New Phytologist

Research

Seeing herbaria in a new light: leaf reflectance spectroscopy unlocks trait and classification modeling in plant biodiversity collections

Dawson M. White[1,2], Jeannine Cavender-Bares[1,2], Charles C. Davis[1,2], J. Antonio Guzmán Q.[2], Shan Kothari[3], Jorge M. Robles[4] and José Eduardo Meireles[5]

[1]Harvard University Herbaria, 22 Divinity Ave, Cambridge, MA 02138, USA; [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA; [3]Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada; [4]Department of Biological Sciences, California State University San Marcos, San Marcos, CA 92078, USA; [5]School of Biology and Ecology and University of Maine Herbarium, University of Maine, 5735 Hitchner Hall, Orono, ME 04469, USA

Authors for correspondence:
Dawson M. White
Email: dawson.white@gmail.com

Jeannine Cavender-Bares
Email: jcavender@fas.harvard.edu

José Eduardo Meireles
Email: jose.meireles@maine.edu

Received: *10 January 2025*
Accepted: *12 June 2025*

New Phytologist (2025)
doi: 10.1111/nph.70357

Key words: biodiversity digitization, extended specimen, functional traits, herbarium spectroscopy, partial least squares regression, plant diversity, plant phenomics, taxonomic discrimination.

Summary

- Reflectance spectroscopy is a rapid method for estimating traits and discriminating species. Spectral libraries from herbarium specimens represent an untapped resource for generating broad phenomic datasets across space, time, and taxa.
- We conducted a proof-of-concept study using trait data and spectra from herbarium specimens up to 179 yr old, alongside data from recently dried and pressed leaves. We validated model accuracy and transferability for trait prediction and taxonomic discrimination.
- Trait models from herbarium spectra predicted leaf mass per area (LMA) with R^2 = 0.94 and %RMSE = 4.86%. Models for LMA prediction were transferable between herbarium and pressed spectra, achieving R^2 = 0.88, %RMSE = 8.76% for herbarium to pressed spectra, and R^2 = 0.76, %RMSE = 10.5% for the reverse transfer. Discriminant models classified leaf spectra from 25 species with 74% accuracy, and classification probabilities were significantly associated with several herbarium specimen quality metrics.
- The results validate herbarium spectral data for trait prediction and taxonomic discrimination, and demonstrate that trait modeling can benefit from the complementary use of pressed-leaf and herbarium-leaf spectral datasets. These promising advancements help to justify the spectral digitization of plant biodiversity collections and support their application in broad ecological and evolutionary investigations.

Introduction

The urgency of global biodiversity assessment is driving the application of reflectance spectroscopy as a broadly informative technology for advancing systematic knowledge of plant diversity at scales ranging from molecules to continents (Serbin *et al.*, 2014; Cavender-Bares *et al.*, 2017, 2025; Meireles *et al.*, 2020a). This powerful approach offers a rapid method for characterizing leaf traits and discriminating taxa by capturing spectral signals that integrate structural, chemical, and physiological information from plants studied in laboratory, herbarium, and field settings (Costa *et al.*, 2018; Serbin & Townsend, 2020; Kothari & Schweiger, 2022).

Despite its potential, the spectral-based taxonomic and phenotypic characterization of plant diversity faces significant challenges. Limited access to material from remote geographic regions and uncommon taxa results in spectral datasets that are both biased and highly sparse (Meireles *et al.*, 2020a), even more so than global plant trait databases (Jetz *et al.*, 2016). Addressing this limitation requires extensive, costly, and time-intensive fieldwork. Additionally, the lack of linkage between leaf spectral data and voucher specimens complicates spatiotemporal precision and reliability as inevitable taxonomic and nomenclatural changes occur.

A promising path forward for bridging this impasse across the plant tree of life lies in leveraging the *c.* 400 million dried plant specimens stored in over 3500 herbaria worldwide (Thiers, continuously updated; Heberling, 2022; Kothari *et al.*, 2023b). This wealth of plant specimens has long been a key resource for researchers studying plant diversity and ecological and evolutionary processes across spatial and temporal scales (Davis, 2023). Indeed, herbarium collections anchor every species definition and are the physical foundation of our taxonomic understanding of plant and fungal diversity. They also include specimens that are rare, extinct, or regionally extirpated.

Several studies have now demonstrated the utility of pressed leaves (i.e. collected, dried, pressed, and stored in newsprint) for spectra-based trait prediction and taxonomic discrimination, offering a positive outlook for extending these applications to the more variable conditions of herbarium specimens (Durgante *et al.*, 2013; Costa *et al.*, 2018; Kothari *et al.*, 2023b; Hernández-Leal *et al.*, 2025). In contrast to pressed leaves, herbarium specimens typically reflect a much broader array of collection and preservation protocols – many of which are minimally documented – and are stored for considerably longer periods (Box 1). As such, herbarium specimens represent a much wider range of tissue variability with respect to their biological factors as well as processing and degradation. Modern spectroradiometers (350–2500 nm) are highly sensitive to the physical and chemical characteristics of scanned tissues, requiring careful standardization to ensure data quality and interoperability (Meireles *et al.*, 2020a). As such, the differences in collection and processing protocols among herbarium specimens, as well as mounting techniques, chemical treatments, long-term storage conditions, and age are expected to introduce spectral noise and reduce comparability across datasets (Kühn *et al.*, 2024). Herbarium specimens thus present unique challenges for reflectance spectroscopy, as they carry multiple layers of variation beyond natural biological differences – complicating both data interpretation and model transferability.

Within the new and rapidly evolving field of spectral biology, the application of reflectance spectroscopy to herbarium specimens is still in its early stages. For example, Kühn *et al.* (2025) demonstrated that herbarium spectra could be used to detect historical changes in leaf nitrogen, phosphorus, and carbon concentrations associated with shifts in agricultural management practices. In this issue, Neto-Bradley *et al.* (2025) have assessed taxonomic discrimination in *Lithocarpus*, a taxonomically challenging clade with largely homogeneous leaf and vegetative morphology, providing insights into methodologies and classification limits. Building on these efforts, the present study aims to evaluate the extent to which herbarium specimens can be used for estimating leaf traits and species classification using reflectance spectra.

Here, we extend the experimental framework established by Kothari *et al.* (2023b) for pressed leaves to investigate the utility of herbarium specimens for leaf trait prediction and species discrimination. We targeted 25 of the most well-sampled species from the Kothari *et al.* dataset for spectral measurement at the Harvard University Herbaria (HUH), enabling direct comparison between pressed and herbarium spectra. We focused on predicting leaf mass per area (LMA) because it was the best-performing trait in pressed-leaf models and is minimally invasive. LMA can be directly measured without altering specimens if detached leaves are available in specimen packets (see Fig. 1b). We also used this framework to evaluate the transferability and 'generalizability' of trait prediction models from pressed leaves to herbarium spectra and vice versa, as a proxy to understand how herbarium variation and degradation affect spectral information and models. Finally, we investigated whether herbarium specimen qualities – including age, greenness, and the presence of glue – were correlated with the probability of correct taxonomic classification.

Our validation approach highlights practical considerations – such as trait range, model transferability, and specimen quality – that influence the reliability of spectral inferences from herbarium specimens. These findings inform future efforts to develop and apply spectral models across diverse herbarium collections.

Materials and Methods

Sampling design

We reanalyzed the pressed-leaf spectral dataset from Kothari *et al.* (2022b), which includes 618 leaf samples representing 67 species of North American trees, shrubs, and herbs, along with one Australian species included as a complementary pressed-leaf spectral and trait dataset. This dataset includes the values for 22 leaf traits assayed for each sample. Spectral data from Kothari *et al.* (2023b) were collected using a PSR+ spectroradiometer equipped with a leaf clip optical probe (Spectral Evolution Inc.) on pressed voucher specimens following 6 months to 3 yr of storage. While Kothari *et al.* (2023b) used spectra averaged across replicates for each individual (see Kothari *et al.*, 2022), the present study instead analyzes the original, unaveraged spectra (see Data Availability Statement).

To enable a comparison with the pressed-leaf dataset, we generated a corresponding herbarium dataset from specimens housed at the HUH for 25 of the 68 species analyzed by Kothari *et al.* (2023b). A comparison of individuals and numbers of spectral measurements from each dataset is provided in Table 2. Specimen metadata were obtained from the Global Biodiversity Information Facility (GBIF.org) database using the R package RGBIF v.3.8.0 (Chamberlain & Boettinger, 2017; R Core Team, 2023; Chamberlain *et al.*, 2024). We targeted collections from New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont), contrasting with the geographic focus on Ontario and Quebec in Kothari *et al.* (2023b).

To select herbarium specimens for measurement, we first inspected all specimens per species and selected those holding loose leaves in packets. If we were not able to get a minimum of 15 specimens with loose leaves, we obtained permission from Lisa Standley (curator of the New England Botanical Club Herbarium) and Michaela Schmull (Director of Collections for the HUH) to detach one leaf for measuring spectra and LMA. If multiple leaves were available, we selected leaves without any sign of glue, but otherwise sampled specimens randomly with respect to the visual quality and degree of degradation.

Spectral measurement protocol

Specimens were measured using a Spectra Vista Corporation HR 1024i spectroradiometer (350–2500 nm spectral range) with a fiber-optic cable connected to the LC-RP Pro Leaf Clip/

Box 1. Pressed vs herbarium specimens

Fig. 1 Spectral measurements of pressed (a) and herbarium (b) leaf specimens. Pressed leaves (a; photograph by J. Cavender-Bares) are unmounted and easily scanned, while herbarium specimens are mounted and more variable in preservation. In (b), a detached leaf fragment from the packet is measured on a black background to avoid spectral interference (specimen from A: Herbarium of the Arnold Arboretum of Harvard University; photograph by D.M. White).

Table 1 Summary of differences in storage, age, collection, and preservation methods, contamination risk, and spectral integrity between pressed and herbarium specimens.

Characteristic	Pressed tissues	Herbarium tissues
Age	Months to decades (0.5–3 yr in Kothari *et al*., 2023a,b)	Years to centuries (1–179 yr in this study)
Storage	Loose in newsprint	Mounted on archival sheets, can have loose tissues in packets
Variability in collection and preservation	Low; few collectors, consistent processing	High; many collectors, variable processing
Spectral contamination risk	Low	High (paper, glues, pesticides)
Presumed spectral integrity	Good	Variable; more noise due to aging, mounting, and storage effects

Recent advances have shown that reflectance spectra from recently dried leaves can produce accurate predictive models for taxonomic discrimination and leaf traits – comparable to those based on fresh tissue (Durgante *et al*., 2013; Costa *et al*., 2018; Kothari *et al*., 2023b). These results support extending spectral analyses to herbarium specimens, which span a broad range of ages and preservation conditions. While pressed and herbarium specimens share many features, key differences in storage, processing, and preservation justify their comparison as distinct sample types (Table 1).

Pressed specimens are typically prepared using standard herbarium protocols – collected, pressed in newsprint, and dried – and are usually associated with ongoing research projects. These specimens are relatively young (from months to a few decades), stored loosely in paper, and easily accessible for measurement on both leaf surfaces (Fig. 1a). They often serve as taxonomic vouchers and are often intended for future herbarium accessioning.

Herbarium specimens, by contrast, represent decades to centuries of collecting history. Their preservation is more variable, influenced by differences in field and processing techniques, storage environments (Forman & Bridson, 1989), and the use of chemicals such as glues, pest treatments, or chemical preservatives (Bieker *et al*., 2020), all of which can influence spectral signals. Many specimens have also been transferred between institutions, adding further variability.

A major distinction is that herbarium specimens are generally mounted on archival paper (Fig. 1b) – often glued – which can complicate spectral measurement due to interference from adhesives and backing materials (Neto-Bradley *et al*., 2025). Measuring such specimens often requires selecting loose tissues from packets or inserting nonreflective black backgrounds when mounting allows. Some herbaria store specimens unmounted in newsprint, more similar to pressed collections.

Pressed leaves thus represent a more uniform and accessible subset of the broader variability found in herbarium collections. They are a valuable resource for spectral model development and offer a critical intermediate between fresh tissues and historical collections. Their consistency and accessibility make them ideal for establishing transferable models that bridge *in vivo* trait measurements with the preserved diversity in global herbarium collections.

Table 2 Sampling design for herbarium and pressed datasets.

Species	Family	Herbarium		Pressed	
		n individuals	*n* spectra	*n* individuals	*n* spectra
Acer rubrum L.	Sapindaceae	20	72	72	302
Acer saccharinum L.	Sapindaceae	20	69	21	93
Acer saccharum Marshall	Sapindaceae	22	81	41	195
Acer spicatum Lam.	Sapindaceae	20	75	1	3
Agonis flexuosa (Willd.) Sweet	Myrtaceae	15	86	67	351
Betula papyrifera Marshall	Betulaceae	16	63	21	98
Betula populifolia Marshall	Betulaceae	21	96	86	403
Claytosmunda claytoniana (L.) Metzgar & Rouhan	Osmundaceae	18	56	1	7
Fagus grandifolia Ehrh.	Fagaceae	21	63	26	119
Helianthus divaricatus L.	Asteraceae	16	54	1	3
Myrica gale L.	Myricaceae	19	57	1	2
Osmunda regalis L.	Osmundaceae	20	72	1	2
Ostrya virginiana (Mill.) K.Koch	Betulaceae	20	60	1	4
Phalaris arundinacea L.	Poaceae	18	57	6	21
Phragmites australis (Cav.) Trin. ex Steud.	Poaceae	18	57	11	34
Populus grandidentata Michx.	Salicaceae	19	63	21	104
Populus tremuloides Michx.	Salicaceae	17	83	102	512
Prunus pensylvanica L.f.	Rosaceae	21	69	2	5
Prunus serotina Ehrh.	Rosaceae	20	63	1	5
Quercus rubra L.	Fagaceae	19	67	26	125
Rubus idaeus L.	Rosaceae	22	72	9	46
Rubus odoratus L.	Rosaceae	16	54	1	3
Solidago altissima L.	Asteraceae	19	57	6	29
Solidago gigantea Aiton	Asteraceae	21	63	7	35
Spiraea latifolia (Aiton) Borkh.	Rosaceae	22	81	2	4

For the pressed dataset, species with 11 or fewer individuals were excluded from the taxonomic classification analysis, following the approach of Kothari *et al*. (2023b).

Reflectance Probe with a narrow-angle lens, which reduced the measurement area to a 6 × 4 mm diameter ellipse. Throughout this manuscript, we refer to spectral 'measurements' as the method of reflectance data acquisition obtained using a contact probe with a fixed field of view. Before spectral measurements, the instrument was turned on for a minimum of 15 min with the reflectance probe lamp set to low, to allow the light source to warm and the sensors to cool. At the beginning of each session, the lamp was switched to high and a white reference measurement on a white Spectralon® reference panel was taken, followed by three spectral measurements of the white Spectralon® reference panel, followed by three measurements of our black background material: black cardstock sprayed with three coats of Krylon® Camouflage Black Matte spray paint (acrylic alkyd, water-based paint; product #K04290777). All measurements were made with an integration time of 2 s.

For one to two leaves per specimen, one leaf at a time was placed on top of the black background, and three measurements were made of the leaf lamina on the adaxial surface. The reflectance probe was rotated slightly and moved a few millimeters between each measurement to capture variability within each leaf across a small leaf area. Following Kothari *et al.* (2023b), we targeted leaf regions that avoided the midvein, prominent secondary veins, or regions with disease, fungus, or other damage. To further avoid possible contamination of light reflected from the bench, the leaves on top of the cardstock were placed on top of a 5-mm felt pad coated with the matte black spray paint (visible in Fig. 1b). After each specimen's measurements, a second white reflectance measurement was taken; all white and black target measurements were recorded for future monitoring of instrument and optics quality control (not described here).

Trait measurements

Leaf weight, area, and thickness were recorded for each measured leaf to validate LMA predictions from spectra. After spectral measurements were made, petioles were removed at the point of contact with the leaf lamina or at the midpoint of acuminate leaf bases. Leaf blade weight was measured in milligrams using a Sartorius Practum64-1S Analytical Balance. Petioles were stored in glassine envelopes and labeled with leaf numbers. Leaf area was measured using the LeafByte® app on an iPhone 15 with 5 or 10 cm^2 calibration dots. LMA was calculated in kilograms per square meter ($kg\ m^{-2}$).

Spectra preprocessing

We used the SPECTROLAB v.0.0.18 R package (Meireles *et al.*, 2017) to combine herbarium spectra files with their associated metadata and to smooth sensor overlap regions at

991.3 nm and 1902.5 nm with a 5-nm interpolation region. To ensure compatibility with downstream analyses and comparability of results across datasets, we reprocessed and reanalyzed the pressed leaf spectra of Kothari *et al.* (2022, 2023b), which were in 1 nm resolution instead of the *c.* 1.5 nm resolution of the herbarium dataset. We resampled reflectance spectra of both datasets to 5 nm intervals using the Full Width Half Maximum (FWHM) method in the CWT R package (Guzmán, 2024). The FWHM method was chosen as it is the standard function applied to downsampling spectra.

With the goal of optimizing the transferability of models across spectral datasets, the resampled reflectance spectra in each dataset were then transformed using two methods: vector normalization and continuous wavelet transformation (CWT). Vector normalization of the spectra was implemented as a method to reduce the impact of differences in illumination geometry between spectrometers, which can impact the magnitude of reflectance. This method was applied using the 'normalize' function of *SpectroLab*. Continuous wavelet transformation was implemented as a method to isolate scales that capture spectral features, potentially enhancing the prediction of leaf traits and the transferability of models (Guzmán & Sanchez-Azofeifa, 2021). This method is based on the premise that the leaf reflectance spectra can be expressed as a combination of wave-like functions (wavelets) of varying scales (widths), enhancing fine spectral features at lower scales and broader spectral patterns at larger scales (Rivard *et al.*, 2008). We applied this transformation on the resampled leaf reflectance from both datasets using a second-order Gaussian derivative wavelet function with a variance of 1. The selection of the wavelet function and its variance was done assuming that individual spectral features follow ideal Gaussian distributions (Rivard *et al.*, 2008). The choice of wavelet scales can impact the predictive performance of predictive models (Guzmán & Sanchez-Azofeifa, 2021). Based on exploratory analysis, scales 2^2, 2^3, and 2^4 were computed and summed to form the summed-wavelet spectra used for predicting leaf traits. The CWT transformation was implemented using the 'cwt' function from the *CWT* package in R (Guzmán, 2024).

The resulting reflectance spectra (e.g. reflectance, vector-normalized, and summed-wavelet) were trimmed to a range of 450–2400 to remove noisy regions at the spectrum's edges (Supporting Information Fig. S1), as has been done in other studies (Guzmán & Sanchez-Azofeifa, 2021; Ji *et al.*, 2024). We also subdivided the data into different spectral regions: 450–1300 nm as the visible and near-infrared (VNIR+) region ('+' because 1100–1300 nm is in the short-wave infrared) that could be noisier due to pigment degradation (Fourty *et al.*, 1996), and the 1350–2400 nm short-wave infrared region (SWIR).

Prediction of leaf traits

Using the processed spectra and the measured LMA (kg m^{-2}) from each of the pressed and herbarium datasets across the VNIR+ (450–1300 nm), SWIR, and full-range spectral regions, we built predictive models using partial least squares (PLS) regression implemented with the *pls* and *caret* R packages. Metadata and spectral data were split into training (75%) and validation (25%) datasets using a stratified sampling approach based on growth form, mirroring Kothari *et al.* (2023b). We generated 1000 model segments by randomly selecting individual measurements for each specimen using a custom data segmentation function. This procedure ensured that measurements from each specimen were never split between both the training and validation datasets while capturing the variability within specimens and any rare spectral features that might be removed by the averaging of spectra.

Model optimization was performed using a custom tuning function that used cross-validation with the 'oscorepls' method. The predictive residual sum of squares (PRESS) metric was used to evaluate the models during cross-validation, and the optimal number of components for the PLS regression models was selected as the smallest value whose PRESS value was within one SD of the minimum PRESS value.

Final models were constructed using the optimal number of components and validated on the independent test datasets. We evaluated our predictions using the full ensemble of model segments, averaged to each individual, and predictions of LMA were compared to observed values to calculate residuals and evaluate performance. The model performance was evaluated by estimating the coefficient of determination (R^2), the bias, the root mean squared error (RMSE), and the percentage RMSE (%RMSE = RMSE/range of 0.99 and 0.01 quantiles). We calculated variable importance in projection (VIP) values to estimate the most informative spectral regions and extracted model coefficients used in external predictions and tests of model transferability between pressed and herbarium specimens.

To directly evaluate transferability, we applied model coefficients derived from one (herbarium/pressed) LMA dataset to the spectra of the other. We then assessed transfer prediction accuracy by calculating residuals and comparing predicted vs observed values. This approach allowed us to test the generalizability of LMA models and the compatibility between herbarium and pressed-leaf spectral data.

Lastly, we used the trait values beyond LMA from Kothari *et al.* (2023b), including carbon, calcium, carotenoids, cellulose, Chl*a*, nitrogen, and solubles, to generate PLSR models in the same manner. We generated model coefficients and predicted trait values from the herbarium leaf spectra for these traits. To assess the generalizability and trait value consistency of model transfers for the traits for which we had no observed herbarium trait values, we compared the distributions of predicted herbarium trait values against the observed values from Kothari *et al.* (2023b).

Taxonomic discrimination

To test the viability of models classifying herbarium leaf spectral measurements into taxa, we applied partial least squares discriminant analysis (PLS-DA) and linear discriminant analysis (LDA) to the reflectance spectra of the full-range herbarium spectral dataset. We tested both the PLS-DA and LDA algorithms because they are both commonly applied classification

algorithms. PLS-DA uses partial least squares regression to reduce dimensionality and optimize feature selection, making it suitable for spectral datasets, especially in scenarios with few samples compared to many predictors (high-dimensional low-sample-size problems; Geladi & Kowalski, 1986; Carrascal *et al.*, 2009; Serbin & Townsend, 2020). This method requires researchers to specify the number of components used by the model to balance between improving accuracy and avoiding overfitting to the training dataset. LDA, by contrast, assumes normally distributed data and separates classes by maximizing variance between groups, offering robust classification in well-distributed datasets without the need to specify a number of components.

Classification models were built using the CARET, PLS, and PLSVARSEL packages in R (Kuhn, 2008; Mehmood *et al.*, 2012; Liland *et al.*, 2024). First, spectral data were preprocessed by splitting the dataset into 10 individuals per species selected for training and the rest for validation, ensuring balanced representation across species. The same segmentation process as above was employed to generate 1000 data segments for iterative training and testing across spectral measurements.

For PLS-DA, model tuning was performed with the PLS method and optimized by the classification accuracy metric. We generated final models across our 1000 data segments by selecting the number of components returning the highest classification accuracy. LDA models were generated with the 'LDA' method optimized by the accuracy metric.

Model performance was assessed using the independent test datasets by generating confusion matrices to calculate accuracy, sensitivity, and specificity metrics. We also generated VIP scores from the models to identify the most influential spectral regions for distinguishing taxa and extracted and saved coefficients from the PLS-DA models for generating class predictions and prediction probabilities from all specimens for an analysis of factors that influence classification success.

Analysis of specimen predictors on taxonomic descrimination

To evaluate the biotic and herborization factors influencing the success of PLS-DA classification, we utilized the full ensemble of 1000 optimized PLS-DA models trained on the full-spectrum herbarium dataset of 25 species. To evaluate classification performance, we used two related but distinct metrics: classification probability and classification accuracy (also referred to as probability of correct classification). Classification probability refers to the value calculated by the PLS-DA model for each reflectance spectrum for each predicted class. This continuous value (ranging from 0 to 1) is calculated from the coefficients of the PLS-DA model and reflects the model's internal confidence in its classification, enabling probabilistic analysis of how specimen characteristics influence prediction strength. By contrast, classification accuracy describes the overall probability that measurements from a given class – or from all classes collectively – are correctly classified. It summarizes the model's performance at the group or dataset level.

Using custom R scripts, we computed classification probabilities for all classes for all 1690 herbarium leaf measurements across the ensemble of models and used these values to examine the effect of specimen predictors, specimen characteristics believed to affect spectra, and model performance, on model confidence at the measurement level. Specifically, we conducted a series of comparisons and independent regressions of classification probabilities against four categorical variables (specimen quality, glue presence, observed damage, and leaf developmental stage) and five numerical variables (age, Julian day of collection, nearest taxon distance, LMA, and greenness index). All specimens were scored by JMR with initial input from DMW. Descriptions of predictor variables are provided in Table 3.

To estimate nearest taxon distance, a phylogram was made using TIME TREE 5 (timetree.org; Kumar *et al.*, 2022) with modifications following results from V.PHYLOMAKER v.2 (Jin & Qian, 2022) to add *Phragmites australis* as sister to *Phalaris arundinacea* at 39.8 million years (Myr) and add *Betula populifolia* as sister to *Betula papyrifera* at 39.7 Myr. Greenness index, which measures the relative difference in reflectance between green light (550 nm) and red light (690 nm; see equation in Table 3), was selected over other commonly used vegetation indices, such as normalized difference vegetation index, green normalized difference vegetation index, and Chl/carotenoid index, due to its significant correlation with the independent estimate of specimen quality (Fig. S2).

Relationships and regressions were visualized using the GGPLOT2 package in R (Wickham, 2016), and significant differences in classification probabilities between correct and incorrect classes were assessed using *t*-tests as implemented in the 'ggsignif' function in GGPLOT2.

To evaluate predictors of classification accuracy, we performed logistic regression and random forest analyses. Classification probabilities were averaged across all models, and the class with the highest average probability was assigned as the predicted class. The binary measure of correct or incorrect classification was used as the response variable in both analyses. The logistic regression model was implemented with the 'glm' function in the STATS R package and using a binomial error structure. Random Forest analysis, performed using the RANDOMFOREST R package (Liaw & Wiener, 2002), quantified predictor importance based on mean decrease in accuracy and Gini impurity metrics.

Results

Trait prediction and model transferability

Spectral profiles of 25 species from the HUH have a similar shape but lower magnitudes compared to pressed leaves (Fig. 2a). Within herbarium spectra, we also observe notable variation in the coefficient of variation of reflectance within the visible (450–700 nm) and SWIR regions (specifically *c.* 1900–2400; Fig. S3). Models trained on herbarium spectra using all combinations of spectral transformations (untransformed, vector-

Table 3 Metadata predictors from herbarium specimens recorded for each leaf and used to evaluate model utility.

Metadata predictor	Class	Description
Leaf developmental stage	Young	Thin leaves with underdeveloped venation, prone to bruising, may appear darker; measurements usually have lower reflectance. Collection date is informative.
	Mature	Typically thick leaves, with potential color differences between adaxial and abaxial surfaces.
	Senescent (not observed)	Discolored leaves, often associated with aging. Collection date may help confirm senescence.
Leaf damage	None	No visible damage to any leaves on the herbarium sheet. Damage includes factors like herbivory, burning during specimen drying, or any physical damage before or after collection.
	Minor	Physical damage visible on some leaves on the specimen, but no damage on the measured leaf.
	Medium	Damage visible on measured leaves, but no damage is present in the measured target area.
	Major	Damage is visible in the measured target area.
Specimen quality	Good	A well-pressed and dried specimen with leaves that are flat as they would appear *in vivo*. Specimen presents minimal discoloration.
	Medium	Leaves show some discoloration and/or curling that may indicate wilting caused by a delay in specimen pressing and drying.
	Poor	Highly degraded specimen, with discoloration, mold, or curling/rugosity from wilting. These factors were likely caused by delayed or inadequate specimen pressing and preservation in the field before drying.
Glue	Present	Glue expected in the measured target area.
	Absent	No glue expected in the measured target area.
Green index	(Numerical)	Green index = $\text{Reflectance}_{550nm} - \text{Reflectance}_{690nm}/\text{Reflectance}_{550nm} + \text{Reflectance}_{690nm}$
Age	(Numerical)	Years since specimen was collected (median = 91)
Day of year	(Numerical)	Julian day of collection
Leaf mass per area	(Numerical)	kg m^{-2}
Nearest taxon distance	(Numerical)	Estimated age (in Myr) of most recent common ancestor shared between predicted taxon and nearest sampled species.

normalized, and CWT) and wavelength ranges (full, VNIR+, and SWIR) had performance Pearson's correlation coefficient values (R^2) between 0.91 and 0.94, as compared to the pressed models with R^2 values between 0.93 and 0.95 (validation tests in Table 4; full statistics in Table S1).

Overall, the best herbarium validation models according to R^2 and %RMSE were the full-range, vector-normalized models, but the models using untransformed reflectance values were only slightly less accurate. For the nontransformed reflectance dataset, pressed LMA models performed similarly to the herbarium LMA models (pressed R^2 = 0.94, %RMSE = 6.29%; herbarium R^2 = 0.93, %RMSE = 5.18%, Fig. 3a,b). After full-range models, SWIR models generally performed slightly better than VNIR+ in the herbarium models, but the reverse was true with the pressed models (Table S1).

As expected, the performance of models was reduced when they were transferred and validated with the other (herbarium or pressed) LMA dataset, but the CWT and nontransformed reflectance models could still accurately predict observed LMA (Tables 4, S1; Fig. 3b,c). The best transfer model was for the full-range CWT dataset (herbarium to pressed R^2 = 0.88, %RMSE = 8.76%; pressed to herbarium R^2 = 0.76, %RMSE = 10.53%). The shifted slope of an ordinary least squares regression of predicted values highlights a systematic difference in models between datasets (0.91 in Fig. 3c; 1.25 in Fig. 3d; transfer tests in Table 4). Models based on the VNIR+ spectra also performed well for untransformed reflectance and CWT datasets, but SWIR-based models showed reduced performance (Table S1). Contrasting with their improved performance in internal validation tests, the models based on vector-normalized spectra performed less well than the other two datasets, yet showed best performance for models in the SWIR range (Tables 4, S1).

The compatibility of the models is further illustrated by the similarity of VIP values for reflectance spectra (Fig. 2c). The VIP plots reveal considerable differences between herbarium and pressed models in the visible and (less so) NIR regions, but the relative values across wavelengths in the SWIR region are similar. This same pattern applies to the model coefficients (Fig. 2c). The CWT models show a similar pattern across the visible, NIR, and SWIR regions, with higher similarity among the peaks and overall closer magnitudes (Fig. 2d,f). The CWT models have the most clearly defined peaks and highlight informative spectral regions throughout the spectral range (Fig. 2d; peaks = VIS: 500, 545, 590, 640, 670, 695 nm; NIR: 730 nm; SWIR: 1200, 1400, 1440, 1655, 1705, 1875, 1920, 2225, 2295 nm).

To extend the inference of the utility of transferring trait models, we applied seven additional pressed-leaf trait models to predict traits from the herbarium spectra for 25 species (Fig. 4; validation results in Table S2). The predicted trait distributions from herbarium spectra closely align with observed distributions from the pressed dataset, highlighting the potential of these models for cross-dataset applications. Predicted values for key traits, including LMA, carbon fractions, and carotenoids, generally showed contiguous distributions with substantial overlap between datasets. This overlap demonstrates the general utility of the spectral models in maintaining rank-order consistency across species. However, notable discrepancies were observed for some traits and taxa. For example, carbon predictions showed differing distributions for many species, and several traits differed

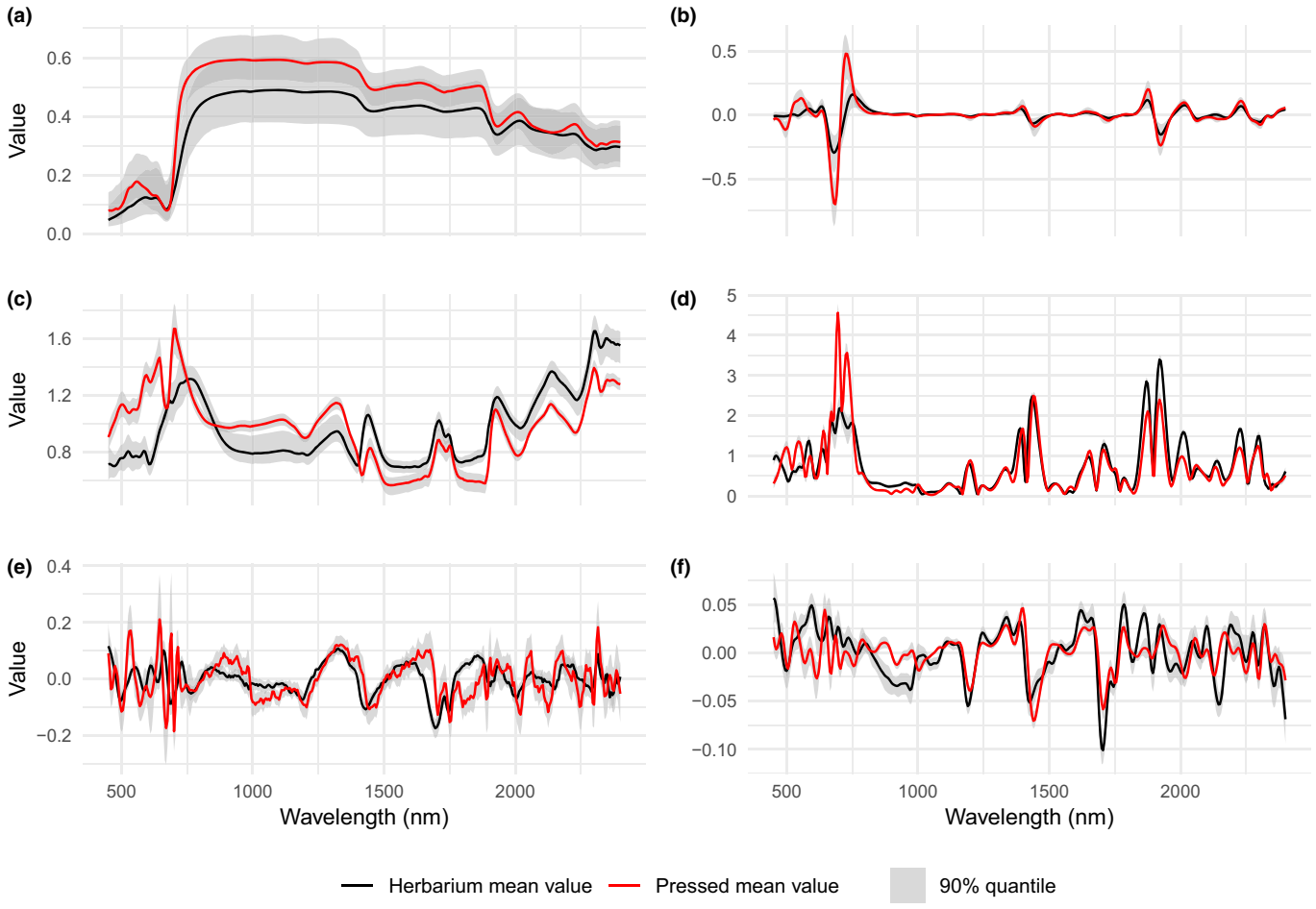

Fig. 2 Plots of reflectance and continuous wavelet transformation (CWT) values for herbarium and pressed leaf datasets, associated variable importance in projection (VIP) metrics, and model coefficients for leaf mass per area models. Black lines represent mean herbarium data, and red lines represent mean pressed leaf data, with 90% quantiles plotted in gray bands. Panels show the data for (a) untransformed reflectance across all samples, (b) CWT-transformed reflectance across all samples, (c) VIP values for reflectance data across 1000 model iterations, (d) VIP values for CWT data across 1000 model iterations, (e) reflectance model coefficients across 1000 iterations, and (f) CWT model coefficients across 1000 iterations.

Table 4 Performance metrics for LMA models (full range) averaged across 1000 model segments.

Test	Model	Spectra	Transform	*n*	*n* components	R^2	%RMSE	RMSE (kg m^{-2})	BIAS	Slope	Intercept
Validation	Herbarium	Herbarium	CWT	220	10	0.93 ± 0.01	5.31 ± 0.15	0.01 ± 0.00	0.00 ± 0.00	0.97 ± 0.02	0.00 ± 0.00
Validation	Herbarium	Herbarium	Reflectance	220	14	0.93 ± 0.01	5.18 ± 0.15	0.01 ± 0.00	0.00 ± 0.00	0.98 ± 0.02	0.00 ± 0.00
Validation	Herbarium	Herbarium	Normalized	220	14	0.94 ± 0.01	4.86 ± 0.20	0.01 ± 0.00	0.00 ± 0.00	1.01 ± 0.01	0.00 ± 0.00
Validation	Pressed	Pressed	CWT	212	8	0.94 ± 0.00	6.34 ± 0.10	0.01 ± 0.00	0.00 ± 0.00	1.03 ± 0.02	0.00 ± 0.00
Validation	Pressed	Pressed	Reflectance	212	16	0.94 ± 0.00	6.29 ± 0.12	0.01 ± 0.00	0.00 ± 0.00	1.02 ± 0.02	0.00 ± 0.00
Validation	Pressed	Pressed	Normalized	212	13	0.95 ± 0.00	6.01 ± 0.12	0.01 ± 0.00	0.00 ± 0.00	1.00 ± 0.02	0.00 ± 0.00
Transfer	Herbarium	Pressed	CWT	609	14	0.88 ± 0.03	8.76 ± 0.02	0.02 ± 0.00	0.00 ± 0.01	0.91 ± 0.06	0.00 ± 0.01
Transfer	Herbarium	Pressed	Reflectance	609	14	0.91 ± 0.01	10.99 ± 0.02	0.02 ± 0.00	−0.01 ± 0.01	0.82 ± 0.02	0.00 ± 0.01
Transfer	Herbarium	Pressed	Normalized	609	14	0.91 ± 0.01	78.48 ± 50.44	0.14 ± 0.09	0.12 ± 0.11	1.47 ± 0.05	0.13 ± 0.16
Transfer	Pressed	Herbarium	CWT	479	8	0.76 ± 0.05	10.53 ± 0.01	0.02 ± 0.00	0.00 ± 0.00	1.25 ± 0.06	−0.02 ± 0.01
Transfer	Pressed	Herbarium	Reflectance	479	16	0.66 ± 0.07	13.13 ± 0.02	0.02 ± 0.00	0.01 ± 0.01	1.13 ± 0.1	0.01 ± 0.00
Transfer	Pressed	Herbarium	Normalized	479	13	0.51 ± 0.09	781.00 ± 242.86	1.18 ± 0.37	−1.18 ± 0.37	0.41 ± 0.09	−0.41 ± 0.06

substantially for *Agonis flexuosa* and the two grasses (*Phalaris arundinacea* and *Phragmites australis*) species, reflecting the limits of model generalizability in these cases (Fig. 4). Discrepancies were especially pronounced where pressed datasets included only a single individual per species. Nonetheless, the lack of unrealistic trait values and the general correspondence of trait distributions across datasets is a positive result for the generalizability of pressed and herbarium models.

These results taken together provide robust support for the utility of herbarium spectra for trait estimation, both for models built

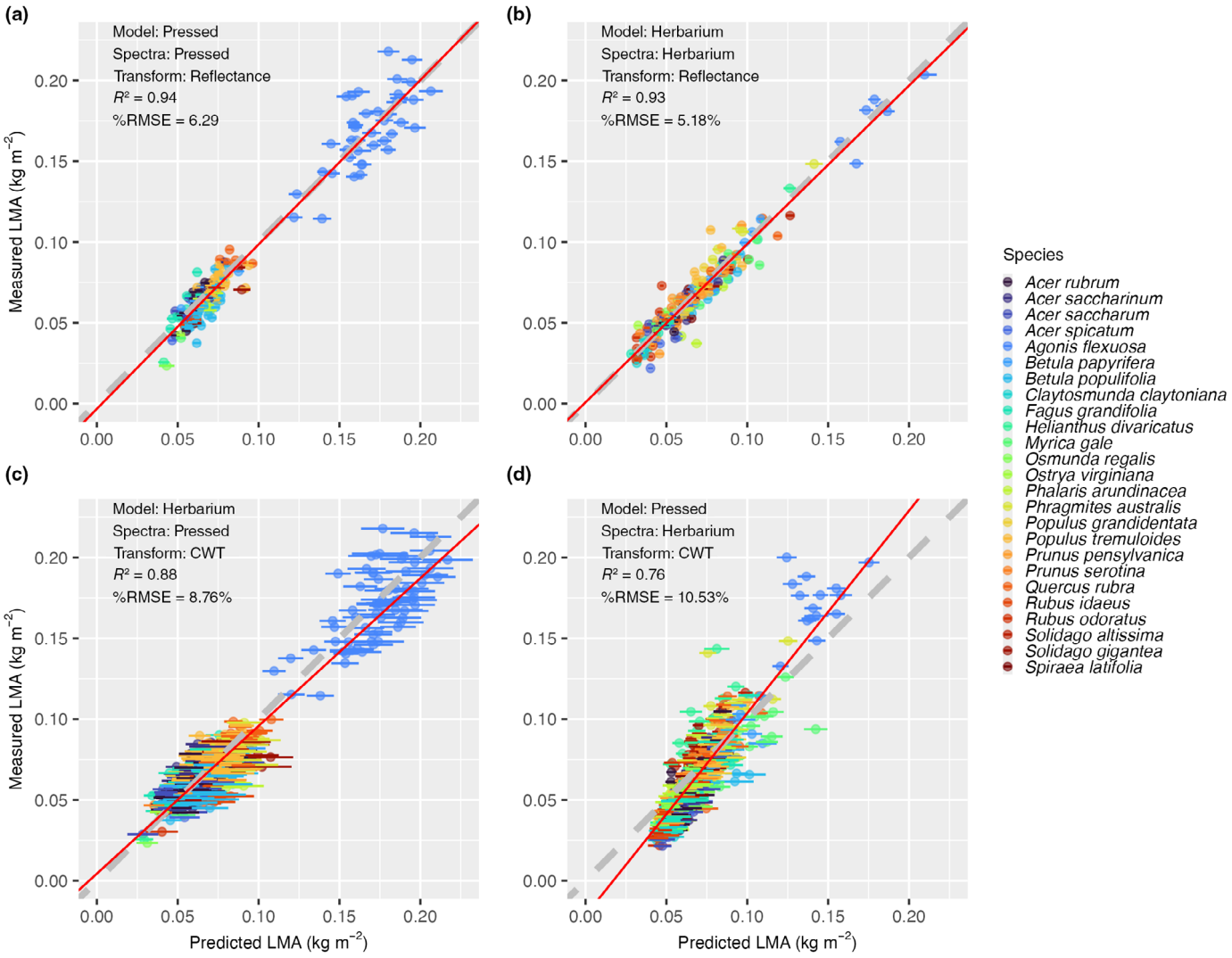

Fig. 3 Validation and model transfer results for leaf mass per area (LMA) per individual across 25 species. Error bars represent the SD in predictions across 1000 model iterations. Linear regressions of observed vs predicted values averaged across iterations are shown in red lines for comparison with the gray 1 : 1 dashed lines. Individual plots show the results for full-range spectra (450–2500 nm) of (a) pressed models from untransformed reflectance values, (b) herbarium models from untransformed reflectance values, (c) transfer of continuous wavelet transformation (CWT) herbarium models to CWT pressed leaf spectra, and (d) transfer of CWT pressed models to CWT herbarium leaf spectra.

from herbarium-derived trait datasets as well as for the transfer of pressed leaf models built from trait values measured in living plants.

Taxonomic discrimination

To evaluate the utility of reflectance spectra for taxonomic discrimination, we applied LDA and PLS-DA models across datasets at two taxonomic levels: species and genus. To ensure direct comparability of results, we also analyzed the pressed-leaf dataset for the 10 species for which 20 or more individuals were sampled (Table 2). Performance metrics, including accuracy, precision, and balanced accuracy, were compared to assess the classification capabilities of each approach.

Pressed datasets outperformed herbarium datasets in classification accuracy, precision, and balanced accuracy, yet herbarium spectra still provided reliable classification models (Table 5). In the 10-species dataset, pressed specimens achieved accuracies of 91.7 ± 2% (LDA) and 81.1 ± 2% (PLS-DA), while herbarium specimens achieved 71.9 ± 2% (LDA) and 58.0 ± 2% (PLS-DA).

For the 25-species dataset, herbarium spectra achieved 74.3 ± 1% accuracy with PLS-DA, outperforming LDA's 64.4 ± 2%. The confusion matrix (Fig. 5) shows that most classification errors occurred between congeneric species, highlighting challenges in distinguishing closely related taxa. Some species, such as *Osmunda regalis* and *Quercus rubra*, were frequently misclassified as *Betula* species. Notably, *Solidago gigantea* had a correct classification rate of only 39%, with 51% of its measurements misclassified as *Solidago altissima*. The VIP plots are consistent across species and emphasize key spectral regions in the visible, near-infrared, and shortwave infrared (SWIR) ranges (Fig. S4).

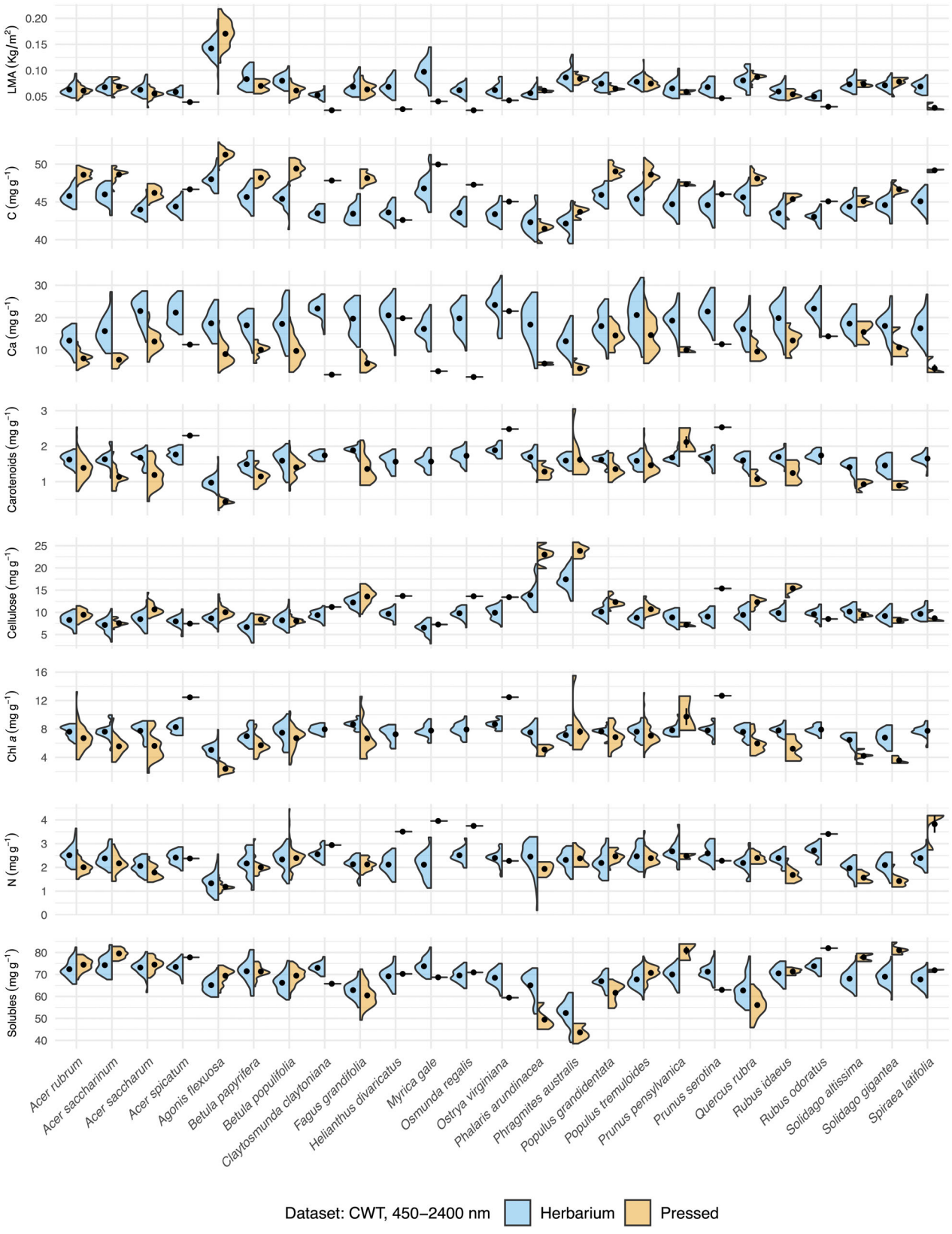

Fig. 4 Comparison of observed trait distributions from pressed leaves with predicted values obtained by applying continuous wavelet transformation (CWT) pressed models to spectra from herbarium leaves. Panels display the distributions for eight traits across 25 species. Mean values are indicated with black dots.

Table 5 Performance metrics of classification analyses.

Dataset	Rank	Model	Classes	*n* components	Accuracy ± SD (%)	Precision ± SD (%)	Balanced accuracy ± SD (%)
Herbarium	Species	LDA	10 spp	N/A	71.9 ± 2	72.1 ± 20	84.2 ± 10
Herbarium	Species	PLS-DA	10 spp	15	58 ± 2	58.5 ± 23	76.5 ± 13
Pressed	Species	LDA	10 spp	N/A	91.7 ± 2	86.6 ± 20	96.3 ± 3
Pressed	Species	PLS-DA	10 spp	15	81.1 ± 2	73.2 ± 22	91.7 ± 6
Herbarium	Genus	LDA	6 genera	N/A	89.3 ± 2	87.8 ± 14	92.8 ± 7
Herbarium	Genus	PLS-DA	6 genera	13	82.1 ± 1	79.5 ± 21	86.8 ± 12
Pressed	Genus	LDA	6 genera	N/A	96.9 ± 1	94.6 ± 10	98.3 ± 2
Pressed	Genus	PLS-DA	6 genera	13	89.8 ± 1	85.2 ± 13	93.9 ± 5
Herbarium	Species	LDA	25 spp	N/A	64.4 ± 2	67.2 ± 19	82 ± 9
Herbarium	Species	PLS-DA	25 spp	24	74.3 ± 1	75.3 ± 15	87.3 ± 8
Herbarium	Genus	LDA	17 genera	N/A	75.3 ± 2	76.5 ± 18	86.4 ± 8
Herbarium	Genus	PLS-DA	17 genera	27	84.9 ± 1	87 ± 8	90.9 ± 9

N/A indicates values are not applicable for this analysis type (LDA).

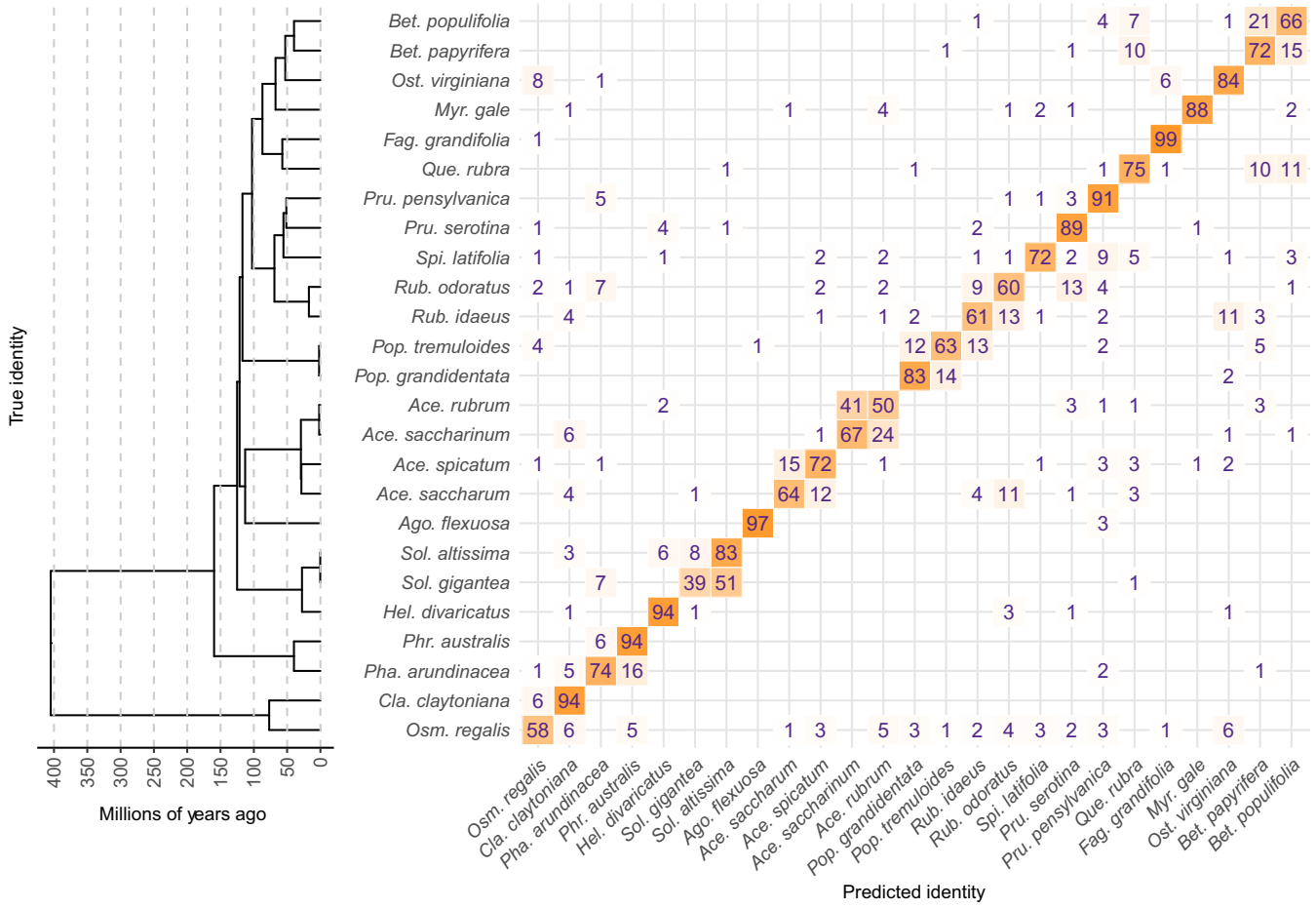

Fig. 5 Phylogram and confusion matrix summarizing the validation of herbarium specimen classification using partial least squares discriminant analysis. The left panel shows a phylogram representing the evolutionary relationships among species, scaled by millions of years. The right panel displays a confusion matrix where rows represent true species identities and columns represent predicted species identities. Tile colors indicate the percentage of observations of each pair of true and predicted identities, with darker shades representing higher percentages. Numbers within tiles show rounded percentages. Mean accuracy for the validation is 74.3%.

In the six-genera dataset, pressed specimens achieved with 96.9 ± 1% (LDA) and 89.8 ± 1% (PLS-DA) accuracy, while herbarium specimens achieved 89.3 ± 2% (LDA) and 82.1 ± 1% (PLS-DA) accuracy. In the more complex 17-genus dataset, herbarium spectra performed better with PLS-DA (84.9 ± 1%) compared to LDA (75.3 ± 2%). Similarly, PLS-DA outperformed

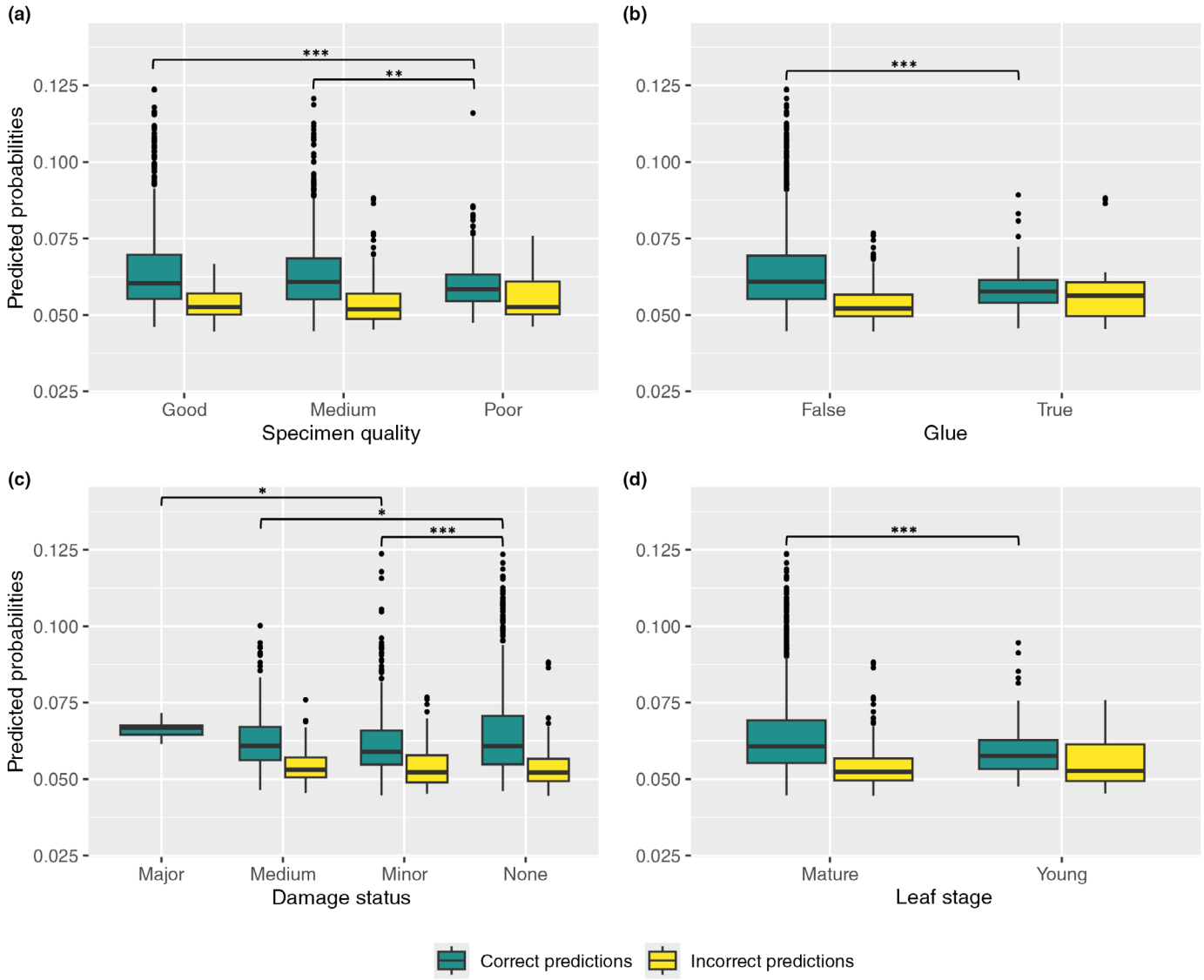

Fig. 6 Comparison of distributions of probabilities of assignment of each measurement to a specific class for correctly (true-positive) or incorrectly classified (false-positive) specimens by leaf characteristics (see Table 3). (a) Specimen quality observations primarily reflecting discoloration or tissue degradation. (b) The presence or absence of mounting glue on the leaf. (c) Visible biotic contamination, pre- or post-collection damage to leaves. (d) Leaf phenological stage. Significant pairwise differences among correct or incorrect classes were determined using *t*-tests and indicated with the codes: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Note that there were no significant differences among classification probabilities for incorrect predictions.

LDA in the 17-genus dataset, with herbarium models achieving 84.9 ± 1% for PLS-DA compared to 75.3 ± 2% for LDA.

The VIP plots comparing herbarium and pressed datasets reveal consistent peaks across the visible, near-infrared, and short-wave infrared (SWIR) regions, reflecting the spectral regions most important for PLS-DA classification (Fig. S5).

Assessing herborization factors on taxonomic discrimination

To evaluate the influence of specimen factors on PLS-DA classification performance, we analyzed the classification probabilities across all 1690 herbarium spectral measurements using the full-spectrum 25-species dataset (Fig. S6). Logistic regression and independent *t*-tests revealed significant relationships between classification probabilities and several categorical and numerical predictor variables.

The probabilities of correct classifications varied significantly with specimen quality, glue presence, leaf damage, and leaf phenological development (Fig. 6). Leaves with good ($P < 0.001$) or medium quality ($P < 0.01$) had higher probabilities for correct classifications than those with poor quality, but there was not a significant difference between good and medium quality specimens. Following expectations, specimens without mounting glue had significantly higher probabilities than those with glue ($P < 0.001$). Mature leaves exhibited higher classification probabilities than young leaves ($P < 0.001$). Probabilities of correct classifications for specimens with no damage were significantly higher than those

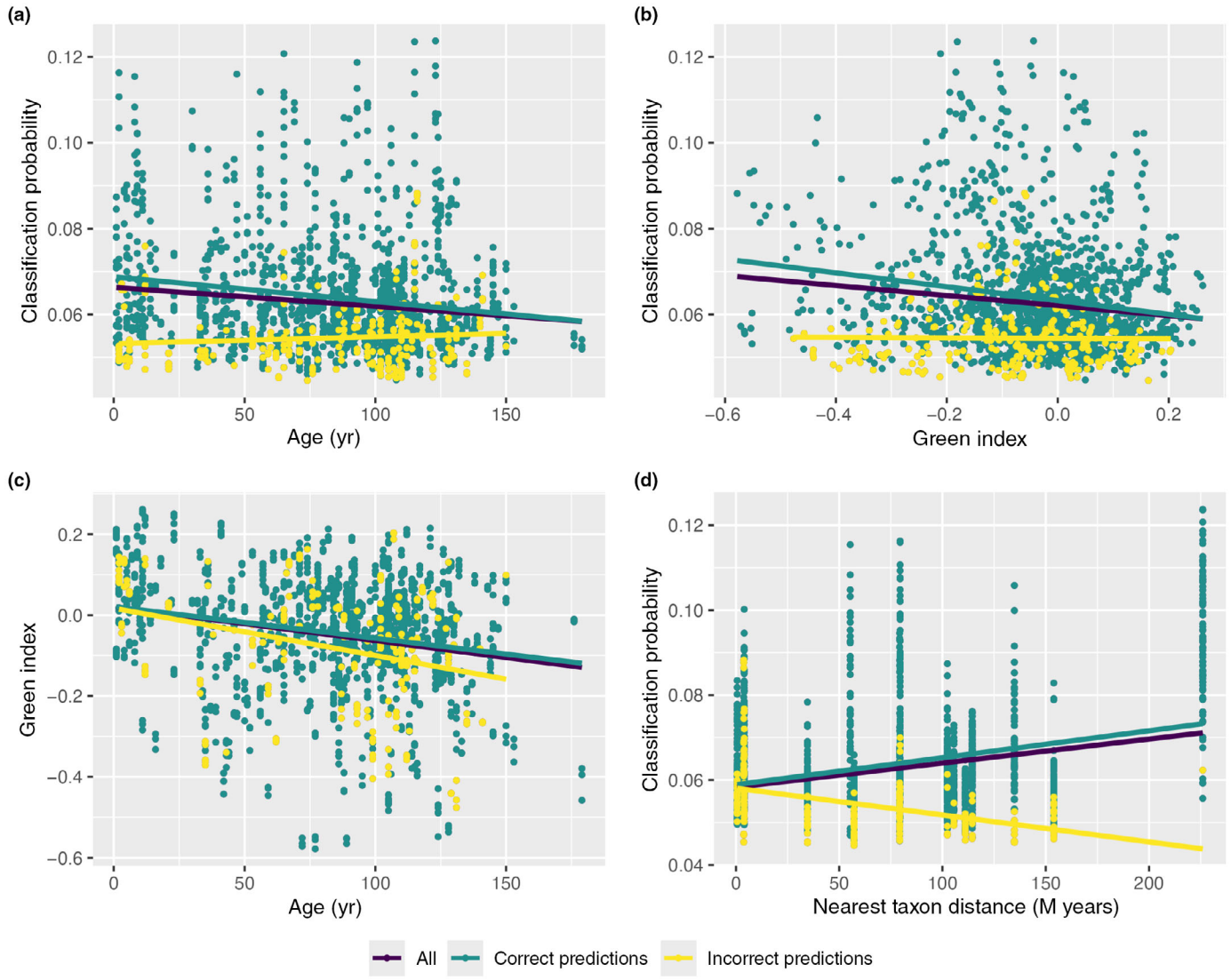

Fig. 7 Relationships between numeric predictor variables and classification outcomes. (a) Relationship between (a) age (years) and classification probability, (b) green index and classification probability, (c) specimen age (years) and green index, and (d) nearest taxon distance (M years) and classification probability. Points represent individual observations colored by correct vs incorrect status. Solid lines represent linear regression fits for each dataset.

with minor damage ($P < 0.001$) and medium damage ($P < 0.05$). Classification probabilities also differed between minor (with the lowest mean probability) and major damage (with the highest mean probability; $P < 0.05$). This is because, contrary to expectations, the two specimens (six measurements) scored with major damage were correctly predicted, and with high classification probabilities. These were *Populus tremuloides* spectra, and this species had a low classification accuracy of 63%. The probabilities of incorrect classifications – which represent false-positive classifications with higher probabilities than true-positives – did not significantly differ across damage classes (Fig. 6).

Numerical predictors also had significant relationships with classification probabilities (Fig. 7). Specimen age was negatively correlated with classification probability, suggesting reduced model performance for older specimens (Fig. 7a). The age of the sampled specimens ranged from 1 to 179 yr with a median age of 91 yr (Fig. S7). The green index was also negatively correlated with classification probabilities, indicating that greener leaves were associated with lower model performance (Fig. 7b). The relationship between age and green index revealed that older specimens generally exhibited lower green index values, consistent with expected tissue degradation over time (Fig. 7c).

Classification probabilities increased with greater phylogenetic distance to the nearest taxon (Fig. 7d), an expected relationship that corroborates the results of the confusion matrix. Conversely, the probability of a false positive classification decays with phylogenetic distance to the predicted class (Fig. S8). Leaf mass per area also shows a strong positive correlation with classification probability (Fig. S9), with the caveat of covariation with species composition. *Agonis flexuosa* was classified with an overall

Table 6 Logarithmic regression of all predictors.

	Estimate	SE	z Value	Pr(> |z|)	Sig.
(Intercept)	1.18E+01	3.60E+02	3.29E-02	9.74E-01	
Nearest Taxon Distance	8.15E-03	1.69E-03	4.83E+00	1.35E-06	***
Age	1.05E-02	2.32E-03	4.55E+00	5.43E-06	***
Glue: present	−9.19E-01	2.14E-01	−4.30E+00	1.72E-05	***
Green index	2.29E+00	6.06E-01	3.78E+00	1.54E-04	***
Leaf kg m^{-2}	1.30E+01	4.40E+00	2.97E+00	3.02E-03	***
Quality: medium	−5.38E-01	1.94E-01	−2.78E+00	5.45E-03	***
Quality: poor	−7.35E-01	2.93E-01	−2.51E+00	1.21E-02	**
Julian day	4.38E-03	2.50E-03	1.75E+00	7.93E-02	.
Leaf stage: young	−2.85E-01	2.68E-01	−1.07E+00	2.87E-01	
Damage: medium	−1.26E+01	3.60E+02	−3.51E-02	9.72E-01	
Damage: minor	−1.25E+01	3.60E+02	−3.47E-02	9.72E-01	
Damage: none	−1.24E+01	3.60E+02	−3.44E-02	9.73E-01	

Significance values are indicated with the codes: ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; the '.' denotes marginal significance ($P < 0.1$) and blank values indicate non-significant correlations ($P \geq 0.1$).

accuracy of 97%, and LMA values for this species are much higher than other species in the dataset.

Logistic regression taking into account phylogeny (Table 6) further supported these factors as important in classification success. As expected, the most influential metric in classification success is nearest taxon distance, but the next most significant predictors were age, green index, absence of glue, and specimen quality. Finally, there is a weak positive relationship between calendar day of specimen collection and classification success (Table 6) or classification probability (Fig. S10). This relationship indicates that species collected early in the growing season are somewhat more likely to be misclassified than those collected at later dates. Random forest models generally corroborated these results, but optimized LMA, age, and the green index as more significant factors than nearest taxon distance (Table S3).

These results highlight the critical influence of specimen metadata on PLS-DA classification performance. Factors such as tissue quality, as measured by the green index, and phylogenetic distinctiveness strongly impact classification success. By contrast, older specimens, poor-quality leaves, and the presence of glue reduce classification probabilities, underscoring the importance of these metadata for optimizing model performance.

Discussion

As the largest scientific repositories of plant diversity, herbaria offer exceptional resources for investigations of plant biology, but their suitability for reflectance spectroscopy-based inferences remains largely unknown. The wide variety of collection and processing methods, as well as specimen age and storage, differentiate herbarium plant tissues from freshly collected plant tissues, leading to uncertainties in their relevance for plant trait prediction and taxonomic classification. A positive outlook has come from recent investigations of pressed leaves on the order of months to years old (i.e. collected, pressed, dried, stored in newspaper), which have demonstrated the robust application of spectra for both applications (Durgante *et al.*, 2013; Lang *et al.*, 2017; Costa *et al.*, 2018; Kothari *et al.*, 2023b; Hernández-Leal *et al.*, 2025). Our study has extended this discovery, clearly demonstrating that herbarium specimens retain enough morphological and anatomical integrity to be useful for these same spectra-based inferences. Here, we outline the insights from this study in the context of promises and challenges for reflectance spectroscopy of herbarium specimens.

Trait prediction

Leaf mass per area is consistently one of the most accurately modeled traits across studies (SLA of Costa *et al.*, 2018; Serbin *et al.*, 2019; Kothari *et al.*, 2023a,b) and is a key indicator of plant resource-use strategies within the leaf economics spectrum (Wright *et al.*, 2004; Díaz *et al.*, 2016). Overall, the herbarium LMA models performed nearly as well as the pressed leaf models. Among herbarium models, those based on normalized spectra performed slightly better than those based on untransformed reflectance (normalized, full-range $R^2 = 0.94$; %RMSE = 4.86% vs reflectance $R^2 = 0.93$; %RMSE = 5.18%), suggesting that variation in measured spectral magnitudes may not be useful. Continuous wavelet transformation showed similar performance ($R^2 = 0.93$; %RMSE = 5.31%), indicating that preserving the overall shape and relative magnitudes of reflectance spectra is important for trait prediction. At the same time, CWT, normalized, and untransformed reflectance spectra showed nearly identical predictive performance in the pressed dataset.

Improving the generalizability of models is a critical step toward global-scale trait modeling across temporal scales (Serbin *et al.*, 2019; Kothari *et al.*, 2023a; Ji *et al.*, 2024). While our models demonstrated promising transferability between pressed and herbarium specimens, their performance varied depending on spectral preprocessing. The CWT-transformed models showed the best overall performance statistics among transfer tests, and herbarium models transferred to pressed spectra worked better than the reverse transfer. This pattern may be a consequence of the broader spectral variability in the herbarium dataset. Although our experiment focused on comparing model transferability between pressed

and herbarium spectra, future work should explore the benefits of training models on combined datasets. Traits like LMA appear more amenable to general modeling (Serbin *et al.*, 2019; Kothari *et al.*, 2023a), but other traits may require more tailored, taxon-specific approaches. As our predictions for additional traits showed (Fig. 4), taxonomic context matters, and herbarium collections offer a valuable platform for testing model behavior across phylogenetic and geographic gradients.

Our herbarium-derived models are likely to perform well for predicting LMA in both pressed and herbarium leaves from the same genera and within the temperate broadleaf and mixed forests of North America. They may also generalize to other taxa with LMA values that fall within the modeled range (0.025–-0.18 kg m^{-2}). However, extending these models to new regions and taxa will require further validation.

In this context, the inclusion of the Australian species *Agonis flexuosa*, which exhibits unusually high LMA values, illustrates the importance of balanced trait sampling for effective PLSR model training. When *Agonis* was excluded from the pressed dataset, model performance decreased (R^2 = 0.69, %RMSE = 11.92%; Fig. S11A; Table S4). This can be attributed to the smaller spread of trait values in relation to residuals in the pressed data, but also due to less training data on LMA values (Fig. S12). This is evidenced by the reduced performance of the transfer test of the pressed-leaf model to the herbarium spectra, especially at higher LMA values (R^2 = 0.60, %RMSE = 20.18%; Fig. S11D; Table S4). However, when we excluded *Agonis* from the herbarium dataset, model performance remained similar (R^2 = 0.90, % RMSE = 8.83%; Fig. S11B; Table S4).

These findings support a general strategy for herbarium-based trait modeling: build models using taxonomically and geographically diverse training data with balanced representation of trait values (Burnett *et al.*, 2021). Following the strategy of Kothari *et al.* (2023b), we partitioned our data as a 70/30 split into training and validation datasets subset by growth form (Fig. S13), but a trait-stratified proportional or other method to ensure balanced trait representation in data splitting and cross validation steps may lead to even better model performance (Joseph & Vakayil, 2022).

A key challenge in advancing herbarium-based trait modeling is that model construction and validation will require some amount of destructive sampling. Estimating traits such as nitrogen, carbon, and carbon fractions can require substantial amounts of material – up to 500 mg of dry leaf tissue (Schweiger *et al.*, 2018; Kothari *et al.*, 2024). To mitigate specimen loss, researchers should prioritize sampling from unmounted duplicates or bulk collections, with the goal of maximizing trait variation while achieving broad, balanced representation across major clades and preservation conditions.

Pressed leaves represent a critical resource in this context. They offer access to relatively well-preserved tissue with known preservation histories, making them ideal for model development and for studying trait degradation over time. Our results show that models trained on pressed leaves have the potential to accurately model traits of herbarium leaf tissues, providing a link to trait values as they may have existed *in vivo*. This not only improves our confidence in trait predictions, but also reduces the need for further destructive sampling of irreplaceable collections. Integrating pressed leaves into trait modeling pipelines will strengthen the foundation for scaling spectral trait prediction across global herbaria.

Together, these findings highlight both the potential and limitations of herbarium specimens for trait modeling. While traits like LMA can be predicted with high accuracy, extending this success to other traits and taxa will require strategic sampling for continued refinement of models. Building generalizable models across the tree of life will depend on thoughtful integration of specimen conditions with the phylogenetic and environmental components of phenotypic variability.

Taxonomic discrimination

Our results show that herbarium-based taxonomic discrimination models perform reasonably well, but with lower accuracy than their pressed-leaf counterparts. This is likely due to better tissue integrity and fewer preservation artifacts affecting spectral information in pressed leaves. LDA models tended to outperform PLS-DA models in cases with fewer classes, while PLS-DA performed better in the 17-genus and 25-species herbarium datasets. Across both PLS-DA and LDA analyses, misclassifications occurred most frequently between closely related species, such as *Acer*, *Betula*, and *Solidago*, reflecting underlying phenotypic and biochemical similarity. Notably, *Solidago altissima* was more often classified as its congener, a finding consistent with the positive correlation between classification probability and nearest taxon distance (Fig. 7). This suggests that spectral discrimination becomes more difficult among closely related taxa, where spectral features are more similar; a pattern that has been found in fresh leaf spectra (Schweiger *et al.*, 2018).

A major challenge in spectral classification lies in the relationship between model complexity and performance. As shown in Table 5, models with fewer species classes achieved higher classification accuracy, while accuracy generally declined as the number of species included in the model increased – a well-documented limitation of discriminant analysis approaches (Meireles *et al.*, 2020b). Spectral resolution is another important consideration, as our method of down sampling and smoothing spectra to 5 nm intervals could have reduced classification accuracy due to the loss of small spectral features. However, higher spectral resolution would also introduce interpolation issues and complicate cross-instrument data integration. Similarly, increasing the number of replicate measurements per leaf may enhance model robustness. Previous studies (e.g. Durgante *et al.*, 2013) have focused on averaging multiple spectral measurements, which differs from our iterative approach that retains information at the level of individual measurements.

Beyond accuracy and performance, a fundamental limitation of discriminant and supervised classification models is that they cannot identify unknown taxa outside of the trained species pool. Existing ordination approaches based on reflectance data are often too noisy for reliable clustering or taxonomic inference, but principal components analysis of FTIR spectra has been successfully used to resolve taxa (Damasco *et al.*, 2019). In addition, other

methods could be useful for dimensionality reduction and exploration of taxon clustering (e.g. UMAP, t-SNE). A promising future direction is the development of probabilistic classification frameworks capable of flagging outliers or uncertain specimens. Another alternative is to predict traits from individual spectra and explore phenotypic clustering in multidimensional trait space (Schweiger *et al.*, 2021). This strategy could reveal natural groupings based on shared ecological function, even when traditional taxonomic resolution is elusive, and provides a complementary framework for leveraging spectral data to uncover structure within herbarium collections (Hernández-Leal *et al.*, 2025).

The effects of herborization on spectral inferences

The herborization process encompasses the collection, processing steps, and time-sensitive effects of storage, and presents a wide range of variables that influence the spectral properties of plant tissues. Our analyses here indicate that most of the expected effects of herborization and aging of plant tissues negatively affect the classification probabilities and performance of discriminant models.

We assessed specimen preservation conditions using visual indicators of specimen quality, such as discoloration, wilting, pathogen presence, and signs of poor initial drying, as well as evidence of physical damage (e.g. herbivory, tearing, or burning). Specimens categorized as medium or poor quality were significantly associated with lower classification probabilities (Fig. 6) and reduced classification accuracy (Tables 6, S3), confirming that visual degradation correlates with diminished model performance. Logistic regression analyses further supported this pattern, identifying specimen quality, glue presence, and low greenness index values as significant predictors of reduced classification success (Table 6). These findings were reinforced by random forest analyses, which ranked LMA, specimen age, greenness index, and nearest taxon distance as the most important predictors of model performance. Specimen quality and glue presence were also influential, albeit to a lesser degree. These results collectively highlight the critical role of both biological traits and preservation history in classification success using spectral data from herbarium specimens.

While specimen age and greenness are intuitively expected to correlate – since younger specimens often appear greener – the relationship between these variables and spectral performance is more complex. Past studies in DNA sequencing suggest that age alone is a poor predictor of preservation quality (Erkens *et al.*, 2008; Brewer *et al.*, 2019; Forrest *et al.*, 2019; White *et al.*, 2021), and our findings echo this. Instead, specimen processing methods during the early stages of preservation – namely, how quickly and efficiently the specimen was dried, as well as the stability of long-term storage conditions – may play a more important role in long-term tissue integrity than age. These will be important factors to discern in future studies.

Greenness, driven largely by residual Chl, strongly affects spectral signatures in the visible range. Although green tissues may indicate good preservation, high Chl content can also obscure informative spectral features. Conversely, its absence – as seen in older or less green leaves – may enhance the visibility of structural and chemical features that are useful for classification or trait modeling (Kothari *et al.*, 2023b). Thus, while greenness remains a useful preservation indicator, its influence on spectral quality is objective-dependent and nonlinear.

Preservation variables are not fully independent, and their combined effects can be complex. Differences among herbaria related to specimen treatment, mounting practices, and storage conditions, such as relative humidity, are further expected to generate variation among spectral datasets. Standardized metadata and mounting practices, such as using herbarium mounting tape instead of glue, are likely to be important in minimizing these effects.

Finally, the assessment of specimen quality and damage involves some degree of subjectivity. Even identifying glue residues can be nuanced. As herbarium digitization scales up, training technicians to score these factors consistently will be vital for ensuring data quality and interoperability across institutions.

Seeing herbaria in a new light

As herbaria face mounting vulnerabilities – from chronic underfunding to institutional threats of closure – the need to unlock new scientific value from these collections has never been greater (Davis, 2024; Thiers, 2024). The results of this study underscore the promise of reflectance spectroscopy as a powerful, scalable tool for extracting functional and taxonomic information from preserved plant specimens. As part of the growing field of spectral biology (Cavender-Bares *et al.*, 2025), this approach offers not only a new lens on plant diversity but also the opportunity to better understand how specimen processing and preservation influence data quality. Given the high sensitivity of spectral instruments to both biological and technical variation, reflectance spectroscopy is uniquely positioned to help illuminate the effects of herborization and even inform best practices for specimen care and long-term preservation.

While trait prediction and species classification remain foundational applications, the integration of spectral data with genomic, morphological, and spatial datasets will enable deeper insights into species delimitation, phenotypic evolution, community assembly, and biogeography. These opportunities are particularly compelling when viewed through the lens of the Global Metaherbarium – a growing digital infrastructure that connects specimen metadata, images, and extended datasets (Hedrick *et al.*, 2020; Davis, 2023). As this field advances, reflectance spectroscopy will continue to reveal new dimensions of plant diversity, transforming how we study, use, and preserve the world's herbarium collections.

Acknowledgements

We thank the curatorial and digitization staff at the Harvard University Herbaria for past and present efforts to curate and digitize the Herbaria, with special thanks to Michaela Schmull and Lisa Standley for project support and protocol development. This project was funded by the NSF BII DBI-2021898 (JCB and

JEM), NSF DEB-2442433 (JEM), the Harvard University Herbaria Postdoctoral Research Fellowship (DMW), and an NSF REU 2150058.

Competing interests

None declared.

Author contributions

JC-B, CCD, JAGQ, SK, JEM and DMW conceptualized the project. JC-B, JAGQ, JEM and DMW developed the methodology. JAGQ, JMR and DMW curated the data. JMR and DMW conducted the investigation. DMW performed the formal analysis. JC-B and DMW secured funding and managed the project. JC-B and SK provided resources. JAGQ, JEM and DMW developed the software. JC-B, JEM and DMW supervised the project. JAGQ and DMW validated the results. JAGQ, JEM and DMW created visualizations. JEM and DMW wrote the original draft, and all authors reviewed and edited the final version of the manuscript.

ORCID

Jeannine Cavender-Bares https://orcid.org/0000-0003-3375-9630
Charles C. Davis https://orcid.org/0000-0001-8747-1101
J. Antonio Guzmán Q. https://orcid.org/0000-0002-0721-148X
Shan Kothari https://orcid.org/0000-0001-9445-5548
José Eduardo Meireles https://orcid.org/0000-0002-2267-6074
Jorge M. Robles https://orcid.org/0009-0008-7742-2419
Dawson M. White https://orcid.org/0000-0002-0670-9390

Data availability

All analysis codes used in this study are publicly available at GitHub (github.org/Erythroxylum/herbarium-spectra). The spectral and metadata files generated and analyzed in this study have been deposited in the Plant Diversity and Ecology collection of the Harvard Dataverse repository and are publicly available at doi: 10.7910/DVN/LXPHBC.

References

Bieker VC, Sánchez Barreiro F, Rasmussen JA, Brunier M, UK N, Martin MD. 2020. Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Molecular Ecology Resources* **20**: 1206–1219.

Brewer GE, Clarkson JJ, Maurin O, Zuntini AR, Barber V, Bellot S, Biggs N, Cowan RS, Davies NMJ, Dodsworth S *et al.* 2019. Factors affecting targeted sequencing of 353 nuclear genes from Herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* **10**: 1102.

Burnett AC, Anderson J, Davidson KJ, Ely KS, Lamour J, Li Q, Morrison BD, Yang D, Rogers A, Serbin SP. 2021. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *Journal of Experimental Botany* **72**: 6175–6189.

Carrascal LM, Galván I, Gordo O. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **118**: 681–690.

Cavender-Bares J, Gamon JA, Hobbie SE, Madritch MD, Meireles JE, Schweiger AK, Townsend PA. 2017. Harnessing plant spectra to integrate the biodiversity sciences across biological and spatial scales. *American Journal of Botany* **104**: 966–969.

Cavender-Bares J, Meireles JE, Pinto-Ledezma J, Reich PB, Schuman M, Townsend PA, Trowbridge A. 2025. Spectral biology across scales in changing environments. *Ecology* **105**: e70078.

Chamberlain SA, Boettiger C. 2017. *R Python, and Ruby clients for GBIF species occurrence data*. doi: 10.7287/peerj.preprints.3304v1.

Chamberlain S, Oldoni D, Barve V, Desmet P, Geffert L, Mcglinn D, Ram K, Waller J. 2024. *RGBIF: interface to the global biodiversity information facility API*. R package v.3.81. [WWW document] URL https://CRAN.R-project.org/package=rgbif.

Costa FR, Lang C, Almeida DR, Castilho CV, Poorter L. 2018. Near-infrared spectrometry allows fast and extensive predictions of functional traits from dry leaves and branches. *Ecological Applications* **28**: 1157–1167.

Damasco G, Daly DC, Vicentini A, Fine PVA. 2019. Reestablishment of *Protium cordatum* (Burseraceae) based on integrative taxonomy. *Taxon* **68**: 34–46.

Davis CC. 2023. The herbarium of the future. *Trends in Ecology & Evolution* **38**: 412–423.

Davis CC. 2024. Collections are truly priceless. *Science* **383**: 1035.

Díaz S, Kattge J, Cornelissen JHC, Wright IJ, Lavorel S, Dray S, Reu B, Kleyer M, Wirth C, Colin Prentice I *et al.* 2016. The global spectrum of plant form and function. *Nature* **529**: 167–171.

Durgante FM, Higuchi N, Almeida A, Vicentini A. 2013. Species spectral signature: discriminating closely related plant species in the Amazon with near-infrared leaf-spectroscopy. *Forest Ecology and Management* **291**: 240–248.

Erkens RHJ, Cross H, Maas JW, Hoenselaar K, Chatrou LW. 2008. Assessment of age and greenness of herbarium specimens as predictors for successful extraction and amplification of DNA. *Blumea – Biodiversity, Evolution and Biogeography of Plants* **53**: 407–428.

Forman L, Bridson D, eds. 1989. *The herbarium handbook*. Kew, UK: Royal Botanic Gardens.

Forrest LL, Hart ML, Hughes M, Wilson HP, Chung K-F, Tseng Y-H, Kidner CA. 2019. The limits of Hyb-Seq for herbarium specimens: Impact of preservation techniques. *Frontiers in Ecology and Evolution* **7**: 439.

Fourty T, Baret F, Jacquemoud S, Schmuck G, Verdebout J. 1996. Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote Sensing of Environment* **56**: 104–117.

Geladi P, Kowalski BR. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* **185**: 1–17.

Guzmán JA. 2024. CWT: continuous wavelet transformation for spectroscopy. R package v.0.2.1. [WWW document] URL https://CRAN.R-project.org/package=CWT.

Guzmán JA, Sanchez-Azofeifa GA. 2021. Prediction of leaf traits of lianas and trees via the integration of wavelet spectra in the visible-near infrared and thermal infrared domains. *Remote Sensing of Environment* **259**: 112406.

Heberling JM. 2022. Herbaria as big data sources of plant traits. *International Journal of Plant Sciences* **183**: 87–118.

Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA, Cook JA, Blackburn DC *et al.* 2020. Digitization and the future of natural history collections. *Bioscience* **70**: 243–251.

Hernández-Leal MS, Guzmán JA, Rodríguez AG, Cavender-Bares J. 2025. Foliar spectral signatures reveal adaptive divergence in live oaks (*Quercus* section Virentes) across species and environmental niches. EcoEvoRxiv. doi: 10.32942/X2BK90.

Jetz W, Cavender-Bares J, Pavlick R, Schimel D, Davis FW, Asner GP, Guralnick R, Kattge J, Latimer AM, Moorcroft P *et al.* 2016. Monitoring plant functional diversity from space. *Nature Plants* **2**: 1–5.

Ji F, Li F, Hao D, Shiklomanov AN, Yang X, Townsend PA, Dashti H, Nakaji T, Kovach KR, Liu H *et al.* 2024. Unveiling the transferability of PLSR models for leaf trait estimation: lessons from a comprehensive analysis with a novel global dataset. *New Phytologist* **243**: 111–131.

Jin Y, Qian H. 2022. V.PhyloMaker2: an updated and enlarged R package that can generate very large phylogenies for vascular plants. *Plant Diversity* **44**: 335–339.

Joseph VR, Vakayil A. 2022. SPlit: an optimal method for data splitting. *Technometrics* **64**: 166–176.

Kothari S, Beauchamp-Rioux R, Blanchard F, Crofts AL, Girard A, Guilbeault-Mayers X, Hacker PW, Pardo J, Schweiger AK, Demers-Thibeault S *et al.* 2023a. Predicting leaf traits across functional groups using reflectance spectroscopy. *New Phytologist* **238**: 549–566.

Kothari S, Beauchamp-Rioux R, Laliberté E, Cavender-Bares J. 2022. Pressed-leaf CABO spectra from herbarium project. Dataset. Ecological Spectral Information System (EcoSIS). doi: 10.21232/KS7MbtCK.

Kothari S, Beauchamp-Rioux R, Laliberté E, Cavender-Bares J. 2023b. Reflectance spectroscopy allows rapid, accurate and non-destructive estimates of functional traits from pressed leaves. *Methods in Ecology and Evolution* **14**: 385–401.

Kothari S, Hobbie SE, Cavender-Bares J. 2024. Rapid estimates of leaf litter chemistry using reflectance spectroscopy. *Canadian Journal of Forest Research* **54**: 978–991.

Kothari S, Schweiger AK. 2022. Plant spectra as integrative measures of plant phenotypes. *Journal of Ecology* **110**: 2536–2554.

Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* **28**: 1–26.

Kühn P, Proß T, Römermann C, Wesche K, Bruelheide H. 2024. Using near-infrared spectroscopy to predict nitrogen and phosphorus concentrations of herbarium specimens under different storage conditions. *Plant Methods* **20**: 19.

Kühn P, Umazekabiri R, Römermann C, Bruelheide H, Wesche K. 2025. Nitrogen content of herbarium specimens from arable fields and mesic meadows reflect the intensifying agricultural management during the 20th century. *Journal of Ecology* **113**: 555–569.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution* **39**: msac174.

Lang C, Almeida DRA, Costa FRC. 2017. Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). *Forest Ecology and Management* **406**: 219–227.

Liland KH, Mevik B-H, Wehrens R, Hiemstra P. 2024. *pls: partial least squares and principal component regression.* R package v.2.8-5. [WWW document] URL http://CRAN.R-project/package=pls.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* **2**: 18–22.

Mehmood T, Liland KH, Snipen L, Sæbø S. 2012. A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **118**: 62–69.

Meireles JE, Cavender-Bares J, Townsend PA, Ustin S, Gamon JA, Schweiger AK, Schaepman ME, Asner GP, Martin RE, Singh A. 2020a. Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytologist* **228**: 485–493.

Meireles JE, O'Meara B, Cavender-Bares J. 2020b. Linking leaf spectra to the plant tree of life. In: Cavender-Bares J, Gamon JA, Townsend PA, eds. *Remote sensing of plant biodiversity.* Cham, Switzerland: Springer International Publishing, 155–172.

Meireles JE, Schweiger A, Cavender-Bares J. 2017. *spectrolab: class and methods for hyperspectral data.* R package v.0.02. doi: 10.5281/zenodo.3934575.

Neto-Bradley BM, Bonnet P, Goeau H, Joly A, Cavender-Bares J, Coomes DA. 2025. Using reflectance spectra and Pl@ntNet to identify herbarium specimens: a case study with Lithocarpus. *New Phytologist.*

R Core Team. 2023. *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rivard B, Feng J, Gallie A, Sanchez-Azofeifa A. 2008. Continuous wavelets for the improved use of spectral libraries and hyperspectral data. *Remote Sensing of Environment* **112**: 2850–2862.

Schweiger AK, Cavender-Bares J, Kothari S, Townsend PA, Madritch MD, Grossman JJ, Gholizadeh H, Wang R, Gamon JA. 2021. Coupling spectral and resource-use complementarity in experimental grassland and forest communities. *Proceedings of the Royal Society B: Biological Sciences* **288**: 20211290.

Schweiger AK, Cavender-Bares J, Townsend PA, Hobbie SE, Madritch MD, Wang R, Tilman D, Gamon JA. 2018. Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nature Ecology & Evolution* **2**: 976–982.

Serbin SP, Singh A, McNeil BE, Kingdon CC, Townsend PA. 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecological Applications* **24**: 1651–1669.

Serbin SP, Townsend PA. 2020. Scaling functional traits from leaves to canopies. In: Cavender-Bares J, Gamon JA, Townsend PA, eds. *Remote sensing of plant biodiversity.* Cham, Switzerland: Springer International, 43–82.

Serbin SP, Wu J, Ely KS, Kruger EL, Townsend PA, Meng R, Wolfe BT, Chlus A, Wang Z, Rogers A. 2019. From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. *New Phytologist* **224**: 1557–1568.

Thiers BM. continuously updated. *Index Herbariorum: a global directory of public herbaria and associated staff.* New York Botanical Garden's virtual herbarium. [WWW document] URL https://sweetgum.nybg.org/science/ih/.

Thiers BM. 2024. Strengthening partnerships to safeguard the future of Herbaria. *Diversity* **16**: 36.

White DM, Huang J-P, Jara-Munoz OA, Madriñán S, Ree RH, Mason-Gamer RJ. 2021. The origins of coca: museum genomics reveals multiple independent domestications from progenitor *Erythroxylum gracilipes*. *Systematic Biology* **70**: 1–13.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis.* New York: Springer-Verlag New York.

Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JHC, Diemer M *et al.* 2004. The worldwide leaf economics spectrum. *Nature* **428**: 821–827.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Comparison of raw reflectance spectra across the full spectral range (350–2500 nm).

Fig. S2 Green index compared to specimen quality observation.

Fig. S3 Spectral means, variability, and CV per species.

Fig. S4 VIP herbarium PLS-DA 25 species.

Fig. S5 VIP pressed vs herbarium.

Fig. S6 Confusion matrix from coefficient-based predictions.

Fig. S7 Distribution of absolute age of specimens.

Fig. S8 Regression of phylogenetic distance of predicted class against probability of classification for misclassifications.

Fig. S9 Regression of LMA against classification probabilities.

Fig. S10 Linear and polynomial regressions of collection Julian day against classification probabilities.

Fig. S11 Herbarium validation results for leaf mass per area (LMA) without *Agonis flexuosa*.

Fig. S12 Distribution of leaf mass per area (LMA) values across training and validation datasets input in PLSR models.

Fig. S13 Distribution of leaf mass per area (LMA) values across growth forms in the herbarium and pressed leaf datasets.

Table S1 Performance metrics for all LMA trait models (full, VNIR, and SWIR).

Table S2 Validation results for predicting eight traits across three pressed leaf models (full-range reflectance spectra, full-range CWT spectra, and 1350–2500 nm vector-normalized spectra).

Table S3 Random Forest model of variable importance of all predictors.

Table S4 Performance metrics for the herbarium reflectance LMA trait model (full, VNIR, and SWIR) built without *Agonis flexuosa*.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.