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

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

Authors for correspondence: Dawson M. White, dawson.white@gmail.com; Jeannine Cavender-Bares, jcavender@fas.harvard.edu

Affiliations:
[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, 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.

Summary:

- Reflectance spectroscopy is a non-destructive, rapid, and robust method for estimating functional traits and distinguishing species. Spectral reflectance libraries generated from herbarium specimens are an untapped and promising resource for generating broad phenomic datasets across space, time, and species.
- We conducted a proof-of-concept study using functional trait data and spectra from recently dried, pressed leaves, alongside data from herbarium specimens up to 179 years old. We assessed the utility and transferability of these datasets for functional trait prediction and taxonomic discrimination.
- Herbarium spectra discriminated species with 74% accuracy and predicted leaf mass per area (LMA) with R^2=0.92 and %RMSE=5.8%. 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.
- The results demonstrate the feasibility of using herbarium spectral data for functional trait prediction and taxonomic discrimination. This success provides methodological guidance for advancing the global Metaherbarium and integrating spectral reflectance into next-generation digitization efforts for plant biodiversity collections.

Plain language summary: Reflectance spectroscopy applied to herbarium collections offers a transformative method to generate phenomic data across the plant tree of life. Despite preservation challenges, we demonstrate its reliability in predicting functional traits and facilitating taxonomic discrimination in specimens with a median age of 91 years. We further provide guidelines for researchers and collections managers to collect and scale spectra effectively.

Introduction

The vision of herbarium spectral scanning

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, In Review.; Meireles *et al.*, 2020). This powerful approach provides a rapid and non-destructive method to characterize leaf functional traits and identify taxa through spectral signals, integrating structural, chemical, and physiological information from plants in various contexts, including the lab, herbarium, and field (Box 1; Costa *et al.*, 2018; Serbin & Townsend, 2020; Kothari & Schweiger, 2022). Despite its potential, the spectral-based taxonomic and functional 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.*, 2020), 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 undermines geographic and temporal precision and compromises reliability as inevitable taxonomic and nomenclatural changes occur.

A promising path forward for filling these gaps across the plant tree of life lies in leveraging the approximately 400 million dried plant specimens stored in over 3,500 herbaria worldwide (Thiers, 2020; Heberling, 2022; Kothari *et al.*, 2023). Herbarium collections are the physical foundation of our scientific understanding of plant and fungal diversity and anchor every species definition. Herbarium collections also include specimens that are rare, extinct, or regionally extirpated. By grounding plant taxonomy and spatial data with diverse extended datasets, the wealth of global herbarium specimens have long been a key resource for researchers studying plant diversity and ecological and evolutionary processes across spatial and temporal scales (Davis, 2023).

The integration of reflectance spectroscopy into herbarium digitization pipelines, with appropriate and standardized modifications, would result in the rapid generation of spectral data directly linked to physical specimens. This would create a new dimension of phenomic data integrated with other extended specimen datasets and tremendously improve capacities for taxonomic, ecological, and evolutionary investigations. Integrating spectral data with herbarium specimens would not only expand trait datasets but also deepen our understanding of plant functional diversity across space and time.

However, until now, the promise of this innovative use of herbarium reflectance spectroscopy is tempered by the reality that we have yet to empirically prove the utility of herbarium specimens, given the potential for tissue degradation with mounting, preservation, and storage. Spectral scans with modern

spectroradiometers (350–2,500 nm) are highly sensitive to all physical aspects of scanned materials and require rigorous standardized protocols in the field and laboratory to ensure high data quality and aggregation from different sources and sensors. Herbarium specimens present unique challenges given the extra variables beyond the normal biological variation that has affected their tissues (Fig. 2).

77

Box 1: Learning from spectra

Reflectance spectroscopy measures light reflected from a material — typically from the visible, near-infrared, and short-wave infrared wavelengths (350–2,500 nm) — across hundreds of specific wavelength bands (Fig. 1). These scans reveal absorption features associated with the structural and chemical properties of the target material, such as leaf tissue (Curran, 1989). Spectral scans are fast (two to three seconds), require minimal digital resources (30–40 KB per file), and enable powerful predictive modeling using chemometric and machine-learning approaches such as partial least squares regression, discriminant analysis, and neural networks.

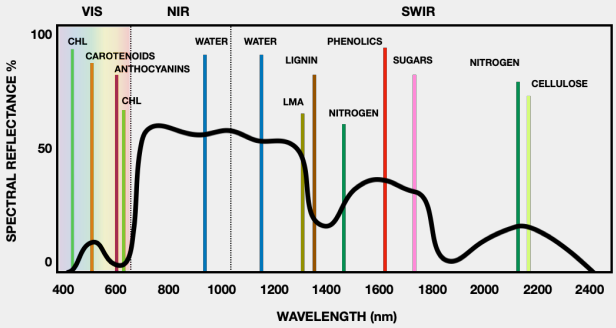

Fig. 1. *Typical fresh leaf spectrum showing the percentage of light reflected across the visible (VIS), Near-Infrared (NIR) and Short-Wave Infrared light regions and highlighting a few absorption features associated with leaf chemistry and structure. Redrawn from Cavender-Bares et al, in review.*

Spectra have been widely used to predict leaf functional traits — such as leaf mass per area (LMA) and leaf nitrogen content — that reveal plant resource use strategies and ecological roles, offering insights into species interactions, community assembly, and ecosystem functions such as productivity and disturbance resistance (Wright *et al.*, 2004; Díaz *et al.*, 2016). Estimating traits is critical for understanding biodiversity responses to global change, refining predictive models of ecosystem function, and monitoring plant strategies and resource availability across scales efficiently (Díaz *et al.*, 2016; Funk *et al.*, 2017).

Reflectance spectroscopy also offers a robust, non-destructive method for taxonomic classification by capturing the unique spectral profiles of different taxa (Meireles *et al.*, 2020), allowing researchers to distinguish species, populations, and even hybrids with accuracies comparable to genetic barcoding (72–100%; Abasolo *et al.*, 2013; Lang *et al.*, 2017; Stasinski *et al.*, 2021).

By integrating signals from leaf structural, chemical, and physiological traits, spectra facilitate species identification and offer insights into ecological and evolutionary processes (Cavender-Bares *et al.*, 2016, Cavender-Bares et al., in review.; Cotrozzi *et al.*, 2017; Meireles *et al.*, 2020; Kothari & Schweiger, 2022).

Recent advances demonstrate that spectral data taken from both fresh and recently dried leaf samples can be effectively used to generate accurate predictive models for taxonomic discrimination and functional trait values (Durgante *et al.*, 2013; Costa *et al.*, 2018; Kothari *et al.*, 2023).

78

The plant tissues in herbaria have been altered by dynamic collection, processing, and preservation techniques with variable storage conditions and durations, ranging from months to centuries. Variability in preservation methods across institutions globally and over time have inevitably led to subtle changes in tissue structure and composition that create challenges to aggregating and comparing data. Preservation and pest removal have in many cases introduced contamination in the form of chemical residues or post-mortem biological agents (Bieker *et al.*, 2020). There is also the problem that plant specimens are glued to herbarium paper, and reflectance scans of glued portions of these specimens will be mixed with the chemical components and structure of these materials (see Neto-Bradley *et al.*). As herbaria are just beginning to incorporate spectral scanning into their digitization pipelines, there is urgency in communicating the unique challenges of herbarium spectral scanning to establish standardized protocols that ensure data quality and compatibility.

Several studies have now used pressed leaves (*i.e.* collected, dried, pressed, and stored in newsprint) for trait prediction and taxonomic classification with remarkable success (Durgante *et al.*, 2013; Costa *et al.*, 2018; Kothari *et al.*, 2023), yet this paper, along with the investigation by Neto-Bradley *et al.* presented in this special issue, are the first to investigate actual herbarium specimens. As such, our primary questions are: To what extent can herbarium spectra complement fresh or pressed leaf spectra in estimating traits and classifications? How do specimen-specific qualities like age or preservation techniques influence the utility of spectral data? What next steps are necessary to optimize sampling strategies, scanning protocols, and digital infrastructure for global data integration?

This study addresses whether herbarium specimens can serve as a reliable resource for reflectance spectroscopy and exploring the biological signal integrity of these samples. We do this by building off of the trait prediction and taxonomic classification work by Kothari *et al.* (2023), where 618 leaf samples representing 67 species of North American trees, shrubs, and herbs plus one Australian species were assayed for 22 functional traits before being reflectance scanned. The dried and pressed vouchers of these samples were scanned using a PSR+ spectroradiometer with leaf clip (Spectral Evolution Inc.) after six months to three years of storage time. Reanalysis of this dataset at 5 nm resampled bands (for comparability) confirmed that pressed-leaf spectra excelled at predicting LMA ($R^2 = 0.94$; %RMSE = 6.29%), carbon ($R^2 = 0.79$; %RMSE = 10.10%), and cellulose ($R^2 = 0.77$; %RMSE = 7.80%), along with moderate success for water-related traits, some nutrients, and pigments. Taxonomic classification models achieved excellent accuracy (91%) for 10 species ($N > 20$), demonstrating the viability of pressed-leaf spectra for functional trait estimation and species identification without destructive sampling.

Here, we have targeted 25 of the most well-sampled species in the pressed-leaf dataset for scanning at the Harvard University Herbaria. This provides a comparative framework to test the utility of herbarium specimens for predicting leaf mass per area (LMA) and for taxonomic classification. Leaf mass

per area was the best-predicted trait for the pressed-leaf models and does not require destructive sampling to measure on herbarium specimens, when detached leaves are available. Additionally, this approach permitted evaluation of the transferability of the pressed-leaf trait estimation models, which predict the values of functional traits as they existed in vivo, to our herbarium spectra. Our goal here was to understand just how different the models generated from pressed or herbarium leaves would be as a proxy for understanding the changes herbarium leaf tissues undergo during preservation and storage. Finally, we investigated whether specific herbarium specimen qualities, such as age, greenness, and the presence of glue, were correlated with the success of taxonomic classification.

Our work seeks to form and refine a vision of herbarium-based spectral scanning as a promising method for inferring plant traits within and across undersampled clades, and to begin to build spectral databases that will be used in a variety of biodiversity science applications. This will also clarify the limitations and opportunities inherent in using herbarium spectra, emphasizing the need to refine methodologies and outline the scaled use of herbarium spectra in biodiversity science. By leveraging this powerful technology with the amazing plant diversity collections in herbaria, we can establish a next generation of digital resources that will be efficiently applied to pressing challenges in ecological and evolutionary research.

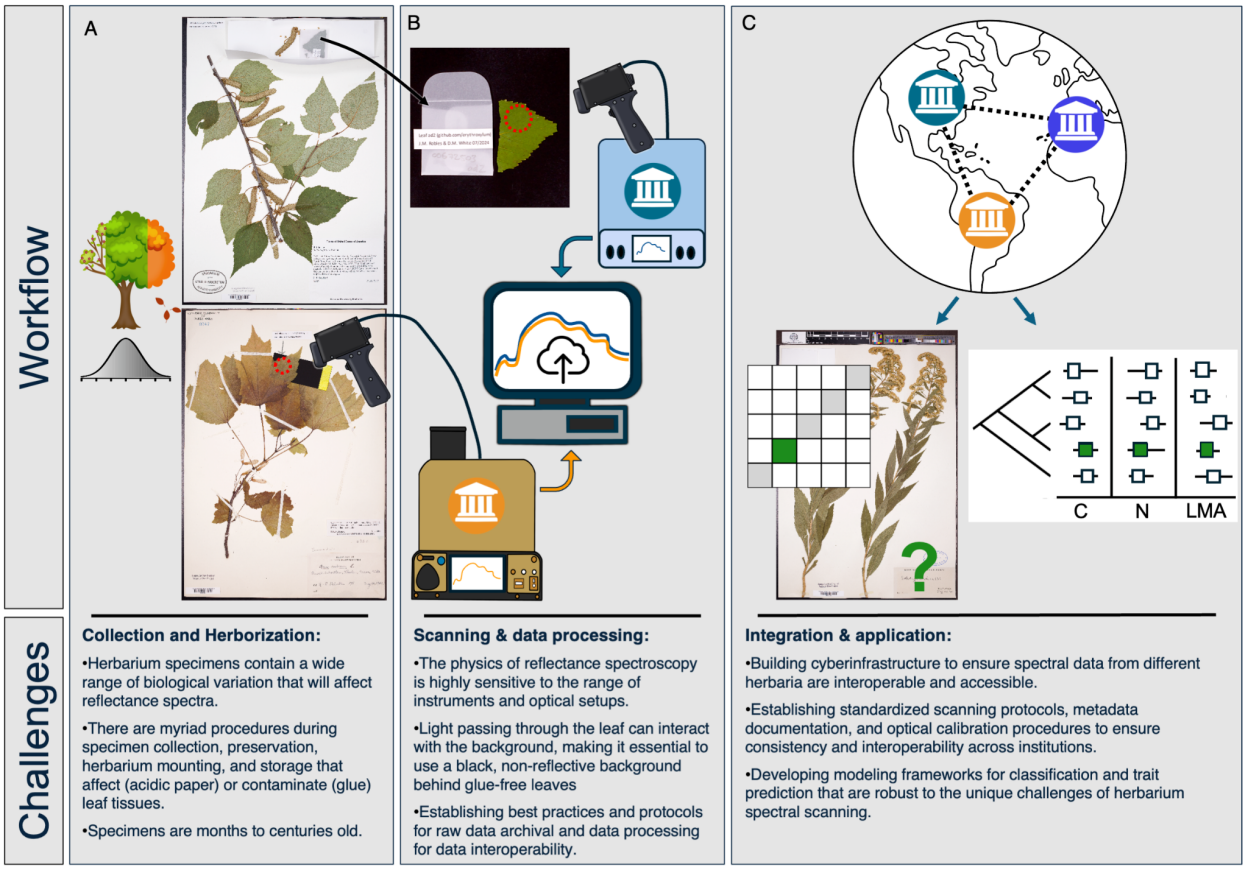

Fig. 2: **The herbarium spectral scanning workflow from specimen collection to global integration, and challenges**. **A)** The biological variation, specimen collection, and herborization factors affecting leaf spectra. Herbaria capture natural variation of plant tissues due to ontogeny, phenology, and plastic responses to abiotic and biotic factors. Along with this natural variation, the range of protocols employed to collect, preserve, and store specimens in herbaria influence tremendous variation in tissue preservation. Herbarium users observe a range of green (top) to brown (bottom) to black leaves and other characteristics reflecting tissue degradation or damage. Specimens have been mounted to different types of herbarium paper via glue (top), tape (bottom), or sewing. Loose leaf fragments held in envelopes (top) might be the only source of glue-free tissue for spectral applications. These specimens are to be preserved in perpetuity, but tissues will continue to change as they are used for research activities and experience environmental fluctuations as they age within herbaria across the globe. Top specimen from A: Herbarium of the Arnold Arboretum of Harvard University; bottom specimen from ECON: The Economic Herbarium of Oakes Ames of Harvard University. **B)** Scans taken with different instruments will inevitably be different. Spectroradiometers vary in spectral range, resolution, and signal-to-noise ratio, and even measurements taken with the same instrument model are affected by the optical components — e.g., light sources, lenses, and probe geometry — and instrument settings (e.g. integration time). The background against which leaves are scanned significantly affects the spectral signal and we currently do not have robust methods to unmix leaf spectral signals from the herbarium paper and glue, so herbarium leaves should be scanned against a standardized black background. This is done using detached leaves stored in labeled packets (top) or by carefully sliding a black background under glue-free leaves taped or sewn onto the herbarium sheet (bottom). Spectral data processing — e.g. scaling, band resampling, and applying mathematical transformations (e.g., wavelet transformations) — can be useful

but also introduce unintended biases and variation, so the raw spectra should always be made publicly available. **C)** The vision of integrating herbarium spectral data into a global Metaherbarium, enabling interoperable networks that connect institutions and support transformative scientific applications, such as functional trait estimation and taxonomic discrimination. Achieving this vision relies on standardized scanning protocols, optical procedures, and instrumentation, along with robust cyberinfrastructure to aggregate and harmonize spectral data across herbaria, ensuring compatibility for downstream analyses. These efforts lay the groundwork for advancing modeling frameworks capable of addressing biological variation across the plant tree of life, while accounting for spectral changes introduced during herborization and specimen aging. Such advancements will bridge herbarium collections with ecological and evolutionary research, ushering in a new era of integrative biodiversity science. Specimen from NEBC: The New England Botanical Club Herbarium.

Methods

Sampling design

Harvard University Herbaria (HUH) collections metadata for 68 species were obtained from the Global Biodiversity Information Facility (GBIF.org) database using the R package rgbif v.3.8.0 (Chamberlain *et al.*, 2024). To minimize geographic variation in traits that could affect trait comparisons, we targeted collections from New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont). This contrasts with the geographic scope of Ontario and Quebec in Kothari *et al.* (2023). The one exception in both datasets is the species *Agonis flexuosa* from Australia. We first inspected all specimens per species and selected specimens 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 remaining herbaria, to detach one leaf for scanning against the black background and measuring LMA. If multiple leaves were available, we selected leaves without any sign of glue, but otherwise tried to randomly sample with respect to the visual quality and degree of degradation across specimens.

Scanning Protocol

Specimens were scanned using a Spectra Vista Corporation HR 1024i spectroradiometer (350–2,500 nm spectral range) with a fiber optic cable connected to the LC-RP Pro Leaf Clip/Reflectance Probe with a narrow-angle lens, which reduced the scanning aperture to a 6 mm x 4 mm ellipse. The instrument was turned on for a minimum of 10 minutes prior to scanning to allow the light source to warm and the sensors to cool. At the beginning of each session, black card stock sprayed with three coats of Krylon® Camouflage Black Matte spray paint was target-scanned three times to record a black background spectrum for downstream quality control (not applied here) and then we took a reflectance scan on a white Spectralon® reference panel. For each sample, we placed the leaf on top of the spray-painted black cardstock and took three two-second scans of the adaxial surface. As an extra precaution against scanning light reflected from the herbarium bench, the cardstock and leaves were scanned on top of a 5 mm felt pad coated with the same matte black spray paint. We targeted regions of the leaf that avoided the midvein, prominent secondary veins, or regions with disease, fungus, or other damage. Up to two leaves per specimen were scanned. Finally, a second target scan of the white Spectralon® panel was taken for downstream quality control (not applied here).

Trait Measurements

Leaf weight, area, and thickness were recorded for each scanned leaf to validate leaf mass per area (LMA) predictions from spectra. After scanning, petioles were removed at the point of contact with the leaf lamina or slightly above 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 with labeled scan numbers. Leaf area was measured using the LeafByte® app on an iPhone 15 with five or 10 cm^2 calibration dots.

Spectra Preprocessing

To ensure compatibility with downstream analyses and comparability of results across datasets, we reprocessed and reanalyzed the raw pressed leaf spectra of Kothari *et al.* (2023) in addition to the herbarium leaf spectra. Raw spectra files were processed using the SpectroLab v. 0.0.18 R package (Meireles *et al.*, 2017). We resampled reflectance spectra to 5 nm intervals using the Full-Width Half-Maximum (FWHM) method in the CWT R package (Guzmán Q., 2024) to make the spectral resolution consistent, as the two datasets were generated on different sensors with different resolutions.

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 (CWT) 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 Q. & 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 large 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 and applying a variance of 1. The choice of wavelet scales can impact the predictive performance of predicting models (Guzmán Q. & 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 Q., 2024).

The resulting reflectance spectra (e.g., reflectance, vector normalized, and summed-wavelet) were trimmed to a range of 450–2,400 to remove noisy regions at the spectrum's edges. The 400–450 nm range

was removed because reflectance values in this region differed substantially between the herbarium and pressed leaf datasets. We also subdivided the data into different spectral regions: 450–1,300nm as the visible and near-infrared (VNIR+) region ("+" because 1,100–1,300 nm is in the short-wave infrared) that could be noisier due to pigment degradation (Fourty *et al.*, 1996), and the 1,350–2,400 nm short-wave infrared region.

Prediction of LMA

Using the processed spectra and the measured leaf mass per area (LMA; kg m^{-2}) from each of the pressed and herbarium datasets across the VNIR+ (450-1,300 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 (Liland *et al.*, 2024b; Kuhn *et al.*, 2024). 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.* (2023). We generated 1,000 model segments by randomly selecting individual scans for each specimen using a custom data segmentation function. This procedure ensured that scans from each specimen were never split among both the training and validation datasets while capturing the variability within specimens and avoiding bias introduced 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 standard deviation 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). Finally, we calculated variable importance in projection (VIP) values to estimate the most informative spectral regions, and extracted model coefficients for making predictions across external datasets – permitting our tests of model transfer between the pressed models to herbarium spectra and vice versa. Lastly, we used the trait data presented in Kothari *et al.* (2023) to generate PLSR models in the same manner as for LMA in order to use the model coefficients and intercept to predict trait values from the herbarium leaf spectra for carbon, calcium, carotenoids, cellulose, chlorophyll A, LMA, nitrogen, and solubles. To assess the accuracy of model transfers for these

other traits for which we have no observed herbarium trait values, we compared the distributions of predicted herbarium trait values against the observed values from Kothari *et al.* (2023).

To further evaluate transferability, we applied model coefficients derived from one dataset to spectra from the other. Using the transformed reflectance data, predictions were generated, and their accuracy was assessed by calculating residuals and comparing predicted vs. observed values. This step validated the applicability of standardized coefficients across datasets and quantified the degree of compatibility between herbarium and pressed leaf spectra.

Taxonomic Classification

To test the viability of models classifying herbarium leaf scans 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, without any vector normalization or continuous wavelet transformations. We tested both the PLS-DA and LDA algorithms because, although they are both implemented by different research groups, we are not aware of any studies that have directly compared their results. PLS-DA uses partial least squares regression to reduce dimensionality and optimize feature selection, making it suitable for high-dimensional datasets like spectra, especially in scenarios with few samples compared to many predictors (high-dimensional low-sample-size problems). This method requires researchers to specify the number of components used by the model, demanding a careful balance between improving accuracy and avoiding overfitting to the training dataset. LDA, in 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. LDA is also computationally much lighter than PLS-DA.

Classification models were built using the *caret*, *pls*, and *plsVarSel* packages in R (Liland *et al.*, 2024a,b; Kuhn *et al.*, 2024). First, spectral data were preprocessed by splitting the dataset into ten 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 1,000 data segments for iterative training and testing across spectral scans.

For PLS-DA, model tuning was performed with the PLS method and optimized by the classification accuracy metric. We generated final models across our 1,000 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 variable importance in projection (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 classification

To evaluate the biotic and herborization factors influencing the success of PLS-DA classification, we utilized the full ensemble of 1,000 optimized PLS-DA models trained on the full-spectrum herbarium dataset of 25 species. Using the model coefficients, x-values, and y-values (intercepts), we computed classification probabilities for all 1,690 herbarium leaf scans across the ensemble of models. Probabilities were averaged across all models, and the predicted class was determined as the one with the highest average probability.

These predictions were integrated with specimen metadata to conduct a series of comparisons and independent regressions of classification probabilities against categorical variables (specimen quality, glue presence, observed damage, and leaf developmental stage) and numerical variables (age, Julian day of collection, nearest taxon distance, LMA, and greenness index). Descriptions of predictor variables are provided in Table 1.

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.PhyloMaker2 (Jin & Qian, 2022) to add *Phragmites australis* as sister to *Phalaris arundinacea* at 39.8 My and add *Betula populifolia* as sister to *Betula papyrifera* at 39.7 My. Greenness index, which measures the relative difference in reflectance between green light (550 nm) and red light (690 nm; see equation in Table 1), was selected over other commonly used vegetation indices, such as normalized difference vegetation index, green normalized difference vegetation index, and chlorophyll/carotenoid index, due to its significant correlation with the independent estimate of specimen quality (Fig. S1). Relationships and regressions were visualized using the ggplot2 package in R (Wickham *et al.*, 2024), and significant differences in classification probabilities between correct and incorrect classes were assessed using t-tests as implemented by the 'ggsignif' function in ggplot2.

Logistic regression and random forest analyses were performed to further evaluate significant relationships between specimen metadata and classification accuracy. The logistic regression model, implemented with the 'glm' function in the stats R package (R Core Team, 2023), employed a binomial error structure and included all predictors. For the random forest analysis, the randomForest R package (Breiman *et al.*, 2024) was used to quantify predictor importance based on mean decrease in accuracy and Gini impurity metrics.

Table 1: Metadata predictors of herbarium specimens and descriptions of classes.

Metadata predictor	Class	Description
Leaf Developmental Stage	**Young**	Thin leaves with under-developed venation, prone to bruising, may appear darker, scans usually have lower reflectance. Collection date is important.
	Mature	Typically thick leaves, with potential color differences between adaxial and abaxial surfaces.
	Senescent (Not observed)	Discolored leaves, often associated with aging. Collection date helps confirm senescence.
Leaf Damage	**None**	No visible pre- or post-mortem damage to leaves.
	Minor	Damage present on some leaves but not affecting scanned regions.
	Medium	Minor damage visible on scanned leaves, but likely not in the scanned regions.
	Major	Significant damage is visible in the scanned regions.
Specimen Quality	**Good**	Resembles a freshly pressed specimen.
	Medium	Shows some discoloration but is otherwise intact.
	Poor	Highly degraded specimen, with discoloration, mold, or rugosity from wilting.
Glue	**Present**	Specimen preparation involved glue application.
	Absent	No glue was used during specimen preparation.
Green Index	**(Numerical)**	Green Index= $Reflectance_{550nm}$ - $Reflectance_{690nm}$ / $Reflectance_{550nm}$ + $Reflectance_{690nm}$
Age	**(Numerical)**	Years since specimen was collected (mean = 94)
Day of Year	**(Numerical)**	Julian day of collection
Leaf Mass per Area	**(Numerical)**	Kg m^{-2}
Nearest Taxon Distance	**(Numerical)**	Estimated age (in millions of years) of most recent common ancestor shared between predicted taxon and nearest sampled species.

Results

Trait prediction & model transferability

Spectral profiles of 25 species scanned from the Harvard University Herbaria have similar profiles but lower magnitudes compared to pressed leaves (Fig. 3A). 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 ~1,900-2,400); Fig. S2). Models trained on herbarium spectra using all combinations of spectral transformations (untransformed, vector-normalized, and CWT) and wavelength ranges (full, VNIR+, and SWIR) had excellent performance (validation tests in Table 2; full statistics in Table S1).

Overall, the best 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 non-transformed reflectance values, pressed LMA models performed only slightly better than the herbarium LMA models (pressed $R^2 = 0.942$, %RMSE = 6.29%; herbarium $R^2 = 0.891$, %RMSE = 6.62%, Fig. 4A and B). After full-range models, SWIR models performed better than VNIR+ (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 non-transformed reflectance models could still accurately predict observed LMA (Table 2; Table S1; Fig. 4B and 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. 4C and 1.25 in Fig. 4D; transfer tests in Table 2). Models based on the VNIR+ spectra also performed well for 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 worse than the other two spectral datasets, yet showed best performance for models in the SWIR range (Table 2; Table S1).

353 Table 2: Performance metrics for LMA models averaged across 1,000 model segments.

test	model	spectra	transform	N	Ncomponents	R^2	%RMSE	RMSE (Kg m^{-2})	BIAS	slope	intercept
validation	herbarium	herbarium	CWT	220	10	0.81 ± 0.01	7.92 ± 0	0.011 ± 0	0 ± 0	1.00 ± 0.02	0 ± 0
validation	herbarium	herbarium	reflectance	220	14	0.89 ± 0.01	6.62 ± 0	0.01 ± 0	0 ± 0	0.99 ± 0.02	0 ± 0
validation	herbarium	herbarium	normalized	220	14	0.92 ± 0.01	5.79 ± 0	0.009 ± 0	0 ± 0	1.01 ± 0.01	0 ± 0
validation	Pressed	Pressed	CWT	212	8	0.94 ± 0	6.34 ± 0	0.011 ± 0	0 ± 0	1.03 ± 0.02	0 ± 0
validation	Pressed	Pressed	reflectance	212	16	0.94 ± 0	6.29 ± 0	0.011 ± 0	0 ± 0	1.02 ± 0.02	0 ± 0
validation	Pressed	Pressed	normalized	212	13	0.95 ± 0	6.01 ± 0	0.011 ± 0	0 ± 0	1.00 ± 0.02	0 ± 0
transfer	herbarium	pressed	CWT	609	14	0.88 ± 0.03	8.76 ± 0.02	0.015 ± 0	0 ± 0.01	0.91 ± 0.06	0.00 ± 0.01
transfer	herbarium	pressed	reflectance	609	14	0.91 ± 0.01	10.99 ± 0.02	0.019 ± 0	-0.01 ± 0.01	0.82 ± 0.02	0.00 ± 0.01
transfer	herbarium	pressed	normalized	609	14	0.91 ± 0.01	78.48 ± 0.5	0.136 ± 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	0.016 ± 0	0 ± 0	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	0.01 ± 0.01	1.13 ± 0.1	0.01 ± 0
transfer	pressed	herbarium	normalized	479	13	0.51 ± 0.09	781.00 ± 2.43	1.178 ± 0.37	-1.18 ± 0.37	0.41 ± 0.09	-0.41 ± 0.06

354
355

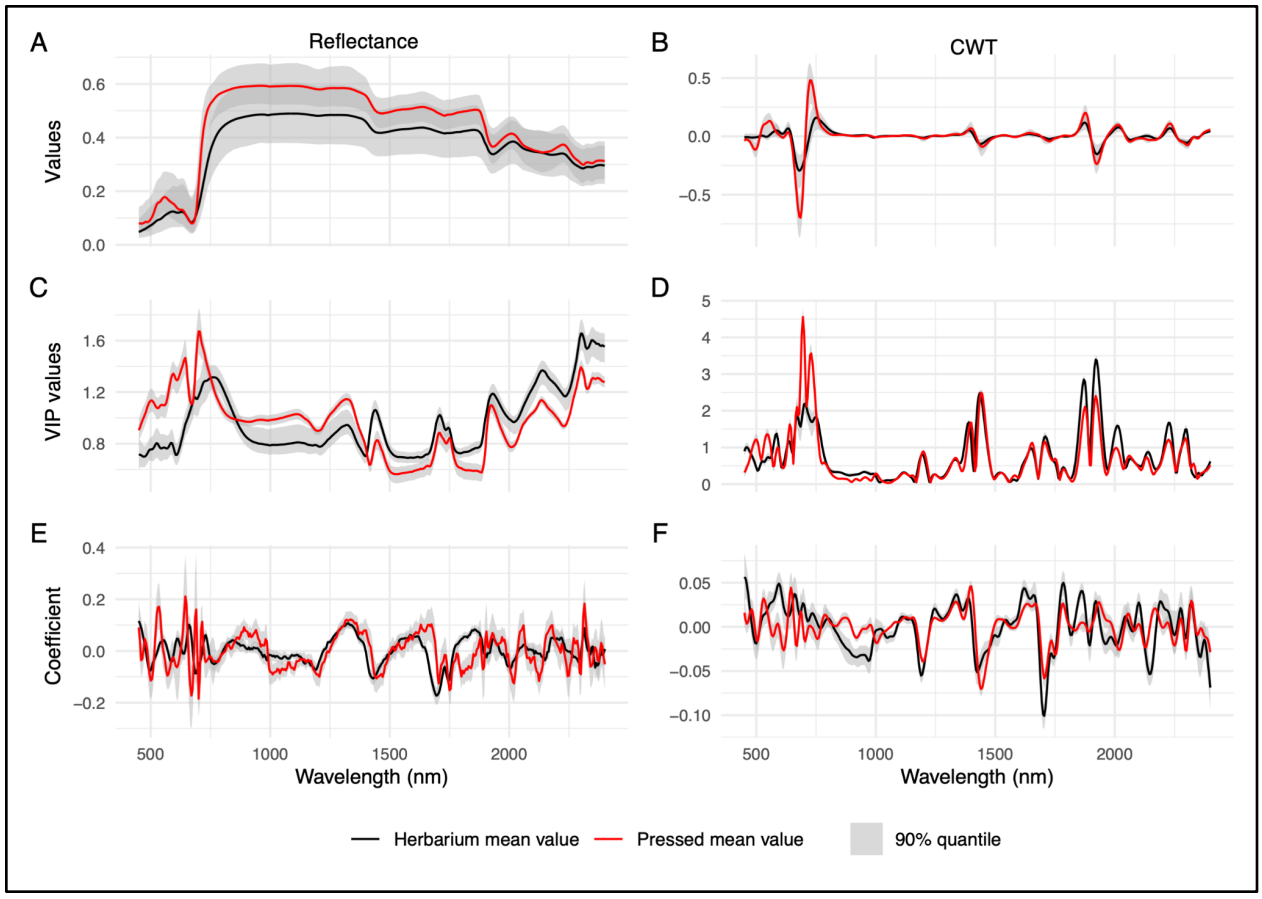

Fig. 3: Plots of reflectance and CWT values for herbarium and pressed datasets, plus associated variable importance in projection (VIP) metrics and model coefficients for LMA models. Black lines represent mean herbarium data and red lines represent mean pressed data, with 90% quantiles plotted in gray bands. Panels show the data for (A) the reflectance data across all samples, (B) the CWT transformed reflectance data across all samples, (C) VIP values for reflectance data across 1,000 model iterations, (D) VIP values for CWT data across 1,000 model iterations, (E) Reflectance model coefficients across 1,000 iterations, (F) CWT model coefficients across 1,000 iterations.

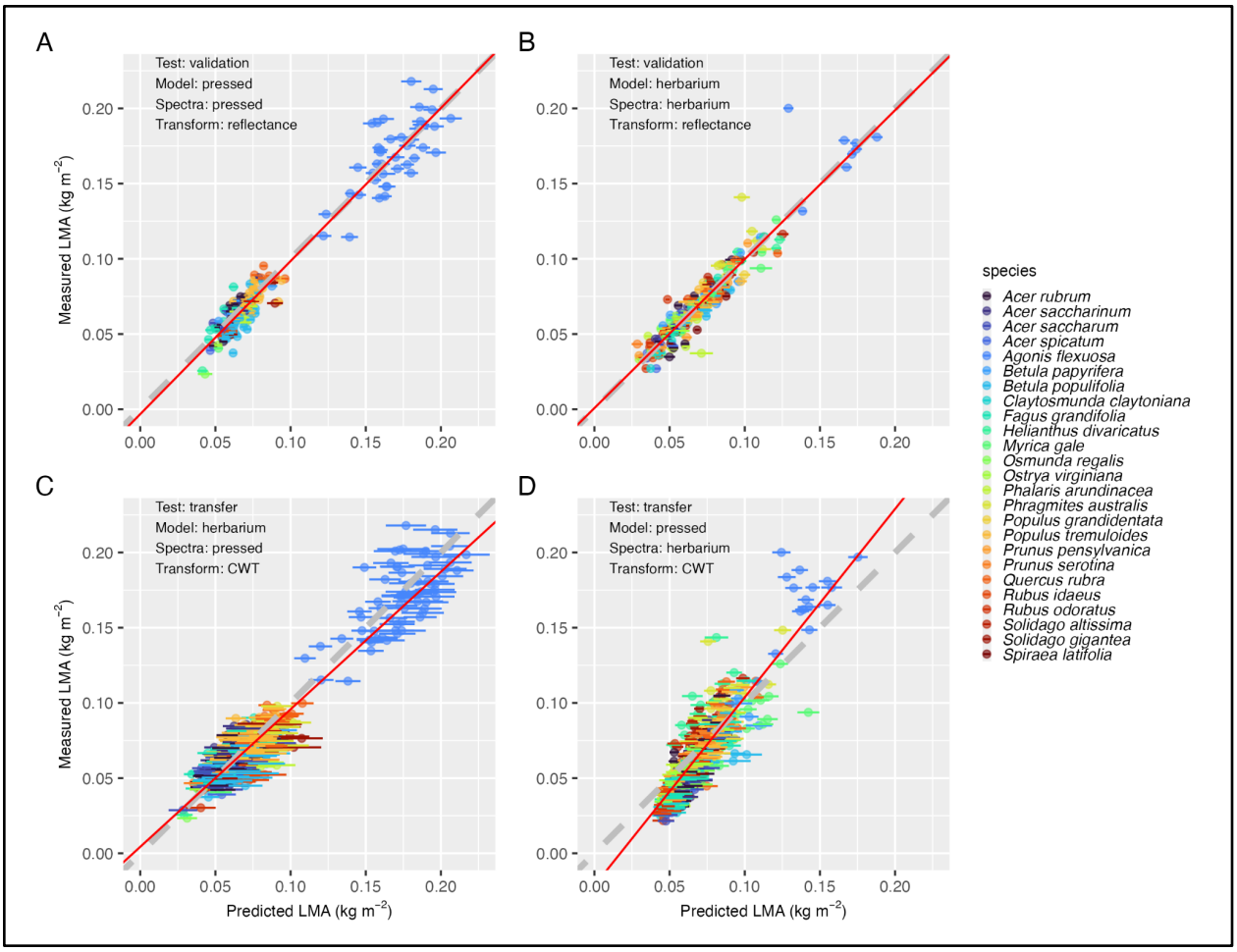

Fig. 4: Validation and model transfer results for leaf mass per area (LMA) per individual across 25 species. Error bars represent the standard deviation in predictions across 1,000 model iterations. Linear regressions of observed versus predicted values averaged across iterations are shown in red lines for comparison with the gray 1:1 dashed lines. Individual plots show the results for full-range spectra (450-2,500 nm) of (A) pressed models from untransformed reflectance values, (B) herbarium models from untransformed reflectance values, (C) transfer of CWT herbarium models to CWT pressed spectra, and (D) transfer of CWT pressed models to CWT herbarium spectra.

The compatibility of the models is further illustrated by the similarity of variable importance in projection (VIP) values for reflectance spectra (Fig. 3C). The VIP plots reveal considerable differences between herbarium and pressed models in the visible and (less-so) NIR regions, but the relative values in the SWIR region are similar. This same pattern applies to the model coefficients (Fig. 3C). The CWT models follow a similar pattern across the visible, NIR, and SWIR regions, yet with fewer differences among peak patterns and overall closer magnitudes (Fig. 3D and F). The CWT models have the most clearly defined peaks and highlight informative spectral regions throughout the spectral range (Fig. 3D peaks = VIS: 500 nm, 545 nm, 590 nm, 640 nm, 670 nm, 695 nm; NIR: 730 nm; SWIR: 1,200 nm, 1,400 nm, 1,440 nm, 1,655 nm, 1,705 nm, 1,875 nm, 1,920 nm, 2,225 nm, 2,295 nm).

To extend this 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. 5; 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 leaf mass per area (LMA), carbon fractions, and carotenoids, generally showed contiguous distributions with substantial overlap between datasets. This overlap shows the robustness of the spectral models in maintaining rank-order consistency across species. Discrepancies arose, particularly where pressed datasets included only a single individual per species and model generalizability was limited, but the lack of unrealistic trait values and the general correspondence of distributions across these additional traits is a surprisingly positive result.

These results taken together provide robust support for the utility of herbarium spectra for trait estimation both for models built from herbarium-derived trait datasets as well as for the transfer of pressed leaf models built from trait values measured in living plants.

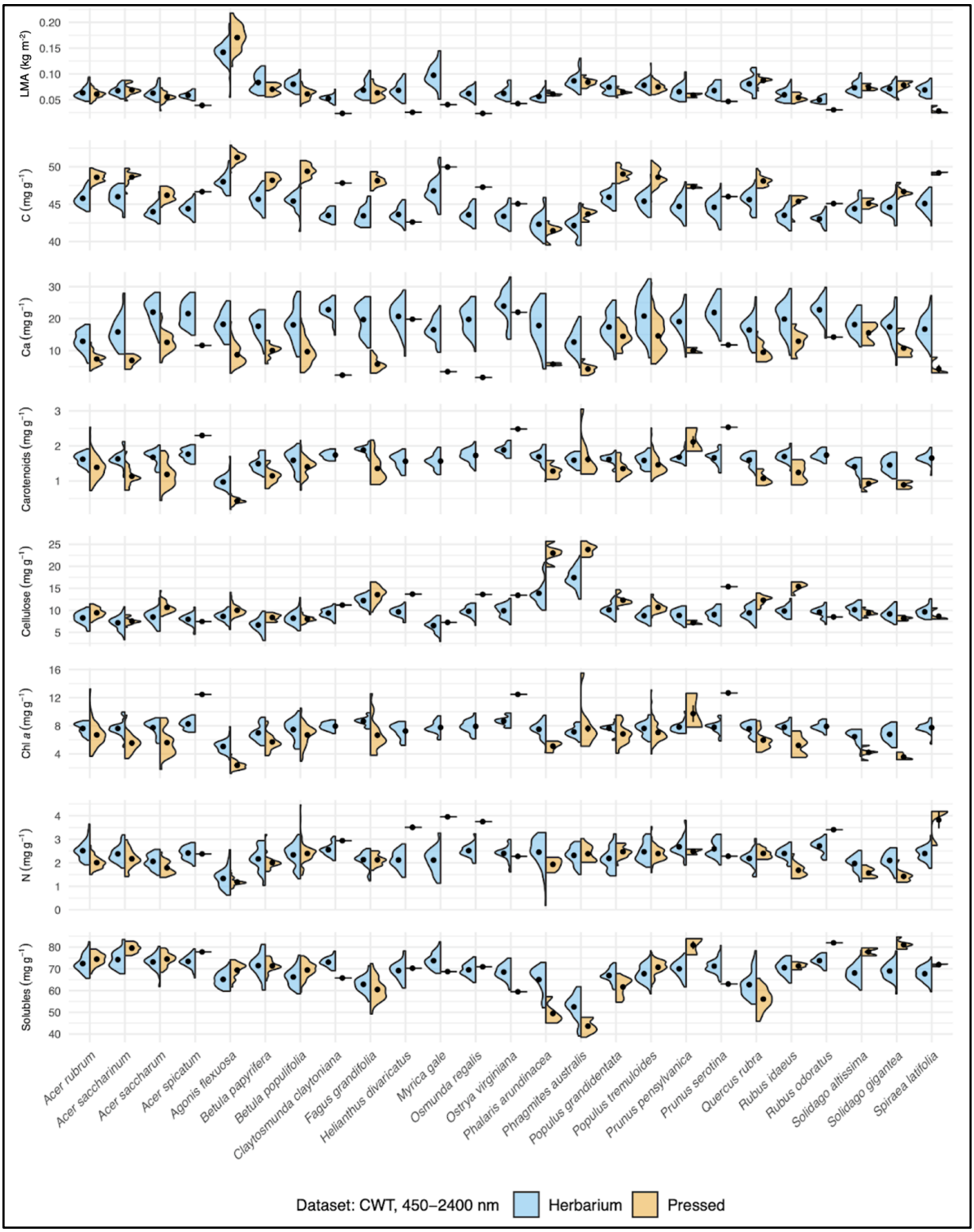

Fig. 5: Comparison of observed trait distributions from pressed leaves with predicted values obtained by applying continuous wavelet transformation (CWT) pressed models to herbarium spectra, as shown in Fig. 2C. Panels display the distributions for eight traits across 25 species. Mean values are indicated with black dots.

Taxonomic Classification

To evaluate the utility of reflectance spectra for taxonomic classification, we applied linear discriminant analysis (LDA) and partial least squares discriminant analysis (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 ten species for which 20 or more individuals were sampled. 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 3). 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 scans misclassified as *Solidago altissima*. The variable importance in projection (VIP) plots are consistent across species and emphasize key spectral regions in the visible, near-infrared, and shortwave infrared (SWIR) ranges (Fig. S3).

At the genus level, pressed specimens achieved near-perfect accuracy in the six-genera dataset, with 96.9 ± 1% (LDA) and 89.8 ± 1% (PLS-DA), while herbarium specimens achieved 89.3 ± 2% (LDA) and 82.1 ± 1% (PLS-DA). 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 LDA in the 17-genus dataset, 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 shortwave infrared (SWIR) regions, reflecting the spectral regions most important for PLS-DA classification (Fig. S4).

Table 3: Performance metrics of classification analyses

Dataset	**Rank**	**Model**	**Samples**	**Ncomponents**	**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	PLSDA	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	PLSDA	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	PLSDA	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	PLSDA	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	PLSDA	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	PLSDA	17 genera	27	84.9 ± 1	87 ± 8	90.9 ± 9

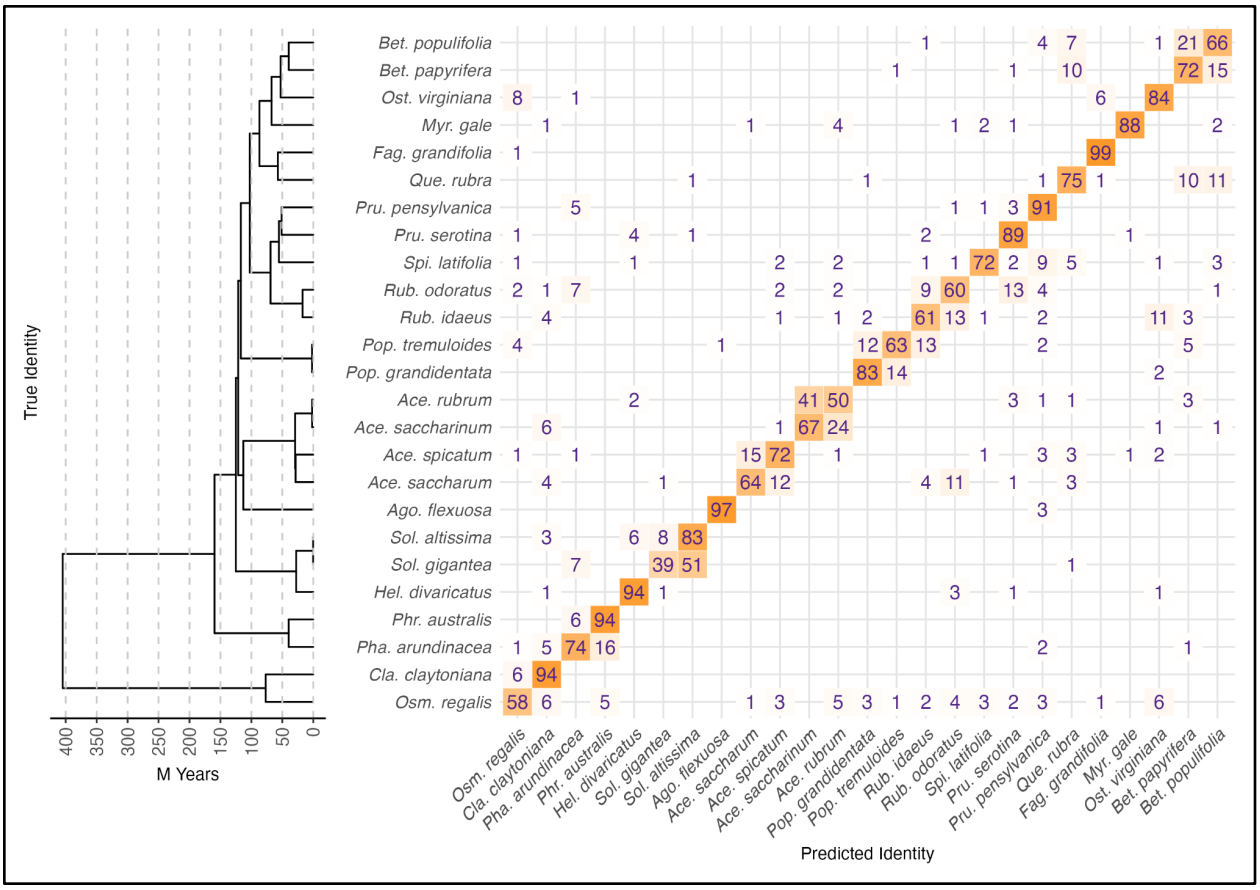

Fig. 6: Phylogram and confusion matrix summarizing the validation of herbarium specimen classification using Partial Least Squares Discriminant Analysis (PLS-DA). The left panel shows a phylogram representing the evolutionary relationships among species scaled by millions of years. The right panel displays 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%.

Assessing herborization factors on classification

To evaluate the influence of specimen factors on PLS-DA classification performance, we analyzed the classification probabilities across all 1,690 herbarium scans using the full-spectrum 25-species dataset (Fig. S5). 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. 7). Leaves with good ($p < 0.001$) or medium quality ($p < 0.01$) had higher predicted probabilities for correct classifications compared to 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 predicted probabilities than those with glue ($p < 0.001$). Mature leaves exhibited higher predicted probabilities compared to young leaves ($p < 0.001$). Probabilities of correct classifications for specimens with no damage were significantly higher than those with minor damage ($p < 0.001$) and medium damage ($p < 0.05$), as well as between minor and major - but in the opposite direction of expected ($p < 0.05$). The few specimens with major damage had curiously high classification probabilities. The probabilities of incorrect classifications – which represent the individual false-positive classifications that had higher probabilities than true-positives – did not significantly differ across classes within these variables (Fig. 7).

Numerical predictors also had significant relationships with classification probabilities (Fig. 8). Age negatively correlated with classification probabilities, suggesting reduced model performance for older specimens (Fig. 8A). The age of the sampled specimens ranged from one to 179 years with a median age of 91 years (Fig. S6). The green index was also negatively correlated with classification probabilities, indicating that greener leaves were related to worse model performance (Fig. 8B). 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. 8C).

Classification probabilities increased with greater phylogenetic distance to the nearest taxon (Fig. 8D), 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. S7). Leaf mass per area also shows a strong positive correlation with classification probability (Fig. S8) with the caveat of covariation with species composition. *Agonis flexuosa* was classified with an overall accuracy of 97% and LMA values for this species are outstanding within this dataset.

Logistic regression taking into account phylogeny (Table 4) 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 with classification success (Table 4)

and classification probability (Fig. S9) with the calendar day of specimen collection. This provides weak support that species collected early in the growing season were 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 integrity, as measured by the green index, and phylogenetic distinctiveness strongly impact classification success. In contrast, older specimens, poor-quality leaves, and the presence of glue reduce classification probabilities, underscoring the importance of these metadata for optimizing model performance.

Table 4: Logarithmic regression of all predictors.

	Estimate	**Std. Error**	**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	

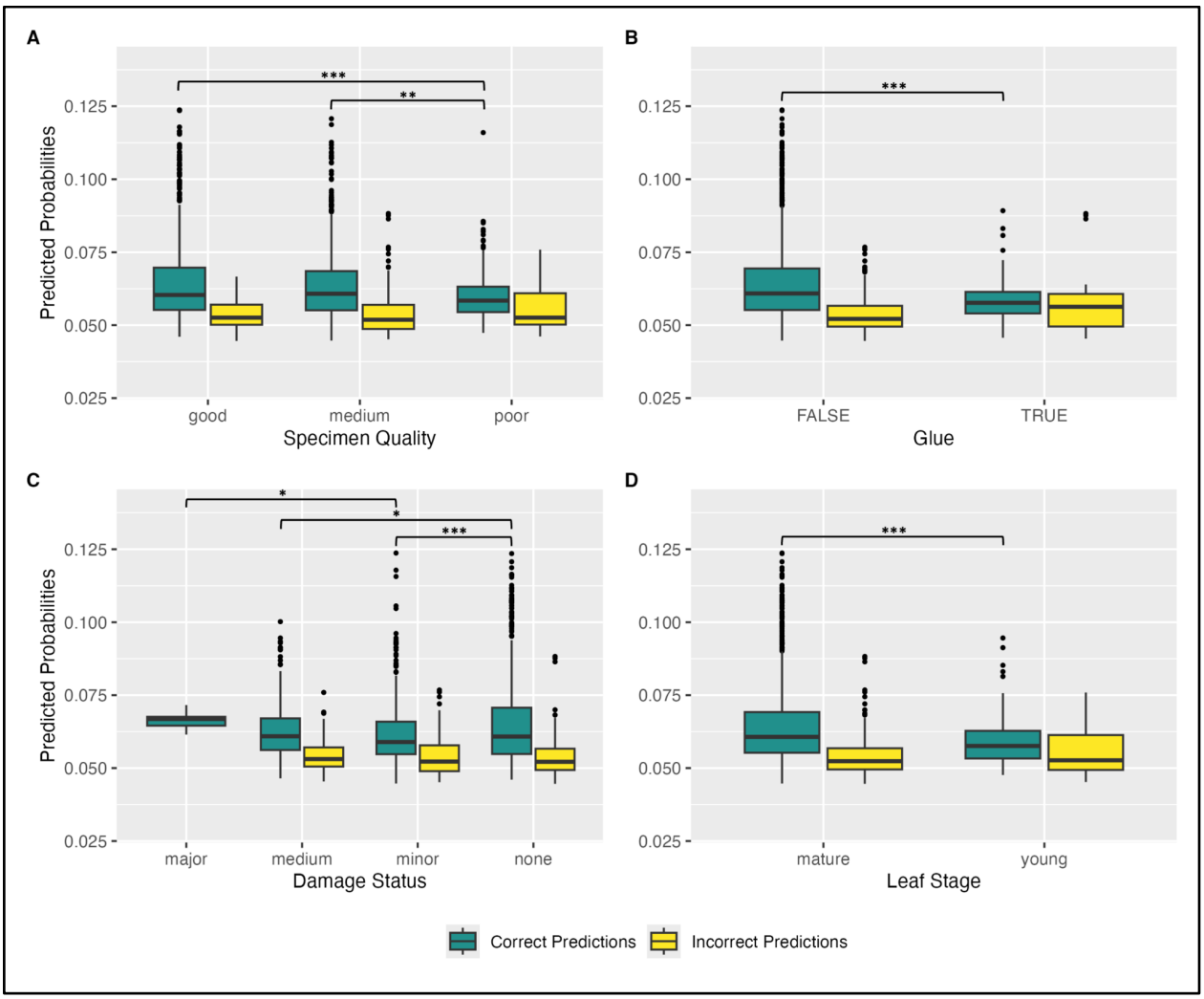

Fig. 7: Comparison of distributions of probabilities of assignment of each scan to a specific class for correctly (true-positive) or incorrectly classified (false-positive) specimens by leaf characteristics (see Table 1). A) Specimen quality observations primarily reflecting discoloration or tissue degradation. B) The presence or absence of mounting glue on the leaf. C) Visible biotic contamination, pre- or post-mortem damage to leaves. D) Leaf phenological development. Significant pairwise differences among correct or incorrect classes were determined using t-tests and indicated with the codes: * ($p<0.05$), ** ($p<0.01$), and *** ($p<0.001$). Note there were no significant differences among classification probabilities for incorrect predictions.

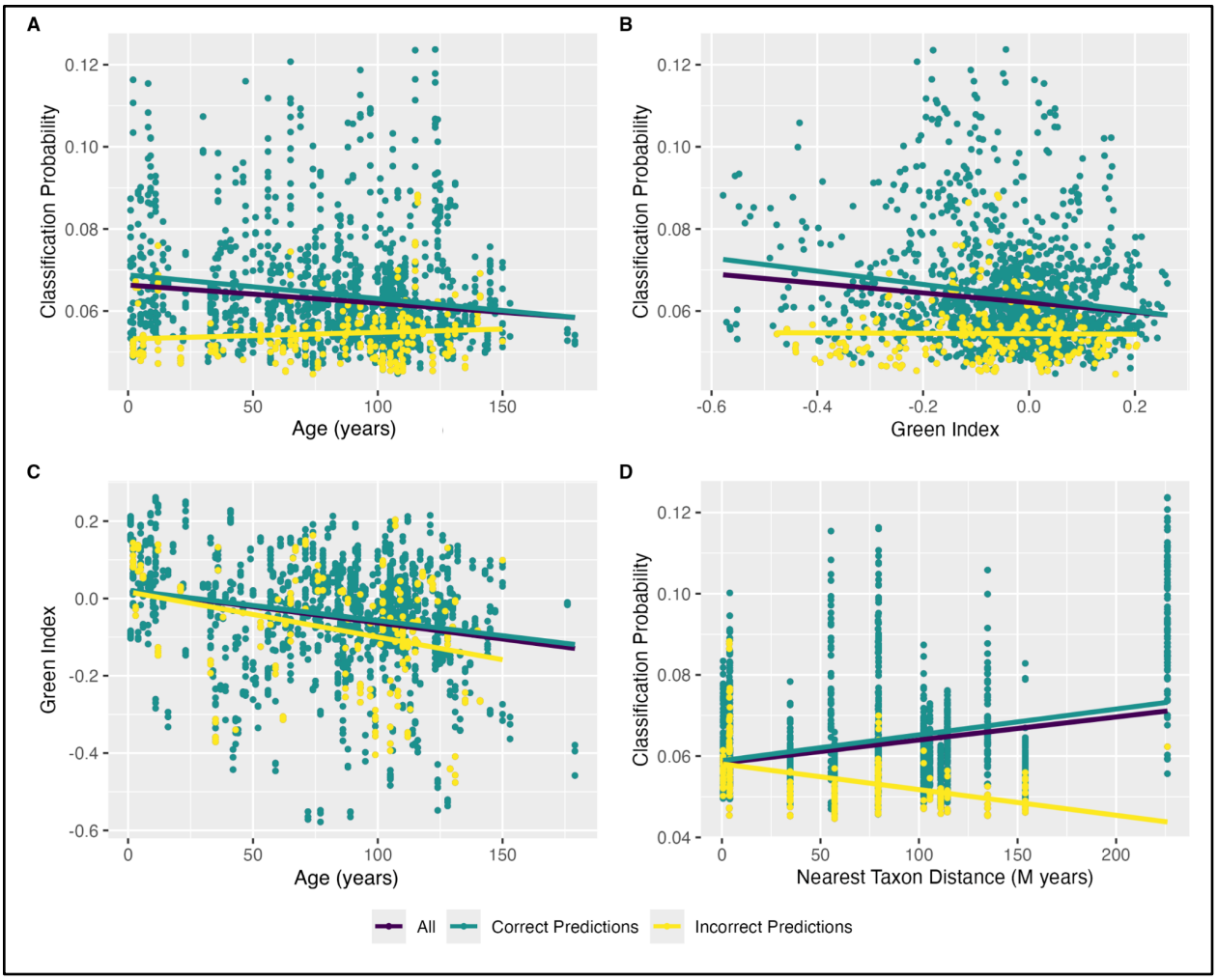

Fig. 8: Relationships between numeric predictor variables and classification outcomes. (A) Relationship between age (years) and classification probability, (B) relationship between green index and classification probability, (C) relationship between age (years) and green index, and (D) relationship between nearest taxon distance (NTD, M years) and classification probability. Points represent individual observations colored by correct versus incorrect status. Solid lines represent linear regression fits for each dataset.

Discussion

As the largest scientific repositories of plant diversity, herbaria offer exceptional resources for investigations of plant biology, but their utility is shaped by the condition of preserved tissues. Recent advances have proven the utility of spectra from pressed leaves (*i.e.* collected, pressed, dried, stored in newspaper) on the order of months to years old for taxonomic classification (Durgante *et al.*, 2013; Lang *et al.*, 2017; Kothari *et al.*, 2023) and functional trait estimation (Lang *et al.*, 2017; Kothari *et al.*, 2023). Our study has extended this discovery to clearly demonstrate that herbarium specimens retain enough morphological and anatomical integrity to be useful for the same spectra-based inferences.

The success of this proof-of-concept highlights the potential of reflectance spectroscopy as a valuable addition to herbarium digitization pipelines (Hedrick *et al.*, 2020; Davis, 2023). To fully realize this potential, the collections community must work collaboratively to establish standardized protocols that ensure the compatibility of spectra collected across institutions (Fig. 2). Advancing standardized protocols will require clear communication of the fundamental concepts of reflectance spectroscopy and the myriad yet understudied factors influencing spectra from herbarium specimens. While our study has identified many of these factors, much remains to be explored. Here, we aim to outline these considerations and challenges as a foundation for the advancement of herbarium reflectance spectroscopy.

Considerations and challenges for herbarium reflectance spectroscopy

Biological variation

The goal of phenomic assessments is to characterize biological variation, but researchers must recognize that spectral scans capture the cumulative effects of all factors—both natural and artifactual—that influence reflected light. Before considering the effects of specimen processing and storage, researchers should record metadata for developmental, phenological, and ecological factors that might influence leaf structure and physiology.

For instance, herbarium specimens representing species with asynchronous flowering or fruiting may disproportionately contain young, developing leaves or reproductive structures, leading to spectra that do not adequately represent mature leaves (Fig. 2A). This variation introduces potential biases that researchers should consider when interpreting spectral data. Alternatively, given the considerable changes in leaf traits caused by phenology, traits will not be accurately estimated by models trained on one phenological stage if the spectra were sampled from a different phenological stage. However, predictive models show promise in accommodating such biological variation as long as models are trained to “see” the full range of variation (Lang *et al.*, 2015).

In our study, mature leaves and young leaves were classified correctly the same proportion of times because both of these traits were modeled (see Results). For traits, models should be trained from datasets that span the range of trait values expected in testing datasets.

Additionally, tissue heterogeneity within a single leaf must be considered. Researchers should avoid scanning leaf midribs, as their higher proportion of vascular tissue will influence spectra. Sampling protocols should prioritize mature laminar leaf regions to ensure that spectral measurements accurately reflect the traits of interest.

Herborization

The herborization process, which encompasses the preservation and storage of collected specimens, presents a wide range of variables affecting plant tissues (Fig. 2A). During collection, plants are pressed flat in newspaper shortly after collection, ideally before wilting. If drying cannot be performed within 24–48 hours, specimens are typically soaked in 50–95% ethanol and sealed to prevent fungal growth. Historically, other preservatives like formaldehyde have also been used. Drying methods vary, ranging from forced-air systems or industrial ovens (30–70ºC for 15–48 hours) to passive drying in arid environments (Bridson *et al.*, 1998; Forrest *et al.*, 2019). Other analog methods necessary in more remote locations, such as drying over hot coals, apply even hotter and more variable temperatures and are the cause of the occasional encounter with a partially burnt specimen. Improperly dried specimens, on the other hand, may exhibit discoloration, structural degradation, or fungal growth.

We assessed the impact of visual cues about the quality of herbarium specimens via our metadata collection on specimen quality - our general interpretation of degradation as interpreted from discoloration, wilting, pathogens, or signs of poor initial preservation - and on specimen damage. Medium and poor herbarium quality classes were inferred to be significantly negatively correlated with correct predictions (Table 4; Table S3), showing that apparent specimen degradation does indeed translate to reduced model accuracy. The damage metadata was primarily used to annotate specimens with obvious tissue alterations such as herbivory or burning that affected part of the specimen but usually not the scanned leaves. Apart from the curious result of the six scans from two specimens of *Populus tremuloides* with major damage having high classification probability yet low species classification accuracy (63%), the more damaged leaves also followed expectations of reduced model performance. However, this was not a significant factor in the logistic regression nor the random forest analysis.

After drying, specimens are transported, frozen for one to two weeks to eliminate insects, and then mounted on herbarium sheets using glue, tape, or sewing. Herbarium sheets are typically made from acid-free, lignin-free paper or cardboard, designed to prevent chemical reactions that could damage specimens over time (Drobnic, 2008). Acid-free paper resists yellowing and brittleness, as acidic

materials in the paper will otherwise degrade both the sheet and the plant material, altering the specimen's color and structural integrity. However, not all herbaria have access to these archival-quality sheets. In such cases, locally available paper may be acidic and will accelerate specimen degradation.

While sewing specimens to the herbarium sheets is the most durable and secure method, it is highly labor-intensive and impractical for most herbaria with limited staff. Consequently, glues are the most widely used adhesive in the United States. Many institutions rely on water-based glues such as Elmer's® Glue-All or Jade 403® because they are easy to apply and generally effective. Glues are either painted or sprayed onto the specimen's backside and blotted dry before pressing the specimen against the sheet for adhesion. However, not all glues are ideal for long-term preservation. Some glues contain acidic or unstable components that can break down over time, causing discoloration, loss of adhesion, or chemical reactions that further degrade both the sheet and specimen. It is common to observe old glues or stains from previous adhesives on older specimens that have been remounted through time. As such, there are a variety of adhesives that have been used through the decades that need to be accounted for if scanned, and furthermore, the leaves of some specimens may contain multiple layers of different adhesives.

Due to their potential for direct contamination of spectra, glue is the most significant contaminating source for herbarium spectra. Our study demonstrated a clear reduction in the probability of correct classifications when glue is present on the leaf (Figure 7B) and significant impact of glue on classification success (Table 4). We have tried and are aware of efforts to 'unmix' or 'subtract' the glue or paper spectra from the leaf using different spectral libraries of these contaminants, but thus far we are not aware of any solutions to isolate the leaf signal from a spectral profile that contains these extra materials. As detailed below, it is critical to standardize scanning backgrounds to ensure protocol consistency and data interoperability.

After mounting, labels containing collection data and envelopes or 'packets' for loose tissues are attached to the sheets before specimens are stored in herbarium cabinets. Although large herbaria in the global north often maintain temperature and humidity controls, daily, seasonal, and annual humidity fluctuations remain a significant challenge worldwide, potentially accelerating specimen degradation.

Sampling herbarium specimens

When selecting specimens for sampling, researchers must choose among different herbarium specimens and then among different leaves. The selected leaves should be mature and free from biotic pathogens or contaminants (unless such factors are within the scope of the investigation), yet otherwise representative of the normal variation in the specimen (Fig. 2A).

Our results on the age and greenness of the specimens might appear counterintuitive since these two variables are, intuitively, negatively correlated; younger specimens tend to retain more greenness, and green specimens are often assumed to represent better-preserved tissues. However, the relationship between age and greenness, and their combined effect on methodological success, is more complex. While DNA sequencing success – a methodology also reliant on tissue preservation – is frequently assumed to decline with age, studies frequently show weak or no correlation between specimen age and sequencing outcomes (Erkens *et al.*, 2008; Forrest *et al.*, 2019; White *et al.*, 2021). Instead, specimen processing methods during the early stages of preservation appear to play a much more critical role in tissue preservation.

Greenness, influenced by the presence of chlorophyll, has been shown to significantly affect optical properties, particularly in the visible spectrum. While green tissues may indicate good preservation, high chlorophyll concentrations in tissues can also mask other spectral features that might be more informative for downstream applications like functional trait prediction or classification modeling. The absence of chlorophyll, as seen in older or less green specimens, could enhance the detection of structural and biochemical features that are less visible in green leaves (Kothari *et al.* 2023). Thus, while greenness remains an important indicator of specimen preservation, its role in spectral data acquisition and prediction success may depend on the specific objectives of the study.

Light transmitted through scanned leaves may reflect from the background (glue, paper, and even lab benches), which 'contaminates' the spectrum, resulting in erroneously high measurements of leaf reflectance. The degree of contamination depends on the optical thickness of the leaves, which governs how much light is transmitted. Our analysis identified a significant positive correlation of LMA with classification success, suggesting that thicker leaves that scatter more light perform better for prediction. We believe it is critical that researchers avoid such contamination by scanning leaves against a black background that absorbs nearly all transmitted light, which is currently being implemented through two different protocols (Fig. 2B). First, if the specimen has been sewn or taped, it may be possible to slide a thin black sheet between the attached leaf and the paper. Second, and preferable, herbarium specimens with loose leaves available in packets may be selected and those leaves checked for glue before being scanned in a leaf clip or against a black background.

Selection of these black backgrounds is a critical component of standardization that is just now being evaluated experimentally. Researchers are currently using EVA foam and other black plastic (Flavia Durgante, pers. comm.), black card stock painted with Krylon® Camouflage Matte Black spray paint (Aaron Lee, pers. comm.), and SpectralBlack® foil (Samantha Bazan, Thomas Couvreur, pers. comm), plus the black backgrounds of manufacturer leaf clips for portable spectroradiometers (Malvern

Panalytical, Spectral Vista Corporation, Spectral Evolution). The identification and adoption of a universal background standard represents one of the most important objectives of protocol development.

Instrumentation

Beyond the tissues and backgrounds, instrumentation variables play a critical role in shaping spectral profiles, adding complexity to the use of herbaria for spectral data collection (Fig. 2B). A standard scanning protocol begins by allowing the instrument's light source to warm up and stabilize. During this time, researchers should record appropriate metadata, including a standardized filename reflecting the herbarium accession number. Once the instrument is ready, a white reference calibration standard is scanned to establish 100% reflectance across wavelengths. Regularly calibrating the instrument with a clean white reference standard for every specimen scanned ensures accuracy in reflectance calculations from radiance. With the calibration complete, the instrument is ready to scan the sample leaf tissue placed on a black background.

Additional factors influencing spectral data quality include the sensitivity of fiber optics, the duration of scans, the geometry of the optical measurement setup (e.g. the angle of incidence) and the potential for light sources to heat and alter leaf properties during scanning. These factors can also change with leaf properties, such as highly rugose leaves or small, round leaves or needles. Variations in fiberoptic alignment or quality can impact the signal-to-noise ratio, requiring careful handling and regular replacement. Regular replacement of calibration standards and routine instrument maintenance, such as cleaning and recalibrating sensors, are essential to sustaining instrument performance. Users select integration time, during which instruments perform multiple rapid scans. Prolonged scanning times can improve signal quality but may introduce heat effects on the specimen, which must be avoided.

Researchers may choose to process raw spectra with resampling and normalization or transformation using derivatives or continuous wavelet transforms (CWT) to standardize datasets from different instruments. We caution against band resampling at higher resolution than was measured by the instrument as this could introduce artificial data to the spectrum. In our study, we used 5 nm resampling for band spacing to harmonize the differences in spectral resolution between sensors (i.e., Spectral Evolution PSR+ and SVC HR-1024i) and reduce the number of correlated bands for predicting models. The raw, 1.4 nm bandwidths did return a higher overall accuracy of taxonomic classification.

Researchers will also typically trim the 350–400 nm and 2,400–2,500 nm regions because they might have a lot of noise. Additionally, we further trimmed our data in later stages of analysis after noticing reflectance and VIP differences between herbarium datasets between the pressed and in the 400-450 nm range. However, the raw spectral data should be archived so future users can choose and improve upon data preprocessing steps.

The success of data aggregation for herbarium spectral scans depends on the adoption of standardized protocols. To this end, we are actively collaborating with individuals from diverse institutions to establish robust and universally applicable methodologies for herbarium spectroscopy. These efforts aim to ensure consistency and reproducibility across studies, paving the way for expanded applications in plant biology and global ecology.

Data aggregation

The success of herbarium reflectance spectroscopy hinges on robust data aggregation practices that ensure consistency, interoperability, and accessibility across institutions (Fig. 2C). Metadata standardization is critical for harmonizing datasets, as it facilitates the integration of phenomic data with associated specimen metadata, such as taxonomy, collection locality, and ecological context. By adopting common metadata schemas and persistent identifiers (DOIs), researchers can link spectral data directly to digital databases, fostering seamless collaboration and data reuse. Experience gained from successful protocol standardization and data aggregation initiatives (e.g. Darwin Core and iDigBio; Wieczorek *et al.*, 2012; Soltis, 2017) can be leveraged to implement a strategy for herbarium spectroscopic data.

The development of cyberinfrastructure has been pivotal in enabling large-scale aggregation of spectral data. Platforms like iDigBio and GBIF provide centralized repositories for biodiversity data, but dedicated cyberinfrastructure for spectral datasets, integrated with existing platforms, will be essential for advancing collections-based research. These systems should support real-time synchronization of available data from herbarium institutions, cross-referencing, and retrieval for global accessibility. Any dedicated spectral cyberinfrastructure platforms will require Application Programming Interfaces (APIs) to enable researchers to query, retrieve, and contribute spectral datasets programmatically; facilitating the large-scale synthesis of data. Existing cyberinfrastructure developed specifically for spectral data and models, such as EcoSis and EcoSML (Wagner *et al.*, 2019), can be leveraged for herbarium data or used as models for developing new infrastructure.

Ensuring data quality controls is another foundational aspect of data aggregation. These controls include rigorous preprocessing of spectral datasets (e.g., noise removal, calibration) and standardization of scanning protocols to maintain the highest possible consistency across instruments, collections, and institutions. Routine validation processes will ensure that aggregated data meet the necessary standards for reproducibility and analysis. Finally, the implementation of analysis engines capable of handling high-dimensional datasets will be transformative. These engines should integrate spectral data with complementary datasets, such as genomic or spatial data, and provide tools for advanced modeling and visualization. Open-source analysis platforms with user-friendly interfaces will democratize access to these tools and foster collaboration across disciplines.

Scaling herbarium reflectance spectroscopy

As we consider the broader utility of herbarium specimens for estimating traits, the successful transfer of pressed-leaf models onto herbarium spectra offers a powerful method for reconstructing traits as they would have existed in vivo. This approach not only enhances the ecological relevance of trait predictions from herbarium specimens but also circumvents the need for destructive sampling of these invaluable collections. By preserving specimen integrity, reflectance spectroscopy provides a non-destructive, scalable, and integrative methodology for linking historical plant traits to modern ecological and evolutionary studies (Costa *et al.* 2018; Kothari *et al.* 2023).

Advancing collections-based spectroscopy for plant biology and global ecology is increasingly critical, particularly given the current vulnerabilities faced by herbaria and collection facilities. Despite their pivotal role in biodiversity research, herbaria often face devaluation and threats of closure (Thiers, 2024; Davis, 2024). At the same time, advances in digital technologies and biological data networks are unlocking unprecedented opportunities for their use (Meineke *et al.*, 2018; Lang *et al.*, 2019; Hedrick *et al.*, 2020; Bakker *et al.*, 2020; Heberling, 2022; Davis, 2023). Over the past fifteen years, targeted funding initiatives, such as those supported by the U.S. National Science Foundation, have facilitated the creation of comprehensive digital databases containing specimen images, metadata, and extended datasets like DNA sequences (Soltis, 2017). Platforms like iDigBio and GBIF now aggregate these resources into a global "Metaherbarium"—an integrated digital repository of plant diversity and distributions (Davis, 2023). This Metaherbarium is already driving transformative, global-scale research in biodiversity, ecology, and evolution, offering fresh insights into how plant life responds to environmental change (Meineke *et al.*, 2018; Davis, 2023).

As the momentum for leveraging collections continues to grow, the balance between innovative use and long-term preservation becomes increasingly vital (Davis *et al.*, 2024). The non-destructive nature of reflectance spectroscopy aligns perfectly with this goal. Spectral data files can be seamlessly integrated into digital specimen databases through persistent identifiers, linking these phenomic datasets directly to physical specimens and their associated metadata. By avoiding physical alteration, reflectance spectroscopy ensures the continued preservation of herbarium specimens for future research.

While functional trait prediction and taxonomic classification remain the most well-established applications of spectral reflectance, the future promises exciting new integrations with complementary datasets. In particular, the merging of spectral libraries with genomic, phenotypic, and spatial datasets offers unprecedented opportunities for addressing ecological and evolutionary questions. For example, these integrated datasets could facilitate species delimitation or provide novel insights into community assembly and functional biogeography. The combination of spectral reflectance libraries with "omics" and other synthetic datasets holds tremendous potential for generating new inferences about plant

738 diversity, distributions, and functional traits across temporal and spatial scales (Cavender-Bares *et al.*,
739 2017; Davis, 2023). Such advancements will continue to transform biodiversity science, bridging
740 historical collections with cutting-edge methodologies to address the pressing ecological and evolutionary
741 questions of our time.

References:

Abasolo M, Lee DJ, Raymond C, Meder R, Shepherd M. **2013**. Deviant near-infrared spectra identifies *Corymbia* hybrids. *Forest Ecology and Management* **304**: 121–131.

Bakker FT, Bieker VC, Martin MD. **2020**. Herbarium Collection-Based Plant Evolutionary Genetics and Genomics. *Frontiers in Ecology and Evolution* **8**.

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

Breiman L, Cutler A, Liaw A, Wiener M. **2024**. randomForest: Breiman and Cutlers Random Forests for Classification and Regression.

Bridson DM, Forman L, Royal Botanic Gardens K. **1998**. *The herbarium handbook*. Richmond, Surrey: Royal Botanic Gardens, Kew.

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, Couture JJ, Kaproth MA, Kingdon CC, Singh A, Serbin SP, Center A, Zuniga E, Pilz G, *et al.*2016. Associations of leaf spectra with genetic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity. *Remote Sensing* **8**: 17–17.

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

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.

Costa FRC, Lang C, Almeida DRA, 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.

Cotrozzi L, Couture JJ, Cavender-Bares J, Kingdon CC, Fallon B, Pilz G, Pellegrini E, Nali C, Townsend PA. **2017**. Using foliar spectral properties to assess the effects of drought on plant water potential. *Tree Physiology* **37**: 1582–1591.

Curran PJ. **1989**. Remote sensing of foliar chemistry. *Remote Sensing of Environment* **30**: 271–278.

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

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

Davis CC, Sessa E, Paton A, Antonelli A, Teisher JK. **2024**. Guidelines for the effective and ethical sampling of herbaria. *Nature Ecology & Evolution*: 1–8.

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.

Drobnic J. **2008**. Modern techniques of herbarium protection. *Scripta Facultatis Rerum Naturalium Universitatis Ostraviensis* **186**.

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

784 greenness of herbarium specimens as predictors for successful extraction and amplification of DNA.
785 *Blumea - Biodiversity, Evolution and Biogeography of Plants* **53**: 407–428.

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

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

792 **Funk JL, Larson JE, Ames GM, Butterfield BJ, Cavender-Bares J, Firn J, Laughlin DC, Sutton-**
793 **Grier AE, Williams L, Wright J**. **2017**. Revisiting the Holy Grail: using plant functional traits to
794 understand ecological processes. *Biological Reviews* **92**: 1156–1173.

795 **GBIF.org**. GBIF Home Page. *http://www.gbif.org/*

796 **Guzmán Q. JA**. **2024**. CWT: Continuous Wavelet Transformation for Spectroscopy.

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

800 **Heberling JM**. **2022**. Herbaria as Big Data Sources of Plant Traits. *International Journal of Plant*
801 *Sciences* **183**: 87–118.

802 **Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA,**
803 **Cook JA, Blackburn DC, *et al.*2020**. Digitization and the Future of Natural History Collections.
804 *BioScience* **70**: 243–251.

805 **Jetz W, Cavender-Bares J, Pavlick R, Schimel D, Davis FW, Asner GP, Guralnick R, Kattge J,**
806 **Latimer AM, Moorcroft P, *et al.*2016**. Monitoring plant functional diversity from space. *Nature Plants*
807 **2**: 1–5.

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

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

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

815 **Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, R**
816 **Core Team, *et al.*2024**. caret: Classification and Regression Training.

817 **Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB**. **2022**.
818 TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution* **39**:
819 msac174.

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

823 **Lang C, Costa FRC, Camargo JLC, Durgante FM, Vicentini A**. **2015**. Near infrared spectroscopy
824 facilitates rapid identification of both young and mature Amazonian tree species. *PLoS ONE* **10**:
825 e0134521–e0134521.

826 **Lang PLM, Willems FM, Scheepens JF, Burbano HA, Bossdorf O**. **2019**. Using herbaria to study

global environmental change. *New Phytologist* **221**: 110–122.

Liland KH, Mehmood T, Sæbø S. **2024a**. plsVarSel: Variable Selection in Partial Least Squares.

Liland KH, Mevik B-H, Wehrens R, Hiemstra P. **2024b**. pls: Partial Least Squares and Principal Component Regression.

Meineke EK, Davis CC, Davies TJ. **2018**. The unrealized potential of herbaria for global change biology. *Ecological Monographs* **88**: 505–525.

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

Meireles JE, Schweiger A, Cavender-Bares J. **2017**. Spectrolab: class and methods for hyperspectral data. *R package version 0.0* **2**.

Neto-Bradley BM, Bonnet P, Goeau H, Joly A, Cavender-Bares J, Coomes DA. 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.

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: Springer International Publishing, 43–82.

Soltis PS. **2017**. Digitization of herbaria enables novel research. *American Journal of Botany* **104**: 1281–1284.

Stasinski L, White DM, Nelson PR, Ree RH, Meireles JE. **2021**. Reading light: leaf spectra capture fine-scale diversity of closely related, hybridizing arctic shrubs. *New Phytologist* **232**: 2283–2294.

Thiers B. **2020**. Index Herbariorum: A worldwide index of 3,100 herbaria and 12,000 associated staff where a total of 390 million botanical specimens are permanently housed.

Thiers BM. **2024**. Strengthening Partnerships to Safeguard the Future of Herbaria. *Diversity* **16**: 36.

Wagner EP, Merz J, Townsend PA. **2019**. EcoSIS: A Spectral Library and the Tools to Use It. **2019**: B11F-2396.

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, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D, Brand T van den, *et al.*2024. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. **2012**. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE* **7**: e29715.

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.

869

870

871

Acknowledgements

We thank the curatorial and digitization staff at the Harvard University Herbaria (HUH) for past and present efforts to curate and digitize the Herbaria, with special thanks to Michaela Schmull and Lisa Standley for support and protocol development. Nico Gascon and Sydney Kaye assisted with specimen imagery. This project was funded by the NSF BII DBI-2021898 (JCB and JEM), the Harvard University Herbaria Postdoctoral Research Fellowship (DMW), and an NSF REU award 2150058 (JMR).

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.

Data Availability Statement

All analysis codes used in this study are publicly available at GitHub (github.org/Erythroxylum/herbarium-spectra). Spectral data have been deposited in the EcoSIS repository and can be accessed at https://ecosis.org/.