Title

Specimen preservation affects herbarium-derived leaf traits, reflectance spectra, and predictive models

Authors

Riley P. Fortier[1*], Ryan S. Fuller[1], and Matthew W. Austin[1]

Affiliations

1. Herbarium, Missouri Botanical Garden, St. Louis, MO, USA

[*]**Corresponding author contact information:**

Riley P. Fortier

Fortier.riley@gmail.com

503-729-0401

ORCIDs

Riley P. Fortier: https://orcid.org/0000-0003-4032-7676

Ryan S. Fuller: https://orcid.org/0000-0003-2457-0975

Matthew W. Austin: https://orcid.org/0000-0002-1231-9081

Summary

- Herbaria are increasingly used as sources of functional trait data and reflectance spectra used in models for trait prediction and species classification. However, it is currently unknown how leaf preservation protocols may affect these data.
- Using a factorial experimental design, we tested how alcohol preservation, drying intensity, and their interaction affected foliar traits (leaf mass per area (LMA), effective leaf width (ELW), thickness, area, length), leaf reflectance spectra, and predictive models. We also tested the effect on accuracy when predictive models are trained on spectra from one preservation treatment and tested on spectra from another treatment.
- Alcohol preservation significantly affected all leaf traits, causing greater shrinkage with drying, whereas drying intensity had minimal effects. Both alcohol and drying had marked effects on reflectance spectra and on the accuracy of predictive models. Transferring models between preservation treatments significantly reduced prediction accuracy for both traits and species, although this effect was modest for traits.
- Our study highlights the importance of considering specimen preservation in herbarium-based studies. By demonstrating the effect of preservation on models that use reflectance spectra to predict traits and classify species, our study can be used to refine predictive models built from herbarium-based spectra.

Key words

Ethanol, hyperspectral, leaf shrinkage, PLSR, PLS-DA, spectroscopy

Introduction

The world's herbaria house over 400 million preserved plant specimens (Thiers, 2025). For centuries, these have been vital in understanding global biodiversity through their core use in the fields of taxonomy and phylogenetics. More recently, herbaria have been recognized as vast repositories for data on functional traits (i.e., measurable, morpho-physiological characteristics that influence plant fitness (Violle *et al.*, 2007)), situating herbarium specimens as valuable resources for ecological research at broad spatiotemporal scales (Heberling & Isaac, 2017; Meineke *et al.*, 2018; Heberling, 2022). This value of herbarium specimens comes not only from direct functional trait measurements, but also from the prediction of traits using reflectance spectroscopy (Cavender-Bares *et al.*, 2025). A technique principally employed on fresh leaf material, reflectance spectroscopy has increasingly been shown to reliably predict traits and taxonomic identities of dried leaf material without requiring destructive sampling (Durgante *et al.*, 2013; Costa *et al.*, 2018; Kothari *et al.*, 2023; White *et al.*, 2025; Cavender-Bares *et al.*, 2025). However, the world's herbarium specimens were collected across multiple centuries by thousands of botanists using many different collection and preservation protocols (Kozlov *et al.*, 2021). Consequently, there is substantial uncertainty surrounding the effects of specimen preservation on herbarium-derived functional trait measurements, reflectance spectra, and the accuracy of models built from these spectra to predict traits and taxonomic identities (Heberling, 2022; Cavender-Bares *et al.*, 2025).

While the botanical collection process generally follows a standard set of steps – i.e., specimen sampling, pressing, and drying – the exact methodology varies by country, institution, and/or collector. For example, some botanists often put specimens in makeshift drying ovens directly in the field, which can be difficult to maintain at steady temperatures and require varying drying durations (Davies *et al.*, 2024). Without ready access to drying ovens, many botanists preserve plants in bundles of newspaper soaked with a diluted concentration of ethanol or another form of alcohol to prohibit the quick degradation of leaves by pathogens, a technique primarily used in the humid tropics (Hodge, 1947; Mori, 2011; Davies *et al.*, 2024). Once transported from the field to an herbarium, specimens undergo drying in ovens (if not fully dried already). Drying protocols can vary by institution and by botanist, with herbarium handbooks recommending oven temperatures anywhere between 25-60 °C and for durations of 12 hours to

multiple days (Lipsen & Tan, 2023; Davies *et al.*, 2024). Despite these variable collection practices, the effects of specimen preservation techniques on herbarium-based datasets is largely unknown. In other words, we do not fully understand how alcohol preservation and drying protocols may influence herbarium-derived functional trait measurements and reflectance spectra, highlighting an important gap in knowledge and suggesting a potential caveat in herbarium-based ecological research.

To reliably use herbarium-derived functional traits in ecological studies, we must understand how variable collection practices affect trait measurements of herborized tissue. One known effect of herborization is leaf shrinkage caused by drying (Blonder *et al.*, 2012; Perez *et al.*, 2020). Correction factors can be applied to model fresh leaf traits (Heberling, 2022), providing an estimation of a fresh trait value during the time of a specimen's collection. However, several factors influence the amount of shrinkage caused by drying. For example, leaves of evergreen species shrink less than those of deciduous species, as do leaves of woody versus herbaceous species (Blonder *et al.*, 2012). How the preservation of leaves in alcohol affects trait measurements and leaf shrinkage is less fully understood. One study suggested that preserving leaves in an alcohol treatment for seven weeks can lead to more shrinkage in leaf length and width after drying than non-treated leaves (Parnell *et al.*, 2013). Their study, however, only included a few temperate species, left plants in drying ovens for somewhat unusual durations (with some drying treatments of up to 15 weeks), and did not test for statistical significance between treatments. Because preserving leaves in alcohol is more common in the tropics (Mori, 2011) and typical drying protocols call for periods of days rather than weeks (Davies *et al.*, 2024), there are still major uncertainties about how alcohol and drying treatments affect trait measurements.

Particularly understudied is the effect of specimen preservation on foliar spectra, which can be used to train predictive models. Predicting traits and/or taxa from spectra involves training a multivariate statistical model to learn the relationship between reflectance spectra and a response variable such as a trait value (e.g. leaf thickness) or taxonomic rank (e.g. genus or species). The accuracy of trained models are then assessed by comparing model predictions to observed, ground truth data (Kothari *et al.*, 2023; White *et al.*, 2025; Hernández-Leal *et al.*, 2025). White et al. (2025) found that model accuracy tends to decrease with specimen age,

damage, and degradation, and when models are transferred between mounted specimens and unmounted leaf material (e.g., when a model is trained on mounted leaves and tested on unmounted leaves). However, there is a dearth of information on how preserving leaves in alcohol and different drying conditions affect raw reflectance values and the performance of predictive models. In particular, it is unknown whether a model trained on spectra from specimens preserved following one methodology (e.g., no alcohol and a moderate drying temperature) can be reliably transferred to specimens preserved following a different methodology (e.g., with alcohol and a high drying temperature), or alternatively, whether model prediction accuracy might decrease. So far, no study has evaluated how alcohol preservation in the field and variable lengths of drying ultimately affect reflectance spectra nor the reliability of spectral-based models in predicting traits and taxonomic names (Cavender-Bares *et al.*, 2025).

This study tested how different preservation techniques affected pressed and dried leaves. Specifically, we addressed three aims related to how alcohol preservation and drying protocols affect herbarium-derived traits:

1. **Leaf functional traits**: We hypothesized that alcohol preservation would affect functional traits, since a previous study suggested that the use of alcohol led to increased leaf shrinkage in drying ovens (Parnell *et al.*, 2013). This would result in larger measurements for some traits (e.g., LMA, leaf thickness) in alcohol-treated versus non-alcohol treated leaves, and smaller measurements for other traits (e.g., leaf area). We also hypothesized that drying would have no significant effect on leaf shrinkage.
2. **Reflectance spectra**: We hypothesized that both alcohol and drying would lead to differences in reflectance in the visible spectrum, as the use of alcohol is known to cause leaf discoloration in herbarium specimens and excessive drying can singe or burn leaves (Mori, 2011; Davies *et al.*, 2024). We also hypothesized that there would be differences in reflectance at higher wavelengths in the near infrared (NIR) and short-wave infrared (SWIR) regions, since reflectance within these regions is typically associated with functional trait measurements (Kothari *et al.*, 2023; White *et al.*, 2025).
3. **Predictive models**: We hypothesized that models predicting trait values and discriminating species from spectra would have comparable accuracy regardless of alcohol or drying treatment. We also hypothesized that models trained on leaf spectra

from one treatment (e.g., alcohol preservation) and tested on leaf spectra from another treatment (e.g., no alcohol preservation) would lead to decreased accuracy.

Our findings will inform future studies and protocols that use herbarium specimens for functional trait measurements and spectral-based predictive modeling. As a result, this study will help to refine downstream uses of herbarium specimens, ensuring that herbaria can continue to be used as vast data repositories for taxonomic, ecological, and evolutionary research.

Methods:

Botanical Collection and Experimental Design:

In September 2025 and January 2026, we collected leaves from plants in the living collection of the Missouri Botanical Garden (MBG) in St. Louis, Missouri, USA (38° 36' 45" N, 90° 15' 33" W). We collected from a total of 33 trees representing 25 species, which included 16 trees belonging to 8 temperate species and 17 trees belonging to 17 tropical species (**Table S1**). Collections were made from temperate species growing in the outdoor collection and from tropical species growing in MBG's climate-controlled tropical plant conservatory. The temperate species were chosen because they commonly occur in natural forests throughout their respective ranges and are taxa regularly featured in ecological research. We were more limited with our selection of tropical species due to collecting in a conservatory, so we chose plants that had ample material from which to collect and that represented taxa commonly found in tropical forests. For each tree, we targeted healthy branches near to the edge of crowns and then selected 30 healthy, pre-senescent leaves with minimal damage or pathogens.

The 30 leaves from each tree were randomly divided into groups of five leaves, and each group was assigned to one of six experimental treatments in a 2 x 3 factorial design (**Fig. S1**). This resulted in 990 leaves distributed among the six treatments. The first factor was a drying treatment, with the control representing a standard drying protocol (D_C: between 24-48 hours in a drying oven or until leaves reached a constant mass at ~32° C) and a treatment representing an extended drying protocol (D_E: one week in the oven at ~55° C). The second factor was an alcohol (ethanol) treatment with a control (A_C: no alcohol preservation), preservation in ~50% alcohol (A_{50}), and preservation in ~80% alcohol (A_{80}). For the alcohol treatments, we loosely

pressed leaves in newspaper, stored them in plastic bags, and applied enough alcohol to saturate, but not submerge, all contents, following Mori (2011). We left leaves preserved in alcohol for between 2-3 weeks prior to drying. We attempted to replicate field conditions as closely as possible for the alcohol treatments. The six treatment combinations are coded as illustrated in **Fig S1**.

Trait measurements:

We measured functional traits on fresh leaves immediately after field collection and again after drying. We chose traits that were easy to measure on herbarium specimens, that relate to plant fitness, and that are known to respond to environmental change (Wright *et al.*, 2004). For these reasons, they are traits that are commonly used in herbarium-based ecological studies (Heberling, 2022). For both fresh and dry leaves, we scanned the adaxial side of each leaf with a digital flatbed scanner (Canon CanoScan LiDE 100) and used ImageJ (Schindelin *et al.*, 2012) to measure leaf area, length, and effective leaf width (ELW). ELW is the diameter of the largest circle that can be inscribed within the margins of a leaf and is an important trait related to leaf thermoregulation (Leigh *et al.*, 2017). We also measured leaf thickness with a Mitutoyo Digital Micrometer (0.001 mm resolution) three times on the lamina of each leaf – near the base, middle, and tip – between the midvein and margin, avoiding large veins when possible. We took each leaf's mean thickness for subsequent analyses. We excluded one species, *Melastoma beccarianum*, from all analyses pertaining to leaf thickness due to its dense venation prohibiting accurate measurements. On dried leaves only, we massed each leaf with a U.S. Solid Analytical Balance (0.1 mg resolution) to calculate leaf mass per fresh area (LMA_{fresh}) and leaf mass per dry area (LMA_{dry}) expressed in g m^{-2}. Petioles were excluded from all trait measurements, and protocols follow Pérez-Harguindeguy et al. (2013). To evaluate how well our trait measurements capture the global range of plant traits, we compared our measurements to those from the TRY plant trait database (Kattge *et al.*, 2020).

Spectral measurements and processing:

Spectral reflectance was measured on dried leaves using a Spectral Evolution NaturaSpec Spectroradiometer (wavelength range 350-2500 nanometers (nm)) with a handheld surface probe. The surface probe's field of view is a 12x15 mm diameter ellipse. Before each measurement session, the spectroradiometer was powered on for at least 15 minutes in order to

reach an optimal operating temperature. In order to ensure accurate reflectance function, we scanned a white Spectralon reference standard. Spectral measurements followed a modified version of the International Herbarium Spectral Digitization (IHerbSpec) protocol (Cavender-Bares et al. 2025, IHerbSpec 2026). We took two measurements per leaf, one on the adaxial surface and one on the abaxial surface, targeting clean areas of the lamina that were free of damage or pathogens. All foliar spectral measurements were taken with a black background (< 4% reflectance) beneath the leaf. Spectra were trimmed to retain wavelengths 450-2400 nm to reduce noise at the ends of spectra (Burnett *et al.*, 2021; Ji *et al.*, 2024). All subsequent data analyses were performed in R statistical software (R Core Team, 2024).

Data analysis – Measured traits:

We first analyzed how each treatment affected leaf functional traits. For each trait, we calculated the ratio of each leaf's dry measurement to its fresh measurement. Accordingly, values <1 indicate that the dry measurement was less than the fresh measurement, and vice versa. We then performed a two-way analysis of variance (ANOVA) on each trait to determine if there were significant differences in these ratios between treatments. We then calculated Tukey's Honest Significant Differences (HSD) to identify which treatment combinations were significantly different from one another.

To further explore how drying and alcohol treatments affected dry trait measurements, we ran a generalized linear mixed model (GLMM) on each leaf trait. We modeled dry trait measurements as a function of fresh trait measurements, with drying and alcohol treatments as fixed effects. We also included the two-way interactions between fresh trait measurement and both preservation treatments to test if each treatment has a greater effect on larger leaves than on smaller leaves, or vice versa. Individual nested within species was a random effect. We then evaluated how the amount of trait variation resulting from preservation treatments compares to that which occurs naturally within a species using variance component decomposition (VCD). To do the VCD, we built additional GLMMs for each trait with dry measurement as the response and fresh measurement as a fixed effect, this time including drying and alcohol treatments as random effects instead of fixed effects. We then used the *VarCorr* function from the R package 'lme4' (Bates *et al.*, 2015) to calculate the amount of variance explained by the random effects relative to the total amount of variance within a model.

Data analysis – Reflectance spectra:

To evaluate treatment effects on reflectance spectra, we performed a two-way ANOVA at each wavelength, testing whether alcohol, drying, or their interaction significantly affected mean reflectance. We assessed significance after adjusting p-values using the Benjamini-Hochberg method for multiple comparisons (Benjamini & Hochberg, 1995). To further elucidate differences between the two alcohol treatments (A_{50} and A_{80}), we ran another ANOVA at each wavelength without the interaction term and performed pairwise contrasts to discern how spectra from each of the two alcohol treatments compare to the control and to each other. As before, we adjusted these p-values using the Benjamini-Hochberg method.

Since reflectance spectra represent highly dimensional multivariate data, with reflectance at a given wavelength being dependent on surrounding wavelengths, we also ran a permutational multivariate ANOVA (PERMANOVA) to test for treatment effects on the overall structure of spectra. We first calculated pairwise Bray-Curtis dissimilarity between each pair of spectra to generate a dissimilarity matrix. We used this dissimilarity matrix in the PERMANOVA to test the effects of alcohol treatment, drying treatment, and their interaction across 999 permutations and stratified permutations by individual tree. Then, to test if a significant PERMANOVA was due to differences in means (centroids) or differences in dispersion (mean distance to centroid) between groups, we tested for homogeneity of dispersion via PERMDISP. PERMDISP tests the null hypothesis that dispersion within groups is equal. Finally, we ran a non-metric multidimensional scaling (NMDS) to visualize spectral differences between treatments. The NMDS was run in two dimensions which yielded a stress value of 0.09, signifying that the two dimensions were sufficient to adequately capture the multivariate space encompassing all reflectance spectra. We used the R package ‘vegan’ (Oksanen *et al.*, 2025) for the PERMANOVA and NMDS analyses.

Data analysis – Trait prediction using PLSR:

We used partial least squares regressions (PLSR) to evaluate how accurately reflectance spectra can be used to predict leaf traits. In PLSR training and evaluation, we used the first and second derivatives of spectral reflectance, based on preliminary tests yielding better model results compared to raw reflectance. These models were built only for leaf thickness and LMA, traits that previous studies have shown PLSR models to predict with moderate to high accuracy

(Costa *et al.*, 2018; Kothari *et al.*, 2023; White *et al.*, 2025; Hernández-Leal *et al.*, 2025). For both traits, we built models to predict only dry measurements.

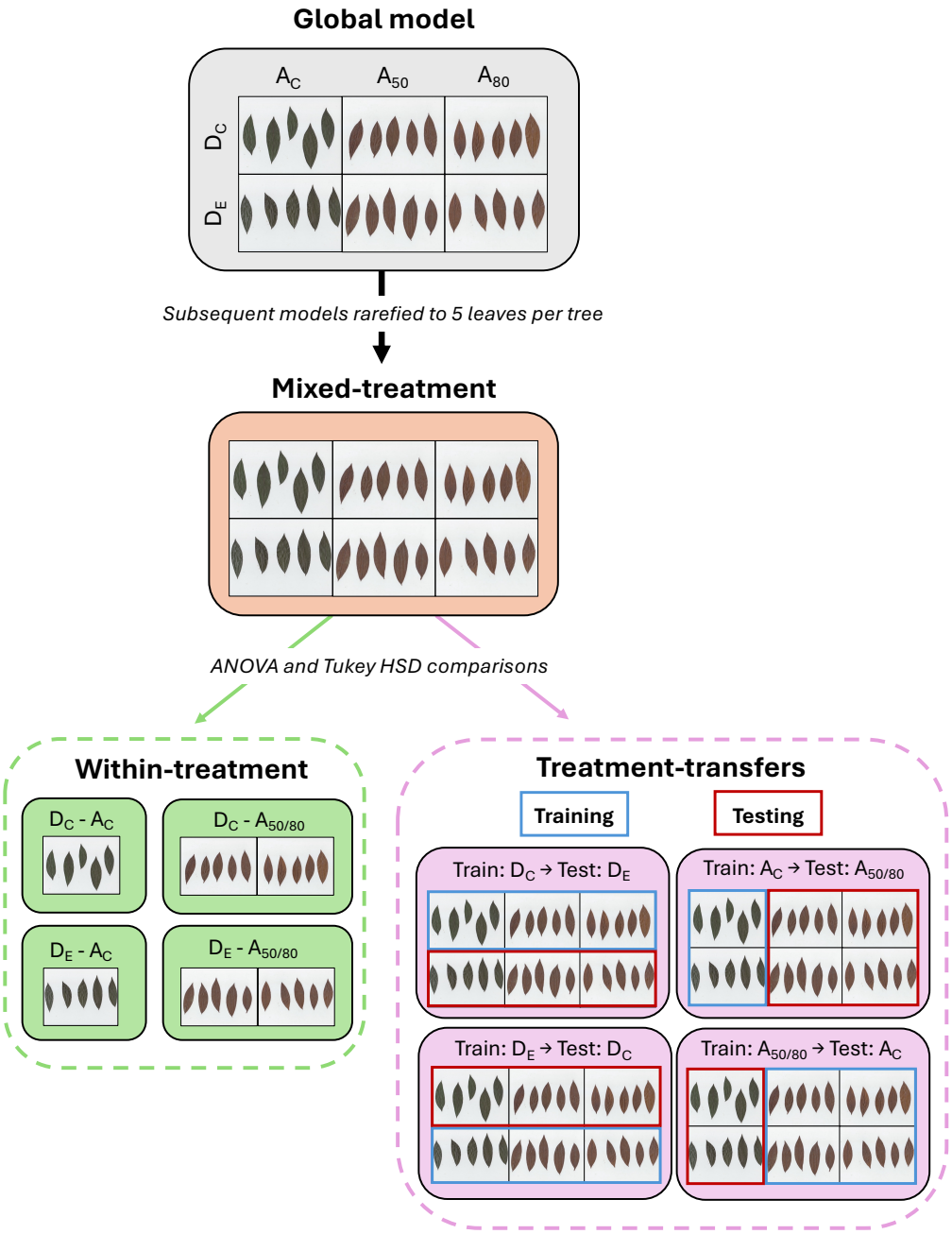
For both leaf thickness and LMA, we first fit a global model, using all leaf spectra across all treatments. Data were split into training (60%) and testing (40%) sets, stratified by species, ensuring that each treatment combination for a given tree would not be split between training and testing. We fit the model using the minimum number of components for which the root mean squared error (RMSE) of prediction from 10-fold cross-validation fell within one standard deviation of the global minimum. We ran 100 model iterations, and performance was evaluated by calculating the mean and standard deviations of R^2, bias, RMSE, and the percentage RMSE (%RMSE) across all 100 iterations (Burnett et al. 2021).

To understand how alcohol and drying treatments affect model performance, we compared the global model to models restricted to leaves from single treatment combinations ("within-treatment" models, described below). Because larger training datasets inherently yield higher model accuracy (Burnett et al. 2021), directly comparing the global model (which used all available data) to smaller within-treatment models is not appropriate. Therefore, we ran a reduced version of the global model that rarefied the training data to match the sample size of the smallest within-treatment model by randomly selecting five leaves per tree within each of 100 model iterations prior to data splitting. We refer to this rarefied version as the "mixed-treatment" model, as it retains leaves from across all treatment combinations but at a reduced sample size. As with the global model, 60% of the data went into training and 40% into testing, stratified by individual tree.

We then built "within-treatment" PLSR models for each treatment combination separately. Due to finding minimal differences between the two alcohol treatments for both traits and reflectance spectra (see Results), we grouped the two alcohol treatments together for these and subsequent PLSR models (hereafter $A_{50/80}$). Two treatment combinations consisted of only five leaves per tree (D_C-A_C and D_E-A_C). To enable comparison between models, for the other two treatment combinations, we randomly selected five leaves per tree prior to splitting. To test for treatment effects on model accuracy, we performed two-way ANOVAs on the mean R^2 and %RMSE values for the mixed- and within-treatment models, then calculated Tukey's HSD to investigate significant differences in model performance.

Next, we built "treatment-transfer" models that tested the accuracy of models trained on spectra from one treatment and used to predict traits using spectra from another treatment. Specifically, we built alcohol treatment-transfer models (trained on the A_C treatment and tested on the $A_{50/80}$ treatment, and vice versa) and drying treatment-transfer models (trained on the D_C treatment and tested on the D_E treatment, and vice versa). In each model, we randomly selected five leaves per tree while ensuring that 60% of these were from the treatment included in the training set (e.g. the A_C treatment) while 40% were from the treatment included in the testing set (e.g. $A_{50/80}$ treatment). This resulted in a total of four treatment-transfer models for each trait. We evaluated differences between treatment-transfer models and compared their accuracy to the mixed-treatment model by performing ANOVAs and calculating Tukey's HSD as described above. **Fig. 1** illustrates the modeling workflow, including leaf selection for each model and model comparisons.

Figure 1: Conceptual diagram of the modeling framework. Each box filled with a solid color represents either a global, mixed-treatment, within-treatment, or treatment-transfer model. The modeling framework was run as PLSR models for predicting LMA and leaf thickness, and as PLS-DA models for predicting species. PLSR models were iterated 100 times while PLS-DA models were iterated 25 times. Within-treatment and treatment-transfer models combined A_{50} and A_{80} treatments. When there were more than five leaves per tree, the mixed-treatment, within-treatment, and treatment-transfer models each randomly sampled five leaves per tree for model training and evaluation to standardize sample sizes and enable direct comparisons between models.



303

304

305 *Data analysis – Species discrimination using PLS-DA:*

306 We trained models to discriminate species based on leaf spectra using partial least squares
307 discriminant analysis (PLS-DA), starting with a global model. We divided the data into training
308 (60%) and testing (40%) sets, stratified by species, ensuring that no treatment combination for a

given tree would be split between training and validation. We trained models using 10-fold cross-validation repeated three times. Since some species in the dataset were represented by a greater number of samples than others, which can bias classification algorithms (Sun *et al.*, 2009), we performed a two-step process to balance size classes. We first downsampled the data so that each species was represented by the minimum number of samples for a species. We then chose the number of components (*n*) that maximized Cohen's kappa (κ). Next, we upsampled the data so that species with fewer samples would match the maximum sample size of a species and restricted the range of components tested to be no more than *n*. We used the calibrated model from the upsampling step to classify species in the testing set. We ran 25 model iterations and assessed model performance using the mean classification accuracy and mean κ across all iterations.

We then performed mixed-treatment, within-treatment, and treatment-transfer PLS-DA models using the exact same workflow as the PLSR modeling (**Fig. 1**). That is, we randomly selected five leaves from each tree prior to data splitting, trained and tested PLS-DA models on these rarefied datasets, and compared the models' classification accuracy and κ using an ANOVA and Tukey's HSD. PLSR and PLS-DA models were built using the R packages 'caret' and 'pls' (Kuhn, 2008; Liland *et al.*, 2026).

Results

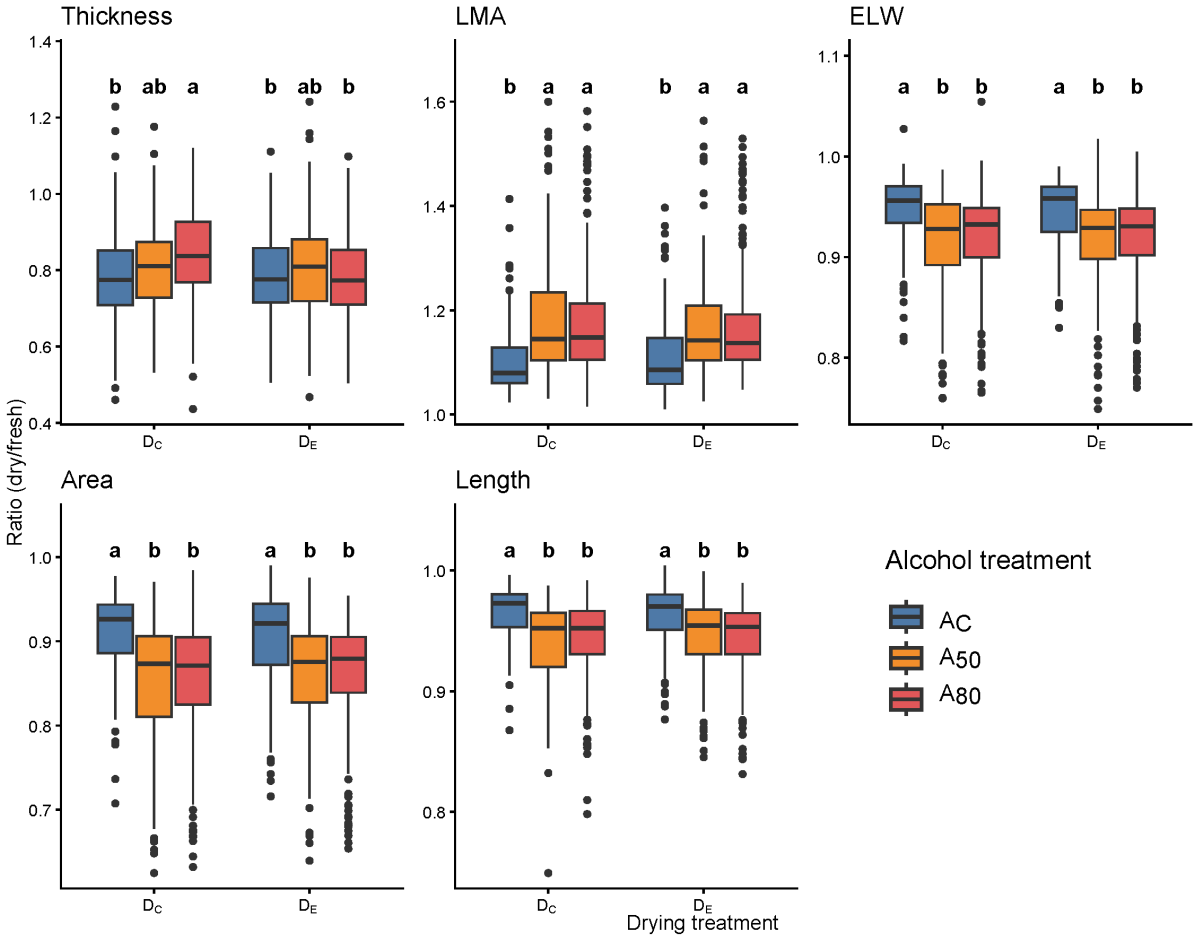
Leaf traits

Our fresh trait measurements cover a substantial portion of global plant trait space as reported on the TRY plant trait database (Kattge *et al.*, 2020) for all traits except for leaf area (**Table S2**). However, our trait values do not cover the higher end of global trait space, especially for leaf area, length, and ELW.

On average, dry measurements of leaf thickness and LMA were higher while ELW, area, and length were lower than their respective fresh measurements (**Fig. 2**). ANOVAs revealed significant effects of drying and alcohol treatments on the ratio of each trait's dry to fresh measurement. Specifically, there were significant main and interactive effects of drying and alcohol on leaf thickness (**Fig. 2 & Table S3**). Tukey's HSD revealed that leaf thickness was significantly higher in the D_C-A_{80} treatment than in the D_C-A_C, D_E-A_C, or D_E-A_{80} treatments. For

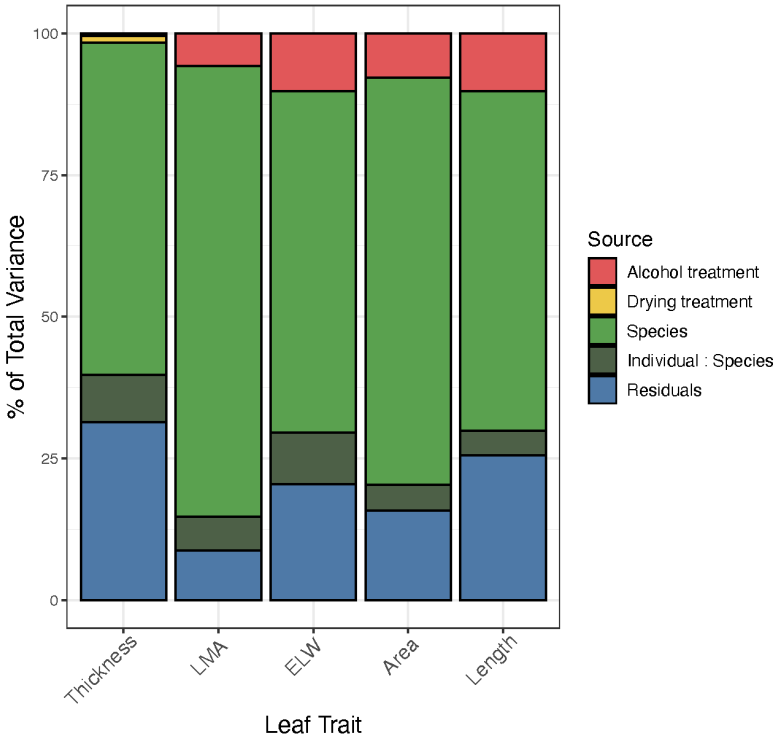
all other leaf traits, there was only a significant main effect of alcohol, which led to increased LMA and decreased ELW, leaf area, and leaf length (**Fig. 2, Table S3**). Tukey's HSD showed no significant differences between A_{50} and A_{80} treatments for all traits except for leaf thickness (**Fig. 2**).

Figure 2: Ratios of dry to fresh measurements for leaf thickness, leaf mass per area (LMA), effective leaf width (ELW), leaf area, and leaf length, illustrating the effects of drying and alcohol treatments. Bars are the medians, boxes are the inter-quartile ranges (IQR), and whiskers are 1.5 * IQR. Dots represent outliers. Boxes are color coded based on alcohol treatment. In each panel, letters indicate significantly different groups as determined by a two-way ANOVA and Tukey's HSD.



GLMMs revealed dry measurements to be highly correlated with fresh measurements for all traits (**Fig. S2, Table S4**). For all traits, there was a significant effect of alcohol treatment and no effect of drying treatment on dry trait measurements. Meanwhile, there were significant interactions between fresh trait measurements and alcohol treatments for all traits except leaf length, indicating that alcohol preservation had a pronounced effect on larger leaves. There was no significant interaction between fresh trait measurement and drying treatment for any trait (**Table S4**). When preservation treatments were included as random effects in a GLMM for each trait, VCD showed that species contributed the majority of each model's variance (>58%), followed by residuals contributing the next largest portion (8-31%). Alcohol treatment and individual each contributed between 4%-9%, while drying treatment generally contributed the least amount of variance (<2%) (**Fig. 3, Table S5**).

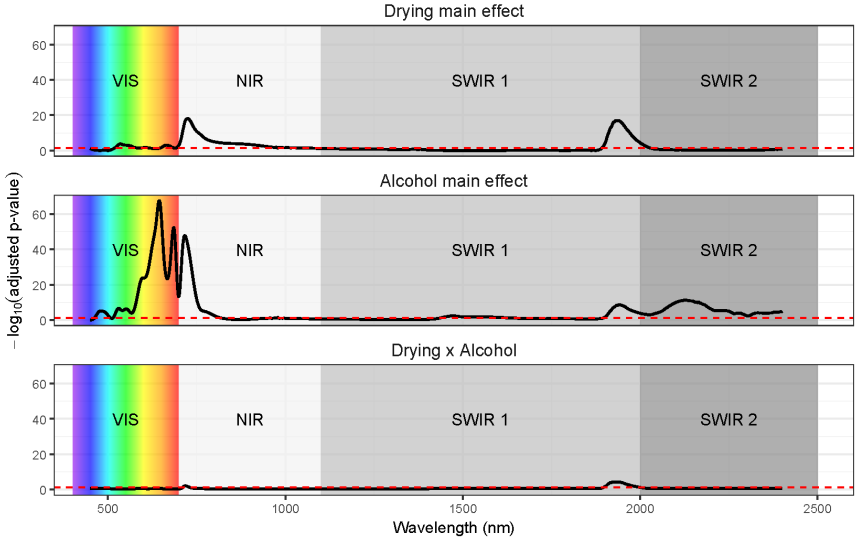
Figure 3: Amount of variance explained by random effects and residuals in generalized linear mixed models (GLMMs) for five leaf traits. One GLMM was run for each trait. Sources of variance are color coded.



Reflectance spectra

Leaf reflectance spectra displayed significant differences between treatments at certain regions throughout the electromagnetic spectrum. Drying and alcohol treatments both significantly affected reflectance in nearly the entire visible region between 450-700 nm, in the NIR region between 700-815 nm, and in the shortwave infrared SWIR region between 1900-2030 nm. There was also a significant effect of drying in the NIR region between 815-1075 nm and of alcohol in the SWIR region between 1440-1620 nm and between 2030-2400 nm. The interaction between drying and alcohol had a relatively weak but significant effect in the SWIR region between 1900-2000 nm (**Figs. 4, S4**). Pairwise contrasts revealed weak but significant spectral differences between A_{50} and A_{80} in a portion of the visible region (**Fig. S5**).

Figure 4: Effects of drying, alcohol, and their interaction on leaf reflectance from 450-2400 nm. Black lines are curves showing the p-values of a two-way ANOVA performed at each wavelength. P-values were adjusted for multiple comparisons and -$\log_{10}$ transformed for visualization. The dashed red line indicates a p-value corresponding to 0.05, where everything above the line is a significant effect (p-value <0.05) of drying (top), alcohol (middle), or the interaction between drying and alcohol (bottom). Plots are annotated with four different wavelength regions: visible spectrum (VIS), near infrared (NIR), short-wave infrared 1 (SWIR 1), and short-wave infrared 2 (SWIR2).



A PERMANOVA revealed a significant effect of alcohol, drying, and their interaction on reflectance spectra (p-values < 0.001) (**Table S6**). The PERMDISP was significant for both drying and alcohol, indicating a greater dispersion of spectra in multivariate space for the two groups relative to the control (**Table S7**). This was supported by the NMDS which revealed considerable overlap between spectra, but with greater variability in the drying and alcohol treatments than in the control, as evidenced by their wider distributions along the first NMDS axis (**Fig. S6**). Spectra of one species, *Piper jacquemontianum* (Piperaceae), were generally dissimilar to the spectra of all other species.

Trait prediction using PLSR

The global PLSR model showed high accuracy in predicting LMA (R^2 = 0.88, %RMSE = 8.79) and leaf thickness (R^2 = 0.83, %RMSE = 9.68) (**Fig. 5, Table S8**). The mixed-treatment model, despite a much lower sample size, also had high accuracy for predicting both LMA (R^2 = 0.86, %RMSE = 9.5) and leaf thickness (R^2 = 0.81, %RMSE = 10) (**Fig. S7, Table S8**). The within-treatment models all had significantly higher accuracy than the mixed-treatment model. Among these, all models had $R^2 \geq 0.82$ and %RMSE ≤ 9.7 for both leaf traits, with the most accurate model being from the D_C-A_C treatment combination (**Figs. 6 & S8, Table S8**). Treatment-transfer models were less accurate than the mixed- and within-treatment models (**Figs. 7 & S9, Table S9**). Among the treatment-transfer models, the highest accuracy was for predicting LMA using the D_E treatment for training and the D_C treatment for testing (R^2 = 0.82, %RMSE = 11.1), while the lowest accuracy was for models predicting leaf thickness using the $A_{50/80}$ treatment for training and the A_C treatment for testing (R^2 = 0.61, %RMSE = 22).

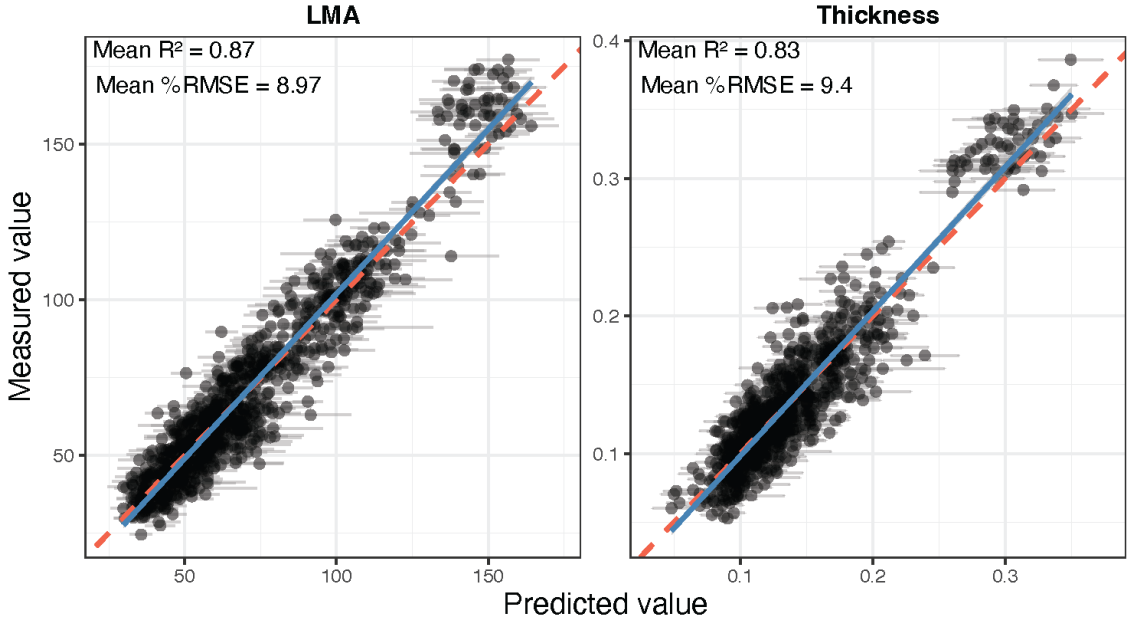
Species classification using PLS-DA

The global PLS-DA model showed moderate accuracy in classifying species (accuracy = 0.63, κ = 0.62) (**Fig. 5, Table S10**). As expected due to a smaller sample size, the mixed-treatment model had slightly lower accuracy (accuracy = 0.58, κ = 0.56) (**Fig. S10, Table S10**). The within-treatment models all had high classification accuracy (**Figs. 6 & S11, Table S10**). The most accurate model was for the D_E-$A_{50/80}$ treatment (accuracy and κ = 0.89) and was significantly higher than the other within-treatment models which had accuracy between 0.82-

0.86 and κ between 0.81-085 (**Fig. 6**). Treatment-transfer models showed considerably lower accuracy than each of the mixed- and within-treatment models (**Figs. 7 & S12, Table S11**). Among the treatment-transfer models, the two models with the lowest classification accuracies were both of the alcohol treatment transfers, with accuracy ≤ 0.35 and κ ≤ 0.32. An ANOVA and Tukey's HSD found these to perform significantly worse than either of the drying treatment transfers (**Fig. 7**).

Figure 5: Results of the global predictive models. A) Scatterplots for the PLSR models for predicting LMA (left) and leaf thickness (right). Each dot is a leaf, with horizontal lines indicating the standard error of predictions across 100 model iterations. The blue lines indicate the relationship between predicted and measured trait values according to an ordinary least-squares regression. The dashed red line indicates a 1:1 relationship. B) Confusion matrix for the PLS-DA model for classifying species. Rows represent true species identities while columns represent predicted species identities. Squares and numbers on the diagonal represent the percentage of correct classifications for each species. Squares and numbers off the diagonal represent misclassifications by the model.

A) PLSR trait prediction



B) PLS–DA species classification

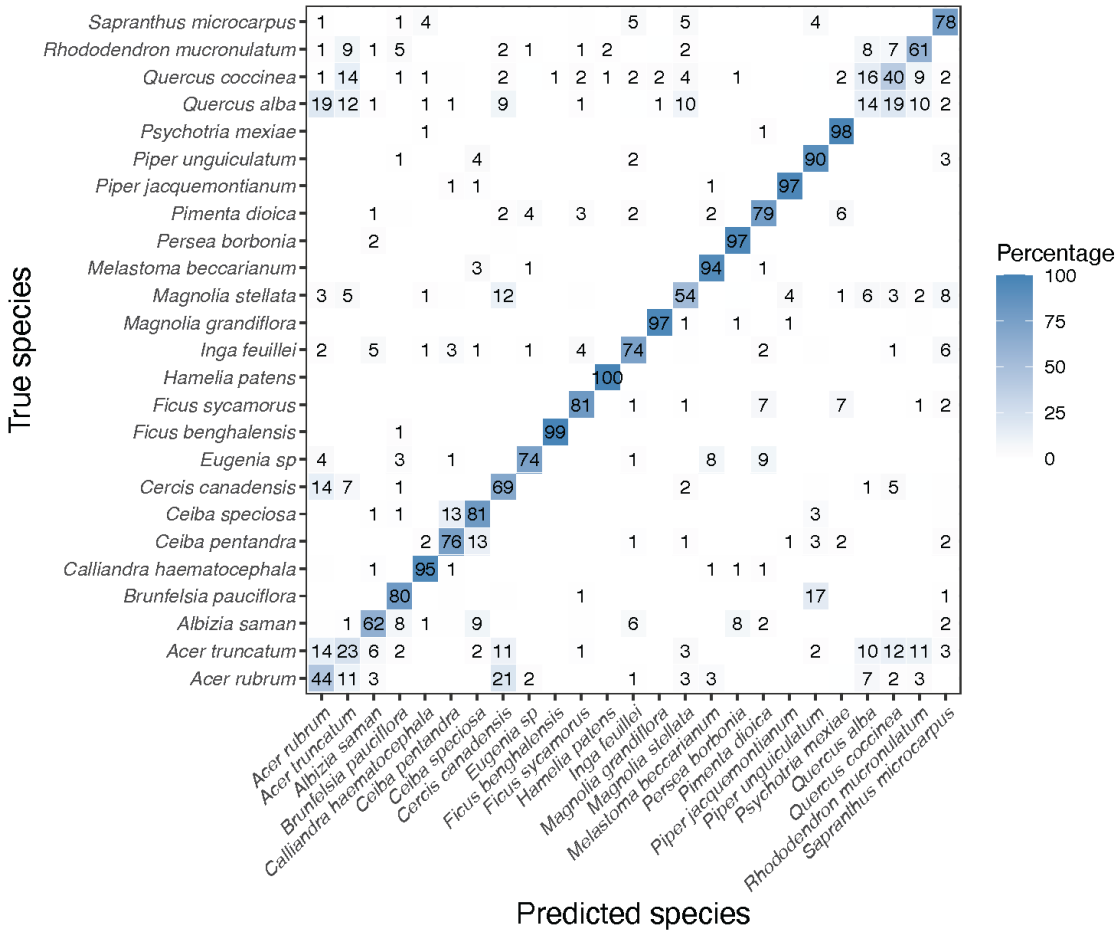


Figure 6: Mixed and within-treatment model performances. Each model is color coded. **A)** Shows R^2 and %RMSE across 100 iterations of PLSR models to predict LMA (left) and leaf thickness (right). **B)** Shows mean accuracy and Cohen's kappa across 25 iterations of PLS-DA models. Letters indicate statistically significant groups according to a two-way ANOVA and Tukey's HSD post-hoc test performed separately for LMA and leaf thickness predictions and species classification.

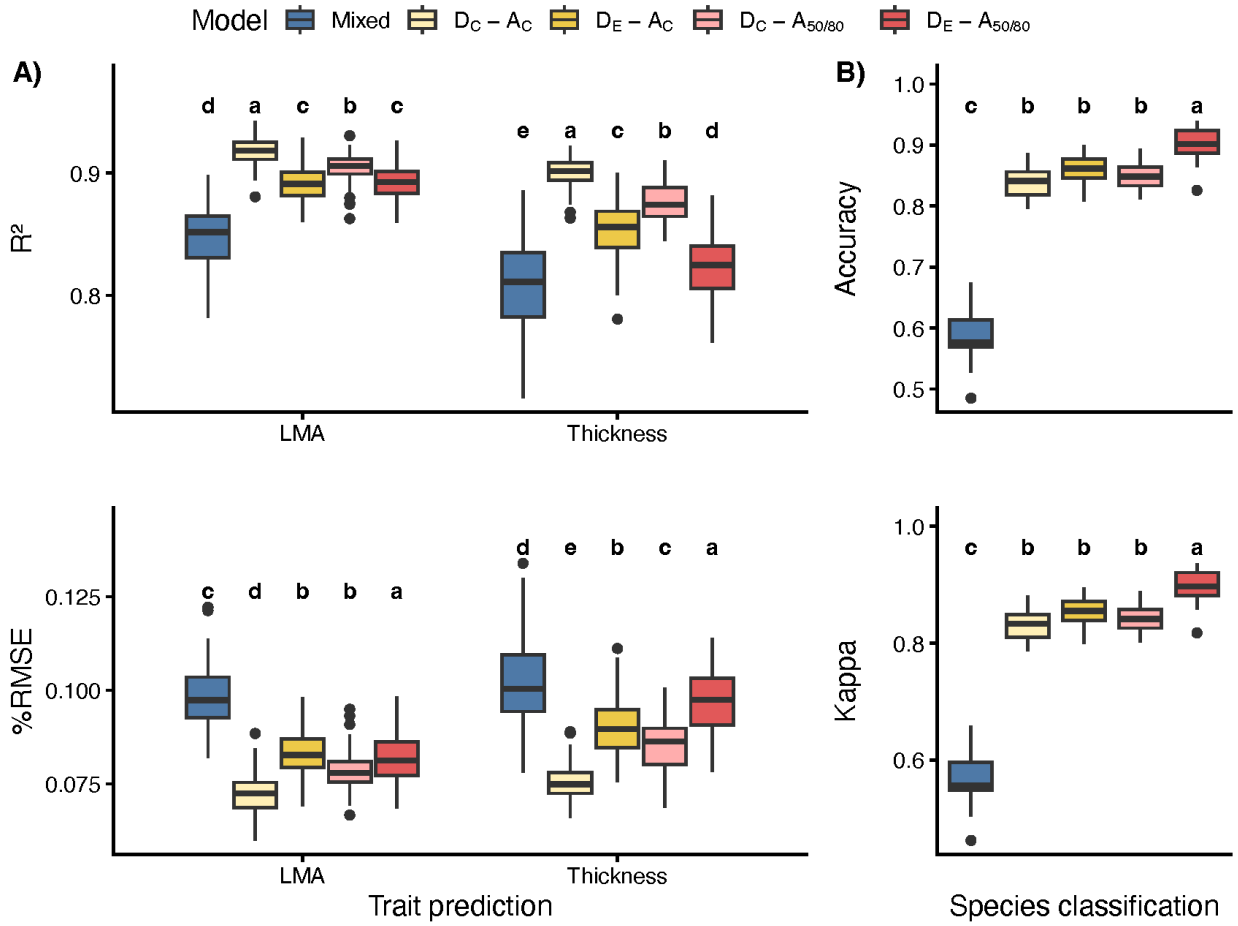
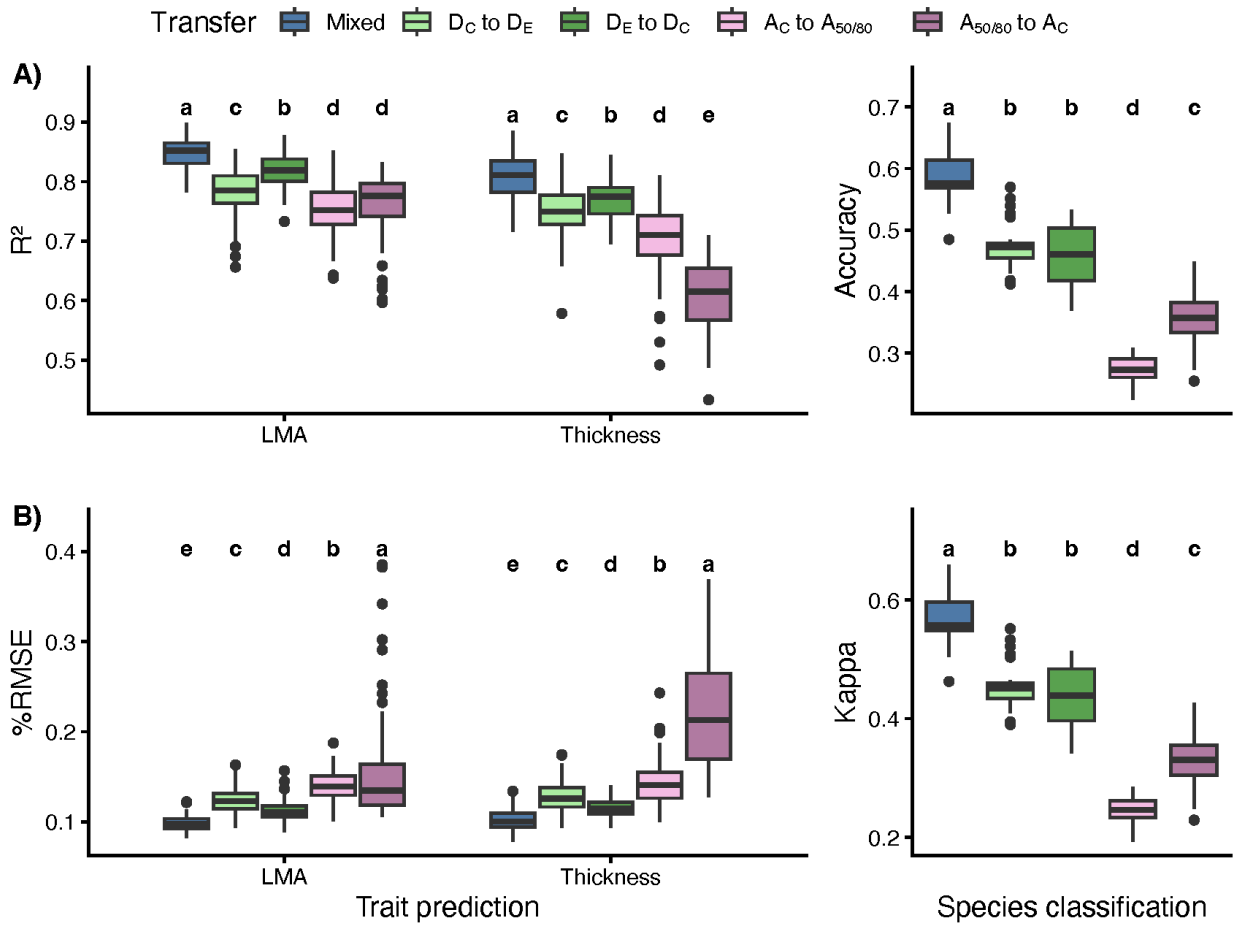


Figure 7: Mixed and treatment-transfer model performances. Each model is color coded. **A)** Shows R^2 and %RMSE across 100 iterations of PLSR models to predict LMA and leaf thickness. **B)** Shows mean accuracy and Cohen's kappa across 25 iterations of PLS-DA models. Letters indicate statistically significant groups according to a two-way ANOVA and Tukey's HSD post-hoc test performed separately for LMA and leaf thickness predictions and species classification.



Discussion

Herbarium specimens are invaluable resources for taxonomic, ecological, and evolutionary studies. Indeed, they underpin our current understanding of plant species descriptions and enable the investigation of species occurrences and functional traits across broad spatiotemporal scales (Queenborough, 2017; Perez *et al.*, 2020; Heberling, 2022). Reflectance spectroscopy extends and strengthens the value of herbarium specimens, allowing

for the prediction of functional traits and aiding in taxonomic discrimination from historical collections (Cavender-Bares *et al.*, 2025). However, it was previously unknown how the diverse preservation histories of specimens affect herbarium-derived functional trait measurements and leaf reflectance spectra. Here, we showed that drying and alcohol treatments significantly affected leaf traits and spectral reflectance and also had marked effects on predictive models that use spectra to predict traits and to classify species.

Preservation effects on leaf traits and spectra

Alcohol treatments affected leaf traits more than drying treatments. Specifically, alcohol preservation led to increased leaf shrinkage, as evidenced by higher dry to fresh ratios of leaf thickness and LMA and lower ratios of ELW, leaf area, and leaf length (**Fig. 2**). These results corroborate Parnell et al. (2013) who found similar patterns for leaf and floral morphological measurements. We found that the effect of alcohol increases with leaf size, as indicated by the significant interaction between fresh trait measurement and alcohol treatment in a GLMM for most traits (**Table S4**). However, the mechanisms behind this effect are unclear. It may be due to ethanol's degradation of leaf structural compounds (e.g., lignin or cellulose), which would jeopardize the structural integrity of leaves and lead to increased leaf shrinkage in drying ovens. The shrinkage effects of ethanol have been observed in biomass refineries, where a pre-treatment of ethanol is commonly applied to hasten the conversion of lignin and cellulose to fuels (Lancefield *et al.*, 2017). Because we did not directly measure lignin, cellulose, or other structural compounds, we could not directly test for this effect in an herbarium setting, so future research would benefit from investigating the effect of alcohol on such traits.

There were clear effects of alcohol on herbarium-derived leaf traits. However, are these effects large enough to dissuade researchers from utilizing herbarium specimens in plant functional trait analyses? In our study, the amount of trait variation explained by drying and alcohol treatments was less than the amount of variation within a species (**Fig. 3, Tables S1, S5**). Specifically, when included as random effects in GLMMs for each trait, alcohol and drying treatments consistently explained less variance in each model than the model's residuals. The residuals likely resulted from unaccounted leaf-level trait variation in the model, which could be caused by environmental or stochastic effects. Therefore, the effect of specimen preservation on leaf traits is likely small when compared to naturally occurring intraspecific trait variation.

Indeed, factors such as light exposure and ontogeny can generate large amounts of intraspecific trait variation (Martin *et al.*, 2020; Barton, 2024). Because many of our study species were represented by a single individual tree, the effects of alcohol and drying may be further reduced when working with many herbarium specimens per species. Specifically, trait variation due to environmental effects would be magnified with additional sampling of herbarium specimens, whereas trait variation due to different methods of leaf preservation would be less magnified. As a result, herbarium-derived trait datasets likely include greater natural trait variation than variation induced by alcohol preservation or drying protocols. Still, the effects of alcohol and drying should not be discounted as they introduce an additional dimension of trait variation in herbarium-based studies. For these reasons, researchers should be mindful of the level of precision needed for their research question and, when comparing two sets of herbarium specimens with distinct preservation histories, consider how preservation methods may affect the results (Perez *et al.*, 2020). In summary, we argue that the effect of preservation on trait measurements is minimal, and while researchers measuring traits from herbarium specimens should acknowledge this modest range of uncertainty, we advocate for the continued use of herbaria as sources of trait data.

There are many traits that we did not measure and that could also be affected by alcohol or drying treatments. Beyond the potential effect of alcohol on leaf structural compounds, alcohol may affect additional traits such as nutrient concentrations, secondary metabolites, or pigments such as carotenoids or anthocyanins. Yet, these effects remain unknown, highlighting the need for continued research on leaf preservation effects on functional and chemical traits.

Just as preservation treatments affected functional traits, there were significant treatment effects on leaf reflectance spectra. Both drying and alcohol treatments significantly affected wavelengths in nearly the entire visible region, along with affecting smaller sections within the NIR and SWIR regions (**Figs. 4, S3-S5**). Alcohol's role in discoloration of herbarium specimen leaves has long been acknowledged (Mori, 2011; Davies *et al.*, 2024), but until now had not been quantified. Although the mechanisms driving these results are unclear, the significant treatment effects in the NIR and SWIR regions could relate to their effects on leaf traits themselves. A previous study found the same wavelengths that were affected by alcohol and drying treatments in our study (~750 nm and ~1900 nm) to be important regions in predicting LMA, leaf structural

compounds, and the concentrations of several micronutrients from reflectance spectra (Kothari *et al.*, 2023). We found that alcohol significantly affected LMA measurements, which could translate to an effect on reflectance at these wavelengths. However, we cannot rule out that the effects on reflectance found in our study could also be caused by alcohol or drying affecting other traits that we did not measure.

Along with differences in reflectance in specific spectral regions, there were significant differences in the overall structure of spectra between treatments as revealed by a PERMANOVA (**Table S6**). Specifically, spectra were more variable in the drying and alcohol treatments than in their control groups, as evidenced by their significant effects in the PERMDISP (**Table S7**) and by the wider distributions across the NMDS axes in spectra belonging to these groups (**Fig. 5**). This greater variability suggests that alcohol and drying introduced "noise" into reflectance spectra, resulting in less uniform spectra than those captured from untreated leaves.

Preservation effects on trait prediction and species classification

PLSR models built from herbarium-derived reflectance spectra reliably predict many leaf functional traits. LMA prediction is generally the most accurate compared to prediction of other leaf traits (Costa *et al.*, 2018; Kothari *et al.*, 2023; White *et al.*, 2025; Hernández-Leal *et al.*, 2025), with R^2 as high as 0.96 and %RMSE as low as 5.02 (Kothari *et al.*, 2023). Our study provides additional evidence that reflectance spectra can be used to reliably predict LMA (**Fig. 5**). Fewer studies have tested the ability of PLSR models to predict leaf thickness from reflectance spectra, but those that did have reported moderate accuracy, with $R^2 < 0.65$ and %RMSE > 28 (Costa *et al.*, 2018; Hernández-Leal *et al.*, 2025). In comparison, our study's global model had considerably higher accuracy in predicting leaf thickness, likely due in part to a greater sample size (**Fig. 5**). Our study therefore strenghtens support for using PLSR models built from herbarium-derived reflectance spectra to predict leaf thickness.

For both LMA and leaf thickness, there were clear differences in the accuracy of our mixed-treatment, within-treatment, and treatment-transfer models. These differences revealed a clear pattern: models tended to increase in accuracy when restricted to leaves from within a single treatment. This was evident in the significantly lower accuracy of the mixed-treatment model, which combined alcohol treated and non-treated leaves, compared to each of the within-treatment models (**Fig. 6, Table S9**). Furthermore, prediction accuracy of our models

significantly decreased when they were transferred from one treatment to another, yet remained modestly high (all $R^2 > 0.61$) (**Figs. 7, Table S9**). These findings are consistent with a previous study that also found decreased accuracy for models that were transferred between qualitatively distinct groups of leaves. Specifically, White et al. (2025) found that when PLSR models were trained on spectra from herbarium specimens and tested on spectra from more recently collected, unmounted leaves, accuracy decreased when compared to models that were trained and tested on the same dataset. These findings highlight the limitations of models that are transferred between leaves that underwent distinct preservation protocols. PLSR model accuracy suffers when training data has a narrower range of trait values than that which occurs in a testing dataset (Burnett *et al.*, 2021). The same could be true for when spectra used in model training are less variable than or have distinct profiles from spectra used in testing. As a result, to maximize the accuracy of PLSR, models should be trained on datasets with the same spectral variability as would be expected in the testing dataset. For example, if a researcher is interested in predicting traits on herbarium specimens with presumably different preservation histories, they should train models using spectra from specimens with a diverse assemblage of preservation histories.

Previous studies have shown that PLS-DA models trained on herbarium-derived spectra are highly accurate in discriminating species, with accuracy capable of exceeding 90% (Kothari *et al.*, 2023; Hernández-Leal *et al.*, 2025). In contrast, we found 67% accuracy in a global PLS-DA model across 25 iterations, likely due in part to a limited sample size. The limited sample size in our study resulted in PLS-DA models including spectra of the same individual trees in both training and testing. While it is standard practice to avoid having individuals shared between training and testing sets (Cavender-Bares et al. 2025), our leaf sampling and experimental design led to a low number of individuals sampled per species, requiring us to split data at the individual tree level. The somewhat low accuracy of our global PLS-DA model was also likely due to a greater number of species than in previous studies that found higher accuracy (Kothari *et al.*, 2023; Hernández-Leal *et al.*, 2025). Indeed, model accuracy steadily decreases when trained on a higher number of species (Neto-Bradley *et al.*, 2025). However, the goal of our study was not to test the ability of PLS-DA models to classify species, but rather how leaf preservation treatments affect model accuracy.

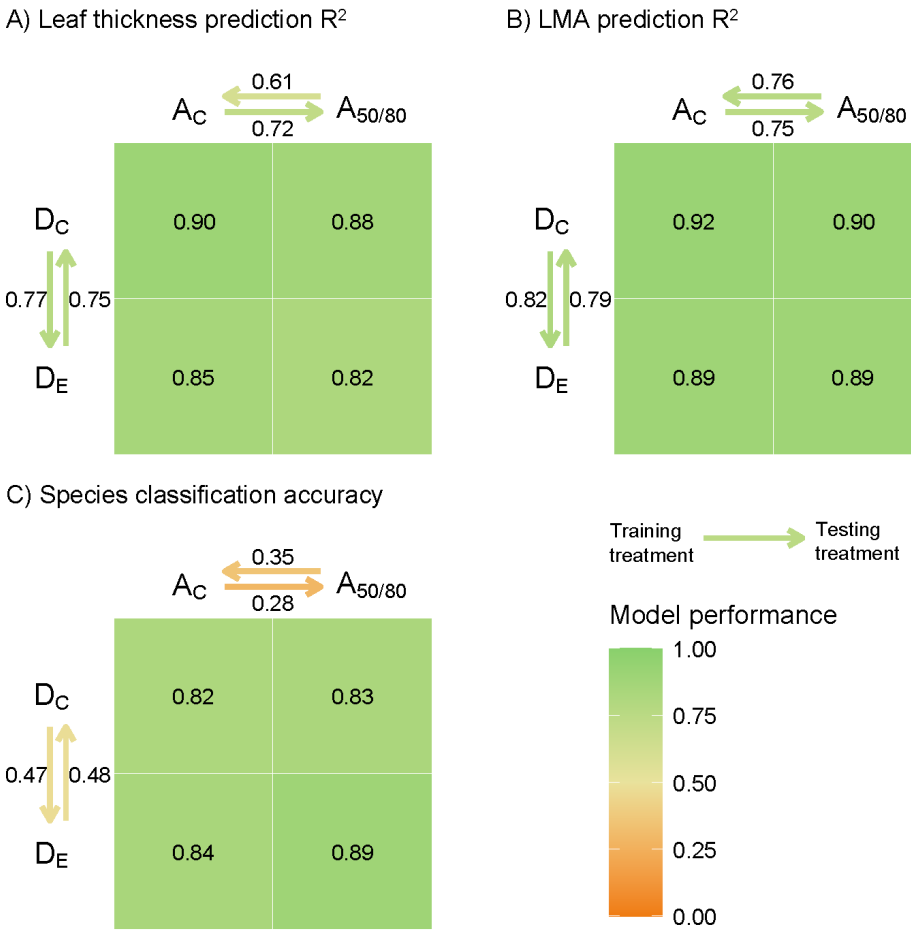
We found important differences between the mixed-treatment, within-treatment, and treatment-transfer PLS-DA models. Among these, the within-treatment models were most accurate in classifying species, with accuracy generally increasing with more intense preservation treatments (e.g., longer drying durations and the use of alcohol). In fact, the D_E-$A_{50/80}$ treatment model was the most accurate (**Fig. 6, Table S10**). While the reason for this is unclear, it may be due to leaf discoloration leading to improved accuracy of PLS-DA models. A previous study found that herbarium specimens with higher green indices were associated with lower classification model performance (White *et al.*, 2025). In our study, alcohol and drying treatments both contributed to leaf discoloration, as evidenced by the significant effects of preservation treatments on reflectance in the visible region which led to discolored leaves (**Figs. 4 & S1**). Similar reflectance across species at wavelengths associated with green pigments may therefore obscure other spectral features important for species classification and lead to ambiguous decision boundaries between species, thus reducing model accuracy.

The difference in accuracy between the mixed-treatment, within-treatment, and treatment-transfer models suggests that PLS-DA models perform worse when there is inconsistency in leaf preservation methods between training and testing datasets. Indeed, the treatment-transfer models had lower performance than any of the mixed-treatment or within-treatment models (**Figs. 7**). Further, the mixed-treatment model had lower performance than the within-treatment models (**Figs. 6**). Within each iteration of the mixed-treatment model, data splitting may have resulted in the model being trained on alcohol-treated leaves and tested on non-treated leaves (or vice versa) for a given species. These results suggest that species classification models trained on herbarium-derived spectra will perform best when training and testing are restricted to specimens that underwent similar preservation histories. However, the preservation history of specimens is often unknown, which complicates the prospect of developing models restricted to one preservation type.

For both PLSR and PLS-DA models, accuracies of treatment-transfer models significantly decreased relative to the mixed-treatment and within-treatment models (**Fig. 8**). However, this decrease in accuracy for PLSR models was relatively modest, with treatment-transfer models retaining $R^2 > 0.6$, suggesting that transferring models between leaves with distinct preservation histories can still yield reliable trait predictions. In contrast, the decrease in

accuracy for the PLS-DA models was dramatic. Indeed, PLS-DA treatment-transfer models all had accuracies < 0.48, which are potentially too low for the predictions to be useful. Ultimately, future researchers should be mindful of this limitation for PLS-DA models and consider this decrease in classification accuracy when designing studies.

Figure 8: Model performances across all within-treatment and treatment-transfer models. Panel A) shows mean R^2 across 100 iterations of PLSR models predicting leaf thickness, panel B) shows mean R^2 across 100 iterations of PLSR models predicting LMA, and panel C) shows mean accuracy across 25 iterations of PLS-DA models classifying species. In each panel, within-treatment models are shown in a 2 x 2 grid while treatment-transfer models are indicated by the arrows between treatments. Arrows point from training treatments towards testing treatments. Grid squares and arrows are color coded by R^2 or accuracy values.



Refining future herbarium-based studies

Our study highlights that spectral datasets derived from herbarium specimens with varying preservation histories are generally reliable for predicting leaf traits but may reduce accuracy when classifying species. However, there are still steps researchers can take to help maximize the utility of herbarium specimens and to refine future analyses. For example, botanists should specify if alcohol was used to preserve leaves in the field. Because many botanists now automate the generation of herbarium labels from spreadsheets (Davies *et al.*, 2024), an additional field indicating the details of alcohol use would require minimal extra effort and could be applied to specimen labels in bulk. Furthermore, field stations would benefit from investing in drying ovens, even rudimentary ones. Because we found minimal drying effects on leaf traits, reflectance spectra, and predictive modeling, moderate differences in drying durations or temperatures in field ovens would likely have minimal impact on resulting herbarium specimens. Once dried and stored in an herbarium, there are additional steps researchers can take to refine scanning protocols and analyses. For example, when scanning specimens, researchers should take note of leaf discoloration and try to capture a wide range of leaf coloration. In fact, these and similar notes are encouraged by the IHerbSpec protocol for measuring spectral reflectance of herbarium specimens (IHerbSpec, 2026). If both green and discolored leaves are included in a scanning session, this may suggest that the specimens underwent different preservation methods – either they were dried immediately (leaves are still green in color) or that alcohol was used (leaves are discolored/brown). As a result, models built from these scans would likely perform better in predicting leaf traits or in classifying taxonomic ranks. In addition to data collection refinements, the development of additional types of predictive models may help alleviate the reduction in accuracy that is induced by variable preservation histories. For example, predictive models could be trained to detect when specimens were preserved in alcohol and any specimens detected to have undergone alcohol preservation could have a correction factor applied to their spectra before being used to train a species classifier.

Conclusions

Herbaria are critical repositories for data on plant phenotype, with specimen-derived traits increasingly used for ecological inference and species delimitation (Heberling, 2022). Here, we show that alcohol preservation and drying intensity – two vital, yet variable components of

specimen collection protocols – significantly affected measurements of leaf phenotype. Alcohol increased leaf shrinkage during preservation, which significantly affected a suite of traits related to leaf size. Both alcohol and drying significantly affected reflectance in the visible, NIR, and SWIR regions of the electromagnetic spectrum. Notably, these effects on reflectance only modestly decreased the accuracy of using spectra for predicting traits related to leaf size, yet considerably decreased the accuracy of using spectra for predicting species. These results have important implications for the international botanical community. First, herbarium users ranging from ecologists to taxonomists should be mindful of the effect of preservation on foliar traits (Blonder *et al.*, 2012). In particular, botanists should be cognizant of alcohol preservation causing increased leaf shrinkage. Second, there is a growing community of scientists harnessing herbarium specimens for reflectance spectra (Cavender-Bares et al. 2025). The finding that species prediction accuracy is significantly affected by variable preservation histories may necessitate the development of identification models that account for preservation history in model training and testing. Finally, we recommend that information on specimen preservation be included as standard annotation on botanical collections, thus enabling downstream users of herbarium specimens to account for preservation history in their work. By accounting for specimen preservation history, the global botanical community will be empowered to more accurately investigate natural and environmentally induced phenotypic variation.

Acknowledgements

This study was conducted as part of the MBG's Revolutionizing Species Identification project. We thank Kate Freeman, Madison Yocks, Rebecca Sucher, and Sophia Marsh for assistance in collecting leaves from the MBG living collection.

Competing interests

None declared.

Author contributions

RPF and MWA conceived and designed the study. RPF and RSF acquired data. RPF conducted data analysis. RPF, RSF, and MWA interpreted the data and results. RPF wrote the first draft of the manuscript which was reviewed, edited, and approved by RPF, RSF, and MWA.

Data availability

All data and code to reproduce the analyses are available in a GitHub repository at: https://github.com/rpfortier/specimen_preparation_experiment

References

Barton KE. 2024. The ontogenetic dimension of plant functional ecology. *Functional Ecology* 38: 98–113.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57: 289–300.

Blonder B, Buzzard V, Simova I, Sloat L, Boyle B, Lipson R, Aguilar-Beaucage B, Andrade A, Barber B, Barnes C, *et al.* 2012. The leaf-area shrinkage effect can bias paleoclimate and ecology research. *American Journal of Botany* 99: 1756–1763.

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.

Cavender-Bares J, White DM, Ahlstrand NI, Austin MW, Bastianelli D, Bazan S, Boughalmi K, Cardinal-McTeague W, Chacón-Madrigal E, Couvreur TLP, *et al.* 2025. Next-generation specimen digitization: capturing reflectance spectra from the world's herbaria for modeling plant biology across time, space, and taxa. *New Phytologist*: nph.70645.

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.

Davies NMJ, Drinkell C, Utteridge TMA (Eds). 2024. *The Herbarium Handbook*. Richmond, UK: Royal Botanic Gardens, Kew.

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.

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

Heberling JM, Isaac BL. 2017. Herbarium specimens as exaptations: New uses for old collections. *American Journal of Botany* 104: 963–965.

Hernández-Leal MS, Guzmán Q. JA, González-Rodríguez A, Cavender-Bares J. 2025. Foliar spectral signatures reveal adaptive divergence in live oaks (*Quercus* section *Virentes*) across species and environmental niches. *New Phytologist* 248: 370–388.

Hodge WH. 1947. The use of alcohol in plant collecting. *Rhodora* 49: 207–210.

IHerbSpec, 2026. Protocol for Spectral Digitization of Herbarium Specimens, v.1.2.1. Zenodo. https://doi.org/10.5281/zenodo.18451589

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.

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

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

Kozlov MV, Sokolova IV, Zverev V, Zvereva EL. 2021. Changes in plant collection practices from the 16th to 21st centuries: implications for the use of herbarium specimens in global change research. *Annals of Botany* 127: 865–873.

Kuhn M. 2008. Building Predictive Models in *R* Using the caret Package. *Journal of Statistical Software* 28: 1–26.

Lancefield CS, Panovic I, Deuss PJ, Barta K, Westwood NJ. 2017. Pre-treatment of lignocellulosic feedstocks using biorenewable alcohols: towards complete biomass valorisation. *Green Chemistry* 19: 202–214.

Leigh A, Sevanto S, Close JD, Nicotra AB. 2017. The influence of leaf size and shape on leaf thermal dynamics: does theory hold up under natural conditions? *Plant, Cell & Environment* 40: 237–248.

Liland KH, Mevik B-H, Wehrens R, Hiemstra P. 2026. pls: Partial Least Squares and Principal Component Regression.

Lipsen LPJ, Tan D. 2023. *Pressed plants: making a herbarium*. Victoria: Royal British Columbia Museum.

Martin RE, Asner GP, Bentley LP, Shenkin A, Salinas N, Huaypar KQ, Pillco MM, Ccori Álvarez FD, Enquist BJ, Diaz S, *et al.* 2020. Covariance of Sun and Shade Leaf Traits Along a Tropical Forest Elevation Gradient. *Frontiers in Plant Science* 10: 1810.

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

Mori SA. 2011. *Tropical Plant Collecting: From the Field to the Internet* (SA Mori, A Berkov, CA Gracie, and EF Hecklau, Eds). Florianopolis, Brazil: TECC Editora.

Neto-Bradley BM, Bonnet P, Goëau 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*: nph.70258.

Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin, PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, *et al.* 2025. vegan: Community Ecology Package.

Parnell J, Rich T, McVeigh A, Lim A, Quigley S, Morris D, Wong Z. 2013. The effect of preservation methods on plant morphology. *TAXON* 62: 1259–1265.

Perez TM, Rodriguez J, Mason Heberling J. 2020. Herbarium-based measurements reliably estimate three functional traits. *American Journal of Botany* 107: 1457–1464.

Pérez-Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poorter H, Jaureguiberry P, Bret-Harte MS, Cornwell WK, Craine JM, Gurvich DE, *et al.* 2013. New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 61: 167.

Queenborough S. 2017. Collections-based Studies of Plant Functional Traits. In: Friis I, Balslev H, eds. Proceedings of an international symposium held at the Royal Danish Academy of Sciences and Letters. Copenhagen: Scientia Danica, 223–236.

R Core Team. 2024. R: A language and environment for statistical computing.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9: 676–682.

Sun Y, Wong AKC, Kamel MS. 2009. Classification of Imbalanced Data: A Review. *International Journal of Pattern Recognition and Artificial Intelligence* 23: 687–719.

Thiers BM. 2025. *The World's Herbaria 2024: A Summary Report Based on Data from Index Herbariorum*. New York Botanical Garden.

Violle C, Navas M, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E. 2007. Let the concept of trait be functional! *Oikos* 116: 882–892.

White DM, Cavender-Bares J, Davis CC, Guzmán Q. JA, Kothari S, Robles JM, Meireles JE. 2025. Seeing herbaria in a new light: leaf reflectance spectroscopy unlocks trait and classification modeling in plant biodiversity collections. *New Phytologist*: nph.70357.

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.