bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Disentangling the roles of inter and intraspecific variation on leaf trait distributions across the eastern United States

Sergio Marconi[1,2], Benjamin G. Weinstein[2], Jeremy W. Lichstein[3], Stephanie A. Bohlman[1], Aditya Singh[4], Ethan P. White[2]

[1]School of Forest, Fisheries, & Geomatics Sciences, University of Florida, Gainesville, Florida, USA
[2]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida, USA
[3]Department of Biology, University of Florida, Gainesville, Florida, USA
[4]Department of Agricultural & Biological Engineering, University of Florida, Gainesville, FL 32611, USA

*Sergio Marconi

Email: sergio.marconi@weecology.org

Author Contributions: Sergio Marconi, Ben Weinstein and Ethan White designed the experiment and developed the methods; Sergio Marconi performed the analysis; Ethan White, Jeremy Lichstein, Aditya Singh and Stephanie Bohlman supervised the work, helped with experimental design and advice on data analysis and technical aspects. All authors contributed to editing the manuscript.

Competing Interest Statement: No conflict of interests.

Keywords: leaf traits, intra-species variability, multilevel joint-distribution, cross-scale.

Abstract

Functional traits are central to how organisms perform and influence ecosystem function. Although phylogenetic constraints and environmental conditions are both known to affect trait distributions, data limitations have resulted in large scale studies modeling traits either as species weighted averages (ignoring intraspecific variation) or as a function of the environment (ignoring phylogenetic constraints). As a result, large scale predictions for trait distributions do not include key drivers, likely resulting in biased predictions, and cannot be used to assess the relative contributions of inter- and intraspecific variation.

To address these limitations, we developed a joint model integrating phylogenetic and environmental information to understand and predict the distribution of eight leaf traits across the eastern United States. This joint model explained 68% of trait variation, outperforming both species-only and environment-only models, with variance attributable to phylogeny alone (23%), the environment alone (18%), and their overlapping effects (26%). The importance of phylogenetic constraints and the environment varied by trait, with some traits associated predominantly with environmental variation and others with phylogeny.

To make predictions more continuously across the eastern USA we combined this model with data from the large-scale Forest Inventory and Analysis survey to estimate traits for ~1.2 million trees. The combined model exhibited significant deviations in predictions from both species-only and environment-only models with variation in the direction and magnitude of these differences among ecoregions. These predictions demonstrate the importance of modeling both intra- and interspecific variation to understand and predict large scale gradients in species and ecosystem traits.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Significance Statement

Large scale trait studies typically rely on species level average trait values, but recent research suggests that intra-species variation plays a key role in the ecology of species across their range of distribution. We developed a new continental scale approach that integrates inter- and intra-specific variation on leaf traits by leveraging the interaction between the environment and the phylogenetic relationships across species. Using this approach to make predictions for over a million trees covering the Eastern United States shows that combining intra- and inter-species variation is crucial for capturing patterns at the continental scale. This approach represents a novel state of the art method with the potential to become the new standard for understanding traits at large scales.

Introduction

Global change is expected to cause extensive changes in terrestrial ecosystems, driving unprecedented redistribution of plant species and their associated traits (Pecl et al., 2017, Diaz, S., and M. Cabido, 2001). Plant functional traits are involved in key ecosystem processes from local community assembly (McGill et al., 2006, Sterck et al., 2011) to global land-surface biogeochemical cycles, and these processes are interconnected across scales (Reichstein et al., 2014, Peaucelle et al., 2019). Biotic interactions, micro-climate, and soil conditions can affect species co-occurrence and influence local trait distributions (Bruelheide et al., 2018, Simpson et al., 2016), and variation in climate within species ranges can affect realized niches and drive plastic trait responses (Chave, 2013). Given their central role across multiple levels of organization, understanding and predicting how traits vary within and among species across different scales and environments is essential for monitoring and conserving present and future ecosystem function (Violle et al., 2014).

Plant traits vary geographically through a combination of interspecific shifts in species abundances and intraspecific trait variation (Valladares et al., 2014, Münzbergová et al., 2017). Fully understanding how traits respond to the environment across wide geographic areas requires approaches that integrate both of these sources of trait variation. However, this is challenging because individual level trait data are geographically and taxonomically limited, making it hard for traditional methods to identify the relative importance of intra- and interspecific trait variation at broad spatial scales (Henn et al., 2018). In addition, to effectively predict broad-scale geographic patterns requires the ability to make predictions for species not included in trait datasets but widely distributed across the continent. Nonetheless, most current approaches to understanding and predicting large scale trait variation circumvent these limitations by focusing on either environmental predictors (ignoring species effect on traits) or species-level trait averages (ignoring the environment).

Environmental models develop empirical relationships between trait values and environmental variables, either ignoring species identity (e.g., Ordonez et al., 2009) or treating it as a random effect (e.g., Moles et al. 2014). This approach does not directly model known phylogenetic signals in traits variation driven by biological, physical, and historical constraints (Wright et al., 2004, Anderegg et al., 2018), instead assuming that the environmental model implicitly captures relevant changes in species distribution and abundance. Moreover, because the information about species identity is missing from predictions, these models cannot be used to explicitly identify the amount of traits variation uniquely ascribable to inter- intra species (but see Moles et al. 2014) and potentially yield biased predictions for assemblage level trait values unless the environmental model fully captures relevant shifts in species abundance.

Due to these limitations, it has been suggested that species-level models are more informative than modeling the relationship between environment and traits directly (Clark, 2016; Yang et al., 2018, Swenson et al., 2010, Zhang et al., 2018). Species-level models assume that environmental drivers affect trait distributions indirectly by shaping community structure and species abundance and therefore that species distributions are the best predictor of traits and associated ecosystem function. These approaches estimate large scale trait distributions using community weighted means based on species-level average trait values (Swenson 2010, Clark 2016, Wieczynski et al., 2019, Swenson 2017, Stahl et al., 2014) or broad plant functional types

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Butler et al., 2017). These models can be used to make predictions for traits over large areas by leveraging large scale forest vegetation surveys (e.g., forest inventories). However, this approach ignores intraspecific variation, which can be larger than interspecific variation for broadly distributed species (Niinemets, 2015, Messier et al., 2017).

The above approaches have made important contributions to our understanding of trait variation, but neither approach explicitly incorporates both key drivers. This limitation potentially results in biased predictions and prevents assessment of the relative importance of inter- and intraspecific variation on trait distributions across a continuum of geographical scales. To address this limitation, we developed a model that combined species phylogenetic information (from the Tree of Life; Hinchliff et al., 2015) and environmental drivers (climate, elevation, slope and terrain aspect) with large scale leaf trait data from the National Ecological Observatory Network (NEON) (National Ecological Observatory Network, 2020). We jointly modeled eight leaf traits: nitrogen (N%), carbon (C%), chlorophyll A (ChlA%), chlorophyll B (ChlB%), carotenoids (Crt%), leaf mass per area (LMA, g m^{-2}), lignin (%) and cellulose (%). We compared this combined model to models based on only environmental drivers or only species information. We integrated the combined model with USA Forest Inventory and Analysis (FIA)(USDA Forest Service, 2001, Smith et al., 2002) and Daymet data (Thornton et al., 2018) to make predictions for ~1.2 million trees across the eastern USA, compared these predictions to species-only and environment-only approaches to assess the influence of model differences on large scale prediction, and analyzed the relative contribution of environmental factors and phylogeny to quantifying leaf trait variation across ecoregions.

Results and Discussion

Model evaluation. We built the species, environment, and combined models using leaf trait data from NEON and evaluated their explanatory power using the Bayesian R^2 of the predicted values for 88 out-of-sample test trees. For all 8 leaf traits, the combined species and environment model explained the largest amount of variance in the held-out test data (average R^2 across the 8 traits = 0.64), substantially outperforming both the environment-only (average R^2 = 0.35) and species-only (average R^2 = 0.52) models (Figure 1, Figure S.2). The combined model had the highest performance for predicting LMA (R^2 = 0.84) and the lowest performance for predicting ChlA (R^2 = 0.53). Uncertainty in predictions was accurately estimated across all traits, with mean 95% coverage values ranging from 94.3% to 98.8% (Figure S.3). The importance of different environmental drivers also varied among traits, supporting an important role of climate in driving leaf economics in local communities (Ordoñez et al. 2009). Precipitation and temperature were mainly important for traits involved in photosynthesis (N%, Chlorophyll A, B and Carotenoids), generally having a positive effect on their concentration (with the exception of Chlorophyll B, Figure S.4). Net radiation showed a negative effect on N%, while vapor pressure generally had a negative effect on pigments but positive on traits associated with leaf toughness and durability (cellulose, lignin and C%). Elevation was the most important topographic predictor and the strongest environmental driver of LMA patterns (Figure S.4), consistent with previous studies (Reich & Oleksyn, 2004, Hedin 2004, Poorter et al., 2009, Kitajima et al.,2016). The joint model structure also captured the strong correlations among LMA and N% widely described by the leaf economics spectrum (Reich et al, 1997, Wright et al., 2004). A hierarchical clustering of traits based on model residuals supported two major trait classes (Figure S.5): traits mainly involved in photosynthesis (Croft et al., 2017) and traits involved in leaf structure.

Role of species and environmental effects in predicting traits. We used variance partitioning on the out-of-sample test data to explore the relative contributions of inter- and intraspecific trait variation. Variance uniquely explained by species effects is by definition independent of environmental effects and thus represents pure interspecific variation. Similarly, variance uniquely explained by environmental drivers is independent of species effects (i.e., shifts in species composition) and thus represents intraspecific trait variation. Variance explained by the combined species-environment model but not uniquely ascribable to either species or environmental effects may include variation that could be explained by either species or environmental effects alone, as well as variation that can only be explained by a combined approach. Variance explained by non-combined approaches (i.e., species-only or environment only models) is difficult to interpret,

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

because these approaches ignore the joint variance component, do not allow for variance partitioning and may therefore overestimate the pure species and pure environmental variance components. On average, interspecific variation (pure species effect) accounted for 29% of the total explained trait variation across the 8 traits, intraspecific variation (pure environment effect) accounted for 12%, and joint species-environment effects accounted for 23%.
The relative importance of inter- and intraspecific effects varied widely among traits. Species effects explained most of the variation in N% and LMA. For these traits, often used in large scale species and trait distribution studies, species distributions may contain more information about traits than direct predictions from the environment, as previously suggested by species- average models. In contrast, the intraspecific component of the model accounted for as much or more of the variance than the interspecific one for pigments and also uniquely explained a non-negligible amount of variance for C% and lignin. Note that our analysis only quantifies intraspecific variation in upper canopy (sun-lit) leaves across space, and thus does not include intraspecific variation from sun to shade across canopy light gradients (Osnas et al., 2018, Niinemets et al., 2015). Furthermore, 62% of species in our analysis were only sampled within a single NEON site, thus representing only a small fraction of the true environmental range. Thus, our analysis likely underestimates the true level of intraspecific trait variation.

Model transferability. To determine the appropriateness of using the model to make predictions outside of NEON sites, we tested the performance of the combined model on novel locations and novel species using independent data from the Botanical Information and Ecology Network (BIEN) (Enquist et al., 2009) and a subset of TRY (Kattge et al., 2020) leaf trait datasets. These datasets include trait data on LMA, N% and C% for 62 species, including 27 species not present in the training data (Figure S.1, supplement S1). The combined model showed good transferability to other data sources (mean R^2 = 0.54, 95% coverage = 91%, Figure S.6). The inclusion of full phylogenetic relationships in addition to environmental predictors yielded successful model transfer to species not sampled at NEON sites (mean R^2 = 0.4, 95% coverage = 94%, Figure 2). This was possible because phylogenetic relationships allow for borrowing strength from closely related species sampled across a wider geographic range (Evans et al., 2016). The model is therefore suitable for large scale application.

Predicting large scale trait variation. To understand large scale variation in traits using our combined model and to compare it to the species-only and environment-only approaches, we integrated each of the three models with data on tree species abundance and topographic data from ~30,000 Forest Inventory and Analysis (FIA) plots (~1.2 million trees) and climate data from DayMet (Figure 3, S.7, S.8, S.9) to predict leaf traits across the eastern USA. Predictions from the combined model differed from species- and environment-only models at scales ranging from ecoregions to the entire eastern USA (Figure 4). These differences were complex, with different regions exhibiting shifts of different magnitudes and directions from either species- or environment-only approaches (Figure 4, Figure S.10, Supplement 4). 80% of ecoregion-trait-model combinations exhibited significant differences in predicted plot level traits between the combined and species- or environment-only models ($p < 0.0001$ in paired t-tests). These results showed that neither predictions from species-only nor environment-only produced traits distributions statistically overlapping with the combined model, suggesting that both species and environment contain non-overlapping biological information at large scales. Also, 93% of ecoregion-trait combinations exhibited different trait distributions predicted by the species-only and environment-only models ($p < 0.0001$ in a paired t-test), suggesting that in most cases the environment and species affects traits distribution in different ways and magnitude. This demonstrates the importance of a combined modeling approach for prediction as well as inference.
In addition to improving plot level trait predictions, one of the essential reasons for including both environmental drivers and species information in large scale trait models is that this combined approach allows for predictions of intraspecific variation (Figure 5). Species-level models, by definition, do not account for intraspecific variation; and although environment-only models include intraspecific variation, they also include trait variation due to shifts in species composition. This interspecific variation must be separated from intraspecific variation to predict the latter. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

joint-hierarchical structure of our model makes this possible for hundreds of unique species-trait combinations. This is important because the predicted degree of intraspecific variation varied widely in both form and magnitude across species. For example, the magnitude of intraspecific variation in N% ranged from less than 10% of the total observed N% variation for species with limited geographic ranges (e.g., *Populus heterophylla, Sabal palmetto, and Gleditsia aquatica*), to ~17% for conifers with broader geographic ranges (e.g. *Pinus palustris, Pinus taeda and Pinus echinata*), to over 60% of the total variation for broadleaf species wide-ranging across the eastern USA (like *Cercis canadensis, Betula lenta, and Carpinus caroliniana*).
The response of N% to temperature varied from the globally observed bell-shaped pattern (Reich & Oleksyn, 2004) to no relationship (only 91 out of 180 species showed meaningful relationships) and varied in range and magnitude depending on the species (Figure 5a, Figure S.11). While evaluating the accuracy of large scale predictions of intraspecific variation is challenging due to data limitations, the combined model produced realistic ranges of intraspecific variation when compared with available independent data for widely distributed species. Specifically, the intraspecific variation of N% predicted from the combined model showed ranges similar to those observed in independent data for three of the most widespread and abundantly sampled species in the NEON, FIA and TRY datasets (*Abies balsamea, Acer rubrum*, and *Fagus grandifolia*). For these species (Figure 5b-d), the combined model predictions showed N% covering a large range of values (1.12 N% on average), comparable to the average difference between evergreen needleleaf and broadleaf deciduous species in eastern USA (1.13 from NEON field data). This intraspecific variation holds fundamental information about how trees adapt to different environmental conditions (Albert et al., 2010, Henn et al., 2018) and is not captured by approaches relying only on species average trait values or only on the environment with no information about species.

Conclusions and Implications. Both species and environmental effects are fundamental to understanding the drivers and distribution of plant traits and their response to global change. Combining species and environmental effects in a single model is challenging due to raw data limitations. However, as we show here, a combined approach is possible with available large scale datasets and results in better prediction than models using either species average traits or the environment in isolation. Our results show that both interspecific variation driven by shifts in species' abundances and intraspecific variation related to the environment play key roles in shaping trait distributions at near-continental scales. The influence of these different components varies by species, trait, and ecoregion. Capturing this complexity requires models that account for intra- and interspecific trait variation, as in our combined modeling approach. This advance in large scale trait modeling was made possible by recent increases in open data on traits, phylogeny, and forest composition. Our approach overcomes previous data limitations by integrating these different sources of biological information to create a single integrated model for making large scale predictions of leaf traits that can be applied to species not included in the training data. This transferability relies on phylogenetic and environmental signals in the trait data. The advances in large scale prediction provided by our combined modeling approach should enable future work aiming to better understand the causes and ecosystem-level consequences of intra- and interspecific trait variation across space and time.

Materials and Methods

Data. We used provisional data from the National Ecological Observatory Network (NEON), the Botanical Information and Ecology Network (BIEN) and TRY to link information on leaf traits, species identity, and approximate locations for individual trees. We used Foliar Physical and Chemical Properties (DP1.10026.001)(and Vegetation Structure data (DP1.10098.001) from NEON to build joint trait distribution models with environmental drivers (climate and topography) alone, phylogenetic drivers (species identity and phylogeny) alone, and both (combined model). Linking the two different NEON datasets produced individual tree data with stem geolocation and measures of eight leaf traits (LMA, chlorophyll A and B, carotenoids, lignin, cellulose, C, N) for 542 trees in 21 sites across the USA (Figure S.1). Since foliar trait concentrations can vary significantly with phenology and position within the canopy (Niinemets et al., 2015), foliar samples

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

were collected at the yearly "peak of greenness" season and from the sunlit portion of the canopy. We tested the generalizability of our approach outside of NEON sites by evaluating it on independent (out of sample) data available from the Botanical Information and Ecology Network (BIEN) and the TRY dataset (Kattge et al., 2020, Appendix S1). These two datasets provide measures for a subset of traits including C, N and LMA for a total of 223 individual trees. We used data from the Open Tree of Life (Redelings BD, Holder, 2017) to measure phylogenetic distance between species. Data for environmental drivers included average monthly climate data from 1995 to 2015 (Appendix S1) extracted from Daymet (Thornton et al., 2018) and topographic variables (i.e. elevation, slope and aspect) reported in the NEON and FIA datasets. For three common eastern USA tree species (Acer rubrum, Fagus grandifolia, and Abies balsamea), we used all publicly available leaf N% data from the TRY database (Kattge et al., 2020) to quantify intraspecific variation in leaf N% across each species' geographic range within the USA. We selected these three species for the following reasons: (1) *Abies balsamea* is the needleleaf species with the most leaf N% data in TRY for the USA; (2) *Fagus grandifolia* is the broadleaf species with the most leaf N% data in TRY for the USA; and (3) *Acer rubrum* occurs throughout much of the eastern USA in a wide variety of habitats (e.g., from xeric to mesic; Burns and Honkala 1990) and has abundant leaf N% data in TRY. We used forest survey data from the Forest Inventory and Analysis (FIA) database (https://www.fia.fs.fed.us/) to estimate trait distributions for all individual trees surveyed in the FIA across the eastern USA from 2016 to 2019. We used Lv.3 ecoregions and Lv. 2 ecoprovinces as defined by the Environmental Protection Agency (McMahon 2001, Omernik et al., 2014) to analyze traits distribution at different scales.

Overview of Models. Here, we briefly summarize the statistical modeling framework. Additional details are in Appendix S2. We modeled the joint multivariate distribution of the eight leaf traits (the response variables) using three different approaches: (1) Environment-only model using climate and topography as fixed effects; (2) Phylogeny-only model using species as random effects, with covariances among the random effects structured by the phylogenetic tree for all woody species detected in the FIA; (3) Combined model including both environmental and species/phylogenetic effects. In the environment-only and combined models, environmental effects were fit using a generalized additive modeling approach, which allows for taking into account non-linear relationships between independent and dependent variable In the species-only and combined models, phylogenetic relationships across species were modeled by including species as a random effect and calculating their correlation structure from cross-species cophenetic distance (Paradis et al., 2019). The distance matrix was used to estimate the correlation structure across species and genuses, allowing for borrowing strength from widely sampled species to rare or unsampled taxa (de Villemeruil & Nakagawa, 2014). We used multivariate normal families, simple splines, and weakly informative priors in all cases (Appendix S2). To reduce the number of climate predictors and to avoid problems with collinearity, we calculated a specific PCA for each climate variable (net radiation, precipitation, vapor pressure, maximum and minimum temperature) using a 30 years' time series of monthly averages, from 1985 to 2015. We used the first component of each PCA as climate variables in the environment-only and combined models. To quantify uncertainty in predictions we used the Bayesian R^2 (Gelman et al., 2018). We used variance partitioning to quantify the relative effects of environment and phylogeny on intra- and interspecific trait variation. See additional methods details in Appendix S1-S3. Code for reproducing analyses is available on Zenodo (https://zenodo.org/badge/latestdoi/353383665).

Acknowledgments

This work was supported by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4563 to E. P. White and by the National Science Foundation through grant 1926542 to E. P. White, S. A. Bohlman, A. Zare, D. Z. Wang, and A. Singh; by the NSF Dimension of Biodiversity program grant (DEB-1442280) and USDA/NIFA McIntire-Stennis program (FLA-FOR-005470) to S. A. Bohlman; by the University of Florida Biodiversity Institute (UFBI) and Informatics Institute (UFII) Graduate Fellowship to Sergio Marconi. There was no additional external funding received for this study.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.C., Clark, T.D., Colwell, R.K., Danielsen, F., Evengård, B. and Falconi, L., 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. Science, 355(6332), p.eaai9214
2. Diaz, S., and M. Cabido. 2001. Vive la difference: plant functional diversity matters to ecosystem processes. Trends in Ecology & Evolution 16:646–655.
3. McGill, B.J., Enquist, B.J., Weiher, E. and Westoby, M., 2006. Rebuilding community ecology from functional traits. Trends in ecology & evolution, 21(4), pp.178-185.
4. Reichstein, M., Bahn, M., Mahecha, M.D., Kattge, J. and Baldocchi, D.D., 2014. Linking plant and ecosystem functional biogeography. Proceedings of the National Academy of Sciences, 111(38), pp.13697-13702.
5. Peaucelle, M., Bacour, C., Ciais, P., Vuichard, N., Kuppel, S., Peñuelas, J., Belelli Marchesini, L., Blanken, P.D., Buchmann, N., Chen, J. and Delpierre, N., Covariations between plant functional traits emerge from constraining parameterization of a terrestrial biosphere model. Global Ecology and Biogeography.
6. Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S.M., Botta-Dukát, Z., Chytrý, M., Field, R., Jansen, F. and Kattge, J., 2018. Global trait–environment relationships of plant communities. Nature ecology & evolution, 2(12), p.1906.
7. Simpson, A.H., Richardson, S.J. and Laughlin, D.C., 2016. Soil–climate interactions explain variation in foliar, stem, root and reproductive traits across temperate forests. Global Ecology and Biogeography, 25(8), pp.964-978.
8. Chave, J., 2013. The problem of pattern and scale in ecology: what have we learned in 20 years?. Ecology letters, 16, pp.4-16.
9. Violle, C., Reich, P.B., Pacala, S.W., Enquist, B.J. and Kattge, J., 2014. The emergence and promise of functional biogeography. Proceedings of the National Academy of Sciences, 111(38), pp.13690-13696.
10. Valladares, F., Matesanz, S., Guilhaumon, F., Araújo, M.B., Balaguer, L., Benito-Garzón, M., Cornwell, W., Gianoli, E., van Kleunen, M., Naya, D.E. and Nicotra, A.B., 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecology letters, 17(11), pp.1351-1364.
11. Münzbergová, Z., Hadincová, V., Skálová, H. and Vandvik, V., 2017. Genetic differentiation and plasticity interact along temperature and precipitation gradients to determine plant performance under climate change. Journal of Ecology, 105(5), pp.1358-1373.
12. Henn, J.J., Buzzard, V., Enquist, B.J., Halbritter, A.H., Klanderud, K., Maitner, B.S., Michaletz, S.T., Pötsch, C., Seltzer, L., Telford, R.J. and Yang, Y., 2018. Intraspecific trait variation and phenotypic plasticity mediate alpine plant species response to climate change. Frontiers in Plant Science, 9, p.1548.
13. Ordoñez, Jenny C., et al. "A global study of relationships between leaf traits, climate and soil measures of nutrient fertility." Global Ecology and Biogeography 18.2 (2009): 137-149.
14. Moles, A.T., Perkins, S.E., Laffan, S.W., Flores-Moreno, H., Awasthy, M., Tindall, M.L., Sack, L., Pitman, A., Kattge, J., Aarssen, L.W. and Anand, M., 2014. Which is a better predictor of plant traits: temperature or precipitation?. Journal of Vegetation Science, 25(5), pp.1167-1180.
15. Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H., Diemer, M. and Flexas, J., 2004. The worldwide leaf economics spectrum. Nature, 428(6985), p.821.
16. Anderegg, L. D., Berner, L. T., Badgley, G., Sethi, M. L., Law, B. E., & HilleRisLambers, J. (2018). Within-species patterns challenge our understanding of the leaf economics spectrum. Ecology letters, 21(5), 734-744.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

17. Clark, J. S. (2016). Why species tell more about traits than traits about species: predictive analysis. Ecology, 97(8), 1979-1993.
18. Yang, J., Cao, M. and Swenson, N.G., 2018. Why functional traits do not predict tree demographic rates. Trends in ecology & evolution, 33(5), pp.326-336
19. Swenson, N. G., & Weiser, M. D. (2010). Plant geography upon the basis of functional traits: an example from eastern North American trees. *Ecology*, *91*(8), 2234-2241.
20. Zhang, Y.J., Sack, L., Cao, K.F., Wei, X.M. and Li, N., 2017. Speed versus endurance tradeoff in plants: Leaves with higher photosynthetic rates show stronger seasonal declines. Scientific reports, 7, p.42085.
21. Wieczynski, D.J., Boyle, B., Buzzard, V., Duran, S.M., Henderson, A.N., Hulshof, C.M., Kerkhoff, A.J., McCarthy, M.C., Michaletz, S.T., Swenson, N.G. and Asner, G.P., 2019. Climate shapes and shifts functional biodiversity in forests worldwide. Proceedings of the National Academy of Sciences, 116(2), pp.587-592.
22. Swenson, N.G., Weiser, M.D., Mao, L., Araújo, M.B., Diniz-Filho, J.A.F., Kollmann, J., Nogués-Bravo, D., Normand, S., Rodríguez, M.A., García-Valdés, R. and Valladares, F., 2017. Phylogeny and the prediction of tree functional diversity across novel continental settings. Global Ecology and Biogeography, 26(5), pp.553-562.
23. Stahl, U., Reu, B. and Wirth, C., 2014. Predicting species' range limits from functional traits for the tree flora of North America. Proceedings of the National Academy of Sciences, 111(38), pp.13739-13744.
24. Niinemets, Ü., Keenan, T.F. and Hallik, L., 2015. A worldwide analysis of within-canopy variations in leaf structural, chemical and physiological traits across plant functional types. New Phytologist, 205(3), pp.973-993.
25. Messier, J., Lechowicz, M. J., McGill, B. J., Violle, C., & Enquist, B. J. (2017). Interspecific integration of trait dimensions at local scales: The plant phenotype as an integrated network. Journal of Ecology, 105(6), 1775– 1790. https://doi.org/10.1111/1365-2745.12755
26. Hinchliff, Cody E., et al. "Synthesis of phylogeny and taxonomy into a comprehensive tree of life." Proceedings of the National Academy of Sciences 112.41 (2015): 12764-12769.
27. National Ecological Observatory Network. 2020. Provisional data downloaded from http://data.neonscience.org on 2 Jan 2020. Battelle, Boulder, CO, USA
28. USDA Forest Service, 2001. Forest Inventory and Analysis National Core Field Guide, Volume I: Field Data Collection Procedures For Phase 2 Plots, Version 1.5. US Department of Agriculture, Forest Service, Washington, DC.
29. Smith, W.B., 2002. Forest inventory and analysis: a national inventory and monitoring program. Environmental pollution, 116, pp.S233-S242.
30. Thornton, M.M., P.E. Thornton, Y. Wei, B.W. Mayer, R.B. Cook, and R.S. Vose. 2018. Daymet: Annual Climate Summaries on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1343
31. Ordoñez, Jenny C., et al. "A global study of relationships between leaf traits, climate and soil measures of nutrient fertility." Global Ecology and Biogeography 18.2 (2009): 137-149.
32. Reich, P.B., Oleksyn., J., 2004. Global patterns of plant leaf N and P in relation to temperature and latitude. Proceedings of the National Academy of Sciences, 101 (30) 11001-11006; DOI: 10.1073/pnas.0403588101
33. Hedin, L.O., 2004. Global organization of terrestrial plant–nutrient interactions. Proceedings of the National Academy of Sciences, 101(30), pp.10849-10850.
34. Poorter, H., Niinemets, Ü., Poorter, L., Wright, I.J. and Villar, R., 2009. Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis. New phytologist, 182(3), pp.565-588.
35. Kitajima K, Wright SJ, Westbrook JW (2016) Leaf cellulose density as the key determinant of inter- and intra-specific variation in leaf fracture toughness in a species-rich tropical forest. Interface Focus 6:20150100.
36. Reich, P.B., Walters, M.B. & Ellsworth, D.S. (1997). From tropics to tundra: Global convergence in plant functioning. Proc. Natl. Acad. Sci. U.S.A., 94, 13730–13734.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

37. Croft, H., Chen, J.M., Luo, X., Bartlett, P., Chen, B. and Staebler, R.M. (2017), Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Glob Change Biol, 23: 3513-3524. doi:10.1111/gcb.13599
38. Osnas, J.L., Katabuchi, M., Kitajima, K., Wright, S.J., Reich, P.B., Van Bael, S.A., Kraft, N.J., Samaniego, M.J., Pacala, S.W. and Lichstein, J.W., 2018. Divergent drivers of leaf trait variation within species, among species, and among functional groups. Proceedings of the National Academy of Sciences, 115(21), pp.5480-5485.
39. Enquist, B.J., Condit, R., Peet, R.K., Schildhauer, M. and Thiers, B., 2009. The Botanical Information and Ecology Network (BIEN): Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. Salt Lake City, UT: iPlant Collaborative.
40. Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Tautenhahn, S., Werner, G.D., Aakala, T., Abedi, M. and Acosta, A.T., 2020. TRY plant trait database–enhanced coverage and open access. Global change biology.
41. Evans, M.E., Merow, C., Record, S., McMahon, S.M. and Enquist, B.J., 2016. Towards process-based range modeling of many species. Trends in Ecology & Evolution, 31(11), pp.860-871.
42. Albert, C.H., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P. and Lavorel, S., 2010. Intraspecific functional variability: extent, structure and sources of variation. Journal of Ecology, 98(3), pp.604-613.
43. Burns, R.M. & Honkala, B.H. (1990). Silvics of North America: 1. Conifers; 2. Hardwoods. Agriculture Handbook 654. U.S. Department of Agriculture, Forest Service, Washington, DC.
44. McMahon, G., Gregonis, S. M., Waltman, S. W., Omernik, J. M., Thorson, T. D., Freeouf, J. A., ... & Keys, J. E. (2001). Developing a spatial framework of common ecological regions for the conterminous United States. Environmental Management, 28(3), 293-316.
45. Omernik, J. M., & Griffith, G. E. (2014). Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environmental management, 54(6), 1249-1266.
46. Paradis, E., Blomberg, S., Bolker, B., Brown, J., Claude, J., Cuong, H.S. and Desper, R., 2019. Package 'ape'. Analyses of phylogenetics and evolution, version, 2(4).
47. De Villemeruil P. & Nakagawa, S. (2014) General quantitative genetic methods for comparative biology. In: Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice(ed. Garamszegi L.) Springer, New York. pp. 287-303.
48. Gelman, Ben Goodrich, Jonah Gabry & Aki Vehtari. (2018). R-squared for Bayesian regression models, The American Statistician. https://doi.org/10.1080/00031305.2018.1549100.

Figures and Tables

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

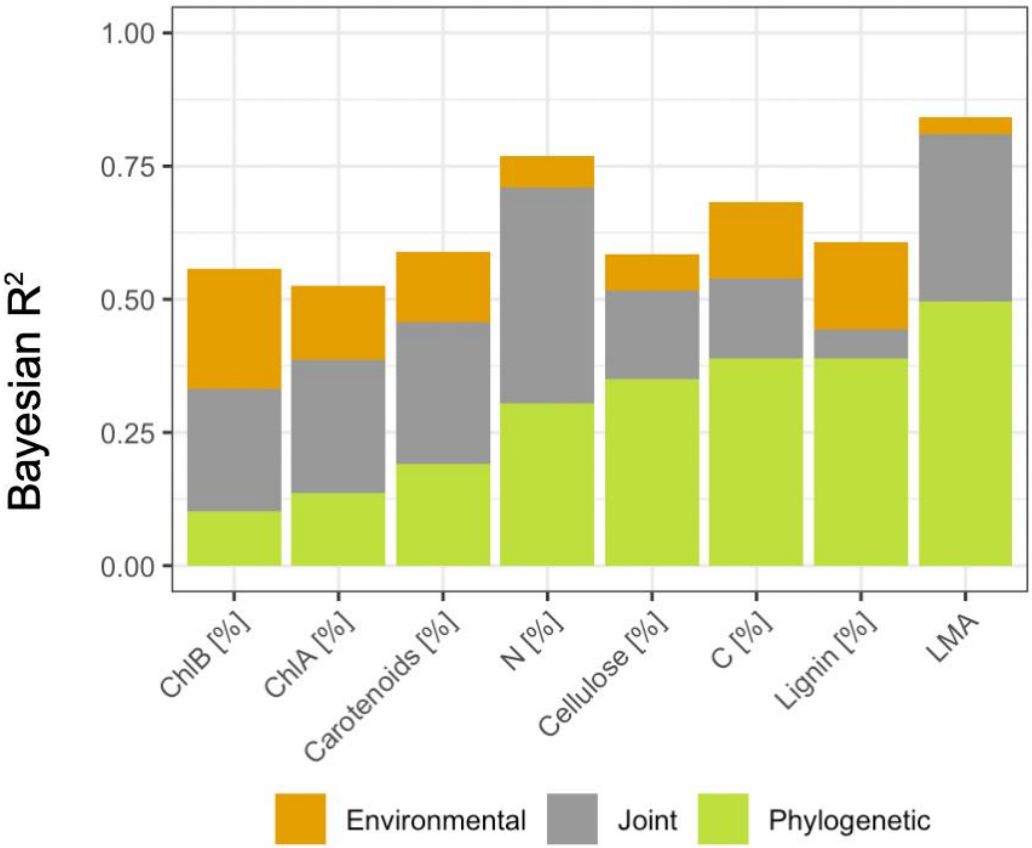

Figure 1. Variance partitioning between pure species, pure environmental (intraspecific variation), and joint species-environment effects based on the R^2 for independent validation data. Pure environmental variation is independent of species variation, and vice versa. Joint variation is jointly explained by both species and environmental effects and cannot be uniquely ascribed to either. LMA is leaf dry mass per area, and all other traits are % mass

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

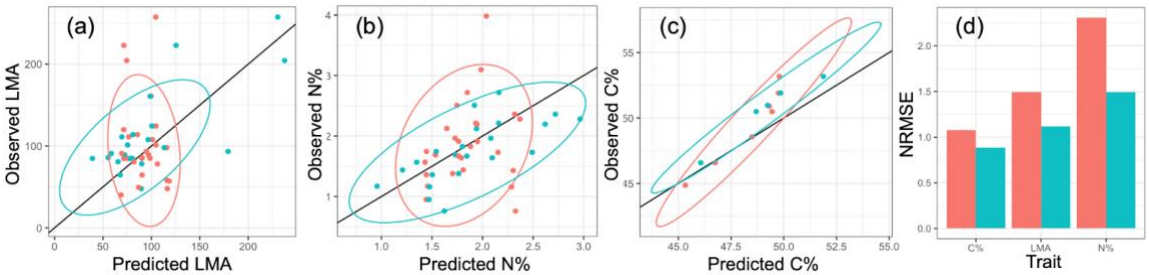

Figure 2. Observed vs. predicted values from the combined model for (a) LMA (g m^{-2}), (b) N%, and (c) C%, either assuming phylogenetically-structured covariance among species random effects (azure), or assuming independent species random effects (orange). Ellipses are 95% confidence ellipses. (d) RMSE (normalized by standard deviation) showing the difference between accounting for (azure) or ignoring (orange) phylogenetic relationships when predicting traits from untrained species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

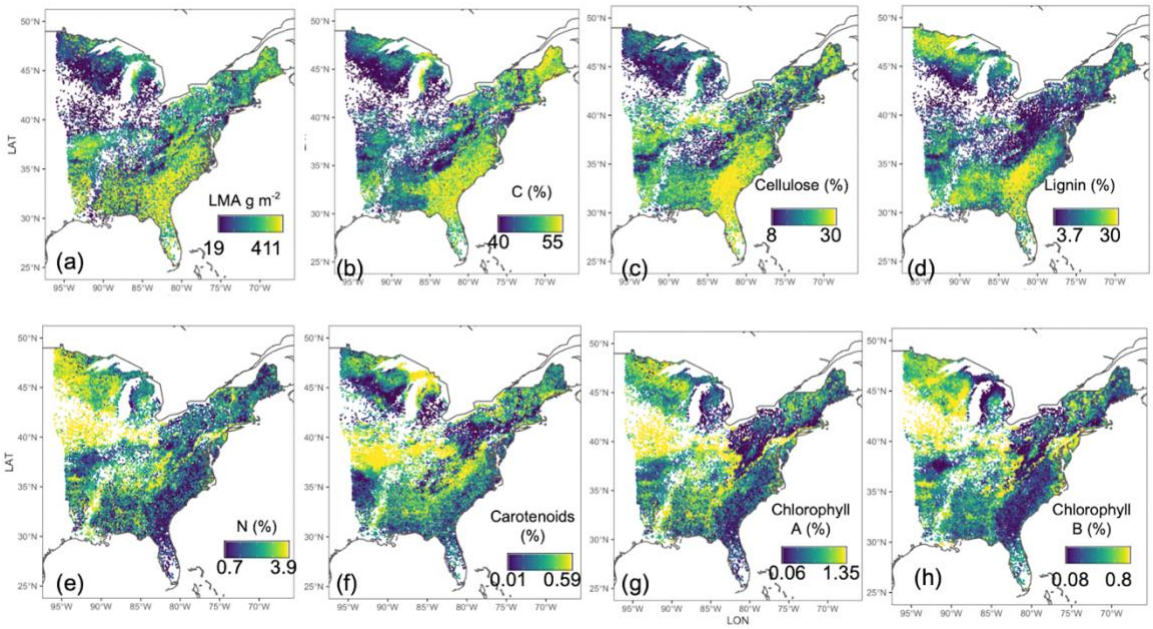

Figure 3. Trait distributions across the eastern USA for leaf structural (LMA, lignin, cellulose and %C), macro-nutrient (N%) and pigment (carotenoids, chlorophyll A and B) traits. Values represent species-abundance-weighted trait averages at each FIA plot (n = 30,331 plots).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

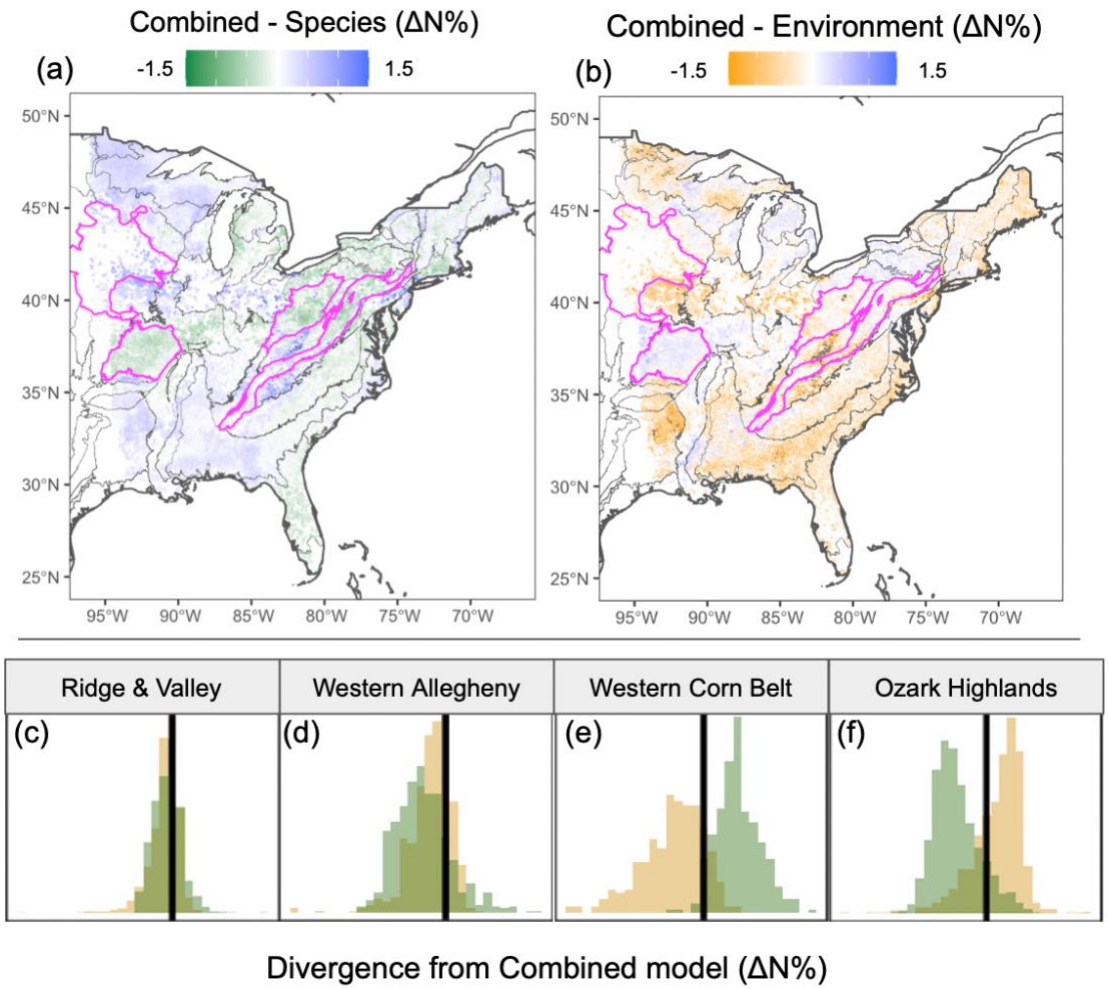

Figure 4. Divergence in N% between the combined model and the species-only (a) and environment-only (b) models. Dots represent FIA plots where predicted N% was higher in the combined (blue), environment (orange), or species (green) model, or plots where little to no divergence ($p > 0.0001$) was found (white). (c-f) Examples of divergence patterns in four ecoregions (see the outlined ecoregions in a-b, labeled c-f according to their corresponding histogram): divergence between the combined model and the environmental model (yellow) or the species model (green). Vertical black line represents FIA plots where the divergence from the combined model is 0. (c) combined, species, and environment models capturing similar information (ΔN%~ 0); (d) species and environment models similar to each other but diverging from the combined model; (e-f) divergence among all three models.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.01.438064; this version posted April 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

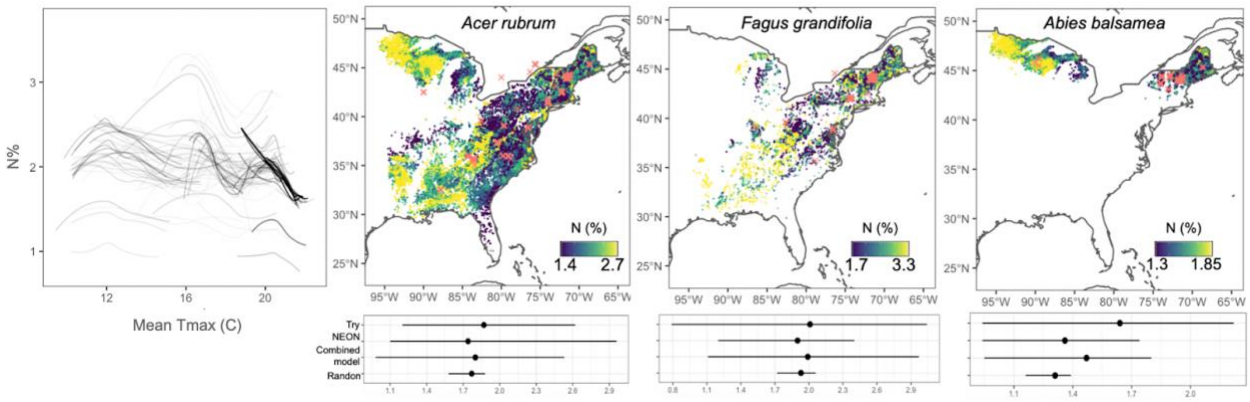

Figure 5. Intraspecific variation of predicted leaf N% for tree species in eastern USA. (a) Trends of the median N% prediction vs. maximum temperature for the 91 species showing significant effect of temperature on N% distribution (the darkness of the curve increases with the strength of the relationship between N% and temperature). Trends are estimated using GAMs and may be prone to artifacts at the extremes of their tails. (b-d) Geographic patterns and range of intraspecific N% for two widespread deciduous broadleaf (Acer rubrum and Fagus grandifolia) and one widespread needleleaf evergreen (Abies balsamea) tree species. Mean (points) and 99% intraspecific ranges (lines) in b-d are from: TRY data (within the eastern USA), NEON data (within the eastern USA), the combined model, and random error from the species-only model. Red crosses represent geolocation of TRY field data.