Bruno Garcia Luize,21 Evlyn Márcia Moraes de Leão Novo,22 Percy Núñez Vargas,23 Thiago Sanna Freire Silva,24 Eduardo Martins Venticinque,25 Angelo Gilberto Manzatto,26 Neidiane Farias Costa Reis,27 John Terborgh,28,29 Katia Regina Casula,27 Euridice N. Honorio Coronado,30,7 Juan Carlos Montero,31,3 Beatriz S. Marimon,32 Ben Hur Marimon-Junior,32 Ted R. Feldpausch,33,7 Alvaro Duque,34 Chris Baraloto,35 Nicolás Castaño Arboleda,11 Julien Engel,15,35 Pascal Petronelli,36 Charles Eugene Zartman,3 Timothy J. Killeen,37 Rodolfo Vasquez,18 Bonifacio Mostacedo,38 Rafael L. Assis,39 Jochen Schöngart,16 Hernán Castellanos,40 Marcelo Brilhante de Medeiros,41 Marcelo Fragomeni Simon,41 Ana Andrade,42 José Luís Camargo,42 Layon O. Demarchi,16 William F. Laurance,29 Susan G.W. Laurance,29 Emanuelle de Sousa Farias,43,44 Maria Aparecida Lopes,45 José Leonardo Lima Magalhães,46,47 Henrique Eduardo Mendonça Nascimento,3 Helder Lima de Queiroz,48 Gerardo A. Aymard C.,49 Roel Brienen,7 Juan David Cardenas Revilla,3 Flávia R.C. Costa,3 Adriano Quaresma,16 Ima Célia Guimarães Vieira,5 Bruno Barçante Ladvocat Cintra,50 Pablo R. Stevenson,51 Yuri Oliveira Feitosa,52 Joost F. Duivenvoorden,53 Hugo F. Mogollón,54 Leandro Valle Ferreira,5 James A. Comiskey,55,56 Freddie Draper,57,35 José Julio de Toledo,58 Gabriel Damasco,59 Nállarett Dávila,60 Roosevelt García-Villacorta,61,62 Aline Lopes,16,63 Alberto Vicentini,12 Janaína Costa Noronha,64 Flávia Rodrigues Barbosa,64 Rainiellen de Sá Carpanedo,64 Thaise Emilio,65,12 Carolina Levis,66,67 Domingos de Jesus Rodrigues,64 Juliana Schietti,3 Priscila Souza,3 Alfonso Alonso,56 Francisco Dallmeier,56 Vitor H.F. Gomes,68,69 Jon Lloyd,70 David Neill,71 Daniel Praia Portela de Aguiar,16 Alejandro Araujo-Murakami,72 Luzmila Arroyo,72 Fernanda Antunes Carvalho,12,73 Fernanda Coelho de Souza,12,7 Dário Dantas do Amaral,5 Kenneth J. Feeley,74,75 Rogerio Gribel,3 Marcelo Petratti Pansonato,3,76 Jos Barlow,77 Erika Berenguer,78,77 Joice Ferreira,47 Paul V.A. Fine,59 Marcelino Carneiro Guedes,79 Eliana M. Jimenez,80 Juan Carlos Licona,31 Maria Cristina Peñuela Mora,81 Carlos A. Peres,82 Boris Eduardo Villa Zegarra,83 Carlos Cerón,84 Terry W. Henkel,85 Paul Maas,2 Marcos Silveira,86 Juliana Stropp,87 Raquel Thomas-Caesar,88 Tim R. Baker,7 Doug Daly,89 Kyle G. Dexter,90,91 John Ethan Householder,13 Isau Huamantupa-Chuquimaco,23 Toby Pennington,33,91 Marcos Ríos Paredes,92 Alfredo Fuentes,93,94 José Luis Marcelo Pena,95 Miles R. Silman,96 J. Sebastián Tello,94 Jerome Chave,97 Fernando Cornejo Valverde,98 Anthony Di Fiore,99 Renato Richard Hilário,58 Juan Fernando Phillips,100 Gonzalo Rivas-Torres,101,102 Tinde R. van Andel,2,103 Patricio von Hildebrand,104 Edelcilio Marques Barbosa,3 Luiz Carlos de Matos Bonates,3 Hilda Paulette Dávila

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Doza,92 Émile Fonty,105,15 Ricardo Zárate Gómez,106 Therany Gonzales,107 George Pepe Gallardo Gonzales,92 Jean-Louis Guillaumet†,108 Bruce Hoffman,109 André Braga Junqueira,110 Yadvinder Malhi,111 Ires Paula de Andrade Miranda,3 Linder Felipe Mozombite Pinto,92 Adriana Prieto,112 Agustín Rudas,112 Ademir R. Ruschel,47 Natalino Silva,113 César I.A. Vela,114 Vincent Antoine Vos,115 Egleé L. Zent,116 Stanford Zent,116 Bianca Weiss Albuquerque,16 Angela Cano,51,117 Diego F. Correa,51,118 Janaina Barbosa Pedrosa Costa,79 Bernardo Monteiro Flores,119 Milena Holmgren,120 Marcelo Trindade Nascimento,121 Alexandre A. Oliveira,76 Hirma Ramirez-Angulo,122 Maira Rocha,16 Veridiana Vizoni Scudeller,123 Rodrigo Sierra,124 Milton Tirado,124 Maria Natalia Umaña,125 Geertje van der Heijden,126 Emilio Vilanova Torre,122,127 Corine Vriesendorp,17 Ophelia Wang,128 Kenneth R. Young,129 Manuel Augusto Ahuite Reategui,130 Cláudia Baider,131,76 Henrik Balslev,132 Sasha Cárdenas,51 Luisa Fernanda Casas,51 William Farfan-Rios,23,133,94 Cid Ferreira,3 Reynaldo Linares-Palomino,56 Casimiro Mendoza,134,135 Italo Mesones,59 Armando Torres-Lezama,122 Ligia Estela Urrego Giraldo,34 Daniel Villarroel,72 Roderick Zagt,136 Miguel N. Alexiades,137 Karina Garcia-Cabrera,96 Lionel Hernandez,40 William Milliken,65 Walter Palacios Cuenca,138 Susamar Pansini,27 Daniela Pauletto,139 Freddy Ramirez Arevalo,140 Adeilza Felipe Sampaio,27 Elvis H. Valderrama Sandoval,141,140 Luis Valenzuela Gamarra,18 Gerhard Boenisch,142 Jens Kattge,143 Nathan Kraft,144 Aurora Levesley,7 Karina Melgaço,7 Georgia Pickavance,7 Lourens Poorter,67 Hans ter Steege,2,145

* Corresponding author: e.t.pos@uu.nl

† Deceased 01-2018

Affiliations

1Ecology & Biodiversity Group, Utrecht University, Padualaan 8, Utrecht, 3584 CH, The Netherlands

2Naturalis Biodiversity Center, PO Box 9517, Leiden, 2300 RA, The Netherlands

3Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Petrópolis, Manaus, AM, 69067-375, Brazil

4Programa Professor Visitante Nacional Sênior na Amazônia - CAPES, Universidade Federal Rural da Amazônia, Av. Perimetral, s/n, Belém, PA, Brazil

5Coordenação de Botânica, Museu Paraense Emílio Goeldi, Av. Magalhães Barata 376, C.P. 399, Belém, PA, 66040-170, Brazil

6EMBRAPA – Centro de Pesquisa Agroflorestal de Roraima, BR 174, km 8 – Distrito Industrial, Boa Vista, RR, 69301-970, Brazil

7School of Geography, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK

8Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud-BIOMAS, Universidad de las Américas, Campus Queri, Quito, Ecuador

9Keller Science Action Center, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL, 60605-2496, USA

10Departamento de Botânica, Instituto de Pesquisas Científicas e Tecnológicas do Amapá - IEPA, Rodovia JK, Km 10, Campus do IEPA da Fazendinha, Amapá, 68901-025, Brazil

11Herbario Amazónico Colombiano, Instituto SINCHI, Calle 20 No 5-44, Bogotá, DC, Colombia

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12Coordenação de Pesquisas em Ecologia, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Petrópolis, Manaus, AM, 69067-375, Brazil
13Dep. of Wetland Ecology, Institute of Geography and Geoecology, Karlsruhe Institute of Technology - KIT, Josefstr.1, Rastatt, D-76437, Germany
14Biogeochemistry, Max Planck Institute for Chemistry, Hahn-Meitner Weg 1, Mainz, 55128, Germany
15AMAP, IRD, Cirad, CNRS, INRA, Université de Montpellier, Montpellier, F-34398, France
16Coordenação de Dinâmica Ambiental, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Petrópolis, Manaus, AM, 69067-375, Brazil
17Science and Education, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL, 60605-2496, USA
18Jardín Botánico de Missouri, Oxapampa, Pasco, Peru
19Applied Ecology Research Group, School of Life Sciences, Anglia Ruskin University, East Road, Cambridge, CB1 1PT, UK
20ICNHS, Universidade Federal de Mato Grosso, Av. Alexandre Ferronato, 1200, Sinop, MT, 78557-267, Brazil
21Departamento de Ecologia, Universidade Estadual Paulista - UNESP – Instituto de Biociências – IB, Av. 24 A, 1515, Bela Vista, Rio Claro, SP, 13506-900, Brazil
22Divisao de Sensoriamento Remoto – DSR, Instituto Nacional de Pesquisas Espaciais – INPE, Av. dos Astronautas, 1758, Jardim da Granja, São José dos Campos, SP, 12227-010, Brazil
23Herbario Vargas, Universidad Nacional de San Antonio Abad del Cusco, Avenida de la Cultura, Nro 733, Cusco, Cuzco, Peru
24Biological and Environmental Sciences, University of Stirling, Stirling, FK9 4LA, UK
25Centro de Biociências, Departamento de Ecologia, Universidade Federal do Rio Grande do Norte, Av. Senador Salgado Filho, 3000 , Natal, RN, 59072-970, Brazil
26Departamento de Biologia, Universidade Federal de Rondônia, Rodovia BR 364 s/n Km 9,5 - Sentido Acre, Unir, Porto Velho, RO, 76.824-027, Brazil
27Programa de Pós- Graduação em Biodiversidade e Biotecnologia PPG- Bionorte, Universidade Federal de Rondônia, Campus Porto Velho Km 9,5 bairro Rural, Porto Velho, RO, 76.824-027, Brazil
28Department of Biology and Florida Museum of Natural History, University of Florida, Gainesville, FL, 32611, USA
29Centre for Tropical Environmental and Sustainability Science and College of Science and Engineering, James Cook University, Cairns, Queensland, 4870, Australia
30Instituto de Investigaciones de la Amazonía Peruana (IIAP), Av. A. Quiñones km 2,5, Iquitos, Loreto, 784, Peru
31Instituto Boliviano de Investigacion Forestal, Av. 6 de agosto #28, Km. 14, Doble via La Guardia, Casilla 6204, Santa Cruz, Santa Cruz, Bolivia
32Programa de Pós-Graduação em Ecologia e Conservação, Universidade do Estado de Mato Grosso, Nova Xavantina, MT, Brazil
33Geography, College of Life and Environmental Sciences, University of Exeter, Rennes Drive, Exeter, EX4 4RJ, UK
34Departamento de Ciencias Forestales, Universidad Nacional de Colombia, Calle 64 x Cra 65, Medellín, Antioquia, 1027, Colombia

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

35International Center for Tropical Botany (ICTB) Department of Biological Sciences, Florida International University, 11200 SW 8th Street, OE 243, Miami, FL, 33199, USA
36Cirad UMR Ecofog, AgrosParisTech,CNRS,INRA,Univ Guyane, Campus agronomique, Kourou Cedex, 97379, France
37Agteca-Amazonica, Santa Cruz, Bolivia
38Facultad de Ciencias Agrícolas, Universidad Autónoma Gabriel René Moreno, Santa Cruz, Santa Cruz, Bolivia
39Natural History Museum, University of Oslo, Postboks 1172, Oslo, 0318, Norway
40Centro de Investigaciones Ecológicas de Guayana, Universidad Nacional Experimental de Guayana, Calle Chile, urbaniz Chilemex, Puerto Ordaz, Bolivar, Venezuela
41Prédio da Botânica e Ecologia, Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Av. W5 Norte, Brasilia, DF, 70770-917, Brazil
42Projeto Dinâmica Biológica de Fragmentos Florestais, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Petrópolis, Manaus, AM, 69067-375, Brazil
43Laboratório de Ecologia de Doenças Transmissíveis da Amazônia (EDTA), Instituto Leônidas e Maria Deane, Fiocruz, Rua Terezina, 476, Adrianópolis, Manaus, AM, 69060-001, Brazil
44Programa de Pós-graduação em Biodiversidade e Saúde, Instituto Oswaldo Cruz - IOC/FIOCRUZ, Pav. Arthur Neiva – Térreo, Av. Brasil, 4365 – Manguinhos, Rio de Janeiro, RJ, 21040-360, Brazil
45Instituto de Ciências Biológicas, Universidade Federal do Pará, Av. Augusto Corrêa 01, Belém, PA, 66075-110, Brazil
46Programa de Pós-Graduação em Ecologia, Universidade Federal do Pará, Av. Augusto Corrêa 01, Belém, PA, 66075-110, Brazil
47Embrapa Amazônia Oriental, Trav. Dr. Enéas Pinheiro s/nº, Belém, PA, 66095-100, Brazil
48Diretoria Técnico-Científica, Instituto de Desenvolvimento Sustentável Mamirauá, Estrada do Bexiga, 2584, Tefé, AM, 69470-000, Brazil
49Programa de Ciencias del Agro y el Mar, Herbario Universitario (PORT), UNELLEZ-Guanare, Guanare, Portuguesa, 3350, Venezuela
50Instituto de Biociências - Dept. Botanica, Universidade de Sao Paulo - USP, Rua do Matão 277, Cidade Universitária, São Paulo, SP, 05508-090, Brazil
51Laboratorio de Ecología de Bosques Tropicales y Primatología, Universidad de los Andes, Carrera 1 # 18a- 10, Bogotá, DC, 111711, Colombia
52Programa de Pós-Graduação em Biologia (Botânica), Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Petrópolis, Manaus, AM, 69067-375, Brazil
53Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Sciencepark 904, Amsterdam, 1098 XH, The Netherlands
54Endangered Species Coalition, 8530 Geren Rd., Silver Spring, MD, 20901, USA
55Inventory and Monitoring Program, National Park Service, 120 Chatham Lane, Fredericksburg, VA, 22405, USA
56Center for Conservation and Sustainability, Smithsonian Conservation Biology Institute, 1100 Jefferson Dr. SW, Suite 3123, Washington, DC, 20560-0705, USA
57Department of Global Ecology, Carnegie Institution for Science, 260 Panama St., Stanford, CA, 94305, USA

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

58Universidade Federal do Amapá, Ciências Ambientais, Rod. Juscelino Kubitschek km2, Macapá, AP, 68902-280, Brazil
59Department of Integrative Biology, University of California, Berkeley, CA, 94720-3140, USA
60Biologia Vegetal, Universidade Estadual de Campinas, Caixa Postal 6109, Campinas, SP, 13.083-970, Brazil
61Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, 215 Tower Road, Ithaca, NY, 14850, USA
62Peruvian Center for Biodiversity and Conservation (PCBC), Iquitos, Peru
63Department of Ecology, University of Brasilia, Brasilia, DF, 70904-970, Brazil
64ICNHS, Federal University of Mato Grosso, Av. Alexandre Ferronato 1200, Setor Industrial, Sinop, MT, 78.557-267, Brazil
65Natural Capital and Plant Health, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK
66Programa de Pós-Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Petrópolis, Manaus, AM, 69067-375, Brazil
67Forest Ecology and Forest Management Group, Wageningen University & Research, Droevendaalsesteeg 3, Wageningen, P.O. Box 47, 6700 AA, The Netherlands
68Escola de Negócios Tecnologia e Inovação, Centro Universitário do Pará, Belém, PA, Brazil
69Universidade Federal do Pará, Rua Augusto Corrêa 01, Belém, PA, 66075-110, Brazil
70Faculty of Natural Sciences, Department of Life Sciences, Imperial College London, Silwood Park, South Kensington Campus, London, SW7 2AZ, UK
71Ecosistemas, Biodiversidad y Conservación de Especies, Universidad Estatal Amazónica, Km. 2 1/2 vía a Tena (Paso Lateral), Puyo, Pastaza, Ecuador
72Museo de Historia Natural Noel Kempff Mercado, Universidad Autónoma Gabriel Rene Moreno, Avenida Irala 565 Casilla Post al 2489, Santa Cruz, Santa Cruz, Bolivia
73Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Departamento de Genética, Ecologia e Evolução, Av. Antônio Carlos, 6627 Pampulha, Belo Horizonte, MG, 31270-901, Brazil
74Department of Biology, University of Miami, Coral Gables, FL, 33146, USA
75Fairchild Tropical Botanic Garden, Coral Gables, FL, 33156, USA
76Instituto de Biociências - Dept. Ecologia, Universidade de Sao Paulo - USP, Rua do Matão, Trav. 14, no. 321, Cidade Universitária, São Paulo, SP, 05508-090, Brazil
77Lancaster Environment Centre, Lancaster University, Lancaster, Lancashire, LA1 4YQ, UK
78Environmental Change Institute, University of Oxford, Oxford, Oxfordshire, OX1 3QY, UK
79Empresa Brasileira de Pesquisa Agropecuária, Embrapa Amapá, Rod. Juscelino Kubitschek km 5, Macapá, Amapá, 68903-419, Brazil
80Grupo de Ecología y Conservación de Fauna y Flora Silvestre, Instituto Amazónico de Investigaciones Imani, Universidad Nacional de Colombia sede Amazonia, Leticia, Amazonas, Colombia
81Universidad Regional Amazónica IKIAM, Km 7 via Muyuna, Tena, Napo, Ecuador
82School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

83Direccíon de Evaluación Forestal y de Fauna Silvestre, Av. Javier Praod Oeste 693, Magdalena del Mar, Peru
84Escuela de Biología Herbario Alfredo Paredes, Universidad Central, Ap. Postal 17.01.2177, Quito, Pichincha, Ecuador
85Department of Biological Sciences, Humboldt State University, 1 Harpst Street, Arcata, CA, 95521, USA
86Museu Universitário / Centro de Ciências Biológicas e da Natureza / Laboratório de Botânica e Ecologia Vegetal, Universidade Federal do Acre, Rio Branco, AC, 69915-559, Brazil
87Institute of Biological and Health Sciences, Federal University of Alagoas, Av. Lourival Melo Mota, s/n, Tabuleiro do Martins, Maceio, AL, 57072-970, Brazil
88Iwokrama International Centre for Rain Forest Conservation and Development, Georgetown, Guyana
89New York Botanical Garden, 2900 Southern Blvd, Bronx, New York, NY, 10458-5126, USA
90School of Geosciences, University of Edinburgh, 201 Crew Building, King's Buildings, Edinburgh, EH9 3JN, UK
91Tropical Diversity Section, Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh, Scotland, EH3 5LR, UK
92Servicios de Biodiversidad EIRL, Jr. Independencia 405, Iquitos, Loreto, 784, Peru
93Herbario Nacional de Bolivia, Universitario UMSA, Casilla 10077 Correo Central, La Paz, La Paz, Bolivia
94Center for Conservation and Sustainable Development, Missouri Botanical Garden, P.O. Box 299, St. Louis, MO, 63166-0299, USA
95Universidad Nacional de Jaén, Carretera Jaén San Ignacio Km 23, Jaén, Cajamarca, 06801, Peru
96Biology Department and Center for Energy, Environment and Sustainability, Wake Forest University, 1834 Wake Forest Rd, Winston Salem, NC, 27106, USA
97Laboratoire Evolution et Diversité Biologique, CNRS and Université Paul Sabatier, UMR 5174 EDB, Toulouse, 31000, France
98Andes to Amazon Biodiversity Program, Madre de Dios, Madre de Dios, Peru
99Department of Anthropology, University of Texas at Austin, SAC 5.150, 2201 Speedway Stop C3200, Austin, TX, 78712, USA
100Fundación Puerto Rastrojo, Cra 10 No. 24-76 Oficina 1201, Bogotá, DC, Colombia
101Colegio de Ciencias Biológicas y Ambientales-COCIBA & Galapagos Institute for the Arts and Sciences-GAIAS, Universidad San Francisco de Quito-USFQ, Quito, Pichincha, Ecuador
102Department of Wildlife Ecology and Conservation, University of Florida, 110 Newins-Ziegler Hall, Gainesville, FL, 32611, USA
103Biosystematics group, Wageningen University, Droevendaalsesteeg 1, Wageningen, 6708 PB, The Netherlands
104Fundación Estación de Biología, Cra 10 No. 24-76 Oficina 1201, Bogotá, DC, Colombia
105Direction régionale de la Guyane, ONF, Cayenne, F-97300, French Guiana
106PROTERRA, Instituto de Investigaciones de la Amazonía Peruana (IIAP), Av. A. Quiñones km 2,5, Iquitos, Loreto, 784, Peru
107ACEER Foundation, Jirón Cusco N° 370, Puerto Maldonado, Madre de Dios, Peru

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

108Departement EV, Muséum national d'histoire naturelle de Paris, 16 rue Buffon, Paris, 75005, France
109Amazon Conservation Team, Doekhieweg Oost #24, Paramaribo, Suriname
110Institut de Ciència i Tecnologia Ambientals, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
111Environmental Change Institute, Oxford University Centre for the Environment, Dyson Perrins Building, South Parks Road, Oxford, England, OX1 3QY, UK
112Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Apartado 7945, Bogotá, DC, Colombia
113Instituto de Ciência Agrárias, Universidade Federal Rural da Amazônia, Av. Presidente Tancredo Neves 2501, Belém, PA, 66.077-830, Brazil
114Escuela Profesional de Ingeniería Forestal, Universidad Nacional de San Antonio Abad del Cusco, Jirón San Martín 451, Puerto Maldonado, Madre de Dios, Peru
115Universidad Autónoma del Beni José Ballivián, Campus Universitario Final, Av. Ejercito, Riberalta, Beni, Bolivia
116Laboratory of Human Ecology, Instituto Venezolano de Investigaciones Científicas - IVIC, Ado 20632, Caracas, DC, 1020A, Venezuela
117Cambridge University Botanic Garden, 1 Brookside., Cambridge, CB2 1JE, UK
118School of Agriculture and Food Sciences - ARC Centre of Excellence for Environmental Decisions CEED, The University of Queensland, St. Lucia, QLD 4072, Australia
119University of Campinas, Plant Biology Department, Rua Monteiro Lobato, 255, Cidade Universitária Zeferino Vaz, Barão Geraldo, Campinas, São Paulo, CEP 13083-862, Brazil
120Resource Ecology Group, Wageningen University & Research, Droevendaalsesteeg 3a, Lumen, building number 100, Wageningen, Gelderland, 6708 PB, The Netherlands
121Laboratório de Ciências Ambientais, Universidade Estadual do Norte Fluminense, Av. Alberto Lamego 2000, Campos dos Goyatacazes, RJ, 28013-620, Brazil
122Instituto de Investigaciones para el Desarrollo Forestal (INDEFOR), Universidad de los Andes, Conjunto Forestal, 5101, Mérida, Mérida, Venezuela
123Departamento de Biologia, Universidade Federal do Amazonas - UFAM – Instituto de Ciências Biológicas – ICB1, Av General Rodrigo Octavio 6200, Manaus, AM, 69080-900, Brazil
124GeoIS, El Día 369 y El Telégrafo, 3° Piso, Quito, Pichincha, Ecuador
125Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, 48109, USA
126University of Nottingham, University Park, Nottingham, NG7 2RD, UK
127School of Environmental and Forest Sciences, University of Washington, Seattle, WA, 98195-2100, USA
128Environmental Science and Policy, Northern Arizona University, Flagstaff, AZ, 86011, USA
129Geography and the Environment, University of Texas at Austin, 305 E. 23rd Street, CLA building, Austin, TX, 78712, USA
130Medio Ambiente, PLUSPRETOL, Iquitos, Loreto, Peru
131The Mauritius Herbarium, Agricultural Services, Ministry of Agro-Industry and Food Security, Reduit, 80835, Mauritius
132Department of Bioscience, Aarhus University, Building 1540 Ny Munkegade, Aarhus C, Aarhus, DK-8000, Denmark

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

133Living Earth Collaborative, Washington University in Saint Louis, St. Louis, MO, 63130, USA
134Escuela de Ciencias Forestales (ESFOR), Universidad Mayor de San Simon (UMSS), Sacta, Cochabamba, Bolivia
135FOMABO, Manejo Forestal en las Tierras Tropicales de Bolivia, Sacta, Cochabamba, Bolivia
136Tropenbos International, Lawickse Allee 11 PO Box 232, Wageningen, 6700 AE, The Netherlands
137School of Anthropology and Conservation, University of Kent, Marlowe Building, Canterbury, Kent, CT2 7NR, UK
138Herbario Nacional del Ecuador, Universidad Técnica del Norte, Quito, Pichincha, Ecuador
139Instituto de Biodiversidade e Floresta, Universidade Federal do Oeste do Pará, Rua Vera Paz, Campus Tapajós, Santarém, PA, 68015-110, Brazil
140Facultad de Biologia, Universidad Nacional de la Amazonia Peruana, Pevas 5ta cdra, Iquitos, Loreto, Peru
141Department of Biology, University of Missouri, St. Louis, MO, 63121, USA
142Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, P.O. Box 10 01 64, Jena, 07701, Germany
143Functional Biogeography, Max-Planck-Institute for Biogeochemistry, P.O. Box 10 01 64, Jena, 07701, Germany
144Department of Ecology and Evolutionary Biology, UCLA, 621 Charles E. Young Drive South, Box 951606, Los Angeles, CA, 90095, USA
145Systems Ecology, Vrije Universiteit Amsterdam, De Boelelaan 1087, Amsterdam, 1081 HV, The Netherlands

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract In a time of rapid global change, the question of what determines patterns in species abundance distribution remains a priority for understanding the complex dynamics of ecosystems. The constrained maximization of information entropy provides a framework for the understanding of such complex systems dynamics by a quantitative analysis of important constraints via predictions using least biased probability distributions. We apply it to over two thousand hectares of Amazonian tree inventories across seven forest types and thirteen functional traits, representing major global axes of plant strategies. Results show that constraints formed by regional relative abundances of genera explain almost ten times more of local relative abundances then constraints based on either directional or stabilizing selection for specific functional traits, although the latter does show clear signals of environmental dependency. These results provide a quantitative insight by inference from large-scale data using cross-disciplinary methods, furthering our understanding of ecological dynamics.

Key words: maximum entropy, information theory, biodiversity, Amazon rainforest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Drivers of species distributions and their predictions have been a long-standing search in ecology, with approaches varying from deterministic [1–6] to neutral [7–11] and almost everything in between (e.g. near-neutral, continuum or emergent-neutral: [12,13]). Most models are based on prior assumptions of processes that drive community dynamics. The Maximum Entropy Formalism (hereafter called MEF), however, makes no such a-priori assumptions in generating predictions of distributions, including those of species abundances [14,15,24,16–23]. It is a useful construct to infer processes driving community dynamics given the constraints imposed by prior knowledge (e.g. functional traits or summed regional abundances), as it does not include any bias introduced by potentially unjustified assumptions (*29*). Quantifying the relative importance of these distinct constraints can thus provide additional answers to understand the complexity of community dynamics (see Supporting Materials SM: boxes S1-S3). This is especially so because, although many different tests are available that link variation in taxon abundances to 1) trait variation, 2) taxon turnover between habitats or environments and 3) the distance decay of similarities between samples, none quantify the importance of these relative to each other. The MEF as applied here, however, is capable of and designed to do exactly this by decomposing variation to separate information explained by each of these aspects in a four-step model (Box S2) (*25*). Its application to an unprecedented large tree inventory database on genus level taxonomy consisting of > 2,000 1-ha plots distributed over Amazonia [25] and a genus trait database of 13 key functional traits representing global axes of plant strategies [26] allows us to advance the study of Amazonian tree community dynamics from a new cross-disciplinary perspective.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Principles from information theory [14,15,27] can be used in an ecological setting to predict the most likely abundance state for each taxon while simultaneously maximizing entropy based on constraints. Maximization of entropy allows quantifying the information yield for each constraint and therefor identifies which constraints reduce entropy the most. Here we specifically use Shipley's mathematical model to quantify probabilities, similar to earlier studies [28,29].

Predictive power of the four-step model

Using a uniform prior and both CWM and CWV as constraints accounted for 23% on average of total deviance between observed and predicted relative abundances (measured by R^2_{KL} values, see Box S2 equation 5). Filtered by forest type this was 36% for podzol forests, *várzea* 25%, *igapó* 23%, swamp forests 34%, 24% and 21% for Guyana Shield and Pebas *terra firme* respectively and 20% for Brazilian Shield *terra firme* forests (see Table S1 for detailed decomposition). Using observed metacommunity relative abundances as prior regardless of CWM or CWV values accounted on average 58% for the combined dataset with all forest types between 50 and 60%, except for the Guyana Shield *terra firme* with 63%. Including both trait constraints and the metacommunity prior performed slightly better for the combined dataset (average 62%), with a minimum of 56% for *igapó* forests and a maximum of 66% for the Guyana Shield *terra firme* forests. To compensate for spurious relationships between regional abundances and local trait constraints, regardless of selection, explanatory power was regarded relative to model bias yielding the pure trait and metacommunity effects (Box S3, Fig. 2 and Table S1). This lowered the proportion of information accounted for and yielded average pure metacommunity

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

effects of 43% for the overall dataset ranging between 30 and 48% for each forest type separately with pure trait effects explaining only 5% of information for the combined dataset on average with for each forest type between 3 and 8%. Although the latter was lowered substantially, the explanatory power did appear to be strongly dependent on forest type. Supplementary material provides additional results relating to the predictive power of each model as well as the spatial gradient between the pure trait and metacommunity effect ratios (Figs. S2-3).

Direction and strength of selection of trait-based constraints

Each trait showed significant differences in lambda when compared between forest types (Fig. S1). Scatterplots of CWM trait values versus lambda show that, in general, higher lambda values correspond with higher CWM trait values (Figure S7), although the relationships are complex. Greater trait dissimilarity also correlated positively with proportion of deviance attributable to pure trait effects (Pearson R of .23; Fig. S9), supporting a trait-based selection. All traits except leaf nitrogen content also showed reduction of variance accompanied a strong difference in community weighted mean values associated with lambdas (either positive or negative), in line with expectations of trait-based selection (Fig. S10).

Effect of regional metacommunity prior

There was a remarkable similar mean 21% decrease of the information explained purely by the metacommunity prior for each forest type (Fig. 3). It should be noted there is an obvious risk that when sampling size is increased, this also includes more environmental heterogeneity as samples are coming from a variety of localities potentially leading to changing composition. If this were the case, however, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

regional prior (q_i from Fig. S1 and Box S2) would also change, as taxa might be abundant in some places but rare or absent in others. As the metacommunity effect is the explained information that remains relative to any trait effects (i.e. information unique to the neutral prior) and the pure trait effects are the explained information remaining after correcting for pure metacommunity effects (Box S3) this effect should then be accompanied by an increase in pure trait effect for each sample. This was not observed, not even within the different forest types. Instead, the trait effect gradually went up and then remained constant (Fig. S4).

Discussion

The MEF emerges from a well-founded theoretical and empirical body of ecology and evolutionary biology, regarding natural selection, migration and population dynamics [16,22,29,42,43]. From an ecological point of view it can be used to quantify the relative association between directional or stabilizing selection for functional traits versus the importance of relative regional abundance regardless of these traits by imposing these as constraints. Our results show that pure trait effects, on average, explained only 5% of the information when all forest types were taken together whereas the pure metacommunity effect, however, explained almost ten times more with an average value of 43%. Greater trait dissimilarity was positively associated with higher pure trait effects, indicating trait-based selection, although the assumed influence of dispersal regardless of these traits appeared to confer more information explaining tree genus composition of the Amazon rainforest. The strength and direction of selection indicated clear selective pressure for life history strategies of either growth or protection, depending on forest type (see appendix S-A for a detailed exploration of ecological interpretation).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Despite showing clear patterns in environmental selection and dispersal effects, there was a large proportion of information left unexplained (44% on average). Potentially, local demographic stochasticity could weaken any link between functional traits measured and regional abundances of genera. This would, however, mean that almost half of the information contained in relative abundances are the result of random population dynamics and are not structurally governed. Alternatively, this could be due to functional traits reflective of processes not taken into account in this study, such as traits reflective of interactions between trophic levels. Another and at least equally likely hypothesis for (local) unexplained information is that when scaling up, the ratio of genus richness to total abundance decreases rather rapidly at first but levels out as at some point relatively non-overlapping habitats are included in the regional abundance distributions and more genera are included again due to the different habitats. This would result in a change of the regional abundance distribution (the prior) to which each local community is compared, resulting in higher local unexplained information. Further study into these aspects could provide additional insight, although as of yet the data necessary for these scales is lacking.

Although the initial explanatory power of the metacommunity prior differed between forest types, the decay pattern was very similar. As the effects of either traits or the metacommunity are measured in the goodness-of-fit predictions on local relative abundances, this implies that at small spatial scales the surrounding regional abundances provide better estimators than functional traits, while at larger spatial scales this shifts to the traits. The ecological translation would be that on small spatial scales, local communities share similar environmental conditions leaving dispersal and drift acting in changing community composition, at least for genus level taxonomy. As the potential regional pool is increased, more and more environmental

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

heterogeneity and non-overlapping regions are likely to be introduced. The more gradual decline of *terra firme* forests can then arguably be attributed to these forests having the largest relative surface area of Amazonia (even for the separate subregions), potentially giving these forests an almost continuous metacommunity without gaps, resulting in a more gradual transition from metacommunity to trait relative importance. The fact the metacommunity effects do not change anymore after certain distances would indicate the effect of dispersal potentially occurs over very large distances. It should be noted that as these calculations are done at community and genus level, they do not measure single dispersal events but rather the effect of dispersal on community composition much deeper in time. In other words, this effect suggests more than a dispersal event every now and then. Instead, it argues for prolonged mixing of forests on large geographical and temporal scales, supported by recent findings demonstrating a lack of geographical phylogenetic structure of lineages for Amazonian tree genera [44].

Using an unprecedented scale of data and applying the Maximum Entropy Formalism from information theory we show that constraints formed by regional relative abundances of genera explain almost ten times more of local relative abundances then constraints based on either directional or stabilizing selection for specific functional traits, although the latter does show clear signals of environmental dependency. There is, however, still much to be explored due to the large unexplained effects and analyses on finer taxonomic (i.e. species level) and environmental (e.g. microhabitat) scales could resolve these issues. The relatively large effects of the regional pool of genera over great distances does suggest an important role for long term dispersal and mixing of Amazonian trees, especially for the Amazonian interior.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Empirical data

The ATDN (ter Steege et al.) consists of over 2000 tree inventory plots distributed over the Amazon basin and the Guiana Shield, collectively referred to as Amazonia. Only those plots with trees ≥ 10 cm diameter at breast height were used, leaving 2011 plots with a mean of 558 individuals per plot identified to at least genus level. Most plots used are 1 ha in size (1414) with 492 being smaller (minimum size of .1 ha) and 105 larger (maximum size of 80 ha). Genera have been standardized to the W3 Tropicos database ("Tropicos Missouri Botanical Garden") using the Taxonomic Name Resolution Service (TNRS [32]). After filtering based on above criteria and solving nomenclature issues, 1,121,935 individuals belonging to over 828 genera remained. Plots were distributed over seven abiotically different forest types: Podzol forests (PZ), *Igapó* (IG, black water flood forests), *Várzea* (VA white water flood forests), Swamp (SW) and *Terra firme* forests (TF) with subregions BS (Brazilian Shield), GS (Guyana Shield) and PB (Pebas).

Functional traits and trait imputation

Constraints were formed by community weighted means (CWM) and variance of functional traits (CWV), related to key ecological life history aspects on which natural selection potentially operates (Table 1). According to principles of natural selection, CWM values will be biased towards favourable trait values for that particular environment in the case of directional selection, as taxa with these traits will be more abundant due to environmental selection while stabilizing selection would decrease CWV values [33]. For many traits it has been shown earlier that the interspecific variability was larger than the intraspecific variability, allowing the use of data from

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

565 different sources to at least calculate a mean species trait value [34]. Genus trait
566 values were computed as genus-level means of species values if known within the
567 genus and considered constant for each genus. Genus level of taxonomy was used as
568 the available trait database had the most information on this taxonomic level.
569 Unknown values for traits were estimated by Multiple Imputation with Chained
570 Equations (MICE) by delta adjustment, subtracting a fixed amount (delta), with
571 sensitivity of this adjustment to the imputations of the observed versus imputed data
572 analysed using density plots (Fig. S8) and a linear regression model [35]. Procedure
573 was done using the *mice* package available on the R repository [36] under predictive
574 mean matching (*pmm* setting, 50 iterations). Results showed imputations were stable
575 and showed near identical patterns with each imputation scenario (see Figs S5-6 and
576 Table S2). After imputation, all trait values were transformed to Community
577 Weighted Means (CWM) of each trait (*J*) for each plot (*K*) ($\overline{T}_{JK}$) as $\overline{T}_{JK} =$
578 $\sum_{i=1}^{S} t_{ij} ra_{ik}$ with *ra* the relative abundance of the ith genus in the kth plot, following
579 earlier uses [37].
580

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

581 *MEF procedure predictions and ecological inference*

582 Figure 1 provides a schematic procedure overview, box S1 provides an overview of
583 important terms and Boxes S2-3 mathematical details. Initially, a maximally
584 uninformative prior is specified, where q_i (Box S1 equation 1) equals 1/S and trait
585 constraints are randomly permuted multiple times among genera to test whether
586 inclusion of specified constraints significantly changes derived probability
587 distributions (see also Roxburgh & Mokany, 2010). Subsequently, the same prior is
588 used but now observed trait CWM or CWV belonging to specific genera are used as
589 constraints. Third, observed regional abundances are used as prior with permutated
590 trait constraints and finally both observed regional abundances and observed trait
591 CWM/CWV are used as prior and constraints. *Maxent2* [28,29], an updated version
592 of the *maxent* function currently in the FD library of R [39] provided the
593 computational platform. Proportions of uncertainty explained by each model are given
594 by the Kullback-Leibler divergence R^2_{KL}, a generalization of the classic R^2 goodness
595 of fit [29]. Pure trait, pure metacommunity, joint metacommunity-trait and
596 unexplained effects are calculated as proportions of total biologically relevant
597 information (Box S1 and Box S2). Data was rarefied to smallest sample size (swamp
598 forests; 28) and calculations bootstrapped 25 times. Results indicated no significant
599 change compared to using all data, hence the total dataset was used for all analyses.
600

601 *Strength and direction of selection*

602 Predictions of genus relative abundances are computed as a function of traits reflected
603 in the CWM or CWV values and a series of constants (λ_{jk}: the Lagrange Multipliers).
604 Each multiplier quantifies the association between a unit of change for a particular
605 trait *j* and a proportional change in predicted relative abundance p_{ik} (the ith genus in

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the kth community) considering all other traits are constant, formally described as: $\partial p_{ik} / \partial t_{ij} = \lambda_{jk} p_{ik} (1 - p_{ik})$ (see appendix 1 from Sonnier, Navas, Fayolle, & Shipley, 2012). Positive values indicate larger trait values associated with higher abundances (positive selection), negative values indicate the opposite (negative selection) with changes proportional to lambda. Values approximating zero indicate no association between specific traits and relative abundances of species. Decomposing λ_{jk} and comparing by means of a One-Way Analysis of Variance for each trait separately between forest types allows studying both the strength and direction of selection in different habitats. Note that this is done for the same constraint between forest types, as lambda values for each constraint do not scale linearly between different constraints.

Estimation of metacommunity size

Iteratively increasing the regional species pool taken into account as prior in concentric circles of a fixed radius of 50 km allows estimating the spatial effect of metacommunity size. The relationship between pure metacommunity effect and radius of metacommunity size was fitted using a smoothing loess regression (function *loess* and *predict;* R-package *stats* [41] with span set at 0.1). Fits subsequently were used to predict values of metacommunity effect based on geographical distance to visualize general patterns for each forest type. Exponential decay of pure metacommunity effect was described using a self-start asymptotic regression function (*SSasymp*) of the form $y(t) \sim y_f + (y_0 - y_f) e^{-exp(log(\alpha))t}$ (*nls* from *stats,* R Core Team, 2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

632 **References and notes:**

633 [1] D. Tilman, *Resource Competition between Plankton Algae: An Experimental*
634 *and Theoretical Approach*, Ecology **58**, 338 (1977).

635 [2] D. Tilman, *Resource Competition and Community Structure* (Princeton
636 university press, 1982).

637 [3] R. H. MacArthur, *Geographical Ecology: Patterns in the Distribution of*
638 *Species* (Princeton University Press, 1972).

639 [4] J. M. Chase and M. a Leibold, *Spatial Scale Dictates the Productivity-*
640 *Biodiversity Relationship.*, Nature **416**, 427 (2002).

641 [5] P. Chesson, *Mechanisms of Maintenance of Species Diversity*, Annu. Rev.
642 Ecol. Syst. **31**, 343 (2000).

643 [6] D. Tilman and S. W. Pacala, *The Maintenance of Species Richness in Plant*
644 *Communities*, Species Diversity in Ecological Communities.

645 [7] S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography.*
646 (Princeton Monographs in Population Biology. Princeton University Press,
647 Princeton, New Jersey, USA, 2001).

648 [8] I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan, *Neutral Theory and*
649 *Relative Species Abundance in Ecology.*, Nature **424**, 1035 (2003).

650 [9] J. Rosindell and S. J. Cornell, *Universal Scaling of Species-Abundance*
651 *Distributions across Multiple Scales*, Oikos **122**, 1101 (2012).

652 [10] G. Bell, *The Distribution of Abundance in Neutral Communities.*, Am. Nat.
653 **155**, 606 (2000).

654 [11] G. Bell, *The Co-Distribution of Species in Relation to the Neutral Theory of*
655 *Community Ecology*, Ecology.

656

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[12] R. a Chisholm and S. W. Pacala, *Niche and Neutral Models Predict Asymptotically Equivalent Species Abundance Distributions in High-Diversity Ecological Communities.*, Proc. Natl. Acad. Sci. U. S. A. **107**, 15821 (2010).

[13] D. Tilman, *Niche Tradeoffs, Neutrality, and Community Structure: A Stochastic Theory of Resource Competition, Invasion, and Community Assembly.*, Proc. Natl. Acad. Sci. U. S. A. **101**, 10854 (2004).

[14] E. T. Jaynes, *Information Theory and Statistical Mechanics I*, Phys. Rev. **106**, 620 (1957).

[15] E. T. Jaynes, *Information Theory and Statistical Mechanics. II*, Phys. Rev. **108**, 171 (1957).

[16] B. Shipley, D. Vile, and E. Garnier, *From Plant Traits to Plant Communities: A Statistical Mechanistic Approach to Biodiversity*, Science (80-.). **314**, 812 (2006).

[17] B. J. McGill, *A Renassaince in the Study of Abundance*, Solid State Phys. **314**, 770 (2006).

[18] S. Pueyo, F. He, and T. Zillio, *The Maximum Entropy Formalism and the Idiosyncratic Theory of Biodiversity*, Ecol. Lett. **10**, 1017 (2007).

[19] R. C. Dewar and A. Porté, *Statistical Mechanics Unifies Different Ecological Patterns*, J. Theor. Biol. **251**, 389 (2008).

[20] B. Shipley, *Inferential Permutation Tests for Maximum Entropy Models in Ecology*, Ecology **91**, 2794 (2010).

[21] J. Bertram and R. C. Dewar, *Combining Mechanism and Drift in Community Ecology: A Novel Statistical Mechanics Approach*, Theor. Ecol. **8**, 419 (2015).

[22] J. Harte, *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics* (OUP Oxford, 2011).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

683 [23] J. Harte and E. A. Newman, *Maximum Information Entropy: A Foundation for*
684 *Ecological Theory*, Trends Ecol. Evol. **29**, 384 (2014).
685 [24] J. P. O'Dwyer, A. Rominger, and X. Xiao, *Reinterpreting Maximum Entropy in*
686 *Ecology: A Null Hypothesis Constrained by Ecological Mechanism*, Ecol. Lett.
687 **20**, 832 (2017).
688 [25] H. ter Steege, N. C. A. Pitman, D. Sabatier, C. Baraloto, R. P. Salomão, J. E.
689 Guevara, O. L. Phillips, C. V Castilho, W. E. Magnusson, J.-F. Molino, A.
690 Monteagudo, P. Núñez Vargas, J. C. Montero, T. R. Feldpausch, E. N. H.
691 Coronado, T. J. Killeen, B. Mostacedo, R. Vasquez, R. L. Assis, J. Terborgh, F.
692 Wittmann, A. Andrade, W. F. Laurance, S. G. W. Laurance, B. S. Marimon,
693 B.-H. Marimon, I. C. Guimarães Vieira, I. L. Amaral, R. Brienen, H.
694 Castellanos, D. Cárdenas López, J. F. Duivenvoorden, H. F. Mogollón, F. D. de
695 A. Matos, N. Dávila, R. García-Villacorta, P. R. Stevenson Diaz, F. Costa, T.
696 Emilio, C. Levis, J. Schietti, P. Souza, A. Alonso, F. Dallmeier, A. J. D.
697 Montoya, M. T. Fernandez Piedade, A. Araujo-Murakami, L. Arroyo, R.
698 Gribel, P. V. A. Fine, C. A. Peres, M. Toledo, G. A. Aymard C, T. R. Baker, C.
699 Cerón, J. Engel, T. W. Henkel, P. Maas, P. Petronelli, J. Stropp, C. E. Zartman,
700 D. Daly, D. Neill, M. Silveira, M. R. Paredes, J. Chave, D. de A. Lima Filho, P.
701 M. Jørgensen, A. Fuentes, J. Schöngart, F. Cornejo Valverde, A. Di Fiore, E.
702 M. Jimenez, M. C. Peñuela Mora, J. F. Phillips, G. Rivas, T. R. van Andel, P.
703 von Hildebrand, B. Hoffman, E. L. Zent, Y. Malhi, A. Prieto, A. Rudas, A. R.
704 Ruschell, N. Silva, V. Vos, S. Zent, A. A. Oliveira, A. C. Schutz, T. Gonzales,
705 M. Trindade Nascimento, H. Ramirez-Angulo, R. Sierra, M. Tirado, M. N.
706 Umaña Medina, G. van der Heijden, C. I. A. Vela, E. Vilanova Torre, C.
707 Vriesendorp, O. Wang, K. R. Young, C. Baider, H. Balslev, C. Ferreira, I.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mesones, A. Torres-Lezama, L. E. Urrego Giraldo, R. Zagt, M. N. Alexiades, L. Hernandez, I. Huamantupa-Chuquimaco, W. Milliken, W. Palacios Cuenca, D. Pauletto, E. Valderrama Sandoval, L. Valenzuela Gamarra, K. G. Dexter, K. Feeley, G. Lopez-Gonzalez, M. R. Silman, S. P. Hubbell, F. He, R. Condit, L. Borda-de-Agua, J. Kellner, H. Ter Steege, G. A. Black, T. H. Dobzhansky, C. Pavan, J. M. Pires, T. Dobzhansky, G. A. Black, M. J. G. Hopkins, M. J. Costello, R. M. May, N. E. Stork, P. Haripersaud, H. ter Steege, J.-J. de Granville, H. Chevillotte, M. Hoff, D. P. Bebber, M. A. Carine, J. R. Wood, A. H. Wortley, D. J. Harris, G. T. Prance, G. Davidse, J. Paige, T. D. Pennington, N. K. Robson, R. W. Scotland, B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, E. P. White, R. J. Warren, D. K. Skelly, O. J. Schmitz, M. A. Bradford, N. C. A. Pitman, J. W. Terborgh, M. R. Silman, P. N. V, D. A. Neill, C. E. Cerón, W. A. Palacios, M. Aulestia, N. C. A. Pitman, M. R. Silman, J. W. Terborgh, F. D. Lozano, M. W. Schwartz, M. W. Schwartz, D. Simberloff, J. E. Richardson, R. T. Pennington, T. D. Pennington, P. M. Hollingsworth, T. L. Couvreur, F. Forest, W. J. Baker, S. Cavers, C. W. Dick, D. H. Janzen, S. A. Mangan, S. A. Schnitzer, E. A. Herre, K. M. Mack, M. C. Valencia, E. I. Sanchez, J. D. Bever, W. Balée, D. G. Campbell, C. Levis, P. F. de Souza, J. Schietti, T. Emilio, J. L. P. V. Pinto, C. R. Clement, F. R. C. Costa, D. A. Posey, E. Montoya, V. Rull, N. D. Stansell, M. B. Abbott, S. Nogué, B. W. Bird, W. A. Díaz, C. Gomez-Navarro, C. Jaramillo, F. Herrera, S. L. Wing, R. Callejas, C. H. McMichael, D. R. Piperno, M. B. Bush, M. R. Silman, A. R. Zimmerman, M. F. Raczka, L. C. Lobato, H. ter Steege, P. P.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

733 Haripersaud, O. S. Bánki, F. Schieving, S. J. Phillips, R. P. Anderson, R. E.
734 Schapire, S. J. Phillips, M. Dudik, C. A. Quesada, J. Lloyd, L. O. Anderson, N.
735 M. Fyllas, M. Schwarz, C. I. Czimczik, B. Rollet, H. ter Steege, N. C. Pitman,
736 O. L. Phillips, J. Chave, D. Sabatier, A. Duque, J. F. Molino, M. F. Prévost, R.
737 Spichiger, H. Castellanos, P. von Hildebrand, R. Vásquez, P. M. Fearnside, D.
738 Mouillot, D. R. Bellwood, C. Baraloto, J. Chave, R. Galzin, M. Harmelin-
739 Vivien, M. Kulbicki, S. Lavergne, S. Lavorel, N. Mouquet, C. E. Paine, J.
740 Renaud, W. Thuiller, G. Lopez-Gonzalez, S. L. Lewis, M. Burkitt, O. L.
741 Phillips, P. J. M. Maas, L. Y. T. Westra, H. Rainer, A. Q. Lobão, R. H. J.
742 Erkens, P. V. A. Fine, D. C. Daly, G. V. Muñoz, I. Mesones, K. M. Cameron,
743 K. J. Feeley, M. R. Silman, M. Dufrene, P. Legendre, J. Bunge, M. Fitzpatrick,
744 J. Bunge, L. Woodard, D. Böhning, J. A. Foster, S. Connolly, H. K. Allen, J.-P.
745 Z. Wang, B. G. Lindsay, A. Chao, R. K. Colwell, C.-W. Lin, N. J. Gotelli, C.
746 X. Mao, R. K. Colwell, U. Brose, N. D. Martinez, R. J. Williams, I. J. Good, A.
747 Chao, S.-M. Lee, I. Rocchetti, J. Bunge, D. Bohning, R. A. Fisher, A. S.
748 Corbet, C. B. Williams, F. W. Preston, J.-P. Wang, P. V. A. Fine, and R. H.
749 Ree, *Hyperdominance in the Amazonian Tree Flora*, Science **342**, 1243092
750 (2013).

751 [26] S. Díaz, J. Kattge, J. H. C. Cornelissen, I. J. Wright, S. Lavorel, S. Dray, B.
752 Reu, M. Kleyer, C. Wirth, I. Colin Prentice, E. Garnier, G. Bönisch, M.
753 Westoby, H. Poorter, P. B. Reich, A. T. Moles, J. Dickie, A. N. Gillison, A. E.
754 Zanne, J. Chave, S. Joseph Wright, S. N. Sheremet Ev, H. Jactel, C. Baraloto,
755 B. Cerabolini, S. Pierce, B. Shipley, D. Kirkup, F. Casanoves, J. S. Joswig, A.
756 Günther, V. Falczuk, N. Rüger, M. D. Mahecha, and L. D. Gorné, *The Global*
757 *Spectrum of Plant Form and Function*, Nature **529**, 167 (2016).

758

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[27] C. E. Shannon, *A Mathematical Theory of Communication*, Bell Syst. Tech. J. **27**, 379 (1948).

[28] B. Shipley, *Measuring and Interpreting Trait-Based Selection versus Meta-Community Effects during Local Community Assembly*, J. Veg. Sci. **25**, 55 (2014).

[29] B. Shipley, C. E. T. Paine, and C. Baraloto, *Quantifying the Importance of Local Niche-Based and Stochastic Processes to Tropical Tree Community Assembly*, Ecology **93**, 760 (2012).

[30] H. ter Steege, A. Monteagudo, A. F. Sampaio, A. Ruschel, A. Prieto, A. Quaresma, A. Rudas, A. Vicentini, A. Araujo-Murakami, A. A. de Oliveira, A. Alonso, A. Fuentes, A. Lopes, A. Duque, A. S. de Andrade, A. Cano, A. G. Manzatto, A. Di Fiore, A. Torres-Lezama, B. S. Marimon, B. H. M. Jr, B. W. Albuquerque, B. Laurance, B. Magnusson, B. Mostacedo, B. Villa, B. Hoffman, B. Cintra, B. G. Luize, C. Cerón, C. Peres, C. Castilho, C. Levis, C. Mendoza, C. Vela, C. E. Zartman, C. Baraloto, C. Ferreira, C. Baider, C. Vriesendorp, D. Cárdenas, D. Praia, D. Sabatier, D. Villarroel, D. Pauletto, D. D. do Amaral, D. Neill, D. Correa, D. Reyes, D. de A. L. Filho, D. Rodrigues, D. Daly, E. M. Barbosa, E. M. Venticinque, E. Zent, E. M. Jimenez, E. Valderrama, E. de S. Farias, E. V. Torre, E. Berenguer, E. Honorio, E. Almeida, E. M. M. de L. Novo, F. A. Carvalho, F. C. de Souza, F. C. Valverde, F. Costa, F. Wittmann, F. D. de A. Matos, F. Dallmeier, F. R. Arevalo, G. Damasco, G. Lopez-Gonzalez, G. van der Heijden, G. P. G. Gonzales, G. Aymard, G. Rivas, H. Beltrán, H. Balslev, H. E. M. Nascimento, H. Castellanos, H. P. D. Doza, H. Ramirez-Angulo, H. Mogollón, I. Amaral, I. Vieira, I. P. de A. Miranda, I. Huamantupa-Chuquimaco, I. Mesones, J. Costa,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

J.-F. Molino, J.-L. Guillaumet, J. Chave, J. Comiskey, J. Ricardo, J. Schöngart, J. Terborgh, J. Ferreira, J. Lloyd, J. Duivenvoorden, J. Barlow, J. F. Ramos, J. J. de Toledo, J. L. Camargo, J. L. M. Pena, J. R. da S. Guimarães, J. C. Licona, J. C. Montero, J. D. C. Revilla, J. E. Guevara, J. F. Phillips, J. Schietti, J. Stropp, J. Engel, K. Garcia, K. R. Casula, K. Feeley, K. R. Young, K. Dexter, L. Silva, L. Demarchi, L. Ferreira, L. E. Urrego, L. F. M. Pinto, L. Hernandez, L. Barbosa, L. V. Gamarra, L. Casas, L. C. de M. Bonates, L. de S. Coelho, L. Arroyo, M. Rocha, M. Piedade, M. Ahuite, M. C. Guedes, M. de J. V. Carim, M. Medeiros, M. P. Pansonato, M. Simon, M. T. Nascimento, M. dos Santos, M. R. Paredes, M. Silveira, M. C. P. Mora, M. N. U. Medina, M. P. Martins, M. V. Irume, M. Alexiades, M. Silman, M. Aulestia, M. Tirado, N. Dávila, N. Targhetta, N. Silva, N. F. C. Reis, N. C. Arboleda, N. Pitman, O. Bánki, O. Phillips, O. Wang, P. Stevenson, P. Petronelli, P. von Hildebrand, P. Fine, P. Maas, P. N. Vargas, P. M. Jørgensen, P. Souza, R. L. de Assis, R. Salomão, R. Thomas-Caesar, R. R. Hilário, R. Linares-Palomino, R. Z. Gómez, R. Foster, R. Zagt, R. Vasquez, R. Sierra, R. Brienen, R. Gribel, R. García, S. Cárdenas, S. Mori, S. Zent, S. Pansini, S. Laurance, T. Feldpausch, T. Henkel, T. Emilio, T. Gonzales, T. S. F. Silva, T. Baker, T. Killeen, T. van Andel, T. Gardner, T. Pennington, V. Vos, W. Palacios, W. Campelo, W. Farfan, W. Milliken, W. Nauray, Y. Malhi, and Y. Feitosa, *Http://Atdn.Myspecies.Info/*, (unpublished).

[31] *Tropicos Missouri Botanical Garden*, (unpublished).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[32] B. Boyle, N. Hopkins, Z. Lu, J. A. Raygoza Garay, D. Mozzherin, T. Rees, N. Matasci, M. L. Narro, W. H. Piel, S. J. McKay, S. Lowry, C. Freeland, R. K. Peet, and B. J. Enquist, *The Taxonomic Name Resolution Service: An Online Tool for Automated Standardization of Plant Names.*, BMC Bioinformatics **14**, 16 (2013).

[33] J. Loranger, F. Munoz, B. Shipley, and C. Violle, *What Makes Trait–Abundance Relationships When Both Environmental Filtering and Stochastic Neutral Dynamics Are at Play?*, Oikos **127**, 1735 (2018).

[34] E. Kazakou, C. Violle, C. Roumet, M. L. Navas, D. Vile, J. Kattge, and E. Garnier, *Are Trait-Based Species Rankings Consistent across Data Sets and Spatial Scales?*, J. Veg. Sci. **25**, 235 (2014).

[35] S. Van Buuren, *Flexible Imputation of Missing Data* (Chapman and Hall/CRC, 2018).

[36] S. van Buuren and K. Groothuis-Oudshoorn, *Mice: Multivariate Imputation by Chained Equations in R*, J. Stat. Softw. **45**, 1 (2011).

[37] B. Shipley, D. C. Laughlin, G. Sonnier, and R. Otfinowski, *A Strong Test of a Maximum Entropy Model of Trait-Based Community Assembly*, Ecology **92**, 507 (2011).

[38] S. H. Roxburgh and K. Mokany, *On Testing Predictions of Species Relative Abundance from Maximum Entropy Optimisation*, Oikos **119**, 583 (2010).

[39] E. Laliberté and B. Shipley, *FD: Measuring Functional Diversity from Multiple Traits, and Other Tools for Functional Ecology. R Package Version 1.0-11*, URL Http//CRAN. R-Project. Org/Package= FD [Accessed 9 January 2012] (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[40] G. Sonnier, M. L. Navas, A. Fayolle, and B. Shipley, *Quantifying Trait Selection Driving Community Assembly: A Test in Herbaceous Plant Communities under Contrasted Land Use Regimes*, Oikos **121**, 1103 (2012).

[41] R Core Team, *R: A Language and Environment for Statistical Computing*.

[42] B. Shipley, *From Plant Traits to Vegetation Structure. Chance and Selection in the Assembly of Ecological Communities* (Cambridge University Press, 2010).

[43] M. Favretti, *Remarks on the Maximum Entropy Principle with Application to the Maximum Entropy Theory of Ecology*, Entropy **20**, (2018).

[44] K. G. Dexter, M. Lavin, B. M. Torke, A. D. Twyford, T. A. Kursar, P. D. Coley, C. Drake, R. Hollands, and R. T. Pennington, *Dispersal Assembly of Rain Forest Tree Communities across the Amazon Basin*, Proc. Natl. Acad. Sci. **114**, 2645 (2017).

[45] E. Weiher, A. van der Werf, K. Thompson, M. Roderick, E. Garnier, and O. Eriksson, *Challenging Theophrastus: A Common Core List of Plant Traits for Functional Ecology*, J. Veg. Sci. **10**, 609 (1999).

[46] K. G. Raghothama, *Phosphate Acquisition*, Annu. Rev. Plant Physiol. Plant Mol. Biol. **50**, 665 (1999).

[47] L. Poorter, M. Van De Plassche, S. Willems, and R. G. A. Boot, *Leaf Traits and Herbivory Rates of Tropical Tree Species Differing in Successional Status*, Plant Biol. **6**, 746 (2004).

[48] A. A. Agrawal, *Macroevolution of Plant Defense Strategies*, Trends Ecol. Evol. **22**, 103 (2007).

[49] J. I. Sprent and R. Parsons, *Nitrogen Fixation in Legume and Non-Legume Trees*, F. Crop. Res. **65**, 183 (2000).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

856 [50] F. Shah, C. Nicolas, J. Bentzer, M. Ellstrom, M. Smits, F. Rineau, B. Canback,
857 D. Floudas, R. Carleer, G. Lackner, J. Braesel, D. Hoffmeister, B. Henrissat, T.
858 Johansson, D. S. Hibbett, F. Martin, P. Persson, and A. Tunlid,
859 *Ectomycorrhizal Fungi Decompose Soil Organic Matter Using Oxidative*
860 *Mechanisms Adapted from Saprotrophic Ancestors*, New Phytol. 1705 (2017).
861 [51] J. V. Colpaert, J. H. L. Wevers, E. Krznaric, and K. Adriaensen, *How Metal-*
862 *Tolerant Ecotypes of Ectomycorrhizal Fungi Protect Plants from Heavy Metal*
863 *Pollution*, Ann. For. Sci. **68**, 17 (2011).
864 [52] C. Poschenrieder, B. Gunsé, I. Corrales, and J. Barceló, *A Glance into*
865 *Aluminum Toxicity and Resistance in Plants*, Sci. Total Environ. **400**, 356
866 (2008).
867 [53] H. F. Howe and J. Smallwood, *Ecology of Seed Dispersal Ecology of Seed*
868 *Dispersal*, **13**, 201 (2013).
869 [54] C. Baraloto, C. E. T. Paine, L. Poorter, J. Beauchene, D. Bonal, A. M.
870 Domenach, B. Hérault, S. Patiño, J. C. Roggy, and J. Chave, *Decoupled Leaf*
871 *and Stem Economics in Rain Forest Trees*, Ecol. Lett. **13**, 1338 (2010).
872 [55] C. Fortunel, P. V. A. Fine, and C. Baraloto, *Leaf, Stem and Root Tissue*
873 *Strategies across 758 Neotropical Tree Species*, Funct. Ecol. **26**, 1153 (2012).
874 [56] N. Fyllas and S. Patino, *Basin-Wide Variations in Foliar Properties of*
875 *Amazonian Forest: Phylogeny, Soils and Climate*, Biogeosciences 2677 (2009).
876 [57] N. J. B. Kraft, R. Valencia, and D. D. Ackerly, *Functional Traits and Niche-*
877 *Based Tree Community Assembly in an Amazonian Forest.*, Science **322**, 580
878 (2008).
879 [58] F. M. J. Ohler, *Phytomass and Mineral Content in Untouched Forest*, (1980).
880

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

881 [59] C. E. T. Paine, C. Baraloto, J. Chave, and B. Hérault, *Functional Traits of*
882 *Individual Trees Reveal Ecological Constraints on Community Assembly in*
883 *Tropical Rain Forests*, Oikos **120**, 720 (2011).
884 [60] T. L. Pons, K. Perreijn, C. Van Kessel, and M. J. A. Werger, *Symbiotic*
885 *Nitrogen Fixation in a Tropical Rainforest: 15N Natural Abundance*
886 *Measurements Supported by Experimental Isotopic Enrichment*, New Phytol.
887 **173**, 154 (2007).
888 [61] J. Thompson, J. Proctor, V. Viana, W. Milliken, J. A. Ratter, and D. A. Scott,
889 *Ecological Studies on a Lowland Evergreen Rain Forest on Maraca Island,*
890 *Roraima, Brazil. I. Physical Environment, Forest Structure and Leaf*
891 *Chemistry*, J. Ecol. 689 (1992).
892 [62] J. Kattge, S. Díaz, S. Lavorel, I. C. Prentice, P. Leadley, G. Bönisch, E.
893 Garnier, M. Westoby, P. B. Reich, I. J. Wright, J. H. C. Cornelissen, C. Violle,
894 S. P. Harrison, P. M. Van Bodegom, M. Reichstein, B. J. Enquist, N. A.
895 Soudzilovskaia, D. D. Ackerly, M. Anand, O. Atkin, M. Bahn, T. R. Baker, D.
896 Baldocchi, R. Bekker, C. C. Blanco, B. Blonder, W. J. Bond, R. Bradstock, D.
897 E. Bunker, F. Casanoves, J. Cavender-Bares, J. Q. Chambers, F. S. Chapin, J.
898 Chave, D. Coomes, W. K. Cornwell, J. M. Craine, B. H. Dobrin, L. Duarte, W.
899 Durka, J. Elser, G. Esser, M. Estiarte, W. F. Fagan, J. Fang, F. Fernández-
900 Méndez, A. Fidelis, B. Finegan, O. Flores, H. Ford, D. Frank, G. T. Freschet,
901 N. M. Fyllas, R. V. Gallagher, W. A. Green, A. G. Gutierrez, T. Hickler, S. I.
902 Higgins, J. G. Hodgson, A. Jalili, S. Jansen, C. A. Joly, A. J. Kerkhoff, D.
903 Kirkup, K. Kitajima, M. Kleyer, S. Klotz, J. M. H. Knops, K. Kramer, I. Kühn,
904 H. Kurokawa, D. Laughlin, T. D. Lee, M. Leishman, F. Lens, T. Lenz, S. L.
905 Lewis, J. Lloyd, J. Llusià, F. Louault, S. Ma, M. D. Mahecha, P. Manning, T.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

906 Massad, B. E. Medlyn, J. Messier, A. T. Moles, S. C. Müller, K. Nadrowski, S.
907 Naeem, Ü. Niinemets, S. Nöllert, A. Nüske, R. Ogaya, J. Oleksyn, V. G.
908 Onipchenko, Y. Onoda, J. Ordoñez, G. Overbeck, W. A. Ozinga, S. Patiño, S.
909 Paula, J. G. Pausas, J. Peñuelas, O. L. Phillips, V. Pillar, H. Poorter, L. Poorter,
910 P. Poschlod, A. Prinzing, R. Proulx, A. Rammig, S. Reinsch, B. Reu, L. Sack,
911 B. Salgado-Negret, J. Sardans, S. Shiodera, B. Shipley, A. Siefert, E. Sosinski,
912 J. F. Soussana, E. Swaine, N. Swenson, K. Thompson, P. Thornton, M.
913 Waldram, E. Weiher, M. White, S. White, S. J. Wright, B. Yguel, S. Zaehle, A.
914 E. Zanne, and C. Wirth, *TRY - a Global Database of Plant Traits*, Glob. Chang.
915 Biol. **17**, 2905 (2011).

916

917 **Author Contributions:**

918 E.T. Pos and H. ter Steege designed the experiment. E.T.Pos wrote R-scripts,
919 analysed the results and took the lead in writing the manuscript, H. ter Steege
920 supervised the writing and provided regular feedback both for the manuscript and the
921 interpretation of the results. All other authors provided feedback on the manuscript
922 and provided their data from the Amazon Tree Diversity Network or trait data.
923 Authors E.T. Pos to L.V. Gamarra provided tree inventory data, authors G. Boenisch,
924 J. Kattge, N. Kraft, A. Levesley, K. Melgaço, G. Pickavance, L. Poorter provided data
925 on functional traits, C. Baraloto, J. Lloyd, A. A. Oliveira and H. ter Steege provided
926 both tree inventory and functional trait data. **Competing interests:** Authors declare
927 no competing interests. **Data and materials availability:** R scripts are available on
928 the github repository of E.T. Pos (EdwinTPos). The data that support the findings of
929 this study are available from The Amazon Tree Diversity Network (ATDN) upon
930 reasonable request.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements:

This paper is the result of the work of hundreds of different scientists and research institutions in the Amazon over the past 80 years. Without their hard work this analysis would have been impossible. The 25-ha Long-Term Ecological Research Project of Amacayacu is a collaborative project of the Instituto Amazónico de Investigaciones Científicas Sinchi and the Universidad Nacional de Colombia Sede Medellín, in partnership with the Unidad de Manejo Especial de Parques Naturales Nacionales and the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute (CTFS). The Amacayacu Forest Dynamics Plot is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. We acknowledge the Director and staff of the Amacayacu National Park for supporting and maintaining the project in this National Park. We would also like to thank Prof. John Harte, Prof. Bill Shipley and the anonymous reviewers for their constructive feedback during earlier review process.

Funding: HtS and RS were supported by grant 407232/2013-3 - PVE - MEC/MCTI/CAPES/CNPq/FAPs; CB was supported by grant FAPESP 95/3058-0 - CRS 068/96 WWF Brasil - The Body Shop; DS, JFM, JE, PP and JC benefited from an “Investissement d’Avenir” grant managed by the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01); HLQ/MAP/JLLM received financial supported by MCT/CNPq/CT-INFRA/GEOMA #550373/2010-1 and # 457515/2012-0, and JLLM were supported by grant CAPES/PDSE # 88881.135761/2016-01 and CAPES/Fapespa #1530801; The Brazilian National Research Council (CNPq) provided a productivity grant to EMV (Grant 308040/2017-1); Floristic identification

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

955 in plots in the RAINFOR forest monitoring network have been supported by the
956 Natural Environment Research Council (grants NE/B503384/1, NE/ D01025X/1,
957 NE/I02982X/1, NE/F005806/1, NE/D005590/1 and NE/I028122/1) and the Gordon
958 and Betty Moore Foundation; B.M.F. is funded by FAPESP grant 2016/25086-3.
959 BSM, BHMJ and OLP were supported by grants CNPq/CAPES/FAPS/BC-Newton
960 Fund #441244/2016-5 and FAPEMAT/0589267/2016; TWH was funded by National
961 Science Foundation grant DEB-1556338.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437717; this version posted April 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE 1

Functional trait	**Units**	**Mean**	**SD**	**Est %**	**Associated challenge**
Wood density (*WD*)	g/cm^3	0.63	0.17	30	Longevity [45]
Seed Mass Class (*SMC*)	categorical (1-8)	4.3	1.4	31	Dispersal, Fecundity, Establishment [45]
Specific Leaf Area (*SLA*)	mm^2/mg	15	5.9	41	Establishment, Plasticity, Disturbance [45]
Leaf nitrogen content (*N*)	mg/g	22.3	7.30	41	Photosynthetic capacity [45]
Leaf phosphorus content (*P*)	mg/g	1	0.77	50	Limited available P for metabolism [46]
Leaf carbon content (*C*)	mg/g	468	38.1	54	Herbivore resistance (C:N) [47]
Latex	1=no, 2 =yes	1.2	0.43	46	Herbivore resistance [48]
Resin	1=no, 2 =yes	1.1	0.35	58	Herbivore resistance [48]
Root Nodules (*Nodules*)	1=no, 2 =yes	1.1	0.28	0	Nitrogen fixation [49]
Ectomycorrhiza (*EctoMyco*)	1=no, 2 =yes	1.01	0.11	0	Organic N fixation [50], heavy metal pollution [51]
Aluminum accumulation (*AlAcc*)	1=no, 2 =yes	1.1	0.21	3	Heavy metal pollution [52]
Fleshy Fruits (*Fleshy*)	1=no, 2 =yes	1.6	0.50	7	Dispersal (*specificity*) [53]
Winged seeds (*Wings*)	1=no, 2 =yes	1.2	0.42	39	Dispersal (*limitation*) [53]

Table 1. Overview of used functional traits. Mean and standard deviation (SD) are calculated after predictive mean matching (percentage of estimated values is given by Est (%)). Associated challenge indicates different aspects of life history and selective environment related to specific functional traits, sources are given in the footnote. For specific methodology of measurement protocols and calculation for each trait we refer to the original sources of the data: Chris Baraloto (TRY), Adalardo de Oliveira (unpublished data), L. Poorter (unpublished data), J. Lloyd (TRY), Van der Sande and Mazzei (unpublished data), Van der Sande and Poorter (unpublished data), [54–62]

FIGURE 1

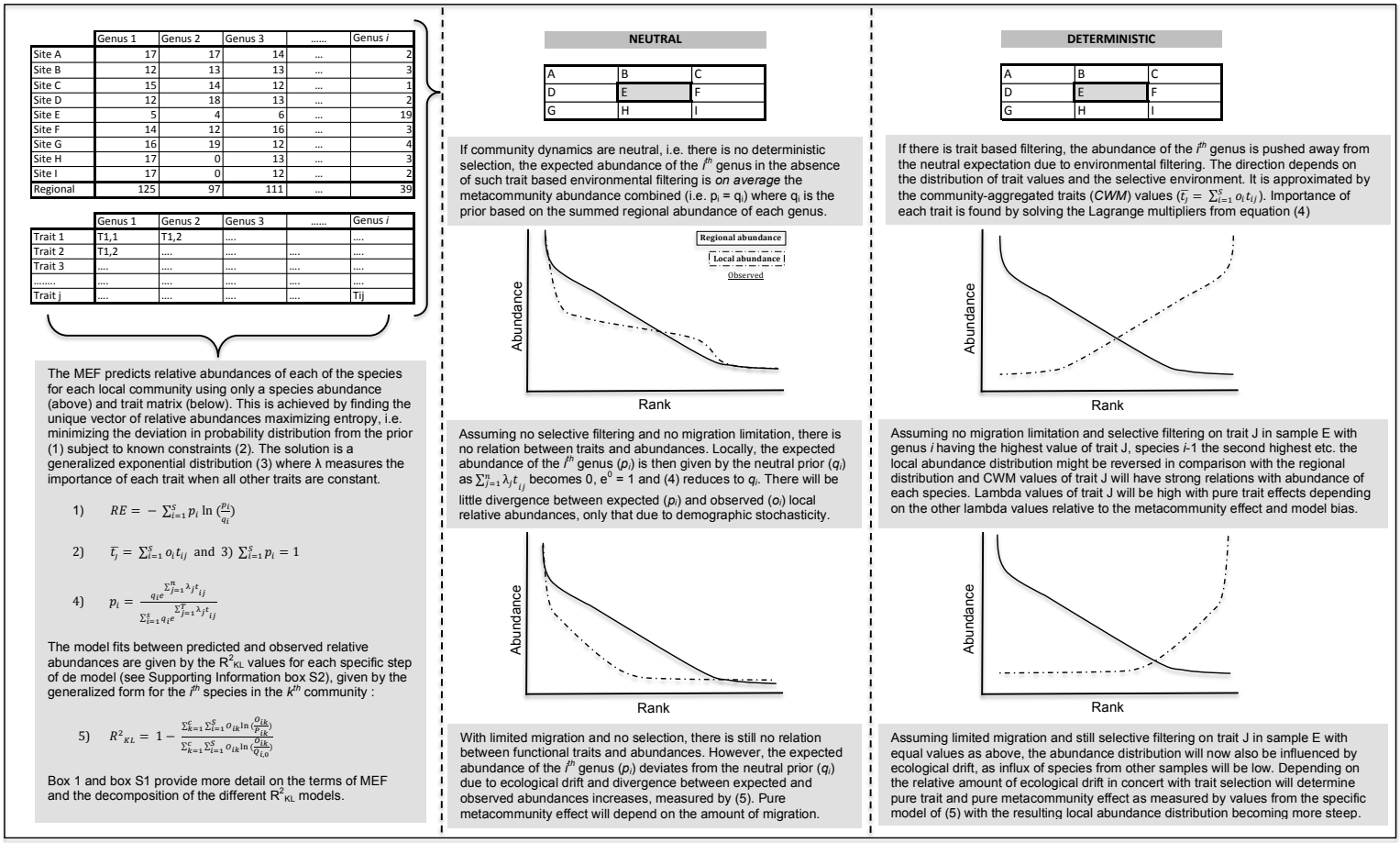

Fig. 1. Schematic depiction of the MEF procedure. Left panel shows a genus abundances per site and a functional trait matrix per genus, bottom half outlines calculations. Middle and right panel show different scenarios of neutral and deterministic dynamics under infinite or limited migration. Dashed and solid lines indicate local and regional abundance distributions respectively.

FIGURE 2

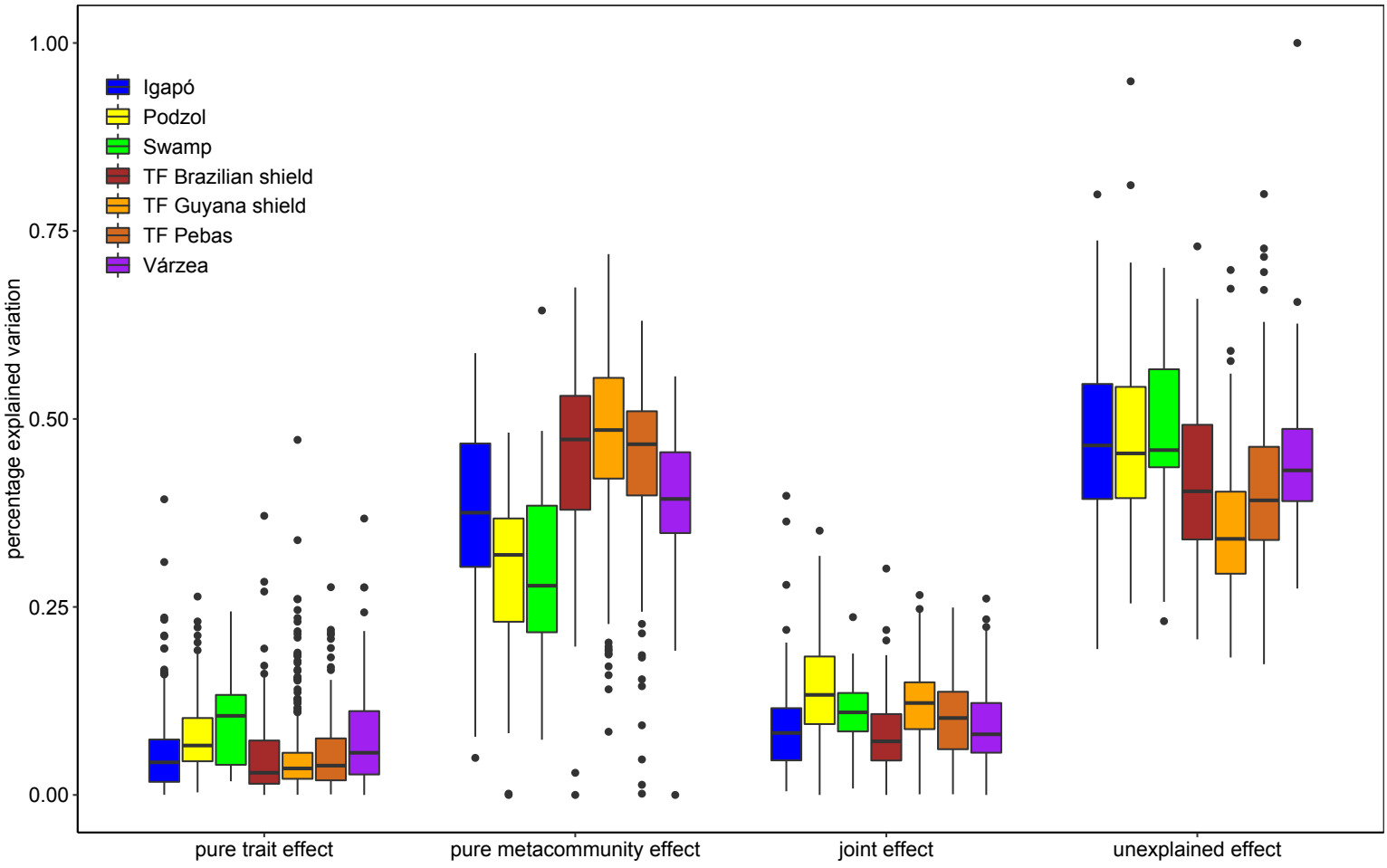

Fig. 2. Visual representation of pure trait, pure metacommunity, hybrid model and the remaining unexplained information for each separate forest type.

Abbreviations indicate different types: igapó (IG), podzol (PZ), swamp (SW), Brazilian shield terra firme (TFBS), Guiana Shield terra firme (TFGS), Pebas terra firme (TFPB) and várzea (VA). Boxplots show median value of pure effects over all samples, with lower and upper hinges corresponding to 25th and 75th percentiles. Whiskers extends from hinge to largest or smallest value no further than 1.5 * IQR from hinge. Points beyond this range are plotted individually.

FIGURE 3

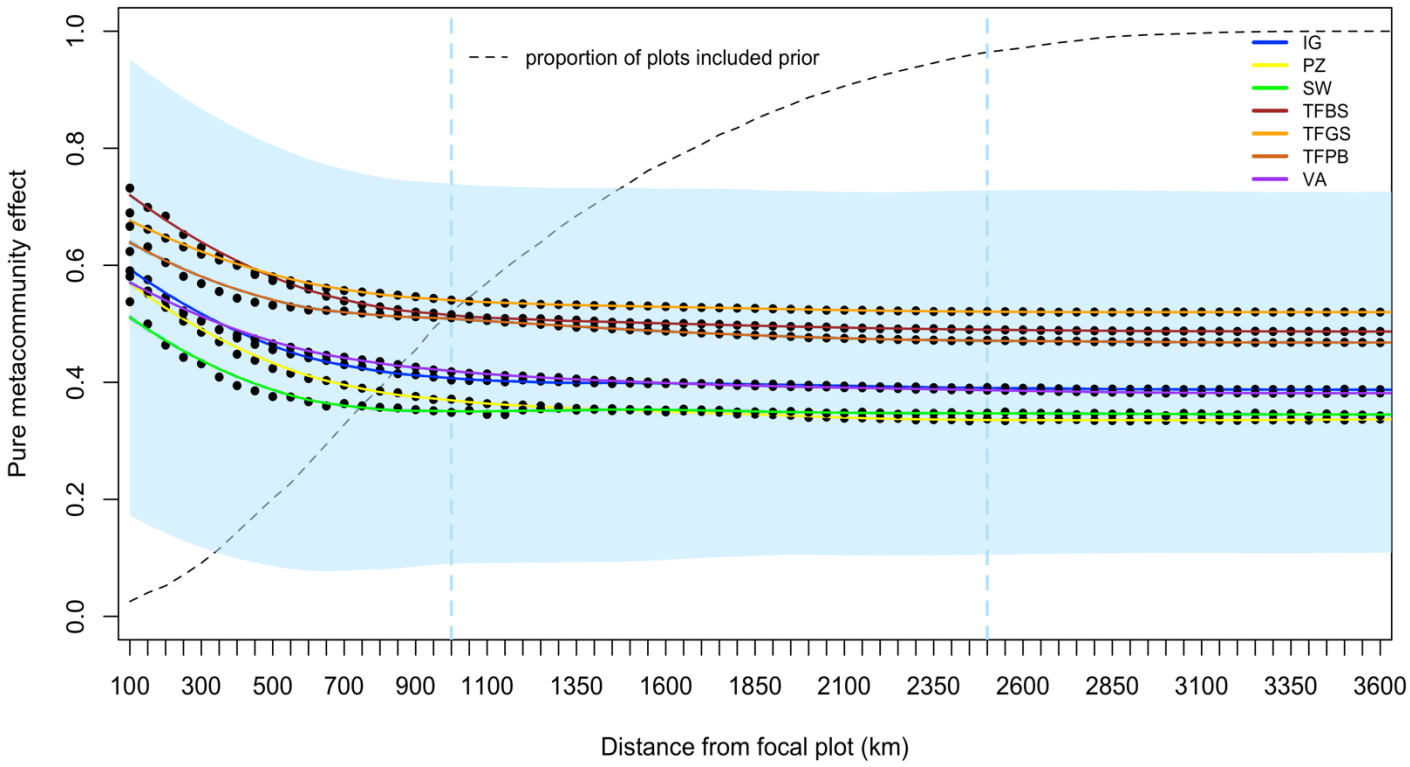

Fig. 3. Distance decay of pure metacommunity effect. X-axis represents radius of metacommunity prior. Dashed line indicates mean number of plots for that distance included as metacommunity prior. Y-axis represents information unique to metacommunity prior taken relative to model bias. Solid lines indicate predictions from loess regression based on all points with different colours indicating forest types with abbreviations as in main text. Blue shading reflects minimum and maximum loess regression predicted values.

BOX S1

Entities:
The basic unit of the MEF model can exist in different states. If the system under study is a collection of genera existing at a site, then each entity is a single genus.

States:
Classification of different ways any entity can exist. In the same collection of taxa, states of each entity (i.e. genus) are their specific abundance at that site. Microstates are the exact arrangement in time and space for the states of the entities in the system. Macrostates are the description of entities among the possible states in the system under study without regard to the spatial or temporal arrangement of these entities. I.e. observing a relative abundance distribution, but not the actual dispersal and germination of individuals.

Traits, attributes or properties:
Each entity possesses measurable properties whose values will probably differ between states. For example, genera differ in average wood density, seed mass, height etcetera.

Maximally uninformative prior:
All the information concerning states before constraints are introduced. Called maximally uninformative as preferably all empirical information is introduced in the form of constraints as to have the maximal gain of information regarding the different traits.

Prior distribution:
Prior distribution of expected states for the entities which can be incorporated as a constraint in addition to the traits, being either the observed relative abundance of each entity in the summed sample (i.e. the metacommunity) or a maximally uninformed (uniform) distribution. The former would be a neutral prior (expected local abundance is equal to the abundance in the larger metacommunity).

Community-weighted means:
The average trait value (i.e. measurable property such as wood density) of entities (such as genera) weighted by the relative abundance of each entity at a specific site

Box S1. Different ingredients necessary for analyses using MEF. Definitions of the most important terms used in the MEF analyses and throughout the main text to provide the necessary framework of understanding.

BOX S2

The Maximum Entropy Formalism works on the basis of a conceptual model called the CATS (*Community Assembly by Trait Selection*) and makes use of three inputs:

i) A **trait matrix** containing the measured functional traits of each of the S total genera in the total regional pool, these can be of either discrete or continuous form.

ii) A **vector of *n* community weighted trait values**, estimating the average trait value over all individuals in the local community for each of the traits

iii) A **prior probability distribution** specifying the regional abundance distribution, quantifying potential contributions of the regional pool of recruits to the structure of local communities.

Using these three sources of information, the model predicts relative abundances (p_i) in the form of Bayesian probabilities for each genus in each local community without assuming any a priori relations or processes. This is achieved by finding the vector of relative abundances maximizing entropy:

$$1)\ RE = -\sum_{i=1}^{S} p_i \ln\left(\frac{p_i}{q_i}\right)$$

with q_i the regional species pool abundance of species i and RE (Relative Entropy) subject to the known constraints for j traits and i species.:

$$2)\ \overline{t_j} = \sum_{i=1}^{S} o_i t_{ij} \text{ and } 3)\ \sum_{i=1}^{S} p_i = 1$$

The solution is a generalized exponential distribution where the λ values measure the importance of each trait when all other traits are constant:

$$4)\ p_i = \frac{q_i e^{\sum_{j=1}^{n} \lambda_j t_{ij}}}{\sum_{i=1}^{s} q_i e^{\sum_{j=1}^{T} \lambda_j t_{ij}}}$$

Note that when all λ values are zero, i.e. there is no trait based selection, $p_i = q_i$

The final step is to measure the proportion of total deviance accounted for between observed and predicted relative abundances for each of the four-step solution. These are the R^2_{KL} values, a generalization of the classic R^2 index of maximum likelihood estimation using the Kullback-Leibler index [16]:

i) $\overline{\mathbf{R^2_{KL}}}$**(u):** fit of model bias, the model null hypotheses given a uniform prior (i.e. equal distribution in the regional pool of recruits).

ii) $\mathbf{R^2_{KL}}$**(u, t):** fit using again a uniform prior but including traits as constraints.

iii) $\overline{\mathbf{R^2_{KL}}}$**(m):** fit using the metacommunity prior but excluding traits as constraints

iv) $\mathbf{R^2_{KL}}$**(m, t):** fit using the metacommunity prior and including traits as constraints

The general form of the $\mathbf{R^2_{KL}}$ divergence is calculated by:

$$5)\ R^2_{KL} = 1 - \frac{\sum_{k=1}^{c}\sum_{i=1}^{S} O_{ik} \ln\left(\frac{O_{ik}}{P_{ik}}\right)}{\sum_{j=k}^{c}\sum_{i=1}^{S} O_{ik} \ln\left(\frac{O_{ij}}{Q_{i,0}}\right)}$$

With the following parameters:

O_{ik} as the observed relative abundances of the ith genus in the kth community,

P_{ik} the accompanying predicted values for the specific model of the four solution step as described in the main text and,

$Q_{i,0}$ the predicted relative abundances given only the maximum uninformative prior.

Further details on the calculation of all separate $\mathbf{R^2_{KL}}$ values and accompanying pure trait, pure metacommunity, joint information and biologically unexplained information can be found in the SOM (box S2).

Box S2. Mathematical description of the Maximum Entropy Formalism for the four-step solution. Left panel shows the necessary ingredients and basic formulation of the Maximum Entropy Formalism. Right side panel shows decomposition of the proportion of total deviance accounted for between observed and predicted relative abundances for each of the four-step solution.

BOX S3

The purpose of using MEF is to decompose the deviance between observed and predicted relative abundances using the four-step solution as described in the main text. The values generated are described below. The R^2_{KL} value is a generalization of the classic R^2 index of maximum likelihood estimation using the Kullback-Leibler index for a non-linear regression including a multinomial error structure [16–18]*. In essence, it is a way of measuring the proportion of total deviance accounted for by that specific model from one of the four steps:*

$\bar{R}^2_{KL}(u)$: fit of model bias, the model null hypotheses given a uniform prior and permuted traits
$R^2_{KL}(u, t)$: fit using a uniform prior but including observed traits as constraints
$\bar{R}^2_{KL}(m)$: fit using the metacommunity prior but excluding observed traits as constraints
$R^2_{KL}(m, t)$: fit using the metacommunity prior and including observed traits as constraints

1) The increase in the explained deviance due to traits can be calculated either by

$\Lambda R^2_{KL}(t|\varphi) = R^2_{KL}(u, t) - \bar{R}^2_{KL}(u)$
Increase in explained deviance due to traits beyond that due solely to model bias

or **$\Lambda R^2_{KL}(t|m) = R^2_{KL}(m, t) - \bar{R}^2_{KL}(m)$**
Increase in explained deviance due to traits beyond contributions made by the meta-community

2) The increase in explained deviance due dispersal mass effects via the metacommunity can be calculated by either:

$\Lambda R^2_{KL}(m|\varphi) = \bar{R}^2_{KL}(m) - \bar{R}^2_{KL}(u)$
Increase in explained deviance (if any) due to the metacommunity beyond that due to model bias

or **$\Lambda R^2_{KL}(m|t) = R^2_{KL}(m, t) - R^2_{KL}(u, t)$**
Increase in explained deviance due to the meta-community given traits, relative to the explained deviance due only to the traits: i.e. information unique to neutral prior

3) And finally the joint information and the biologically unexplained information:

$\Lambda R^2_{KL}(m+t) = \Lambda R^2_{KL}(m|\varphi) - \Lambda R^2_{KL}(m|t) = \Lambda R^2_{KL}(t|\varphi) - \Lambda R^2_{KL}(t|m)$
Joint information gain, or increase in explained deviance due to both the metacommunity prior and the constraints based on the traits

1- **$\Lambda R^2_{KL}(m,t)$**
Biologically unexplained variation

From these values the pure trait, pure metacommunity, joint effect and biologically unexplained variation can be calculated by the following calculations:

Pure trait effects: **$\Lambda R^2_{KL}(t|m) / (1-\bar{R}^2_{KL}(u))$**
Pure metacommunity effects: **$\Lambda R^2_{KL}(m|t) / (1-\bar{R}^2_{KL}(u))$**
Joint metacommunity and trait effects: **$\Lambda R^2_{KL}(m+t) / (1-\bar{R}^2_{KL}(u))$**
Unexplained effects: 1-**$\Lambda R^2_{KL}(m,t) / (1-\bar{R}^2_{KL}(u))$**

Box S3. Detailed decomposition of the four-step solution from the MEF. Mathematical description of the decomposition based on the constraints and prior distributions (both uniform and neutral) for each of the steps from the four-step solution to measure the proportion of total deviance accounted for by each specific model from one of the four steps.

FIGURE S1

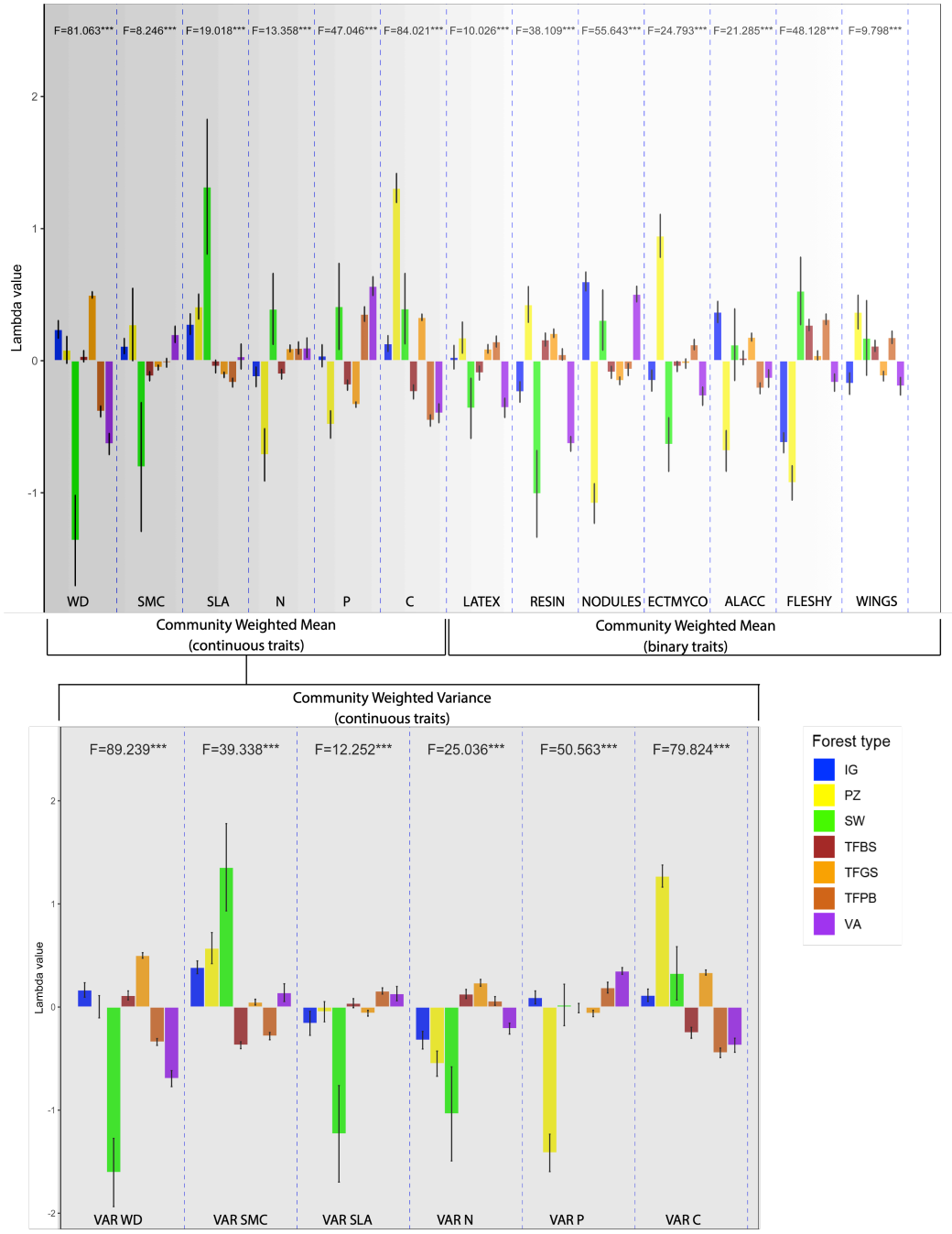

Fig. S1. Mean lambda values with standard error bars for each functional trait and compared between forest types. Forest types are *igapó* (IG), podzol (PZ), swamp (SW), Brazilian shield *terra firme* (TFBS), Guiana Shield *terra* firme (TFGS), Pebas *terra* firme (TFPB) and *várzea* (VA). Differences were tested with a one way analysis of variance with significance levels corresponding to: ns non-significant, * $p < .05$, ** $p < .01$ and *** $p < .001$. Traits used were wood density (WD), seed mass class (SMC), specific leaf area (SLA), nitrogen (N), phosphorus (P) and carbon (C) leaf content with the prefix of VAR for the variance of continuous traits. Latex, Resin, Nodules, Ectomycorrhiza (EctoMyco), the ability to accumulate aluminum (AlAcc), and the presence/absence of fleshy fruits (Fleshy) and winged seeds (Wings) were are all binary traits.

FIGURE S2

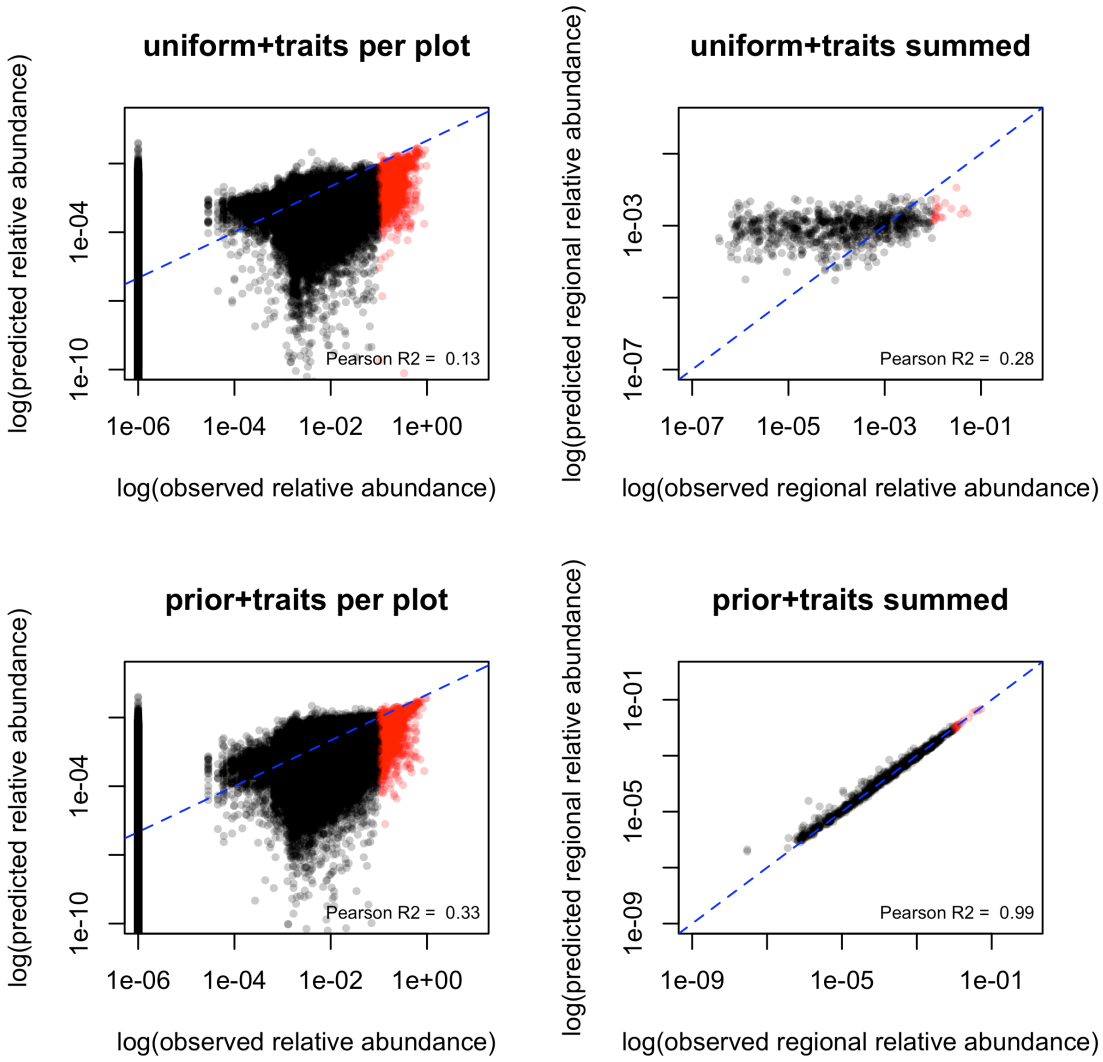

Fig. S2. Observed relative abundances for each genus in all plots plotted against predicted relative abundance per plot (left) and summed (right) using only the traits as constraints in combination with a uniform prior (top) or the hybrid model using both traits and the metacommunity relative abundance as prior (bottom) on a log-log scale. Top figures show predictions using only a uniform prior, left separate for all plots and right for all genera summed over all plots. Bottom figures show predictions using the regional prior, again separate for all plots and genera (left) and summed over all plots for each genus (right). Red points indicate taxa with observed relative abundances over 1e-1. Lines show the x=y prediction and R^2 values correspond to the Pearson's correlation coefficient. Reported R2 value is equal to the Pearson correlation coefficient (R) between the observed and predicted relative abundances defined as one minus the ratio of the error sum of squares to the total sum of squares.

FIGURE S3

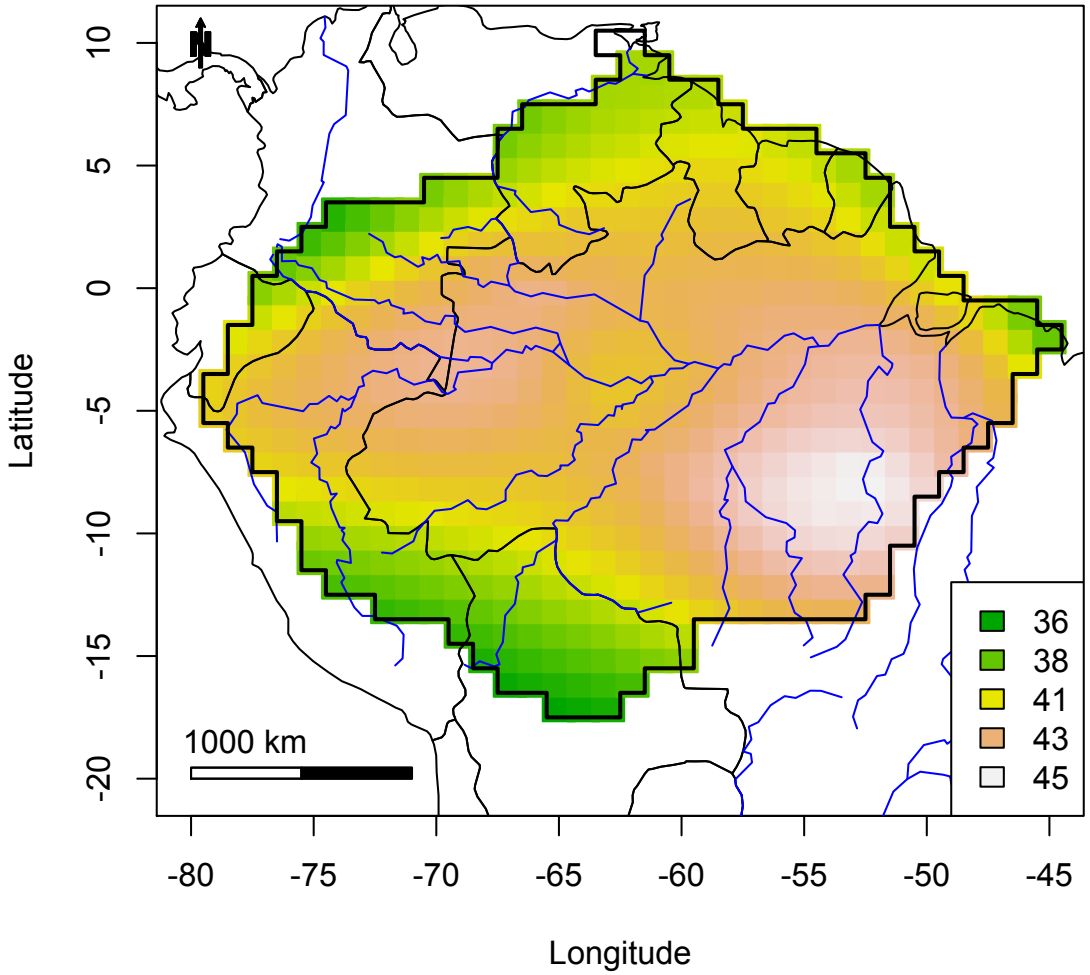

Fig. S3. Spatial gradient in pure trait relative to pure metacommunity effect. Map showing the ratio between the pure metacommunity effect and the pure trait effects for each plot. Ratio was calculated per plot by dividing the pure metacommunity effect by (10*pure trait effects+1). Values for projection on the map using a loess regression were multiplied by 1000 to allow clearer differentiation. Squares show the predictions from loess regression (color depending on value). Map shows interior of the Amazon having weaker trait effects relative to metacommunity effects whereas on the edges of the Amazon this pattern is reversed.

FIGURE S4

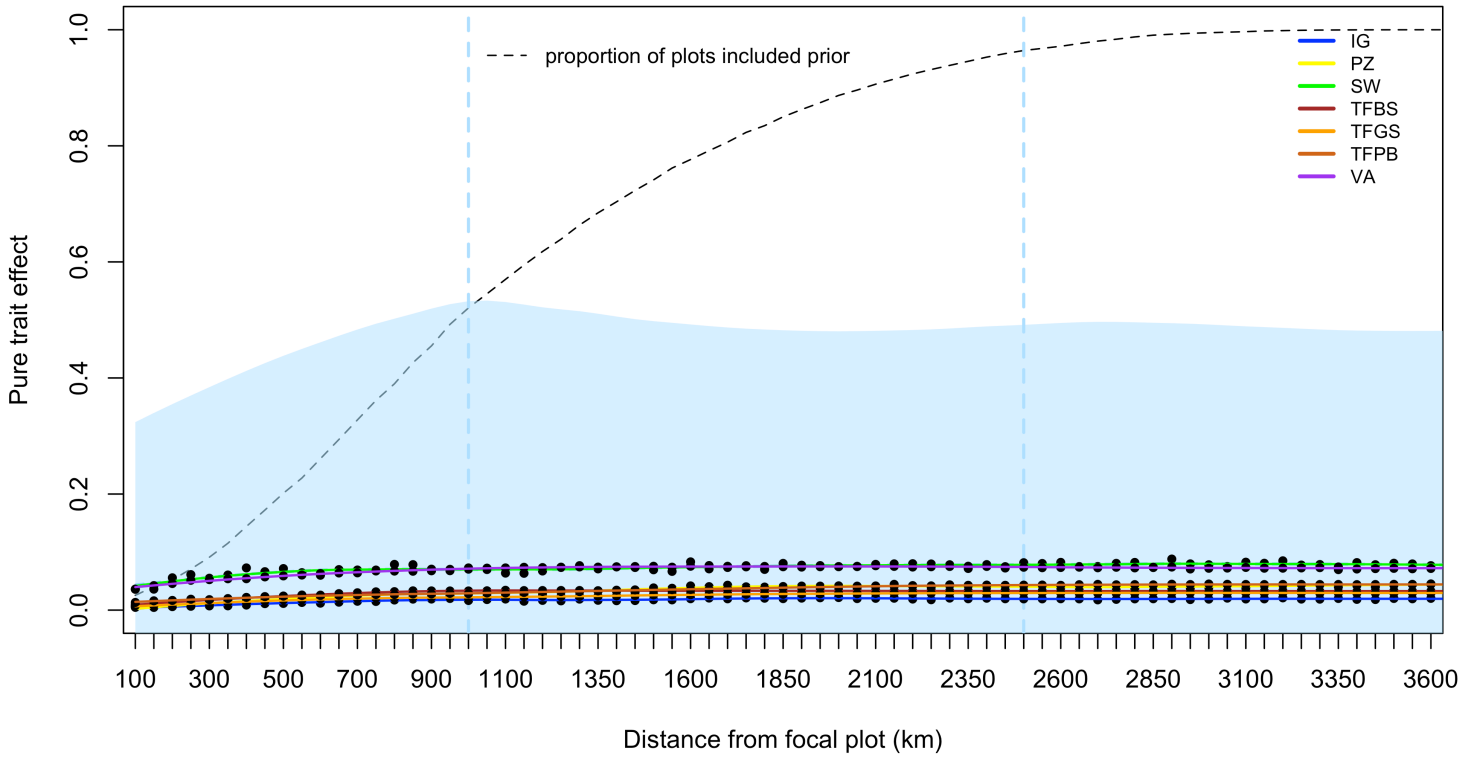

Fig. S4. Distance decay of pure trait effect for each forest type separately and the overall dataset. X-axis represents the radius of the metacommunity prior; i.e. the first 100 km consists of just a few plots and at 3800 km all plots are taken into account. Y-axis represent the pure trait effect, i.e. the increase in explained deviance due to traits beyond contributions made by the meta-community and relative to the model bias (see also Box S2). Colors indicate the different forest types with abbreviations as in main text. Lines indicate the predictions following from the loess regression based on all points. Blue vertical lines indicate the 1000 and 2500 km boundary points. Blue shading reflects maximum values for that distance of the whole dataset.

FIGURE S5

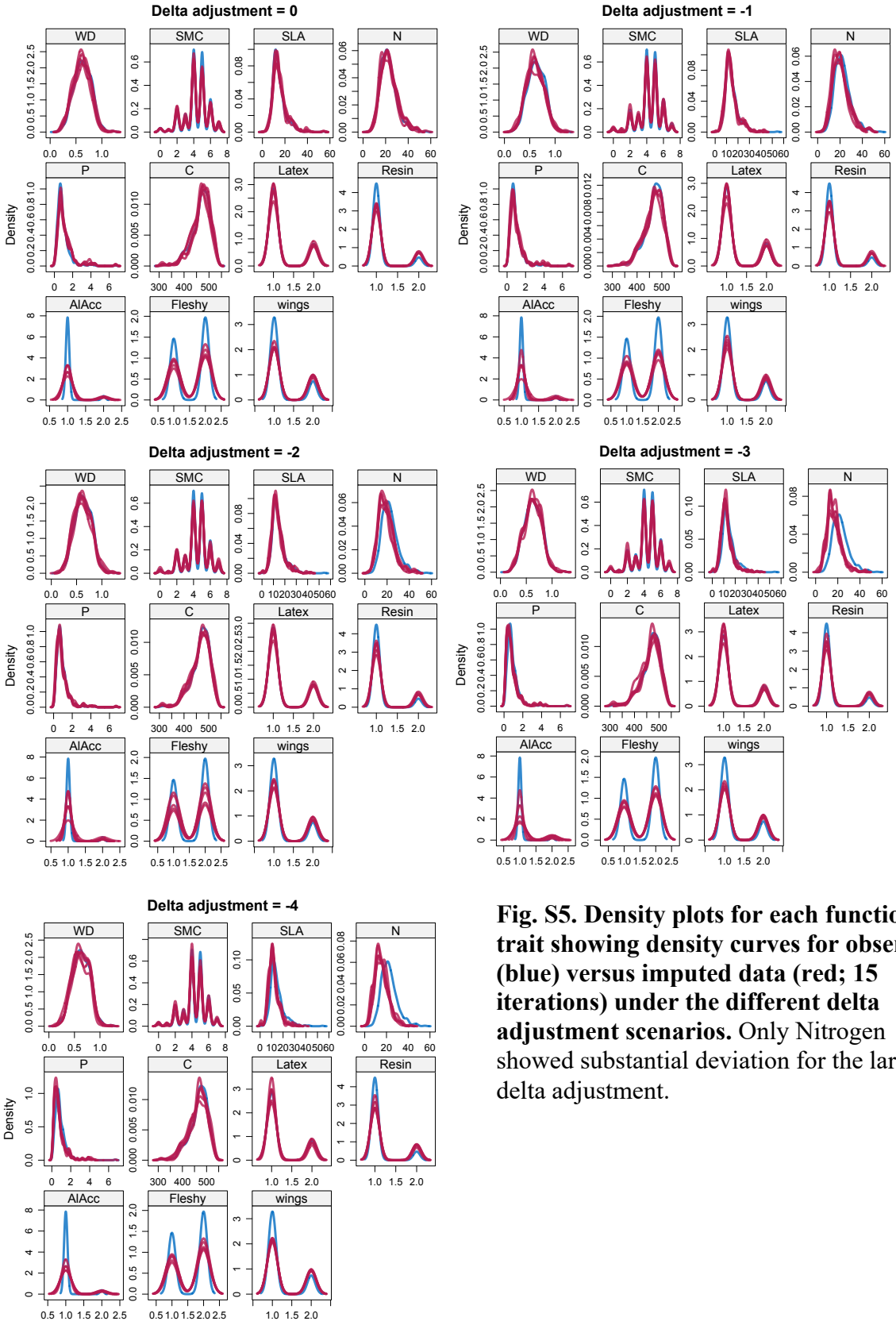

Fig. S5. Density plots for each functional trait showing density curves for observed (blue) versus imputed data (red; 15 iterations) under the different delta adjustment scenarios. Only Nitrogen showed substantial deviation for the larger delta adjustment.

FIGURE S6

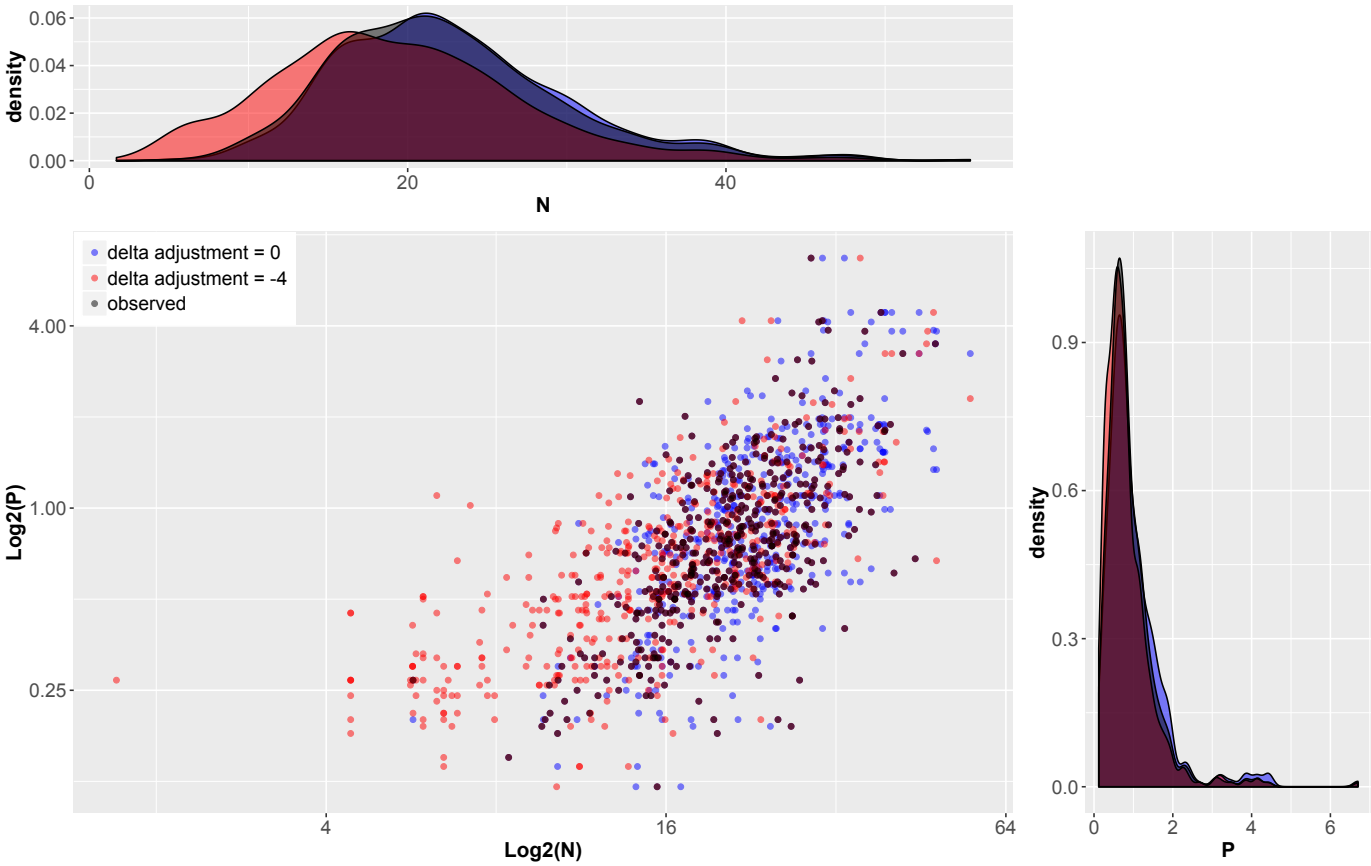

Fig. S6. Scatterplot of observed vs. imputed leaf nitrogen and phosphorus content under different delta adjustment scenarios. As Leaf Nitrogen content showed substantial deviation in the larger delta adjustment scenario it was here plotted versus Leaf Phosphorus content, which showed hardly any deviation.

FIGURE S7

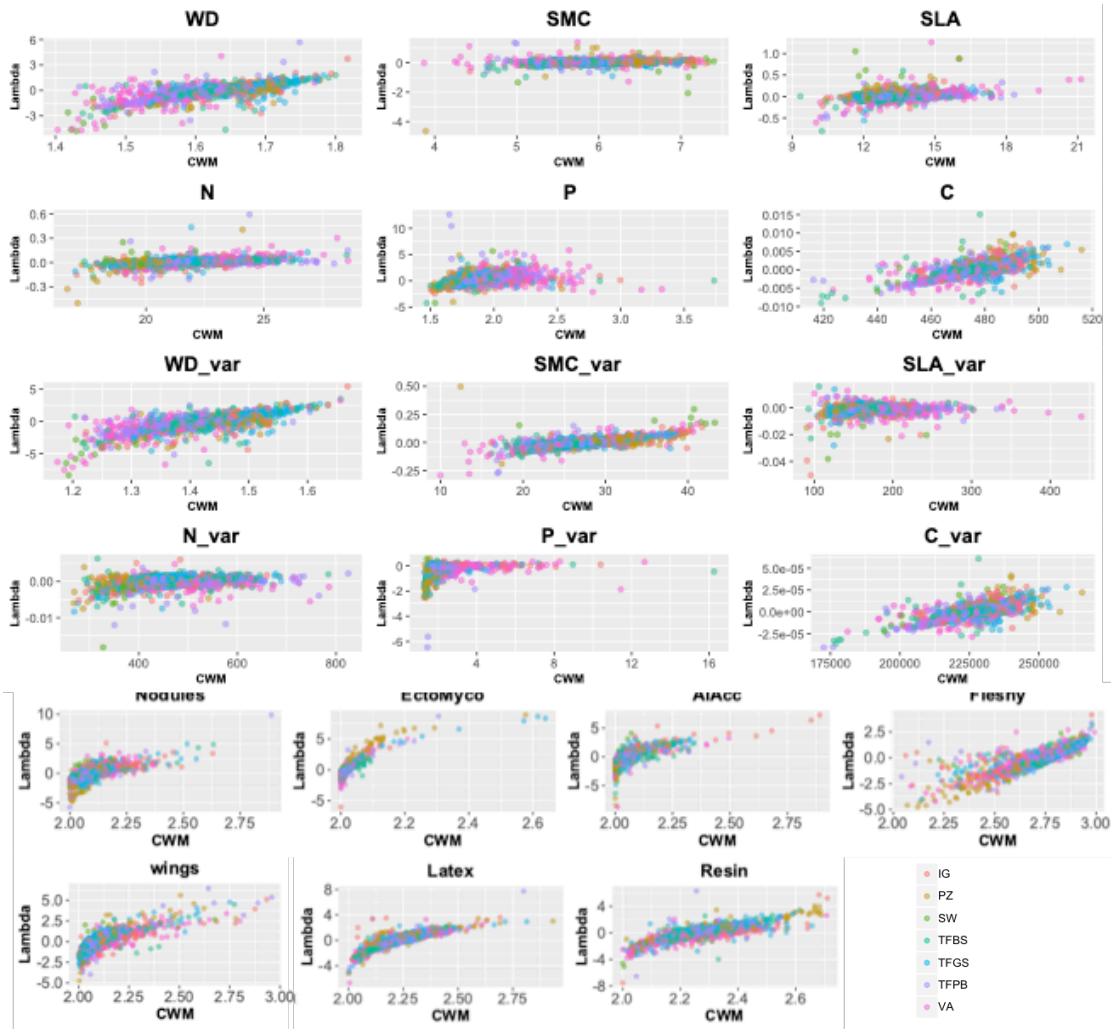

Fig. S7 Scatterplots of the CWM values against lambda values, colored according to forest type. Titles are abbreviations for functional traits as used in the main text. Plots show in some cases these are clearly correlated (e.g. wood density, seedmassclass and C) but for many others not (e.g. SLA, Latex or Nodules).

FIGURE S8

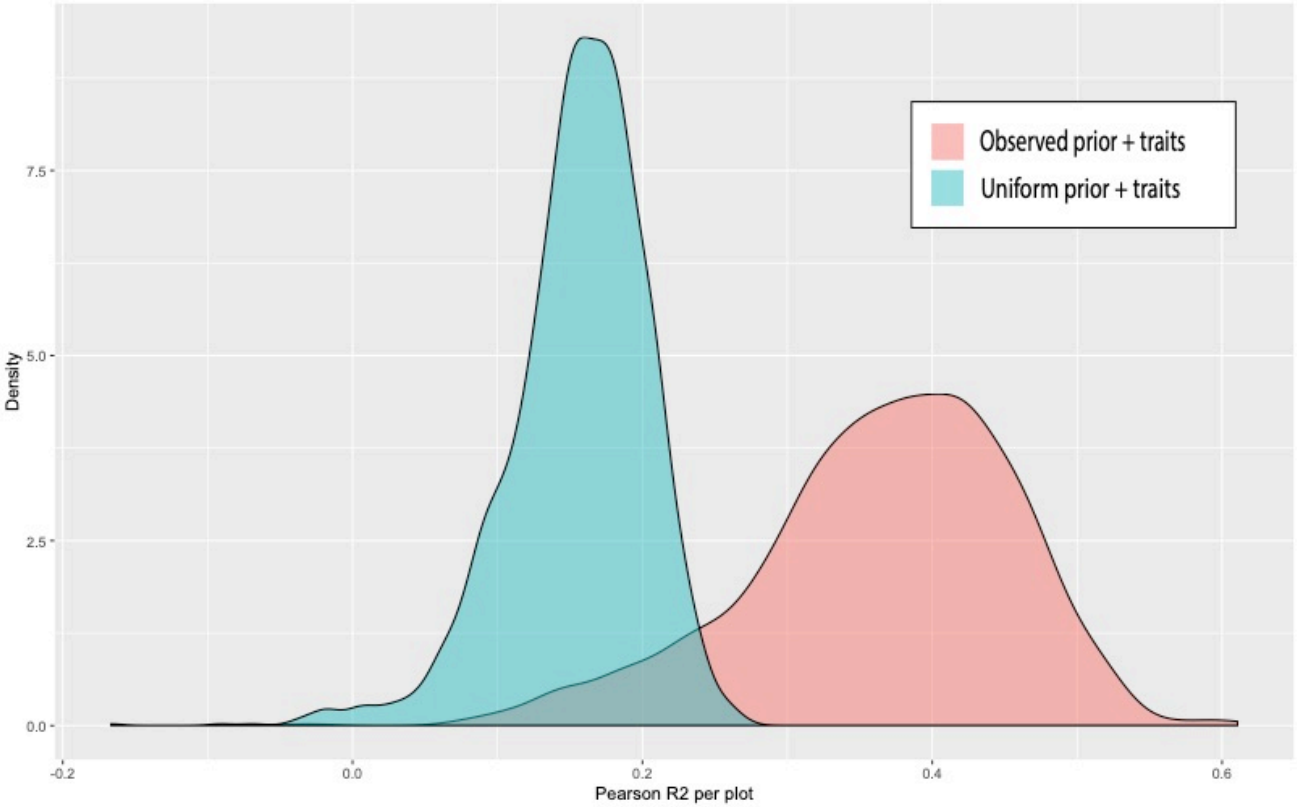

Fig. S8 Density plot of the per-plot-Pearson correlation coefficient between predicted relative abundances of each genus. Models either used a uniform prior and functional traits (blue) or the actual observed prior and functional traits (red), results show a large increase in accuracy for the latter.

FIGURE S9

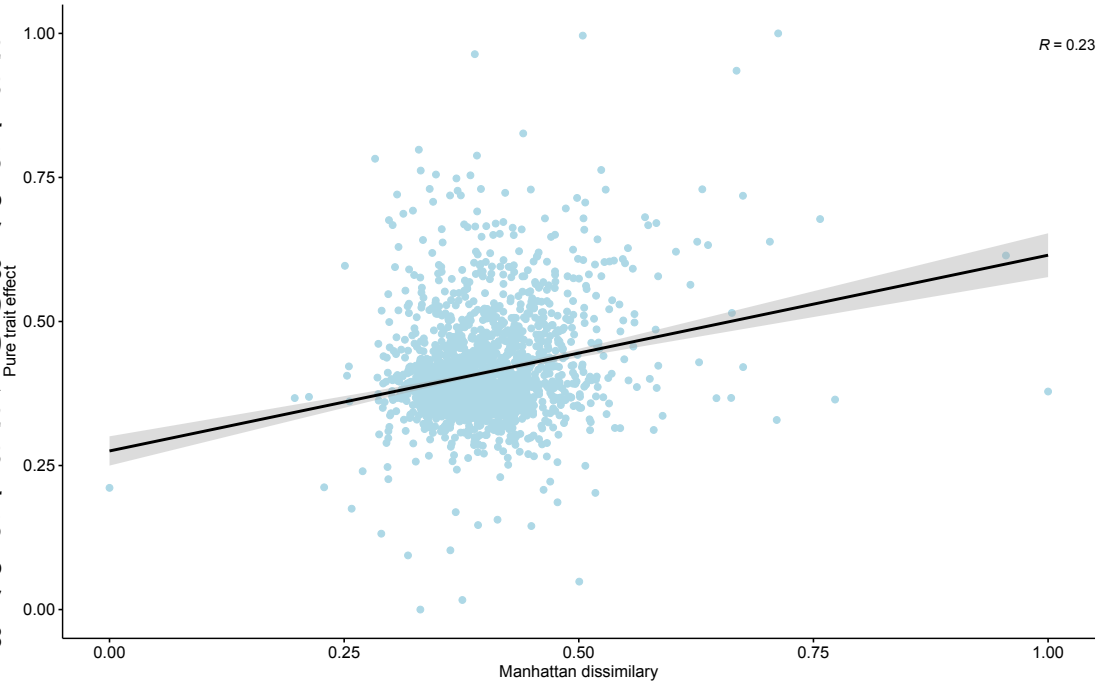
R = 0.23
1.00
0.75
0.50
0.25
0.00
Pure trait effect
0.00
0.25
0.50
0.75
1.00
Manhattan dissimilary

FIGURE S10

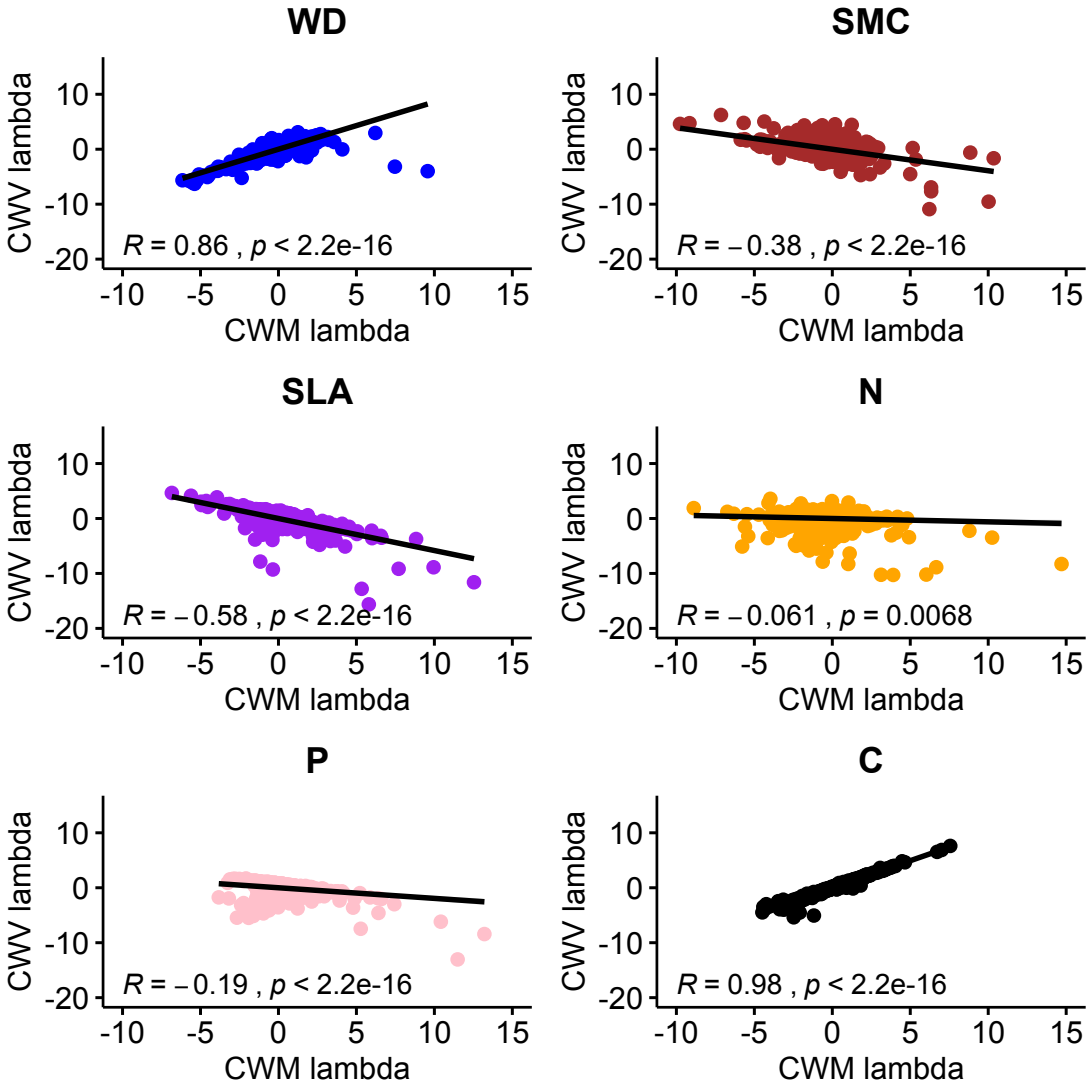
WD
SMC
SLA
N
P
C
CWV lambda
CWM lambda
R = 0.86 , p < 2.2e-16
R = – 0.38 , p < 2.2e-16
R = – 0.58 , p < 2.2e-16
R = – 0.061 , p = 0.0068
R = – 0.19 , p < 2.2e-16
R = 0.98 , p < 2.2e-16

TABLE S1

Table S1. Decomposition of results from the various maximum entropy models, combined and separated by forest type (PZ podzol, IG i*gapó*, VA *várzea*, SW Swamp, TF *terra firme* with subregions BS (Brazilian Shield), GS (Guyana Shield) and PB (Pebas formation). Top rows indicate estimated proportions (R^2_{KL}) of total information reflective of variation in local relative abundance explained for by the various models. Middle rows indicate the specific information gain from any one of the used models relative to the model bias. Bottom rows show the actual effects of traits, the metacommunity and the joint information relative to the model bias.

	Forest types							
Explained proportions	PZ	VA	IG	SW	TFBS	TFGS	TFPB	Combined
$\overline{R}^2_{KL}(u)$ *model bias fit*	0.20	0.14	0.15	0.23	0.11	0.11	0.09	0.12
$\overline{R}^2_{KL}(m)$ *pure neutral model fit*	0.54	0.53	0.54	0.57	0.56	0.63	0.56	0.58
$R^2_{KL}(u,t)$ *pure trait model fit*	0.36	0.25	0.23	0.34	0.20	0.24	0.21	0.23
$R^2_{KL}(m,t)$ *hybrid model fit*	0.60	0.59	0.56	0.60	0.60	0.66	0.60	0.62
Increase in explained deviance								
$\Lambda R^2_{KL}(m\|\varphi)$ *metacommunity effect beyond model bias*	0.34	0.39	0.39	0.34	0.45	0.52	0.47	0.45
$\Lambda R^2_{KL}(t\|\varphi)$ *trait effect beyond model bias*	0.15	0.11	0.08	0.11	0.10	0.13	0.11	0.11
$\Lambda R^2_{KL}(t\|m)$ *trait effect beyond metacommunity effect*	0.06	0.06	0.02	0.03	0.04	0.03	0.05	0.04
$\Lambda R^2_{KL}(m\|t)$ *metacommunity effect relative to trait effects*	0.24	0.34	0.33	0.27	0.40	0.43	0.40	0.38
$\Lambda R^2_{KL}(m+t)$ *joint effect of metacommunity and traits*	0.09	0.05	0.06	0.07	0.06	0.10	0.07	0.07
$1-\Lambda R^2_{KL}(m,t)$ *unexplained effects*	0.40	0.41	0.44	0.40	0.40	0.34	0.40	0.38
Biologically relevant information								
Pure trait effect *Information from traits, relative to bias*	0.08	0.07	0.03	0.05	0.05	0.04	0.05	0.05
Pure metacommunity effect *Information from metacommunity, relative to bias*	0.30	0.39	0.38	0.33	0.45	0.48	0.44	0.43
Joint effect *Information from joint effect, relative to bias*	0.12	0.06	0.07	0.10	0.06	0.11	0.07	0.08
Unexplained information *Left over information not explained, relative to bias*	0.50	0.48	0.52	0.52	0.45	0.38	0.44	0.44

TABLE S2

Table S2. Summary statistics overview for the linear models of the various scenarios under the delta adjustment technique as described in the main text. Rows indicate the different delta adjustments used with the columns representing the standard summary statistics of the linear model comparing the imputed versus observed trait values. Results showed similar patterns with each imputation scenario, indicating a robust imputation procedure.

	Summary statistics linear model lm(imputed ~ observed)						
Scenario	Intercept	Std. error	T value	Pr.	R^2	Adj. R^2	Sigma
Delta 0	-.33	.07	-4.85	1.42e-06	.32	.33	.63
Delta -2.5	-.34	.06	-5.93	4.31e-09	.37	.37	.58
Delta -5	-.16	.04	-3.57	3.77e-04	.40	.40	.53
Delta -7.5	.64	.09	-7.33	5.18e-13	.42	.42	.88
Delta -10	0.09	.04	3.17	1.57e-03	.47	.47	.48

S-A Ecological interpretation of the MEF results

A number of functional traits associated with low nutrient conditions (e.g. ectomycorrhiza) and life history strategies suited for protection against herbivores (e.g. latex and high leaf C content) were clearly positively associated with abundance in nutrient poor environments (podzols) in terms of community weighted mean values, indicated by the positive lambda values. In contrast, community weighted means for fleshy fruits and high leaf N and P content were negatively associated with abundance on these soils. Nodulation was also negatively associated with abundance on poor soils, supporting earlier results [1]. The ability to accumulate aluminium was positively associated with abundance on soils commonly associated with higher aluminium content such as *igapó* (strong positive effects) and *terra firme* soils (a minor, yet positive effect). In contrast, it was strongly negatively associated with abundance for podzol, *várzea* and swamp forests. Traits such as SLA or winged fruits also showed strong patterns dependent on forest type.

Signals of quantitative environmental selection were found to be highest for podzol forests, whereas its counterpart in the form of the dispersal mass effect from the regional pool of genera had the lowest value. Podzol forests, having extremely nutrient poor soils could reflect a much stronger selective environment than any of the other forest types. *Terra firme* forests, presumably reflective of a less strong selective environment in terms of resource availability, showed the opposite, with less than half of the pure trait effect in comparison with podzol forests (even when rarefied to accommodate for different sample sizes). Traits associated with protection against herbivores such as latex [2] and high leaf carbon content showed higher values associated with greater abundance and overall lower variance on podzol soils, whereas traits indicative of investment in growth and photosynthetic ability such as high foliar concentrations of P and N [3] showed strong negative associations on nutrient poor soils for both community weighted means and variance. The ability to accumulate aluminium was also strongly positively associated with relative abundance on the more nutrient but also often aluminium enriched soils of *terra firme* and in some cases aluminium rich *igapó* forests. Lambda values also showed strong negative lambda values for wood density in swamp and forests in both community weighted means and variance, fitting high tree mortality and many individuals belonging to pioneer species in especially the western Amazonian swamp forests. *Várzea* and Pebas *terra firme* forests showed a similar response. As the Pebas consists mainly of Andean sediments it has higher nutrient content, promoting lower wood density, supported by our results whereas *várzea* forests are also often flooded. There were also traits that showed no specific (strong) signal of selection on certain forest types (either positive or negative), such as latex on *igapó* and ectomycorrhiza on *várzea* (see Fig. 1 for all lambda values). Plotting lambda values for CWM and CWV constraints of the continuous traits showed WD and C were both strongly positively correlated indicating strong directional selection for lower trait values accompanied by a reduction in trait variance. SMC, SLA and leaf P content, however, showed a negative correlation with higher lambda values for CWM values associated with lower trait variance (Figure S11). None of the traits showed a reduction in variance without a change in the CWM, suggesting directional selection is more likely than stabilizing selection, even though the overall information yield remains low. Interestingly, terra firme forests in general showed the smallest lambda values overall (positive or negative).

This may be indicative of either more pronounced demographic stochasticity or ecological drift eliminating the association between traits and relative abundance. Lower effects of selection in general or more (random) variation due to the larger species pool in comparison with other forest types, however, could also be the result of mixing heterogeneous microenvironments into a single environmental class. Support for such heterogeneity within terra firme forests having influence on distribution of functional traits on valleys or plateaus has recently been found [4]. In addition, natural but also anthropogenic [5] disturbance history affects biotic community composition and can lead to changes in tree community through time, blurring relationships between traits and relative abundances. It should further be noted that, although for terra firme forests we were able to make a distinction by subregion, true within forest type heterogeneity was not taken into account. This might cause an underestimation of the deterministic effect but as of yet cannot be corrected for on this scale and is worth to be investigated in future studies. In addition, podzol forests have a smaller connected surface area and accompanying smaller number of genera in comparison with terra firme forests, adding to the calculated stronger trait effects [6,7]. When more detailed understanding and knowledge of these functional traits would be provided, this would most likely increase the explanatory power of the MEF. The fact, however, that we do not have a very specific knowledge of these interactions and specific traits is precisely the reason why the MEF can provide additional insight.

It should be noted that for species level analyses any micro environmental gradients might prove to also show (stronger) selection at local scales [8,9], as it has been shown that most variation in community composition, due to selection in regard to habitat filtering and niche conservatism, is found at lower taxonomic levels, such as between species within genera [10,11]. In contrast, theoretically it has been shown and tested that immigration numbers are actually very robust across taxonomic scales [12], validating our results of the metacommunity importance using genus level taxonomy. Spatial patterns of metacommunity effects, showing shallowest declines in the centre, also support the suggestion that high diversity of the Amazonian interior could be explained by influx of recruits due to large (overlapping) ranges. This mid-domain effect [13], however, would also predict lower species richness for the edges due to lower range overlap, assuming a closed community. This is not the case, as there is a strong species richness gradient from West (rich) to Eastern Amazonian forests (poor) [14]. The lower metacommunity effect for the edges then is most likely not due to less absolute influx of genera, but rather less influx from the Amazonian tree community. Influx from the species-rich Andes could account for the high diversity [15], yet low Amazonian metacommunity effect for Western Amazonian forests. In contrast, South Eastern parts of Amazonia receive influx from tree species-poor biomes (i.e. the Cerrado) resulting in lower diversity but also low metacommunity effect for Amazonian trees in this region.

REFERENCES SUPPLEMENTARY MATERIAL

[1] H. ter Steege, N. C. a Pitman, O. L. Phillips, J. Chave, D. Sabatier, A. Duque, J.-F. Molino, M.-F. Prévost, R. Spichiger, H. Castellanos, P. von Hildebrand, and R. Vásquez, *Continental-Scale Patterns of Canopy Tree Composition and Function across Amazonia.*, Nature **443**, 444 (2006).

[2] A. A. Agrawal, *Macroevolution of Plant Defense Strategies*, Trends Ecol. Evol. **22**, 103 (2007).

[3] J. C. Ordoñez, P. M. Van Bodegom, J. P. M. Witte, I. J. Wright, P. B. Reich, and R. Aerts, *A Global Study of Relationships between Leaf Traits, Climate and Soil Measures of Nutrient Fertility*, Glob. Ecol. Biogeogr. **18**, 137 (2009).

[4] R. S. Oliveira, F. R. C. Costa, E. van Baalen, A. de Jonge, P. R. Bittencourt, Y. Almanza, F. de V. Barros, E. C. Cordoba, M. V. Fagundes, S. Garcia, Z. T. T. M. Guimaraes, M. Hertel, J. Schietti, J. Rodrigues-Souza, and L. Poorter, *Embolism Resistance Drives the Distribution of Amazonian Rainforest Tree Species along Hydro-Topographic Gradients*, New Phytol. (2018).

[5] C. Levis, B. M. Flores, P. A. Moreira, B. G. Luize, R. P. Alves, J. Franco-Moraes, J. Lins, E. Konings, M. Peña-Claros, F. Bongers, F. R. C. Costa, and C. R. Clement, *How People Domesticated Amazonian Forests*, Front. Ecol. Evol. **5**, (2018).

[6] J. E. Guevara, G. Damasco, C. Baraloto, P. V. A. Fine, M. C. Peñuela, C. Castilho, A. Vincentini, D. Cárdenas, F. Wittmann, N. Targhetta, O. Phillips, J. Stropp, I. Amaral, P. Maas, A. Monteagudo, E. M. Jimenez, R. Thomas, R. Brienen, Á. Duque, W. Magnusson, C. Ferreira, E. Honorio, F. de Almeida Matos, F. R. Arevalo, J. Engel, P. Petronelli, R. Vasquez, and H. ter Steege, *Low Phylogenetic Beta Diversity and Geographic Neo-Endemism in Amazonian White-Sand Forests*, Biotropica **48**, 34 (2016).

[7] H. ter Steege, N. C. A. Pitman, D. Sabatier, C. Baraloto, R. P. Salomão, J. E. Guevara, O. L. Phillips, C. V Castilho, W. E. Magnusson, J.-F. Molino, A. Monteagudo, P. Núñez Vargas, J. C. Montero, T. R. Feldpausch, E. N. H. Coronado, T. J. Killeen, B. Mostacedo, R. Vasquez, R. L. Assis, J. Terborgh, F. Wittmann, A. Andrade, W. F. Laurance, S. G. W. Laurance, B. S. Marimon, B.-H. Marimon, I. C. Guimarães Vieira, I. L. Amaral, R. Brienen, H. Castellanos, D. Cárdenas López, J. F. Duivenvoorden, H. F. Mogollón, F. D. de A. Matos, N. Dávila, R. García-Villacorta, P. R. Stevenson Diaz, F. Costa, T. Emilio, C. Levis, J. Schietti, P. Souza, A. Alonso, F. Dallmeier, A. J. D. Montoya, M. T. Fernandez Piedade, A. Araujo-Murakami, L. Arroyo, R. Gribel, P. V. A. Fine, C. A. Peres, M. Toledo, G. A. Aymard C, T. R. Baker, C. Cerón, J. Engel, T. W. Henkel, P. Maas, P. Petronelli, J. Stropp, C. E. Zartman, D. Daly, D. Neill, M. Silveira, M. R. Paredes, J. Chave, D. de A. Lima Filho, P. M. Jørgensen, A. Fuentes, J. Schöngart, F. Cornejo Valverde, A. Di Fiore, E. M. Jimenez, M. C. Peñuela Mora, J. F. Phillips, G. Rivas, T. R. van Andel, P. von Hildebrand, B. Hoffman, E. L. Zent, Y. Malhi, A. Prieto, A. Rudas, A. R. Ruschell, N. Silva, V. Vos, S. Zent, A. A. Oliveira, A. C. Schutz, T. Gonzales, M. Trindade Nascimento, H. Ramirez-Angulo, R. Sierra, M. Tirado, M. N. Umaña Medina, G. van der Heijden, C. I. A. Vela, E. Vilanova Torre, C. Vriesendorp, O. Wang, K. R. Young, C. Baider, H. Balslev, C. Ferreira, I. Mesones, A. Torres-Lezama, L. E. Urrego Giraldo, R. Zagt, M. N. Alexiades, L. Hernandez, I. Huamantupa-Chuquimaco, W. Milliken, W. Palacios Cuenca, D. Pauletto, E. Valderrama Sandoval, L. Valenzuela Gamarra, K. G. Dexter, K.

Feeley, G. Lopez-Gonzalez, M. R. Silman, S. P. Hubbell, F. He, R. Condit, L. Borda-de-Agua, J. Kellner, H. Ter Steege, G. A. Black, T. H. Dobzhansky, C. Pavan, J. M. Pires, T. Dobzhansky, G. A. Black, M. J. G. Hopkins, M. J. Costello, R. M. May, N. E. Stork, P. Haripersaud, H. ter Steege, J.-J. de Granville, H. Chevillotte, M. Hoff, D. P. Bebber, M. A. Carine, J. R. Wood, A. H. Wortley, D. J. Harris, G. T. Prance, G. Davidse, J. Paige, T. D. Pennington, N. K. Robson, R. W. Scotland, B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, E. P. White, R. J. Warren, D. K. Skelly, O. J. Schmitz, M. A. Bradford, N. C. A. Pitman, J. W. Terborgh, M. R. Silman, P. N. V, D. A. Neill, C. E. Cerón, W. A. Palacios, M. Aulestia, N. C. A. Pitman, M. R. Silman, J. W. Terborgh, F. D. Lozano, M. W. Schwartz, M. W. Schwartz, D. Simberloff, J. E. Richardson, R. T. Pennington, T. D. Pennington, P. M. Hollingsworth, T. L. Couvreur, F. Forest, W. J. Baker, S. Cavers, C. W. Dick, D. H. Janzen, S. A. Mangan, S. A. Schnitzer, E. A. Herre, K. M. Mack, M. C. Valencia, E. I. Sanchez, J. D. Bever, W. Balée, D. G. Campbell, C. Levis, P. F. de Souza, J. Schietti, T. Emilio, J. L. P. V. Pinto, C. R. Clement, F. R. C. Costa, D. A. Posey, E. Montoya, V. Rull, N. D. Stansell, M. B. Abbott, S. Nogué, B. W. Bird, W. A. Díaz, C. Gomez-Navarro, C. Jaramillo, F. Herrera, S. L. Wing, R. Callejas, C. H. McMichael, D. R. Piperno, M. B. Bush, M. R. Silman, A. R. Zimmerman, M. F. Raczka, L. C. Lobato, H. ter Steege, P. P. Haripersaud, O. S. Bánki, F. Schieving, S. J. Phillips, R. P. Anderson, R. E. Schapire, S. J. Phillips, M. Dudik, C. A. Quesada, J. Lloyd, L. O. Anderson, N. M. Fyllas, M. Schwarz, C. I. Czimczik, B. Rollet, H. ter Steege, N. C. Pitman, O. L. Phillips, J. Chave, D. Sabatier, A. Duque, J. F. Molino, M. F. Prévost, R. Spichiger, H. Castellanos, P. von Hildebrand, R. Vásquez, P. M. Fearnside, D. Mouillot, D. R. Bellwood, C. Baraloto, J. Chave, R. Galzin, M. Harmelin-Vivien, M. Kulbicki, S. Lavergne, S. Lavorel, N. Mouquet, C. E. Paine, J. Renaud, W. Thuiller, G. Lopez-Gonzalez, S. L. Lewis, M. Burkitt, O. L. Phillips, P. J. M. Maas, L. Y. T. Westra, H. Rainer, A. Q. Lobão, R. H. J. Erkens, P. V. A. Fine, D. C. Daly, G. V. Muñoz, I. Mesones, K. M. Cameron, K. J. Feeley, M. R. Silman, M. Dufrene, P. Legendre, J. Bunge, M. Fitzpatrick, J. Bunge, L. Woodard, D. Böhning, J. A. Foster, S. Connolly, H. K. Allen, J.-P. Z. Wang, B. G. Lindsay, A. Chao, R. K. Colwell, C.-W. Lin, N. J. Gotelli, C. X. Mao, R. K. Colwell, U. Brose, N. D. Martinez, R. J. Williams, I. J. Good, A. Chao, S.-M. Lee, I. Rocchetti, J. Bunge, D. Bohning, R. A. Fisher, A. S. Corbet, C. B. Williams, F. W. Preston, J.-P. Wang, P. V. A. Fine, and R. H. Ree, *Hyperdominance in the Amazonian Tree Flora*, Science **342**, 1243092 (2013).

[8] L. H. M. Cosme, J. Schietti, F. R. C. Costa, and R. S. Oliveira, *The Importance of Hydraulic Architecture to the Distribution Patterns of Trees in a Central Amazonian Forest*, New Phytol. **215**, 113 (2017).

[9] S. E. Russo, S. J. Davies, D. A. King, and S. Tan, *Soil-Related Performance Variation and Distributions of Tree Species in a Bornean Rain Forest*, J. Ecol. **93**, 879 (2005).

[10] K. J. Gaston, *Species-Range Size Distributions: Products of Speciation, Extinction and Transformation: Philosophical Transactions of the Royal Society of London Series B*, **353**, 219 (1998).

[11] R. E. Ricklefs, H. Qian, and P. S. White, *The Region Effect on Mesoscale Plant*

Species Richness between Eastern Asia and Eastern North America, **2**, 129 (2004).

[12] F. Munoz, B. R. Ramesh, and P. Couteron, *How Do Habitat Filtering and Niche Conservatism Affect Community Composition at Different Taxonomic Resolutions?*, Ecology **95**, 2179 (2014).

[13] R. K. Colwell, C. Rahbek, and N. J. Gotelli, *The Mid-Domain Effect and Species Richness Patterns:What Have We Learned so Far?*, Am. Nat. **163**, E1 (2004).

[14] H. ter Steege, N. Pitman, D. Sabatier, H. Castellanos, P. Van der Hout, D. C. Daly, M. Silveira, O. Phillips, R. Thomas, J. V. A. N. Essen, H. Mogollon, and W. Morawetz, *A Spatial Model of Tree a -Diversity and Tree Density for the Amazon*, Biodivers. Conserv. **12**, 2255 (2003).

[15] T. F. Rangel, N. R. Edwards, P. B. Holden, J. A. F. Diniz-Filho, W. D. Gosling, M. T. P. Coelho, F. A. S. Cassemiro, C. Rahbek, and R. K. Colwell, *Modeling the Ecology and Evolution of Biodiversity: Biogeographical Cradles, Museums, and Graves. Science*, Science (80-.). **in press**, (2018).

[16] B. Shipley, *Measuring and Interpreting Trait-Based Selection versus Meta-Community Effects during Local Community Assembly*, J. Veg. Sci. **25**, 55 (2014).

[17] B. Shipley, C. E. T. Paine, and C. Baraloto, *Quantifying the Importance of Local Niche-Based and Stochastic Processes to Tropical Tree Community Assembly*, Ecology **93**, 760 (2012).

[18] A. Colin Cameron and F. A. G. Windmeijer, *An R-Squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models*, J. Econom. **77**, 329 (1997).