A Protocol for Standardizing Measurements and Enabling Global Harmonization of Herbarium Leaf Reflectance Spectra

Dawson M. White,[1,2,20] Flavia M Durgante,[3] Matthew W. Austin,[4] J. Antonio Guzmán Q.,[2] Jonathan A. Kennedy,[1] Cornelius Onyedikachi Nichodemus,[5] Natalie I. Ahlstrand,[6] Chris Baraloto,[7] Billur Bektaş,[8] Khalil Boughalmi,[9] Warren Cardinal-McTeague,[10] Thomas L.P. Couvreur,[9] Riley P. Fortier,[4] Ryan S. Fuller,[4] Olwen M. Grace,[11] Rykkar Jackson,[10] Kevin Karbstein,[12,13] Shan Kothari,[14] Aaron K. Lee,[15] Jéssica Lira Viana,[16] Elizabeth Martinez-Victoria,[2] José Eduardo Meireles,[5] Thomas H. Murphy,[7] Carlos Rodrigues-Vaz[9,17], Daniel Spalink,[18] Hanna Tuomisto,[16,19] Caroline C. Vasconcelos,[3] Jeannine Cavender-Bares[1,2,20]

[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]Herbarium, National Institute of Amazonian Research, Manaus, Brazil

[4]Herbarium, Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO 63110 USA

[5]School of Biology and Ecology, University of Maine, Orono, ME 04469, USA

[6]Natural History Museum Denmark, University of Copenhagen, Copenhagen, 1354, Denmark

[7]International Center for Tropical Botany, Florida International University, Miami, FL 33133, USA

[8]Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zurich, 8092, Zurich, Switzerland

[9]DIADE, Univ Montpellier, CIRAD, IRD, Montpellier, France

[10]Department of Forest and Conservation Sciences, University of British Columbia, 2424 Main Mall, Vancouver BC, V6T 1Z4, Canada

[11]Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, United Kingdom

[12]Professorship of Integrative Plant Taxonomy, Friedrich Schiller University of Jena, Jena, Germany

[13]Department of Natural History, Florida Museum of Natural History, Gainesville, FL 32611, USA

[14]Department of Renewable Resources, University of Alberta, Edmonton, AB T6G 2E3, Canada

[15]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108, USA

[16]Department of Biology, Aarhus University, Aarhus, 8000, Denmark

[17]Muséum National d'Histoire Naturelle, Institut de Systématique, Évolution, et Biodiversité (ISYEB), Centre National de la Recherche Scientifique, Sorbonne Université, École Pratique des Hautes Études-Paris, Sciences et Lettres, Université des Antilles, C.P. 39, 57 rue Cuvier, 75005 Paris, France.

[18]S.M. Tracy Herbarium, Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX 77843, USA

[19]Department of Biology, University of Turku, 20014 Turku, Finland

[20]Authors for correspondence (dawson.white@gmail.com; jcavender@fas.harvard.edu)

***Abstract*—**Reflectance spectroscopy offers a powerful approach to integrate high-throughput phenotypic data from herbarium specimens into the digital landscapes of ecology, evolution, and systematics. Because inconsistencies in instrumentation and measurement practices can increase noise and limit dataset compatibility, the International Herbarium Spectral Digitization Working Group (IHerbSpec) has published the *Protocol for Spectral Digitization of Herbarium Specimens*, currently in version 1.2.1, as an open resource that will continue to evolve through community use and collaboration (https://iherbspec.github.io/protocol). The protocol defines a stepwise measurement workflow, standardized filename conventions, structured metadata tables with controlled vocabularies, and guidance on tissue selection, materials, and instrumentation quality control to ensure that newly generated spectral datasets are robust and comparable. By embedding standardized practices at the point of data collection, it provides a scalable foundation for data synthesis and new quantitative insights into plant diversity across taxonomic, geographic, and temporal scales.

***Keywords*—**Data harmonization, data standards, extended specimen, FAIR principles, Herbarium digitization, phenomics

The more than 400 million botanical collections stored in global herbaria represent a vast and irreplaceable repository of plant diversity, together with rich environmental and ecological metadata (Davis, 2023; Thiers, 2025). These specimens underpin our understanding of organismal morphology, taxonomy, and biogeography, and are increasingly used to generate extended genetic and phenotypic datasets that expand their scientific impact and potential to address societal challenges (Lendemer et al., 2020; Kunkel et al., 2025).

Leaf reflectance spectra—here referring to point measurements made using a contact probe (Fig. 1b)—are now well established as a high-throughput means of linking phenotype, environment, and evolutionary history at different scales (Serbin et al., 2014; Cavender-Bares et al., 2016, 2025a; Vance et al., 2016; Kothari and Schweiger, 2022; Stefanski et al., 2025). Spectral measurements across the visible, near-, and shortwave infrared regions (350–2500 nm) capture integrated optical signals related to tissue structure and chemistry (Curran, 1989; Jacquemoud et al., 1996). Spectral reflectance data have demonstrated utility in discriminating taxa and cytotypes (Abasolo et al., 2013; Durgante et al., 2013; Lang et al., 2017; Stasinski et al., 2021; Buono and Albach, 2023; Vasconcelos et al., 2025), estimating leaf traits (Costa et al., 2018; Kothari et al., 2023), and revealing adaptive differentiation and evolutionary processes (Cavender-Bares et al., 2016; Meireles et al., 2020; Fine et al., 2021; Hernández-Leal et al., 2025). Importantly, multiple studies have now demonstrated that these same applications extend to herbarium specimens that are decades to centuries old (Kühn et al., 2024; Boughalmi et al., 2025; Neto-Bradley et al., 2025; White et al., 2025). These advances have signaled herbarium spectral digitization as a robust approach for phenotyping plant biodiversity at unprecedented spatio-temporal scales.

The scientific potential of herbarium spectral digitization lies in the opportunity to represent plant phenotypic diversity across regions, taxa, and time in harmonized datasets

generated across herbaria (Cavender-Bares et al., 2025b; Paton et al., 2025). Yet, coordinated standards for data collection, metadata, and databasing are needed to avoid the risk that spectral data become fragmented into collections that are difficult to find and harmonize (Wilkinson et al., 2016), or are excessively noisy due to undocumented differences among measurement setups. Comparable challenges in plant trait research have been mitigated through standardized measurement and databasing protocols (Kattge et al., 2011; Enquist et al., 2016; Pérez-Harguindeguy et al., 2016).

To document botanical variation present in herbaria through spectral measurements—and to ensure that these data are interpretable and reusable—standards must explicitly capture both the measurement environment and the characteristics of the specimens themselves to enable assessments of data quality. Spectral reflectance measurements do not represent a single, invariant biological signal comparable to a DNA sequence (Cruickshank and Munck, 2011). Instead, they integrate the molecular and structural properties of all the materials that interact with incident light, and this includes not only the target plant tissue but also the background material and potential contaminants, such as mounting materials or epiphylls. Furthermore, herbarium tissues exhibit wide variation in age, preservation histories, mounting techniques, and states of degradation (Kühn et al., 2024; Cavender-Bares et al., 2025b; White et al., 2025).

To address these challenges at an early stage of data generation, we established the International Herbarium Spectral Digitization Working Group (IHerbSpec), a global consortium dedicated to advancing best practices for herbarium-based reflectance spectroscopy and fostering coordinated, cross-institutional collaboration.

Here, we present the *Protocol for the Spectral Digitization of Herbarium Specimens* (the IHerbSpec Protocol for short), a community-developed standard that defines minimum

requirements, recommended practices, a step-by-step workflow, metadata schema, and guidelines for quality control and data management. The IHerbSpec Protocol is a living document deployed at iherbspec.github.io with versioned releases archived on Zenodo (IHerbSpec, 2026). It is designed to foster collaboration and provide a foundation for interoperable spectral datasets that can be confidently aggregated, reused, and integrated into the global biodiversity data ecosystem for ecological and evolutionary research.

Methods and Results

Knowledge exchange and consensus building

The IHerbSpec Protocol was developed collaboratively by the International Herbarium Spectral Digitization Working Group (IHerbSpec), established in fall 2024. The group was assembled to represent a cross-section of the herbarium and spectral biology communities, including herbarium-based curators and researchers interested in spectral digitization and members of the ASCEND Biology Integration Institute (NSF-DBI-2021898; spectralbiology.org) interested in advancing the use of preserved plant specimens.

Working group activities began with a systematic comparison and consolidation of existing herbarium spectroscopy workflows conducted through a series of virtual webinars between December 2024 and April 2025. Ten distinct workflows were presented, several of which are published (Durgante et al., 2013; Laliberté and Soffer, 2018; Kothari et al., 2023; Quinteros Casaverde et al., 2024; Mersni et al., 2025; White et al., 2025). Members also shared approaches to specimen digitization and data management at their institutions. These exchanges provided a common foundation for identifying shared practices, sources of variability, and priorities for coordination. The process culminated in an in-person workshop at the Harvard

University Herbaria in May 2025, where remaining points of divergence were discussed and resolved collectively.

Decisions during protocol development were made by consensus rather than formal voting. Agreement was reached through iterative discussion, refinement of draft text, and collective review, with the explicit goal of protocol compliance. The resulting protocol is a community-developed standard authored by 31 members from 23 institutions across three continents.

Publication and overview of the protocol

The IHerbSpec Protocol is written using Quarto (Allaire et al., 2025) and developed through a public GitHub repository (github.com/IHerbSpec/iherbspec.github.io) where community feedback is organized and incorporated via Discussions and Issues tools. The source files are compiled to HTML and published as a living document at iherbspec.github.io, ensuring that the online version reflects the most current updates. The GitHub repository is linked to Zenodo (https://doi.org/10.5281/zenodo.18451589), where versioned releases, including a compiled PDF, are permanently archived and assigned persistent DOIs. This structure maintains a single, version-controlled source of the protocol that supports transparent revision and community input, while generating both web and citable PDF outputs to ensure long-term preservation and stable scholarly reference. The protocol is distributed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

The broad goal of the IHerbSpec Protocol is for its standards to be implemented across global contexts and scales—from small, hypothesis-driven projects to large institutional digitization initiatives. To accommodate this range of use, the protocol incorporates both *required* and *recommended* components within its measurement workflow and metadata schema,

allowing projects to adapt implementation to local capacity while maintaining a shared structural framework. This tiered design ensures that the protocol can function as both a practical guide for independent researchers and a scalable foundation for coordinated, cross-institutional efforts.

The protocol is organized into modular sections that separate procedural guidance, documentation standards, and supporting rationale. It begins with a preface describing its development and consensus process, followed by six core sections and two appendices detailing workflow, metadata, file conventions, instrumentation, tissue selection, and sampling thresholds (Table 1).

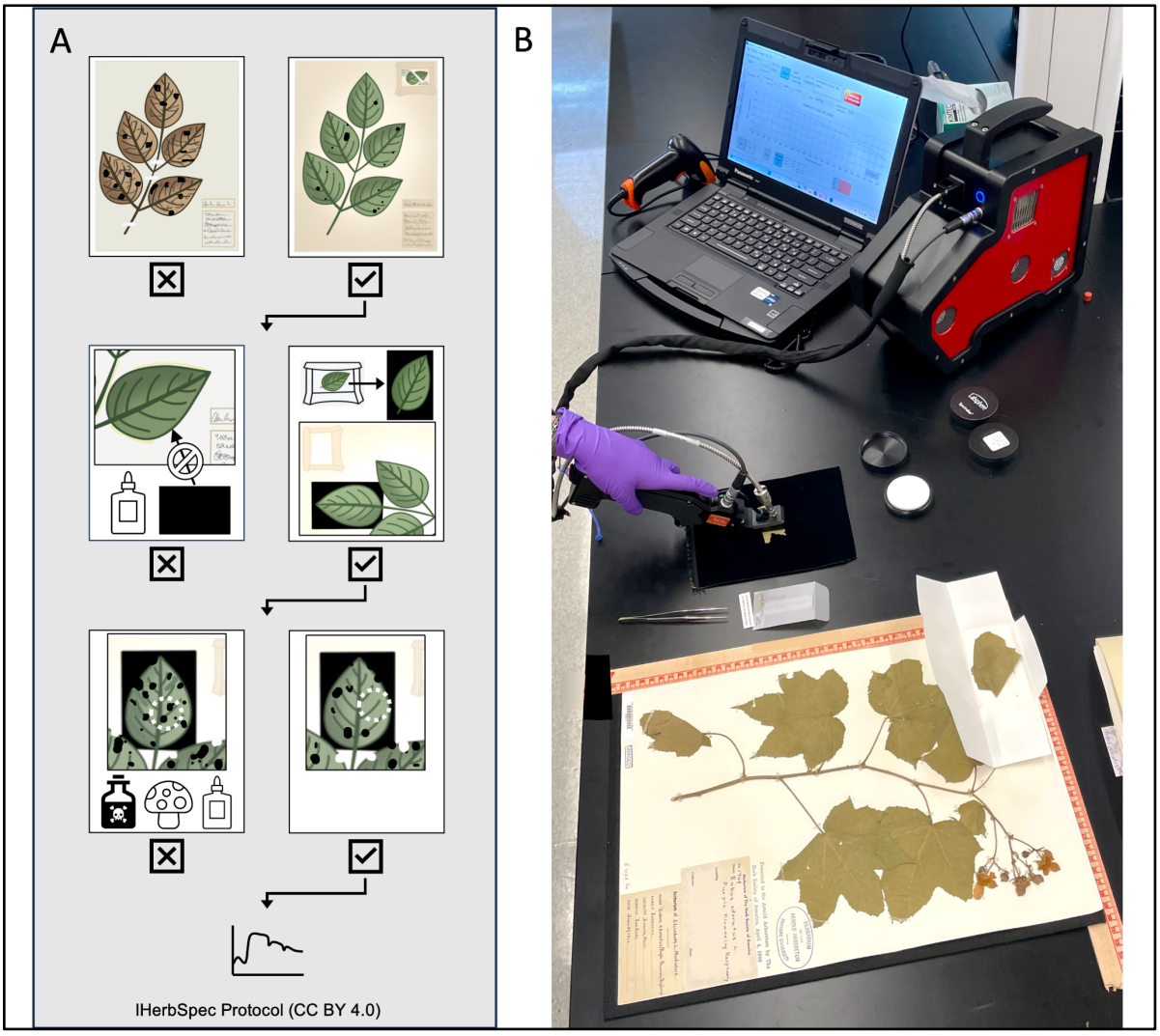

Fig. 1. Specimen and tissue selection workflow and bench measurement setup. (A) Specimen and tissue selection decision tree. The first step (top) prioritizes specimens that are generally suitable for measurement (mature, intact, flat, and with minimally contaminated areas), unless the specimen is of unique importance. The second step prioritizes tissues that permit insertion of a thin non-reflective black background behind the target tissue. As stated in Part I of the IHerbSpec Protocol, insertion of a black background is required whenever possible to minimize spectral contamination from mounting materials. However, this is not always feasible for certain mounted or otherwise valuable specimens; in such cases, measurements may still be

collected provided that background conditions and potential sources of contamination are explicitly documented in the metadata. The final step emphasizes selecting tissue areas that are as representative and uncontaminated as possible (e.g., avoiding glue, epiphylls, or herbivory where feasible). (B) Bench setup for spectral measurement. Figure A reproduced from IHerbSpec Protocol (CC BY 4.0).

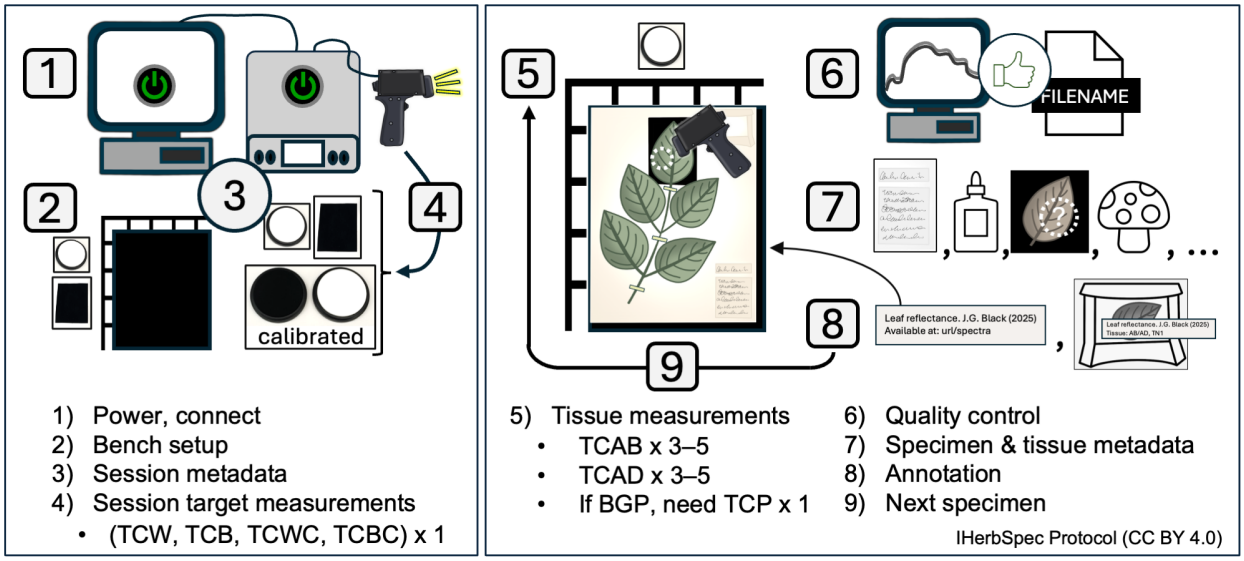

Fig. 2. Schematic of the IHerbSpec Protocol v1.2 measurement workflow. Steps 1–4 illustrate instrument and materials setup, measurement of white and black references and calibrated standards, and recording of session-level metadata associated with these activities. Steps 5–9 illustrate the tissue measurement sequence, filename and data quality validation, scoring of specimen- and tissue-level metadata, and annotation of project information on herbarium specimens. The protocol requires a minimum of three and a recommendation of five or more measurements (see Appendix I of the protocol) for both adaxial and abaxial leaf surfaces across the specimen, providing sufficient tissue is available and suitable for measurement. Plants with small, terete, or otherwise challenging leaves for measurement might not be able to meet this requirement (see *Vision for Future Versions* section). Figure reproduced from IHerbSpec Protocol (CC BY 4.0).

Informatics standards and metadata

The IHerbSpec metadata schema (Parts 3 and 4) was developed to support discovery, harmonization, and reuse of spectral data across projects, consistent with FAIR principles (Wilkinson et al., 2016). During the in-person and virtual meetings, the working group discussed informatics needs for spectral digitization, including digital storage, integration with collection management systems, data entry interfaces, and suitable formats for data creation, preservation, and dissemination. An informatics subgroup reviewed existing biodiversity data standards, particularly those maintained by Biodiversity Information Standards (TDWG; https://www.tdwg.org), including Darwin Core (Wieczorek et al., 2012) and its MeasurementOrFact extension.

However, these frameworks did not adequately capture the session–specimen–tissue hierarchy required for herbarium spectral measurements. The IHerbSpec metadata schema was therefore developed as a relational model encompassing session-, specimen-, and tissue-level information, with explicit attention to measurement configuration and tissue characteristics (https://iherbspec.github.io/protocol/part4-metadata.html#section4-1). This approach aligns with the goals of the BIOFAIR data network, which emphasizes community-driven, domain-aware metadata standards that support interoperability, machine readability, and reuse across biological research infrastructures (Kunkel et al., 2025).

To support different project needs, the metadata schema is designed as a simple spreadsheet (“flat file”) for data capture and dissemination, and as a structured model suitable for integration into relational databases. The spreadsheet with all metadata fields is provided on the IHerbSpec website for immediate use or adaptation https://iherbspec.github.io/protocol/tables.

Although customized to capture the contextual detail needed for quality control, interpretation, and downstream aggregation of herbarium spectral data, the metadata schema was designed with future integration as a Darwin Core extension in mind. Where possible, existing Darwin Core terms are reused (e.g., `scientificName`, `identifiedBy`), and newly defined fields for spectral-specific information (e.g., `tissueDevelopmentalStage`, `hasGlue`, `measurementFlags`) follow Darwin Core naming conventions and semantics. This design supports immediate usability while laying the groundwork for potential future development of a Darwin Core extension for spectral data.

Data sharing and project-level data integration

Data sharing and long-term preservation were treated as core design considerations alongside measurement and metadata standards. Projects are strongly encouraged to archive and share unprocessed spectral data files together with a metadata spreadsheet. Although tabular files (e.g., *.csv*) containing specimens in rows and spectral bands and metadata fields in columns are more user-friendly for data consumers, they necessarily reflect prior processing (e.g., radiance-to-reflectance conversion) and may obscure other metadata such as exact wavelength centers, spectral resolution, detector overlap regions, and measurement conditions. Raw spectral files and their associated headers preserve the most complete record of measurement context and instrument configuration, supporting future harmonization across instruments and projects. Should the metadata spreadsheet become detached from the spectral files, the recommended full filename structure defined in Part 3 (https://iherbspec.github.io/protocol/part3-filenames.html) maintains minimum traceability to the project, specimen, tissue class, and background conditions. When tabular datasets are also provided to enhance usability, all applied processing

steps should be clearly documented in the data archive to maintain transparency and reproducibility.

To support discovery, citation, and long-term accessibility, the working group established the IHerbSpec Dataverse (https://dataverse.harvard.edu/dataverse/iherbspec/) as a primary project-level repository for herbarium spectral data. The Dataverse is freely available, widely used in the research community, and supports persistent identifiers (DOIs) and versioning, and user-defined licensing to communicate how data can be used. Submitted datasets may be minimally curated to help ensure conformance with the protocol's required components and metadata fields.

The IHerbSpec Dataverse is designed to accommodate datasets generated both within and outside formal institutional digitization pipelines, including projects based on specimens that lack barcodes, images, or fully curated database records. In this way, the Dataverse serves as a mechanism to support the aggregation, preservation, and reuse of herbarium spectral data and to link it to the corresponding herbarium specimens across projects and institutions. In addition, IHerbSpec and the IHerbSpec Dataverse act as signposts that enhance the discoverability of herbarium spectral datasets that are stored in other repositories.

Embedding quality control and interoperability into the protocol

The IHerbSpec Protocol encodes quality control throughout its structure, beginning with the definition and overview of the protocol's *required* versus *recommended* elements in each section (Part 1), extending through its workflow and metadata standards (Parts 2, 3, and 4), and reinforced by additional measurement and tissue selection guidelines (Parts 5 and 6; Table 1; Fig. 1). By distinguishing required from recommended elements, the protocol defines the conditions under which spectral data can be considered interoperable while preserving flexibility

252 across projects with different goals and constraints. These guidelines are optimized for large-
253 scale spectral digitization efforts intended to maximize interoperability and data aggregation;
254 projects may justify deviations from recommended conditions, provided that such deviations are
255 transparently documented. Compliance therefore signals that data meet baseline quality
256 expectations and include sufficient metadata to support evaluation, reuse, and integration.

257 To support consistency across institutions, the working group identified several
258 components requiring explicit standardization. These include the measurement and archiving of
259 white and black reference standards, the use of non-reflective black backgrounds to minimize
260 contamination from mounting materials (Fig. 3), a minimum number of measurements per
261 specimen, and standardized metadata and file naming conventions.

262 A key quality-control decision concerned sampling intensity. While a single carefully
263 collected spectrum per leaf surface can adequately represent a specimen, it does not capture
264 variance. The protocol therefore requires a minimum of three representative measurements per
265 abaxial and adaxial leaf surface per specimen, when feasible, with on-screen verification of the
266 integrity of each measurement (Fig. 2; Fig. 4). This threshold enables calculation of wavelength-
267 specific means and variances, and supports detection of anomalous measurements. Five
268 measurements per surface are recommended (or more if feasible given the taxon and project
269 scope) to better capture within-specimen spectral variation. Given the importance of inspecting
270 spectral data quality and consistency, the protocol emphasizes that replicate measurements can
271 be taken from any combination of multiple leaves or a single spot on a leaf depending on the
272 availability of suitable material. Appendix I of the protocol further discusses these decisions
273 (https://iherbspec.github.io/protocol/appendix1.html).

Beyond measurement procedures, the hierarchical metadata schema (session-, specimen-, and tissue-level metadata) is designed to communicate the essential contextual information required to assess data quality and interoperability. Instrument-related fields—such as `instrumentModel`, `opticalSetupDescription`, and `measurementSettings`—are explicitly required because different instruments and optical configurations can introduce batch effects if not documented and harmonized (Meireles et al., 2020). Additional required fields such as `tissueDevelopmentalStage`, `hasGlue`, and `hasNonGlueContamination` (e.g. chemical preservatives, insecticides), and recommended fields such as `measurementFlags` (e.g. `AlcoholPresent`) and `tissueNotes`, are tied specifically to the measurement area to enable precise downstream filtering. Finally, to support training and reduce subjectivity, Appendix II of the protocol provides additional guidelines for scoring tissue characteristics and scoring examples linked to images for 51 specimens from 11 herbaria (See IHerbSpec Protocol Appendix II: https://iherbspec.github.io/protocol/appendix2.html). Metadata should also be subject to quality checks in line with the quality assurance protocols applied to other data in the digitization pipeline.

By thoroughly documenting methodological choices, measurement environments, and specimen characteristics, the IHerbSpec protocol—through its required and recommended elements and structured metadata schema—supports quality control by enabling datasets to be filtered and harmonized based on documented conditions. This documentation allows data users to implement and refine their own quality control procedures while accommodating spectral measurements from challenging specimens (including taxa with small or structurally complex leaves, or leaves glued to herbarium mounting paper) that may not meet ideal conditions but

297 remain scientifically important. In doing so, the protocol promotes inclusive data capture while
298 facilitating data filtering by quality-related characteristics; thereby promoting integration across
299 projects and herbaria globally.

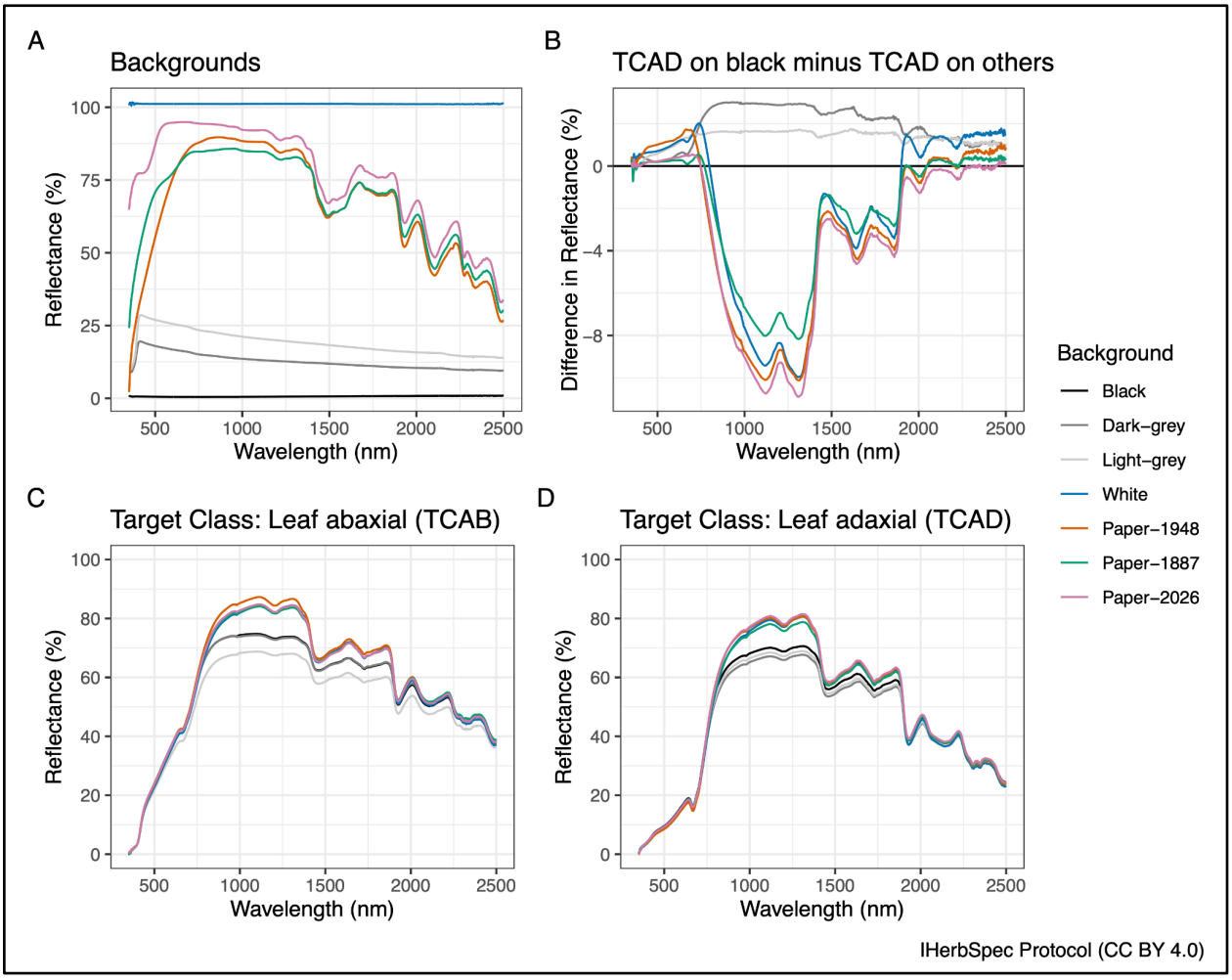

Fig. 3. The effect of background on tissue spectral profiles. The IHerbSpec Protocol requires placement of a low-reflectance black background beneath tissues to minimize background contamination in reflectance data. (A) Reflectance spectra of reference and background materials measured during a session, including the white reference panel (White, colored blue), the black background (Black), gray backgrounds, and representative herbarium mounting papers. These background materials exhibit distinct spectral signatures, particularly outside the visible range (400–700 nm). Plots show full-range (350–2500 nm), unprocessed reflectance data measured with the Spectral Evolution NaturaSpec and Surface Probe. (B) Difference spectra showing adaxial leaf reflectance measured on a black background minus

reflectance measured at the same leaf surface location on alternative backgrounds, illustrating wavelength-dependent background effects on tissue measurements. (C–D) Reflectance profiles of the same leaf measured on different backgrounds for two target classes: leaf abaxial (TCAB; panel C), and leaf adaxial (TCAD; panel D) surfaces. Measurements taken over black and gray backgrounds consistently show lower reflectance than those taken over white reference material and white to yellowed herbarium papers. Across panels, background effects are most pronounced in the 800–1800 nm region, corresponding mostly to the near-infrared (1100–2000 nm), with weaker but detectable effects in the visible (400–700 nm) and longer-wavelength SWIR (2000–2500 nm).

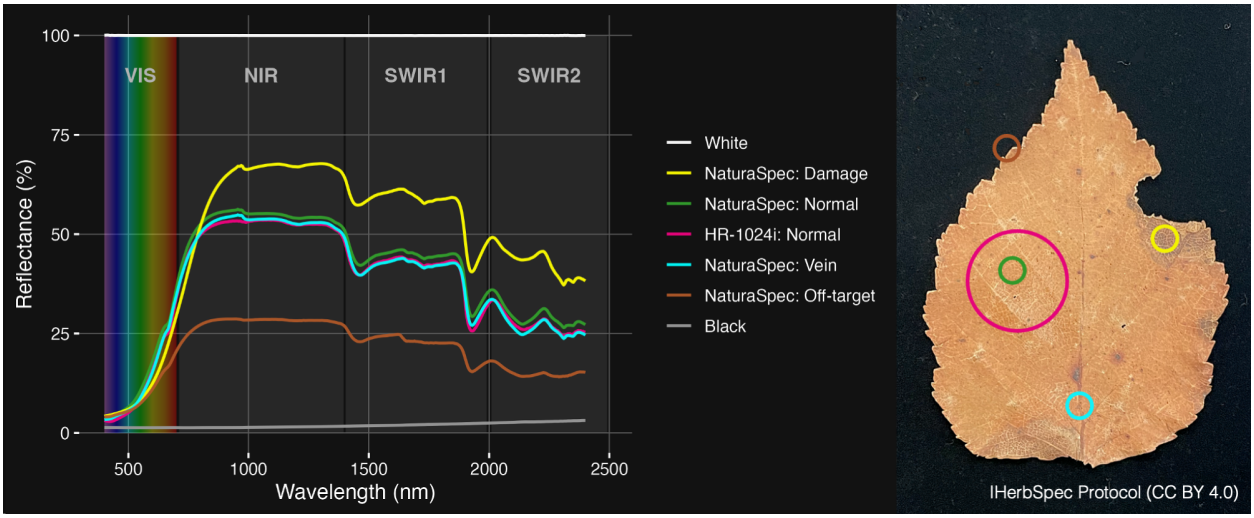

Fig. 4. Instrument and spatial heterogeneity in herbarium leaf reflectance measurements. Reflectance spectra collected from the same dried leaf (*Betula papyrifera;* NEBC00636882; tissue condition scoring in IHerbSpec Protocol Appendix II) using two spectroradiometers with different optical configurations: a Spectral Evolution NaturaSpec with a 2 mm field of view and an SVC HR-1024i with a 22 mm field of view. Spectra were trimmed to 400–2400 nm and plotted with the *ggplot2* R package (Wickham, 2016). Colored circles on the leaf image indicate measurement locations corresponding to the plotted spectra (circles not drawn to scale). NaturaSpec measurements include normal lamina tissue, midvein, pathogen-damaged tissue, off-target measurement including some black background, and reference measurements of the black background and white standard. The SVC measurement represents a standard lamina measurement integrating reflectance over a substantially larger sampling area. Spectral regions are shaded to indicate the visible range (VIS, 400–700 nm; rainbow band), near-infrared (NIR, ~700–1400 nm), and shortwave infrared (SWIR1, ~1400–1900 nm; SWIR2, ~2000–2500 nm). Differences among spectra illustrate how optical configuration and fine-scale

335 tissue heterogeneity influence reflectance profiles across the visible to shortwave infrared
336 spectrum.
337

Embedding spectra within institutional digitization workflows

A central goal in developing the protocol was ensuring its suitability for adoption under real-world institutional digitization workflows. Although expanded metadata capture will slow specimen throughput, the seventeen required metadata fields are essential for quality control filtering and harmonizing spectral data across institutions in accordance with FAIR principles (Wilkinson et al., 2016). Where possible, institutional software for digitization ("front ends") can streamline metadata capture by automatically propagating repetitive elements (e.g., session-level fields like `projectId` and `instrumentModel`).

The protocol's filename conventions and hierarchical metadata schema (Parts 3 and 4; Table 1) help prevent spectral data files from becoming detached from their project, specimen, or measurement conditions, safeguarding their long-term discoverability, interpretability, and reuse alongside the enduring stewardship of the specimens themselves. Thorough metadata capture also honors the substantial institutional effort invested in specimen digitization.

Critically, each specimen and corresponding extended dataset must be associated with a stable, traceable identifier. In the protocol, this consists of two values that are recorded in required specimen-level metadata fields and incorporated as filename components: `herbariumCode` defines the herbarium that houses the specimen, and `specimenId` defines the specimen itself. The latter can be, for example, the Darwin Core-compatible occurrence or catalogue identifier, a barcode, an accession number, or a unique combination of collector name and collector's number (e.g., "Spruce3251"). Maintaining resolvable links among specimen records and associated datasets aligns with extended specimen network models integration frameworks that facilitate repository connectivity and cross-platform interoperability (Kunkel et al., 2025).

In practice, taking spectral measurements is most straightforward on unmounted specimens or free tissues stored in herbarium packets, where all tissue surfaces are accessible, black backgrounds can be easily used, and tissues can be thoroughly inspected for quality. IHerbSpec therefore recommends that specimen preparation workflows include storing detached leaf material in specimen packets, a practice that may already be in place to support destructive sampling and dissection. Institutions may also consider integrating spectral capture with barcode assignment prior to mounting, ensuring stable linkage between specimen and data and enabling access to the whole specimen for spectral measurement.

Importantly, incorporating spectral readings into digitization workflows—whether alongside imaging or specimen label transcription—will require a dedicated workstation and trained personnel. The guidelines embedded in the protocol (Parts 5, 6, and Appendix II) were designed to be readable and to support training and consistent implementation. For specimens already databased, spectral digitization can proceed in coordinated batches, with suitability of each specimen and tissue assessed prior to measurement (Fig. 1) and any deviations from recommended conditions explicitly documented in the tissue-level metadata fields.

Scaling spectral digitization into the global metaherbarium

Realizing the potential value of herbarium spectral data will require integration with established biodiversity data infrastructures and extended specimen networks. As described above, spectral datasets must be anchored to stable specimen identifiers within the IHerbSpec metadata schema to enable resolvable links across institutional repositories, GBIF, iDigBio, and related aggregators (Kunkel et al., 2025).

Currently, extended specimen integration can be implemented through outward linking models like the Darwin Core Resource Relationship extension within existing biodiversity

platforms or collection management systems. For example, Symbiota-based portals allow associations with external resources like spectral datasets hosted in the IHerbSpec Dataverse to be presented on specimen records. This approach allows spectral datasets to remain hosted in domain-specific repositories while maintaining discoverability through primary biodiversity portals. In parallel, publishing spectral metadata through mechanisms such as GBIF's Integrated Publishing Toolkit enables the presence of spectral data to be indexed and searchable at the aggregator level, even when full spectral files are stored elsewhere. IHerbSpec is actively working toward the development of a Darwin Core extension to formally associate spectral datasets with specimen records—further supporting standardized publication, indexing, and interoperability across global biodiversity platforms.

Together, the IHerbSpec Protocol is designed to support both institutional embedding of spectral data and project-level data repositories as parallel pathways for scaled herbarium spectral digitization. This dual-model approach supports participation across a wide range of organizational capacities while maintaining shared standards for data quality, interoperability, and reuse, ensuring that spectral datasets contribute meaningfully to the evolving global metaherbarium (Davis, 2023).

DISCUSSION

Taken together, herbaria represent the largest scientific database of plant diversity across taxonomic, geographic, and temporal scales (Thiers, 2025). Spectral digitization transfers high-dimensional phenotypic information from these collections directly into the digital landscape of modern biodiversity science. Here, we present a community-developed protocol that standardizes measurement, metadata capture, and data sharing to ensure that these inherently expansive collections can be mobilized as interoperable spectral datasets across institutions and projects.

The development of shared spectral datasets and analytical frameworks strengthens connections among scientific communities. For example, spectral data can be used to model structural and chemical traits that would be difficult or impossible to measure directly, but that can provide valuable information for systematists, ecologists and ecophysiologists (Jetz et al., 2016; Cavender-Bares et al., 2025b). Spectral species libraries derived from dried specimens can also advance conservation by supporting taxonomic identification and trait estimations of preserved material collected during field surveys—enabling more robust evaluation of species composition and functional diversity. This is especially important for tackling issues of taxonomic uncertainty in megadiverse tropical regions and for reducing latitudinal biases in biodiversity knowledge, given that tropical herbaria hold a substantial proportion of the world's plant diversity (Delves et al., 2024; Thiers, 2024; Zhigila et al., 2025).

Another goal for herbarium spectral digitization is to create a bridge to remote sensing. As modeling approaches advance, translating dry-leaf spectra into estimates of fresh functional traits may enable more direct integration with airborne and satellite observations (e.g., (Asner and Martin, 2008; Singh et al., 2015; Wang et al., 2019, 2020; Sapes et al., 2024). Standardized spectral libraries from herbaria have the potential to contribute geographically distributed reference datasets to calibrate canopy-level trait models. In turn, the advancement of remotely sensed biodiversity patterns can guide targeted botanical exploration, herbarium collection, and floristic and systematic research.

Finally, expanding spectral digitization enhances both the accessibility and enduring relevance of herbarium collections. Rather than rendering physical collections obsolete, digital extensions underscore their continued necessity: advances in technology and digital architectures will continue to open new avenues for generating biodiversity datasets; these datasets will be

verifiable and repeatable when they are anchored to vouchered specimens that can be revisited and remeasured using novel technologies. By embracing new analytical applications—while maintaining rigorous standards for specimen stewardship—the herbarium community can demonstrate that collections are not static archives of past utility, but dynamic research infrastructures essential to addressing ongoing ecological, evolutionary, and conservation challenges. In doing so, spectral digitization provides new pathways to safeguard and steward collections in perpetuity as well as contribute to their digital repatriation (McAlvay et al., 2021; Pinheiro et al., 2024) under CARE principles (Carroll et al., 2020). Together, these advances extend the scientific and societal reach of natural history collections and reinforce their central role in addressing global challenges.

Limitations

Some aspects of metadata capture, including assessment of tissue condition and developmental stage, involve subjective judgment and may vary among operators. The protocol provides guidelines and examples in Appendix II of the IHerbSpec Protocol (https://iherbspec.github.io/protocol/appendix2.html) aimed at minimizing subjectivity. Differences in instrumentation and optical configuration can also introduce batch effects, even when protocols are followed. Continued development of validation, harmonization, and anomaly-detection methods will therefore be essential for cross-institutional analyses. In addition, although the IHerbSpec metadata schema was designed for immediate community use, it has not yet undergone formal governance within established biodiversity standards frameworks; coordination with Biodiversity Information Standards (TDWG) will be necessary to ensure long-term alignment.

Projects aimed at developing predictive models for traits such as leaf nitrogen and phosphorus concentrations, carbon fractions (e.g., lignin and cellulose), or micronutrients may require targeted destructive sampling of selected specimens to generate ground-truth data for model building and validation. This is because the taxonomic and ecological breadth over which current models can be reliably applied remains uncertain, and more trait data are needed to improve model generalizability (Cavender-Bares et al., 2025b). Herbarium specimens can be reliable sources of traits such as LMA, leaf thickness, and stable carbon or nitrogen isotopes (Körner et al., 2016; Perez et al., 2020). However, traits observed and estimated from herbarium specimens may not accurately approximate true trait values of living leaf tissue, limiting their direct biological interpretation. Therefore, we envision data for model building and validation from recently-dried, pressed leaves (e.g., (Serbin et al., 2014; Kothari et al., 2023) or silica-dried leaf material (e.g., (Kothari et al., 2024) as well as advancements in modeling from projects conducting herborization experiments (e.g., (Quinteros Casaverde et al., 2024).

Although spectral measurements are meant to be non-destructive, specimen handling and use of a contact probe do entail some risk of damaging the specimens themselves, either through excess illumination or mechanical damage, summarized in Table 2. These risks can be mitigated through training of spectral digitizers and review of protocol guidelines.

Finally, measuring very small, thin, or structurally complex photosynthetic tissues (e.g., graminoids, cushion plants, succulents, terete or needle-like leaves) may require justified deviations from the protocol's minimum sampling requirements, such as collecting fewer replicate measurements when tissue size and/or optical geometry limits the feasibility of making multiple high quality measurements.

Vision for Future Versions

The IHerbSpec Protocol provides a standardized measurement approach designed to evolve alongside advances in instrumentation, digital infrastructure, and analytical methodologies. Continued community use and empirical evaluation across taxa, specimen conditions, instruments, and project scales will be essential for refining metadata standards, improving data harmonization, and expanding guidance for measurements.

The current protocol focuses primarily on vascular plant leaves, reflecting the extensive body of work on leaf spectra in plant physiology, functional ecology, and remote sensing (Wessman et al., 1988; Gamon and Surfus, 1999; Cavender-Bares et al., 2025a). Although the protocol does define tissue class codes for additional plant structures (see IHerbSpec Protocol *Table 4.6: Tissue Descriptor Codes*; e.g., `bract`, `OuterBark`, `PhotosyntheticSucculentStem`), measurement frameworks for these other tissues, as well as other herbarium collections like lichens and bryophytes, remain to be developed (but see (Hadlich et al., 2018; Guzmán Q. et al., 2020; Stasinski, 2021). Establishing specific protocols for these materials represents a key direction for future community coordination.

Ongoing development is facilitated through the public GitHub repository (https://github.com/IHerbSpec/iherbspec.github.io), which hosts the protocol code and deploys the IHerbSpec website (https://iherbspec.github.io), providing a transparent platform for revision and community engagement. As participation expands, structured governance—such as subcommittees and leadership teams—will be necessary to balance inclusivity with efficient coordination. Researchers, students, and collections professionals interested in contributing are encouraged to visit the IHerbSpec Contact page (https://iherbspec.github.io/contact.html) to learn about joining the working group and participating in community discussions.

ACKNOWLEDGEMENTS

We gratefully acknowledge botanical collectors as well as the herbaria, curators, collection managers, and staff who steward biodiversity collections. We thank Michaela Schmull for helpful comments on the manuscript. This project was supported by the U.S. National Science Foundation Biology Integration Institute ASCEND (DBI-2021898), the Harvard University Herbaria, iDigBio (DBI-2027654), and the Missouri Botanical Garden's *Revolutionizing Species Identification* project. FD is funded by the Amazonas State Research Support Foundation (FAPEAM) *Women in Science Program* (SPECTRA POP project, process number 01.02.016301.04984/2024-17) and the German Federal Ministry of Research, Technology, and Space (BMBF) grant 01LK2101D. JEM is funded by NSF DEB-2442433. BB is funded by ETH Zurich Career Seed Award (25-1 Seed-006). CCV acknowledges support from the International Association for Plant Taxonomy (IAPT), 2025 research grant round. JLV is supported by Danish National Research Foundation grant DNRF179 to HT.

AUTHOR CONTRIBUTIONS

JCB initiated IHerbSpec and the creation of a protocol. IHerbSpec conceptualized the protocol and DMW led its development. DMW wrote this manuscript with contributions from MWA, FMD, JAG, JAK, and CON. KB drafted Table 2 with input from EMV. All authors contributed to protocol development, reviewed the manuscript, and approved the final version.

LITERATURE CITED

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

Allaire, J. J., C. Teague, C. Scheidegger, Y. Xie, C. Dervieux, and G. Woodhull. 2025. Quarto.

Asner, G. P., and R. E. Martin. 2008. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing of Environment* 112: 3958–3970.

Boughalmi, K., P. G. Santacruz Endara, L. A. Bennett, M. Ecarnot, S. Bazan, D. Bastianelli, L. Bonnal, and T. L. P. Couvreur. 2025. Discrimination of Annonaceae using herbarium leaf reflectance spectra under limited sample size conditions. 2025.09.02.673631.

Buono, D., and D. C. Albach. 2023. Infrared spectroscopy for ploidy estimation: An example in two species of Veronica using fresh and herbarium specimens. *Applications in Plant Sciences* 11: e11516.

Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19.

Cavender-Bares, J., J. E. Meireles, J. J. Couture, M. A. Kaproth, C. C. Kingdon, A. Singh, S. P. Serbin, 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., J. E. Meireles, J. Pinto-Ledezma, Peter B. Reich, M. Schuman, P. A. Townsend, and A. Trowbridge. 2025a. Spectral biology across scales in changing environments. *Ecology* 105.

Cavender-Bares, J., D. M. White, N. I. Ahlstrand, M. W. Austin, D. Bastianelli, S. Bazan, K. Boughalmi, et al. 2025b. Next-generation specimen digitization: capturing reflectance spectra from the world's herbaria for modeling plant biology across time, space, and taxa. *New Phytologist*.

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

Cruickshank, R. H., and L. Munck. 2011. It's barcoding Jim, but not as we know it. *Zootaxa* 2933: 55–56.

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

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

Delves, J., J. Albán-Castillo, A. Cano, C. Fernández Aviles, E. Gagnon, P. Gonzáles, S. Knapp, et al. 2024. Small and in-country herbaria are vital for accurate plant threat assessments: A case study from Peru. *Plants, People, Planet* 6: 174–185.

Durgante, F. M., N. Higuchi, A. Almeida, and A. Vicentini. 2013. Species Spectral Signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. *Forest Ecology and Management* 291: 240–248.

Enquist, B. J., R. Condit, R. K. Peet, M. Schildhauer, and B. M. Thiers. 2016. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprint*.

Fine, P. V. A., D. Salazar, R. E. Martin, M. R. Metz, T. M. Misiewicz, and G. P. Asner. 2021. Exploring the links between secondary metabolites and leaf spectral reflectance in a diverse genus of Amazonian trees. *Ecosphere* 12: e03362.

Gamon, J. A., and J. S. Surfus. 1999. Assessing leaf pigment content and activity with a reflectometer. *New Phytologist* 143: 105–117.

Guzmán Q., J. A., K. Laakso, J. C. López-Rodríguez, B. Rivard, and G. A. Sánchez-Azofeifa. 2020. Using visible-near-infrared spectroscopy to classify lichens at a Neotropical Dry Forest. *Ecological Indicators* 111: 105999.

Hadlich, H. L., F. M. Durgante, J. dos Santos, N. Higuchi, J. Q. Chambers, and A. Vicentini. 2018. Recognizing Amazonian tree species in the field using bark tissues spectra. *Forest Ecology and Management* 427: 296–304.

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

IHerbSpec. 2026. Protocol for Spectral Digitization of Herbarium Specimens v1.2.1.

Jacquemoud, S., S. L. Ustin, J. Verdebout, G. Schmuck, G. Andreoli, and B. Hosgood. 1996. Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment* 56: 194–202.

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

Kattge, J., K. Ogle, G. Bönisch, S. Díaz, S. Lavorel, J. Madin, K. Nadrowski, et al. 2011. A generic structure for plant trait databases. *Methods in Ecology and Evolution* 2: 202–213.

Körner, C., S. Leuzinger, S. Riedl, R. T. Siegwolf, and L. Streule. 2016. Carbon and nitrogen stable isotope signals for an entire alpine flora, based on herbarium samples. *Alpine Botany* 126: 153–166.

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

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

589 Kothari, S., and A. K. Schweiger. 2022. Plant spectra as integrative measures of plant
590 phenotypes. *Journal of Ecology* 110: 2536–2554.

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

594 Kunkel, D. M., B. L. Long-Fox, C. Pittman, J. Portmann, M. Sheik, J. M. Bates, A. Bentley, et
595 al. 2025. Integrating biological and environmental data to solve key scientific and societal
596 challenges. *BioScience*: biaf150.

597 Laliberté, E., and R. Soffer. 2018. Measuring spectral reflectance and transmittance (350-2500
598 nm) of small and/or narrow leaves using the Spectra Vista Corporation (SVC) DC-R/T
599 Integrating Sphere V.3. *protocols.io*.

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

603 Lendemer, J., B. Thiers, A. K. Monfils, J. Zaspel, E. R. Ellwood, A. Bentley, K. LeVan, et al.
604 2020. The Extended Specimen Network: A Strategy to Enhance US Biodiversity
605 Collections, Promote Research and Education. *BioScience* 70: 23–30.

606 McAlvay, A. C., C. G. Armstrong, J. Baker, L. B. Elk, S. Bosco, N. Hanazaki, L. Joseph, et al.
607 2021. Ethnobiology Phase VI: Decolonizing Institutions, Projects, and Scholarship.
608 *Journal of Ethnobiology* 41: 170–191.

609 Meireles, J. E., J. Cavender-Bares, P. A. Townsend, S. Ustin, J. A. Gamon, A. K. Schweiger, M.
610 E. Schaepman, et al. 2020. Leaf reflectance spectra capture the evolutionary history of
611 seed plants. *New Phytologist* 228: 485–493.

612 Mersni, J., S. Bazan, E. Baby, V. Bonnal, and D. Bastianelli. 2025. Protocol for NIR spectra
613 acquisition on herbarium specimen.

614 Neto-Bradley, B. M., P. Bonnet, H. Goëau, A. Joly, J. Cavender-Bares, and D. A. Coomes. 2025.
615 Using reflectance spectra and Pl@ntNet to identify herbarium specimens: a case study
616 with Lithocarpus. *New Phytologist*.

617 Paton, A. J., G. K. Ameka, A. Antonelli, A. Asase, R. L. Barrett, A. Bogaerts, D. Cardoso, et al.
618 2025. Life after herbarium digitisation: Physical and digital collections, curation and use.

Plants, People, Planet.

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

Pérez-Harguindeguy, N., S. Díaz, E. Garnier, S. Lavorel, H. Poorter, P. Jaureguiberry, M. S. Bret-Harte, et al. 2016. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 64: 715–716.

Pinheiro, F. da C., R. C. Forzza, P. M. Leitman, and J. Prado. 2024. The REFLORA Program: implementation, repatriation, and creation of the REFLORA Virtual Herbarium as a tool for biodiversity studies. *Biota Neotropica* 24: e20241701.

Quinteros Casaverde, N. L., S. P. Serbin, and D. C. Daly. 2024. Changes in spectral signature of leaves after desiccation: Implications for the prediction of leaf traits and plant-soil interaction in herbarium samples. 2024.01.21.576284.

Sapes, G., L. Schroeder, A. Scott, I. Clark, J. Juzwik, R. A. Montgomery, J. A. Guzmán Q, and J. Cavender-Bares. 2024. Mechanistic links between physiology and spectral reflectance enable previsual detection of oak wilt and drought stress. *Proceedings of the National Academy of Sciences* 121: e2316164121.

Serbin, S. P., A. Singh, B. E. McNeil, C. C. Kingdon, and P. A. Townsend. 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecological Applications* 24: 1651–1669.

Singh, A., S. P. Serbin, B. E. McNeil, C. C. Kingdon, and P. A. Townsend. 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecological Applications* 25: 2180–2197.

Stasinski, L. 2021. Assessing Plant and Lichen Diversity Using Reflectance Spectra. University of Maine, Orono, Maine.

Stasinski, L., D. M. White, P. R. Nelson, R. H. Ree, and J. E. Meireles. 2021. Reading light: leaf spectra capture fine-scale diversity of closely related, hybridizing arctic shrubs. *New Phytologist* 232: 2283–2294.

Stefanski, A., E. E. Butler, L. J. Williams, R. Bermudez, J. A. Guzmán Q., A. Larson, P. A. Townsend, et al. 2025. All the light we cannot see: Climate manipulations leave short and long-term imprints in spectral reflectance of trees. *Ecology* 106: e70048.

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

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

Vance, C. K., D. R. Tolleson, K. Kinoshita, J. Rodriguez, and W. J. Foley. 2016. Near Infrared Spectroscopy in Wildlife and Biodiversity. *Journal of Near Infrared Spectroscopy* 24: 1–25.

Vasconcelos, C. C., I. D. K. Ferraz, F. M. Durgante, F. Wittmann, J. Schöngart, M. T. F. Piedade, J. L. C. Camargo, and M. H. Terra-Araujo. 2025. Ecclinusa nervosa (Sapotaceae, Chrysophylloideae), a New Species Discovered in Central Amazonia. *Systematic Botany* 50: 55–66.

Wang, Z., A. Chlus, R. Geygan, Z. Ye, T. Zheng, A. Singh, J. J. Couture, et al. 2020. Foliar functional traits from imaging spectroscopy across biomes in eastern North America. *New Phytologist* 228: 494–511.

Wang, Z., P. A. Townsend, A. K. Schweiger, J. J. Couture, A. Singh, S. E. Hobbie, and J. Cavender-Bares. 2019. Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. *Remote Sensing of Environment* 221: 405–416.

Wessman, C. A., J. D. Aber, D. L. Peterson, and J. M. Melillo. 1988. Remote sensing of canopy chemistry and nitrogen cycling in temperate forest ecosystems. *Nature* 335: 154–156.

White, D. M., J. Cavender-Bares, C. C. Davis, J. A. Guzmán Q., S. Kothari, J. M. Robles, and J. E. Meireles. 2025. Seeing herbaria in a new light: leaf reflectance spectroscopy unlocks trait and classification modeling in plant biodiversity collections. *New Phytologist*.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

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

Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018.

Zhigila, D. A., R. J. Schmidt-Knapik, B. M. Thiers, S. D. Abdul, S. Abdullahi, A. A. AbdulRahaman, E. I. Aigbokhan, et al. 2025. Biodiversity science is improved when silent herbaria speak. *Plants, People, Planet*.

682 TABLES

683 Table 1: Structure and content of the IHerbSpec Protocol

Protocol Part	Content Description
Preface	Abstract, frontmatter, and details about IHerbSpec.
Part 1: Overview	Overview of the core measurement procedure and associated metadata fields, distinguishing *required* and *recommended* elements of the protocol.
Part 2: Workflow	Stepwise workflow for spectral measurement and metadata collection (Fig. 2).
Part 3: Filename Conventions and Formats	Standardized 'Full' and 'Simple' file-naming conventions for spectral data files (tissue, background, and calibration spectra), designed to link measurements and ensure traceability. The simple filename convention is streamlined for use during measurement sessions and should be converted into full filenames at the point of data archival and sharing to enable interoperability.
Part 4: Metadata and Databasing	Metadata schema for herbarium spectral reflectance data. Three metadata tables are organized into logical groups of session-, specimen-, and tissue metadata. Additional tables describe controlled vocabularies for tissue classes (e.g. *adaxial* leaf, *abaxial* leaf, bract, etc), tissue developmental stage (e.g. *young*, *mature*, *unknown*, etc.), tissue descriptors (e.g. *AlcoholPresent*, *LichenPresent*, etc.), and background and white reference class codes (e.g. *PaperBackground*, *BlackBackground*). The part concludes with guidance for data management, archiving, and sharing, including recommendations to use the IHerbSpec Dataverse community repository.
Part 5: Instrumentations and Materials	Guidance on instrumentation, materials, and measurement configuration.
Part 6: Tissue Selection	Decision tree and considerations for selecting suitable herbarium tissues for spectral measurement.
Appendix I	Data-driven guidance on the number of measurements per specimen.
Appendix II	Examples and guidance for recording biological and preparation-related sources of variation affecting spectral data in the IHerbSpec metadata.

684

685 Table 2: Estimated incidence of accidental specimen damage during contact-based spectroscopy
686 of herbarium collections.

Type of destructive effect	**Estimated relative frequency (500 scanned specimens*)**	**Mitigation measures**
Leaf detachment *(detachment at the petiole–stem junction)*	2-3 / 500	Prefer already detached leaves (e.g., mounted or stored in protective sleeves or packets) when available; insert background materials carefully to avoid mechanical stress at attachment points (leaf base, petiole base), particularly in fragile tissues.
Major cracks *(visible structural fracture across lamina)*	5 / 500	Select flat, structurally intact leaf areas; assess tissue resistance before probe contact; maintain proper probe alignment with minimal pressure (especially when using small ~10 mm probes); avoid lateral movement once contact is made; and use extra care near fragile (e.g., deteriorated margins, damaged areas) or unsupported regions (e.g., areas near prominent veins).
Superficial cracks *(minor surface fissures without full rupture)*	10 / 500	
Probe perforation	2 / 500	
Partial border breakage *(fragment loss along leaf margin)*	5 / 500	
Burning or discoloration of leaf from heat of contact probe	1 / 500	Test instrumentation on thin or delicate leaves prior to measurement; use low light intensity with longer integration times to minimize heat exposure; and regularly clean the probe surface to prevent transfer of particles, dust, or residues between specimens.

687

688 *Data compiled from Boughalmi et al. (2025) and additional spectroscopy campaigns conducted
689 by the same research group. Frequencies represent the number of damaged specimens relative to
690 a cumulative total of 500 herbarium specimens scanned across two spectroscopy campaigns

conducted between 2024 and 2026 at Muséum National d'Histoire Naturelle (P), Naturalis Biodiversity Center (WAG, L), and Université de Montpellier (MPU).

694 FIGURE LEGENDS
695
696 **Fig. 1. Specimen and tissue selection workflow and bench measurement setup.** (A)
697 Specimen and tissue selection decision tree. The first step (top) prioritizes specimens that are
698 generally suitable for measurement (mature, intact, flat, and with minimally contaminated areas),
699 unless the specimen is of unique importance. The second step prioritizes tissues that permit
700 insertion of a thin non-reflective black background behind the target tissue. As stated in Part I of
701 the IHerbSpec Protocol, insertion of a black background is required whenever possible to
702 minimize spectral contamination from mounting materials. However, this is not always feasible
703 for certain mounted or otherwise valuable specimens; in such cases, measurements may still be
704 collected provided that background conditions and potential sources of contamination are
705 explicitly documented in the metadata. The final step emphasizes selecting tissue areas that are
706 as representative and uncontaminated as possible (e.g., avoiding glue, epiphylls, or herbivory
707 where feasible). (B) Bench setup for spectral measurement. Figure A reproduced from
708 IHerbSpec Protocol (CC BY 4.0).
709
710 **Fig. 2. Schematic of the IHerbSpec Protocol v1.2 measurement workflow.** Steps 1–4
711 illustrate instrument and materials setup, measurement of white and black references and
712 calibrated standards, and recording of session-level metadata associated with these activities.
713 Steps 5–9 illustrate the tissue measurement sequence, filename and data quality validation,
714 scoring of specimen- and tissue-level metadata, and annotation of project information on
715 herbarium specimens. The protocol requires a minimum of three and a recommendation of five
716 or more measurements (see Appendix I of the protocol) for both adaxial and abaxial leaf surfaces
717 across the specimen, providing sufficient tissue is available and suitable for measurement. Plants

718 with small, terete, or otherwise challenging leaves for measurement might not be able to meet
719 this requirement (see *Vision for Future Versions* section). Figure reproduced from IHerbSpec
720 Protocol (CC BY 4.0).
721
722 **Fig. 3. The effect of background on tissue spectral profiles.** The IHerbSpec Protocol requires
723 placement of a low-reflectance black background beneath tissues to minimize background
724 contamination in reflectance data. (A) Reflectance spectra of reference and background materials
725 measured during a session, including the white reference panel (White, colored blue), the black
726 background (Black), gray backgrounds, and representative herbarium mounting papers. These
727 background materials exhibit distinct spectral signatures, particularly outside the visible range
728 (400–700 nm). Plots show full-range (350–2500 nm), unprocessed reflectance data measured
729 with the Spectral Evolution NaturaSpec and Surface Probe. (B) Difference spectra showing
730 adaxial leaf reflectance measured on a black background minus reflectance measured at the same
731 leaf surface location on alternative backgrounds, illustrating wavelength-dependent background
732 effects on tissue measurements. (C–D) Reflectance profiles of the same leaf measured on
733 different backgrounds for two target classes: leaf abaxial (TCAB; panel C), and leaf adaxial
734 (TCAD; panel D) surfaces. Measurements taken over black and gray backgrounds consistently
735 show lower reflectance than those taken over white reference material and white to yellowed
736 herbarium papers. Across panels, background effects are most pronounced in the 800–1800 nm
737 region, corresponding mostly to the near-infrared (1100–2000 nm), with weaker but detectable
738 effects in the visible (400–700 nm) and longer-wavelength SWIR (2000–2500 nm).
739

Fig. 4. Instrument and spatial heterogeneity in herbarium leaf reflectance measurements. Reflectance spectra collected from the same dried leaf (*Betula papyrifera;* NEBC00636882; tissue condition scoring in IHerbSpec Protocol Appendix II) using two spectroradiometers with different optical configurations: a Spectral Evolution NaturaSpec with a 2 mm field of view and an SVC HR-1024i with a 22 mm field of view. Spectra were trimmed to 400–2400 nm and plotted with the *ggplot2* R package (Wickham, 2016). Colored circles on the leaf image indicate measurement locations corresponding to the plotted spectra (circles not drawn to scale). NaturaSpec measurements include normal lamina tissue, midvein, pathogen-damaged tissue, off-target measurement including some black background, and reference measurements of the black background and white standard. The SVC measurement represents a standard lamina measurement integrating reflectance over a substantially larger sampling area. Spectral regions are shaded to indicate the visible range (VIS, 400–700 nm; rainbow band), near-infrared (NIR, ~700–1400 nm), and shortwave infrared (SWIR1, ~1400–1900 nm; SWIR2, ~2000–2500 nm). Differences among spectra illustrate how optical configuration and fine-scale tissue heterogeneity influence reflectance profiles across the visible to shortwave infrared spectrum.