

Citizen science plant observations encode global trait patterns

Received: 16 February 2022

Accepted: 9 September 2022

Published online: 20 October 2022

 Check for updates

Sophie Wolf¹✉, Miguel D. Mahecha^{1,2,3}, Francesco Maria Sabatini^{1,3,4,5}, Christian Wirth^{1,3,6,7}, Helge Brüelheide^{1,3,5}, Jens Kattge^{1,3,7}, Álvaro Moreno Martínez^{1,8}, Karin Mora^{1,3} and Teja Kattenborn^{1,3}

Global maps of plant functional traits are essential for studying the dynamics of the terrestrial biosphere, yet the spatial distribution of trait measurements remains sparse. With the increasing popularity of species identification apps, citizen scientists contribute to growing vegetation data collections. The question emerges whether such opportunistic citizen science data can help map plant functional traits globally. Here we show that we can map global trait patterns by complementing vascular plant observations from the global citizen science project iNaturalist with measurements from the plant trait database TRY. We evaluate these maps using sPlotOpen, a global collection of vegetation plot data. Our results show high correlations between the iNaturalist- and sPlotOpen-based maps of up to 0.69 (r) and higher correlations than to previously published trait maps. As citizen science data collections continue to grow, we can expect them to play a significant role in further improving maps of plant functional traits.

Global maps of plant functional traits provide an indispensable foundation for understanding interactions of the environment and the terrestrial biosphere. Information on plant traits, such as tissue properties or morphological characteristics, is urgently needed as input for dynamic global vegetation models^{1,2} or to understand geographic patterns of plant community structure and functional diversity^{3,4}. Yet, the spatial distribution of plant trait observations remains sparse^{5,6}.

Measuring traits directly often involves destructive sampling and expensive laboratory protocols⁷. Plant trait databases, such as the TRY plant trait database^{8,9}, curate these single measurements from numerous independent studies and projects. Although TRY provides an impressive collection of trait measurements, its cover cannot display global trait patterns^{9,10}. Alternative approaches have upscaled trait expressions from trait databases for a few traits^{5,11–14}. However, the resulting maps still feature huge uncertainties as they extrapolate

into unknown areas^{12,15}—uncertainties reflected in often-large discrepancies among the different trait maps¹⁶. The potential of satellite remote sensing to generate large-scale maps of trait distributions has also been explored^{6,17}. Yet, remote sensing retrieves only a few traits^{18–20} and primarily informs on upper canopies, while lower layers may not be well represented²¹.

One potential source to complement existing trait data could be crowd-sourced observations classified via automated species classification²², which is increasingly accessible to a wide audience via smartphone apps (for a recent comparison of apps see Jones²³). Such technology now enables collecting plant observations in very large sample sizes^{24–27}. Recent evaluations of crowd-sourced data demonstrate the potential of using crowd-sourced plant occurrence data to study species distribution and macroecological floristic gradients^{28,29}. iNaturalist, one of the largest and increasingly popular citizen science

¹Remote Sensing Centre for Earth System Research, Leipzig University, Leipzig, Germany. ²Remote Sensing Centre for Earth System Research, Helmholtz Centre for Environmental Research, UFZ, Leipzig, Germany. ³German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. ⁴BIOIME Lab, Department of Biological, Geological and Environmental Sciences (BiGeA), Alma Mater Studiorum University of Bologna, Bologna, Italy. ⁵Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Halle, Germany. ⁶Institute of Systematic Botany and Functional Biodiversity, Leipzig University, Leipzig, Germany. ⁷Max Planck Institute for Biogeochemistry, Jena, Germany. ⁸Image Processing Laboratory, Universitat de València, València, Spain. ✉e-mail: sophie.wolf@uni-leipzig.de

projects, has motivated users to collect and verify a total of over 14 million vascular plant ‘research-grade’ observations³⁰ (Extended Data Fig. 1). Considering the growth rate and the fact that iNaturalist is only one among many citizen science projects that add to the global boost in vegetation data availability, we can expect that such data will play a crucial role in our understanding of global plant functions in ecosystems³¹.

Yet, the actual value of using citizen science plant occurrence data for mapping global trait patterns remains largely unknown. Many citizen science projects, including iNaturalist, have no sampling design, so the resulting datasets cannot be considered to be representative of space, time, vegetation types or taxa. Biases may depend on site popularity or accessibility, cell phone reception, users’ interest in specific plants (for example, noticeable flowers) or temporal patterns in user activity, such as during vacations, weekends or linked to plant phenology^{28,32–35}. However, if we were able—despite all the potential biases—to make use of the functional information encoded in these data and extract reliable trait information from the iNaturalist citizen science observations, we would have a quickly growing data source for biodiversity monitoring and long-term observation at our disposal.

Here, we explore whether opportunistic iNaturalist research-grade occurrence data of vascular plants³⁶ could help map trait patterns on a global scale; we assess whether these data reflect global functional community composition when complemented with trait information from the global plant trait database TRY⁹ and aggregated spatially into global trait grids. As a reference for evaluating the iNaturalist observations, we use trait community-weighted mean (cwm) from the database sPlotOpen^{37,38}. Database sPlotOpen is curated from globally distributed plots with vegetation abundance measurements, balanced over global climate and soil conditions. It provides a cwm for each vegetation plot for 18 traits. These cwm are also derived from TRY measurements. We thus compare the functional trait information derived from spatially and taxonomically biased occurrence samples provided by iNaturalist citizen scientists to professionally sampled environmentally balanced plot-based abundance weighted data.

Results

iNaturalist and sPlotOpen observation density and cover

We matched the iNaturalist species observations of vascular plants ($n = 14,019,405$) with trait measurements from TRY, using the species name as the link. We were able to match 85% of the iNaturalist vascular plant observations and 32% of the species. A total of 55% of the species listed in the TRY database for the traits of interest were matched. This provided us with nearly 12 million observations ($n = 11,895,453$) of 28,500 vascular plant species, for which we have information on at least one trait (see Table 1 for sample sizes of each trait).

The density and distribution of iNaturalist observations after linking to TRY in comparison to the sPlotOpen plot data ($n = 95,104$) show an almost global coverage (Fig. 1); for density before linking to TRY see Extended Data Fig. 2. There is a strong imbalance in the density of observations, with some regions of the world, especially North America and Europe, being over-represented with around 100,000 observations per cell, while others are represented with only one or two observations per cell. The sPlotOpen data, while showing lower coverage, are also, by design, less biased in density towards these regions³⁸.

Density and distribution in climate space (temperature/precipitation) paint a similar picture (Fig. 1): while iNaturalist observations cover a larger portion of this space, it is also more biased in the number of observations, specifically toward the temperate regions with moderate rainfall, that is the temperate deciduous forest and temperate shrubland Whittaker biomes³⁹.

The density of observations varies greatly across terrestrial biomes⁴⁰. Among the 14 biomes, ‘temperate forests’ and ‘Mediterranean forests, woodlands and shrubs’ are best represented by iNaturalist observations with a density of about 0.4 observations per km². Temperate grasslands have a medium density of 0.13 per km², while all other

Table 1 | Plant traits

Trait ^a	TRY trait name ^b	Unit ^b	iNat observations ^c
Leaf area	Leaf area (in case of compound leaves: leaflet, undefined if petiole is included or excluded)	mm ²	4.9×10 ⁶
Specific leaf area (SLA)	Leaf area per leaf dry mass (SLA or 1/LMA: undefined if petiole is included or excluded)	m ² kg ⁻¹	7.8×10 ⁶
Stem specific density (SSD)	SSD or wood density (stem dry mass per stem fresh volume)	gcm ⁻³	3.4×10 ⁶
Leaf C	Leaf carbon (C) content per leaf dry mass	mg g ⁻¹	5.1×10 ⁶
Leaf N	Leaf nitrogen (N) content per leaf dry mass	mg g ⁻¹	7.2×10 ⁶
Leaf P	Leaf phosphorus (P) content per leaf dry mass	mg g ⁻¹	5.2×10 ⁶
Plant height	Plant height vegetative	m	9.8×10 ⁶
Seed mass	Seed dry mass	mg	1.0×10 ⁷
Seed length	Seed length	mm	3.6×10 ⁶
Leaf dry matter content (LDMC)	Leaf dry mass per leaf fresh mass (LDMC)	gg ⁻¹	6.9×10 ⁶
Leaf N	Leaf nitrogen (N) content per leaf area	gm ⁻²	5.9×10 ⁶
Leaf N/P ratio	Leaf nitrogen/phosphorus (N/P) ratio	gg ⁻¹	3.9×10 ⁶
Leaf delta 15N	Leaf nitrogen (N) isotope signature (delta 15N)	ppm	1.0×10 ⁶
Seeds per reproduction unit	Seed number per reproduction unit	-	4.0×10 ⁶
Leaf fresh mass	Leaf fresh mass	g	2.8×10 ⁶
Stem conduit density	Stem conduit density (vessels and tracheids)	mm ⁻²	1.3×10 ⁶
Dispersal unit length	Dispersal unit length	mm	4.8×10 ⁶
Conduit element length	Wood vessel element length; stem conduit (vessel and tracheids) element length	μm	3.0×10 ⁵

^aTraits used in this analysis; selection based on all continuous traits from Bruehl et al.³.

^bTrait name and unit as used in TRY database. ^cNumber of iNaturalist (iNat) observations linked with each trait.

biomes are represented by much fewer observations, with densities ranging from 0.02 in tropical and boreal forests and deserts down to only 0.004 observations per km² in the tundra (Supplementary Table 1).

Global trait maps

From the merged iNaturalist–TRY data we then created global trait distribution maps based on these citizen science observations. We generated a global spatial grid and calculated the mean trait value of each grid cell for a set of 18 traits (Table 1) linked to iNaturalist observations. An example of such a map is shown in Fig. 2 for the trait leaf area (see Extended Data Fig. 3 for all 18 traits).

Spatial correlation of iNaturalist and sPlotOpen

We then investigated the feasibility of using iNaturalist observations, annotated with TRY trait measurements, to produce global trait maps by comparing them to sPlotOpen vegetation plot data. We generated trait maps using the sPlotOpen in the same manner as described above for the iNaturalist observations. We correlated (correlation coefficient r

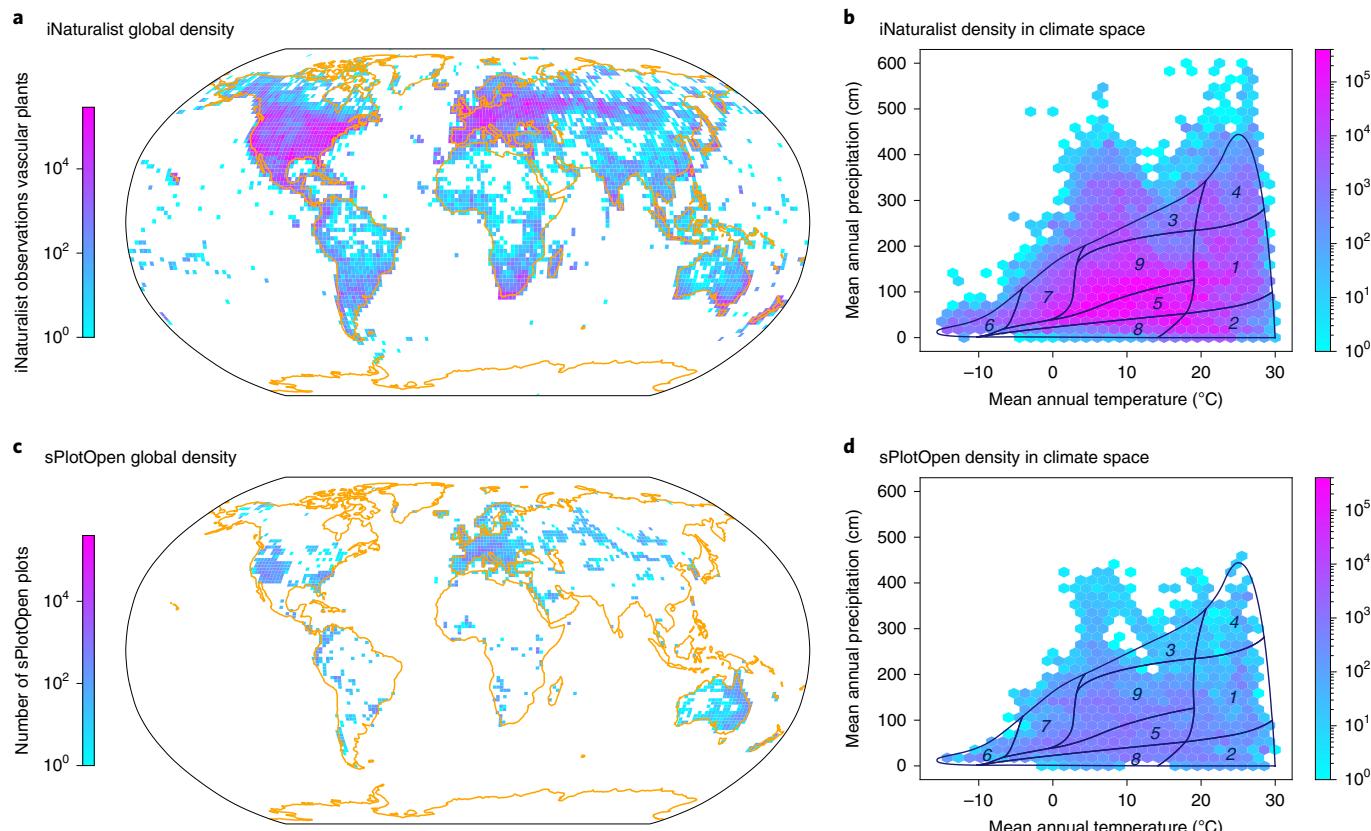


Fig. 1 | Density and distribution of iNaturalist and sPlotOpen datasets.

a, Density of iNaturalist vascular plant observations after linking to TRY database ($n = 11,895,453$); 2° resolution (221 km grid size at the equator). Colour corresponds to the number of observations per cell. For density before linking to TRY see Extended Data Fig. 2. **b**, Density of the iNaturalist in mean annual temperature ($^{\circ}\text{C}$) and precipitation (mm) climate space. Whittaker biomes are

numbered as follows: (1) tropical seasonal forest/savanna, (2) (sub)tropical desert, (3) temperate rain forest, (4) tropical rain forest, (5) woodland/shrubland, (6) tundra, (7) boreal forest, (8) temperate grassland/desert and (9) temperate seasonal forest. **c,d**, Global density of sPlotOpen plots (**c**) and density of the sPlotOpen in mean annual temperature ($^{\circ}\text{C}$) and precipitation (mm) climate space (**d**) ($n = 95,104$).

weighted by area of grid cell) the two trait maps at different resolutions, ranging from 0.06° to 4° (7–440 km at the equator) (Fig. 3a; Methods).

At a 2° resolution we observed the highest correlations for several traits, with a median r of 0.46 over all traits. The following traits exhibited the strongest correlations (r): stem conduit density (0.68), stem specific density (SSD) (0.63), leaf fresh mass (0.59), leaf area (0.59), leaf nitrogen (N) per area (0.59), plant height (0.58) and specific leaf area (SLA) (0.56) (Fig. 3b; see Extended Data Fig. 4 for correlation plots for all traits). At the even coarser 4° resolution, some traits exhibited a higher correlation, such as stem conduit density ($r = 0.69$) or leaf fresh mass ($r = 0.66$), while others, such as plant height and leaf N per area, peaked at the 2° resolution.

We calculated a standard major axis (SMA) regression for each trait (displayed in red lines in Fig. 3b), which showed slopes significantly different from the 1:1 line for some traits ($P \ll 0.01$ for H^0 : slope = 1). This indicates biases in the structure of the datasets: for larger values, iNaturalist observations tend to estimate higher SLA, leaf fresh mass, leaf N per area and stem conduit density in comparison to sPlotOpen. For SSD, this effect was particularly high (slope = 2.57, 95% confidence interval (CI) (2.45, 2.7)). For larger values, iNaturalist observations estimate lower plant height than sPlotOpen. For leaf area, the SMA regression revealed no bias ($P = 0.30$ for H^0 : slope = 1).

In a second approach, instead of aggregating sPlotOpen plots in a grid, we compared every plot cwm to the average trait measurement of all iNaturalist observations within a certain radius around each sPlotOpen plot (Extended Data Fig. 5). In this plot-wise comparison to iNaturalist, the cwm of the sPlotOpen vegetation plots were not

regionally aggregated (compared to the grid-based analysis above). We observed similar trends but a larger scattering in the correlation plots, which is reflected in lower r values (Extended Data Fig. 6).

Differences within biomes

To assess how well iNaturalist observations reflect sPlotOpen plant communities in different terrestrial biomes⁴⁰, the previously mentioned grids were generated for each biome separately and normalized. Figure 4 shows the pixel differences (iNaturalist – sPlotOpen) per biome and top trait (Supplementary Tables 1 and 2 give detailed statistics). In general, iNaturalist better resembles sPlotOpen in predominantly non-forest biomes. In tropical, temperate and boreal forests, SLA is overestimated by iNaturalist observations, with a median of 30%, 26% and 38%, respectively, while leaf N per area is underestimated by 15%, 6% and 20%. In all tropical and subtropical forest biomes and temperate coniferous forest biomes, we additionally observe an underestimation of plant height (ranging from a 55% underestimation in tropical and subtropical coniferous forests to 10% in temperate coniferous forests; see Extended Data Fig. 7). Tropical, subtropical, temperate and montane grasslands, Mediterranean vegetation and deserts, on the whole, exhibit smaller deviations. In most biomes, leaf area and leaf fresh mass are overestimated by 11% and 15% on average.

Comparison to previously published trait products

We also correlated the sPlotOpen maps to published trait maps, which are derived from extrapolation methods^{5,6,11–13} at 0.5° and 2° resolution. For all available published traits maps—leaf N per mass, leaf N per area

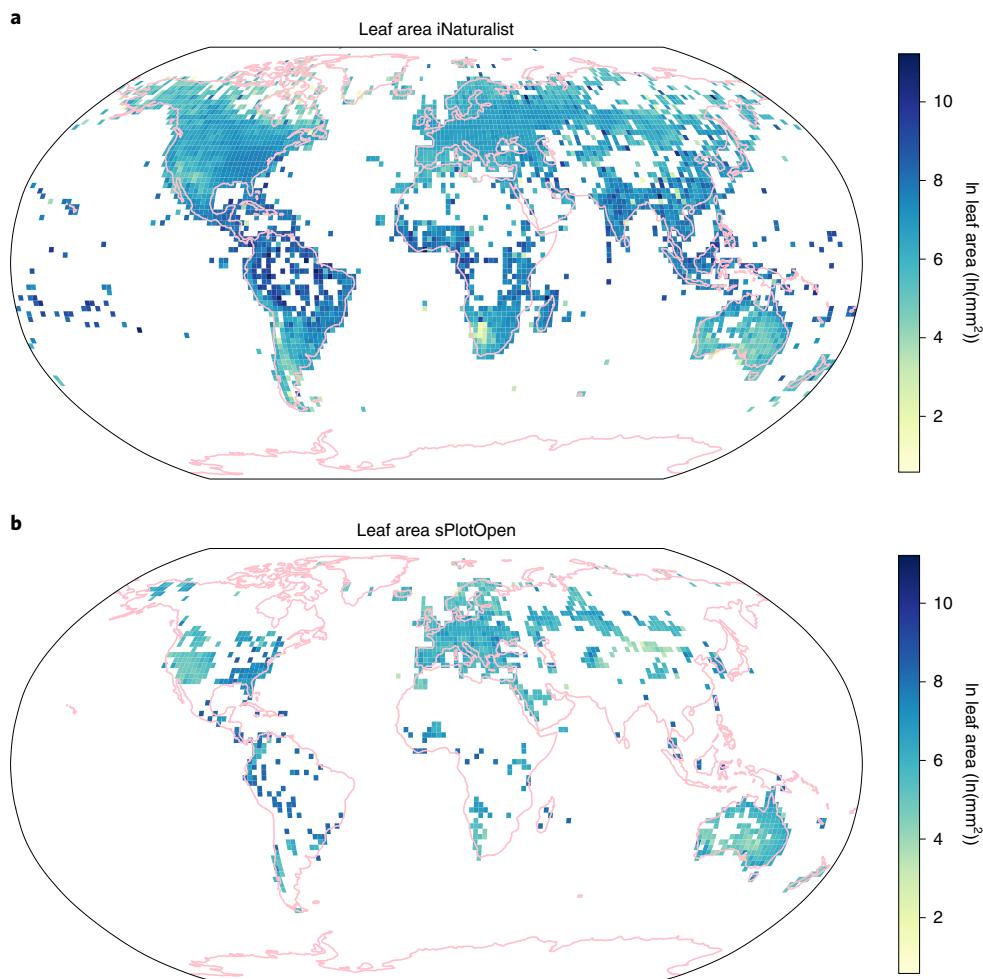


Fig. 2 | Example of a global trait map. **a, b,** In-transformed leaf area (mm^2), using iNaturalist observations (**a**) and sPlotOpen (**b**), respectively. For all iNaturalist trait maps see Extended Data Fig. 3.

and SLA—the weighted correlation (r) was higher for the iNaturalist maps (0.37, 0.59 and 0.56) than for previously published maps based on extrapolation (highest $r=0.25$, 0.50 and 0.51, respectively). Some of these trait maps showed no correlation with sPlotOpen (Table 2).

Discussion

A correlation of up to 0.69 (r) between these two fundamentally different datasets is astounding and unexpectedly high. We did not expect the iNaturalist maps to correlate more strongly with sPlotOpen than previously published trait maps (Table 2), which are all based on extrapolations of TRY data and climate variables. The higher correlation between iNaturalist and sPlotOpen could stem from similar biases in the two datasets. However, we do not assume this explanation to be the most likely: iNaturalist occurrence data are noisy and heterogeneous, shared by users who sample the species they encounter and find interesting; sPlotOpen is a data collection of vegetation plots that were each measured and recorded within a framework of very specific research questions and comes with its own array of biases. And yet, occurrence samples provided by iNaturalist users resemble professionally sampled plot-based abundance data on a level of functional traits.

We found higher correlations between sPlotOpen- and iNaturalist-based trait maps at larger grid cell sizes. We assume that larger grid cell sizes generally compensate for the sparsity present in both datasets, which eventually increases the correspondence across macroecological gradients. When data are spatially aggregated, the

finer spatial details are smoothed out, preserving the more broad climatological gradients (an effect described, for instance, in Joswig et al.⁴¹). In this way, abundance differences of species—which are captured by the cwm trait values in sPlotOpen data but not by the iNaturalist occurrences—become less important, the broader the spatial scale on which they are aggregated.

For several traits (for example, SSD, seed mass, leaf N per area and stem conduit density) the global trait maps presented here are the first published and evaluated global trait maps, to our knowledge. These traits have been previously identified as relevant in the study of functional diversity but no global maps of them existed to date^{3,42}. Other authors have published global trait distribution maps for a subset of traits we present here but they commonly follow no general or independent approach to evaluate the maps^{5,11–13}. This is critical since most of them do extrapolate into unknown space^{43,44}. Here, we evaluate the global trait distribution maps using the sPlotOpen data. This approach may serve as an indicator to judge the reliability of the maps presented here for further use (Fig. 3a). All data and code used for this study are freely available so that functional gradient maps can be updated as the number of iNaturalist observations and trait data in TRY continue to grow.

The mismatches we do see between sPlotOpen and iNaturalist may result from biases and limits within either dataset. Data gaps found in the iNaturalist observations may be ascribed to low population density or accessibility. We can see a greater agreement among iNaturalist

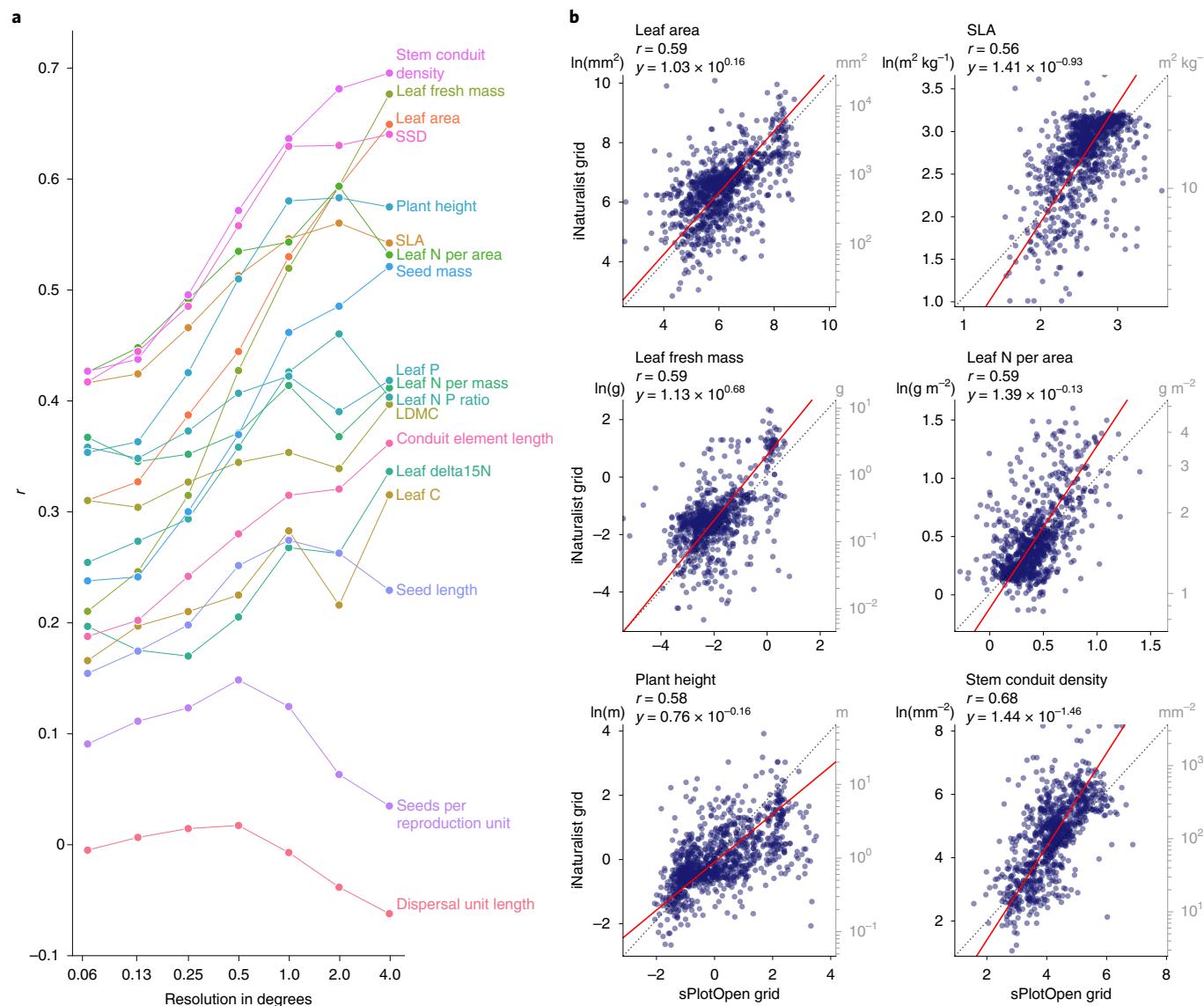


Fig. 3 | Pixel-by-pixel correlation of iNaturalist and sPlotOpen global trait maps. The correlation is quantified using a weighted correlation coefficient (r weighted by grid cell area). **a**, Relationship of r and spatial resolution (0.06°–4° resolution or about 7–440 km grid size at the equator). The lines connecting the points solely enhance readability. **b**, Scatter plots of sPlotOpen map pixel values plotted against the respective iNaturalist map pixel values (at a 2° spatial resolution); displayed here for traits with the highest iNaturalist sPlotOpen agreement (that is, weighted $r > 0.5$ and slope > 0.5 and < 2); see

Extended Data Fig. 4 for correlation plots of all traits. All trait values are ln-transformed, 1:1 line in dotted grey and SMA regression slope in red. For clarity, the secondary y axis on the right shows the raw trait values marked on a log scale, which correspond to the ln-transformed values on the left. Plot extents are the 0.01 and 0.99 quantiles of the data. The 95% CIs for the slope are: leaf area (0.98, 1.08), SLA (1.34, 1.49), leaf fresh mass (1.06, 1.19), leaf N per area (1.32, 1.46), plant height (0.72, 0.8) and stem conduit density (1.37, 1.51).

and sPlotOpen in biomes with a higher iNaturalist observation density (Fig. 4), indicating that more observations indeed lead to a more complete representation of a given plant community. This is supported by the observation that grid cell differences between iNaturalist and sPlotOpen are larger at low observation densities (Extended Data Fig. 8). Given the exponential growth of iNaturalist observations in the past (Extended Data 1), we can expect this dataset to close gaps prospectively and provide enough observations to make stratified sampling possible.

Naturally, biases in sPlotOpen may also contribute to differences between the two datasets. The sPlotOpen is a conglomeration of data collected in various studies with very specific research questions across various ecosystems^{37,38}. Additionally, the different vegetation plots are very heterogeneous in terms of plot size, sampling density or sampling

protocol. As a result, sPlotOpen is unable to reflect a balanced or stratified sample representative for a given region. Nonetheless, we use sPlotOpen plant community data as a baseline reference, as it provides the most representative and extensive information on the global distribution of plant communities. For a limited set of traits, other global maps have been published^{5,6,11–14}; however, since these products usually do not show a strong agreement, they cannot serve as a trustworthy reference (compare Schiller et al.¹⁶). Future studies, especially those using extrapolation (or upscaling) methods, may also compare their maps against sPlotOpen data.

We see in Fig. 3a that for some traits, iNaturalist and sPlotOpen show no or only very low correlation. One explanation for this disagreement could be the true lack of macroecological patterns: if a trait exhibits no macroecological pattern, we cannot expect there to be

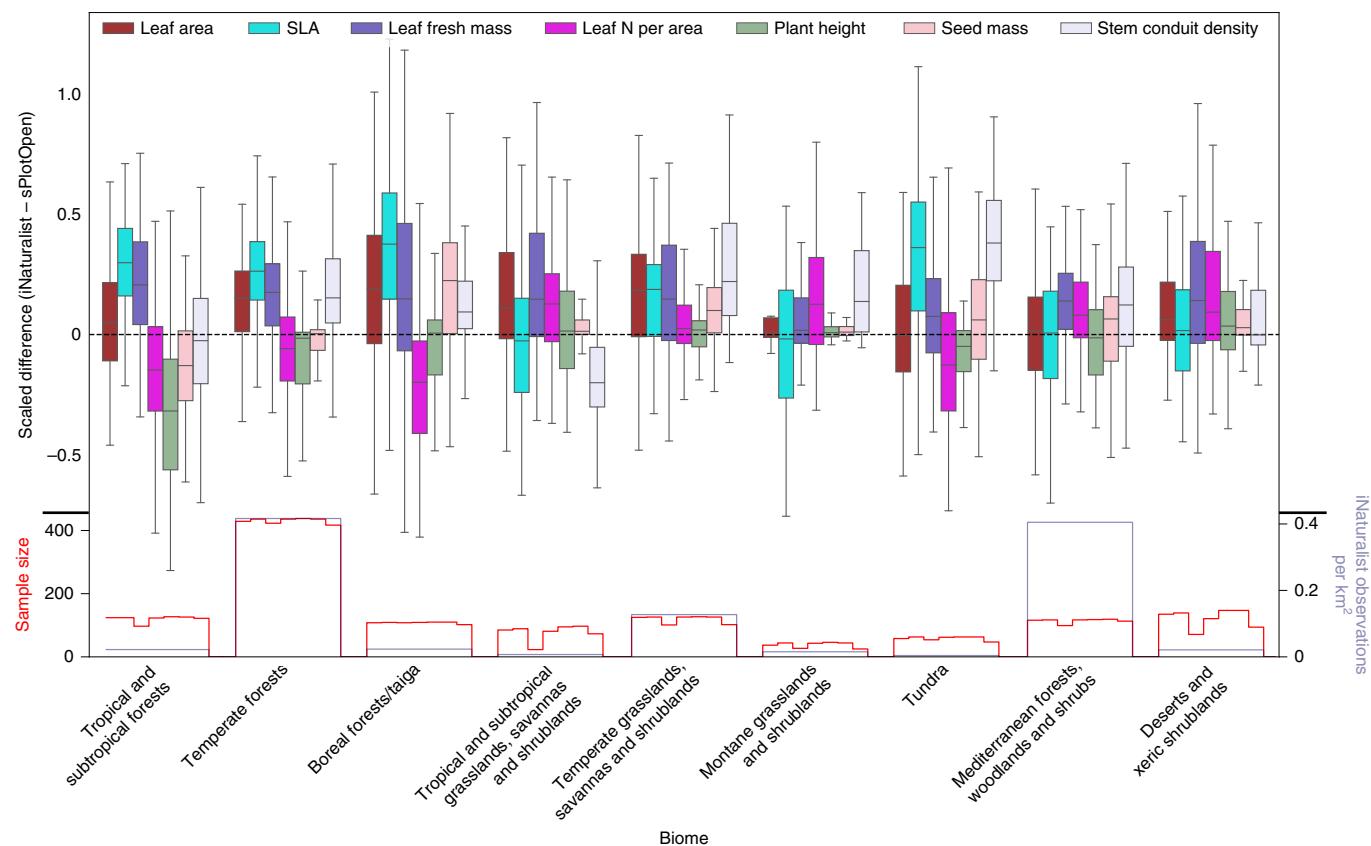


Fig. 4 | Difference between sPlotOpen cwm grid cell averages and iNaturalist observation averages per grid cell for each terrestrial biome. All traits were scaled by range ($-1, 1$) using the 0.05 and 0.95 quantiles. In this figure, all types of tropical and temperate forests were aggregated (for a detailed boxplot over all 14 WWF biomes see Extended Data Fig. 7). The bounds of the box are defined by the first and third quartiles, the centre lines are the medians, the whiskers mark the

1.5 interquartile range and outliers are not shown. The red step-graph shows the sample size n = the number of iNaturalist map pixels that overlap the respective sPlotOpen map per biome and trait. The blue step-graph marks the mean density of iNaturalist vascular plant observations per km² in each biome. For exact sample sizes per biome and trait, see Supplementary Table 1.

Table 2 | Correlation of published maps with sPlotOpen

Dataset	Leaf N mass		Leaf N area		SLA	
	0.5°	2°	0.5°	2°	0.5°	2°
iNaturalist	0.37	0.37	0.53	0.59	0.51	0.56
Butler et al. ¹¹	0.24	0.19	0.41	0.43	0.31	0.34
Boonman et al. ¹²	0.08	-0.02	0.45	0.47	0.48	0.50
Moreno Martínez et al. ⁶	0.25	0.22	0.46	0.50	0.40	0.51
Vallicrosa et al. ¹⁴	0.23	0.09	-0.06	0.08	-	-
Madani et al. ¹³	-	-	-	-	-0.02	0.18
van Bodegom et al. ⁵	-	-	-	-	0.37	0.41

Correlation (r weighted per area of grid cell) of each trait product (aggregated mean ln-values in 0.5° and 2° grid) in relation to sPlotOpen. All—except for the iNaturalist maps—are based on extrapolations. Highest r is marked in bold for each category: for all three traits and in both 0.5° and 2° grids, the iNaturalist-based trait maps correspond most closely to the sPlotOpen grids.

a correlation (a monotonic relationship) between two datasets. The traits may vary greatly on a very small scale, both spatially and temporally, such as leaf N per mass or leaf C. Other traits, such as seed length might simply be less informative than seed mass, for which we see a stronger correlation (0.23). Wood vessel length is a trait only relevant to woody species, which eliminates many observations.

When comparing differences between iNaturalist and sPlotOpen among biomes, we see that iNaturalist observations better represent sPlotOpen vegetation plots in biomes not dominated by trees (that is, grasslands, shrublands and deserts; Fig. 4). Herbaceous species and communities tend to have larger SLA^{3,42}. The overestimation of SLA in forest biomes and the underestimation of plant height in tropical/subtropical and temperate coniferous forest biomes by iNaturalist observations point to iNaturalist users' partiality towards small, herbaceous plants. Users might be inclined to capture not the characteristic dominant species or functional plant types of a plant community but rather the noticeable, yet rare, individuals. This bias could lead to an undersampling of trees in forest biomes¹⁶. The comparison of growth form (tree/shrub/herb) cover between iNaturalist and sPlotOpen supports this observation: tree cover seems to be underestimated and herb cover overestimated by iNaturalist observations in temperate forests (Extended Data Fig. 9). This effect might even be potentiated by the technical difficulty of photographing large shrubs and trees, making it hard for other users to confirm and agree on the species identification from a photo alone, creating a bottleneck as to which observations make it to the 'research-grade' level. Also, trees are weighted higher in the sPlotOpen cwm due to a larger cover per individual plant. The iNaturalist approach does not adjust the weight on the basis of cover but on occurrence and therefore large plants, that is trees, will be weighted less. This, in turn, may result in the observed underestimation of plant height in tropical forest biomes. For traits with a skewed distribution, such as height or SLA in forests, this effect may be amplified.

The overestimation of leaf area by 11% on average in all biomes points to an undersampling of grasses (Fig. 4). This bias might again be the result of the difficulty identifying grass species in general and the added difficulty of identifying grass species from photos to confirm an observation. More knowledge of citizen scientists' sampling behaviour could allow for correcting for these biases^{27,29}.

Both iNaturalist and sPlotOpen are annotated with trait information taken from the TRY database. For this study, average trait measurements from TRY data were simply linked via species name without considering intraspecies trait plasticity. Future studies may include intraspecific trait variation, for example, by considering the relationship with climate^{6,45,46} or using gap-filled trait data or by estimating the trait variability from iNaturalist photos using a machine-learning approach¹⁶. An additional analysis (Extended Data Fig. 10) shows that the trait maps produced by Schiller et al.¹⁶ to check the plausibility of estimating trait values from iNaturalist photos are indeed very similar to the iNaturalist maps based on species classification presented here. These findings provide a promising basis for future work, which may integrate multiple facets of citizen science contribution to global trait maps—from species identifications to photographic records. Also, this approach may be extended to other data sources, such as incorporating all 300 million vascular plant occurrences from the Global Biodiversity Information Facility (GBIF). Using (and understanding) the complete GBIF occurrence data, however, is less trivial: it contains more individual biases, mixed datasets from national or institutional inventories, citizen science and research projects.

The results presented here open up a promising avenue for the use of citizen science data to help fill the spatial gaps in plant trait data. These data can provide invaluable insight, especially in regions where trait distributions have so far only been estimated by extrapolation. As our results showed, the iNaturalist trait maps correspond better to sPlotOpen than do previously published maps (Table 2) at the spatial scales considered here (0.5° and 2° resolution). So far, dynamic global vegetation models incorporate only a very crude representation of functional properties of plants, for example encoded in coarse functional types. Spatially explicit information on plausible ranges of traits, such as leaf N, leaf C, leaf area, SLA, seed mass and SSD, are fundamental to the accurate representation of photosynthesis⁴⁷. Inaccurate data on these traits propagate biases in representing the carbon and coupled water and nutrient cycles. Today, individual-based dynamic vegetation models of forests and grasslands are increasingly capable of modelling vegetation dynamics at the biome scale. Soon they will be of global relevance^{5,47,48} and require reliable global trait information. We call for considering the advent of new citizen science efforts to join forces towards generating global trait maps for scientific purposes.

Already, iNaturalist observations cover a large spectrum in both geographical and climate space (Fig. 1) and can extend our view into unknown environmental areas. Many regions, both spatially and temporally, are still not well represented. BioBlitzes and other observation challenges within the iNaturalist and other citizen science communities have shown the readiness of volunteers to contribute to specific research questions and projects⁴⁹. Effectively designed initiatives can be used to target gaps in vegetation data and stimulate higher environmental awareness among contributors.

Conclusion

Our results suggest that citizen science plant species occurrence data, despite being constrained by several biases and lacking an integrated sampling strategy, can be used to derive plant trait patterns consistent with those generated using plant community data from scientific projects—and even with better approximation than previously published maps that used extrapolation. We present the first maps, to our knowledge, for several traits and provide a way to evaluate such maps using the sPlotOpen data. We found surprisingly high correspondence for multiple traits and were able to identify several systematic

biases. Several geographical regions are not adequately represented in citizen science data but the exponential growth of records suggests that spatial gaps may be filled within the next few years. Plant species occurrences documented by citizen scientists in concert with plant trait expressions curated from professional scientific studies, thus, can provide a promising data stream to uncover global patterns of plant form and function. Integrating this approach with other technologies, such as satellite-based remote sensing, machine learning for upscaling and combining above and below the canopy perspectives, might greatly improve our understanding of Earth's plant functional diversity. Together, our findings may open a new argument to foster and value citizen science approaches, such as iNaturalist, as they offer growing data treasures that can be integrated seamlessly into process-based terrestrial biosphere models.

Methods

iNaturalist research-grade observations

The iNaturalist research-grade data are openly accessible at the GBIF. The vascular plant (Tracheophyta) data were downloaded from www.gbif.org, version 'Darwin core archive' on 5 January 2022³⁶. An iNaturalist observation is granted research-grade when it meets the following criteria: it must include a date, a spatial georeference and a picture or sound; and the observed individual may not be captive or cultivated. Once two-thirds of users agree on a species identification (two-thirds consensus), the observation obtains research-grade and is uploaded to www.gbif.org. iNaturalist contributors have already collected more than 36 million research-grade observations—and more than 14 million such observations of vascular plants alone. Since 2017, iNaturalist has incorporated help identifying an observation using machine learning on both the website and smartphone app. This system of combining user-validation and machine-learning-based species suggestions makes these observations especially interesting as a potential data source in a scientific context. For this study, we used the 'date identified' to calculate the number of iNaturalist observations added every year. We used the species name to retrieve existing knowledge on functional traits from the TRY database—information that iNaturalist data do not include readily—and then used the coordinates to map trait distributions globally.

Linking iNaturalist to TRY

The TRY database contains trait measurements from individual plants and, typically, multiple individual measurements per trait and species. TRY is a heterogeneous collection of measurements, which include various life stages of plant species and a range of different growing conditions. Nonetheless, with well over 11 million trait records of nearly 280,000 plant taxa, this dataset currently provides the only plausible way we can assess plant traits on a global scale⁹. Measurements for the 18 traits of interest (Table 1) across all available species were downloaded from www.try-db.org. We used 18 traits frequently considered when characterizing plant functional gradients and functional diversity³. We ln-transformed (natural logarithm) the trait measurements for a more normal distribution, as shown in Kattge et al.⁹.

We calculated from all these individual measurements an average value per trait and species. The average trait values were then matched with the iNaturalist observations via the species name. The TRY original species names were standardized using the Taxonomic Names Resolution Service (TNRS)⁹. TNRS is based largely on 'The Plant List'. The GBIF backbone for vascular plants, with which the iNaturalist observations were standardized, relies heavily on 'The World Checklist of Vascular Plants' and 'The Leipzig Catalogue of Vascular Plants', both of which are continuations of The Plant List. We assume—since both species lists are harmonized using essentially the same lists—that a simple name matching of the two standardized species names lists already captures the majority of possible matches. By using such an exact match, first against the standardized TRY names and then the original

TRY names, we linked around 84% of the iNaturalist observations. After strict matching, we applied a conservative fuzzy match (rapidFuzz with cutoff=90). We were able to match another 1% of iNaturalist observations with trait information, about 85% of the iNaturalist observations in total. TRY data were not gap-filled.

sPlotOpen vegetation plot data

To test how well the iNaturalist observations resemble plant communities, we used data from sPlotOpen as a reference. The sPlotOpen³⁸ is an open-access and environmentally and spatially balanced subset ($n = 95,000$ plots) of the global sPlot vegetation plots dataset v.2.1 (ref. ³⁷). It was downloaded from the iDiv data repository <https://idat.idiv.de>. For this study, we used the ln-transformed cwm for all 18 traits listed in the sPlotOpen data (Table 1), a selection of traits based on all continuous traits from Bruelheide et al.³. The trait values in sPlotOpen are not measurements recorded on site in the vegetation plots; they are taken from the gap-filled version of the TRY database⁹ and matched via species name to calculate a cwm per plot and trait (see Bruelheide et al.³ for details). Beyond that, we used the geolocation (latitude/longitude) of each vegetation plot.

Distribution and density of observations

The distribution and density of iNaturalist observations and sPlotOpen plots were visualized spatially in a 2° grid. The distribution and density of iNaturalist observations and sPlotOpen vegetation plots in a climate space (precipitation/temperature) were visualized using the Python package ‘hexbin’. Average annual temperature and precipitation for the geolocation of each iNaturalist observation and each sPlotOpen vegetation plot were extracted from the freely available ‘WorldClim bio variables’, downloaded from www.worldclim.org. The Whittaker biome coordinates were taken from the R package ‘plotbiomes’ data <https://github.com/valentinitnelav/plotbiomes>.

Trait maps

For each trait separately, we generated a global spatial grid (latitude/longitude), where each pixel value represents the mean trait value of all iNaturalist observations located within the grid cell. This mean was calculated by spatially aggregating all iNaturalist observations within each grid cell and averaging the linked TRY trait values x_i associated with each observation i . We calculated the arithmetic mean of the ln-transformed trait values. The trait maps based on the sPlotOpen data were generated similarly, only with x_i being the ln-transformed cwm associated with each plot i . We generated the grid maps for all 18 traits and both datasets. These maps are available (see Data availability section) in two versions: one shows the arithmetic mean of the ln-transformed trait values x over all observations n and the other shows the geometric mean given by

$$\exp\left(\frac{1}{n} \sum_{i=1}^n \ln(x_i)\right).$$

Note the conceptual difference between the trait maps derived from sPlotOpen and iNaturalist: while the sPlotOpen trait maps are aggregations of cwm derived from plot-based species abundances, iNaturalist trait maps are derived from mean occurrence values. These means are not weighted by abundance (for example, biomass or cover), since such information is not available for iNaturalist observations. We assume that the observations reflect the species abundance in a community, even though the iNaturalist means are averaged by species occurrences only.

Correlation of iNaturalist and sPlotOpen

We tested several grid sizes (0.06°–4° resolution) to assess at what spatial resolution the iNaturalist trait maps most strongly resemble the trait patterns found in the sPlotOpen trait maps. The mean value

for each trait in each grid cell in the iNaturalist trait map was then correlated to the respective cell in the sPlotOpen trait map. We used a weighted Pearson's correlation coefficient (r) to assess the correlation between the two maps. The weight of each mean trait value corresponds to the area of each grid cell (Supplementary Information).

We calculated slopes and intercepts using a model II regression, specifically the SMA regression since we were not predicting one dataset using the other⁵⁰. The SMA regression slope, 95% CIs and a two-sided t -statistic (H^0 : slope = 1) were calculated using the R package ‘smatr’⁵¹.

Aggregation in buffers around sPlotOpen plots

For the alternative approach of using a buffer around each sPlotOpen vegetation plot, we aggregated all iNaturalist observations around each plot individually within a certain radius or buffer and calculated the mean trait measurement inside the buffer. To obtain more equal area buffers, the latitudinal/longitudinal data were projected into ‘world sinusoidal projection’ (ESRI:54008). The iNaturalist observations within each buffer were aggregated and, from these, the average trait values were calculated. The correlation coefficient (Pearson's) was calculated from sPlotOpen cwm and the corresponding mean trait values of the iNaturalist observations inside the buffer.

Correlating previously published maps with sPlotOpen

We compared the sPlotOpen maps with trait maps published in previous studies. These maps of SLA, leaf nitrogen per mass and leaf nitrogen per area were all spatially extrapolated from sparse TRY observations. We spatially resampled the previously published trait maps^{5,6,11–14} to intersect with the sPlotOpen-based grids and aggregated at a 0.5° and 2° resolution. We then calculated the weighted r , as described above for the iNaturalist maps.

Difference of iNaturalist and sPlotOpen across biomes

The correspondence of average trait expressions between iNaturalist and sPlotOpen cwm was evaluated across terrestrial biomes, as defined in the WWF terrestrial ecoregion map⁴⁰. The WWF terrestrial-ecoregions shape files were downloaded from www.world-wildlife.org/publications/terrestrial-ecoregions-of-the-world. The normalized trait values x' are given by the arithmetic means of each ln-transformed trait value per grid cell, x_{ln} , normalized to a (-1,1) range using the 0.05 and 0.95 quantiles:

$$x' = \frac{\exp(x_{ln}) - q_{0.05}(\exp(x_{ln}))}{q_{0.95}(\exp(x_{ln})) - q_{0.05}(\exp(x_{ln}))}.$$

On the basis of the WWF terrestrial-ecoregions shape files, a separate trait map was calculated for each biome and for the traits that had shown the highest correlation globally—stem conduit density, leaf area, leaf fresh mass, plant height, leaf nitrogen (N) per area and SLA. The iNaturalist aggregated trait value per grid cell was then subtracted from the corresponding sPlotOpen grid cell average (sPlotOpen – iNaturalist); the differences were visualized in boxplots.

Comparing growth forms

We extracted growth form information (tree/shrub/herb) from TRY (Trait-ID 42) to each iNaturalist observation and estimated the coverage of each growth form for all sPlotOpen plots. We chose the most commonly used classification for each species. The average tree/shrub/herb coverage was then calculated for each grid cell and these means correlated.

Reporting summary

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

Data availability

The trait maps in GeoTiff format for both the iNaturalist and sPlotOpen maps are openly available at <https://doi.org/10.5281/zenodo.6671891> (ref. ⁵²). All data used to create and analyse these maps are openly accessible (consult workflow for information on how to download the data).

Code availability

We provide a fully reproducible workflow (<https://doi.org/10.5281/zenodo.6671891>) of all analyses presented here and a script that can be used readily and without much effort to create updated global trait maps using the latest data, as citizen science data continue to grow⁵².

References

1. Sakschewski, B. et al. Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. *Glob. Change Biol.* **21**, 2711–2725 (2015).
2. Berzaghi, F. et al. Towards a new generation of trait-flexible vegetation models. *Trends Ecol. Evol.* **35**, 191–205 (2020).
3. Bruelheide, H. et al. Global trait–environment relationships of plant communities. *Nat. Ecol. Evol.* **2**, 1906–1917 (2018).
4. Joswig, J. S. et al. Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nat. Ecol. Evol.* **6**, 36–50 (2021).
5. van Bodegom, P. M., Douma, J. C. & Verheijen, L. M. A fully traits-based approach to modeling global vegetation distribution. *Proc. Natl Acad. Sci. USA* **111**, 13733–13738 (2014).
6. Moreno Martínez, A. et al. A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sens. Environ.* **218**, 69–88 (2018).
7. Pérez-Harguindeguy, N. et al. New handbook for standardized measurement of plant functional traits worldwide. *Aust. J. Bot.* **23**, 167–234 (2013).
8. Kattge, J. et al. TRY—a global database of plant traits. *Glob. Change Biol.* **17**, 2905–2935 (2011).
9. Kattge, J. et al. TRY plant trait database-enhanced coverage and open access. *Glob. Change Biol.* **26**, 119–188 (2020).
10. Jetz, W. et al. Monitoring plant functional diversity from space. *Nat. Plants* **2**, 16024 (2016).
11. Butler, E. E. et al. Mapping local and global variability in plant trait distributions. *Proc. Natl Acad. Sci. USA* **114**, E10937–E10946 (2017).
12. Boonman, C. C. et al. Assessing the reliability of predicted plant trait distributions at the global scale. *Glob. Ecol. Biogeogr.* **29**, 1034–1051 (2020).
13. Madani, N. et al. Future global productivity will be affected by plant trait response to climate. *Sci. Rep.* **8**, 2870 (2018).
14. Vallicrosa, H. et al. Global distribution and drivers of forest biome foliar nitrogen to phosphorus ratios (N:P). *Glob. Ecol. Biogeogr.* **31**, 861–871 (2022).
15. Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **12**, 1620–1633 (2021).
16. Schiller, C. et al. Deep learning and citizen science enable automated plant trait predictions from photographs. *Sci. Rep.* **11**, 16395 (2021).
17. Aguirre-Gutiérrez, J. et al. Pantropical modelling of canopy functional traits using sentinel-2 remote sensing data. *Remote Sens. Environ.* **252**, 112–122 (2021).
18. Homolova, L. et al. Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* **15**, 1–16 (2013).
19. Van Cleemput, E. et al. The functional characterization of grass-and-shrubland ecosystems using hyperspectral remote sensing: trends, accuracy and moderating variables. *Remote Sens. Environ.* **209**, 747–763 (2018).
20. Kattenborn, T., Fassnacht, F. E. & Schmidlein, S. Differentiating plant functional types using reflectance: which traits make the difference? *Remote Sens. Ecol. Conserv.* **5**, 5–19 (2019).
21. Hauser, L. T. et al. Explaining discrepancies between spectral and in-situ plant diversity in multispectral satellite earth observation. *Remote Sens. Environ.* **265**, 112684 (2021).
22. Wäldchen, J. & Mäder, P. Plant species identification using computer vision techniques: a systematic literature review. *Arch. Comput. Methods Eng.* **25**, 507–543 (2018).
23. Jones, H. G. What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora. *AoB Plants* **12**, plaa052 (2020).
24. Hampton, S. E. et al. Big data and the future of ecology. *Front. Ecol. Environ.* **11**, 156–162 (2013).
25. Wüest, R. O. et al. Macroecology in the age of big data—where to go from here? *J. Biogeogr.* **47**, 1–12 (2020).
26. Mäder, P. et al. The Flora Incognita app—interactive plant species identification. *Methods Ecol. Evol.* **12**, 1335–1342 (2021).
27. Di Cecco, G. J. et al. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. *BioScience* **71**, 1179–1188 (2021).
28. Mahecha, M. D. et al. Crowd-sourced plant occurrence data provide a reliable description of macroecological gradients. *Ecography* **44**, 1131–1142 (2021).
29. Botella, C. et al. Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data. *Methods Ecol. Evol.* **12**, 933–945 (2021).
30. iNaturalist Research-Grade Observations (GBIF, accessed 5 January 2022); <https://www.gbif.org/dataset/50c9509d-22c7-4a22-a47d-8c48425ef4a7>
31. Callaghan, C. T. et al. Three frontiers for the future of biodiversity research using citizen science data. *BioScience* **71**, 55–63 (2020).
32. Dickinson, J. L., Zuckerberg, B. & Bonter, D. N. Citizen science as an ecological research tool: challenges and benefits. *Ann. Rev. Ecol. Evol. Syst.* **41**, 149–172 (2010).
33. Kosmala, M. et al. Assessing data quality in citizen science. *Front. Ecol. Environ.* **14**, 551–560 (2016).
34. Boakes, E. H. et al. Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Sci. Rep.* **6**, 33051 (2016).
35. Bowler, D. E. et al. Temporal trends in the spatial bias of species occurrence records. *Ecography* **2022**, e06219 (2022). <https://doi.org/10.1111/ecog.06219>
36. GBIF Occurrence Download (GBIF, 4 January 2022); <https://doi.org/10.15468/dl.34tjre>
37. Bruelheide, H. et al. sPlot—a new tool for global vegetation analyses. *Journal of vegetation science. J. Veg. Sci.* **30**, 161–186 (2019).
38. Sabatini, F. et al. sPlotOpen—an environmentally balanced, open access, global dataset of vegetation plots. *Glob. Ecol. Biogeogr.* **30**, 1740–1764 (2021).
39. Whittaker, R.H. et al. *Communities and Ecosystems* (Macmillan/Collier Macmillan, 1970).
40. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51**, 933–938 (2001).
41. Joswig, J., Wirth, C. & Schuman, M. Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. *Nat. Ecol. Evol.* **6**, 36–50 (2022).
42. Diaz, S. et al. The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
43. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540 (2020).

44. Meyer, H. & Pebesma, E. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Methods Ecol. Evol.* **12**, 1620–1633 (2021).
45. Schrodt, F. et al. Bhpmpf—a hierarchical Bayesian approach to gap filling and trait prediction for macroecology and functional biogeography. *Glob. Ecol. Biogeogr.* **24**, 1510–1521 (2015).
46. Kuppler, J. et al. Global gradients in intraspecific variation in vegetative and floral traits are partially associated with climate and species richness. *Glob. Ecol. Biogeogr.* **29**, 992–1007 (2020).
47. Scheiter, S., Langan, L. & Higgins, S. I. Next-generation dynamic global vegetation models: learning from community ecology. *New Phytol.* **198**, 957–969 (2013).
48. Taubert, F. et al. Confronting an individual-based simulation model with empirical community patterns of grasslands. *PLoS ONE* **15**, e0236546 (2020).
49. Roger, E. & Klistorner, S. (2016) Bioblitzes help science communicators engage local communities in environmental research. *J. Sci. Commun.* <https://doi.org/10.22323/2.15030206> (2016).
50. Legendre, P. & Legendre, L. *Numerical Ecology* 3rd edn (Elsevier, 2012).
51. Warton, D. I. et al. Smatr 3—an R package for estimation and inference about allometric lines. *Methods Ecol Evol* **3**, 257–259 (2012).
52. Wolf, S. et al. iNaturalist_traits: iNaturalist trait maps version 1 (January 5, 2022) Zenodo <https://doi.org/10.5281/zenodo.6671891> (2022).

Acknowledgements

This study was funded by the National Research Data Infrastructure Germany for Biodiversity, NFDI4Biodiversity, a project by the German Research Foundation (DFG), project no. 442032008. The study is supported by the TRY initiative on plant traits (<http://www.try-db.org>) and the sPlot consortium (<http://www.idiv.de/splot>). The TRY initiative and database are hosted, developed and maintained by J.K. and G. Boenisch (Max Planck Institute for Biogeochemistry, Jena, Germany), currently supported by Future Earth/bioDISCOVERY and the German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig (iDiv, DFG-FZT 118, 202548816). The sPlot is a strategic project of iDiv and is supported by the German Research Foundation (DFG-FZT 118, 202548816). F.M.S. gratefully acknowledges the support of the Italian Ministry of University and Research, under the Maria Levi Montalcini programme. K.M. is funded by iDiv via a FLEPOOL project, funded by DFG (DFG-FZT 118, 202548816). We thank the vegetation scientists, who sampled vegetation plots in the field, digitized them or made them available in databases. We thank

the many iNaturalist citizen scientists, who volunteer their time and expertise to build the research-grade dataset. We thank our reviewers for their helpful comments and suggestions, which substantially strengthened the message and scientific relevance of our manuscript.

Author contributions

S.W. performed the analyses and wrote the manuscript. T.K. conceived and supervised the study. M.D.M. and C.W. supervised the project. F.M.S. and H.B. provided insight into the sPlotOpen data. J.K. contributed insight into the TRY data. Á.M.M. and K.M. contributed knowledge of opportunistic citizen science data. All authors contributed ideas and were involved in writing and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at
<https://doi.org/10.1038/s41559-022-01904-x>.

Supplementary information The online version contains supplementary material available at
<https://doi.org/10.1038/s41559-022-01904-x>.

Correspondence and requests for materials should be addressed to Sophie Wolf.

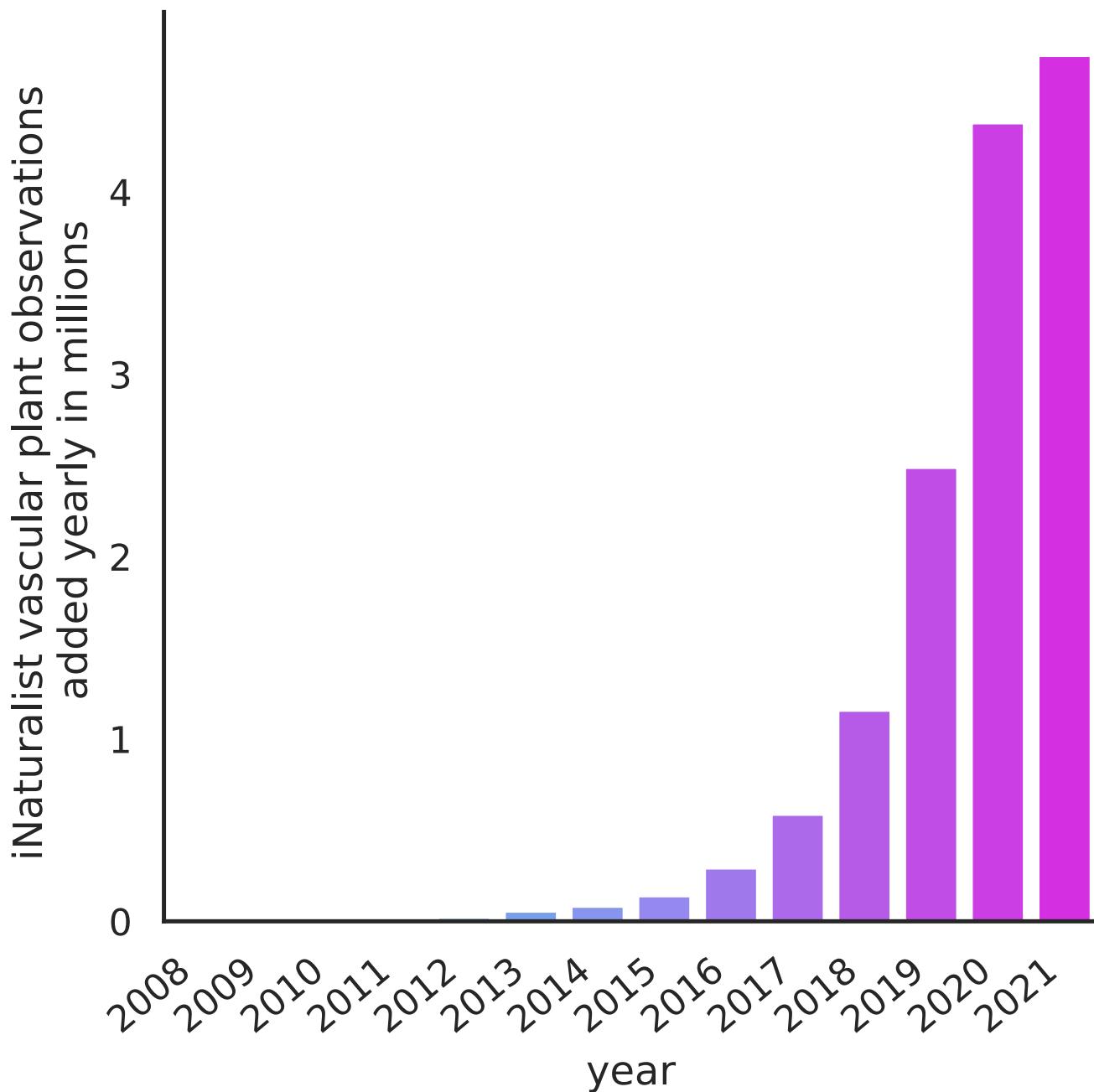
Peer review information *Nature Ecology & Evolution* thanks Angela Moles, Michael Belitz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

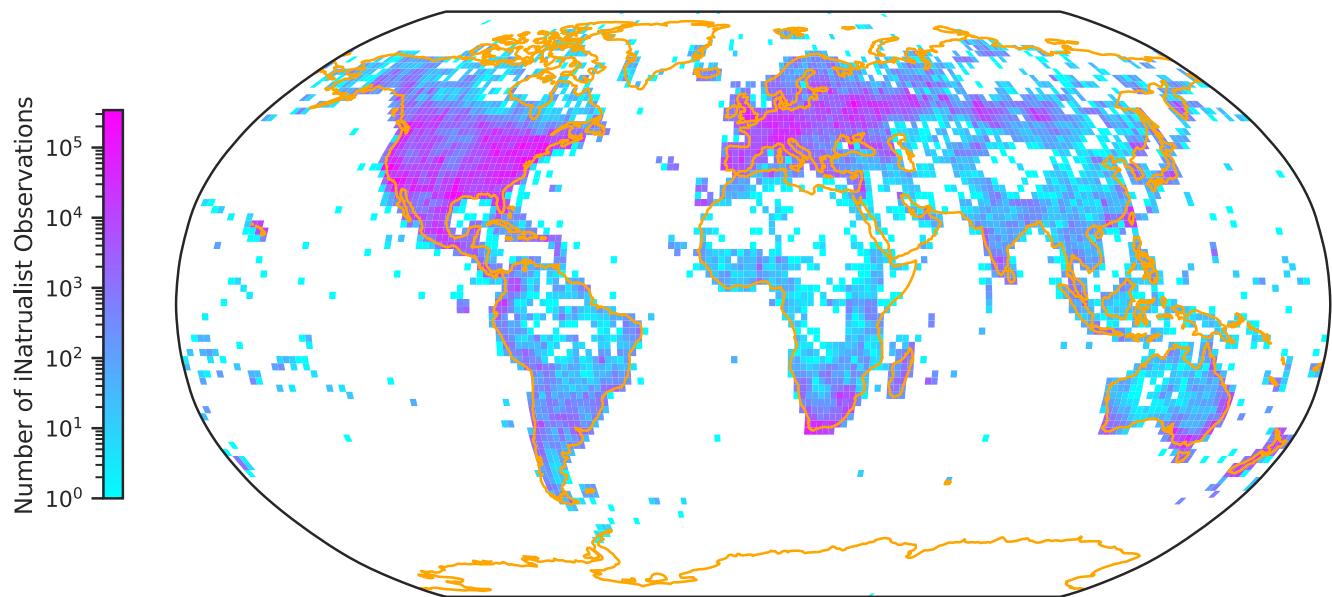
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

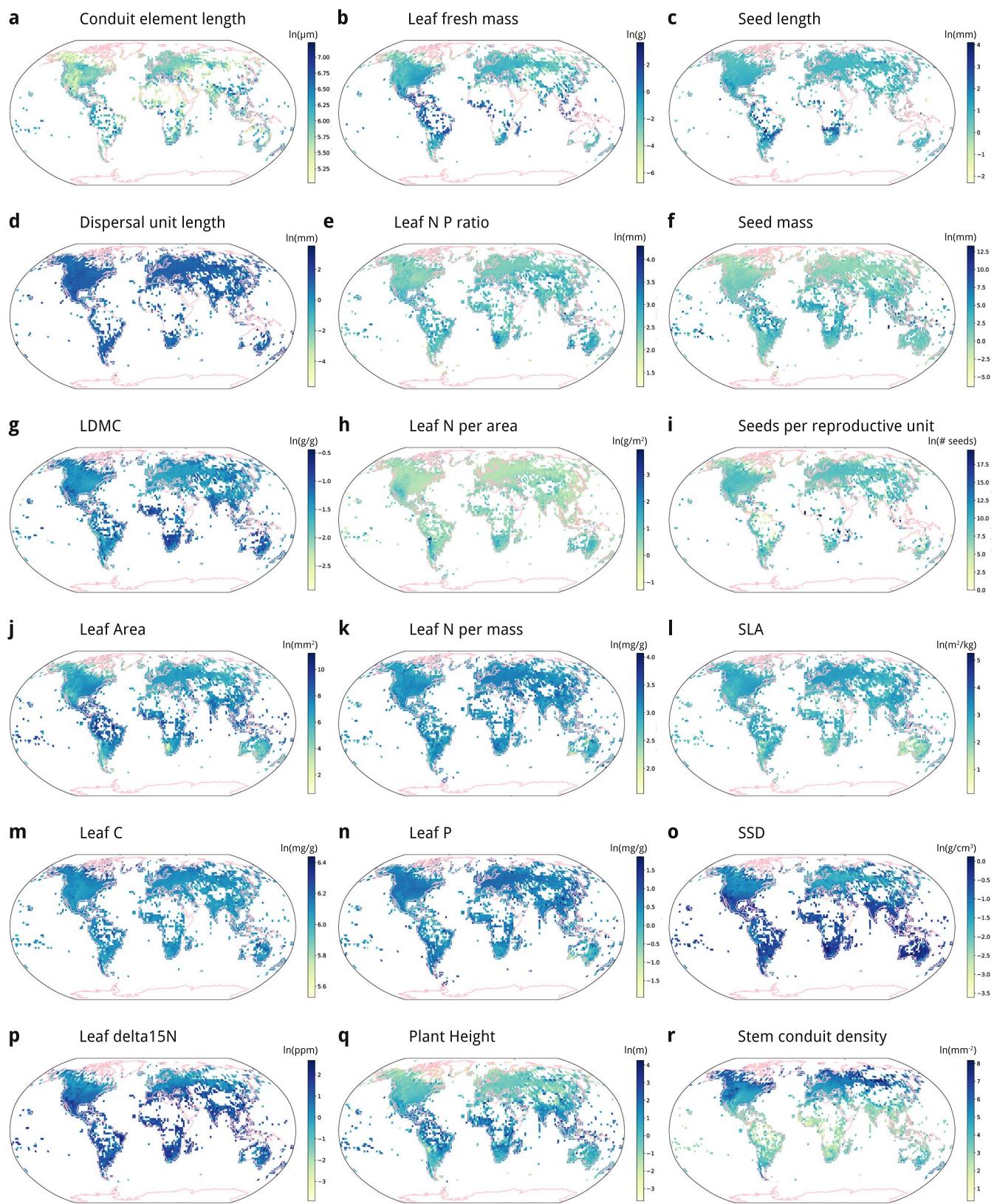
© The Author(s), under exclusive licence to Springer Nature Limited 2022



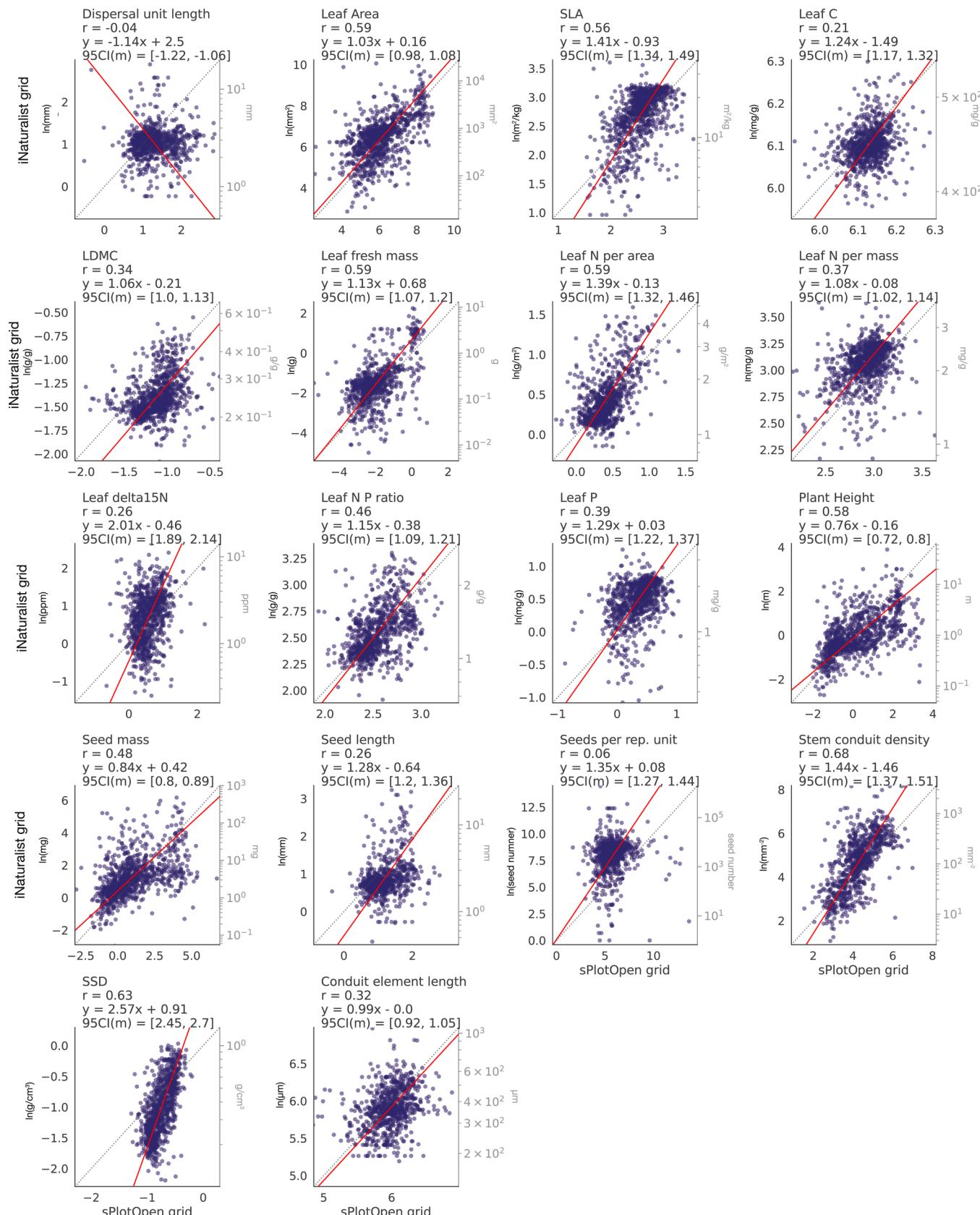
Extended Data Fig. 1 | iNaturalist data growth. Number of iNaturalist vascular plant observations added to “research-grade observations” every year since its foundation in 2008. More observations for 2021 are expected to come in, as the validation process takes time.



Extended Data Fig. 2 | Density of iNaturalist observations before linking to TRY. Density of iNaturalist vascular plant observations before linking to TRY database $n = 14,019,405$ observations; 2° resolution, or 221 km grid size. Colour corresponds to number of observations per cell.

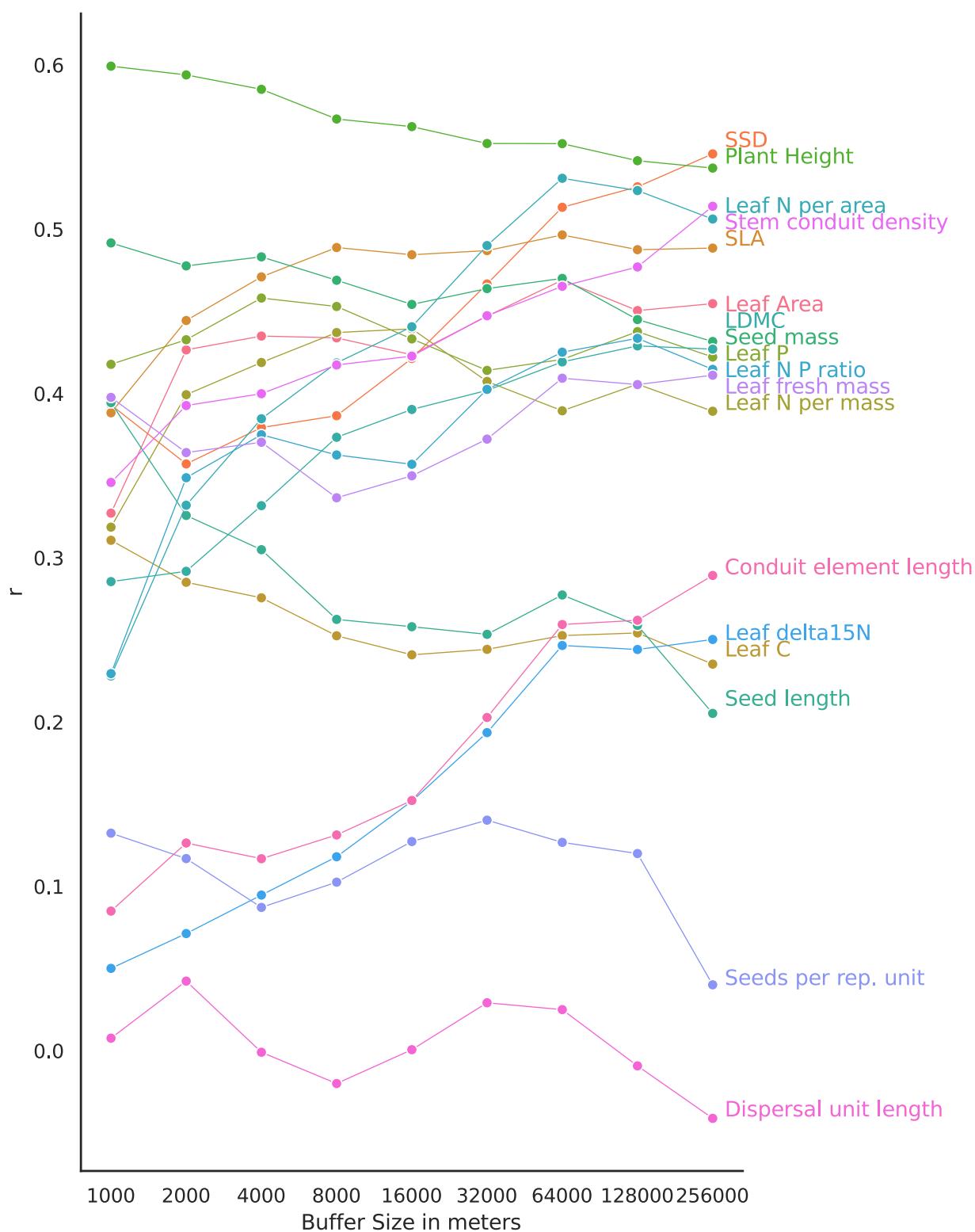


Extended Data Fig. 3 | iNaturalist trait maps. Global trait maps (trait values ln-transformed) using iNaturalist observations linked to TRY, displayed here at 2° resolution. Sample sizes for each trait see Table 1 in main text. For maps in GeoTiff format, refer to the Data availability statement.

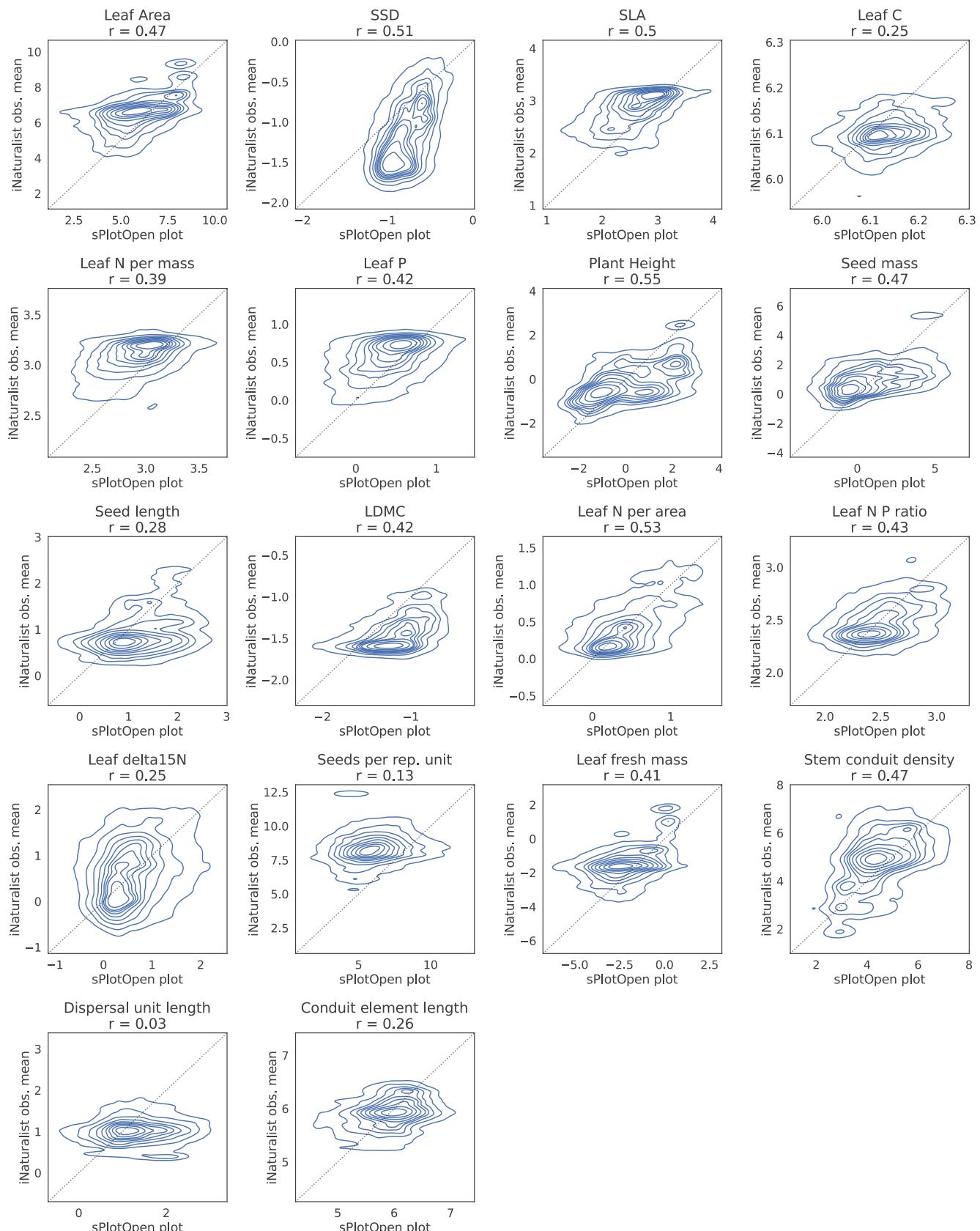


Extended Data Fig. 4 | Correlation scatter plots of iNaturalist and sPlotOpen trait maps. Scatter plots of sPlotOpen map pixel values plotted against the respective iNaturalist map pixel values for all 18 traits at a 2° spatial resolution. Correlation quantified using a weighted correlation coefficient (r weighted by grid cell area). Trait values are ln-transformed, 1:1 line is displayed in dotted

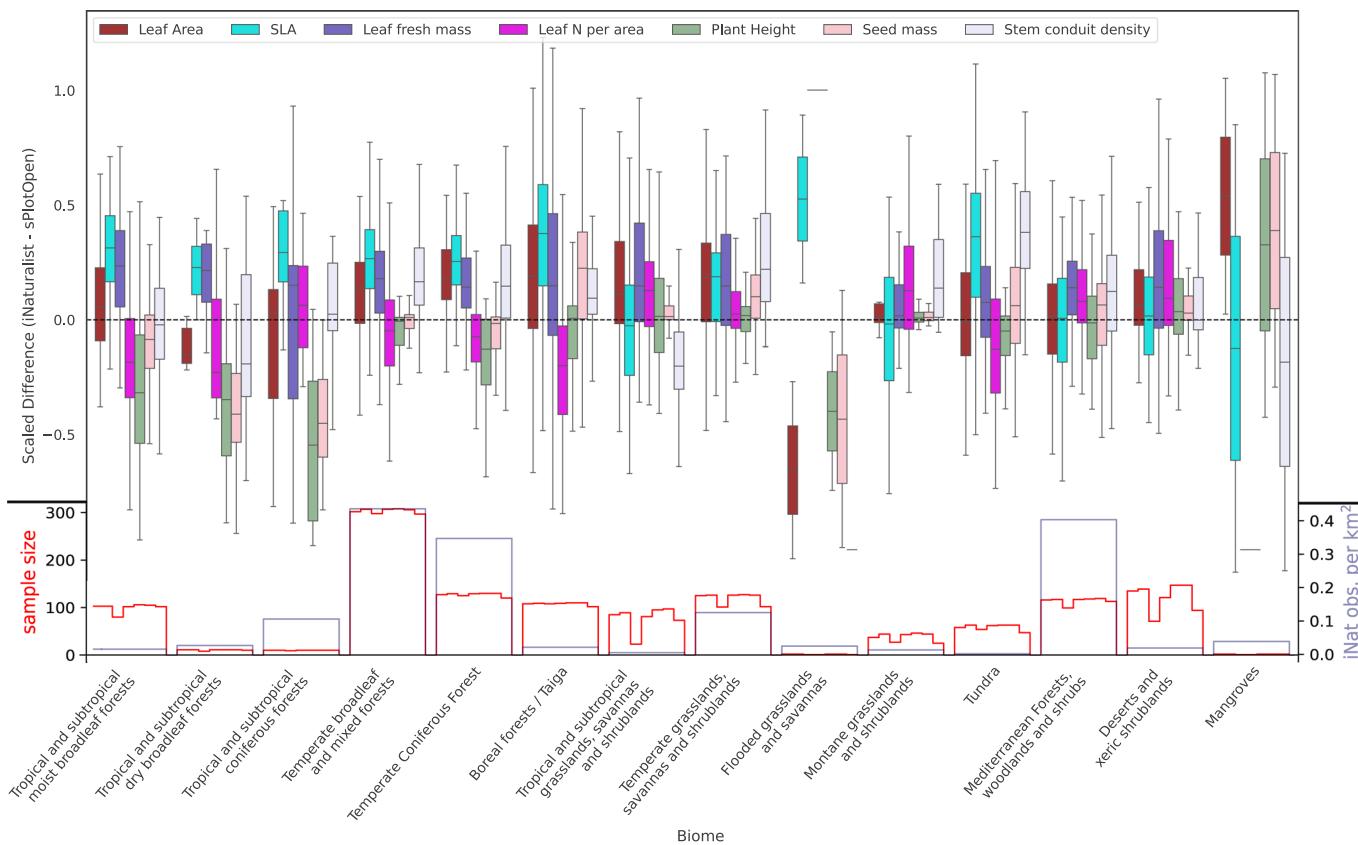
grey, and SMA regression slope in red. For clarity, the secondary y axis on the right shows the raw trait values marked on a log scale, which correspond to the ln-transformed values on the left. Plot extents are the 0.01 and 0.99 quantiles of the data.



Extended Data Fig. 5 | Relationship of r and buffer radius (buffer-based approach). Relationship of r and buffer radius (1 km to 256 km) for alternative approach: aggregation of iNaturalist observations in buffer radius around each vegetation plot. The lines connecting the points solely enhance readability.

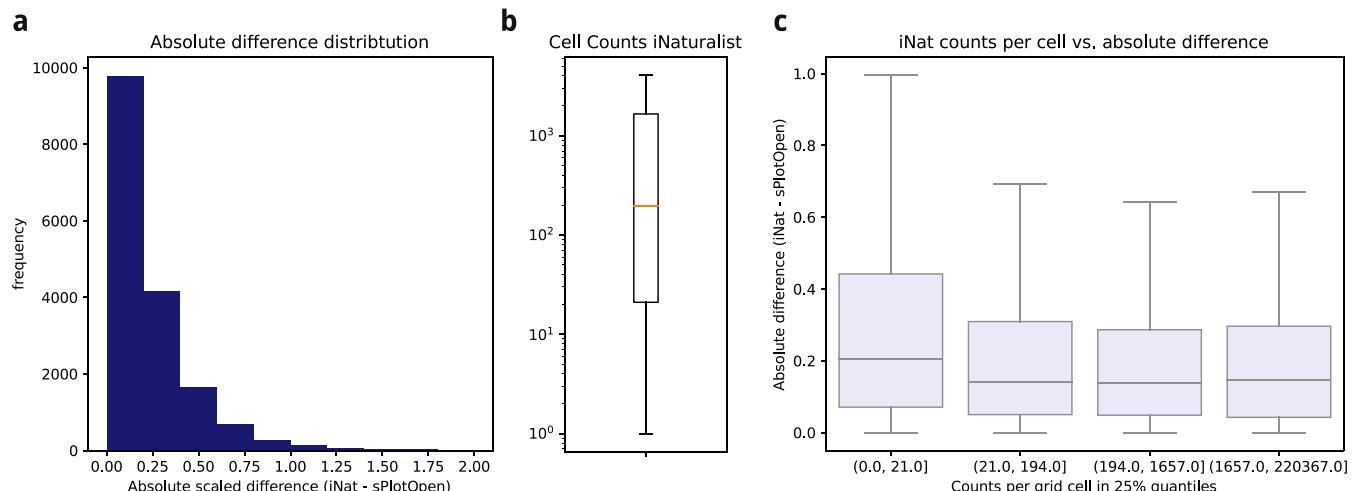


Extended Data Fig. 6 | Correlation density plots of iNaturalist and sPlotOpen using the buffer-based approach. Density plots (KDE plots) of the correlation of each community-weighted mean value in sPlotOpen plot with average trait value of all iNaturalist observations in its vicinity, using the alternative approach with buffers, here using a 64 km radius buffer. r is the Pearson correlation coefficient.


Extended Data Fig. 7 | Differences between biomes, forests not aggregated.

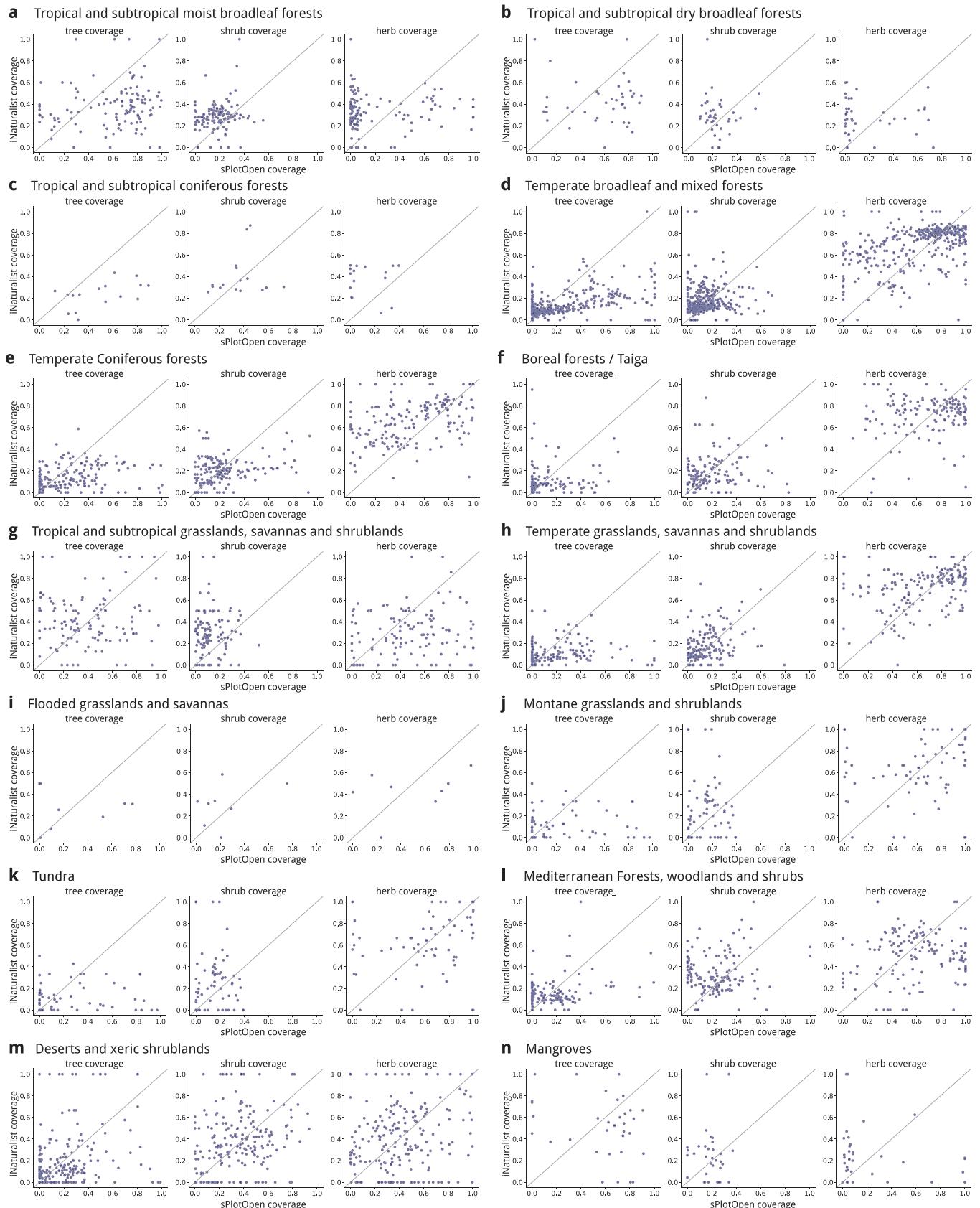
Difference between sPlotOpen and iNaturalist maps for each WWF terrestrial biome. All traits were scaled by range [−1, 1] using the 0.05 and 0.95 quantiles. The bounds of the box are defined by the first and third quartile, the centre lines are the medians, the whiskers mark the 1.5 interquartile range (IQR), outliers

are not shown. The red step-graph shows the sample size n = the number of iNaturalist map pixels that overlap the respective sPlotOpen map per biome and trait. The blue step-graph marks the mean density of iNaturalist vascular plant observations per km^2 in each biome. For exact sample sizes per biome and trait, see Supplementary Information Table S1.

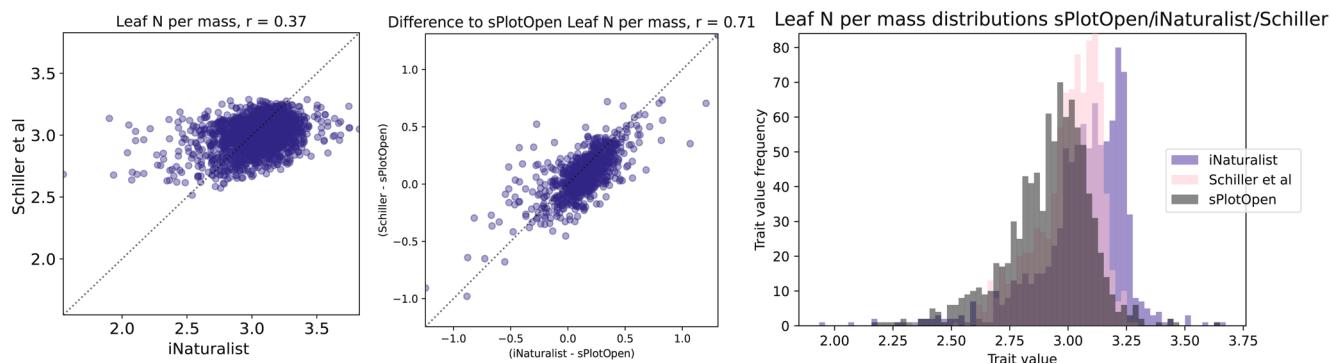
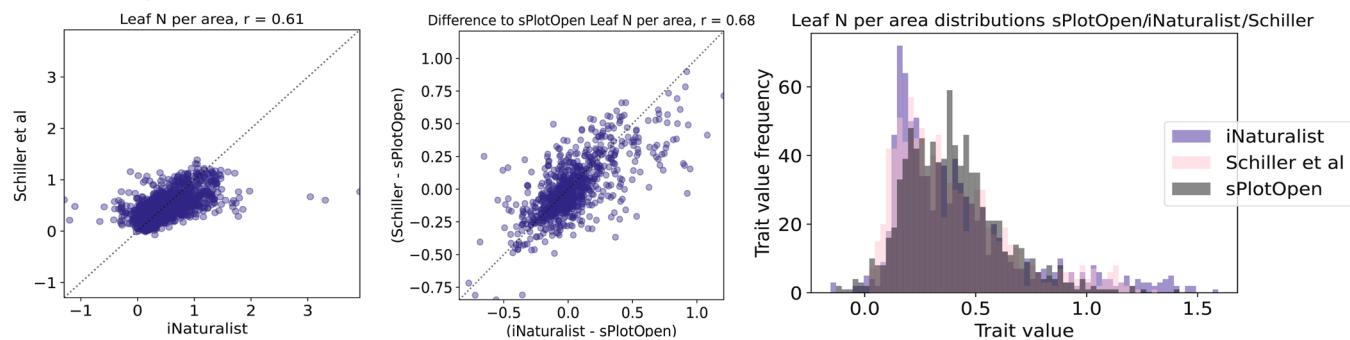
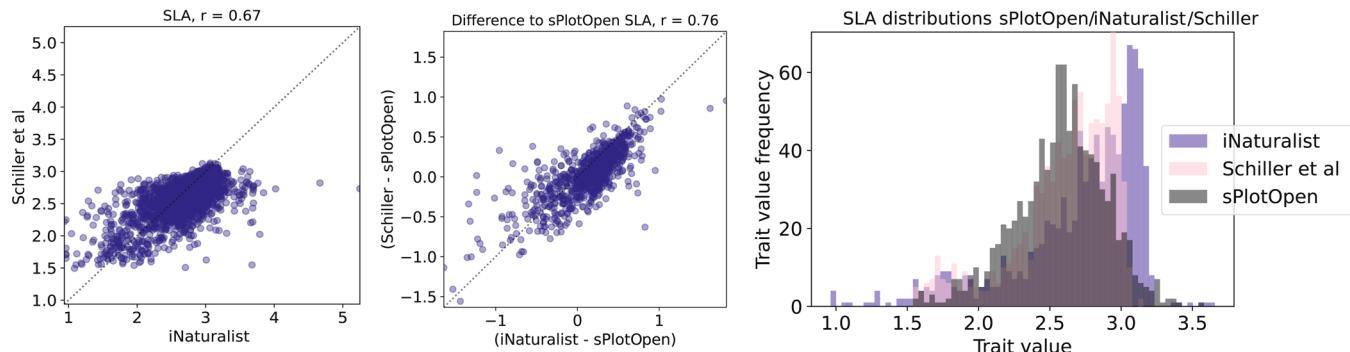


Extended Data Fig. 8 | Differences between iNaturalist and sPlotOpen maps in relationship to iNaturalist observation density. Differences between iNaturalist and sPlotOpen maps in relationship to the number of iNaturalist observations: Comparing number of observations per grid cell to the scaled absolute difference of iNaturalist and sPlotOpen means per respective grid cell at 2° resolution, over all traits with $n = 17024$, or 1037 pixel pairs over 18 traits.
a) Distribution of absolute scaled differences of iNaturalist and sPlotOpen map pixels (scaled using 0.01 and 0.99 quantiles); x-axis range cropped to (0,2).

b) Distribution of iNaturalist observation counts per grid cell, for grid cells that overlap with sPlotOpen map, $n = 17024$. The bounds of the box are defined by the first and third quartile, the centre lines are the medians, the whiskers mark the 1.5 interquartile range (IQR), outliers are not shown. **c)** Distribution of the absolute difference of iNaturalist - sPlotOpen pixels in each bin. The bins are based on 0.25 quantiles of the number of iNaturalist observations within each grid cell, all 18 trait maps combined: each bin has sample size of 4256. The bounds of the box, centre line, and whiskers are defined as in b).



Extended Data Fig. 9 | Growth forms coverage. Correlation of growth forms coverage (tree coverage, shrub coverage, and herb coverage) in iNaturalist and sPlotOpen grid maps in each WWF biomes.

a Leaf N per mass**b Leaf N per area****c SLA**

Extended Data Fig. 10 | Comparison of iNaturalist maps with the Schiller trait maps. Comparison of iNaturalist maps with the Schiller et al (2021) trait maps, which are based on estimating traits from iNaturalist photos. From left to right for each of the three traits **a**) Leaf N per mass [$\ln \text{mg/g}$], **b**) Leaf N per area [$\ln \text{g/m}^2$],

and **c**) SLA [$\ln \text{m}^2/\text{kg}$]: 1. Scatter plot of each corresponding pixel in the two maps, r weighted by grid cell area, black 1:1 line, 2. Scatter plot of the difference of each respective pixel to the sPlotOpen map, 3. Frequency distributions of trait values for iNaturalist maps, the Schiller maps, and the sPlotOpen maps.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
 - Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
 - Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection All data used in this study are openly accessible. Custom Python code in Jupyter notebooks was used to filter and combine the different datasets. The Jupyter notebooks are available at https://github.com/sojwolf/iNaturalist_traits.

Data analysis Custom Python and R code was used to analyze the data in this study. The Jupyter notebooks containing a fully reproducible workflow are available at https://github.com/sojwolf/iNaturalist_traits.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Global trait maps in geotiff format are available at <https://doi.org/10.5281/zenodo.6671891>, along with the fully reproducible workflow. These scripts may be used readily and without much effort to create updated global trait maps using the latest data, as citizen science and crowd sourced data continues to grow.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Our study explores whether plant occurrence data from the citizen science project iNaturalist, complemented with trait information from the global plant trait database TRY, encode global trait patterns. We matched the iNaturalist species observations with trait measurements from TRY, using the species name as the link. From this merged data we then created global trait distribution maps based on these crowd-sourced observations. For each trait, we generated a global spatial grid and calculated the mean of each grid cell for a set of 18 traits. The maps were evaluated using sPlotOpen vegetation plot community weighted means.

Research sample

In this study we use three existing datasets, all of which are openly accessible. The iNaturalist "research-grade" data are available the Global Biodiversity Information Facility (GBIF). All iNaturalist "research grade" observations of all vascular plant were downloaded on January 5, 2022 (<https://doi.org/10.15468/dl.34tjre>). The plant trait data for the 18 traits of interest (ID's 3113, 3117, 4, 13, 14, 15, 3106, 26, 27, 47, 50, 56, 78, 138, 163, 169, 237, 282) across all available species were requested and downloaded from www.try-db.org. The sPlotOpen data were downloaded from the iDiv data repository (<https://idata.idiv.de/ddm/Data>ShowData/3474?version=52>).

Sampling strategy

We used all available 14.6 million iNaturalist "research-grade" vascular plant observations. For the 18 traits of interest ((ID's 3113, 3117, 4, 13, 14, 15, 3106, 26, 27, 47, 50, 56, 78, 138, 163, 169, 237, 282) we used trait information for all available species from the TRY database. We used all sPlotOpen vegetation plots.

Data collection

We used existing open access data acquired and verified by citizen scientists as part of the iNaturalist project. TRY and sPlot are curated data collections of plant traits and vegetation plots, respectively, contributed by countless scientists and scientific projects.

Timing and spatial scale

iNaturalist data have been collected since 2008 at a growing rate and at a global scale.

Data exclusions

Not applicable.

Reproducibility

The fully reproducible workflow is available at https://github.com/sojwolf/iNaturalist_traits.

Randomization

This is a descriptive study using continuous data - we did not group data.

Blinding

We used existing data. The data acquisition was independent of our study. Blinding is not applicable here.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging