Title:

Global trait–environment relationships of plant communities

One Sentence Summary: Trait composition of plant communities across the globe is captured by two main dimensions and is probably shaped by environmental or biotic filtering, but is only weakly related to global climate and soil gradients.

Authors:

Helge Bruelheide[1,2], Jürgen Dengler[2,3,4], Oliver Purschke[1,2], Jonathan Lenoir[5], Borja Jiménez-Alfaro[6,1,2], Stephan M. Hennekens[7], Zoltán Botta-Dukát[8], Milan Chytrý[9], Richard Field[10], Florian Jansen[11], Jens Kattge[2,12], Valério D. Pillar[13], Franziska Schrodt[10,12], Miguel D. Mahecha[2,12], Robert K. Peet[14], Brody Sandel[15], Peter van Bodegom[16], Jan Altman[17], Esteban Alvarez Davila[18], Mohammed A.S. Arfin Khan[19,20], Fabio Attorre[21], Isabelle Aubin[22], Christopher Baraloto[23], Jorcely G. Barroso[24], Marijn Bauters[25], Erwin Bergmeier[26], Idoia Biurrun[27], Anne D. Bjorkman[28], Benjamin Blonder[29,30], Andraž Čarni[31,32], Luis Cayuela[33], Tomáš Černý[34], J. Hans C. Cornelissen[35], Dylan Craven[2,36], Matteo Dainese[37], Géraldine Derroire[38], Michele De Sanctis[21], Sandra Díaz[39], Jiří Doležal[17], William Farfan-Rios[40,41], Ted R. Feldpausch[42], Nicole J. Fenton[43], Eric Garnier[44], Greg R. Guerin[45], Alvaro G. Gutiérrez[46], Sylvia Haider[1,2], Tarek Hattab[47], Greg Henry[48], Bruno Hérault[49,50], Pedro Higuchi[51], Norbert Hölzel[52], Jürgen Homeier[53], Anke Jentsch[20], Norbert Jürgens[54], Zygmunt Kącki[55], Dirk N. Karger[56,57], Michael Kessler[56], Michael Kleyer[58], Ilona Knollová[9], Andrey Y. Korolyuk[59], Ingolf Kühn[36,1,2], Daniel C. Laughlin[60,61], Frederic Lens[62], Jacqueline Loos[63], Frédérique Louault[64], Mariyana I. Lyubenova[65], Yadvinder Malhi[66], Corrado Marcenò[27], Maurizio Mencuccini[67,68], Jonas V. Müller[69], Jérôme Munzinger[70], Isla H. Myers-Smith[71], David A. Neill[72], Ülo Niinemets[73], Kate H. Orwin[74], Wim A. Ozinga[7,75], Josep Penuelas[68,73,76], Aaron Pérez-Haase[77,78], Petr Petřík[17], Oliver L. Phillips[79], Meelis Pärtel[80], Peter B. Reich[81,82], Christine Römermann[2,83], Arthur V. Rodrigues[84], Francesco Maria Sabatini[1,2], Jordi Sardans[68,76], Marco Schmidt[85], Gunnar Seidler[1], Javier Eduardo Silva Espejo[86], Marcos Silveira[87], Anita Smyth[45], Maria Sporbert[1,2], Jens-Christian Svenning[28], Zhiyao Tang[88], Raquel Thomas[89], Ioannis Tsiripidis[90], Kiril Vassilev[91], Cyrille Violle[44], Risto Virtanen[2,92,93], Evan Weiher[94], Erik Welk[1,2], Karsten Wesche[2,95,96], Marten Winter[2], Christian Wirth[2,12,97], Ute Jandt[1,2]

Affiliations:

[1] Martin Luther University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden, Am Kirchtor 1, 06108 Halle, Germany

[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

[3] Zurich University of Applied Sciences (ZHAW), Institute of Natural Resource Sciences (IUNR), Research Group Vegetation Ecology, Grüentalstr. 14, Postfach, 8820 Wädenswil, Switzerland
[4] University of Bayreuth, Bayreuth Center of Ecology and Environmental Research BayCEER, Plant Ecology, Universitätsstr. 30, 95447 Bayreuth, Germany
[5] CNRS, Université de Picardie Jules Verne, UR «Ecologie et Dynamique des Systèmes Anthropisés» (EDYSAN, UMR 7058 CNRS-UPJV), 1 rue des Louvels, 80037 Amiens Cedex 1, France
[6] Research Unit of Biodiversity (CSIC/UO/PA), University of Oviedo, Campus de Mieres, c/ Gonzalo Gutiérrez Quirós s/n 33600 Mieres, Spain
[7] Wageningen Environmental Research (Alterra), Team Vegetation, Forest and Landscape Ecology, PO Box 47, 6700 AA Wageningen, The Netherlands
[8] MTA Centre for Ecological Research, GINOP Sustainable Ecosystems Group, 8237 Tihany, Klebesberg Kuno u. 3, Hungary
[9] Masaryk University, Department of Botany and Zoology, Kotlářská 2, 611 37 Brno, Czech Republic
[10] University of Nottingham, School of Geography, University Park, Nottingham, NG7 2RD, United Kingdom
[11] University of Rostock, Faculty for Agricultural and Environmental Sciences, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
[12] Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Gemany
[13] Universidade Federal do Rio Grande do Sul, Department of Ecology, Porto Alegre, RS, 91501-970, Brazil
[14] University of North Carolina at Chapel Hill, Department of Biology, Chapel Hill, NC 27599-3280, USA
[15] Santa Clara University, Department of Biology, Santa Clara CA 95053, USA
[16] Leiden University, Institute of Environmental Sciences, Department Conservation Biology, Einsteinweg 2, 2333 CC Leiden, The Netherlands
[17] Institute of Botany of the Czech Academy of Sciences, Zámek 1, 252 43 Průhonice, Czech Republic
[18] Escuela de Ciencias Agropecuarias y Ambientales – ECAPMA, Universidad Nacional Abierta y a Distancia –UNAD, Sede José Celestino Mutis, Calle 14S #14, Bogotá
[19] Shahjalal University of Science and Technology, Department of Forestry and Environmental Science, Sylhet, 3114, Bangladesh

[20] University of Bayreuth, Bayreuth Center of Ecology and Environmental Research BayCEER, Department of Disturbance Ecology, Universitätsstr. 30, 95447 Bayreuth, Germany
[21] Sapienza University of Rome, Department of Environmental Biology, P.le Aldo Moro 5, 00185 Rome, Italy
[22] Great Lakes Forestry Centre, Canadian Forest Service, Natural Resources Canada, 1219 Queen St. East, Sault Ste Marie, ON, P6A 2E5, Canada
[23] Florida International University, Department of Biological Sciences, International Center for Tropical Botany (ICTB), 11200 SW 8th Street, OE 243 Miami, FL 33199, USA
[24] Universidade Federal do Acre, Campus de Cruzeiro do Sul, Acre, Brazil.
[25] Ghent University, Faculty of Bioscience Engineering, Department of Applied Analytical and Physical Chemistry, ISOFYS, Coupure Links 653, 9000 Gent, Belgium
[26] University of Göttingen, Albrecht von Haller Institute of Plant Sciences, Vegetation Analysis & Plant Diversity, Untere Karspüle 2, 37073 Göttingen, Germany
[27] University of the Basque Country UPV/EHU, Apdo. 644, 48080 Bilbao, Spain
[28] Aarhus University, Department of Bioscience, Section for Ecoinformatics & Biodiversity, Ny Munkegade 114, 8000 Aarhus C, Denmark
[29] University of Oxford, Environmental Change Institute, School of Geography and the Environment, South Parks Road, Oxford OX1 3QY, United Kingdom
[30] Rocky Mountain Biological Laboratory, PO Box 519, Crested Butte, Colorado, 81224 USA
[31] Scientific Research Center of the Slovenian Academy of Sciences and Arts, Institute of Biology, Novi trg 2, SI 1001 Ljubljana p. box 306, Slovenia
[32] University of Nova Gorica, 5000 Nova Gorica, Slovenia
[33] Universidad Rey Juan Carlos, Department of Biology, Geology, Physics and Inorganic Chemistry, c/ Tulipán s/n, 28933 Móstoles, Madrid, Spain
[34] Czech University of Life Sciences, Faculty of Forestry and Wood Science, Department of Forest Ecology, Kamýcká 1176, 16521, Praha 6 – Suchdol, Czech Republic
[35] Vrije Universiteit Amsterdam, Faculty of Science, Department of Ecological Science, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands
[36] Helmholtz Centre for Environmental Research - UFZ, Dept. Community Ecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany
[37] University of Würzburg, Department of Animal Ecology and Tropical Biology, Am Hubland, 97074 Würzburg, Germany
[38] Cirad, UMR EcoFoG, Campus Agronomique, 97310 Kourou, French Guiana

[39]
Instituto Multidisciplinario de Biología Vegetal (IMBIV), CONICET and FCEFyN, Universidad Nacional de Córdoba, Casilla de Correo 495, 5000 Córdoba, Argentina
[40] Wake Forest University, Department of Biology, Winston Salem, North Carolina, USA
[41] Universidad Nacional de San Antonio Abad del Cusco, Herbario Vargas (CUZ), Cusco, Peru
[42] University of Exeter, College of Life and Environmental Sciences, Geography, Exeter, EX4 4RJ, United Kingdom.
[43] Université du Québec en Abitibi-Témiscamingue, Institut de recherche sur les forêts, 445 Boul. de l'Université, Rouyn-Noranda, Qc J9X 4E5, Canada
[44] CNRS - Université de Montpellier - Université Paul-Valéry Montpellier - EPHE, Centre d'Ecologie Fonctionnelle et Evolutive (UMR5175), 34293 Montpellier Cedex 5, France
[45] University of Adelaide, Terrestrial Ecosystem Research Network, School of Biological Sciences, Adelaide, South Australia, 5005 Australia
[46] Universidad de Chile, Facultad de Ciencias Agronómicas, Departamento de Ciencias Ambientales y Recursos Naturales Renovables, Av. Santa Rosa 11315, La Pintana 8820808, Santiago, Chile
[47] IFREMER (Institut Français de Recherche pour l'Exploitation de la MER) UMR 248 MARBEC (CNRS, IFREMER, IRD, UM), 34203 Sète cedex, France
[48] University of British Columbia, The Department of Geography, 1984 West Mall, Vancouver, BC V6T 1Z2, Canada
[49] INPHB (Institut National Polytechnique Félix Houphouët-Boigny), BP 1093, Yamoussoukro, Ivory Coast
[50] Cirad, University Montpellier, UR Forests & Societies, Montpellier, France
[51] Universidade do Estado de Santa Catarina (UDESC), Departamento de Engenharia Florestal, Av Luiz de Camões, 2090 - Conta Dinheiro, Lages – SC, 88.520 – 000, Brazil
[52] University of Münster, Institute of Landscape Ecology, Heisenbergstr. 2, 48149 Münster, Germany
[53] University of Göttingen, Plant Ecology and Ecosystems Research, Untere Karspüle 2, 37073 Göttingen, Germany
[54] University of Hamburg, Biodiversity, Biocenter Klein Flottbek and Botanical Garden, Ohnhorststr. 18, 22609 Hamburg, Germany
[55] University of Wroclaw, Institute of Environmental Biology, Department of Vegetation Ecology, Przybyszewskiego 63, 51-148 Wrocław, Poland

[56] University of Zurich, Department of Systematic and Evolutionary Botany, Zollikerstrasse 107, 8008 Zurich, Switzerland
[57] Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland.
[58] University of Oldenburg, Institute of Biology and Environmental Sciences, Landscape Ecology Group, Carl-von-Ossietzky Strasse 9-11, 26111 Oldenburg, Germany
[59] Central Siberian Botanical Garden SB RAS, Zolotodolinskaya str. 101, Novosibirsk, 630090, Russia
[60] University of Waikato, Environmental Research Institute, School of Science, Private Bag 3105, Hamilton 3240, New Zealand
[61] University of Wyoming, Department of Botany, Laramie, Wyoming, USA
[62] Leiden University, Naturalis Biodiversity Center, P.O. Box 9517, 2300RA Leiden, The Netherlands
[63] Agroecology, University of Göttingen, Grisebachstrasse 6, 37077 Göttingen, Germany
[64] INRA, VetAgro Sup, UMR Ecosystème Prairial, 63000 Clermont-Ferrand, France
[65] University of Sofia, Faculty of Biology, Department of Ecology and Environmental Protection, 1164 Sofia, 8 Dragan Tsankov Av., Bulgaria
[66] University of Oxford, Environmental Change Institute, School of Geography and the Environment, South Parks Road, Oxford, OX1 3QY, United Kingdom
[67] ICREA Pg. Lluís Companys 23, 08010 Barcelona, Spain
[68] CREAF, Cerdanyola del Vallès, 08193 Barcelona, Catalonia, Spain
[69] Royal Botanic Gardens Kew, Millennium Seed Bank, Conservation Science, Wakehurst Place, Ardingly, RH17 6TN, United Kingdom
[70] AMAP, IRD, CNRS, INRA, Université Montpellier, 34000 Montpellier, France
[71] University of Edinburgh, School of GeoSciences, Edinburgh EH9 3FF, United Kingdom
[72] Universidad Estatal Amazónica, Conservación y Manejo de Vida Silvestre, Paso lateral km 2½ via a Napo Puyo, Pastaza, Ecuador
[73] Estonian University of Life Science, Department of Crop Science and Plant Biology, Kreutzwaldi 1, 51014, Tartu, Estonia
[74] Landcare Research, PO Box 69040, Lincoln 7640, New Zealand
[75] Radboud University Nijmegen, Institute for Water and Wetland Research, 6500 GL Nijmegen, The Netherlands
[76] CSIC, Global Ecology Unit, CREAF-CEAB-UAB, Cerdanyola del Vallès, 08193 Barcelona, Catalonia, Spain

[77] University of Barcelona, Faculty of Biology, Department of Evolutionary Biology, Ecology and Environmental Sciences, Av. Diagonal 643, 08028, Barcelona, Spain
[78] Center for Advanced Studies of Blanes, Spanish Research Council (CEAB-CSIC), 17300 Blanes, Spain
[79] University of Leeds, School of Geography, Leeds LS2 9JT, United Kingdom
[80] University of Tartu, Ülikooli 18, 50090 Tartu, Estonia
[81] University of Minnesota, Department of Forest Recourses, 220F Green Hall, 1530 Cleveland Avenue North, St. Paul, MN 55108, USA
[82] Western Sydney University, Hawkesbury Institute for the Environment, New South Wales 2751, Australia
[83] Friedrich Schiller University Jena, Institute of Systematic Botany, Philosophenweg 16, 07743 Jena, Germany
[84] Universidade Regional de Blumenau, Departamento de Engenharia Florestal, Rua São Paulo, 3250, 89030-000 Blumenau-Santa Catarina, Brazil
[85] Senckenberg Biodiversity and Climate Research Centre (BiK-F), Data and Modelling Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[86] University of La Serena , Department of Biology, La Serena, Chile
[87] Universidade Federal do Acre, Museu Universitário / Centro de Ciências Biológicas e da Natureza / Laboratório de Botânica e Ecologia Vegetal, BR 364, Km 04 - Distrito Industrial - 69915-559 - Rio Branco-AC, Brazil
[88] Peking University, College of Urban and Environmental Sciences, Yiheyuan Rd. 5, 100871, Beijing, China
[89] Iwokrama International Centre for Rain Forest Conservation and Development, 77 High Street, Kingston, Georgetown, Guyana
[90] Aristotle University of Thessaloniki, School of Biology, Department of Botany, 54124 Greece
[91] Bulgarian Academy of Sciences, Institute of Biodiversity and Ecosystem Research, 23 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
[92] Helmholtz Center for Environmental Research – UFZ, Department of Physiological Diversity, Permoserstraße 15, Leipzig 04318, Germany
[93] University of Oulu, Department of Ecology & Genetics, PO Box 3000, FI-90014, Finland
[94] University of Wisconsin - Eau Claire, Department of Biology, Eau Claire, WI 54702-4004, USA

[95] Senckenberg Museum of Natural History Görlitz, P.O. Box 300154, 02806 Görlitz, Germany
[96] TU Dresden, International Institute (IHI) Zittau, Markt 23, 02763 Zittau, Germany
[97] University of Leipzig, Systematic Botany and Functional Biodiversity, Johannisallee 21–23, 04103 Leipzig, Germany

Abstract:

Plant functional traits directly affect ecosystem functions and are fundamental for managing and predicting biodiversity and ecosystem change. Globally, at the species level, plant trait combinations depend on key trade-offs representing different ecological strategies[1], but at the community level trait combinations are expected to be decoupled from these trade-offs because different strategies can facilitate co-existence within communities[2]. A key remaining question is to what extent community-level trait composition is globally filtered and how well it is related to global environmental drivers (macroclimate[3-5], coarse-scale soil properties[6,7]) as compared to local factors (microclimate, fine-scale soil properties, disturbance regime[8], successional dynamics[9]). Here, we perform the first global, plot-level analysis of trait-environment relationships, using a novel database with more than 1.1 million vegetation plots and 26,632 plant species with trait information. Although we found a strong filtering of 17 functional traits, similar climate and soil conditions support communities differing greatly in mean trait values. The two main community trait axes (plant stature and resource acquisitiveness), which capture half of the global trait variation reflect the trade-offs at the species level. However, those axes are weakly associated with climate and soil conditions at the global scale. Similarly, within-plot trait variation does not vary systematically with macro-environment. Beyond the two main trait dimensions, we found a strong correlation between leaf N:P ratio and growing-season warmth, indicating increasing phosphorus limitation towards the tropics. Our results indicate that, at fine spatial grain, macro-environmental drivers are much less important for functional trait composition than has hitherto been assumed from floristic analyses restricted to co-occurrence in large grid cells. Instead, trait combinations seem to be predominantly filtered by local-scale factors such as disturbance, fine-scale soil conditions, niche partitioning or biotic interactions.

Introduction

How climate drives the functional characteristics of vegetation across the globe has been a key question in ecological research for more than a century[10]. While functional information is available for a large portion of the global pool of plant species, we do not know how functional traits of the different species that co-occur in a community are combined, which is what determines their joint effect on ecosystems[6,9,11]. At the species level, Díaz et al.[1] demonstrated that 74% of the global spectrum of six key plant traits determining plant fitness in terms of survival, growth and reproduction can be accounted for by two principal components (PCs). They showed that the functional space occupied by vascular plant species is strongly constrained by trade-offs between traits and converges on a small set of successful trait combinations, confirming previous findings[7,12-14]. While these constraints describe evolutionarily viable ecological strategies for vascular plant species globally, they provide only limited insight into trait composition within communities. There are many reasons why trait composition within communities would produce very different patterns, and indeed much theory predicts this[2,8]. As ecosystem functions and services are ultimately dependent on the traits of the species composing ecological communities, exploring community trait

composition at the global scale can advance our understanding of how climate change and other anthropogenic drivers affect ecosystem functioning.

So far, studies relating trait composition to the environment at continental to global extents have been restricted to coarse-grained species occurrence data (e.g. presence in 1° grid cells[15-17]). Such data capture neither biotic interactions (co-occurrence in large grid cells does not indicate local co-existence), nor local variation in environmental filters (e.g. variation in soil, topography or disturbance regime within grid cells). In contrast, functional composition of ecological communities sampled at fine-grained vegetation plots – with areas of few to a few hundred square meters – is the direct outcome of the interaction between both local and large-scale factors. Here, we present the first global analysis of plot-level trait composition. We combined the 'sPlot' database, a new global initiative incorporating more than 1.1 million vegetation plots from over 100 databases (mainly forests and grasslands; see Methods), with 30 large-scale environmental variables and 18 key plant functional traits derived from TRY, a global plant-trait database (see Methods, Table 2). We selected these 18 traits because they affect different key ecosystem processes and are expected to respond to macroclimatic drivers (Table 1). In addition, they were sufficiently measured across all species globally to allow for imputation of missing values (see Methods). All analyses were confined to vascular plant species and included all vegetation layers in a community, from the canopy to the herb layer (see Methods).

We used this unprecedented fine-resolution dataset to test the hypothesis (Hypothesis 1) that plant communities show evidence of environmental or biotic filtering at the global scale, making use of the observed variation of plot-level trait means and means of within-plot trait variation across communities. Ecological theory suggests that community-level convergence could be interpreted as the result of filtering processes, including environmental filtering and biotic interactions. Globally, temperature and precipitation drive the differences in vegetation between biomes, suggesting strong environmental filtering[6,8] that constrains the number of successful trait combinations and leads to community-level trait convergence. Similarly, biotic interactions may eliminate excessively divergent trait combinations[18,19]. However, alternative functional trait combinations may confer equal fitness in the same environment[2]. If plant communities show a global variation of plot-level trait means higher than expected by chance, and a lower than expected within-plot trait variation (see Figure 1), this would support the view that environmental or biotic filtering are dominant structuring processes of community trait composition at the global scale. A consequence of strong community-level trait convergence, and thus low variation within plots with species trait values centred around the mean, would be that plot-level means will be similar to the trait values of the species in that plot. Hence, community mean trait values should then mirror the trait values of individual species[1].

While Hypothesis 1 addresses the degree of filtering, it does not make a statement on the attribution of driving factors. The main drivers should correlate strongly (though not necessarily linearly[20]) with plot-level trait means and within-plot trait variance. Identifying these drivers has the potential to fundamentally improve our understanding of global trait-environment relationships. We tested the hypothesis (Hypothesis 2) that there are strong correlations between global environmental drivers such as macroclimate and coarse-scale soil

properties and both plot-level trait means and within-plot trait variances[3-7,15-17,20-24] (see Table 1 for expected relationships and Extended Data Table 2 for variables used). Such evidence, although correlative, may contribute to the formulation of novel hypotheses to explain global plant trait patterns.

Results and Discussion

Consistent with Hypothesis 1 and as illustrated in Figure 1, global variation in plot-level trait means was much higher than expected by chance: all traits had positive standardized effect sizes (SESs), which were significantly > 0 for 17 out of 18 traits based on gap-filled data (mean SES = 8.06 standard deviations (SD), Table 2). This suggests that environmental or biotic filtering is a dominant force of community trait composition globally. Also as predicted by Hypothesis 1, within-plot trait variance was typically lower than expected by chance (mean SES = -1.76 SD, significantly < 0 for ten traits but significantly > 0 for three traits; Table 2). Thus, trait variation within communities may also be constrained by filtering.

Trait correlations at the community level were relatively well captured by the first two axes of a Principal Component Analysis (PCA) for both plot-level trait means and within-plot trait variances (Figures 1 and 2). The dominant axes were determined by those traits with the highest absolute SESs of plot-level trait mean trait values (Table 2, mean of CWMs). The PCA of plot-level trait means (Fig. 2) reflects two main functional continua on which community trait values converge: one from short-stature, small-seeded communities such as grasslands or herbaceous vegetation to tall-stature communities with large, heavy diaspores such as forests (the size spectrum), and the other from communities with resource-acquisitive to those with resource-conservative leaves (i.e. the leaf economics spectrum)[12]. The high similarity between this PCA and the one at the species level by Díaz et al.[1] is striking: here at the community level, based on 1.1 million plots, the same functional continua emerged as at the species level, based on 2,214 species. While the trade-offs between different traits at the species level can be understood from a physiological and evolutionary perspective, finding similar trade-offs between traits at the community level was unexpected, as species with opposing trait values can co-exist in the same community. In combination with our finding of strong trait convergence, these results reveal a strong parallel of present-day community assembly to individual species' evolutionary histories.

Surprisingly, we found only limited support for Hypothesis 2. Community-level trait composition was poorly captured by global climate and soil variables. None of the 30 environmental variables accounted individually for more than 10% of the variance in the traits defining the main dimensions in Fig. 2 (Extended Data Fig. 2). The coefficients of determination were not improved when testing for non-linear relationships (see Methods). Using all 30 environmental variables simultaneously as predictors only accounted for 10.8% or 14.0% of the overall variation in plot-level trait means (cumulative variance, respectively, of the first two or all 18 constrained axes in a Redundancy Analysis). Overall, our results show that similar global-scale climate and soil conditions can support communities that differ

markedly in mean trait values and that different climates can support communities with rather similar mean trait values.

The ordination of within-plot variance of the different traits (Fig. 3) revealed two main continua. Variances of plant height and diaspore mass varied largely independently of variances of traits representing the leaf economics spectrum. This suggests that short and tall species can be assembled together in the same community independently from combining species with acquisitive leaves with species with conservative leaves. Global climate and soil variables accounted for even less variation on the first two PCA axes in within-plot trait variances than on the first two PCA axes in plot-level trait means. Only two environmental variables had $r^2 > 3\%$ (Extended Data Fig. 3), whether allowing for non-linear relationships (see Methods) or not, and overall, macro-environment accounted for only 3.6% or 5.0% of the variation (cumulative variance, respectively, of the first two or all 18 constrained axes). Removing species richness effects from within-plot trait variances did not increase the amount of variation explained by the environment (see Methods).

The findings of our study contrast strongly with studies where the variation in traits between species was calculated at the level of the species pool in large grid cells[15,16], suggesting that plot-level and grid cell-level trait composition are driven by different factors[21]. Plot-level trait means and variances may both be predominantly driven by local environmental factors, such as topography (e.g. north- *vs.* south-facing slopes), local soil characteristics (e.g. soil depth and nutrient supply)[6,7,24,25], disturbance regime (including land use[26] and successional status[9,27]) or biotic interactions[18-19,28], while broad-scale climate and soil conditions may only become relevant for the whole species pool in large grid cells. Such differences emphasize the importance of local environment in affecting the communities' trait composition and should be taken into account when interpreting the effect of environmental drivers in functional trait diversity using data on either floristic pools or ecological communities.

We note that the strongest community-level correlations with environment were found for traits not linked to the leaf economics spectrum. Mean stem specific density increased with potential evapotranspiration (PET, $r^2=15.6\%$; Fig. 4a, b), reflecting the need to produce denser wood with increasing evaporative demand. Leaf N:P ratio increased with growing-season warmth (growing degree days above 5°C, GDD5, $r^2=11.5\%$; Fig. 4d), indicating strong phosphorus limitation[29] in most plots in the tropics and subtropics (Fig. 4c, d). This pattern was not brought about by a parallel increase in the presence of legumes, which tend to have relatively high N:P ratios; excluding all species of Fabaceae resulted in a very similar relationship with GDD5 ($r^2=10.0\%$). The global N:P pattern is consistent with results based on traits of single species related to mean annual temperature[30]. The underlying mechanism is the high soil weathering rate at high temperatures and humidity, which in the tropics and subtropics was not reset by Pleistocene glaciation. Thus, phosphorus limitation may weaken the relationships between productivity-related traits and macroclimate (Extended Data Fig. 2). For example, specific leaf area may be low as consequence of low nutrient availability[6-7,24-25] in favourable climates as well as be low as consequence of low temperature and precipitation under favourable nutrient supply. Overall, our findings are relevant in improving Dynamic Global Vegetation Models (DGVMs), which so far have used trait information only from a few calibration plots[22]. The sPlot database provides much-needed empirical data on the

community trait pool in DGVMs[31] and identifies traits that should be considered when predicting ecosystem functions from vegetation, such as stem specific density and leaf N:P ratio.

Our results were surprisingly robust both to the selection of trait data, when comparing different plant formations and when explicitly accounting for the uneven distribution of plots. Using the original trait values measured for the species from the TRY database for the six traits used by Díaz et al.[1] (see Methods), resulted in the same two main functional continua and an overall highly similar ordination pattern (Extended Data Fig. 4) compared to using gap-filled data for 18 traits (Fig. 2). Community-level trait composition was also similarly poorly captured by global climate and soil variables. Single regressions of CWMs with all environmental variables revealed very similar patterns to those based on gap-filled traits (Extended Data Fig. 5). Similarly, subjecting the CWMs based on six original traits to a Redundancy Analysis with all 30 environmental variables accounted only for 20.6% or 21.8% of the overall variation in CWMs (cumulative variance of the first two or all six constrained axes, respectively, Extended Data Fig. 4). These results clearly demonstrate that the imputation of missing trait values did not result in spurious artefacts which may have obscured community trait-environment relationships.

We also assessed whether the observed trait-environment relationships hold for forests and non-forest vegetation independently (see Methods). Both subsets confirmed the overall patterns in trait means (Extended Data Figs. 3-6). The variance in plot-level trait means explained by large-scale climate and soil variables was higher for forest than non-forest plots, probably because forests belong to a well-defined and rather resource-conservative formation, whereas non-forest plots encompass a heterogeneous mixture of different vegetation types, ranging from alpine meadows to semi-deserts, and tend to depend more on disturbance and management, which can strongly affect trait-environment relationships of communities[21]. Finally, to test whether our findings depended on the uneven distribution of plots among the world's different climates and soils, we repeated the analyses in 100 subsets of ~100,000 plots resampled in the global climate space (Extended Data Figs. 7-8). The analyses of the resampled datasets revealed the same patterns and confirmed the impact of PET and GDD5 on stem specific density and leaf N:P ratio, respectively. The correlations between trait means and environmental variables were, however, stronger in the resampled subsets, possibly because the resampling procedure reduced the overrepresentation of the temperate-zone areas with intermediate climatic values.

Our findings have important implications for understanding and predicting plant community trait assembly. First, worldwide trait variation of plant communities is captured by a few main dimensions of variation, which are surprisingly similar to those reported by species-based studies[1,12-14], suggesting that the drivers of past trait evolution, which resulted in the present-day species-level trait spectra[1], are also reflected in the composition of today's plant communities. If species-level trade-offs indeed constrain community assembly, then the present-day contrasts in trait composition of terrestrial plant communities should also have existed in the past and will probably remain, even for novel communities, in the future. Most species in our present-day communities evolved under very variable filtering conditions across the globe, with respect to temperature and precipitation regimes. Therefore, it can be

assumed that future filtering conditions will result in novel communities that follow the same functional continua from short-stature, small-seeded communities to tall-stature communities with large, heavy diaspores and from communities with resource-acquisitive to those with resource-conservative leaves. Second, the main plot-level vegetation trait continua cannot easily be captured by coarse-resolution environmental variables[21]. This brings into question both the use of simple large-scale climate relationships to predict the leaf economics spectra of global vegetation[4,15-16,22] and attempts to derive net primary productivity and global carbon and water budgets from global climate, even when employing powerful trait-based vegetation models[31]. The finding that within-plot trait variances were only very weakly related to global climate or soil variables points to the importance of i) local-scale climate or soil variables, ii) disturbance regimes or iii) biotic interactions for the degree of local trait dispersion[8]. Finally, both our findings on the limited role of large-scale climate in explaining trait patterns and on the prevalence of phosphorus limitation in most plots in the tropics and subtropics call for including local variables when predicting community trait patterns. Even under similar macro-environmental conditions, communities can vary greatly in trait means and variances, consistent with high local variation in species' trait values[5,6,12]. Future research on functional response of communities to changing climate should incorporate the effect of local environmental conditions[24-26] and biotic interactions[18-19] for building reliable predictions of vegetation dynamics.

References

1. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167-171 (2016).

2. Marks, C.O. & Lechowicz, M.J. Alternative designs and the evolution of functional diversity. *Am. Nat.* **167**, 55–67 (2006).

3. Muscarella, R. & Uriarte, M. Do community-weighted mean functional traits reflect optimal strategies? *Proc. R. Soc. B.* **283**, 20152434 (2016).

4. Swenson, N.G. & Weiser, M.D. Plant geography upon the basis of functional traits: An example from eastern North American trees. *Ecology* **91**, 2234–2241 (2010).

5. Moles, A.T. *et al.* Global patterns in plant height. *J. Ecol.* **97**, 923-932 (2009).

6. Ordoñez, J.C. *et al.* A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. *Global Ecol. Biogeogr.* **18**, 137–149 (2009).

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

8. Grime, J.P. Trait convergence and trait divergence in herbaceous plant communities: Mechanisms and consequences. *J. Veg. Sci.* **17**, 255–260 (2006).

9. Garnier, E. *et al.* Plant functional markers capture ecosystem properties during secondary succession. *Ecology* **85**, 2630–2637 (2004).

458 10. Warming, E. *Lehrbuch der ökologischen Pflanzengeographie – Eine Einführung in die*
459 *Kenntnis der Pflanzenvereine.* (Borntraeger, Berlin, 1896).

460 11. Garnier, E., Navas, M.-L. & Grigulis, K. *Plant functional diversity - Organism traits,*
461 *community structure, and ecosystem properties.* (Oxford Univ. Press, 2016).

462 12. Wright, I.J. *et al.* The worldwide leaf economics spectrum. *Nature* **428**, 821-827 (2004).

463 13. Reich, P.B. The world–wide 'fast–slow' plant economics spectrum: a traits manifesto. *J.*
464 *Ecol.* **102**, 275–301 (2014).

465 14. Adler, P.B. *et al.* Functional traits explain variation in plant life history strategies. *Proc.*
466 *Natl. Acad. Sci. USA* **111**, 740–745 (2014).

467 15. Swenson, N.G. *et al.* Phylogeny and the prediction of tree functional diversity across
468 novel continental settings. *Global Ecol. Biogeogr.* **26**, 553–562 (2017).

469 16. Swenson, N.G. *et al.* The biogeography and filtering of woody plant functional diversity
470 in North and South America. *Global Ecol. Biogeogr.* **21**, 798–808 (2012).

471 17. Wright, I.J. *et al.* Global climatic drivers of leaf size. *Science* **357**: 917–921 (2017).

472 18. Mayfield, M.M. & Levine, J.M. Opposing effects of competitive exclusion on the
473 phylogenetic structure of communities. *Ecol. Lett.* **13**, 1085 – 1093 (2010).

474 19. Kraft, N.J.B. *et al.* Community assembly, coexistence and the environmental filtering
475 metaphor. *Funct. Ecol.* **29**, 592–599 (2015).

476 20. Barboni, D. *et al.* Relationships between plant traits and climate in the Mediterranean
477 region: A pollen data analysis. *J. Veg. Sci.* **15**, 635-646 (2004).

478 21. Borgy, B. *et al.* Plant community structure and nitrogen inputs modulate the climate signal
479 on leaf traits. *Global Ecol. Biogeogr.* **26**, 1138-1152 (2017).

480 22. van Bodegom, P.M, Douma, J.C. & Verheijen, L.M. A fully traits-based approach to
481 modeling global vegetation distribution. *Proc. Natl. Acad. Sci. USA* **111**, 13733–13738
482 (2014).

483 23. Moles, A.T. *et al.* Which is a better predictor of plant traits: Temperature or precipitation?
484 *J. Veg. Sci.* **25**, 1167–1180 (2014).

485 24. Ordoñez, J.C. *et al.* Plant strategies in relation to resource supply in mesic to wet
486 environments: Does theory mirror nature? *Am. Nat.* **175**, 225–239 (2010).

487 25. Simpson, A.J., Richardson, S.J. & Laughlin, D.C. Soil–climate interactions explain
488 variation in foliar, stem, root and reproductive traits across temperate forests. *Global Ecol.*
489 *Biogeogr.* **25**, 964-978 (2016).

490 26. Lienin, P. & Kleyer, M. Plant leaf economics and reproductive investment are responsive
491 to gradients of land use intensity. *Agric. Ecosyst. Environ.* **145**, 67-76 (2011).

27. Maire, V. *et al.* Habitat filtering and niche differentiation jointly explain species relative abundance within grassland communities along fertility and disturbance gradients. *New Phytol.* **196**, 497–509 (2012).

28. Craine, J.M. *et al.* Global patterns of foliar nitrogen isotopes and their relationships with climate, mycorrhizal fungi, foliar nutrient concentrations, and nitrogen availability *New Phytol.* **183**, 980–992 (2009).

29. Güsewell, S. N:P ratios in terrestrial plants: variation and functional significance. *New Phytol.* **164**, 243–266 (2004).

30. Reich, P.B. & Oleksyn, J. Global patterns of plant leaf N and P in relation to temperature and latitude, *Proc. Natl. Acad. Sci. USA* **101**, 11001-11006 (2004).

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

Contributions

H.B. and U.J. wrote the first draft of the manuscript, with considerable input by B.J.-A. and R.F.; H.B. carried out most of the statistical analyses and produced the graphs; H.B., O.Pu. and U.J. initiated sPlot as an sDiv working group and iDiv platform; J.De. compiled the plot databases globally; J.De., S.M.H., U.J., O.Pu. and F.J. harmonized vegetation databases; J.De. and B.J.-A. coordinated the sPlot consortium; J.K. provided the trait data from TRY; F.S. performed the trait data gap filling; O.Pu. produced the taxonomic backbone; B.J.-A., G.S. and E. Welk compiled environmental data and produced the global maps; S.M.H. wrote the Turboveg v3 software, which holds the sPlot database; J.L. and T.H. wrote the resampling algorithm. Many authors participated in one or more of the three sPlot workshops at iDiv where the sPlot initiative was conceived and planned, and evaluation of the data and first drafts were discussed. All other authors contributed data. All authors contributed to writing the manuscript.

Acknowledgements

sPlot has been initiated by sDiv, the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118) and now is a platform of iDiv. H.B., J.De., O.Pu, B.J.-A., J.K., D.C., F.M.S., M.W. and C.W. appreciate direct funding through iDiv. For all further acknowledgements see the Electronic Appendix.

Material and Methods

Vegetation Data. The sPlot 2.1 vegetation database contains 1,121,244 plots with 23,586,216 species × plot observations, i.e. records of a species in a plot (https://www.idiv.de/en/sdiv/working_groups/wg_pool/splot.html). This database aims at compiling plot-based vegetation data from all vegetation types worldwide, but with a particular focus on forest and grassland vegetation. Although the initial aim of sPlot was to achieve global coverage, the plots are very unevenly distributed with most data coming from Europe, North America and Australia and an overrepresentation of temperate vegetation types (Extended Data Fig. 1).

For most plots (97.2%) information on the single species' relative contribution to the sum of plants in the plot was available, expressed as cover, basal area, individual count, importance value or per cent frequency in subplots. For the other 2.8% (31,461 plots), for which only presence/absence (p/a) was available, we assigned equal relative abundance to the species (1/species richness). For plots with a mix of cover and p/a information (mostly forest plots, where herb layer information had been added on a p/a basis; 8,524 plots), relative abundance was calculated by assigning the smallest cover value that occurred in a particular plot to all species with only p/a information in that plot. In most cases (98.4%), plot records in sPlot include full species lists of vascular plants. Bryophytes and lichens were additionally identified in 14% and 7% of plots, respectively. After removing plots without geographic coordinates and all observations on bryophytes and lichens, the database contained 22,195,966 observations on the relative abundance of vascular plant species in a total of 1,117,369 plots. The temporal extent of the data spans from 1885 to 2015, but >95% of vegetation plots were recorded later than 1980. Plot size was reported in 65.4% of plots. While forest plots had plot sizes ≥100 m^2, and in most cases ≤1,000 m^2, non-forest plots typically ranged from 5 to 100 m^2.

Taxonomy. To standardize the nomenclature of species within and between sPlot and TRY (see below), we constructed a taxonomic backbone of the 121,861 names contained in the two databases. Prior to name matching, we ran a series of string manipulation routines in R, to remove special characters and numbers, as well as standardized abbreviations in names. Taxon names were parsed and resolved using Taxonomic Name Resolution Service version 4.0 (TNRS[32]; http://tnrs.iplantcollaborative.org; accessed 20 Sep 2015), selecting the best match across the five following sources: i) The Plant List (version 1.1; http://www.theplantlist.org/; Accessed 19 Aug 2015), ii) Global Compositae Checklist (GCC, http://compositae.landcareresearch.co.nz/Default.aspx; accessed 21 Aug 2015), iii) International Legume Database and Information Service (ILDIS, http://www.ildis.org/LegumeWeb; accessed 21 Aug 2015), iv) Tropicos (http://www.tropicos.org/; accessed 19 Dec 2014), and v) USDA Plants Database (http://usda.gov/wps/portal/usda/usdahome; accessed 17 Jan 2015). We allowed for partial matching to the next higher taxonomic rank (genus or family) in cases where full taxon names could not be found. All names matched or converted from a synonym by TNRS were considered accepted taxon names. In cases when no exact match was found (e.g. when alternative spelling corrections were reported), names with probabilities of ≥ 95% or higher were accepted and those with < 95% were examined individually. Remaining non-matching

569 names were resolved based on the National Center for Biotechnology Information's
570 Taxonomy database (NCBI, http://www.ncbi.nlm.nih.gov/; accessed 25 Oct 2011) within
571 TNRS, or sequentially compared directly against The Plant List and Tropicos (accessed
572 September 2015). Names that could not be resolved against any of these lists were left as
573 blanks in the final standardized name field. This resulted in a total of 86,760 resolved names,
574 corresponding to 664 families, occurring in sPlot or TRY or both. Classification into families
575 was carried out according to APGIII[33], and was used to identify non-vascular plant species
576 (~5.1% of the taxon names) which were excluded from the subsequent statistical analysis.

577 **Trait Data.** Data for 18 traits that are ecologically relevant (Table 1) and sufficiently covered
578 across species[34] were requested from TRY[35] (version 3.0) on the 10th August, 2016. We
579 applied gap-filling with Bayesian Hierarchical Probabilistic Matrix Factorization
580 (BHPMF[34,36-37]). We used the prediction uncertainties provided by BHPMF for each
581 imputation to assess the quality of gap-filling and removed all imputations with a coefficient
582 of variation > 1[37]. We obtained 18 gap-filled traits for 26,632 out of a total of 58,065 taxa in
583 sPlot, which corresponds to 45.9% of all species but to 88.7% of all species × plot
584 combinations. Trait coverage of the most frequent species was 77.2% and 96.2% for taxa that
585 occurred in more than 100 or 1,000 plots, respectively. The gap-filled trait data comprised
586 observed and imputed values on 632,938 individual plants, which we $\log_e$ transformed and
587 aggregated by taxon. For those taxa that were recorded at the genus level only, we calculated
588 genus means. Out of 22,195,966 records of vascular plant species with geographic reference,
589 21,172,989 (=95.4%) refer to taxa for which we had gap-filled trait values. This resulted in
590 1,115,785 and 1,099,463 plots for which we had at least one taxon or two taxa with a trait
591 value (99.5% and 98.1%, respectively, of all 1,121,244 plots), and for which trait means and
592 variances could be calculated.

593 As some mean values of traits in TRY were based on a very small number of replicates per
594 species, which results in uncertainty in trait mean and variance calculations[38], we tested to
595 which degree the trait patterns in the dataset might be caused by a potential removal of trait
596 variation by imputation of trait values and additionally carried out all analyses using the
597 original trait data on the same 632,938 individual plants instead of gap-filled data (Extended
598 Data Table 1). The degree of trait coverage of species ranged between 7.0% and 58.0% for
599 leaf fresh mass and plant height, respectively. Across all species, mean coverage of species
600 with original trait values was 21.8%, as compared to 45.9% for gap-filled trait data. Linking
601 these trait values to the species occurrence data resulted in a coverage of species × plot
602 observations with trait values between 7.6% and 96.6% for conduit element length and plant
603 height, respectively (Extended Data Table 1), with a mean of 60.7% as compared to 88.7% for
604 those based on gap-filled traits. Using these original trait values to calculate community-
605 weighted mean (CWM) trait values (see below) resulted in a plot coverage of trait values
606 between 48.2% and 100% for conduit element length and SLA, respectively. Across all plots,
607 mean coverage of plots with original trait values was 89.3%, as compared to 100% for gap-
608 filled trait data (Extended Data Table 1).

609 We are aware that using species mean values for traits excludes the possibility to account for
610 intraspecific variance, which can also strongly respond to the environment[39]. Thus, using one
611 single value for a species is a source of error in calculating trait means and variances.

Environmental Data. We compiled 30 environmental variables (Extended Data Table 2). Macroclimate variables were extracted from CHELSA[40-41], V1.1 (Climatologies at High Resolution for the Earth's Land Surface Areas, www.chelsa-climate.org). CHELSA provides 19 bioclimatic variables equivalent to those used in WorldClim (www.worldclim.org) at a resolution of 30 arc sec (~ 1 km at the equator), averaging global climatic data from the period 1979–2013 and using a quasi-mechanistic statistical downscaling of the ERA-Interim reanalysis[42].

Variables reflecting growing-season warmth were growing degree days above 1°C (GDD1) and 5°C (GDD5), calculated from CHELSA data[43]. We also compiled an index of aridity (AR) and a model for potential evapotranspiration (PET) extracted from the Consortium of Spatial Information (CGIAR-CSI) website (www.cgiar-csi.org). In addition, seven soil variables were extracted from the SOILGRIDS project (https://soilgrids.org/, licensed by ISRIC – World Soil Information), downloaded at 250 m resolution and then resampled using the 30 arc second grid of CHELSA (Extended Data Table 2). We refer to these climate and soil data as "environmental data".

Community trait composition.

For every trait *j* and plot *k*, we calculated the plot-level trait means as community-weighted mean (CWM) according to[9,44]:

$$CWM_{j,k} = \sum_{i}^{n_k} p_{i,k}\, t_{i,j}$$

where n_k is the number of species sampled in plot *k*, $p_{i,k}$ is the relative abundance of species *i* in plot *k*, referring to the sum of abundances for all species with traits in the plot, and $t_{i,j}$ is the mean value of species *i* for trait *j*. This computation was done for each of the 18 traits for 1,115,785 plots. The within-plot trait variance is given by community-weighted variance (CWV)[44,45]:

$$CWV_{j,k} = \sum_{i}^{n_k} p_{i,k}\, (t_{i,j} - CWM_{j,k})^2$$

CWV is equal to functional dispersion as described by Rao's quadratic entropy[46], when using a squared Euclidean distance matrix $d_{i,j,k}$ [47]:

$$CWV_{j,k} = \sum_{i}^{n_k} p_{i,k}\, (t_{i,j} - CWM_{j,k})^2 = FD_Q = \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} p_{i,k}\, p_{j,k}\ {d_{i,j,k}}^2$$

We had CWV information for 18 traits for 1,099,463 plots, as at least two taxa were needed to calculate CWV. We performed the calculations using the 'data.table' package[48] in R.

Assessing the degree of filtering. To analyse how plot-level trait means and within-plot trait variances (based on gap-filled trait data) depart from random expectation, for each trait we calculated standardized effect sizes (SESs) for the variance in CWMs and for the mean in CWVs. Significantly positive SESs in variance of CWM and significantly negative ones in the mean of CWV can be considered a global-level measure of environmental or biotic filtering. To provide an indication of the global direction of filtering, we also report SESs for the mean of CWM trait values. Similarly, to measure how much within-community trait dispersion varied globally, we also calculated SESs for the variance in CWV.

SESs were obtained from 100 runs of randomizing trait values across all species globally. In every run we calculated CWM and CWV with random trait values, but keeping all species abundances in plots. Thus, the results of randomization are independent from species co-occurrences structure of plots[49]. For every trait, the SESs of the variance in CWM were calculated as the observed value of variance in CWM minus the mean variance in CWM of the random runs, divided by the standard deviation of the variance in CWM of the random runs (Fig. 1). SESs for the mean in CWM, the mean in CWV and the variance in CWV were calculated accordingly. Tests for significance of SESs were obtained by fitting generalized Pareto-distribution of the most extreme random values and then estimating *p* values form this fitted distribution[50].

Vegetation trait-environment relationships. Out of the 1,115,785 plots with CWM values, 1,114,304 (99.9%) had complete environmental information and coordinates. This set of plots was used to calculate single linear regressions of each of the 18 traits on each of the 30 environmental variables. We used the 'corrplot' function[51] in R to illustrate Pearson correlation coefficients (see Extended Data Figs. 1-2, 4, 6, 8) and for the strongest relationships produced bivariate graphs and mapped the global distribution of the CWM values using kriging interpolation in ArcGIS 10.2 (Fig. 4). We also tested for non-linear relationships with environment by including an additional quadratic term in the linear model and then report coefficients of determination. As in the linear relationships of CWM with environment, the highest r^2 values in models with an additional quadratic term were encountered between stem specific density and PET (r^2=0.156) and leaf N:P ratio and growing degree days above 5°C (GDD5, r^2=0.118). These were not substantially different from the linear CWM-environment relationships, which had r^2=0.156 and r^2=0.115, respectively (Fig. 4, Extended Data Fig. 2). Similarly, including a quadratic term in the regressions did not increase the CWV-environment correlations. Here, the strongest correlations were encountered between plant height and soil pH (r^2=0.044) and between specific leaf area (SLA) and the volumetric content of coarse fragments in the soil (CoarseFrags, r^2=0.037), which were similar to those in the linear regressions (r^2=0.029 and r^2=0.036, respectively, Extended Data Fig. 3).

To account for a possible confounding effect of species richness on CWV, which may cause low CWV through competitive exclusion of species, we regressed CWV on species richness and then calculated all Pearson correlation coefficients with the residuals of this relationship against all climatic variables. Here, the highest correlation coefficients were encountered

between PET and CWV of conduit element length (r^2=0.038), followed by the relationship of specific leaf area (SLA) and the volumetric content of coarse fragments in the soil (CoarseFrags, r^2=0.034), which were very similar in magnitude to the CWV environment correlations (r^2=0.035 and r^2=0.036, respectively; Extended Data Fig. 3).

The CWMs and CWVs were scaled to a mean of zero and standard deviation of one and then subjected to a Principal Component Analysis (PCA), calculated with the 'rda' function from the 'vegan' package[52]. Climate and soil variables were fitted *post hoc* to the ordination scores of plots of the first two axes, producing correlation vectors using the 'envfit' function. We refrain from presenting any inference statistics, as with > 1.1 million plots all environmental variables showed statistically significant correlations. Instead, we report coefficients of determination (r^2), obtained from Redundancy Analysis (RDA), using all 30 environmental variables as constraining matrix, resulting in a maximum of 18 constrained axes corresponding to the 18 traits. We report both r^2 values of the first two axes explained by environment, which is the maximum correlation of the best linear combination of environmental variables to explain the CWM or CWV plot × trait matrix and r^2 values of all 18 constrained axes explained by environment. We plotted the PCA results using the 'ordiplot' function and coloured the points according to the logarithm of the number of plots that fell into grid cells of 0.002 in PCA units (resulting in approximately 100,000 cells). For further details, see the captions of the figures.

Additionally, we carried out the PCA and RDA analyses, using CWMs based on original trait values (see above). Because of a poor coverage of some traits we confined the analyses with original trait values to the six traits used by Díaz et al.[1], which were leaf area, specific leaf area, leaf N, seed mass, plant height and stem specific density. Using these six traits resulted in 954,459 plots that had at least one species with a trait value for each of the six traits.

Testing for formation-specific patterns. We carried out separate analyses for two 'formations': forest and for non-forest plots. We defined as forest plots that had > 25% cover of the tree layer. However, this information was available for only 25% of the plots in our sPlot database. Thus, we also assigned formation status based on growth form data from the TRY database. We defined plots as 'forest' if the sum of relative cover of all tree taxa was > 25%, but only if this did not contradict the requirement of > 25% cover of the tree layer (for those records for which this information was given in the header file). Similarly, we defined non-forest plots by calculating the cover of all taxa that were not defined as trees and shrubs (also taken from the TRY plant growth form information) and that were not taller than 2 m, using the TRY data on mean plant height. We assigned the status 'non-forest' to all plots that had >90% cover of these low-stature, non-tree and non-shrub taxa. In total, 21,888 taxa out of the 52,032 in TRY which also occurred in sPlot belonged to this category, and 16,244 were classed as trees. The forests and non-forest plots comprised 330,873 (29.7%) and 513,035 (46.0%) of all plots, respectively. We subjected all CWM values for forest and non-forest plots to PCA, RDA and bivariate linear regressions to environmental variables as described above.

The forest plots, in particular, confirmed the overall patterns, with respect to variation in CWM explained by the first two PCA axes (60.5%) and the two orthogonal continua from small to large size and the leaf economics spectrum (Extended Data Fig. 6). The variation explained by macroclimate and soil conditions was much larger for the forest subset than for the total data, with the best relationship (leaf N:P ratio and the mean temperature of the coldest quarter, bio11) having r^2=0.369 and the second next best ones (leaf N:P ratio and GDD1 and GDD5) close to this value with r^2=0.357 (Extended Data Fig. 7) and an overall variation in CWM values explained by environment of 25.3% (cumulative variance of all 18 constrained axes in a RDA). The non-forest plots showed the same functional continua, but with lower total amount of variation in CWM accounted for by the first two PCA axes (41.8%, Extended Data Fig. 8) and much lower overall variation explained by environment. For non-forests, the best correlation of any CWM trait with environment was the one of volumetric content of coarse fragments in the soil (CoarseFrags) and leaf C content per dry mass with r^2=0.042 (Extended Data Fig. 9). Similarly, the cumulative variance of all 18 constrained axes according to RDA was only 4.6%. This shows, on the one hand, that forest and non-forest vegetation are characterized by the same interrelationships of CWM traits, and on the other hand, that the relationships of CWM values with the environment were much stronger for forests than for non-forest formations. The coefficients of determination were even higher than those previously reported for trait-environment relationships for North American forests (between CWM of seed mass and maximum temperature, r^2=0.281)[6].

Resampling procedure in environmental space. In order to achieve a more even representation of plots across the global climate space, we first subjected the same 30 global climate and soil variables as described above, to a Principal Component Analysis (PCA), using the climate space of the whole globe, irrespective of the presence of plots in this space, and scaling each variable to a mean of zero and a standard deviation of one. We used a 2.5 arc minute spatial grid, which comprised 8,384,404 terrestrial grid cells. We then counted the number of vegetation plots in the sPlot database that fell into each grid cell. For this analysis, we did not use the full set of 1,117,369 plots with trait information (see above), but only those plots that had a location inaccuracy of max. 3 km, resulting in a total of 799,400 plots. The resulting PCA scores based on the first two principal components (PC1-PC2) were rasterized to a 100 × 100 grid in PC1-PC2 environmental space, which was the most appropriate resolution according to a sensitivity analysis. This sensitivity analysis tested different grid resolutions, from a coarse-resolution bivariate space of 100 grid cells (10 × 10) to a very fine-resolution space of 250,000 grid cells (500 × 500), iteratively increasing the number of cells along each principal component by 10 cells. For each iteration, we computed the total number of sPlot plots per environmental grid cell and plotted the median sampling effort (number of plots) across all grid cells versus the resolution of the PC1-PC2 space. We found that the curve flattens off at a bivariate environmental space of 100 × 100 grid cells, which was the resolution for which the median sampling effort stabilized at around 50 plots per grid cell. As a result, we resampled plots only in environmental cells with more than 50 plots (858 cells in total).

To optimize our resampling procedure within each grid cell, we used the heterogeneity-constrained random (HCR) resampling approach[53]. The HCR approach selects the subset of

vegetation plots for which those plots are the most dissimilar in their species composition while avoiding selection of plots representing peculiar and rare communities that differ markedly from the main set of plant communities (outliers), thus providing a representative subset of plots from the resampled grid cell. We used the turnover component of the Jaccard's dissimilarity index (β_{jtu}[54]) as a measure of dissimilarity. The β_{jtu} index accounts for species replacement without being influenced by differences in species richness. Thus, it reduces the effects of any imbalances that may exist between different plots due to species richness. We applied the HCR approach within a given grid cell by running 1,000 iterations of randomly selecting 50 plots out of the total number of plots available within that grid cell. Where the cell contained 50 or fewer plots, all were included and the resampling procedure was not run. This procedure thinned out over-sampled climate types, while retaining the full environmental gradient.

All 1,000 random draws of a given grid cell were subsequently sorted according to the decreasing mean of β_{jtu} between pairs of vegetation plots and then sorted again according to the increasing variance in β_{jtu} between pairs of vegetation plots. Ranks from both sortings were summed for each random draw, and the random draw with the lowest summed rank was considered as the most representative of the focal grid cell. Because of the randomized nature of the HCR approach, this resampling procedure was repeated 100 times for each of the 858 grid cells. This enabled us to produce 100 different subsamples out of the full sample of 799,400 vegetation plots subjected to the resampling procedure. Each of these 100 subsamples was finally subjected to ordinary linear regression, PCA and RDA as described above. We calculated the mean correlation coefficient across the 100 resampled data sets for each environmental variable with each trait.

To plot bivariate relationships, we used the mean intercept and slope of these relationships. PCA loadings of all 100 runs were stored and averaged. As different runs showed different orientation on the first PCA axes, we switched the signs of the axis loadings in some of the runs to make the 100 PCAs comparable to the reference PCA, based on the total data set. Across the 100 resampled data sets, we then calculated the minimum and maximum loading for each of the two PCA axes and plotted the result as ellipsoid. We also collected the post-hoc regressions coefficients of PCA scores with the environmental variables in each of the 100 runs, switched the signs accordingly and plotted the correlations to PC1 and PC2 as ellipsoids. The result is a synthetic PCA of all 100 runs. To illustrate the coverage of plots in PCA space, we used plot scores of one of the 100 random runs. Similarly, the coefficients of determination obtained from the RDAs of these 100 resampled sets were averaged.

The mean PCA loadings across these 100 subsets (summarized in Extended Data Fig. 10) were fully consistent with those of the full data set in Fig. 2, with the same two functional continua in plant size and diaspore mass (from bottom left to top right), and perpendicular to that, the leaf economics spectrum. The variation in CWM accounted for by the first two axes was on average 50.9% ± 0.04 standard deviations (SD), and thus, virtually identical with that in the total dataset. In contrast, the variation explained on average by macroclimate and soil conditions (26.5% ± 0.01 SD as average cumulative variance of all 18 constrained axes in the RDAs across all 100 runs) was considerably larger than that for the total dataset, which is also reflected in consistently higher correlations between traits and environmental variables

(Extended Data Fig. 11). The highest mean correlation was encountered for plant height and PET (mean r^2=0.342 across 100 runs). PET was a better predictor for plant height than the precipitation of the wettest months (bio13, mean r^2=0.231), as had been suggested previously[7]. The correlation of PET with stem specific density (mean r^2=0.284) and warmth of the growing season (expressed as growing degree days above the threshold 5°C, GDD5) with leaf N:P ratio (mean r^2=0.250) ranked among the best 12 correlations encountered out of all 540 trait-environment relationships, which confirms the patterns found in the whole data set (compared with Fig. 4). Overall, the coefficients of determination were much closer to the ones reported from other studies with a global collection of a few hundred plots (r^2 values ranging from 36% to 53% based on multiple regressions of single traits with five to six environmental drivers[22]).

References

32. Boyle, B. *et al.* The Taxonomic Name Resolution Service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**, 16 (2013).

33. Bremer, B. *et al.* An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121 (2009).

34. Schrodt, F. *et al.* BHPMF – a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecol. Biogeogr.* **24**, 1510–1521 (2015).

35. Kattge J. *et al.* TRY—a global database of plant traits. *Glob. Change Biol.* **17**, 2905–2935 (2011).

36. Shan, H. *et al.* Gap filling in the plant kingdom - Trait prediction using hierarchical probabilistic matrix factorization. *Proceedings of the 29th International Conference for Machine Learning (ICML 2012)* 1303−1310 (2012).

37. Fazayeli, F. *et al.* Uncertainty quantified matrix completion using Bayesian Hierarchical Matrix factorization. *13th International Conference on Machine Learning and Applications (ICMLA 2014)* 312−317 (2014).

38. Borgy, B. *et al.* Sensitivity of community-level trait–environment relationships to data representativeness: A test for functional biogeography. *Global Ecol. Biogeogr.* **26**, 729–739 (2017).

39. Herz, K. *et al.* Drivers of intraspecific trait variation of grass and forb species in German meadows and pastures. *J. Veg. Sci.* **28**, 705–716 (2017).

40. Karger, D.N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122. doi: 10.1038/sdata.2017.122 (2017)

41. Karger, D.N. *et al.* Climatologies at high resolution for the Earth land surface areas (Version 1.1). *World Data Center for Climate (WDCC) at DKRZ.* http://chelsa-climate.org/downloads/ (2016).

42. Dee, D. P. *et al.* The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.* **137**, 553–597 (2011).

43. Synes, N.W. & Osborne, P.E. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecol. Biogeogr.* **20**, 904–914 (2011).

44. Enquist, B. *et al.* Scaling from traits to ecosystems: developing a general trait driver theory via integrating trait-based and metabolic scaling theories. *Adv. Ecol. Res.* **52**, 249-318 (2015).

45. Buzzard, V. *et al.* Re-growing a tropical dry forest: functional plant trait composition and community assembly during succession. *Funct. Ecol.* **30**, 1006–1013 (2016).

46. Rao, C.R. Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.* **21**, 24–43 (1982).

47. Champely, S. & Chessel, D. Measuring biological diversity using Euclidean metrics. *Environ. Ecol. Stat.* **9**, 167-177 (2002).

48. Dowle, M. *et al.* data.table: Extension of Data.frame. R package version 1.9.6. (2015) https://CRAN.R-project.org/package=data.table

49. Hawkins, B.A. *et al.* Structural bias in aggregated species-level variables driven by repeated species co-occurrences: a pervasive problem in community and assemblage data. *J. Biogeogr.* **44**, 1199–1211 (2017).

50. Knijnenburg, T.A. *et al.* Fewer permutations, more accurate P-values. *Bioinformatics* **25**, i161–i168 (2009).

51. Friendly, M. Corrgrams: Exploratory displays for correlation matrices. *Am. Statistician*, **56**, 316–324 (2002).

52. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.3-3 (2016). https://CRAN.R-project.org/package=vegan

53. Lengyel, A., Chytrý, M. & Tichý, L. Heterogeneity-constrained random resampling of phytosociological databases. *J. Veg. Sci.* **22**, 175–183 (2011).

54. Baselga, A. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecol. Biogeogr.* **21**, 1223-1232 (2012).

55. Garnier, E. *et al.* Towards a thesaurus of plant characteristics: an ecological contribution. *J. Ecol.* **105**, 298-309 (2017). www.top-thesaurus.org

Detailed Acknowledgements

The study has been supported by the TRY initiative on plant traits (http://www.try-db.org). The TRY initiative and database is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany). TRY is currently supported by DIVERSITAS/Future Earth and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig.

Jan Altman was funded by research grants 17-07378S of the Grant Agency of the Czech Republic and long-term research development project no. RVO 67985939.

Isabelle Aubin was funded through Natural Sciences and Engineering Research Council of Canada and Ontario Ministry of Natural Resources and Forestry.

Idoia Biurrun was funded by the Basque Government (IT936-16).

Benjamin Blonder was supported by the UK Natural Environment Research Council (NE/M019160/1).

Anne Bjorkman and Isla Myers-Smith thank the Herschel Island-Qikiqtaruk Territorial Park management, Catherine Kennedy, Dorothy Cooley, Jill F. Johnstone, Cameron Eckert and Richard Gordon for establishing the ecological monitoring programme. Funding was provided by Herschel Island-Qikiqtaruk Territorial Park.

Zoltán Botta-Dukát was supported by project GINOP-2.3.2-15-2016-00019.

Andraž Čarni acknowledges the financial support from the Slovenian Research Agency (research core funding No. P1-0236).

Luis Cayuela was supported by project BIOCON08_044 funded by Fundación BBVA.

Milan Chytrý and Ilona Knollová were supported by the Czech Science Foundation (14-36079G, Centre of Excellence Pladias).

Greg Guerin acknowledges support from the Terrestrial Ecosystem Research Network (Australia).

Alvaro G. Gutiérrez acknowledges FONDECYT 11150835, Project FORECOFUN-SSA PIEF-GA-2010–274798), CONICYT-PAI (82130046).

Pedro Higuchi has been awarded a research grant by the Brazilian National Council for Scientific and Technological Development (CNPq).

Jürgen Homeier received funding from BMBF (Federal Ministry of Education and Science of Germany) and the German Research Foundation (DFG Ho3296-2, DFG Ho3296-4).

Jens Kattge acknowledges support by the Max Planck Institute for Biogeochemistry (Jena, Germany), Future Earth, the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig and the EU H2020 project BACI, Grant No 640176.

Jérôme Munzinger was supported by the French National Research Agency (ANR) with grants INC (ANR-07-BDIV-0008), BIONEOCAL (ANR-07-BDIV-0006) & ULTRABIO

(ANR-07-BDIV-0010), by National Geographic Society (Grant 7579-04), and with fundings and authorizations of North and South Provinces of New Caledonia.

Ülo Niinemets and Meelis Pärtel were supported by the European Commission through the European Regional Development Fund (the Center of Excellence EcolChange). Meelis Pärtel acknowledges funding by the Estonian Ministry of Education and Research (IUT20-29)

Josep Peñuelas would like to acknowledge the financial support from the European Research Council Synergy grant ERC-SyG-2013-610028 IMBALANCE-P

Petr Petřík was supported by long-term research development project RVO 67985939 (The Czech Academy of Sciences).

Oliver Phillips is supported by an ERC Advanced Grant 29158 ("T-FORCES") and is a Royal Society-Wolfson Research Merit Award holder.

Valério D. Pillar has been supported by the Brazil's National Council of Scientific and Technological Development (CNPq, grant 307689/2014-0).

Peter B. Reich was supported by United States Department of Energy (DE-SL0012677), NSF grant IIS-1563950 and two University of Minnesota Institute on the Environment Discovery Grants.

Franziska Schrodt was supported by a University of Minnesota Institute on the Environment Discovery Grant, a German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig grant (50170649_#7) and a University of Nottingham Anne McLaren Fellowship.

Jens-Christian Svenning considers this work a contribution to his VILLUM Investigator project "Biodiversity Dynamics in a Changing World" funded by VILLUM FONDEN.

Cyrille Violle was supported by the European Research Council (ERC) Starting Grant Project "Ecophysiological and biophysical constraints on domestication of crop plants" (Grant ERC-StG-2014-639706-CONSTRAINTS) by the French Foundation for Research on Biodiversity (FRB; www.fondationbiodiversite.fr) in the context of the CESAB project "Assembling, analysing and sharing data on plant functional diversity to understand the effects of biodiversity on ecosystem functioning: a case study with French Permanent Grasslands" (DIVGRASS).

Evan Weiher was funded by NSF DEB-0415383, UWEC-ORSP, and UWEC-BCDT.

We are indebted to Lukas Bruelheide for drawing the icons in Fig. 2 and 3. We would like to thank John Terborgh and Roel Brienen for contributing additional plot data.

951 Table 1: Traits used in this study and their function in the community. Traits are arranged
952 according to the degree to which they should respond to macroclimatic drivers. ↑↓ in the trait
953 column denotes opposing relationships, ⇳ in the description column denotes trade-offs. For
954 trait units, plot-level trait means and within-plot trait variance see Table 2.

Trait	Description	Function	Expected correlation with macroclimate
Specific leaf area, Leaf area, Leaf fresh mass, Leaf N, Leaf P ↑↓ Leaf dry matter content, Leaf N per area, Leaf C	Leaf economics spectrum[12-13,17]: Thin, N-rich leaves with high turnover and high mass-based assimilation rates ⇳ Thick, N-conservative, long-lived leaves with low mass-based assimilation rates	Productivity, competitive ability	Very high[3.4,17,21,23]
Stem specific density	Fast growth ⇳ Mechanical support, Longevity	Productivity, drought tolerance	Very high[3,22]
Conduit element length ↑↓ Stem conduit density	Efficient water transport ⇳ Safe water transport	Water use efficiency	High
Plant height	Mean individual height of adult plants	Competitive ability	High[3,5]
Seed number per reproductive unit ↑↓ Seed mass, Seed length, Dispersal unit length	Seed economics spectrum[23]: Small, well dispersed seeds ⇳ Seeds with storage reserve to facilitate establishment and increase survival	Dispersal, regeneration	Moderate[23-24]
Leaf N:P ratio	P limitation (N:P > 15) N limitation (N:P < 10)[29]	Nutrient supply	Moderate[30]
Leaf nitrogen isotope ratio (leaf $\delta^{15}N$)	Access to N derived from N_2 fixation ⇳ N supply via mycorrhiza	Nitrogen source, soil depth	Moderate[28]

955

956

Table 2: Traits, abbreviation of trait names, identifier in the Thesaurus Of Plant characteristics (TOP)[55], units of measurement, observed values (obs.) standardized effect sizes (SES) and significance (p) of SES for means and variances of both plot-level trait means (community-weighted means, CWMs) and within-plot trait variances (community-weighted variances, CWVs). CWMs and CWVs were based on gap-filled traits for 1,115,785 and 1,099,463 plots, respectively. All trait values were $\log_e$-transformed prior to analysis and observed values are on the $\log_e$ scale. SES are also based on $\log_e$-transformed values. Stem specific density is stem dry mass per stem fresh volume, specific leaf area is leaf area per leaf dry mass, leaf C, N and P are leaf carbon, nitrogen and phosphorus content, respectively, per leaf dry mass, leaf dry matter content is leaf dry mass per leaf fresh mass, leaf delta ^{15}N is the leaf nitrogen isotope ratio, stem conduit density is the number of vessels and tracheids per unit area in a cross section, conduit element length refers to both vessels and tracheids. SESs were calculated by randomizing trait values across all species globally 100 times and calculating CWM and CWV with random trait values, but keeping all species abundances in plots (see Fig. 1). Tests for significance of SES were obtained by fitting generalized Pareto-distribution of the most extreme random values and then estimating p values form this fitted distribution[50]. * indicates significance at $p < 0.05$.

				CWM						CWV					
				mean			variance			mean			variance		
Trait	Abbreviation	TOP	Unit	obs.	SES	*p*	obs.	SES	*p*	obs.	SES	*p*	obs.	SES	*p*
Leaf area	LA	25	mm^2	6.130	-9.75	*	1.691	12.53	*	1.565	-2.59	*	2.448	-0.27	n.s.
Specific leaf area	SLA	50	$m^2\ kg^{-1}$	2.850	9.89	*	0.172	12.88	*	0.150	-1.33	n.s.	0.023	1.10	n.s.
Leaf fresh mass	Leaf.fresh.mass	35	g	-2.125	-13.28	*	1.395	10.83	*	1.520	-2.05	*	2.311	0.01	n.s.
Leaf dry matter content	LDMC	45	$g\ g^{-1}$	-1.294	-5.67	*	0.101	11.52	*	0.130	0.95	n.s.	0.017	6.73	*
Leaf C	LeafC	452	$mg\ g^{-1}$	6.116	-3.77	*	0.003	8.80	*	0.002	-1.78	*	0.000	-0.38	n.s.
Leaf N	LeafN	462	$mg\ g^{-1}$	3.038	4.22	*	0.055	6.29	*	0.063	-3.19	*	0.004	-0.13	n.s.
Leaf P	LeafP	463	$mg\ g^{-1}$	0.535	9.57	*	0.097	2.81	*	0.117	-5.17	*	0.014	-2.11	*
Leaf N per area	LeafN.per.area	481	$g\ m^{-2}$	0.251	-9.06	*	0.075	8.18	*	0.099	-0.28	n.s.	0.010	1.54	n.s.
Leaf N:P ratio	Leaf.N:P.ratio	-	$g\ g^{-1}$	2.444	-11.95	*	0.040	0.40	n.s.	0.081	-2.74	*	0.007	-0.39	n.s.
Leaf $\delta^{15}N$	Leaf.delta15N	-	ppm	0.521	-3.58	*	0.254	6.68	*	0.455	2.82	*	0.207	2.44	*
Seed mass	Seed.mass	103	mg	0.407	-11.19	*	2.987	3.69	*	2.784	-9.06	*	7.750	-2.81	*
Seed length	Seed.length	91	mm	1.069	-4.51	*	0.294	5.50	*	0.365	-4.67	*	0.134	-3.07	*
Seed number per reproductive unit	Seed.num.rep.unit	-		6.179	7.67	*	2.783	4.40	*	5.156	1.44	n.s.	26.588	2.25	*
Dispersal unit length	Disp.unit.length	90	mm	1.225	-2.51	*	0.343	6.50	*	0.451	-3.21	*	0.203	-1.39	n.s.

Plant height	Plant.height	68	m	-0.315	-12.15	*	1.532	13.34	*	1.259	-9.01	*	1.585	9.68	*
Stem specific density	SSD	286	g cm^{-3}	-0.869	-14.93	*	0.041	13.15	*	0.058	2.09	*	0.003	2.99	*
Stem conduit density	Stem.cond.dens	-	mm^{-2}	4.407	15.08	*	0.656	8.45	*	0.975	-0.95	n.s.	0.951	1.10	n.s.
Conduit element length	Cond.elem.length	-	µm	5.946	-7.09	*	0.182	9.14	*	0.367	7.12	*	0.135	5.29	*
Mean SES					-3.50			8.06			-1.76			1.25	
Mean absolute SES					8.66			8.06			3.36			2.43	

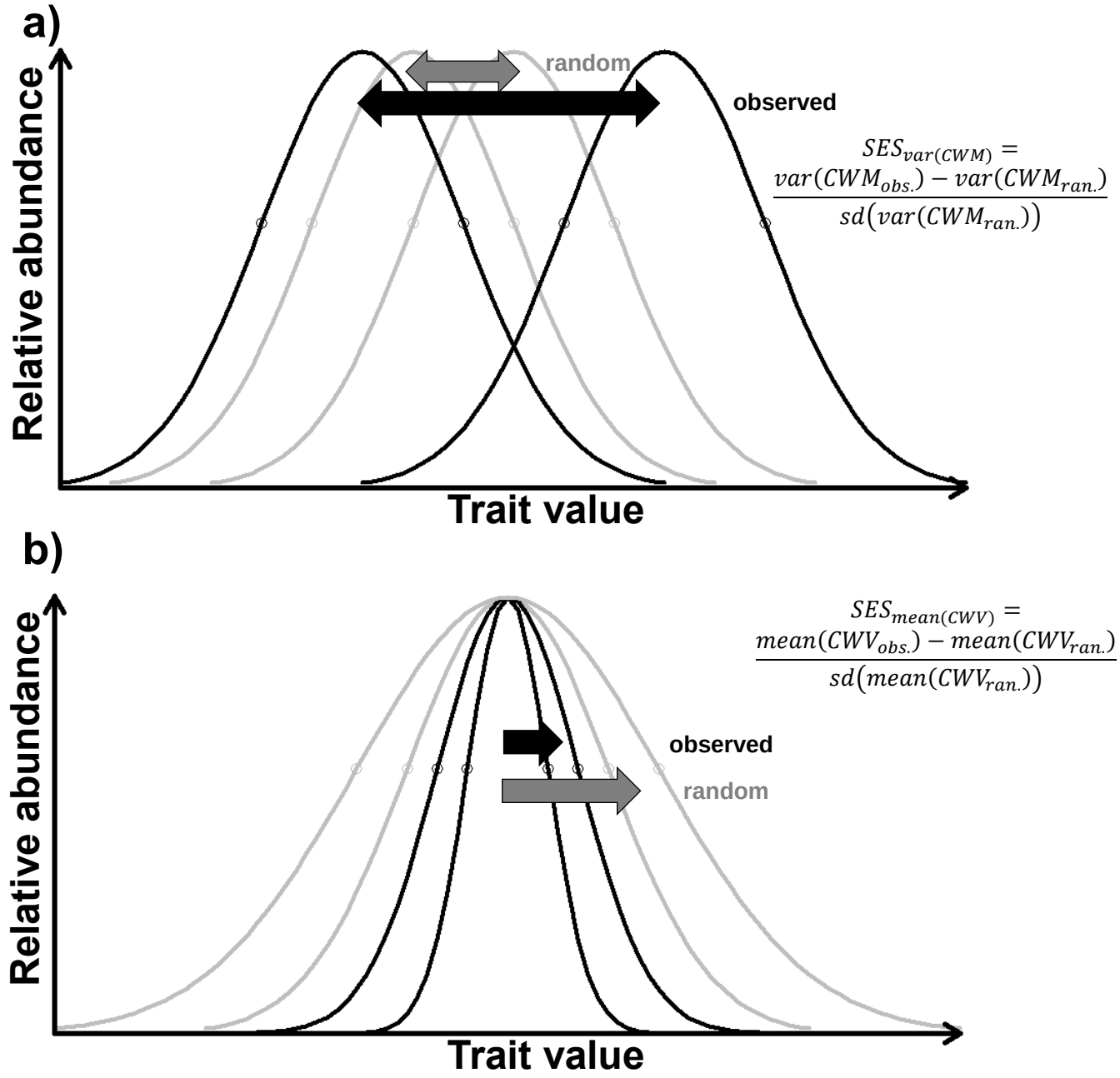


Fig. 1: Conceptual figure to illustrate Hypothesis 1, stating that environmental or biotic filtering of community trait values result in a) higher than expected variation of community-weighted means and b) lower than expected community-weighted variances of trait values. Both figures give an example for a single trait and show the relative abundance of trait values of all species in a plot. Black curves refer to observed plot-level trait values in two exemplary plots, while grey curves show plot-level trait values obtained from randomizing trait values across all species globally (see Methods). Randomization was done 100 times, but only one randomization event is shown. Deviation from random expectation was assessed with standardized effect sizes (SESs) for a) the variance in CWMs and b) for the mean in CWVs. Evidence for filtering is given in a) if the variance in plot-level trait means was higher than expected by chance (SES significantly positive) or b) if within-plot trait variance was typically lower than expected by chance (SES significantly negative, see Methods).

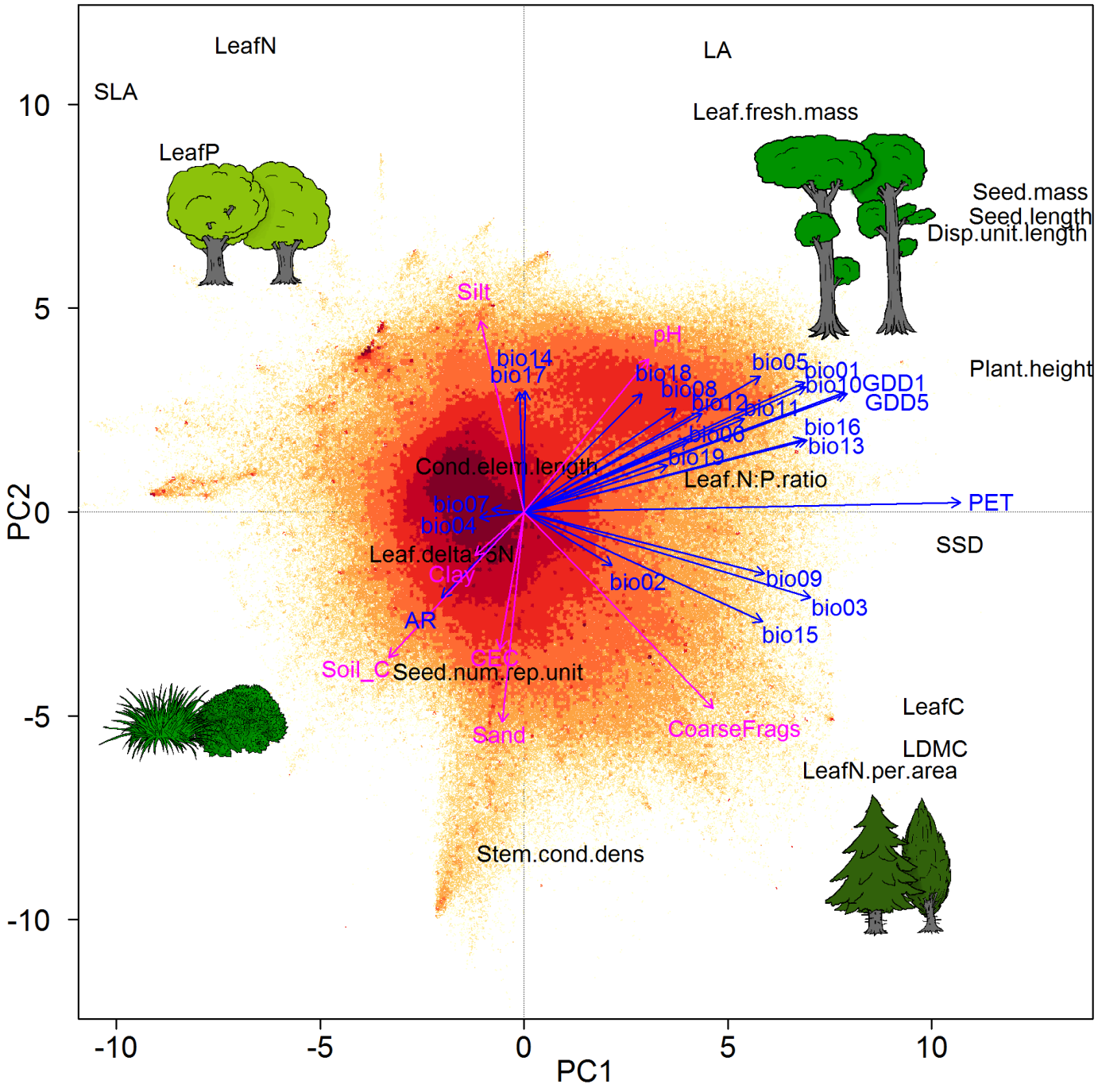


Fig. 2: Principal Component Analysis of global plot-level trait means (community-weighted means, CWMs). The plots (n=1,114,304) are shown by coloured dots, with shading indicating plot density on a logarithmic scale, ranging from yellow with 1–4 plots at the same position to dark red with 251–1142 plots. Prominent spikes are caused by a strong representation of communities with extreme trait values, such as heathlands with ericoid species with small leaf area and seed mass. Post-hoc correlations of PCA axes with climate and soil variables are shown in blue and magenta, respectively. Arrows are enlarged in scale to fit the size of the graph; thus, their lengths show only differences in variance explained relative to each other. Variance in CWM explained by the first and second axis was 29.7% and 20.1%, respectively. The vegetation sketches schematically illustrate the size continuum (short *vs.* tall) and the leaf

economics continuum (low *vs.* high LDMC and leaf N content per area in light and dark green colours, respectively). See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.

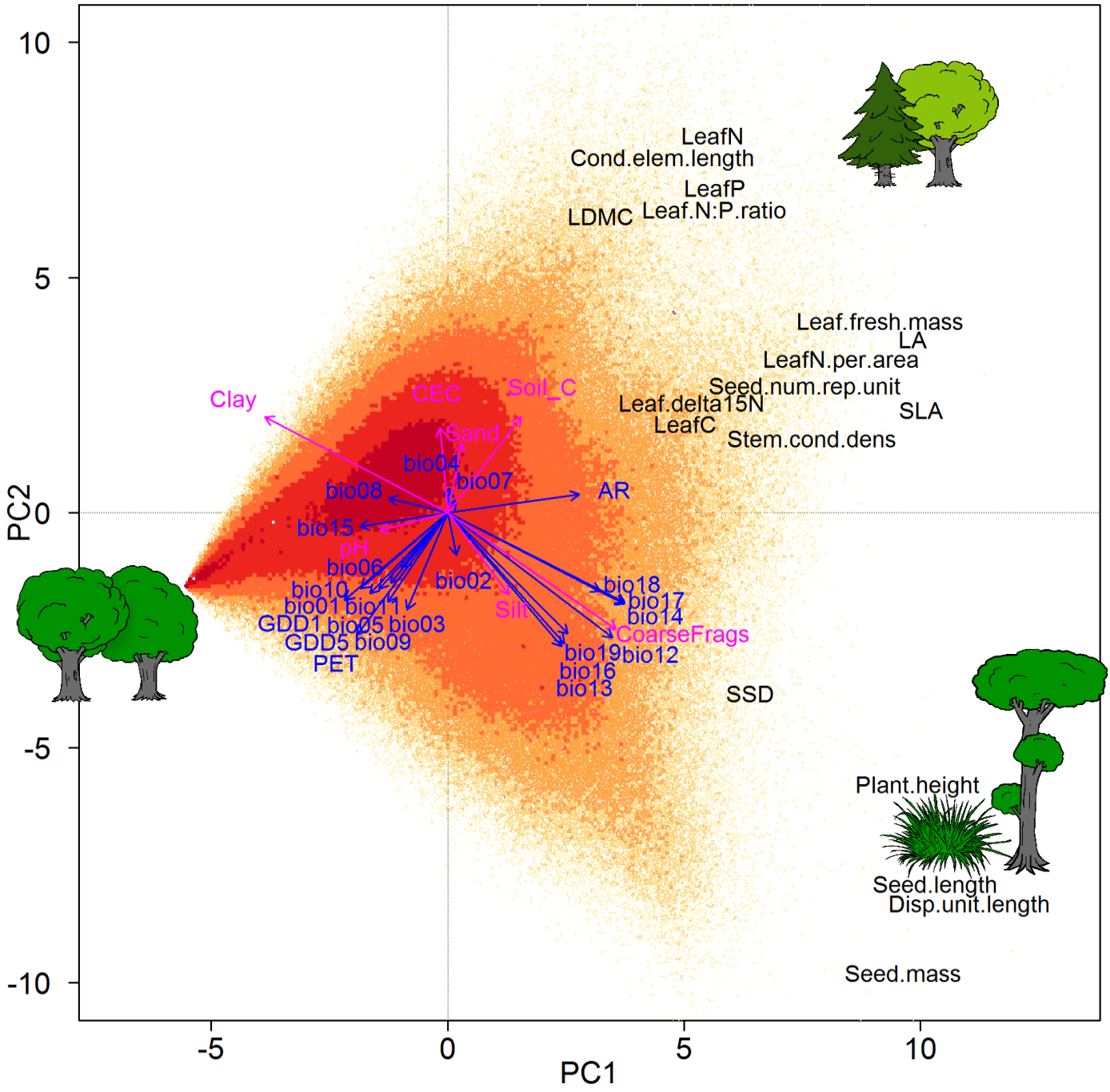


Fig. 3: Principal Component Analysis of global within-plot trait variances (community-weighted variances, CWVs). The plots (n=1,098,015) are shown by coloured dots, with shading indicating plot density on a logarithmic scale, ranging from yellow with 1–2 plots at the same position to dark red with 631–1281 plots. Post-hoc correlations of PCA axes with climate and soil variables are shown in blue and magenta, respectively. Arrows are enlarged in scale to fit the size of the graph; thus, their lengths show only differences in variance explained relative to each other. Variance in CWV explained by the first and second axis was 24.9% and 13.4%, respectively. CWV values of all traits increased from the left to the right, which reflects increasing species richness (r^2 = 0.116 between scores of the first axis and number of species in the communities for which traits were available). The vegetation

sketches schematically illustrate low and high variation in the plant size and leaf economics continua. See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.

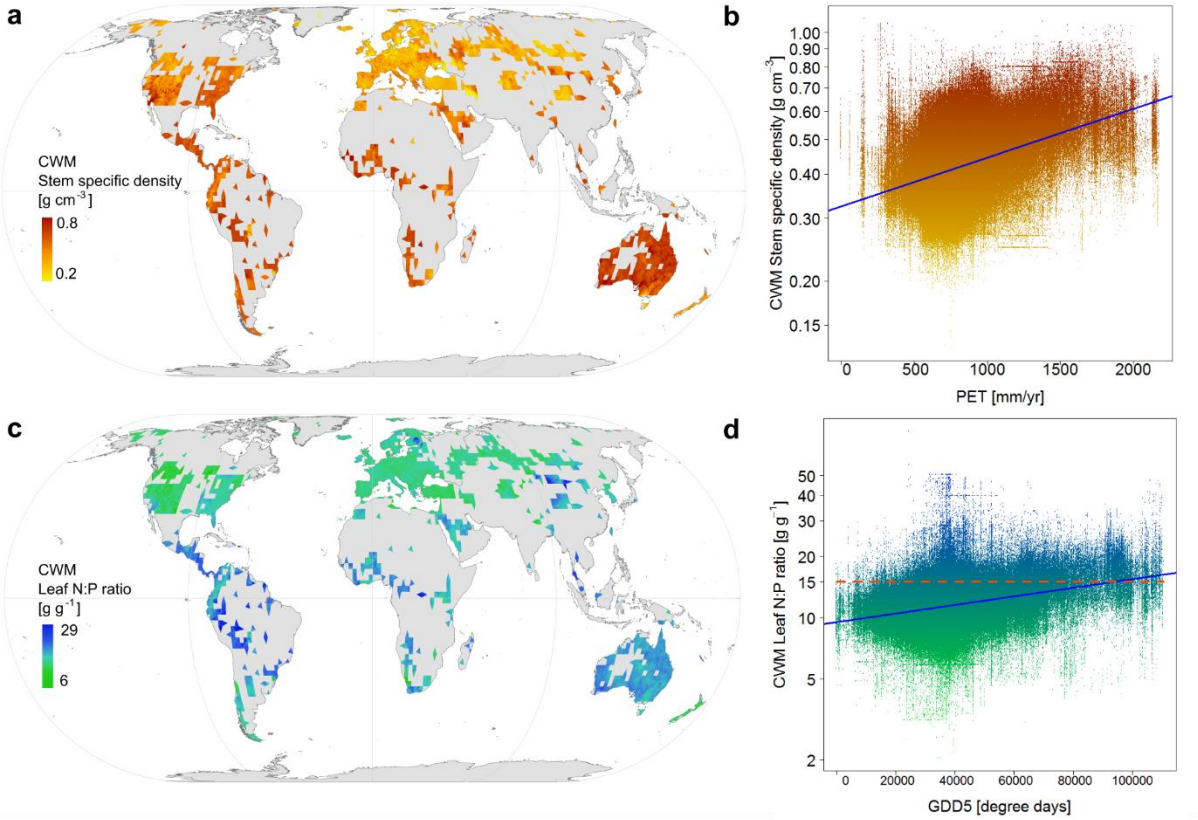


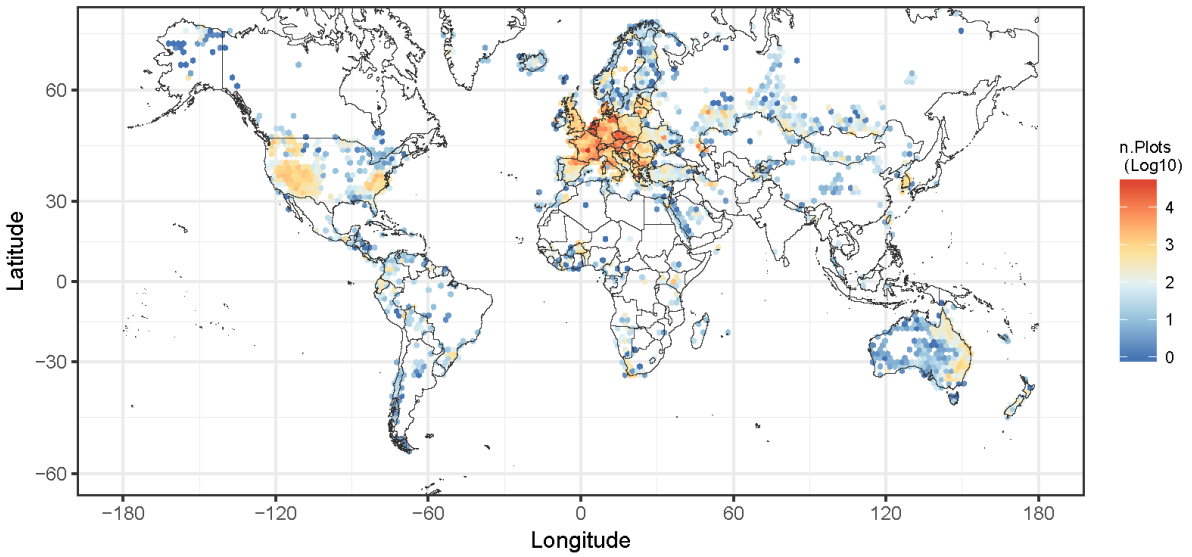
Fig. 4: The two strongest relationships found for global plot-level trait means (community-weighted means, CWMs) in the sPlot dataset. CWM of the natural logarithm of stem specific density [g cm^{-3}] as a) global map, interpolated by kriging within a radius of 50 km around the plots using a grid cell of 10 km, and b) function of potential evapotranspiration (PET, r^2=0.156). CWM of the natural logarithm of the N:P ratio [g g^{-1}] as c) global kriging map and d) function of the warmth of the growing season, expressed as growing degree days over a threshold of 5°C (GDD5, r^2=0.115). Plots with N:P ratios > 15 (of 2.71 on the log_e scale) tend to indicate phosphorus limitation[29] and are shown above the broken line in red colour (90,979 plots, 8.16% of all plots). The proportion of plots with N:P ratios > 15 increases with GDD5 (r^2=0.895 for a linear model on the log response ratio of counts of plots with N:P > 15 and ≤15 counted within bins of 500 GDD5).

Extended Data Table 1: Per cent coverage of the sPlot 2.1 database with original trait values, with respect to species for which original trait values were measured in TRY (of a total of 58,065 species in sPlot 2.1), to species × plot observations for which original trait values were available (of a total of 21,050,514 observations) and to plots (of a total of 1,104,219 plots for which coordinates and environmental information was available). For a comparison with gap-filled trait values, per cent coverage across all species is 45.87%, per cent coverage of all species × plot occurrences is 88.7%, and per cent coverage of plots is 100%.

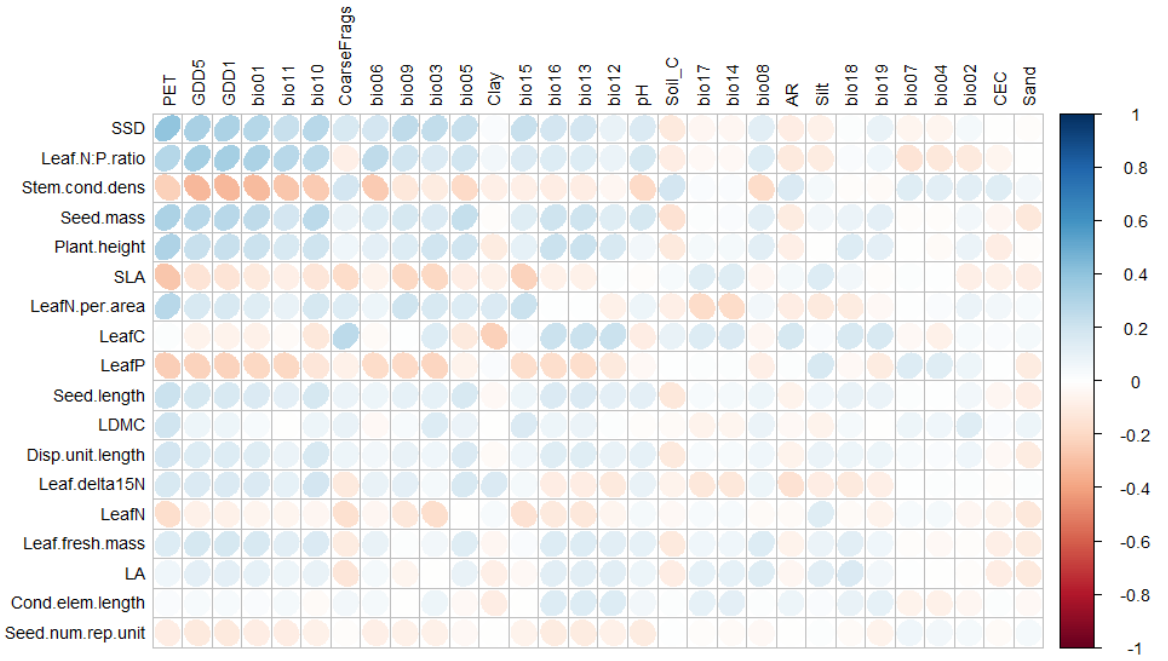
Trait	Abbreviation	Coverage of species %	Coverage of occurrences %	Coverage of plots %
Leaf area	LA	37.38	87.11	99.65
Specific leaf area	SLA	34.66	89.16	100.00
Leaf fresh mass	Leaf.fresh.mass	7.04	47.89	88.79
Leaf dry matter content	LDMC	15.89	81.94	97.78
Leaf C	LeafC	15.14	65.60	95.97
Leaf N	LeafN	28.27	77.57	99.16
Leaf P	LeafP	18.53	60.99	96.54
Leaf N per area	LeafN.per.area	18.51	60.78	94.98
Leaf N:P ratio	Leaf.N:P.ratio	12.53	45.32	93.58
Leaf $\delta^{15}N$	Leaf.delta15N	7.14	11.10	72.28
Seed mass	Seed.mass	59.64	91.18	99.65
Seed length	Seed.length	9.35	75.01	93.82
Seed number per reproductive unit	Seed.num.rep.unit	7.22	72.82	92.71
Dispersal unit length	Disp.unit.length	11.40	81.36	93.82
Plant height	Plant.height	58.03	96.58	99.90
Stem specific density	SSD	22.35	29.26	86.75
Stem conduit density	Stem.cond.dens	15.24	10.88	53.15
Conduit element length	Cond.elem.length	13.18	7.62	48.20

Extended Data Table 2: Environmental variables used as predictors. Climate data were obtained from CHELSA[38,39] (www.chelsa-climate.org), GDD1 and GDD5 were calculated from CHELSA data, based on monthly temperature and precipitation values for the years 1979–2013[40-41]. The index of aridity (AR) and potential evapotranspiration (PET) were extracted from the CGIAR-CSI website (www.cgiar-csi.org). Soil variables were obtained from the SOILGRIDS project (https://soilgrids.org/) and reflect mean values expected at 0.15 m depth.

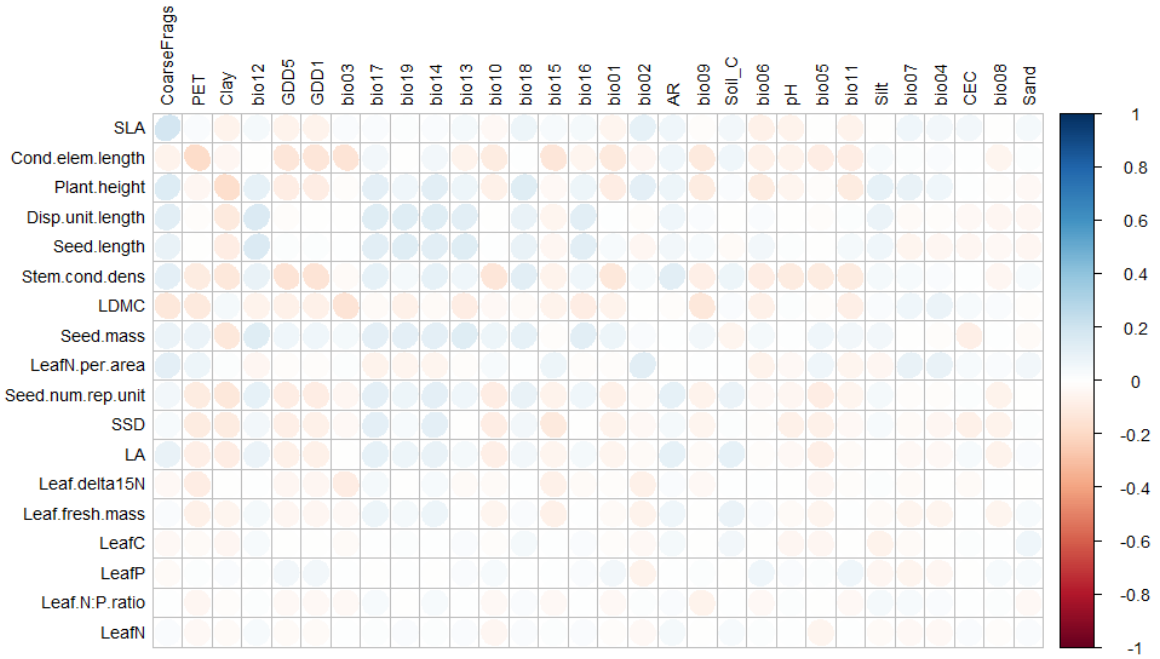
Variable	Abbreviation	Unit	Data sc
Annual Mean Temperature	Bio01	°C*10	CHELSA
Mean Diurnal Range (Mean of monthly (maximum temperature - minimum temperature))	Bio02	°C	CHELSA
Isothermality (bio2/bio7) (* 100)	Bio03	-	CHELSA
Temperature Seasonality (standard deviation of monthly temperature averages)	Bio04	°C*100	CHELSA
Max Temperature of Warmest Month	Bio05	°C*10	CHELSA
Min Temperature of Coldest Month	Bio06	°C*10	CHELSA
Temperature Annual Range (bio5-bio6)	Bio07	°C*10	CHELSA
Mean Temperature of Wettest Quarter	Bio08	°C*10	CHELSA
Mean Temperature of Driest Quarter	Bio09	°C*10	CHELSA
Mean Temperature of Warmest Quarter	bio10	°C*10	CHELSA
Mean Temperature of Coldest Quarter	bio11	°C*10	CHELSA
Annual Precipitation	bio12	mm/year	CHELSA
Precipitation of Wettest Month	bio13	mm/month	CHELSA
Precipitation of Driest Month	bio14	mm/month	CHELSA
Precipitation Seasonality	bio15	coefficient of variation	CHELSA
Precipitation of Wettest Quarter	bio16	mm/quarter	CHELSA
Precipitation of Driest Quarter	bio17	mm/quarter	CHELSA
Precipitation of Warmest Quarter	bio18	mm/quarter	CHELSA
Precipitation of Coldest Quarter	bio19	mm/quarter	CHELSA
Growing degree days above 1°C	GDD1	°C days	calculat
Growing degree days above 5°C	GDD5	°C days	calculat
Index of aridity	AR	(*10,000)	CGIAR-C
Potential evapotranspiration	PET	mm/year	CGIAR-C
Cation exchange capacity of soil	CEC	$cmol_c\ kg^{-1}$	SOILGRI
Soil pH	pH	(*10)	SOILGRI
Coarse fragment volume	CoarseFrags	vol. %	SOILGRI
Soil organic carbon content in the fine earth fraction	Soil_C	$g\ kg^{-1}$	SOILGRI
Clay content (0–2 µm)	Clay	mass fraction %	SOILGRI
Silt content (2–50 µm)	Silt	mass fraction %	SOILGRI
Sand content (50–2000 µm)	Sand	mass fraction %	SOILGRI



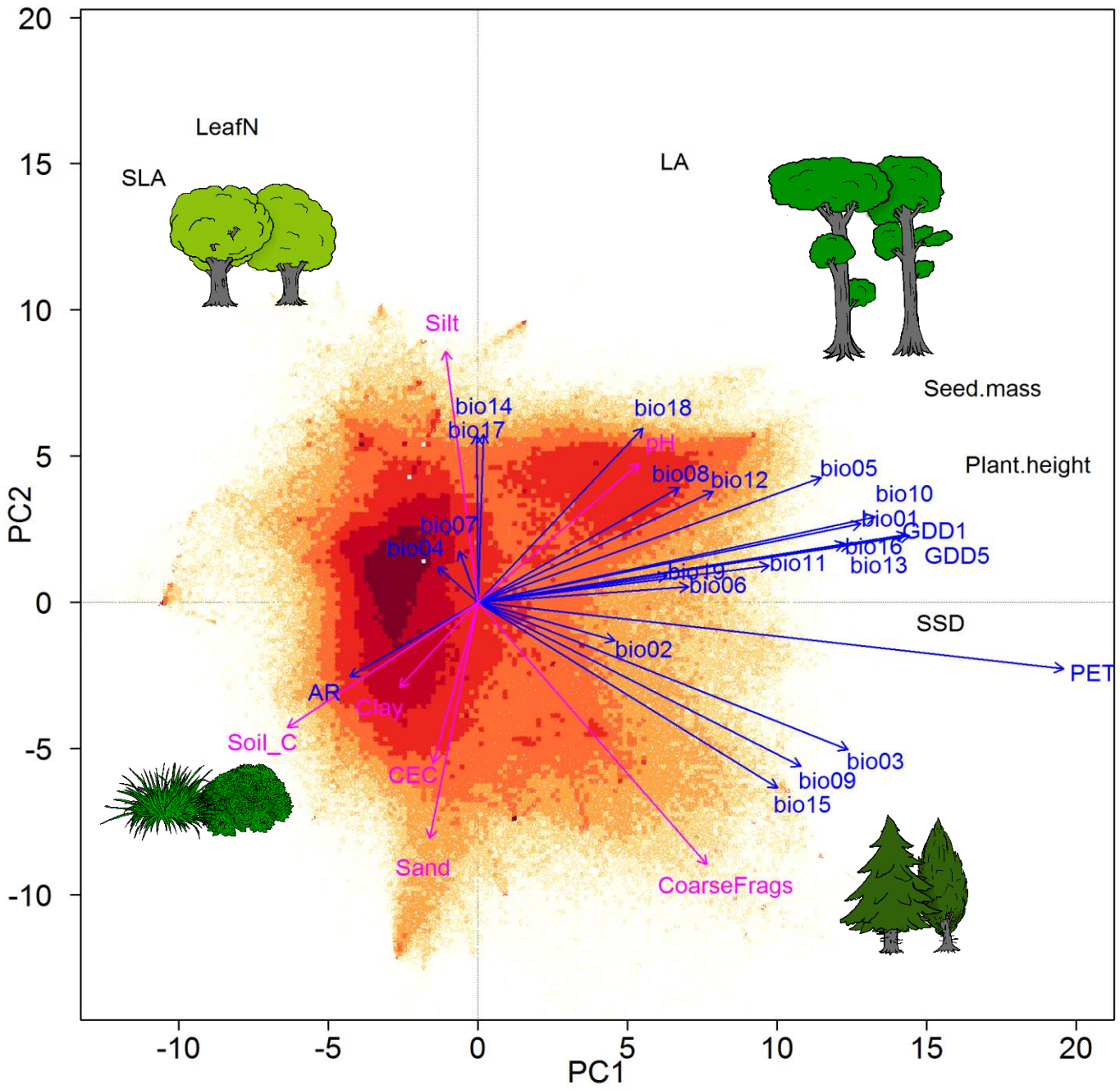
Extended Data Fig. 1: Distribution of plots in sPlot 2.1. The map shows plot density in a Mercator projection with a hexagonal grid with a radius of 120.14 km, corresponding to 5000 km^2 per grid cell at the equator. Hexagons at 60° latitude have a size of 1250 km^2.



Extended Data Fig. 2: Visualisation of the Pearson correlation matrix of plot-level trait means (community-weighted means, CWMs) of all 18 traits (rows) in the entire dataset (n = 1,114,304) with all 30 environmental predictors (columns). Positive correlations are shown in blue, negative ones in red colour, with increasing colour intensity as the correlation value moves away from 0. The eccentricity of the ellipses is scaled to the absolute value of the correlation[51]. Rows and columns are arranged from top to bottom and from left to right according to decreasing absolute correlation values. The highest correlation coefficient (between stem specific density and PET) was 0.395 (r^2=0.156). The best predictors for the plant height and seed mass trade-off were potential evapotranspiration (PET) and growing degree days above 5°C (GDD5), with r^2=0.093 and 0.052 for plant height and r^2=0.099 and 0.074 for seed mass, respectively. The best predictors for traits of the leaf economics spectrum were PET and the seasonality in precipitation (bio15), with r^2=0.078 and 0.051 for specific leaf area (SLA) and r^2=0.039 and 0.024 for leaf dry matter content (LDMC), respectively. See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



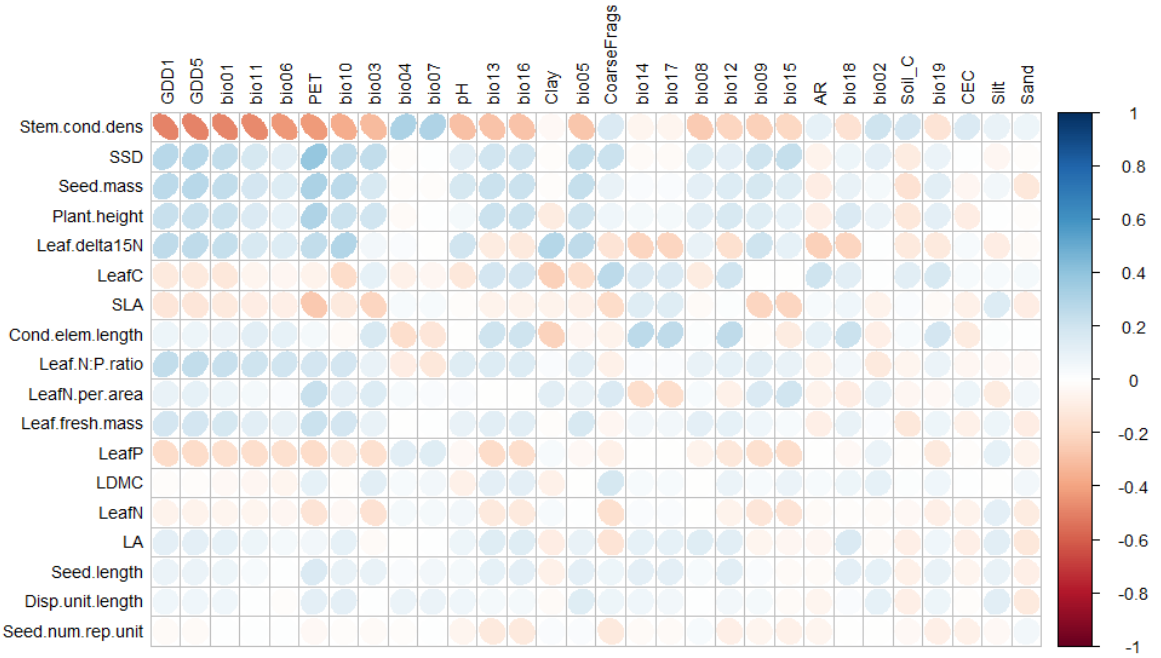
Extended Data Fig. 3: Visualisation of the Pearson correlation matrix of within-plot trait variances (community-weighted variances, CWVs) of all 18 traits (rows) in the entire dataset (n = 1,098,015) with all environmental predictors (columns). Positive correlations are shown in blue, negative ones in red colour, with increasing colour intensity as the correlation value moves away from 0. The eccentricity of the ellipses is scaled to the absolute value of the correlation[51]. Rows and columns are arranged from top to bottom and from left to right according to decreasing absolute correlation values. The highest correlation coefficient was encountered between specific leaf area (SLA) and the volumetric content of coarse fragments in the soil CoarseFrags, r^2=0.036), followed by the correlation of PET to CWV of conduit element length (r^2=0.035). See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



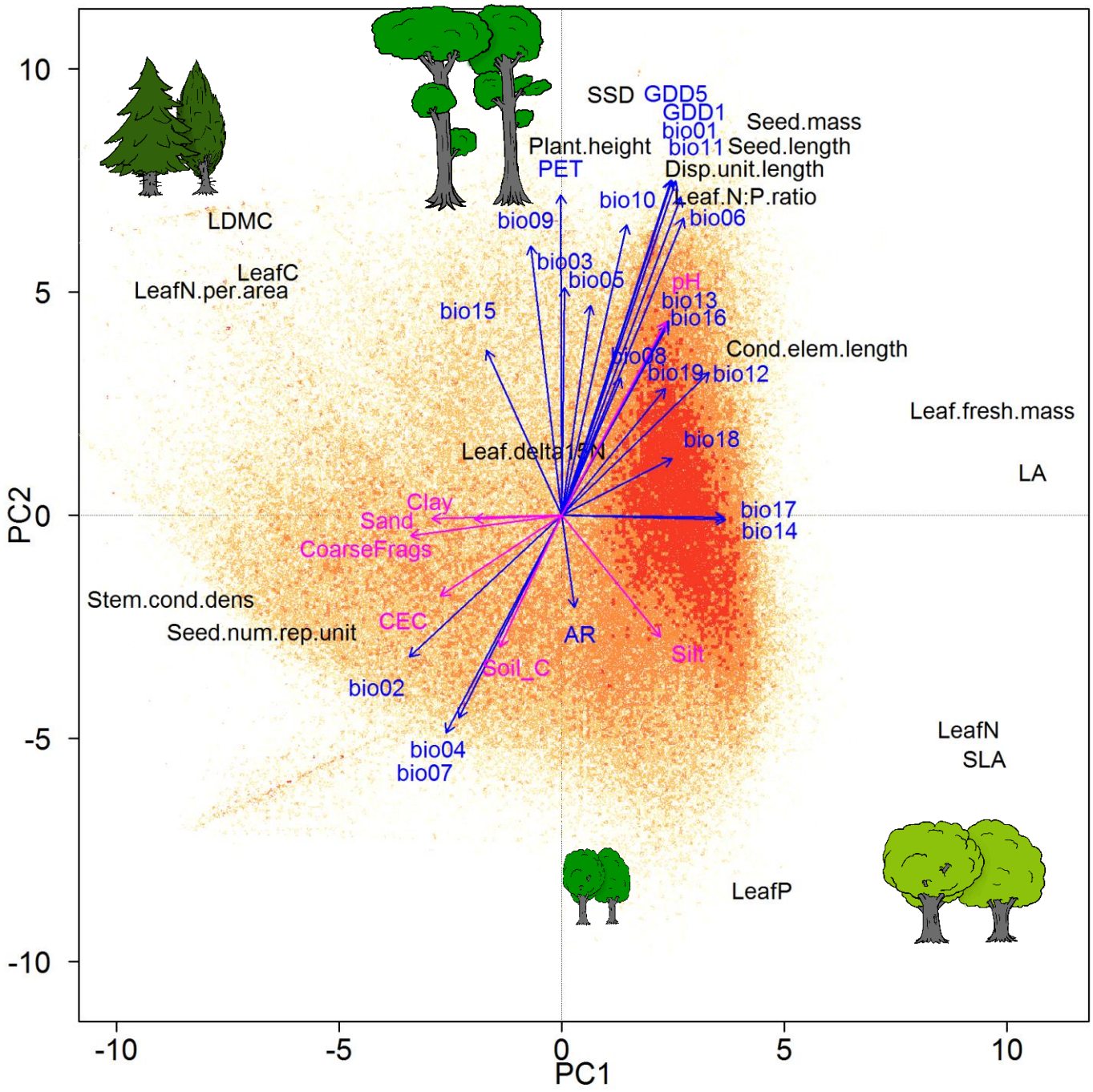
117

118 Extended Data Fig. 4: Principal Component Analysis of global plot-level trait means
119 (community-weighted means, CWMs), based on the original trait values measured for the
120 species from the TRY database for the six traits used by Díaz et al.[1] (leaf area, specific leaf
121 area, leaf N, seed mass, plant height and stem specific density). The plots (n = 954,459) are
122 shown by coloured dots, with shading indicating plot density on a logarithmic scale, ranging
123 from yellow with 1–8 plots at the same position to dark red with 501–1626 plots. Post-hoc
124 correlations of PCA axes with climate and soil variables are shown in blue and magenta,
125 respectively. Arrows are enlarged in scale to fit the size of the graph; thus, their lengths show
126 only differences in variance explained relative to each other. Variance in CWM explained by
127 the first and second axis was 43.5% and 30.9%, respectively. The vegetation sketches
128 schematically illustrate the size continuum (short *vs.* tall) and the leaf economics continuum

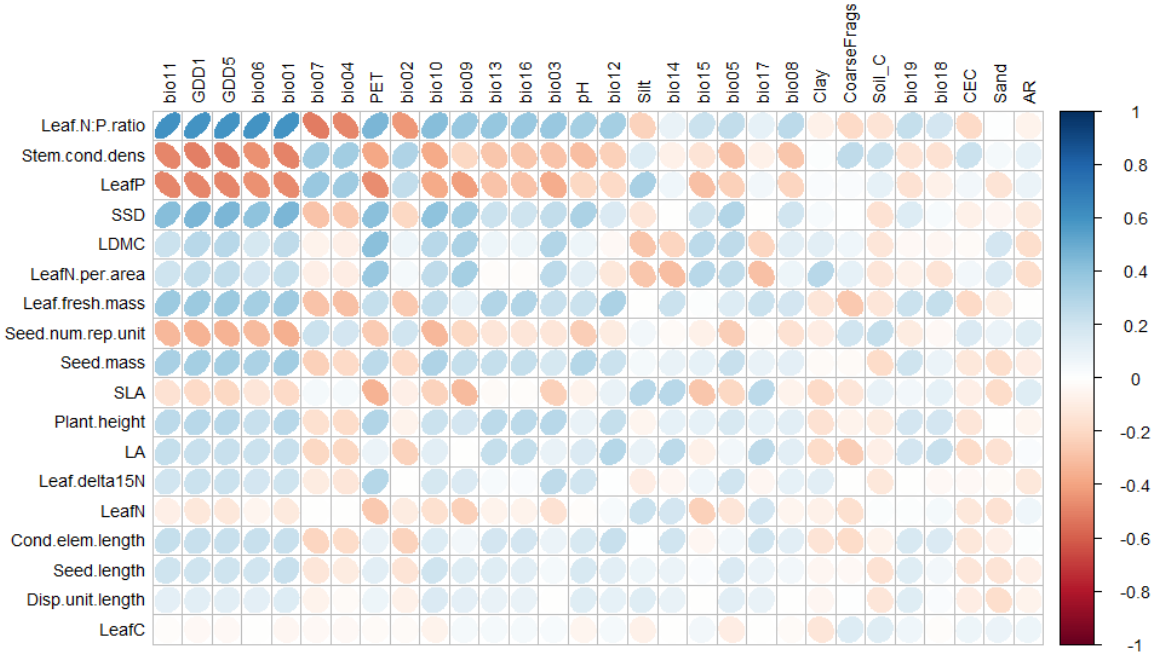
(low *vs.* high SLA and leaf N content per dry mass in dark and light green colours, respectively). See Table 1, 2 and Extended Data Tables 2 for the description of traits and environmental variables and compare with Fig. 2 for the same analyses with 18 traits based on gap-filled trait-data.



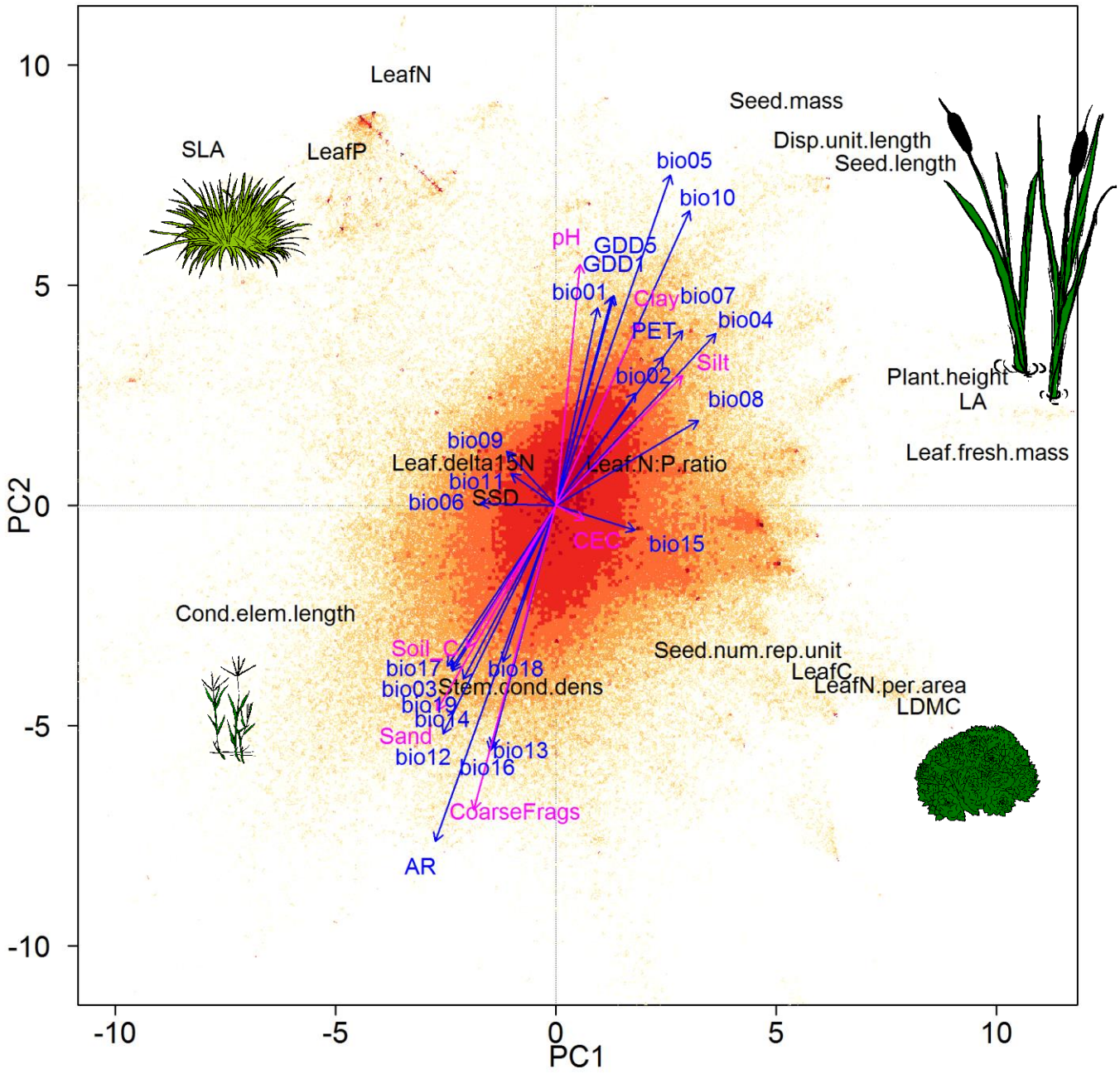
Extended Data Fig. 5: Visualisation of the Pearson correlation matrix of plot-level trait means (community-weighted means, CWMs) of all 18 traits (rows) based on the original trait values measured for the species from the TRY database in the entire dataset (n = 1,104,219) with all 30 environmental predictors (columns). Positive correlations are shown in blue, negative ones in red colour, with increasing colour intensity as the correlation value moves away from 0. The eccentricity of the ellipses is scaled to the absolute value of the correlation[51]. Rows and columns are arranged from top to bottom and from left to right according to decreasing absolute correlation values. The highest correlation coefficient was encountered for Stem conduit density and growing degree days above 1°C (GDD1, r^2=0.242), with similarly high coefficients of determination for growing degree days above 5°C (GDD5), mean annual temperature (bio1) and mean temperature of the coldest quarter (bio 11). There was also a high correlation of stem specific density and PET (r^2=0.152). The best predictors for the plant height and seed mass trade-off were potential evapotranspiration (PET) and growing degree days above 5°C (GDD5), with r^2=0.093 and 0.051 for plant height and r^2=0.099 and 0.074 for seed mass, respectively. The best predictors for traits of the leaf economics spectrum were PET and the seasonality in precipitation (bio15), with r^2=0.068 and 0.047 for specific leaf area (SLA), respectively. See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



Extended Data Fig. 6: Principal Component Analysis of plot-level trait means (community-weighted means, CWM) of forest communities only in the dataset. The plots (n = 330,873) are shown by coloured dots, with shading indicating plot density on a logarithmic scale, ranging from yellow with 1–4 plots at the same position to dark orange with 32–453 plots. Post-hoc correlations of PCA axes with climate and soil variables are shown in blue and magenta, respectively. Arrows are enlarged in scale to fit the size of the graph; thus, their lengths show only differences in variance explained relative to each other. Variance in CWM explained by the first and second axis was 32.9% and 27.6%, respectively. The vegetation sketches schematically illustrate low and high variation in the plant size and leaf economics continua. See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



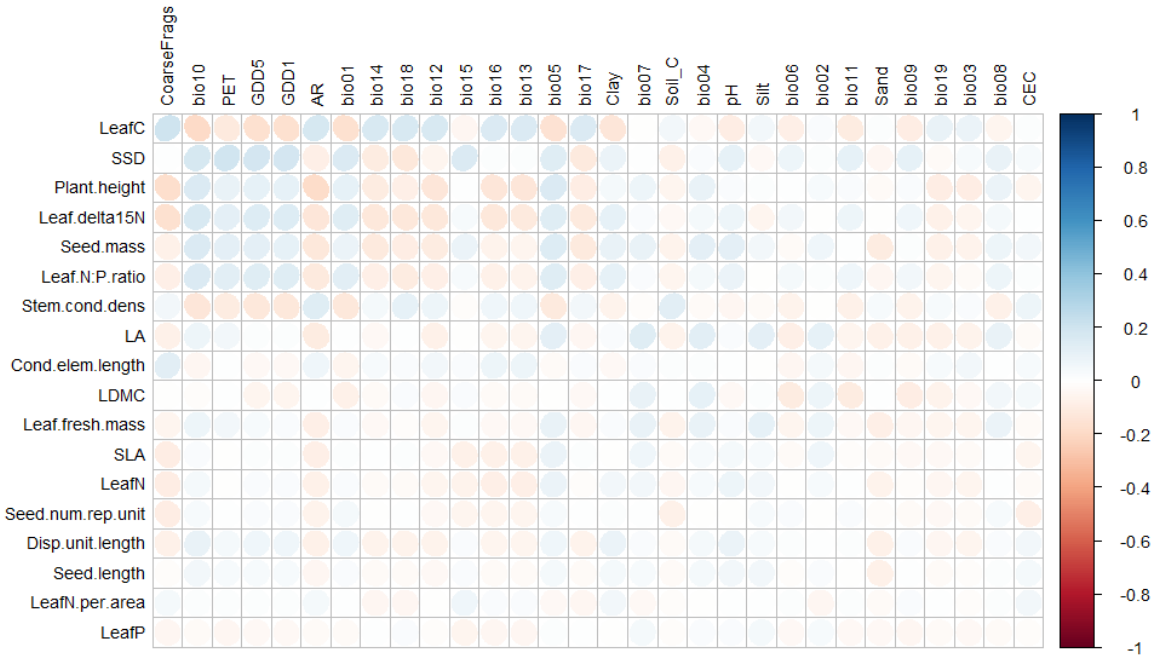
Extended Data Fig. 7: Visualisation of the Pearson correlation matrix of plot-level trait means (community-weighted means, CWMs) of all 18 traits (rows) of forest communities only in the dataset (n = 330,873) with all environmental predictors (columns). Positive correlations are shown in blue, negative ones in red colour, with increasing colour intensity as the correlation value moves away from 0. The eccentricity of the ellipses is scaled to the absolute value of the correlation[51]. Rows and columns are arranged from top to bottom and from left to right according to decreasing absolute correlation values. The highest correlation coefficient (between leaf N:P ratio and the mean temperature of coldest quarter (bio11)) was 0.607 (r^2=0.369). See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



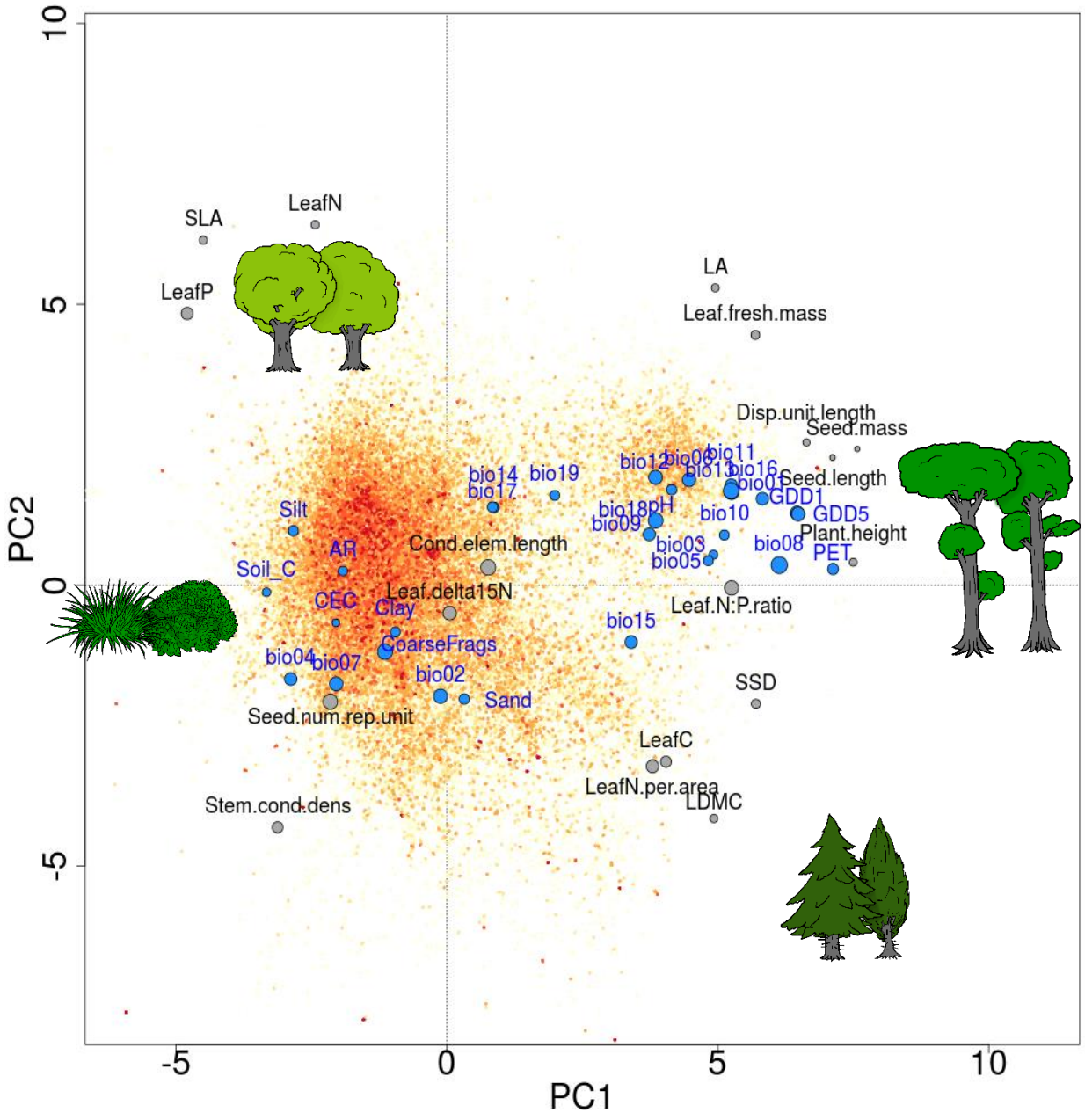
178

179 Extended Data Fig. 8: Principal Component Analysis of plot-level trait means (community-
180 weighted means, CWMs) of non-forest communities only in the dataset. The plots (n =
181 513,035) are shown by coloured dots, with shading indicating plot density on a logarithmic
182 scale, ranging from yellow with 1–4 plots at the same position to dark red with 251–1111
183 plots. Post-hoc correlations of PCA axes with climate and soil variables are shown in blue and
184 magenta, respectively. Arrows are enlarged in scale to fit the size of the graph; thus, their
185 lengths show only differences in variance explained relative to each other. Variance in CWM
186 explained by the first and second axis was 24.3% and 17.5%, respectively. The vegetation
187 sketches schematically illustrate low and high variation in the plant size and leaf economics
188 continua. See Table 2 and Extended Data Table 2 for the description of traits and
189 environmental variables.

190

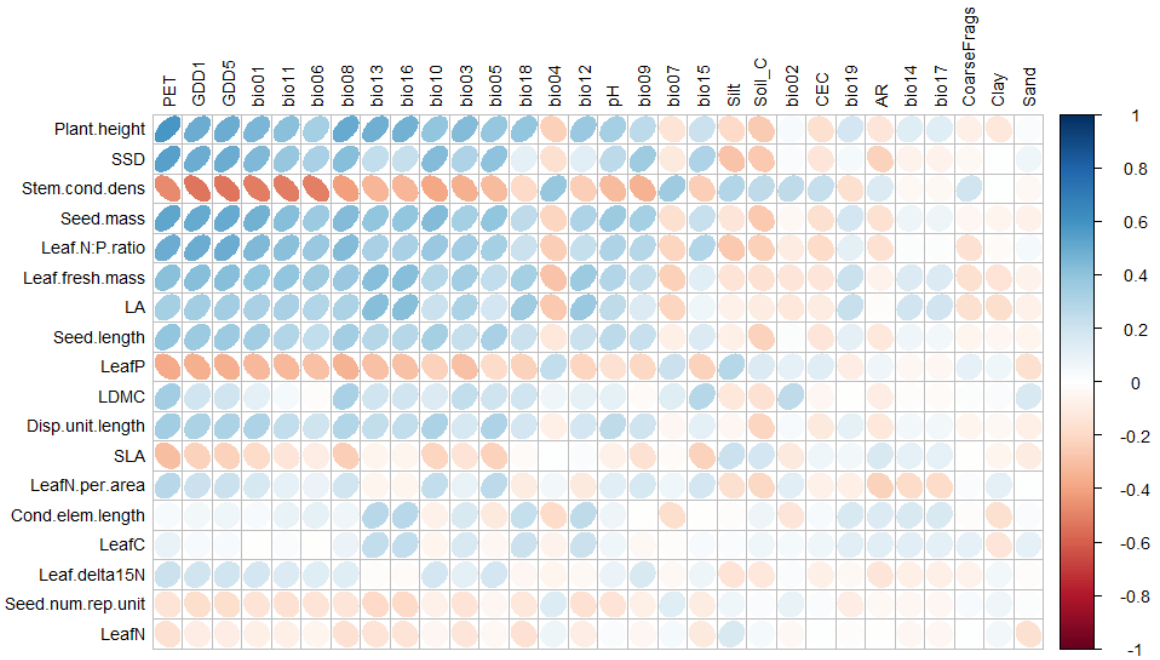


Extended Data Fig. 9: Visualisation of the Pearson correlation matrix of plot-level trait means (community-weighted means, CWMs) of all 18 traits (rows) of non-forest communities only in the dataset (n = 513,035) with all environmental predictors (columns). Positive correlations are shown in blue, negative ones in red colour, with increasing colour intensity as the correlation value moves away from 0. The eccentricity of the ellipses is scaled to the absolute value of the correlation[51]. Rows and columns are arranged from top to bottom and from left to right according to decreasing absolute correlation values. The highest correlation coefficient (between leaf C content per dry mass and the volumetric content of coarse fragments in the soil (CoarseFrags)) was 0.204 (r^2=0.042). See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



Extended Data Fig. 10: Summary of Principal Components Analyses applied to 100 resampled subsets of plot-level trait means (community-weighted means, CWMs) from the entire dataset for all 18 traits in the sPlot dataset. Each subset was resampled from the global environmental space (see Methods) and comprised between 99,342 and 99,400 (mean 99,380) plots. The coloured dots show the plots of one random example of these 100 subsets, with shading indicating plot density on a logarithmic scale, ranging from yellow with 1–3 plots at the same position to red with 10–81 plots in the subset. The loadings of each of the traits are displayed by a grey circle, its radius scaled to the range of loadings on PC1 and PC2 of all 100 runs. Post-hoc regressions of PCA axes with each of the environmental variables are illustrated by blue circles, its radius scaled to the range of correlations with PC1 and PC2. The circles are rather small, indicating that both the loadings and the post-hoc correlations with the environment had very similar values in the different runs. The mean variance in CWM explained by the first and second axis across the 100 runs was 33.4% ± 0.04 sd and 17.5% ± 0.03 sd, respectively. The vegetation sketches schematically illustrate low and high variation

in the plant size and leaf economics continua. See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.



Extended Data Fig. 11: Visualisation of the mean Pearson correlation coefficients of plot-level trait means (community-weighted means, CWMs) of all 18 traits (rows) with all environmental predictors (columns) of the 100 resampled subsets. Each subset was resampled from the global environmental space (see Methods) and comprised between 99,342 and 99,400 (mean 99,379.5) plots. Positive correlations are shown in blue, negative ones in red colour, with increasing colour intensity as the correlation value moves away from 0. The eccentricity of the ellipses is scaled to the absolute value of the correlation[51]. Rows and columns are arranged from top to bottom and from left to right according to decreasing absolute mean correlation values. The highest mean correlation coefficient (between plant height and potential evapotranspiration (PET) was 0.585 (r^2=0.342). See Table 2 and Extended Data Table 2 for the description of traits and environmental variables.