New Phytologist

Commentary

Macroecology of plant diversity across spatial scales

The distribution of vascular plants on Earth has puzzled scientists for at least three centuries. The description of the first patterns of vascular plants species richness – the number of species living in a region at a given time – dates to Alexander von Humboldt (e.g. the 'latitudinal gradient of diversity', von Humboldt, 1806). Since the pioneering work of von Humboldt, followed by early approaches for mapping species richness (e.g. Wulff, 1935), (macro)ecologists worldwide have attempted to refine mapping methods, to quantify plant diversity globally and to disentangle its underlying drivers. However, patterns of species richness are scale dependent, originating from the nonlinearity of the 'species–area relationship' (Arrhenius, 1921). In addition, the multiple determinants of species richness might interact in complex ways, being also habitat dependent and scale dependent (Keil & Chase, 2019). Cai *et al.* (2023), in this issue of *New Phytologist* (Cai *et al.*, 2023; pp. 1432–1445), by benefiting from up-to-date machine learning techniques, present a framework to bring a new quality to the global species and phylogenetic diversity predictions across a range of regional grain sizes.

> *'The high-resolution maps provided by Cai* et al. *can be regarded as a new benchmark for future global biodiversity assessments'.*

Diversity across scales

The 'species–area relationship' is regarded as a universal biological rule. That is, species richness increases nonlinearly with increasing sampling area (grain size). Thus, mapping biological diversity inherently depends on the spatial scale (Keil & Chase, 2019). For instance, Sabatini *et al.* (2022) found anomalies in plot-scale diversity (α-diversity) estimates at the global scale when comparing smaller vs larger plots (10 vs 400 m^2). Machine learning techniques provide a powerful means to model such complex relationships. These modeling techniques relax many assumptions of other statistical models (e.g. linearity), and can handle missing data, collinearity and detect complex nonlinear interactions among multiple predictors without any *a priori* assumption. Cai *et al.* gathered information on regional species and phylogenetic richness across unequally sized geographic regions using the Global Inventory of Floras and Traits database (Weigelt *et al.*, 2020) – a careful collection of regional plant checklists and floras. The apparent disadvantage in compiling a dataset including unequally sized regions turned, in this case, into an advantage. The authors, using machine learning, were able to explicitly model the complex species–area relationship in their dataset. Remarkably, they were able to explain an unprecedent > 80% of variation in both species and phylogenetic regional diversity across a range of grain sizes (7774–209 903 km^2).

The effort by Cai *et al.* to model global patterns of regional diversity across grain sizes meets the current needs and trends of macroecology. Just a few months earlier, Sabatini *et al.* (2022) modeled plant plot-level diversity across grain sizes using a machine learning method. Although both the models by Cai *et al.* and Sabatini *et al.* (2022) predict diversity across specific ranges of grain size, these results cannot be extrapolated to all possible spatial scales. Specifically, there remains an important gap of knowledge on global patterns of species richness at the landscape scale (Fig. 1). Bridging this gap is the next goal for building a solid predictive framework of how biodiversity facets will respond to the challenges imposed by the major drivers of biodiversity loss at all possible spatial scales. For instance, human impact decreases and isolates natural vegetation patches. Therefore, many species might persist in some habitat fragments in a region, but they might be absent in most suitable sites, thereby falling into dark diversity (*sensu* Pärtel *et al.*, 2011). Understanding at which scale such processes are mostly occurring remains a priority for macroecological research.

Drivers of global diversity patterns

Competing hypotheses have been put forth for trying to explain the geography of species richness. The 'water–energy dynamics hypothesis' posits current climate as the major driver of species richness (Hawkins *et al.*, 2003; Kreft & Jetz, 2007). Climate controls the relative flows of water and energy at a given site, and these set the conditions for species coexistence, possibly through constraints on productivity and species-specific adaptations (i.e. environmental filtering). The 'heterogeneity–diversity hypothesis' (von Humboldt & Bonpland, 1807) states that heterogeneous habitats favor species turnover by increasing the available niches. Geographic and historical hypotheses suggest plate tectonics-related events and long-term climate stability as the drivers of species diversification, and thus of species richness (Antonelli, 2017). Cai *et al.* provide the most comprehensive test to date of how this set of hypotheses concurrently explains global patterns of species and phylogenetic regional diversity. They identified water–energy dynamics as the main driver of species and phylogenetic richness, followed by habitat heterogeneity, geographic factors and, ultimately, by past climate legacy.

This article is a Commentary on Cai *et al.* (2023; pp, 1432–1445).

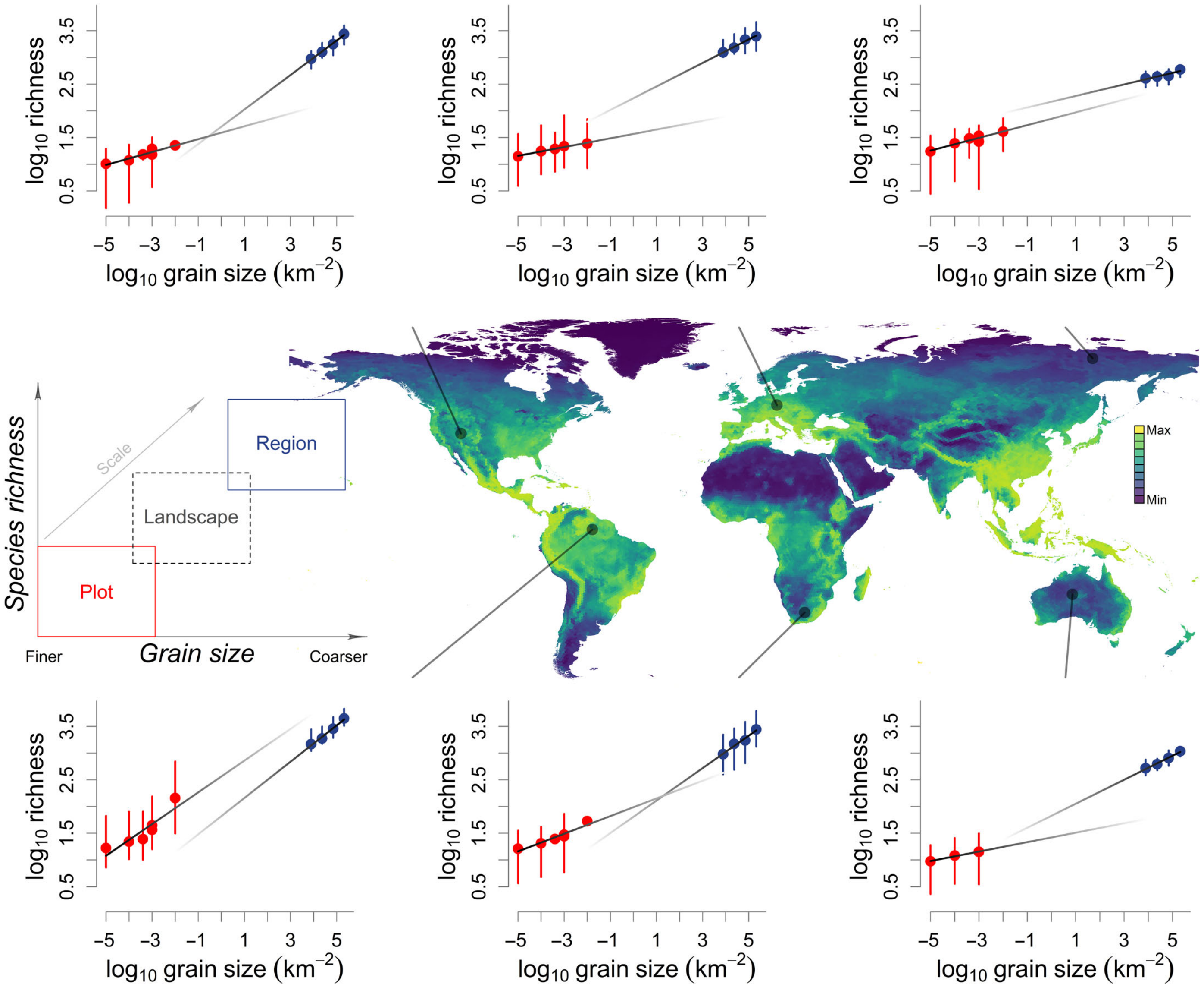

Fig. 1 Vascular plant species richness across spatial scales globally. Major advancements in predicting global patterns of species richness in relation to grain size (left-end panel). Plot-level and regional diversity refer to the advancements made by Sabatini *et al*. (2022), and by Cai *et al*. (2023; pp. 1432–1445) in this issue of *New Phytologist*, respectively. Plot-level and regional diversity are represented as boxes in the left-end panel because the modeling procedures used to produce the maps work across a specific range of grain sizes, as shown for selected regions across the globe (panels above and below the graph connected to selected locations on the central global map obtained from Cai *et al*. and available at: https://gift.uni-goettingen.de/shiny/predictions/ at 7.774 km^2 resolution). Red and blue dots are richness estimates (median with the range within 500 km radius) from Sabatini *et al*. (2022) and from Cai *et al*., respectively. The dashed box in the left-end panel represents the highlighted gap of knowledge in predicting biodiversity at the landscape scales. Extrapolations much beyond the modeled range are risky but might still hint how the diversity in the so far unexplored gap at the landscape scale might look. In some cases, the extrapolated lines cross each other (different slopes), possibly reflecting that, for both plot-level and regional diversity, different habitat types in a region have relatively comparable diversity but different species composition. However, sometimes extrapolations are mostly parallel. Here, the additional landscape-scale data would allow to understand if the reason is a modeling artifact, or if there are some ecological reasons for these differences.

However, habitat heterogeneity was less relevant in explaining phylogenetic diversity in comparison with species richness, and Cai *et al.* suggest that frequent recent speciation events in heterogeneous regions might have increased the number of related species, decoupling the patterns of species and phylogenetic richness in some cases. Interestingly, Cai *et al.* and Sabatini *et al.* (2022) found current climate as the main determinant of regional and plot-level diversity patterns globally, but heterogeneity had a major effect just at the regional scale. Another necessary endeavor of the macroecology of biodiversity is to identify various drivers of species diversity across scales.

The compelling analysis by Cai *et al.* uncovers highly complex interactions among governing diversity drivers along different environmental gradients. By accounting for such interactions in their modeling framework, Cai *et al.* reveal finely tuned global regional diversity patterns progressing considerably from the previous works (e.g. Barthlott *et al.*, 2005; Kreft & Jetz, 2007). The knowledge of such complex interactions among drivers of species and phylogenetic richness strengthens our understanding of how patterns of diversity facets respond to regional conditions, as previously suggested (Kreft & Jetz, 2007; Keil & Chase, 2019). These findings at the regional scale complement the recently explored effects of climatic and historical predictors on plot-level diversity globally (Sabatini *et al*., 2022). The results from Cai *et al.* expose the need for more research in this direction, carried out at various scales, if we want to finally fill the gap in between predictions of local and regional diversity globally (Fig. 1). Various landscape characteristics are likely necessary in addition to previously used diversity drivers.

Despite the remarkable conceptual and methodological framework provided by Cai *et al*., a further step toward predicting patterns of species richness might require inclusion of biologically relevant information in the modeling pipeline. There is evidence that diversity patterns, and their drivers, can differ among species groups with contrasting adaptations to the environment. For instance, Mutke & Barthlott (2005), using a relatively small dataset of 3000 species, and at coarse resolution, observed that global diversity hotspots mostly reflected those of angiosperms, and when focusing on gymnosperms, they found different diversity hotspots for this group (probably associated with different climatic/biogeographic drivers). Similarly, Sabatini *et al.* (2022) revealed some differences in α-diversity patterns for woody and herbaceous vegetation. This evidence suggests that adding relevant, though coarse, biological information to the state-of-the-art modeling pipeline proposed by Cai *et al.* can further improve modeling predictions and shed light on additional diversity hotspots globally. In addition, there is a need for global models of functional diversity, but current availability of data is limiting. Nevertheless, continuously developing public trait databases (e.g. TRY, Kattge *et al*., 2020) and imputation methods might fill some data gaps.

More data needed from naturally species-poor regions

The analysis by Cai *et al.* once again points the attention to the inherent limitations of any global-scale study: the lack of data covering some relevant portions of the globe. Extreme habitats such as deserts and arctic regions, together with entire regions across the Global South, remain under-sampled. Cai *et al.* found inconsistencies among different modeling techniques in predicting species and phylogenetic richness in these naturally species-poor regions. Thus, increasing data availability from naturally species-poor regions with unique climatic, historical, and ecological features remains a priority to further train machine learning models and improve their ability to predict global patterns of diversity facets.

Contribution to future biodiversity research and conservation

Cai *et al.* provide a solid modeling pipeline accompanied by freely available codes. Current machine learning techniques, as Cai *et al.* nicely show, can prove to be essential for disentangling complex diversity–driver relationships that escape our, sometimes simplistic, understanding of how vascular plants, and biodiversity in general, distribute on the Earth. The contribution by Cai *et al.* could serve as a useful template for various future macroecological research questions. The high-resolution maps provided by Cai *et al.* (available at: https://gift.uni-goettingen.de/shiny/predictions/) can be regarded as a new benchmark for future global biodiversity assessments, contributing for the development of targeted environmental policies. For example, combining quantitative information on plant diversity with IUCN conservation status can help identifying high-priority conservation areas. Future biodiversity research will need to expand the predictive ability of this and other approaches (e.g. Sabatini *et al*., 2022) to bridge the gap at the landscape scales (Fig. 1).

Acknowledgements

We thank Lauri Laanisto for valuable comments. GP was supported by the following grants: IJC2020-043331-I funded by MCIN/AEI/10.13039/501100011033, and PID2021-122214NA-I00 funded by MCIN/AEI/10.13039/501100011033 and by FEDER 'ESF Investing in your future'. MP was supported by the Estonian Research Council (PRG609) and the European Regional Development Fund (Centre of Excellence EcolChange).

ORCID

Meelis Pärtel https://orcid.org/0000-0002-5874-0138
Giacomo Puglielli https://orcid.org/0000-0003-0085-4535

Giacomo Puglielli[1]* **and Meelis Pärtel**[2]

[1]Departamento de Biología Vegetal y Ecología, Facultad de Biología, Universidad de Sevilla, 41080 Sevilla, Spain;
[2]Institute of Ecology and Earth Sciences, University of Tartu, J. Liivi 2, 50409 Tartu, Estonia
(*Author for correspondence: email gpuglielli@us.es)

References

Antonelli A. 2017. Biogeography: drivers of bioregionalization. *Nature Ecology & Evolution* 1: 1–2.

Arrhenius O. 1921. Species and area. *Journal of Ecology* 9: 95–99.

Barthlott W, Mutke J, Rafiqpoor MD, Kreft H. 2005. Global centers of vascular plant diversity. *Nova Acta Leopoldina* 92: 61–83.

Cai L, Kreft H, Taylor A, Denelle P, Schrader J, Essl F, van Kleunen M, Pergl J, Pyšek P, Stein A *et al.* 2023. Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytologist* 237: 1432–1445.

Hawkins BA, Field R, Cornell HV, Currie DJ, Guégan J-F, Kaufman DM, Kerr JT, Mittelbach GG, Oberdorff T, O'Brien EM *et al.* 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology* 84: 3105–3117.

von Humboldt A. 1806. *Ideen zu einer Physiognomik der Gewächse.* Tübingen, Germany: Cotta [in German].

von Humboldt A, Bonpland A. 1807. *Essay on the geography of plants.* Chicago, IL, USA: Reprint University of Chicago Press.

Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M *et al.* 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* 26: 119–188.

Keil P, Chase JM. 2019. Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution* 3: 390–399.

Kreft H, Jetz W. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences, USA* **104**: 5925–5930.

Mutke J, Barthlott W. 2005. Patterns of vascular plant diversity at continental to global scales. *Biologiske Skrifter* **55**: 521–531.

Pärtel M, Szava-Kovats R, Zobel M. 2011. Dark diversity: shedding light on absent species. *Trends in Ecology & Evolution* **26**: 124–128.

Sabatini FM, Jiménez-Alfaro B, Jandt U, Chytrý M, Field R, Kessler M, Lenoir J, Schrodt F, Wiser SK, Arfin Khan MAS *et al.* 2022. Global patterns of vascular plant alpha diversity. *Nature Communications* **13**: 4683.

Weigelt P, König C, Kreft H. 2020. GIFT – a global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography* **47**: 16–43.

Wulff EW. 1935. Versuch einer Einteilung der Vegetation der Erde in pflanzengeographische Gebiete auf Grund der Artenzahl. *Repertorium Specierum Novarum Regni Vegetabilis* **12**: 57–83.

Key words: biodiversity, machine learning, macroecology, spatial scale, species richness, vascular plants.