HAL open science

Tree functional traits, forest biomass, and tree species diversity interact with site properties to drive forest soil carbon

Laurent Augusto, Antra Boča

▶ To cite this version:

Laurent Augusto, Antra Boča. Tree functional traits, forest biomass, and tree species diversity interact with site properties to drive forest soil carbon. Nature Communications, 2022, 13 (1), pp.1097. 10.1038/s41467-022-28748-0 . hal-03627098

HAL Id: hal-03627098

https://hal.inrae.fr/hal-03627098v1

Submitted on 1 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-022-28748-0 OPEN

Tree functional traits, forest biomass, and tree species diversity interact with site properties to drive forest soil carbon

Laurent Augusto [1✉] & Antra Boča [2✉]

Forests constitute important ecosystems in the global carbon cycle. However, how trees and environmental conditions interact to determine the amount of organic carbon stored in forest soils is a hotly debated subject. In particular, how tree species influence soil organic carbon (SOC) remains unclear. Based on a global compilation of data, we show that functional traits of trees and forest standing biomass explain half of the local variability in forest SOC. The effects of functional traits on SOC depended on the climatic and soil conditions with the strongest effect observed under boreal climate and on acidic, poor, coarse-textured soils. Mixing tree species in forests also favours the storage of SOC, provided that a biomass over-yielding occurs in mixed forests. We propose that the forest carbon sink can be optimised by (i) increasing standing biomass, (ii) increasing forest species richness, and (iii) choosing forest composition based on tree functional traits according to the local conditions.

[1] INRAE, Bordeaux Sciences Agro, UMR 1391 ISPA, 33882 Villenave d'Ornon, France. [2] Latvia University of Life Sciences and Technologies, Lielā iela 2, Jelgava LV-3001, Latvia. ✉email: laurent.augusto@inrae.fr; antra.boca@llu.lv

Forests have been identified as a major leverage for climate change mitigation because of their capacity to fix atmospheric CO_2 and transform it as organic carbon, in biomass and then in soils[1,2]. On average, trees and soils store similar amounts of carbon in forest ecosystems[2,3]. Nevertheless, standing tree biomass reaches its limit in various regions of the world[4,5] and, contrary to soils, the carbon pool stored in standing biomass is highly vulnerable to many hazards like windthrow, drought, wildfire, pest, and disease (e.g.,[6,7]). In a context of global change-induced disturbances, storing carbon in forest soils thus appears as an approach less vulnerable than storing in the biomass of standing trees[8].

Massive planting of trees has been proposed as an efficient method to improve the carbon balance of lands[9,10]. The quantitative impact of large scale afforestations is however debatable[9,11], and the factors and management approaches affecting carbon sequestration in forests on a large scale have yet to be determined[3]. While the debate about planting more than a trillion trees has resulted in actions taken by international companies (e.g. www.trilliontreecampaign.org)[10], the influence of forest tree composition on soil organic carbon (SOC) is often not part of the debate as it is still elusive despite decades of research[3,12,13] even though forest composition strongly influences ecosystem functioning[14,15]. Apart from the uppermost soil layer known as the *forest floor* (which is the accumulation of decomposing dead foliage), no general and consistent trend has been isolated for non-nitrogen-fixing species[3,12,16–18]. Consequently, this study aimed at the identification of tree species and their characteristics that enlarge SOC pools and increase SOC stability. Due to the high number of extant tree species, and minimal recurrence of individual species among regions of the world, we did not compare species directly but quantified the influence on SOC of the main plant traits of the tree species (e.g., foliage nitrogen content).

Based on published findings, we formulated three hypotheses. First, because the properties of plant debris (such as litterfall and dead roots) strongly influence both their initial decomposition[19–23] and the subsequent stabilisation of organic carbon in soil[24–28], we hypothesised [H1] that plant traits drive SOC, in pool size and in stability. In particular, we anticipated that gymnosperm species increase SOC accumulation whereas angiosperm species favour SOC stabilisation[12]. We further hypothesised [H2] that the imprint of tree species traits, and standing biomass, on SOC is in interaction with site properties[12,29,30] because the way organic matter is accumulated –or not– depends on processes (i.e., decomposition and mineralisation versus microbial processing, chemical binding, and soil aggregation) whose relative importance varies with local conditions[1]. For instance, SOC stabilisation through microbial processing of organic matter is favoured by soil nutrient availability[31–33], implying that the influence of plant nutrient content (notably in nitrogen (N) and phosphorus (P)) is highest in nutrient poor soils[31,34]. We assumed that climate (temperature and water availability), soil fertility, soil texture (sand content and clay content), and soil pH are the site properties that are the most likely to affect the influence of plant traits on SOC[1,12,14,35–37]. Finally, we hypothesised that forest standing biomass [H3a] and tree diversity [H3b] increase the SOC pool size, respectively directly and indirectly. Indeed, high levels of standing biomass imply large fluxes of plant necromass, which in turn can increase the quantity of organic matter in soils[1,38–41]. Similarly, because the so-called mixed forests (i.e., forests composed of several tree species) have on average a higher growth rate than mono-specific forests[42–45], we expected that they store more SOC through their positive effect on standing biomass.

To address these hypotheses, we compiled, aggregated and harmonised published data on forest SOC and tree traits (see "Methods"), and conducted a global analysis based on 454 mature and mono-specific forest stands clustered in 136 sites worldwide. The stands at each site were in comparable conditions, enabling us to compare the influence of 178 different tree species, from 35 families, on SOC pools. Data about mixed stands or SOC stability were extracted from original publications for 29 and 30 sites, respectively.

Results and discussion

At the global scale, the SOC stocks are strongly influenced by climate, soil properties, and nitrogen atmospheric deposition (Supplementary Fig. S1), as already reported in the literature[46–48]. Because of the strong imprint of the environment on SOC when comparing different sites, and because our objective was to quantify the influence of plant traits of different tree species growing in the same conditions, we normalised SOC values relative to the mean C pool of a site before further analysis. Similarly, the functional trait values and the properties of the forest stands (e.g., aboveground biomass) were normalised per site to take into account the site to site variations in tree species composition. This means that the results reported in the present study showed the effect of forest properties (plant trait values and stand biomass) after the influence of the environment was removed. Positive, null, or negative normalised values indicated absolute values that were respectively above, equal, or below the site mean value (see "Data handling and normalisation" in Methods).

Our results showed that the SOC pool size (forest floor plus mineral soil layers) was negatively correlated with many plant functional traits, such as leaf nitrogen (N) content, specific leaf area (SLA), specific root length (SRL), and wood density, with the exception of leaf dry matter content (LDMC) that had a positive effect (Supplementary Fig. S2). Having numerous plant traits that are related to a single response variable is not surprising since all these traits form the well-known *Plant Economics Spectrum* (PES;[49,50]). The PES describes how all plant functions are dependent on each-other, implying trade-offs and high levels of correlation among traits[51], as observed in our study (Supplementary Fig. S3). Observed at the plant scale, these trade-offs among traits are also strong at the organ scale, defining the *Leaf Economics Spectrum*[52,53] and the *Root Economics Spectrum*[54,55]. Interestingly, while plant traits forming the PES had a strong effect on SOC ($|r| \geq 0.25$; see below), the phylogenetic distance among tree species, which represents life history characteristics, only poorly explained the differences in SOC (Supplementary Fig. S4). This lack of consistency was not totally surprising because, whereas the phylogenetic distance may explain the values distribution of a trait[56,57], it also often poorly correlates with important traits[58] and more generally with ecological difference[59].

Even though LDMC, SLA, and SRL had a high influence on SOC pools (Supplementary Table S1), we built our model using firstly the leaf photosynthetic maximum capacity (A_{max}) as it is an integrative trait of plant functioning[51], and because it is the best predictor of SOC (Fig. 1). A model based on A_{max} (Fig. 2A), and to a lower extent forest standing biomass, explained half (48.5%; adjusted r^2) of the variance associated with SOC pool size (48.5% = 37.0% for A_{max} + 11.5% for biomass; Supplementary Figs. S5A and S6; Supplementary Table S2), supporting our hypothesis H1 (see below more results about biomass). This model showed a satisfactory level of consistency using an independent dataset (Supplementary Fig. S5B). The influence of plant A_{max} on SOC was particularly clear in the forest floor (Supplementary Fig. S7A), but was also significant in the mineral soil (Supplementary Fig. S7B; Supplementary Table S1). Using an integrative

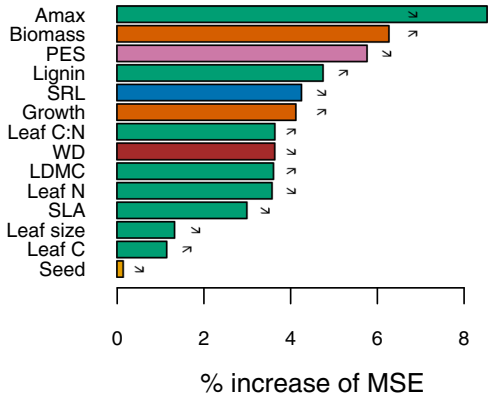

Fig. 1 Main variables explaining the SOC pools at the local scale. The SOC pool was modelled using relative values, normalised to site conditions. The predictors were: (i) leaf traits, in green: maximum photosynthetic capacity (A_{max}), C content, N content, C:N ratio, lignin content, leaf dry matter content (LDMC), leaf size and specific leaf area (SLA); (ii) other plant traits [seed mass (in yellow), wood density (WD; in brown), specific root length (SRL; in blue)]; (iii) the index score of the Plant Economics Spectrum (PES; in violet); and (iv) stand biomass dynamics, in dark orange (standing biomass; tree species growth). The influence of the variables was assessed using the percentage of increase of MSE from the Random Forest approach (see "Methods"). Arrows indicate positive (↗) or negative (↘) effects of the predictors on SOC. Source data are provided as a Source Data files.

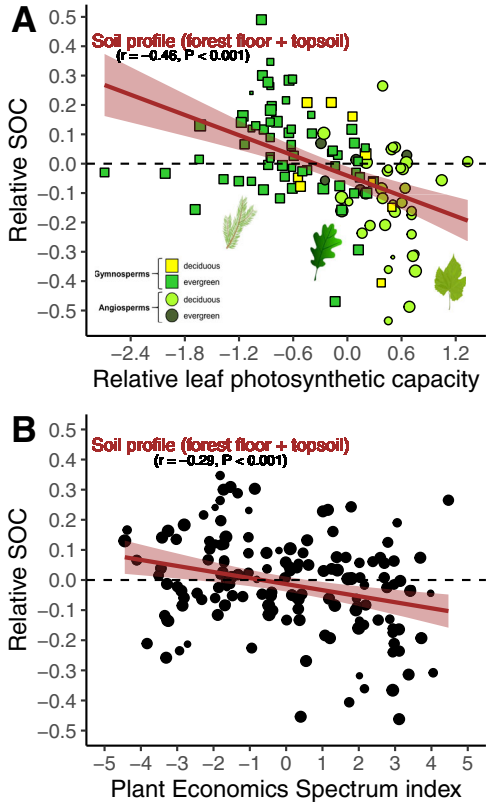

Fig. 2 Global influence of plant functional traits on SOC pool. Relationship between photosynthetic capacity of tree species (A_{max}) and SOC pool (**A**); relationship between the index score of the Plant Economics Spectrum (PES) and SOC pool (**B**). Values are normalised and the symbol size is proportional to data reliability (see "Methods"), which was taken into account as a weighting factor in the regression. A linear regression was fitted (level of confidence of the error band = 0.95). Source data are provided as a Source Data files.

index of the PES (based on the PCA approach with imputation; see Methods for more details) confirmed that not only tree A_{max} influenced SOC pools, but the whole combination of functional traits and species characteristics (Fig. 2B). On average, tree species with low values of leaf A_{max}, N and P content, wood density, SLA and SRL, but high value of LDMC, have large SOC pools. This combination of traits is typical of tree species with a conservative resource strategy (hereafter referred to as *conservative species*), such as many ectomycorrhizal gymnosperms (Supplementary Table S3;[60]), in contrast to arbuscular mycorrhizal angiosperms with an acquisitive resource strategy (hereafter: *acquisitive species*). We found that gymnosperm species store on average more SOC than angiosperm species, which could be attributed to them being generally conservative (Supplementary Fig. S8).

In addition to total SOC content, we studied several metrics of SOC decomposability. Even if they do not take into account all process involved in organic matter preservation such as leaching of dissolved organic carbon, these metrics enabled us to evaluate the effect of tree species on SOC stability. Our results indicated that plant functional types (e.g., mycorrhizal type) and tree traits influenced SOC stability with arbuscular mycorrhizal species, nitrogen-fixing species, and species with high SRL favouring SOC stability (Fig. 3A–C). The absence of clear effect of the spermaphyte type (Fig. 3D) may be explained by the fact that both mycorrhizal types could be found in angiosperms or gymnosperms, and that SRL values of angiosperms overlap with gymnosperms values[61]. The influence of some plant functional types on SOC stability supports the idea that species with nutrient-rich litters (i.e., arbuscular mycorrhizal species and N-fixers)[60] may favour chemical binding to soil minerals, soil aggregation, and microbial processing of organic matter[12,25,62], which in turn increases SOC stability[12,15,19,26]. But, because in our global dataset the SOC pool is on average smaller under these species as compared with under conservative species, it suggests that the influence of tree species on the amount of carbon storage in soils is generally more driven by accumulation of organic matter inherently recalcitrant to decomposition[23] than by SOC long-term stabilisation through chemical binding to soil minerals, or soil aggregation[1].

After finding important effects of tree traits on SOC, we explored whether this relationship could be modulated by environmental characteristics of the sites (e.g., climate or soil properties). Our best model, which considered A_{max} and biomass, was not improved by adding site properties (Supplementary Table S2; Supplementary Fig. S9). Because the original values of A_{max} were not available for all sites, we used the PES index as the predictor for further linear models (Fig. 2B). The main finding was that the control of the PES on SOC was in interaction with past land-use, climate, and soil properties (Supplementary Table S2; Fig. 4 and Supplementary Fig. S10), which supported our initial expectations (hypothesis H2). More precisely, when tree species were on sites with a past history of agricultural land-use or fertilisation, the influence of plant traits was not significant anymore (Supplementary Fig. S11)[63]. Similar to past land-use, the effect of the PES on SOC was also influenced by climate. The PES effect was observed under most climates except under warm-wet conditions (Supplementary Fig. S12A). Finally, the influence of the PES on SOC tended to decrease with soil fertility: finely-textured and neutral soils exhibited only weak trends, while sandy acidic soils showed significant relationships between SOC pools and plant traits (Supplementary Fig. S12B, C). It is probable that both agronomic practices (such as liming and fertilisation) and the tendency of farmers to select the best lands (such as non-acidic, finely textured, soils)[64–66] explain why past land-use and

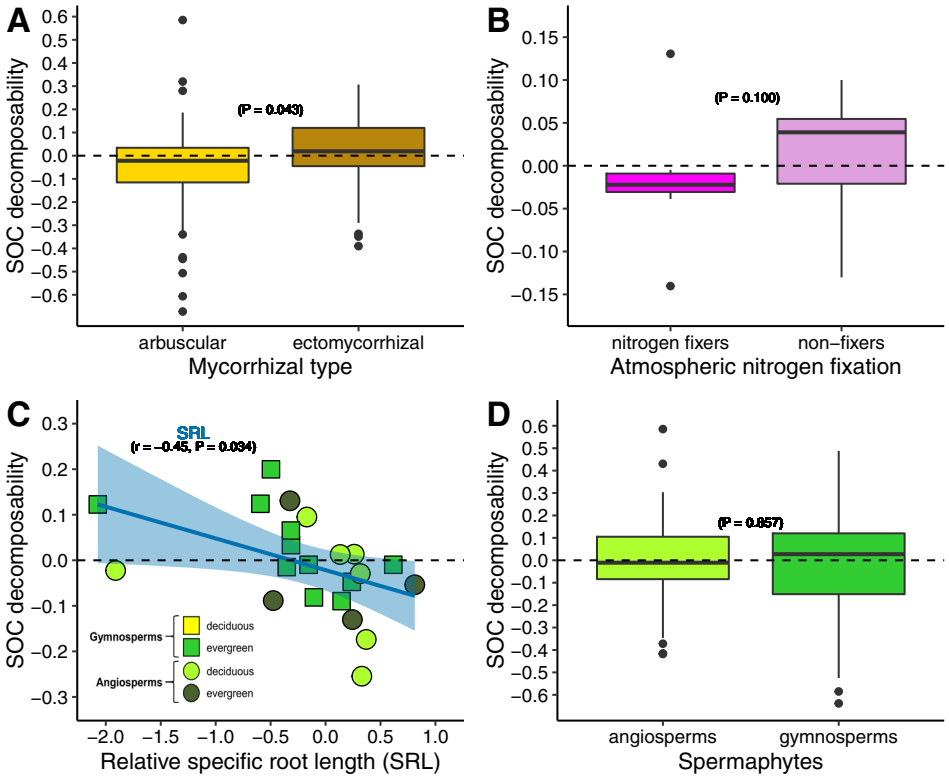

Fig. 3 SOC decomposability as a function of plant functional types and specific root length. The values show the SOC decomposability, which is the opposite of SOC stability. Values are normalised (see "Methods"). Number of values: $n = 49$ & 52 (arbuscular versus ectomycorrhizal), $n = 8$ & 14 (fixers versus non-fixers), $n = 28$ & 31 (angiosperms versus gymnosperms). **A**, **B**, **D**: boxplots represent the median, the first and third quartiles, and 1.5× the inter-quartile range; significant differences tested with pairwise *t* test or Wilcoxon test (two-sided), depending on data structure. **C** a linear regression was fitted (level of confidence of the error band = 0.95). Source data are provided as a Source Data files.

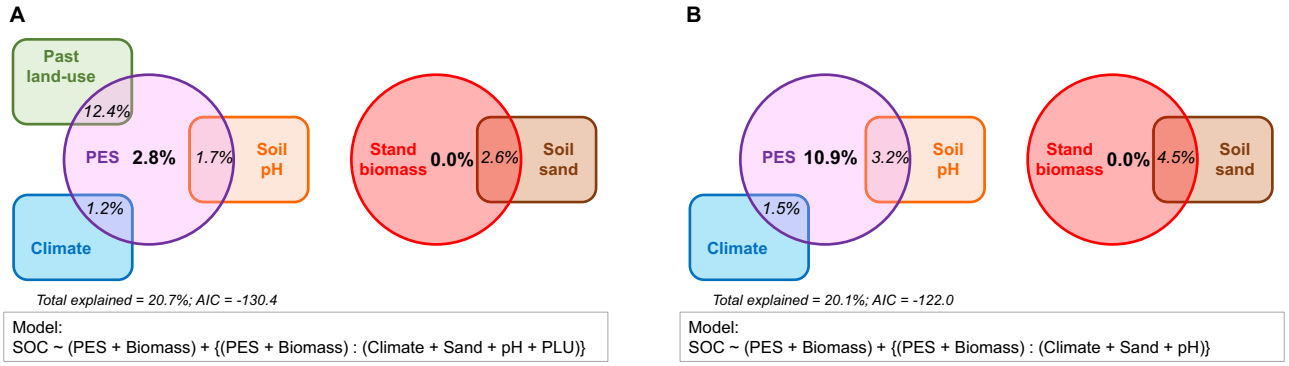

Fig. 4 Modulation of the imprint of tree species on SOC by site properties. The Euler diagrams present the relative importance of factors (and some of their interactions) in explaining the SOC pool. The model used an integrative index of the Plant Economics Spectrum (PES) as predictor (see Methods), which enabled to include a large dataset of sites worldwide but at the expense of the level of variance explained (see Supplementary Table S2 for results with original values from a smaller set of sites). The model was run with (**A**), or without (**B**), the past land-use information (see Methods). PLU = past land-use (classed as "*agriculture*" or "*forest*"; see Supplementary Fig. S11).

soil properties had a confounded effect on the PES-SOC relationship. Overall, the PES control over SOC weakened when site conditions became favourable for biological activity (wet-warm climates or fertile soils). This suggests that the improved SOC storage observed under some tree species may be due to biochemical recalcitrance of their necromass to biological decomposition[20]. When the environmental conditions are favourable for biological activity, they may enable the soil biocenosis to decompose any kind of necromass[41], even if it is in theory less prone to degradation. Conversely, when the environmental conditions become harsher, biological activity is reduced and necromass decomposition becomes more dependent on the biochemical composition of the substrate, leading to higher SOC accumulation under tree species with more recalcitrant necromass, meaning low nutrient contents and high LDMC[12,67], like found for conservative species.

The second factor controlling SOC storage was the standing biomass of the forest (Fig. 1): the higher the biomass, the greater the SOC (Supplementary Fig. S2G). This effect is linked to the observation that a larger biomass represents a larger production of necromass (i.e., litterfall (Supplementary Fig. S13), and dead fine roots), which in turn positively affects SOC accumulation[1,41],

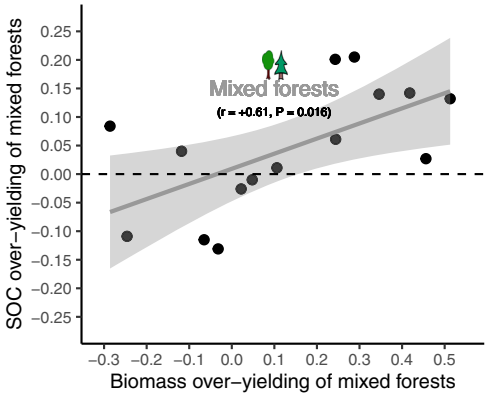

Fig. 5 Increase in SOC sequestration by forest production overyielding induced by tree species mixtures. Values are indices of over-yielding due to tree species mixtures. Negative, zero, or positive values indicate that mixed forests performed worse, equally, or better on average than their respective mono-specific counterparts (i.e., the tree species that compose the mixture). All mixed forests are angiosperm-gymnosperm mixtures, excepted two cases (one angiosperm-angiosperm, one gymnosperm-gymnosperm, but of different tree species). The SOC pool considered is the whole soil profile (forest floor + topsoil). A linear regression was fitted (level of confidence of the error band = 0.95). Source data are provided as a Source Data files.

supporting our hypothesis (H3a). High standing biomass also induces shading of the soil surface with negative consequences on forest floor decomposition[68]. In our study, tree species that produced high amounts of biomass were conservative species, like gymnosperms (Supplementary Fig. S14). This result is logical since, in natural environmental conditions generally characterised by limited supplies of resources, conservative species like gymnosperms can grow as fast as –or even faster than– acquisitive species like many angiosperms[69,70].

In addition to the direct effect of standing biomass on SOC, it is worth highlighting that mixing tree species in a given forest can indirectly increase SOC storage through positive effects on ecosystem productivity (hypothesis H3b). We found that, when mixtures produced more biomass than the relative mono-specific stands, it resulted in more SOC (Fig. 5). However, it is also important to note that mixed forests did not systematically store more SOC than the mono-specific forests ($P = 0.161$, t test = +1.46, $n = 19$), and on average they stored less SOC than the most efficient mono-specific forest at the same site ($P = 0.011$, t test = –2.83, $n = 19$).

Overall, results supported our initial hypotheses, and our study showed that tree species composition of a forest, and standing biomass, have a strong imprint on SOC storage, and should be considered in the current numerous initiatives for massive tree-plantings[10]. A simple recommendation could be to maintain high levels of standing biomass as it influences the storage of SOC positively. However, the carbon pool in trees is exposed to many hazards and it therefore appears reasonable to restrict the high-biomass strategy to regions that are not exposed to frequent disturbances[6].

Our global study particularly showed that the tree species composition of a forest, by determining the dominant plant traits, was the first driver of its capacity to store SOC, but that the role of forest composition was context-dependent[71], meaning that probably no unique global mitigation strategy exists[72]. Indeed, our study suggests that functional traits of tree species do not play an important role for SOC storage in warm-wet regions, like tropical regions. In those regions, the most urgent goal remains to protect or restore the natural composition of tropical forests[73,74], whose surface area is shrinking. They have some of the highest levels of biodiversity worldwide, and are composed of broad-leaved trees that cool the atmosphere due to their intensive evapotranspiration and moderate albedo[73,75].

In non-tropical regions, the most efficient forest compositions to mitigate climate change depend on both climate and soil properties. Under warm-temperate climates or on finely-textured fertile soils, tree species with an acquisitive resource strategy, like many deciduous angiosperms symbiotically associated with arbuscular mycorrhizae, appear to be beneficial. In such environments they perform as well as conservative tree species at storing SOC while often growing fast[69]. In addition, our results suggested that SOC under acquisitive tree species may be more stable, rendering it sequestrated for longer. Finally, even if more studies are needed on this topic[76,77], it seems that promoting deciduous angiosperms under warm-temperate climates may be more interesting than gymnosperms to cool the atmosphere through biophysical effects (evapotranspiration and albedo;[11,77]).

In the least favourable environments for biological activity, which are boreal climates or infertile soils, conservative tree species perform better at storing SOC than acquisitive tree species because of the strong control of the PES on necromass decomposition[60]. Under such conditions, ectomycorrhizal evergreen gymnosperms are more efficient to help mitigate climate change. Beyond their positive effect on SOC accumulation (Supplementary Fig. S8), conservative tree species are adapted to maintain their growth in constrained and competitive environments[69,70]. They are more able to profit from the CO_2 enrichment of the atmosphere than acquisitive species[78], and they only have a slight effect on air temperature through biophysical effects[72,77]. In boreal forests, where SOC constitutes the largest carbon pool of the ecosystem[2] and where conifers are well-adapted to the widely-spread poor soils, forest composition is consequently a major force for enhancing SOC storage.

The way forests can best help mitigate climate change is context-dependent, implying that no single specific or functional tree composition constitutes a practical solution. Instead, maintaining high levels of standing biomass in areas not exposed to frequent disturbances, and restricting tree species to habitats where they are ecologically adapted –such as conservative species in harsh conditions– is one main guideline that emerged from this global study. It also follows that in tropical areas an efficient strategy to help mitigate climate change is to protect the natural forests, and at mid- and high-latitudes (or on fertile soils and on poor soils respectively) it would be to promote broadleaf tree species and needleleaf tree species, respectively. In addition, promoting mixed stands instead of mono-specific stands can improve the carbon sink in forests, provided that the composition of the mixture is well-designed[79].

Because they play a major role in the Earth's carbon cycle, forests can contribute substantially to climate change mitigation if properly protected and managed[80]. Within this perspective, in parallel to actions that reduce CO_2 emissions, the present study showed that choosing the most efficient composition of tree species is a leverage close to hand.

Methods

Data collection: soil organic carbon. The process of data acquisition, selection and harmonisation is illustrated in the Supplementary Fig. S15 and in the Supplementary Tables S4–6. We conducted a systematic review for peer-reviewed journal articles, published before December 2018, from Web of Science, and Google Scholar with the search terms "(tree species OR forest) AND (soil organic carbon OR soil organic matter)". We also used studies listed in two previously published meta-analyses[16,17], or cited in already retained references (including references in English, French, Spanish, Portuguese, or Russian). For inclusion in the analysis we chose studies based on the following criteria: (1) the study reported soil organic carbon (SOC) or soil organic matter (SOM) concentrations or pools, at

least in the topsoil layer and under at least two single-species forest stands; (2) the stands had to be older than 10 years[81]; (3) the stands had not experienced a major disturbance that differed between tree species, for at least 30 years (e.g., we rejected studies that compared natural forests with planted forests that were less than 30 years old); (4) The SOC concentration was <20% (i.e., we did not consider Histosols and peatlands). We considered studies at the scale of individual trees if the effect of other trees could be assumed to be negligible (e.g., tree clusters[82], or lysimeters[83]). We excluded forests used for agricultural production, e.g., coffee. Together with the single-species requirement, this implied that we mainly considered common garden experiments and comparative plantations because spontaneous forests are typically composed of multi-species stands. We also rejected studies that examined highly disturbed reforested sites (mining spoil heaps, harbour sediments, city parks, etc.) as they often present artificial soil conditions. In a given site, we retained only the stands that had not been disturbed (e.g. wildfire[84] or windthrow). For regions where data were scarce (e.g., Africa, Asia (except China), the Tropics, Russia etc.), the selection criteria could be applied with flexibility, e.g., comparisons of spontaneous stands with planted stands. When results from the same study sites were reported in different articles, only one article (the most complete or the most recent one) was included in our database. All studies and data were collegially evaluated to avoid inclusion/exclusion bias. A total of 114 articles from 110 independent author groups, covering 136 sites, and 454 stand-level observations matched the selection criteria and were included (by the two authors, working together) in this study (Supplementary Fig. S16, Source Data file, Supplementary Reference 1). In sites where, besides monospecific stands, two-species mixtures (50–50%) were tested, we also collected data from the mixed stands (29 stands). When data were presented graphically, the values were extracted using the free software WebPlotDigitizer version 4.3 (https://automeris.io/WebPlotDigitizer).

One major bias in soil carbon studies is caused by the comparisons of soil layers with a fixed thickness. Indeed, because the soil bulk density strongly depends on SOC, having two layers with different SOC contents generally implies different bulk density values and different soil mass, which means comparing two SOC pools but from non-equivalent soil layers. Taking bulk density into account improves comparisons done solely with SOC concentration values, but it still produces large errors[85,86]. To overcome this difficulty, it is recommended to use the *equivalent soil mass* approach[85]. We calculated SOC pools in equivalent soil masses (ESM) following a published procedure[87]. We converted SOC concentrations and bulk density values into SOC pools contained in soil layers with equivalent mass. In our study, we chose to describe soil profiles using ESM increments of 1000 $Mg_{\text{-soil}}$ ha^{-1} (see Supplementary Fig. S17 for an example). In practice, ESM.0000-1000 is the SOC content of the uppermost soil layer with a mass of 1000 Mg (dry weight) of soil per hectare, whereas ESM.1000-2000 is the SOC content of the soil layer (also with a soil mass of 1000 Mg ha^{-1}) just below the ESM.0000-1000 layer. The same rationale applies for the soil layers ESM.2000-3000 and so on. During data analysis, we sometimes grouped several ESM layers into one pooled layer. Notably, the pooled layer containing ESM.0000-1000, ESM.1000-2000, and ESM.2000-3000 was referenced to as ESM.0000-3000. To be used, the ESM method requires data for several soil layers. For studies that reported values for only a single topsoil layer depth (e.g., 0–15 cm), we estimated SOC in ESM units by taking into account the soil bulk density and layer thickness. When soil data contained only SOM values, we converted them to SOC values using the simple equation [SOC = SOM / 2.0], which is based on a review study[88].

Information about the soil bulk density (BD) was given for only 36% of the stands so, for stands where BD information was missing, we estimated BD based on soil properties. For the forest floor layer (9% of data), we used 0.1346 as the bulk density, which is the mean value from different studies[89,90]. For the mineral soil layers, we first tested several pedo-transfer functions with a global dataset that contained measured BD values (Supplementary Table S7). It appeared that a function using SOC and soil texture as predictors[91] was the most reliable in soil layers with low SOC concentration values (SOC ≤ 20 mg g^{-1}), whereas a function based only on SOC[89] was more reliable in richer soil layers. We applied these functions accordingly (43% of data). For the remaining data (12% of data) that contained no SOC concentration values but SOC pool values, we built a dedicated function based on data with known values of BD and SOC pool ($n = 288$; $P < 0.001$; $r^2 = 0.42$). It is noticeable that our studied sites were homogeneous in terms of data availability: when values of soil bulk density were missing in a given site, it was for all the site's stands. Because the main focus of our study was the comparison of different tree species in the same site, we assumed that possible errors induced by estimating the soil bulk density did not bias the relative ranking of the several tree species present in this site.

In forest soils, BD value is generally around 1 (i.e., 0.8–1.5 kg L^{-1}; [91]), implying that an ESM layer of 1,000 $Mg_{\text{-soil}}$ ha^{-1} corresponds approximately to a 10 cm thick soil layer. When a study reported SOC values for very thick soil layers, e.g., 30 cm, it was not possible to calculate SOC pools directly for ESM.0000-1000, and so on. To overcome this problem, we used the mean relative distribution of SOC within a given soil mass, which we calculated from the other studies. For example, for a topsoil layer of 3000 $Mg_{\text{-soil}}$ ha^{-1} (ESM.0000-3000), the SOC contained in ESM.0000-1000, ESM.1000-2000 and ESM.2000-3000 was on average 45%, 30%, and 25%, respectively.

Finally, we obtained a dataset with values for nine soil layers: the forest floor and eight layers of the mineral soil horizons (from ESM.0000-1000, which is the uppermost mineral layer, down to ESM.7000-8000, which corresponds to a mean depth of 95 cm (range = 60–100 cm)). Whereas studying deep SOC is highly needed[92–94], we did not do this because it was neither possible nor appropriate. It was not possible because published studies rarely report data for deep soil. In our compilation, data availability was 100% in ESM.0000-1000 but dropped quickly down to 12% in ESM.7000-8000. It was not suitable because of the age of the studied forest stands in comparison to the SOC age along the soil profile. In our study, the mean value of stand age was 44 years (22–54 years between the first and third quartiles). In soils, deep SOC is generally much older than the organic carbon contained in the uppermost layers[95,96]. A recent global analysis showed that around 80% of the SOC that has been incorporated over the last 40 years are in the 0–30 cm soil layer[97]. Taking into account our mean stand age (44 years), the tree species effect on total SOC quantity was not expected to be visible below a depth of 30 cm. For the analysis, we consequently retained data only from four layers (i.e., forest floor, ESM.0000-1000, ESM.1000-2000, and ESM.2000-3000), which we finally grouped into two layers (forest floor and mineral layers hereafter referred to as topsoil (ESM.0000-3000, which was the sum of ESM.0000-1000, ESM.1000-2000, and ESM.2000-3000)). In our study, the SOC content of the cumulated pools in topsoil represented a large proportion (66%) of the whole soil profile. Finally, retaining a limiting value of ~30 cm depth also had the advantage of producing results that could be easily applicable at a large scale because most global soil inventories are carried out using a 0–30 cm basis[98,99].

In addition to values of the total SOC content of soils, we collected information about SOC stability. For this, we collected data based on all approaches (i.e., incubation results, size and density fractions, chemical extractions) because there is no unique and standardised method to quantify SOC stability. The final dataset contained 25 studies that reported metrics for SOC stability at 30 sites.

Data collection: auxiliary data. SOC content at medium or large scales can be controlled by soil properties, vegetation type (in this study: tree species), land-use, topography, soil parent material, and climate[100]. Consequently, we collected auxiliary data related to those factors. In a first step, we collected data about sites, stands, and soil characteristics by extracting information from each publication (or companion publications from the same author group). At the site scale, the collected information was: site name and location (longitude and latitude; to enable the identification of several studies about the same site), elevation, mean annual values of temperature and precipitation (MAT, MAP; °C, mm yr^{-1}), past land-use, study design (e.g., number of blocks; see below the paragraph dedicated to data weighting), fertilisation history, soil name following the USDA classification (conversions from other classification systems were made following published references [101,102]), soil weathering stage and soil parent material[102], topsoil texture (based on the quantitative particle size analysis of the clay, silt and sand fractions, or estimated from the qualitative description and the USDA texture triangle: e.g., "silty loam" corresponds to clay = 150 mg g^{-1}, silt = 650 mg g^{-1}, and sand = 200 mg g^{-1}), and other topsoil properties (e.g., pH, cation exchange capacity and its saturation value, total content in phosphorus). When the topsoil properties were presented at the stand scale, we calculated mean values at the site scale.

At the stand scale, we collected the following information: tree species name (updated to the current nomenclature where necessary), stand age (yr), stand density (stems ha^{-1}), stand aboveground biomass (Mg ha^{-1}), stand aboveground volume (m^3 ha^{-1}), stand basal area (m^2 ha^{-1}), mean value of stem diameter at breast height (cm), mean tree height (cm), flux of litterfall (Mg ha^{-1} yr^{-1}), fine root biomass (Mg ha^{-1}), and fine root production (Mg ha^{-1} yr^{-1}). We also investigated the opportunity to collect more information (stand productivity, understorey composition and abundance, earthworms, etc.), but they were generally lacking in the original studies. To enable comparisons, when possible we calculated the different stand metrics (e.g., missing values of basal area were calculated from stand density and stem diameter). Then, when possible (25% of data), we estimated the missing values of stand biomass using linear regressions with other stand metrics (i.e., based on tree height [$n = 37$; $P < 0.001$; $r^2 = 0.51$] or stem diameter [$n = 50$; $P < 0.001$; $r^2 = 0.45$]), relying on the strong allometric relationships that exist among tree compartments[103–105]. We applied the same rationale to estimate the stand litterfall flux using the fine root production as predictor (3% of data; [$n = 13$; $P = 0.017$; $r^2 = 0.42$]).

In a second step, we complemented our dataset (based on values provided in the original studies) with external sources using the latitude-longitude coordinates of the sites. These external sources were used to include variables that were not reported in the original studies (e.g., nitrogen atmospheric deposition), or to fill in gaps in our datasets when values were missing in the original studies. The extracted variables were: atmospheric nitrogen deposition[106], Köppen-Geiger climate classes[107], mean annual –or monthly– values of precipitation or temperature (MAP, MAT; http://worldclim.org), potential evapotranspiration and aridity index (https://cgiarcsi.community), elevation (https://www2.jpl.nasa.gov/srtm), soil properties[108,109], and soil parent material[110]. The latter was simplified in a four class system (acid, intermediate, mafic, and calcareous)[102]. When possible, we checked the quality of the external sources by comparing them with the values that existed in the original studies (Supplementary Table S8). The consistency was fairly good ($n = 30$–99, slope = 0.75–1.12, $r^2 = 0.71$-0.98) for soil texture and soil acidity (pH or base saturation), and excellent for climate data and site elevation ($n = 79$–110, slope = 0.98–1.03, $r^2 = 0.90$–0.99). Conversely, we considered that

the consistency for other variables (e.g., phosphorus content, cation exchange capacity, exchangeable cations, soil name, etc.) was not high enough and consequently we did not retain these external data. The relationships among the site properties are shown in the Supplementary Fig. S18.

Data collection: plant functional traits. We collected values of the functional traits that define the Plant Economics Spectrum[49], and widely expanded this list by adding plant traits, leaf traits, and root traits (n = 74). The values of plant functional traits were determined based on several approaches. First, we used the TRY trait database[111]. We downloaded the publicly available data related to the 178 tree species included in our SOC database (www.try-db.org; access date: January 2019). This dataset contained 938,978 individual trait values for the 74 studied traits.

In a second stage, we complemented our trait database with data from other published sources like FRED[112], China Plant Trait database[113], ORNL DAAC leaf traits (https://daac.ornl.gov), and TTT[114]. We also used trait values from the references used for building our SOC database as well as 143 other published studies (Supplementary Reference 2). The latter references were included because, when evaluating the data quality of the other sources (including TRY), we realised that some important functional traits had no value for many of the tree species present in our SOC database. To look for additional studies to fill in data gaps, we searched in Google Scholar and Web of Science using specific keywords, for instance: {"cupressus lusitanica" AND (SLA or "specific leaf area")} or {"cupressus lusitanica" AND (leaf OR foliage OR foliar OR needle) AND nitrogen}.

For traits such as leaf life span (classed as deciduous or evergreen) or tree maximum height (in metres), we also consulted the following websites: Wikipedia (in English, French, Spanish, and Portuguese; https://www.wikipedia.org), https://www.conifers.org, http://efloras.org, http://www.fs.fed.us, https://www.gbif.org, http://issg.org/database/welcome/, www.iplantz.com, https://pfaf.org, http://www.tree-guide.com, http://tropical.theferns.info, https://wiki.bugwood.org/Main_Page, http://www.worldagroforestry.org. Missing values for seed mass and wood density were found respectively in the Kew Seed Information Database (http://data.kew.org/sid) and the Tree Functional Attributes and Ecological Database (http://db.worldagroforestry.org/wd). The mycorrhizal status of tree species that were not in the TRY database was complemented using dedicated references [115–117]. When, despite these references, a status was still missing, a value based on the genus or the family was assigned.

In total, the second trait dataset was composed of ~48,700 values, from which we retained 6,611 values (about 71 traits) that were related to the tree species present in our SOC database. This dataset was merged with the TRY dataset.

Data quality was carefully assessed for all traits by checking if *(i)* each category was homogeneously coded (e.g.,: ectomycorrhizal tree species were coded either "E." or "EM"; deciduous species were coded in 12 different manners such as "deciduous", "deciduous type 1", "winter deciduous", etc.), *(ii)* the number of values for a given variable –or category– was high enough to be used (e.g., leaf palatability was discarded as n = 7), *(iii)* the unit used for a numerical trait was the same for all values (e.g., leaf carbon content was found in $g_{\text{-C}}\ g_{\text{-DW}}^{-1}$, $mg_{\text{-C}}\ g_{\text{-DW}}^{-1}$, %, or $mmol_{\text{-C}}\ g_{\text{-DW}}^{-1}$). When necessary, the values of the categorical traits were homogenized (renaming, or merging similar classes). Similarly, the units of the numerical traits were homogenised, and the values were changed accordingly. After this quality assessment stage, we excluded traits for which most of the values were missing (e.g., plant light requirement, or root type). We did the same for traits that presented many classes, which resulted in having only a few tree species per class (e.g., Grime's groups or reproductive phenology), and for traits for which the classes could not be homogenised because of classification inconsistencies among the original data sources (e.g., soil texture requirement of the tree species).

Duplicated values from the same original study, but present in several sources, were removed. When three traits were interconnected (e.g., leaf C content, leaf N content, and leaf C:N ratio) and one value was missing, the latter was calculated based on the two non-missing values. When several values existed for the same trait-species combination, mean values were calculated. Using mean values for functional traits may introduce some additional variance in results because trait values in a given plant species are plastic and acclimate to environmental conditions from site to site[118,119]. However, interspecific trait variation generally exceeds intraspecific trait variation[120,121]. For instance, the interspecific coefficients of variation for SLA, LDMC, leaf N content, and leaf P content (n = 65–144 species) were 73%, 19%, 44%, and 45% in our database, while the intraspecific coefficients of variation of these variables were reported to be ~19%, ~10%, ~18%, and ~22% for two tree species (*Pinus koraiensis* and *Fraxinus mandshurica*;[122]). Similarly, the interspecific coefficient of variation of the leaf photosynthetic maximum rate was very high in our database (CV = 93%, n = 95), which is in line with the global variation of this trait[123]. Consequently, because our study encompassed a wide range of functional plant types studied at the global scale, we assumed that our approach based on mean trait values was reliable due to high interspecific trait variation[124].

We investigated the relationships among functional traits using Principal Component Analyses. When several traits were well-correlated to each other (example 1: leaf dry matter content, leaf strength, leaf thickness, and leaf lifespan; example 2: leaf photosynthetic maximum rate, leaf photosynthesis carboxylation capacity, leaf stomatal conductance, and specific leaf area; example 3: leaf content in C, N, P, K, Ca, Mg and C:N and N:P ratios), we removed the traits with the highest frequency of missing values (in the examples above we removed strength, thickness, lifespan, photosynthesis carboxylation capacity, stomatal conductance, K content, Mg content).

To decrease the number of retained traits even further, and increase the rate of non-missing values, we used a gap filling procedure for five foliage traits (A_{max}, C, N, P, Ca): first, we built relationships among redundant traits (*cf.* example 2, above): when, for a given tree species, the retained trait had a missing value, it was estimated based on the removed traits (e.g., leaf photosynthetic maximum rate estimated based on the values of the leaf photosynthesis carboxylation capacity, or leaf stomatal conductance). Furthermore, we observed that, for a given element (C, N, P, Ca), the contents in leaves, litter, and fine roots were all inter-correlated[56,125,126] while root traits are poorly known[127]. Consequently, we built regressions between various related plant traits, and estimated the missing values for leaf nutrient content based on litterfall nutrient content, or fine root nutrient content (r^2 = 0.44–0.99; estimated data = 1.5–13.7% of data; Supplementary Table S9).

For the leaf C content, the remaining missing values were estimated using the plant genus, family, or spermaphyte group, based on published reviews[56,128,129]. Out of the 454 stands, 4 stands from scarcely represented areas only reported the tree genus but not the species name (i.e., *Acer* spp., *Eucalyptus* spp., *Pinus* spp., and *Quercus* spp.). For these few cases, we used the mean trait values of all the species in that particular genus.

The dataset containing the final trait values could not be provided in this publication because we agreed to the intellectual property guidelines for the TRY initiative, which prohibit the redistribution of the trait data received (see a summary in the Supplementary Table S10). However, all the trait data that we used are publicly available, and thus it is possible to rebuild our dataset following the methods presented above. We also made the trait data available for the reviewers.

Compiling values of plant traits for a global scale study is exposed to two pitfalls: *(i)* it is difficult to have non-missing values for uncommon plant species, and *(ii)* the high number of traits (which are dependent on each other) makes the data interpretation difficult. To circumvent these problems, in addition to the approach described above, we used an approach based on Principal Component Analysis (PCA), which is particularly adequate to synthetically describe datasets with a high level of covariance among variables[130]. First, we produced estimated values for missing values (using the functions *estim_ncpPCA* and *imputePCA* from the *missMDA* package[131]). Then, we performed a PCA analysis (function *PCA*) to produce an integrative index of our explanatory variables. In practice, this integrative index was the score on the first axis of the PCA. We produced two integrative indexes: (i) an index score of the PES (produced by a PCA that took into account the main traits: leaf size, A_{max}, LDMC, SLA, SRL, and leaf contents in lignin, C, P, N, P, Ca, and C:N ratio; Supplementary Fig. S19), and (ii) a Biomass index (based on stand biomass, annual litterfall, tree species maximal height, species growth rate, seed mass, and wood density; Supplementary Fig. S20). Adding categorical variables (such as spermaphytes) either did not change the results or made them worse, and were consequently not used for the PCA.

We used both original trait values and the index values in our study. Original values were mainly used to test the main effects of the studied drivers (i.e., plant traits and stand biomass) because they were not estimated values. Nevertheless, original values were more or less missing, rending difficult any multi-variables analysis. On the other hand, because index values were calculated for most sites, they were used to confirm results on a larger scale, and to test the influence of the environmental conditions (e.g., temperature or soil clay content) on the PES-SOC relationship (see below). However, as these indexes were only proxies, their use substantially decreased the explanatory power of the analyses.

Dataset compilation: phylogenetic distance. We built a phylogenetic tree from all species in our SOC dataset to determine whether life history characteristics, represented by phylogenetic proximity, constitute a characteristic that affects SOC. The phylogenetic distance between two species was estimated, based on relevant literature and using the approach of the *most recent common ancestor*. To do this, we built a phylogenetic tree that contained all the families included in our dataset about soil organic carbon (Supplementary Fig. S21). The distance between angiosperms and gymnosperms was fixed as 350 Myr[132]. Following the same reference [132], the distance between *Cupressales* and *Pinales* was set at 273 Myr. Within the gymnosperms, the distances among clades down to genera were estimated based on a recent study[133]. The distances between two species of the same genus (*Abies*, *Picea*, or *Pinus*) were fixed based on dedicated studies[134–136]. Within the angiosperms, we first determined the phylogenetic distances among families[137]. Then, we used the Angiosperm Phylogeny Website (http://www.mobot.org/MOBOT/research/APweb; accessed in January 2020) and relevant references [132,138–142] to estimate the distances between two species that were in the same family or genus.

In total, our dataset about soil organic carbon contained 787 combinations of tree species pairs (by considering pairs only within the same site). For instance, in a common garden with stands of four different tree species, there are six possible pairs of tree species comparisons. Overall our dataset contained 641 distinct pairs because some pairs were found in several sites. The phylogenetic tree contained 629 non-missing values of phylogenetic distance (i.e. 98.1% of all pairs). For the remaining pairs (n = 12; 1.9%), we estimated the missing values as being half of the crown age of the clade containing the two tree species.

The influence of the phylogenetic distance was studied by investigating whether the absolute difference of soil organic carbon between two tree species within the same site was correlated to their phylogenetic distance.

Data handling and normalisation. We applied weights to each case study in order to take into account the robustness of the study design. The three criteria considered were: the site design (C1), the number of blocks (C2), the number of sampled soil profiles (C3). For each criterion, a weight value was given or calculated (normalised to [0-1]; see below).

For the site design (C1), the weight value (W_S) was fixed as 1.00, 0.66, 0.33, and 0.00 respectively for: (i) common gardens (i.e., stands installed specifically for use as comparisons, following a unique installation protocol), (ii) comparative plantations (i.e., stands of similar age and planted following similar methods, but without the intention to compare), (iii) spontaneous forests (i.e., non-planted stands, fairly mono-specific (>80% of biomass or stem density), and in similar site conditions), and (iv) heterogeneous designs (i.e., comparisons of spontaneous stands with plantations; the case studies based on this design were retained only in regions where reliable data were lacking).

For the number of blocks (C2), the weight value (W_B) was calculated as: $W_B = [\log_{10}(\text{nb.blocks})]$

For the number of soil profiles (C3), the weight value (W_P) was calculated as: $W_P = [\log_{10}(\text{nb.profiles})] / 2$

All weight criteria ranged between 0 and 1. The data weight (W_{data}) was then calculated as the mean of the weight criteria (W_S, W_B, W_P). To avoid having a null statistical weight, W_{data} values lower than 0.05 were corrected to this threshold value. Overall, W_{data} was on average 0.36 (range = 0.05-0.69). Values of W_{data} were declared as a statistical weight factor during the data analyses (when the SAS procedures or R packages enabled applying such a correction).

At the global scale, the SOC content varies a lot among sites[5], which implies that inter-site variability is generally much larger than intra-site variability induced by plant species[18]. To examine the effect of plant traits on SOC pools without having the influence of the site characteristics, we calculated relative responses of SOC. The relative response approach enables to normalise values relatively for each site, removing the direct influence of the site conditions[143]. Similarly, because the list of present tree species varied from site to site, it was not possible to use the absolute values of functional traits as predictors and we consequently used relative values also for traits.

For variables with a continuous distribution (e.g., functional traits such as leaf N content, SOC pool size, or stand biomass), we calculated a relative value for each stand by normalising values relatively to the site[144]. For instance, in a study reporting results from three stands (A, B, C), we calculated the response ratio of each stand following this formula (here for the stand A):

$$RR.A = \log\left(\frac{SOC.A}{SOC.site}\right) \quad (1)$$

where *RR.A* is the relative response ratio for stand A of the study, *SOC.A and SOC.site* are soil organic carbon (equivalent soil mass) values for stand A and the whole site reported in the study (average of all stands at a site). If RR.A < 0, then the species of stand A had a negative effect on SOC accumulation at this site. If RR.A > 0, then the species had on average a positive effect on SOC. If RR.A ≈ 0, then there was no effect of the tree species on SOC. We followed the same formula for all plant traits of interest (see Supplementary Table S11 for an example).

We should note that the dataset in the Source Data file does not contain all possible relative values. Indeed, the values presented are only those that could be calculated when all the tree species of a given site had non-missing values. During data analysis however, if we studied a possible relationship between two variables (e.g., between SLA and SOC), we recalculated all the possible relative values provided that at least two species within the same site had non-missing values.

When the tree species in a given site were studied using a categorical approach with two classes (e.g., angiosperms versus gymnosperms, or N-fixing species versus non-fixing species), we calculated the mean SOC value per class (e.g., for angiosperms). In practice, firstly we recalculated the mean normalised values per class in order to avoid pseudo-replicated values in sites with more than two tree species of the same class. Then, we calculated a log relative response, commonly used in meta-analyses[145]. For instance, with angiosperms and gymnosperms (see also Supplementary Table S11), the relative values of SOC were:

$$RR = \log\left(\frac{SOC.angiosperms}{SOC.gymnosperms}\right) \quad (2)$$

Because the final dataset about mixed forests did not contain all the data necessary to calculate ESM values, we chose to use SOC pool values in this case. We compared the SOC pools (forest floor + mineral soil) under mono-specific and mixed stands by calculating an *overyielding* index, which was a relative response ratio of stand characteristics and SOC pools following the formula:

$$RR.overyielding = \log\left(\frac{AB}{mean(A,B)}\right) \quad (3)$$

With A, B and AB as the absolute values for the two monospecific stands (species A or species B) and the mixed stand (50%-50% mixture of the species A and B). Negative, null, and positive values indicated that the mixed stand had a lower, equal, or higher value than the mean value of the monospecific stands (overyielding corresponds to cases where RR.overyielding > 0).

We also calculated a *transgressive overyielding* index:

$$RR.transgr.overyielding = \log\left(\frac{AB}{max(A,B)}\right) \quad (4)$$

Positive values indicated that the mixed stand had a higher value than the highest value of the monospecific stands.

Data analysis. We did not remove any outliers, and used all the data we had for our analyses (i.e., regressions, correlations, and tests). However, to produce easily readable graphs, we sometimes reshaped the axes scale which made a few ($n = 1$–3) outliers invisible, but they were still taken into account for the statistics. In graphs showing regressions, error areas are confidence of intervals. Boxplots represent the median, the first and third quartiles, and 1.5 × the inter-quartile range.

During data analysis, we used firstly the *Random Forest* approach to identify the main drivers of SOC (*randomForest* R package[146]). We applied this method to absolute values of SOC to determine the drivers at the global scale (i.e., inter-site variability) and to relative values of SOC to determine the drivers at the site scale (i.e., intra-site variability). The identified drivers of SOC at the site scale (i.e., plant traits and stand properties) were used in subsequent analyses.

Then, we examined the effect of tree species on SOC by evaluating relationships between the RR values of plant traits and SOC using simple linear regressions and Spearman's rank correlation coefficients. In a second step, we used a modelling approach based on multiple linear regressions. In practice, we modelled the RR SOC values (i.e., combined pools of forest floor plus ESM.0000-3000) with the most predictive variable (i.e., plant traits and stand properties, in normalised values), and then analysed the possible existence of relationships between the residuals and other variables. The built model was evaluated in two ways: with the *performance* package in R (see below), and using an independent dataset. This later dataset was composed of the sites that had non-missing SOC values for the forest floor, and for the ESM.0000-1000 soil layer (or the ESM.0000-2000 layer), but had no value for the deeper layers (n = 43 stands from 23 distinct sites distributed over 4 continents, under cold or temperate climates, and on 6 different soil types). The calibration dataset and the validation dataset were initially not strictly comparable because the contribution of the forest floor layer to the combined SOC pool (i.e., forest floor + topsoil) was lower for the calibration dataset (i.e., topsoil = ESM.0000-3000; forest floor = 19% of the combined pool on average) than for the validation dataset (23% and 33% for the cases where topsoil = ESM.0000-2000 and ESM.0000-1000), and because the influence of plant traits on carbon pools were of unequal magnitude in the forest floor and the topsoil. Indeed, the normalised values of SOC pools (RR.SOC) for the validation dataset were 11.5% (forest floor + ESM.0000-1000) and 6.4% (forest floor + ESM.0000-2000) higher than for the calibration dataset (forest floor + ESM.0000-3000). We consequently modified the validation dataset by applying a correction factor of 1.115 or 1.064, depending on the case.

The influence of tree species on SOC was also investigated using a categorical approach (e.g., gymnosperms versus angiosperms, evergreen versus deciduous, ectomycorrhizal versus arbuscular mycorrhizal stands, N-fixing species versus non-fixing stands). After data normalisation (see Eq. 2), the recalculated values were used to test the existence of a significant difference between the two classes (Bonferroni test, Wilcoxon test, or Mann-Whitney test; two-sided tests).

Eventually, we tested whether a SOC-trait relationship (which was significant at the global scale) was influenced by site conditions (e.g., atmospheric N deposition, climate, soil properties, past land-use). In practice, SOC was modelled based on trait values, and possible interactions between trait values and site conditions were also tested:

$$\text{SOC} \sim (\text{Trait} \times \text{Biomass}) + \{(\text{Trait} + \text{Biomass}) : (\text{Site.conditions})\}$$

with N deposition, climate, soil pH, soil texture, and past land-use (PLU) as site conditions.

The site conditions were not tested as direct effects because of the way SOC values were normalised. Indeed, the normalisation (see Eq. 1) implied that, for a given site, the mean SOC normalised value was by definition always close to zero. Because all the stands within a given site share the same values of site conditions, the latter variables had no influence on SOC normalised values.

To avoid collinearity among the site descriptors, we simplified the model as follows:

- Climate descriptors (MAT, MAP, aridity, PET, $f_{climate}$, water balance) and site location (latitude, elevation) were plotted in a PCA. It resulted that the axis 1 (52.4% of the variance) was best explained by $f_{climate}$, MAT and MAP. The $f_{climate}$ metric is designed to increase with concomitant increasing MAT values and increasing MAP values, and indicates if the climatic conditions are favourable for biological functioning (enough water and not too cold)[102]. Because $f_{climate}$ explained well the climatic conditions (see above), and because it was well correlated with other climatic descriptors (Supplementary Fig. S22), it was retained as the unique climatic descriptor in the models.

- Atmospheric N deposition was highly correlated to climatic conditions (Supplementary Fig. S18) and was consequently removed from the models.

- Variables that described soil texture (i.e., clay, silt, and sand contents) were logically correlated to each other (Supplementary Fig. S18). Because sand content was well-correlated to both clay content and silt content, we retained sand content as unique descriptor of soil texture.

These simplifications resulted in the following model:

$$\mathrm{SOC} \sim (\mathrm{Trait} \times \mathrm{Biomass}) + \{(\mathrm{Trait} + \mathrm{Biomass}) : (f_{\mathrm{climate}} + \mathrm{Sand} + \mathrm{pH} + \mathrm{PLU})\}$$

However, in our dataset the categories of past land-use (PLU) were not randomly distributed as the "agriculture" PLU category had significantly more favourable conditions (i.e., higher values of f_{climate} and soil pH; lower values of soil sand content) than the "forest" PLU category ($P = 0.003$ to 0.045). To take into account this dependency among site descriptors, the data was also analysed with a model without PLU:

$$\mathrm{SOC} \sim (\mathrm{Trait} \times \mathrm{Biomass}) + \{(\mathrm{Trait} + \mathrm{Biomass}) : (f_{\mathrm{climate}} + \mathrm{Sand} + \mathrm{pH})\}$$

It should be noted that, while our dataset with SOC content data was large enough to be analysed with these models, the other datasets (about SOC stability or the effects of mixed forests) were too small to enable such an analysis. The later datasets were consequently analysed with simpler models, based on data availability.

We tested these models with a R function that uses the AIC as criterion (*ols_step_forward_aic* from the *olsrr* package[147]). We also used mixed models (*lmer* from *lme4*; assigning the site identity as a random effect) and multi-regression (*glmulti* and *stepAIC*). Because all these methods gave very similar results, we presented only those based on the *olsrr* package. In the case of SOC stability, we used a mixed effects model to take into account that this variable was measured with different lab methods (Supplementary Fig. S23). The quality of the models was evaluated with the *performance* package.

Data warehouse, data handling and preliminary analyses were done using SAS (9.4). Final data analyses and graphs were made with R (4.0.3)[148], including the packages *car*, *ggeffects*, *ggplot2*, *ggpubr*, *glmulti*, *lme4*, *missMDA*, *multcomp*, *nparcomp*, *olsrr*, *performance*, *psych*, *magick*, *randomForest*, *RcmdrMisc*, and *VIM*[131,146,148–161]. The pictures used in graphs were creative commons licensed (CC0), and published in dedicated websites (e.g., https://publicdomainvectors.org).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The SOC data generated in this study have been deposited in the https://data.inrae.fr database under accession link https://doi.org/10.15454/LJRFJR. The raw data from the TRY database are protected and are not available due to data privacy laws. The references used to compile SOC data and plant traits are provided in the Supplementary Information.

Code availability

Code of this study is available in the https://data.inrae.fr database under accession link https://doi.org/10.15454/LJRFJR.

Received: 7 March 2021; Accepted: 2 February 2022;
Published online: 01 March 2022

References

1. Jackson, R. B. et al. The ecology of soil carbon: pools, vulnerabilities, and biotic and abiotic controls. *Annu. Rev. Ecol. Evol. Syst.* **48**, 419–445 (2017).
2. Pan, Y. D. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
3. Mayer, M. et al. Tamm Review: Influence of forest management activities on soil organic carbon stocks: A knowledge synthesis. *Ecol. Manag.* **466**, 118127 (2020).
4. Nabuurs, G.-J. et al. First signs of carbon sink saturation in European forest biomass. *Nat. Clim. Chang.* **3**, 792–796 (2013).
5. Lal, R. Forest soils and carbon sequestration. *Ecol. Manag.* **220**, 242–258 (2005).
6. Thom, D. & Seidl, R. Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests. *Biol. Rev.* **91**, 760–781 (2016).
7. Bowman, D. M. J. S. et al. Fire in the Earth system. *Science* **324**, 481–484 (2009).
8. Minasny, B. et al. Soil carbon 4 per mille. *Geoderma* **292**, 59–86 (2017).
9. Bastin, J.-F. et al. The global tree restoration potential. *Science* **365**, 76–79 (2019).
10. Brancalion, P. H. S. & Holl, K. D. Guidance for successful tree planting initiatives. *J. Appl. Ecol.* **57**, 2349–2361 (2020).
11. Naudts, K. et al. Europe's forest management did not mitigate climate warming. *Science* **351**, 597–600 (2016).
12. Augusto, L. et al. Influences of evergreen gymnosperm and deciduous angiosperm tree species on the functioning of temperate and boreal forests. *Biol. Rev.* **90**, 444–466 (2015).
13. Boca, A. & Van Miegroet, H. Can Carbon Fluxes Explain Differences in Soil Organic Carbon Storage under Aspen and Conifer Forest Overstories? *Forests* **8**, (2017).
14. Augusto, L., Ranger, J., Binkley, D. & Rothe, A. Impact of several common tree species of European temperate forests on soil fertility. *Ann. Sci.* **59**, 233–253 (2002).
15. Cornelissen, J. H. C., Aerts, R., Cerabolini, B., Werger, M. J. A. & van der Heijden, M. G. A. Carbon cycling traits of plant species are linked with mycorrhizal strategy. *Oecologia* **129**, 611–619 (2001).
16. Boča, A., Van Miegroet, H. & Gruselle, M.-C. C. Forest overstory effect on soil organic carbon storage: a meta-analysis. *Soil Sci. Soc. Am. J.* **78**, S35–S47 (2014).
17. Lin, G., McCormack, M. L. L., Ma, C. & Guo, D. Similar below-ground carbon cycling dynamics but contrasting modes of nitrogen cycling between arbuscular mycorrhizal and ectomycorrhizal forests. *N. Phytol.* **213**, 1440–1451 (2017).
18. Peng, Y. et al. Tree species effects on topsoil carbon stock and concentration are mediated by tree species type, mycorrhizal association, and N-fixing ability at the global scale. *Ecol. Manag.* **478**, 118510 (2020).
19. Craig, M. E. et al. Tree mycorrhizal type predicts within-site variability in the storage and distribution of soil organic matter. *Glob. Chang. Biol.* **24**, 3317–3330 (2018).
20. Aulen, M., Shipley, B. & Bradley, R. Prediction of in situ root decomposition rates in an interspecific context from chemical and morphological traits. *Ann. Bot.* **109**, 287–297 (2012).
21. Hu, Z. et al. Traits drive global wood decomposition rates more than climate. *Glob. Chang. Biol.* **24**, 5259–5269 (2018).
22. Prieto, I., Stokes, A. & Roumet, C. Root functional parameters predict fine root decomposability at the community level. *J. Ecol.* **104**, 725–733 (2016).
23. Kahl, T. et al. Wood decay rates of 13 temperate tree species in relation to wood properties, enzyme activities and organismic diversities. *Ecol. Manag.* **391**, 86–95 (2017).
24. Barbi, F. et al. Tree species select diverse soil fungal communities expressing different sets of lignocellulolytic enzyme-encoding genes. *Soil Biol. Biochem.* **100**, 149–159 (2016).
25. Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K. & Paul, E. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Glob. Chang. Biol.* **19**, 988–995 (2013).
26. Cotrufo, M. F., Ranalli, M. G., Haddix, M. L., Six, J. & Lugato, E. Soil carbon storage informed by particulate and mineral-associated organic matter. *Nat. Geosci.* **12**, 989–994 (2019).
27. Wang, H. et al. Differential effects of conifer and broadleaf litter inputs on soil organic carbon chemical composition through altered soil microbial community composition. *Sci. Rep.* **6**, 1–9 (2016).
28. Laganiere, J., Boca, A., Van Miegroet, H. & Pare, D. A tree species effect on soil that is consistent across the species' range: the case of aspen and soil carbon in North America. *Forests* **8**, 113 (2017).
29. Verstraeten, G. et al. Tree species effects are amplified by clay content in acidic soils. *Soil Biol. Biochem.* **121**, 43–49 (2018).
30. Nitsch, P., Kaupenjohann, M. & Wulf, M. Forest continuity, soil depth and tree species are important parameters for SOC stocks in an old forest (Templiner Buchheide, northeast Germany). *Geoderma* **310**, 65–76 (2018).
31. Zechmeister-Boltenstern, S. et al. The application of ecological stoichiometry to plant-microbial-soil organic matter transformations. *Ecol. Monogr.* **85**, 133–155 (2015).
32. Soudzilovskaia, N. A. et al. Quantitative assessment of the differential impacts of arbuscular and ectomycorrhiza on soil carbon cycling. *N. Phytol.* **208**, 280–293 (2015).
33. Winsome, T. et al. Plant-microbe interactions regulate carbon and nitrogen accumulation in forest soils. *Ecol. Manag.* **384**, 415–423 (2017).
34. Hartman, W. H., Ye, R., Horwath, W. R. & Tringe, S. G. A genomic perspective on stoichiometric regulation of soil carbon cycling. *ISME J.* **11**, 2652–2665 (2017).
35. Rasmussen, C. et al. Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* **137**, 297–306 (2018).
36. Ribbons, R. R. et al. Context-dependent tree species effects on soil nitrogen transformations and related microbial functional genes. *Biogeochemistry* **140**, 145–160 (2018).
37. Castellano, M. J., Mueller, K. E., Olk, D. C., Sawyer, J. E. & Six, J. Integrating plant litter quality, soil organic matter stabilization, and the carbon saturation concept. *Glob. Chang. Biol.* **21**, 3200–3209 (2015).
38. Di, J. et al. Combinations of soil properties, carbon inputs and climate control the saturation deficit dynamics of stable soil carbon over 17-year fertilization. *Sci. Rep.* **8**, 1–10 (2018).
39. Osland, M. J. et al. Climate and plant controls on soil organic matter in coastal wetlands. *Glob. Chang. Biol.* **24**, 5361–5379 (2018).
40. Russell, A. E., Kivlin, S. N. & Hawkes, C. V. Tropical tree species effects on soil pH and biotic factors and the consequences for macroaggregate dynamics. *Forests* **9**, 1–14 (2018).
41. Frank, D. A., Pontes, A. W. & McFarlane, K. J. Controls on soil organic carbon stocks and turnover among North American ecosystems. *Ecosystems* **15**, 604–615 (2012).

42. Jactel, H. et al. Positive biodiversity-productivity relationships in forests: climate matters. *Biol. Lett.* **14**, 20170747 (2018).
43. Toïgo, M. et al. Overyielding in mixed forests decreases with site productivity. *J. Ecol.* **103**, 502–512 (2015).
44. Williams, L. J., Paquette, A., Cavender-Bares, J., Messier, C. & Reich, P. B. Spatial complementarity in tree crowns explains overyielding in species mixtures. *Nat. Ecol. Evol.* **1**, 1–7 (2017).
45. Zhang, Y., Chen, H. Y. H. & Reich, P. B. Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *J. Ecol.* **100**, 742–749 (2012).
46. Maes, S. L. et al. Litter quality, land-use history, and nitrogen deposition effects on topsoil conditions across European temperate deciduous forests. *Ecol. Manag.* **433**, 405–418 (2019).
47. Paz, C. P. et al. Soil types influence predictions of soil carbon stock recovery in tropical secondary forests. *Ecol. Manag.* **376**, 74–83 (2016).
48. Hartley, I. P., Hill, T. C., Chadburn, S. E. & Hugelius, G. Temperature effects on carbon storage are controlled by soil stabilisation capacities. *Nat. Commun.* **12**, 1–7 (2021).
49. Diaz, S. et al. The global spectrum of plant form and function. *Nature* **529**, 167–U73 (2016).
50. Bruelheide, H. et al. Global trait-environment relationships of plant communities. *Nat. Ecol. Evol.* **2**, 1906–1917 (2018).
51. Reich, P. B. The world-wide 'fast-slow' plant economics spectrum: a traits manifesto. *J. Ecol.* **102**, 275–301 (2014).
52. Wright, I. J. et al. The worldwide leaf economics spectrum. *Nature* **428**, 821–827 (2004).
53. Wright, I. J. et al. Assessing the generality of global leaf trait relationships. *N. Phytol.* **166**, 485–496 (2005).
54. Prieto et al. Leaf carbon and oxygen isotopes are coordinated with the leaf economics spectrum in Mediterranean rangeland species. *Funct. Ecol.* **32**, 612–625 (2018).
55. Roumet, C. et al. Root structure-function relationships in 74 species: evidence of a root economics spectrum related to carbon economy. *N. Phytol.* **210**, 815–826 (2016).
56. Martin, A. R., Doraisami, M. & Thomas, S. C. Global patterns in wood carbon concentration across the world's trees and forests. *Nat. Geosci.* **11**, 915–920 (2018).
57. Mladkova, P., Mladek, J., Hejduk, S., Hejcman, M. & Pakeman, R. J. Calcium plus magnesium indicates digestibility: the significance of the second major axis of plant chemical variation for ecological processes. *Ecol. Lett.* **21**, 885–895 (2018).
58. Liese, R., Alings, K. & Meier, I. C. Root branching is a leading root trait of the plant economics spectrum in temperate trees. *Front. Plant Sci.* **8**, 1–12 (2017).
59. Cadotte, M. W., Davies, T. J. & Peres-Neto, P. R. Why phylogenies do not always predict ecological differences. *Ecol. Monogr.* **87**, 535–551 (2017).
60. Keller, A. B. & Phillips, R. P. Leaf litter decay rates differ between mycorrhizal groups in temperate, but not tropical, forests. *N. Phytol.* **222**, 556–564 (2019).
61. Comas, L. H. & Eissenstat, D. M. Patterns in root trait variation among 25 co-existing North American forest species. *N. Phytol.* **182**, 919–928 (2009).
62. Prescott, C. E. Litter decomposition: what controls it and how can we alter it to sequester more carbon in forest soils? *Biogeochemistry* **101**, 133–149 (2010).
63. Rytter, R.-M. & Rytter, L. Carbon sequestration at land use conversion - Early changes in total carbon stocks for six tree species grown on former agricultural land. *Ecol. Manag.* **466**, 118129 (2020).
64. Badeau, V., Dambrine, E. & Walter, C. Proprietes des sols forestiers français: resultats du premier inventaire systematique. *Etude Gest. des. Sols* **6**, 165–180 (1999).
65. Lucas-Borja, M. E. et al. Effects of different land uses (abandoned farmland, intensive agriculture and forest) on soil hydrological properties in Southern Spain. *Water* **11**, 503 (2019).
66. Yesilonis, I., Szlavecz, K., Pouyat, R., Whigham, D. & Xia, L. Historical land use and stand age effects on forest soil properties in the Mid-Atlantic US. *Ecol. Manag.* **370**, 83–92 (2016).
67. Laughlin, D. C., Leppert, J. J., Moore, M. M. & Sieg, C. H. A multi-trait test of the leaf-height-seed plant strategy scheme with 133 species from a pine forest flora. *Funct. Ecol.* **24**, 493–501 (2010).
68. Austin, A. T., Soledad Mendez, M. & Ballare, C. L. Photodegradation alleviates the lignin bottleneck for carbon turnover in terrestrial ecosystems. *Proc. Natl Acad. Sci. U.S.A.* **113**, 4392–4397 (2016).
69. Augusto, L., Davies, T. J., Delzon, S. & de Schrijver, A. The enigma of the rise of angiosperms: can we untie the knot? *Ecol. Lett.* **17**, 1326–1338 (2014).
70. Kunstler, G. et al. Plant functional traits have globally consistent effects on competition. *Nature* **529**, 204–209 (2016).
71. Seyfried, G. S., Dalling, J. W. & Yang, W. H. Mycorrhizal type effects on leaf litter decomposition depend on litter quality and environmental context. *Biogeochemistry*. **155**, 21–38 (2021).
72. Luyssaert, S. et al. Trade-offs in using European forests to meet climate objectives. *Nature* **562**, 259–262 (2018).
73. Lewis, S. L., Wheeler, C. E., Mitchard, E. T. A. & Koch, A. Regenerate natural forests to store carbon. *Nature* **568**, 25–28 (2019).
74. Gatti, L. V. et al. Amazonia as a carbon source linked to deforestation and climate change. *Nature* **595**, 388–393 (2021).
75. Bonan, G. B. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008).
76. Rana, G. et al. Tree transpiration in a multi-species Mediterranean garden. *Agric. For. Meteorol.* **280**, 1–13 (2020).
77. Schwaab, J. et al. Increasing the broad-leaved tree fraction in european forests mitigates hot temperature extremes. *Sci. Rep.* **10**, 14153 (2020).
78. Terrer, C., Vicca, S., Hungate, B. A., Phillips, R. P. & Prentice, I. C. Mycorrhizal association as a primary control of the CO2 fertilization effect. *Science* **353**, 72–74 (2016).
79. Baeten, L. et al. Identifying the tree species compositions that maximize ecosystem functioning in European forests. *J. Appl. Ecol.* **56**, 733–744 (2019).
80. Bellassen, V. & Luyssaert, S. Managing forests in uncertain times. *Nature* **506**, 153–155 (2014).
81. Schrumpf, M., Schulze, E.-D., Kaiser, K. & Schumacher, J. How accurately can soil organic carbon stocks and stock changes be quantified by soil inventories? *Biogeosciences* **8**, 1193–1212 (2011).
82. Dijkstra, F. A. & Fitzhugh, R. D. Aluminum solubility and mobility in relation to organic carbon in surface soils affected by six tree species of the northeastern United States. *Geoderma* **114**, 33–47 (2003).
83. Quideau, S. A., Graham, R. C., Chadwick, O. A. & Wood, H. B. Organic carbon sequestration under chaparral and pine after four decades of soil development. *Geoderma* **83**, 227–242 (1998).
84. Nunes, L., Coutinho, J., Nunes, L., Rego, F. & Lopes, D. Growth, soil properties and foliage chemical analysis comparison between pure and mixed stands of Castanea sativa Mill. and Pseudotsuga menziesii (Mirb.) Franco, in Northern Portugal. *Syst.* **20**, 496–507 (2011).
85. von Haden, A. C., Yang, W. H. & DeLucia, E. H. Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties. *Glob. Chang. Biol.* **26**, 3759–3770 (2020).
86. Lee, J., Hopmans, J. W., Rolston, D. E., Baer, S. G. & Six, J. Determining soil carbon stock changes: simple bulk density corrections fail. *Agric. Ecosyst. Environ.* **134**, 251–256 (2009).
87. Wendt, J. W. & Hauser, S. An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. *Eur. J. Soil Sci.* **64**, 58–65 (2013).
88. Pribyl, D. W. A critical review of the conventional SOC to SOM conversion factor. *Geoderma* **156**, 75–83 (2010).
89. Augusto, L. et al. Is 'grey literature' a reliable source of data to characterize soils at the scale of a region? A case study in a maritime pine forest in southwestern France. *Eur. J. Soil Sci.* **61**, 807–822 (2010).
90. Federer, C. A., Turcotte, D. E. & Smith, C. T. The organic fraction—bulk density relationship and the expression of nutrient content in forest soils. *Can. J. Res.* **23**, 1026–1032 (1993).
91. De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J. & Muys, B. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Sci. Soc. Am. J.* **69**, 500–510 (2005).
92. Rumpel, C. & Kogel-Knabner, I. Deep soil organic matter-a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**, 143–158 (2011).
93. Lal, R. Digging deeper: A holistic perspective of factors affecting soil organic carbon sequestration in agroecosystems. *Glob. Chang. Biol.* **24**, 3285–3301 (2018).
94. Callesen, I. et al. Carbon storage and nutrient mobilization from soil minerals by deep roots and rhizospheres. *Ecol. Manag.* **359**, 322–331 (2016).
95. Luo, Z., Wang, G. & Wang, E. Global subsoil organic carbon turnover times dominantly controlled by soil properties rather than climate. *Nat. Commun.* **10**, 1–10 (2019).
96. He, Y. et al. Radiocarbon constraints imply reduced carbon uptake by soils during the 21st century. *Science* **353**, 1419–1424 (2016).
97. Balesdent, J. et al. Carbon transfer from atmosphere to deep soil layers over the last 50 years. *Nature* **559**, 599–602 (2018).
98. De Vos, B. et al. Benchmark values for forest soil carbon stocks in Europe: Results from a large scale forest soil survey. *Geoderma* **251**, 33–46 (2015).
99. Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **65**, 10–21 (2014).
100. Wiesmeier, M. et al. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. *Geoderma* **333**, 149–162 (2019).
101. Krasilnikov, P., Marti, J.-J. I., Arnold, R. & Shoba, S. *A handbook of soil terminology, correlation and classification*. (Routledge, 2009).
102. Augusto, L., Achat, D. L., Jonard, M., Vidal, D. & Ringeval, B. Soil parent material - A major driver of plant nutrient limitations in terrestrial ecosystems. *Glob. Chang. Biol.* **23**, 3808–3824 (2017).
103. Chen, G., Hobbie, S. E., Reich, P. B., Yang, Y. & Robinson, D. Allometry of fine roots in forest ecosystems. *Ecol. Lett.* **22**, 322–331 (2019).

104. Enquist, B. J., Brown, J. H. & West, G. B. Allometric scaling of plant energetics and population density. *Nature* **395**, 163–165 (1998).
105. Hoffmann, N., Heinrichs, S., Schall, P. & Vor, T. Climatic factors controlling stem growth of alien tree species at a mesic forest site: a multispecies approach. *Eur. J. For. Res.* early view, (2020).
106. Vet, R. et al. A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus. *Atmos. Environ.* **93**, 3–100 (2014).
107. Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. World map of the Koppen-Geiger climate classification updated. *Meteorol. Z.* **15**, 259–263 (2006).
108. Hengl, T. et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* **12**, 1–40 (2017).
109. Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **6**, 249–263 (2014).
110. Hartmann, J. & Moosdorf, N. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochem. Geophys. Geosystems* **13**, Q12004 (2012).
111. Kattge, J. et al. TRY plant trait database – enhanced coverage and open access. *Glob. Chang. Biol.* **26**, 119–188 (2020).
112. Iversen, C. M. et al. A global fine-root ecology database to address below-ground challenges in plant ecology. *N. Phytol.* **215**, 15–26 (2017).
113. Wang, H. et al. The China plant trait database. *Ecology* **99**, 500 (2018).
114. Thomas, H. J. D. et al. Global plant trait relationships extend to the climatic extremes of the tundra biome. *Nat. Commun.* **11**, 1–12 (2020).
115. Brundrett, M. C. & Tedersoo, L. Resolving the mycorrhizal status of important northern hemisphere trees. *Plant Soil.* **454**, 3–34 (2020).
116. Harley, J. L. & Harley, E. L. A check-list of mycorrhiza in the British flora. *N. Phytol.* **105**, 1–102 (1987).
117. Chaudhary, V. B. et al. MycoDB, a global database of plant response to mycorrhizal fungi. *Sci. data* **3**, 1–10 (2016).
118. Isaac, M. E. et al. Intraspecific trait variation and coordination: root and leaf economics spectra in coffee across environmental gradients. *Front. Plant Sci.* **8**, 1196 (2017).
119. Clough, B. J., Curzon, M. T., Domke, G. M., Russell, M. B. & Woodall, C. W. Climate-driven trends in stem wood density of tree species in the eastern United States: Ecological impact and implications for national forest carbon assessments. *Glob. Ecol. Biogeogr.* **26**, 1153–1164 (2017).
120. Treurnicht, M. et al. Functional traits explain the Hutchinsonian niches of plant species. *Glob. Ecol. Biogeogr.* **29**, 534–545 (2020).
121. Kazakou, E. et al. Are trait-based species rankings consistent across data sets and spatial scales? *J. Veg. Sci.* **25**, 235–247 (2014).
122. Li, T. et al. Intraspecific functional trait variability across different spatial scales: a case study of two dominant trees in Korean pine broadleaved forest. *Plant Ecol.* **219**, 875–886 (2018).
123. Reich, P. B., Walters, M. B. & Ellsworth, D. S. From tropics to tundra: Global convergence in plant functioning. *Proc. Natl Acad. Sci. USA.* **94**, 13730–13734 (1997).
124. Albert, C. H., Grassein, F., Schurr, F. M., Vieilledent, G. & Violle, C. When and how should intraspecific variability be considered in trait-based plant ecology? *Perspect. Plant Ecol. Evol. Syst.* **13**, 217–225 (2011).
125. Pietsch, K. A. et al. Global relationship of wood and leaf litter decomposability: the role of functional traits within and across plant organs. *Glob. Ecol. Biogeogr.* **23**, 1046–1057 (2014).
126. Flores-Moreno, H. et al. Robustness of trait connections across environmental gradients and growth forms. *Glob. Ecol. Biogeogr.* **28**, 1806–1826 (2019).
127. Laliberté, E. Below-ground frontiers in trait-based plant ecology. *N. Phytol.* **213**, 1597–1603 (2017).
128. Ma, S. et al. Variations and determinants of carbon content in plants: a global synthesis. *Biogeosciences* **15**, 693–702 (2018).
129. Pompa-García, M. et al. Tissue carbon concentration of 175 Mexican forest species. *IForest* **10**, 754–758 (2017).
130. Jolliffe, I. T. Principal component analysis: a beginner's guide - I. Introduction and application. *Weather* **45**, 375–382 (1990).
131. Josse, J. & Husson, F. {missMDA}: A Package for Handling Missing Values in Multivariate Data Analysis. *J. Stat. Softw.* **70**, 1–31 (2016).
132. Zanne, A. E. et al. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
133. Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y. & Wang, X.-Q. Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS One* **9**, e107679 (2014).
134. Saladin, B. et al. Fossils matter: improved estimates of divergence times in Pinus reveal older diversification. *BMC Evol. Biol.* **17**, 95 (2017).
135. Ran, J.-H., Shen, T.-T., Liu, W.-J., Wang, P.-P. & Wang, X.-Q. Mitochondrial introgression and complex biogeographic history of the genus Picea. *Mol. Phylogenet. Evol.* **93**, 63–76 (2015).
136. Xiang, Q.-P. et al. Phylogenetic relationships, possible ancient hybridization, and biogeographic history of Abies (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Mol. Phylogenet. Evol.* **82**, 1–14 (2015).
137. Magallon, S., Gomez-Acevedo, S., Sanchez-Reyes, L. L. & Hernandez-Hernandez, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *N. Phytol.* **207**, 437–453 (2015).
138. Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst. Biol.* **54**, 575–594 (2005).
139. Renner, S. S., Grimm, G. W., Schneeweiss, G. M., Stuessy, T. F. & Ricklefs, R. E. Rooting and dating maples (Acer) with an uncorrelated-rates molecular clock: implications for North American/Asian disjunctions. *Syst. Biol.* **57**, 795–808 (2008).
140. Grimm, G. W. & Renner, S. S. Harvesting Betulaceae sequences from GenBank to generate a new chronogram for the family. *Bot. J. Linn. Soc.* **172**, 465–477 (2013).
141. Thornhill, A. H., Ho, S. Y. W., Kuelheim, C. & Crisp, M. D. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Mol. Phylogenet. Evol.* **93**, 29–43 (2015).
142. Hipp, A. L. et al. Genomic landscape of the global oak phylogeny. *N. Phytol.* **226**, 1198–1212 (2020).
143. Elser, J. J. et al. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol. Lett.* **10**, 1135–1142 (2007).
144. Sullivan, P. F., Ellison, S. B. Z., McNown, R. W., Brownlee, A. H. & Sveinbjornsson, B. Evidence of soil nutrient availability as the proximate constraint on growth of treeline trees in northwest Alaska. *Ecology* **96**, 716–727 (2015).
145. Gurevitch, J. & Hedges, L. V. Statistical issues in ecological meta-analyses. *Ecology* **80**, 1142–1149 (1999).
146. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R. N.* **2**, 18–22 (2002).
147. Hebbali, A. olsrr: Tools for Building OLS Regression Models. (2020).
148. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
149. Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P. & Makowski, D. {performance}: An {R} Package for Assessment, Comparison and Testing of Statistical Models. *J. Open Source Softw.* **6**, 3139 (2021).
150. Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research. (2021).
151. Ooms, J. magick: Advanced Graphics and Image-Processing in R. (2021).
152. Fox, J. RcmdrMisc: R Commander Miscellaneous Functions. (2020).
153. Kowarik, A. & Templ, M. Imputation with the {R} Package {VIM}. *J. Stat. Softw.* **74**, 1–16 (2016).
154. Lüdecke, D. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *J. Open Source Softw.* **3**, 772 (2018).
155. Fox, J. & Weisberg, S. *An {R} Companion to Applied Regression*. (Sage, 2019).
156. Calcagno, V. glmulti: Model Selection and Multimodel Inference Made Easy. (2020).
157. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
158. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots. (2020).
159. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using {lme4}. *J. Stat. Softw.* **67**, 1–48 (2015).
160. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous Inference in General Parametric Models. *Biometrical J.* **50**, 346–363 (2008).
161. Konietschke, F., Placzek, M., Schaarschmidt, F. & Hothorn, L. A. {nparcomp}: An {R} software package for nonparametric multiple comparisons and simultaneous confidence intervals. *J. Stat. Softw.* **64**, 1–17 (2015).

Acknowledgements

This study is part of the CARTON project (ANR-19-CE32-0006), supported by the French *Agence Nationale de la Recherche* (ANR). Part of the trait data were provided by the TRY initiative on plant traits. We are thankful to the following colleagues for their help during the data collection and handling, or the interpretation of results: Andreas Altinalmazis-Kondylis, Edouard Davin, Sylvain Delzon, Nicolas Fanin, Alain Franc, Thomas Guzman, Hervé Jactel (ORPHEE experimental platform), Sébastien Lafont, Tania Maxwell, Céline Meredieu, Christophe Moisy, Christophe Nguyen, Jérôme Ogée, Noémie Pousse, Bruno Ringeval, Helga Van Miegroet, Cyrille Violle, and Jean-Pierre Wigneron.

Author contributions

L.A. initiated the study and wrote the first version of the manuscript. L.A. and A.B. contributed to the design of the study, collecting and processing data, interpreting the results, and writing the successive versions of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-28748-0.

Correspondence and requests for materials should be addressed to Laurent Augusto or Antra Boča.

Peer review information *Nature Communications* thanks Rémy Beugnon, Helge Bruelheide, Iván Prieto and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2022