

Reading light: leaf spectra capture fine-scale diversity of closely related, hybridizing arctic shrubs

Lance Stasinski^{1*} , Dawson M. White^{2*} , Peter R. Nelson^{3,4} , Richard H. Ree²  and José Eduardo Meireles^{1,5} 

¹School of Biology and Ecology, University of Maine, Orono, ME 04469, USA; ²Department of Science and Education, Field Museum, Chicago, IL 60605, USA; ³Schoodic Institute, Winter Harbor, ME 04693, USA; ⁴School of Forest Resources, University of Maine, Orono, ME 04469, USA; ⁵Maine Center for Genetics in the Environment, University of Maine, Orono, ME 04469, USA

Authors for correspondence:

Lance Stasinski

Email: lance.stasinski@gmail.com

Dawson M. White

Email: dawson.white@gmail.com

Received: 9 July 2021

Accepted: 6 September 2021

New Phytologist (2021) 232: 2283–2294

doi: 10.1111/nph.17731

Key words: biodiversity detection, classification, *Dryas*, hybridization, leaf reflectance spectroscopy, population structure.

Summary

- Leaf reflectance spectroscopy is emerging as an effective tool for assessing plant diversity and function. However, the ability of leaf spectra to detect fine-scale plant evolutionary diversity in complicated biological scenarios is not well understood.
- We test if reflectance spectra (400–2400 nm) can distinguish species and detect fine-scale population structure and phylogenetic divergence – estimated from genomic data – in two co-occurring, hybridizing, ecotypically differentiated species of *Dryas*. We also analyze the correlation among taxonomically diagnostic leaf traits to understand the challenges hybrids pose to classification models based on leaf spectra.
- Classification models based on leaf spectra identified two species of *Dryas* with 99.7% overall accuracy and genetic populations with 98.9% overall accuracy. All regions of the spectrum carried significant phylogenetic signal. Hybrids were classified with an average overall accuracy of 80%, and our morphological analysis revealed weak trait correlations within hybrids compared to parent species.
- Reflectance spectra captured genetic variation and accurately distinguished fine-scale population structure and hybrids of morphologically similar, closely related species growing in their home environment. Our findings suggest that fine-scale evolutionary diversity is captured by reflectance spectra and should be considered as spectrally-based biodiversity assessments become more prevalent.

Introduction

Biodiversity faces significant threats worldwide (Bellard *et al.*, 2012; Pimm *et al.*, 2014; Isbell *et al.*, 2015), so it is critical that we improve our ability to assess diversity in order to understand and conserve ecological and evolutionary processes (Atwater & Callaway, 2015; Cavender-Bares *et al.*, 2017). Although biodiversity typically is assessed at the species level, it is now clear that fine-scale diversity – the genetic and phenotypic diversity present within species and taxonomic complexes – also warrants evaluation and monitoring. Fine-scale diversity facilitates taxon-specific adaptive potential (Hoffmann *et al.*, 2017) and it can have significant positive effects on community structure and ecosystem productivity (Crawford & Rudgers, 2013; Atwater & Callaway, 2015; Raffard *et al.*, 2019). The influences and consequences of fine-scale diversity are particularly important to consider among Arctic-alpine plants because they frequently exhibit high within-species genetic diversity (Grundt *et al.*, 2006).

In addition to fine-scale diversity, hybrid taxa are another dimension of biodiversity requiring attention because hybrids may exhibit ecologically or evolutionarily significant capabilities, such as the effects of transgression on adaptive potential and invasiveness (Rieseberg *et al.*, 2003, 2007; Abasolo *et al.*, 2012; Dittrich-Reed & Fitzpatrick, 2013; Gallego-Tévar *et al.*, 2018). Alternatively, phenotypically intermediate hybrids can enable the genetic swamping and extinction of parent taxa (Todesco *et al.*, 2016). These low levels of biological organization are most frequently, and appropriately, defined using genomic data, but it is imperative that we advance our methodologies for estimating and monitoring this fine-scale diversity to maximize biodiversity data in a rapidly changing world.

Leaf reflectance spectroscopy is emerging as a powerful tool to assess plant phylogenetic and functional diversity, and to monitor how it changes over space and time (Turner, 2014; Jetz *et al.*, 2016; Cavender-Bares *et al.*, 2017; Schweiger *et al.*, 2018; Serbin *et al.*, 2019; Meireles *et al.*, 2020). This approach is based on the principle that the electromagnetic radiation reflected off leaves (400–2500 nm) carries information about their structural and

*These authors contributed equally to this work.

chemical traits (Curran, 1989; Ustin *et al.*, 2009; Cavender-Bares *et al.*, 2017). We know that pigments absorb in the visible region (VIS; 400–700 nm), whereas light in the near infrared region (NIR; 700–1100 nm) is scattered by leaf anatomical, tissue, water, and surface features, and light in the short-wave infrared region (SWIR; 1400–2500 nm) is scattered and absorbed by anatomical features and biochemicals such as cellulose, phenolics and water (Gates *et al.*, 1965; Carter, 1991; Ustin *et al.*, 2009; Kokaly & Skidmore, 2015; Fang *et al.*, 2017). Thus, leaf spectra are dense, complex and dynamic phenomic datasets influenced by both environmental and genetic factors. They can be effectively leveraged using supervised statistical models to estimate phenotypic traits (Ustin & Gamon, 2010) or estimate diversity (Durgante *et al.*, 2013; Lang *et al.*, 2017; Schweiger *et al.*, 2018; Meireles *et al.*, 2020).

Leaf spectra mostly have been used to differentiate groups at or above the species level, and these taxa are usually classified with high accuracy (Durgante *et al.*, 2013; Lang *et al.*, 2017; Wang & Gamon, 2019; Meireles *et al.*, 2020). The few studies that have assessed fine-scale diversity have shown variable success. Cavender-Bares *et al.* (2016) classified leaves from four *Quercus oleoides* populations with low accuracy, but Madritch *et al.* (2014) were able to classify 79 genotypes of *Populus tremuloides* with moderate to high accuracy. Thus, it is important to continue to test the limits of spectral detection using genomic data and small spatial scales for closely related taxa, populations and hybrids.

We expect traits to be able to separate populations and species, yet hybrid individuals will not necessarily form a phenotypically cohesive group because individuals will inherit and express different sets of parental alleles resulting in variable phenotypes and weakly correlated traits (Guo *et al.*, 2004; Cheng *et al.*, 2011). Although such individuals are difficult to classify statistically using macro-morphological data such as organ sizes and shapes (Field *et al.*, 2009; Abasolo *et al.*, 2012), reflectance spectra are proving to be detailed enough to identify such challenging hybrid taxa. For instance, an analysis of bloodwoods (*Corymbia*), showed that their weedy, morphologically complex hybrids could not be classified accurately using 30 morphological traits, but classifications of these same hybrids using spectral data was 72–100% accurate (Abasolo *et al.*, 2012, 2013). Reflectance spectra also have been used to accurately detect *Citrus* (Páscoa *et al.*, 2018) and *Populus* (Deacon *et al.*, 2017) hybrids.

In this study, we estimated the ability of leaf spectra to detect phylogenetic divergence and low taxonomic levels in a complicated, yet frequently encountered, biological scenario of closely related, co-occurring, hybridizing plants. *Dryas octopetala* L. s.l. has been described as one to nine species and numerous subspecific taxa representing geographically and/or ecologically distinct lineages, many of which form hybrids with intergrading morphologies (Hultén, 1959; Skrede *et al.*, 2006; Elven *et al.*, 2011). Two noteworthy taxa (currently described as species) in this complex are *Dryas ajanensis* Juzepczuk ssp. *beringensis* Jurtzev (hereafter *D. ajanensis*) and *D. alaskensis* A.E. Porsild (in accordance with the *Flora of North America* treatment; Springer & Parfitt, 2014), which were established as ecotypes through the exemplary experiments of McGraw & Antonovics (1983). These species are

very close relatives, possibly sister species, that are believed to have diverged during the Pleistocene (Hultén, 1959). Individuals expressing intermediate leaf morphologies, long presumed to be hybrids, are found in tight contact zones between habitats (Hultén, 1968; Max *et al.*, 1999).

We generated spectral and genomic datasets for *Dryas* individuals, collected across six alpine sites in the interior of Alaska, and used them to determine: (1) Can we use spectra to accurately classify species, hybrids and genetically defined populations? and (2) To what degree does spectral variation correlate with genetic variation at these low taxonomic levels? These closely related, co-occurring, hybridizing taxa provide a challenging system to explore the boundaries of biodiversity detection via spectroscopy.

Materials and Methods

Taxon sampling

We sampled *D. ajanensis*, *D. alaskensis* and putative hybrids from six alpine sites in the interior of Alaska, USA (Fig. 1). *Dryas ajanensis* is an abundant species found in isolated patches on dry, rocky fellfields across boreal North America. It is identified readily by small tomentose leaves (5–15 mm) with rust-colored ‘scales’ (multicellular, feathered hairs) occurring on the midvein on the bottom (abaxial) side of the leaf. *Dryas alaskensis* occurs in wet tundra microhabitats in Alaska and the Yukon Territories. It has larger leaves (15–50 mm) with less pubescence and stipitate glandular trichomes on the abaxial midvein, and sometimes with adaxial wax secretions (Hultén, 1959, 1968; McGraw & Antonovics, 1983). Individual plants displaying both kinds of adaxial midvein pubescence (rarely was this on the same leaf) were identified in the field as hybrids.

We collected c. 10–15 leaves from evenly spaced *Dryas* individuals along a 100-m transect (except that no transect was used at Wickersham Dome B) that traversed wet and dry habitats, and two voucher specimens per taxon were collected at each site and deposited at the Field Museum herbarium. We collected leaves from the following number of individual plants from each site; Bison Gulch (BG: 20 *D. ajanensis*), Eagle Summit (ES: 19 *D. alaskensis*, 21 *D. ajanensis*, two hybrids), Murphy Dome (MD: 20 *D. ajanensis*), Twelve Mile Summit (TM: 22 *D. alaskensis*, 20 *D. ajanensis*, two hybrids), Wickersham Dome A (WDA: 20 *D. ajanensis*) and Wickersham Dome B (WDB: 11 *D. alaskensis*, 16 *D. ajanensis*, five hybrids; Fig. 1; Supporting Information Table S1).

Genetic structure and ancestry

We isolated DNA from silica-dried leaf samples and prepared GBS libraries using the *ApeKI* restriction endonuclease (Elshire *et al.*, 2011). Libraries were prepared and sequenced at the University of Wisconsin DNA Sequencing Center on an Illumina NovaSeq 6000 (2 × 150-bp reads). We aligned reads to the *D. drummondii* genome (GCA_003254865.1) and called single nucleotide polymorphisms (SNPs) using ipyrad (Eaton & Overcast, 2020). A description of steps from DNA isolation to SNP calling is available in Methods S1.

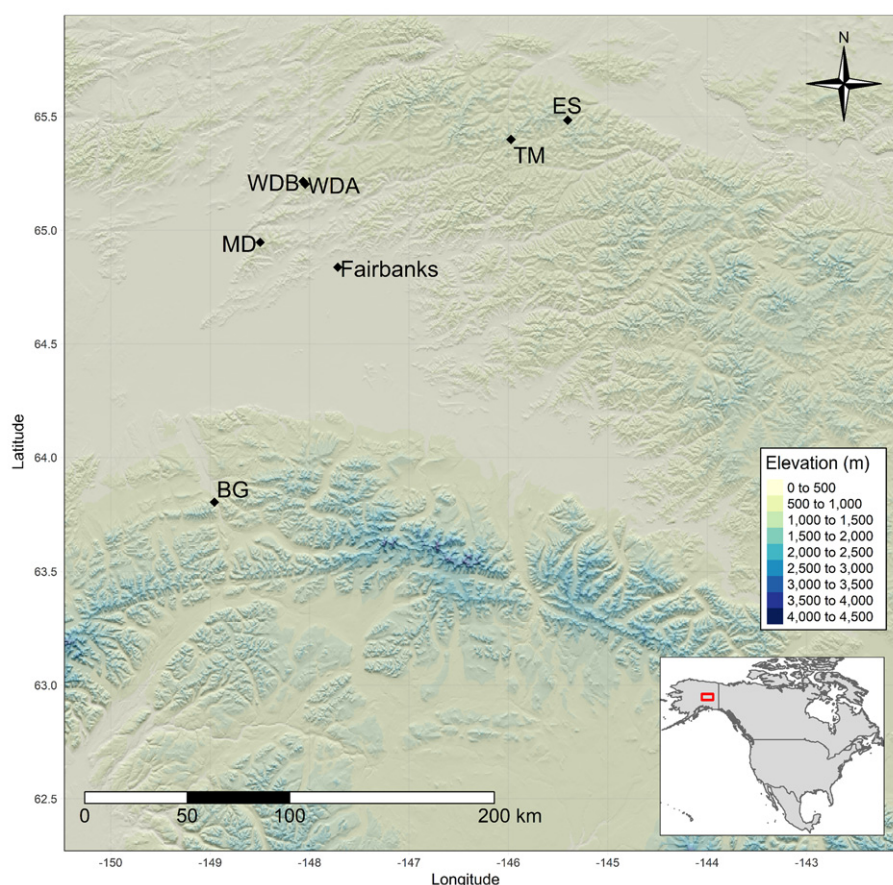


Fig. 1 Map of study area in Alaska, USA. Sampling site acronyms: BG, Bison Gulch; ES, Eagle Summit; MD, Murphy Dome; TM, Twelve Mile Summit; WDA, Wickersham Dome Site A; WDB, Wickersham Dome Site B. The maximum distance between sites was 252 km, the minimum distance was 3 km and the average distance was 118 km. The red rectangle on the inset map of North America represents the study area.

In order to establish the population genetic structure of this system, we first conducted principal components analysis (PCA) in R/ADGENET v.2.1.3 using all single nucleotide polymorphisms (SNPs) to summarize genetic variation and discontinuities among individuals and sites (Jombart, 2008; R Core Team, 2021). Next, we used the model-based clustering method STRUC- TURE to estimate the number of populations in the dataset via each sample's proportional assignment to a set number of inferred ancestral groups (Pritchard *et al.*, 2000). For one to 10 groups (K), we completed 10 replicate runs of 1 million generations plus 500 000 generations of burn-in. STRUC- TURE's Q -value is the proportion of an individual's ancestry from ancestral group K , therefore $K=2$ represents an index of genomic ancestry from *D. ajanensis* or *D. alaskensis*. We then completed separate STRUC- TURE runs for each species to delineate populations. To gauge further support for the separation of lineages by populations or sites, we dropped the hybrid samples and reconstructed a maximum-likelihood phylogeny from the full, concatenated SNP dataset using IQ-TREE v.1.6.12 with the GTR+ASC model of nucleotide evolution and 1000 ultrafast bootstrap pseudoreplicates (Hoang *et al.*, 2018; Minh *et al.*, 2020).

Spectral data

We collected leaves in labeled tea bags, placed them within plastic bags containing ample silica gel desiccant, and left them to dry

for 36–60 h before scanning. Water and water vapor have strong effects on spectral reflectance, especially in the near and short-wave infrared (NIR and SWIR) regions (Carter, 1991), so thoroughly drying the leaves with silica gel reduces the effect of this environmental variable as well as reveals distinctive anatomical and chemical features in these spectral regions (Costa *et al.*, 2018; Páscoa *et al.*, 2018). Although the limitations of fieldwork imposed variability in the duration of drying times, we are confident that these small leaves with thin cuticles were fully desiccated in <24 h of storage in silica gel. This is based on the empirical work of Carrió & Rosselló (2014), as well as our methods to verify the effects of residual water – which are described below and in Methods S2. Reflectance measurements (scans hereafter) were taken after a warm-up period of ≥ 15 min using a PSR+ portable spectroradiometer (Spectral Evolution, Haverhill, MA, USA) with reflectance contact probe (with tungsten halogen light source) and leaf probe clip. A Spectralon® white reference was recorded every five samples to recalibrate, then the leaf clip was reversed to the black background during spectral readings.

We scanned the leaves in three stages owing to the small size of the leaves. First, we arranged two to five leaves across the leaf clip to cover as much of the entire field of view as possible without overlap and with the adaxial leaf surface facing the probe. We then scanned spectral reflectance from 350 to 2500 nm two times to include possible variability from the instrument or slight variations in the output from the light source. For the second stage,

we added one to three leaves to cover more of the exposed black background and took two more scans. We added one to three more leaves for the third stage, resulting in a total of six scans per specimen. The addition of leaves was implemented to explore the two possible shortcomings of measuring leaves that were smaller than the spot size of the reflectance contact probe; those being exposed background that generates noise in the spectra and partially overlapping leaves that alter the shape and magnitude of the reflected spectrum (Neuwirthová *et al.*, 2017).

Spectral analyses

We processed the reflectance spectra by removing erroneous scans that had reflectance values > 1.0 . The resulting dataset consisted of 1045 scans at 1-nm resolution representing 178 individual plants (Table S2). We then trimmed the spectra to a length of 400–2400 nm to remove regions of higher noise (350–399 nm, 2401–2500 nm; Cavender-Bares *et al.*, 2016). All manipulations of the spectra were conducted using R/SPECTROLAB (Meireles *et al.*, 2017). These processed spectra were used in the following analyses, and all six scans (or fewer if scans had reflectance values > 1.0) were used to represent the respective specimen.

We classified species (with and without hybrids), populations, collection sites, and a combination of species identity and collection site (with and without hybrids) using partial least squares discriminant analysis (PLS-DA), a multivariate analysis that classifies observations from PLS regression on indicator variables (Chevallier *et al.*, 2006). This method works well with high dimensional multicollinear datasets such as those acquired via leaf reflectance spectroscopy (Barker & Rayens, 2003; Cavender-Bares *et al.*, 2016). We executed the following PLS-DA procedure independently for each classification unit (species, populations, sites, and species plus site). For classifying species, populations and sites, we partitioned the data such that a random 80% of the spectral data was used as a training set, and the remaining 20% was used for a testing set. However, for classifying the combination of species and location, we split the data so that 50% was used for testing and 50% was used for training because low sample sizes in some of the classes, such as hybrids belonging to a particular location, prevented the use of smaller allocations of data for testing the models. Furthermore, the spectra exhibited large variation at the scale of the individual, so we chose to treat each scan as an independent measurement in these models.

We used a two-step procedure to account for class imbalance. In the first step, we ran 100 iterations of PLS-DA with 10-fold cross-validation repeated three times. Each iteration used an independent partition of the data (see above for how the partition was made) in which the training set was downsampled, and the final model of the respective iteration was tested against the testing set to determine model accuracy. The number of components for which the average overall accuracy was maximized was used as the optimal number of components in the second step. This second step was the same as the first except that the training data was upsampled in each iteration and the number of components used was equal to the number chosen from the first step to

prevent overfitting. Finally, the overall classification accuracy was calculated as the mean accuracy extracted from confusion matrices across all the 100 iterations used with the upsampling procedure (second step).

We evaluated the potential effect of differences in drying times on our classifications by using the same PLS-DA procedure, with 10 iterations instead of 100, to classify the site of collection using spectra with water absorption features and the visible spectrum removed. We assessed site-based classification because it covaried with drying time, and we removed the water absorption features and the visible spectrum because pigments and water content are most likely to be affected by drying (Carter, 1991; Chen *et al.*, 2012). The specific wavelengths removed from the spectra for this analysis were the visible spectrum 400–749 nm and water absorption features at 960–980 nm, 1170–1190 nm, 1235–1255 nm, 1300–1460 nm, 1750–2030 nm, 2040–2060 nm, 2135–2155 nm and 2153–2173 nm (Thenkabail & Lyon, 2016; Wang & Gamon, 2019).

Additionally, we created a PLS beta regression model, implemented in R/PLSRBETA, to predict the proportion of *D. alaskensis* ancestry as quantified by the genetic analyses (Fig. S1a), for each sample using the spectra as predictor variables (Bertrand *et al.*, 2013). The PLS beta regression restricts the response variable to a beta distribution which is necessary to predict the continuous ancestry variable inherently restricted between 0 and 1. We used *D. alaskensis* ancestry and spectral data from the three sites in which both *D. alaskensis* and *D. ajanensis* occurred. The model was assessed after 10-fold cross-validation repeated five times. We chose the optimal number of components (52) as the lowest number of components within two Akaike information criteria (AIC) of the number of components with the lowest AIC. Model RMSE and r^2 statistics were determined by comparing the mean predicted ancestry values from the five repeats to the ancestry values estimated by the genetic analyses.

We determined the phylogenetic signal, or the degree to which closely related individuals resemble each other more than any two randomly drawn individuals from the same phylogenetic tree (Blomberg & Garland, 2002), for each wavelength in the reflected spectrum. We calculated Blomberg's *K*, a measure of phylogenetic signal that compares the similarity of any two individuals to that expected by Brownian motion (Blomberg *et al.*, 2003; Meireles *et al.*, 2020), using the *phylosig* function in the R/PHYTOOLS package with an ultrametric tree, the mean spectra per individual, the SEs associated with those means, and 500 tip-swap simulations per trait (1 nm wavelength; Revell, 2012; Paradis & Schliep, 2019). The result was then compared to the maximum Blomberg's *K*-value estimated from data simulated on a tree with no phylogenetic covariance.

Morphological analyses

We analyzed leaf morphologies and trait correlations among parent species and hybrids to test our hypothesis that hybrids have variable and inconsistent morphologies compared to the parent taxa. We measured four taxonomically informative, though not necessarily spectrally informative, traits from 135 total leaves (five

leaves \times nine individuals \times three taxa total; Springer & Parfitt, 2014; Hultén, 1959; Elven *et al.*, 2011): leaf length (millimeters), abaxial midvein scales (presence or absence), abaxial midvein stipitate glandular trichomes (presence or absence) and degree of adaxial tomentum (glabrous–or nearly so, intermediate or dense). Pubescence and midvein morphology were observed with a digital microscope. We used a multivariate analysis of variance (MANOVA) to test for differences between the traits of parent species and the traits of the hybrids (R Core Team, 2021). We then constructed trait correlation matrices for parent species and hybrids using Pearson product–moment correlations. The correlation structure of the parents was compared to that of the hybrids using a Mantel test with 23 permutations (complete enumeration) in R/VEGAN (Oksanen *et al.*, 2019).

Results

Genetic analyses

We generated a median of 3.08 M reads per sample with a median of 99.8% of these passing quality filters. We then mapped a median of 22 345 loci per sample to the reference genome with an average depth of coverage of 38.7 reads, resulting in 52 325 filtered SNPs from 12 961 GBS loci mapped to 367 *D. drummondii* scaffolds. We subsampled SNPs at least 50 000 bp apart to get a final dataset of 3042 unlinked SNPs.

The STRUCTURE results corroborated this clear separation of *D. alaskensis* and *D. ajanensis* individuals. When two clusters were defined ($K = 2$), the putative morphologically intermediate individuals were inferred to be 31.5–48.1% admixed hybrids (Fig. 2a). When we dropped the hybrids and analyzed the two species independently across several K (Fig. S2), we found *D. alaskensis* was best modeled as two populations, *D. alaskensis*-WDB and *D. alaskensis*-ESTM (Eagle Summit plus Twelve Mile), which is sensible because ES and TM are connected by contiguous habitat (Fig. 1). *Dryas ajanensis* also showed lack of gene flow among sites, being best modeled as four populations: *D. ajanensis*-BG, *D. ajanensis*-ESTM, *D. ajanensis*-MD and *D. ajanensis*-WD (Wickersham Dome A plus Wickersham Dome B; Figs 2a, S2).

The PCA of SNPs also revealed a significant genetic separation of species and sites (Fig. 2b). The first principal component (PC) explained 11.1% of the total genetic variation and clearly separated *D. ajanensis* and *D. alaskensis* individuals, with hybrids falling in the middle, and the second PC separated individuals by site (2.8% of variance explained). In both species, individuals from ES and TM overlapped on both PC axes, supporting those as a single population. *D. ajanensis*-WDA and *D. ajanensis*-WDB individuals had proximal values on these PC axes but overlapped only on a single axis.

Phylogenetic inference corroborated a robust separation of *D. alaskensis* and *D. ajanensis* as well as the six total populations inferred from STRUCTURE (Fig. 2c). Within *D. alaskensis*, all WDB individuals, except a single individual, form a clade with 96% bootstrap support that is set within several ES and/or TM lineages (Fig. 2c). Within the *D. ajanensis* clade, BG, ESTM and

MD all form well-supported monophyletic groups. Individuals from WDB form a nested series of lineages subtending a nearly monophyletic clade of WDA individuals (Fig. 2c).

Overall, the results of our genetic analyses support the delineation of six reproductively isolated groups: *D. alaskensis* is split into *D. alaskensis*-ESTM and *D. alaskensis*-WDB, and *D. ajanensis* is split into *D. ajanensis*-BG, *D. ajanensis*-ESTM, *D. ajanensis*-MD and *D. ajanensis*-WD. The PCA and STRUCTURE show some genetic differentiation among the *D. ajanensis*-WDA vs WDB sites, but as a consequence of their overall similarity and proximity (3 km), as well as the fact that our main conclusions were robust to splitting these sites into two populations, we chose to treat them as a single population.

Spectral analyses

The mean reflectance for *D. ajanensis*, *D. alaskensis* and the hybrids appeared superficially similar (Fig. 3a), yet we were able to train PLS-DA models to classify spectral scans of the three taxa with $92.9 \pm 1.8\%$ accuracy (Fig. 3b; the ‘ \pm ’ indicates one standard deviation (SD)). *Dryas alaskensis* was predicted with $95.9 \pm 2.7\%$ accuracy and *D. ajanensis* with $92.4 \pm 2.6\%$ accuracy. Hybrids alone were predicted with $80.3 \pm 13.9\%$ accuracy (Fig. 3b). All regions of the reflected spectrum were useful for separating these taxa with the VIS and SWIR appearing to be relatively more informative than the NIR according to variable importance calculations – the contribution of coefficients associated with each wavelength weighted proportionally to the reduction in the sums of squares (Fig. S3; Kuhn, 2021). This may indicate that these species are best separated by their leaf pigments, such as chlorophyll, carotenoids and anthocyanins, as well as their lignin, cellulose and phenolic compound contents (Ustin *et al.*, 2009; Kokaly & Skidmore, 2015; Thenkabail & Lyon, 2016). When PLS-DA models were trained without hybrids, *D. alaskensis* and *D. ajanensis* were classified with an overall accuracy of $99.7 \pm 0.4\%$. In contrast to the minor spectral reflectance differences at the species level, the mean spectral reflectance values of populations noticeably varied (Fig. 3c). The PLS-DA models correctly classified scans to their genetically determined population with $97.7 \pm 1.9\%$ to 100% accuracy depending on the population ($98.9 \pm 0.7\%$ overall accuracy; Fig. 3d). See Table S3 for complete overall accuracy statistics and the number of components used for each set of classification models.

We were able to classify both the species identity and collection site, the finest classification resolution, with an overall accuracy of $92.0 \pm 1.2\%$ (Fig. 4). Site alone was predicted with an accuracy of $99.8 \pm 0.3\%$ (Fig. S4). Leaf scans from all *D. alaskensis* sites as well as *D. ajanensis* from BG and MD were correctly assigned with $> 94\%$ accuracy when hybrids were included in the model. Occasionally, one species was misclassified as the other species belonging to the same site with the main contributions to this phenomenon coming from $8.5 \pm 5.9\%$ of *D. ajanensis* from WDB being misclassified as *D. alaskensis* from WDB, which could be a result of the higher levels of admixture at this site (Fig. 2a). Most of the misclassifications were attributed to *D. alaskensis* or *D. ajanensis* being classified as a hybrid from the

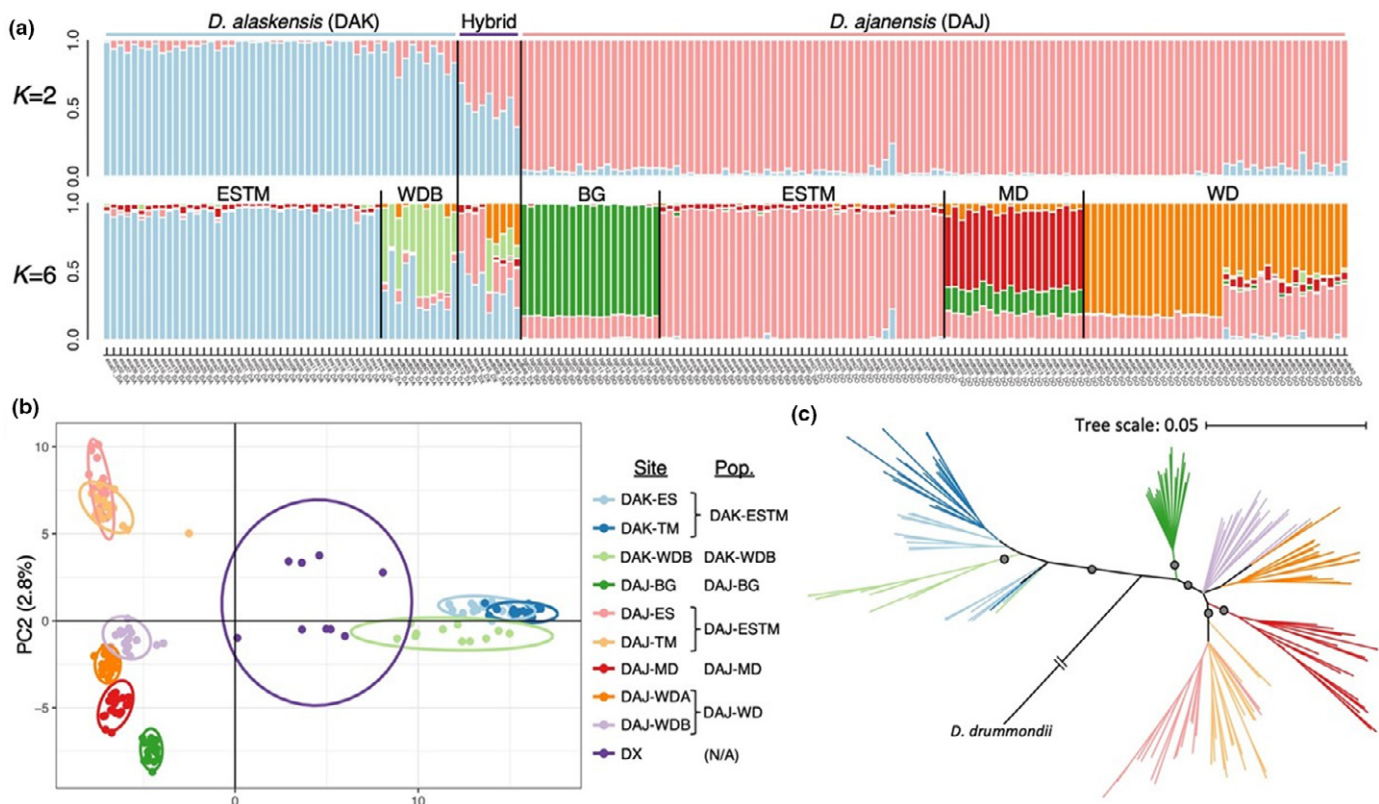


Fig. 2 *Dryas* species and population genetic structure. (a) STRUCTURE bar plots showing the proportion of each individuals' genome assigned to *Dryas alaskensis* (DAK) and *D. ajanensis* (DAJ) species ($K=2$, top), and proportional assignment to the optimal six groups ($K=6$, bottom). (b) Plot of first two principal components, explaining 13.9% of the total genetic variation among individuals with 95% quantile ellipses circumscribing each site for each taxon. (c) Maximum-likelihood cladogram without hybrids; branches with < 70% bootstrap support have been collapsed, backbone nodes with > 95% bootstrap support are indicated by circles. Colors for site and population assignments are provided in legend. Hybrids are abbreviated as DX. Site acronyms: BG, Bison Gulch; ES, Eagle Summit; ESTM, Eagle Summit and Twelve Mile; MD, Murphy Dome; TM, Twelve Mile; WDA, Wickersham Dome A; WDB, Wickersham Dome B; WD, Wickersham Dome A and B.

same site. For example, $6.2 \pm 4.0\%$ of *D. ajanensis* from ES, $9.0 \pm 3.1\%$ of *D. ajanensis* from TM and $9.9 \pm 6.3\%$ of *D. ajanensis* from WDB were misclassified as hybrids from the respective sites. When we modeled species and site without hybrids, individuals were classified with $98.9 \pm 0.8\%$ accuracy (Fig. S5). Most of the error resulted from *D. ajanensis* from ES and WDB being classified as *D. alaskensis* from the respective site ($2.5 \pm 2.9\%$ for ES, $5.0 \pm 6.7\%$ for WDB). When we removed water absorption features and the visible wavelengths from the spectra, our PLS-DA models classified location with $99.8 \pm 0.3\%$ accuracy (51 components; Fig. S6) which indicates that our drying procedure did not create an unintended drought artifact. Furthermore, our analysis of residual leaf water content across sites showed weak spatial autocorrelation (Methods S2; Fig. S7), so drying the leaves for 36–60 h in silica gel before scanning was sufficient to control for this environmental factor.

The PLS beta regression predicted the proportion of *D. alaskensis* ancestry of the samples with mean RMSE = 0.13 and $r^2 = 0.91$ (Fig. 5). Also, our test for phylogenetic signal indicated that all 1-nm wavelengths carried phylogenetic information with the VIS and SWIR being most similar between closely related individuals (Fig. S8). Blomberg's K was greater than the maximum Blomberg's K estimated from the null model and

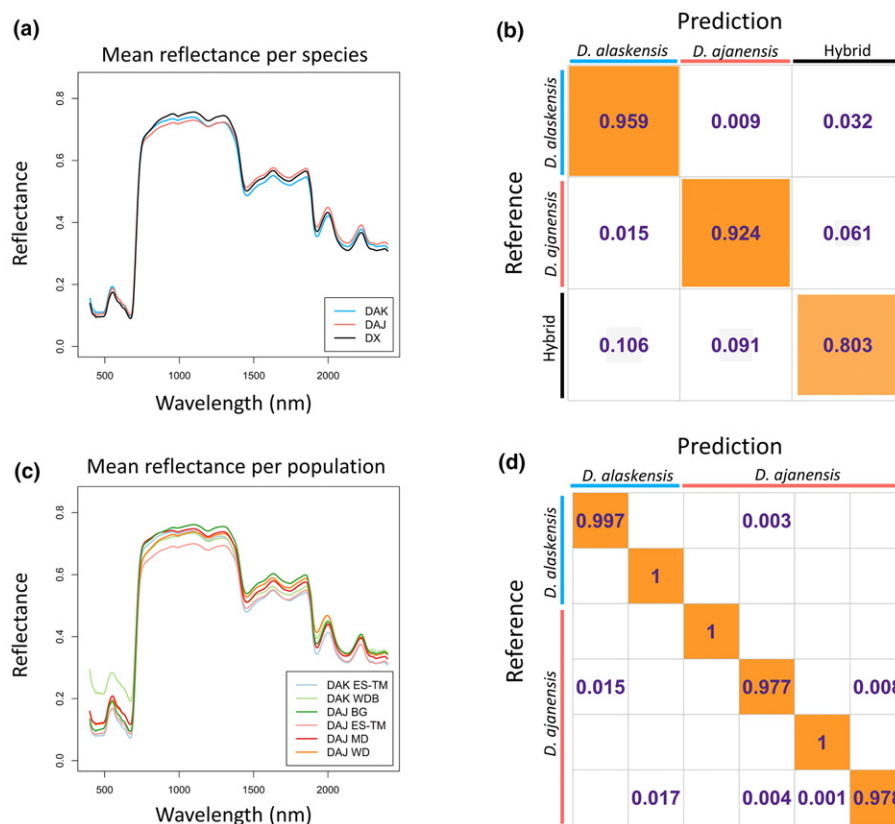
significant for all wavelengths ($P = 0.002$). These results corroborate the evidence from our discriminant analyses that leaf reflectance spectra carry useful genetic information and reflect phylogenetic relationships at these fine levels of biological organization.

Morphological analyses

We evaluated the hypothesis that hybrids exhibit inconsistent morphologies by quantifying a few taxonomically (although not necessarily spectrally) informative morphological traits and their correlations. *Dryas ajanensis* and *D. alaskensis* were clearly separated by abaxial midvein pubescence and by adaxial tomentum, in agreement with previous descriptions (Hultén, 1968), but hybrid individuals are not uniformly intermediate across all traits (Fig. S9). Although both midvein morphologies (glands and scales) were present on hybrid plants observed in the field, individual leaves were not consistent in these traits.

Our analysis of four taxonomically informative leaf traits (leaf length, tomentum, glands, scales) showed high trait correlation coefficients between leaves of parent *D. ajanensis* or *D. alaskensis* individuals, indicating consistent trait expression (absolute values from 0.46 to 0.87; Fig. 6a). Unsurprisingly, the traits of hybrids

Fig. 3 Species and population classification from leaf reflectance spectroscopy. (a) Mean reflectance values for each species. (b) Confusion matrix from the partial least squares discriminant analysis (PLS-DA) model discriminating species. (c) Mean reflectance values for each population. (d) Confusion matrix from the PLS-DA model discriminating populations. For the spectra plots: DAJ, *Dryas ajanensis*; DAK, *D. alaskensis*; and DX, hybrids. In the confusion matrices, the number in each cell represents the proportion of scans from the reference class (row) classified into the predicted class (column). Correctly classified scans fall into the diagonal and misclassifications are off-diagonal. White cells represent zeroes. The proportion of correct identifications are indicated by the size and shade of the orange squares. Site name acronyms: ESTM, Eagle Summit and Twelve Mile; WDB, Wickersham Dome B; BG, Bison Gulch; MD, Murphy Dome; WD, Wickersham Dome A and Wickersham Dome B.



were significantly different from the parent species (MANOVA: $F = 3.7$, $P = 0.007$), but the correlation coefficients between traits in hybrid individuals were much weaker (absolute values from 0.08 to 0.31, and 0.50 for scales and glands; Fig. 6b). Although we do not expect these few traits to drive spectral reflectance values, this supports our hypothesis that traits have segregated differentially among hybrid individuals and thus their overall morphology is not consistently intermediate between parents. The Mantel test confirmed that the correlation structure between traits within hybrids was not the same as the parent species ($r = 0.77$, $P = 0.083$).

Discussion

Biodiversity detection from reflectance spectra has mostly focused on taxa at or above the species level (Ustin & Gamon, 2010; Féret & Asner, 2012; Durgante *et al.*, 2013) despite the importance that fine-scale genetic and phenotypic diversity has for ecological and evolutionary processes (Crawford & Rudgers, 2013; Hoffmann *et al.*, 2017). Thus, our goal was to determine if fine-scale diversity characterized from genomic sequence data over a small geographic scale could be detected via leaf spectral reflectance. We demonstrated that reflectance spectroscopy captures genetic information that can be used to accurately classify leaves to species, hybrids and populations in a taxonomically challenged group of arctic dwarf shrubs.

Genomic sequencing proved an effective method for establishing species and population-level structure in *Dryas*. Max *et al.* (1999)

hypothesized that *Dryas* plants at alpine sites in interior Alaska would have distinct allozyme profiles owing to the effects of drift on isolated mountaintop populations. Initially we were unsure if we would find population genetic structure in this study region because Max *et al.* (1999) found no allozyme differences among sites and very few differences between species, and we expected that widely distributed species like *D. ajanensis* would be genetically cohesive at this spatial scale (Fig. 1). However, our dense sampling and deep sequencing strategy readily revealed the genetic differentiation of *Dryas* species and mountaintop population structure (Fig. 2). Together, our studies confirm there is negligible gene flow across unsuitable habitat at lower, forested elevations. Also, the genomic data showed *D. ajanensis* and *D. alaskensis* were clearly differentiated, and although it would be useful to sample other parts of the range, our results indicated these are valid (plant) species exhibiting morphological and phylogenetic distinction despite recurrent hybrid formation (Baum, 2009).

Several pieces of evidence indicate that our models can recover this fine-scale genetic structure from leaf reflectance spectra. First, we found significant phylogenetic signal in the spectra (Fig. S8), which adds to the body of evidence that spectra convey evolutionary relatedness (Madritch *et al.*, 2014; Cavender-Bares *et al.*, 2016; Schweiger *et al.*, 2018; Meireles *et al.*, 2020). We also demonstrate that the full reflected spectrum is useful for determining phylogenetic relationships below larger taxonomic units, such as families or orders, as is presented by Meireles *et al.* (2020). Second, our partial least squares (PLS) beta regression model accurately estimated the proportion of *D. alaskensis*

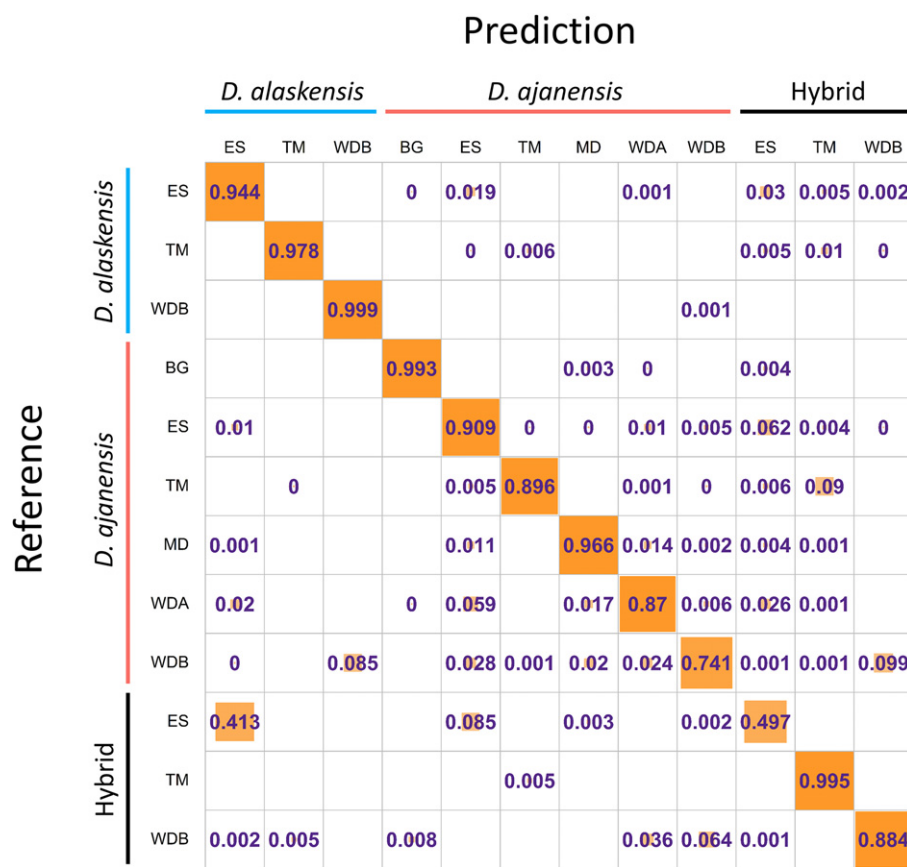


Fig. 4 Confusion matrix from the partial least squares discriminant analysis (PLS-DA) model discriminating samples by species and collection site. Row and column names are acronyms for the site of collection: BG, Bison Gulch; ES, Eagle Summit; MD, Murphy Dome; TM, Twelve Mile; WDA, Wickersham Dome A; WDB, Wickersham Dome B. The numbers in each cell correspond to the proportion of classifications per row averaged over 100 iterations. Larger squares and darker shades of orange indicate larger proportions of scans classified as the corresponding cell. White cells correspond to true zeros, and zeros indicate proportions < 0.001. Correctly identified species and collection site combinations are represented in the diagonal and misclassifications are found on the off-diagonal.

ancestry from the spectra, and the models explained 91% of the variation in the ancestry of the individuals. This is a novel and alternative approach to spectral classification of taxa that is a promising direction for dynamic analyses of hybrid zones and broader analyses of population structure and admixture. This analysis was also successful in revealing spectral similarities within species sampled from multiple sites and the intermediate spectral values of hybrid individuals (Fig. 5).

Third, the genomic resolution of species- and population-level structure was captured very well by the spectral reflectance data, which were used to classify leaves to their populations with 98.9% average accuracy (Fig. 3d). The PLS discriminant analysis (PLS-DA) models successfully classified leaves from populations Eagle Summit plus Twelve Mile (ESTM) and Wickersham Dome (WD) to their correct species with 97.8% (*D. ajanensis*-WD) to 99.7% (*D. alaskensis*-ESTM) accuracy despite these populations spanning two sites (Fig. 3d). Our classification of these six *Dryas* populations was more accurate than the discrimination of four populations of *Quercus oleoides* grown in a common garden (Cavender-Bares *et al.*, 2016). Conversely, our ability to classify these populations is comparable to the classification of *Populus tremuloides* genotypes from aerial imaging spectroscopy (Madritch *et al.*, 2014), which further demonstrates the ability of spectra to detect fine-scale genetic diversity *in situ*.

Lastly, we accurately classified species identities from reflectance spectra of *D. alaskensis* and *D. ajanensis* that co-

occurred (Fig. 3b), and we observed near-perfect classification accuracy (99.7%) when we trained the models without hybrids. We classified these two species of *Dryas* from leaf reflectance spectra more accurately than reported for classifying *Quercus* species (Cavender-Bares *et al.*, 2016); however, the difference in the number of species included in the models (two species of *Dryas* vs 28 species of *Quercus*) and the differences in sample preparation (dry *Dryas* leaves vs fresh *Quercus* leaves) may account for the differences in model accuracy. The accuracy of our models that classified both of the *Dryas* species and the hybrids were less accurate than previous studies that classified aspen (*Populus*; Deacon *et al.*, 2017) and bloodwood (*Corymbia*; Abasolo *et al.*, 2013) species and hybrids.

The most accurate classification model of our study (99.8% overall accuracy) discriminated samples by their collection sites alone (Fig. S4). An initial reaction to this could be that we were classifying the environment as opposed to genetics, but this model actually added the signal from genetics with any signal from site-based environmental factors. The confounding effects of environmental factors on phenotype was directly observed only in the few instances where taxa were misclassified as a different taxon belonging to the same site, which was rare and mostly involved hybrids (Fig. 4). Although it would be desirable to understand the environmental vs genetic components of the observed variability, the unanticipated covariance of genetic structure with mountaintop sites in this *Dryas* system is not

Fig. 5 Predicted *Dryas alaskensis* ancestry from spectral reflectance vs genomic ancestry. The points indicate the mean predicted ancestry for each individual plant, and the bars represent the full range of ancestry predicted from individual scans per plant. The diagonal line represents the 1 : 1 perfect relationship between predicted and actual values. Note: The *D. ajanensis* individual in the top left of the plot had only one scan after the spectral clean-up procedure.

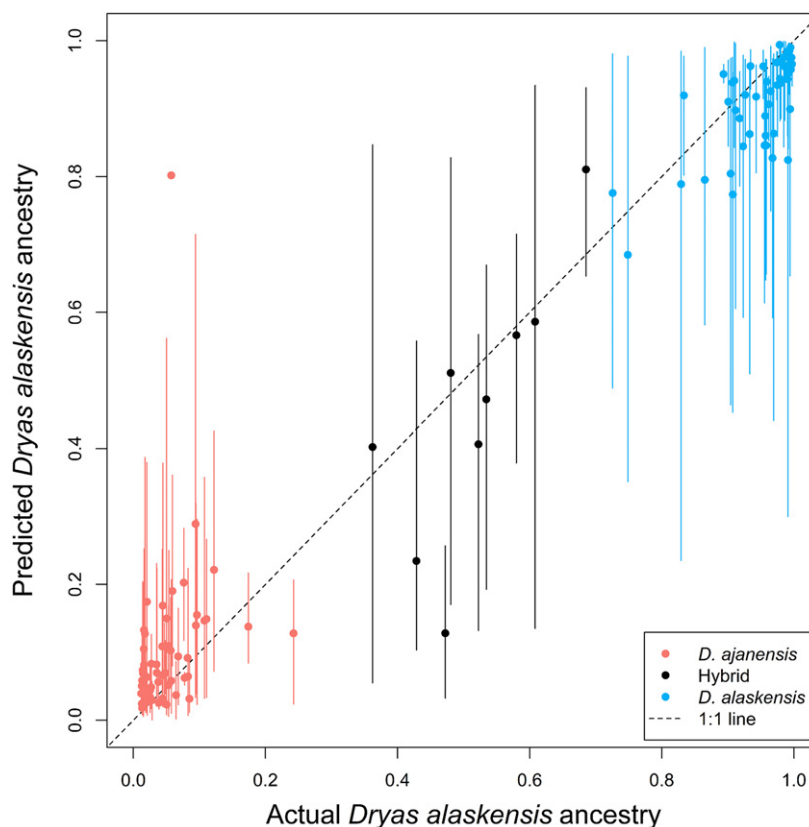
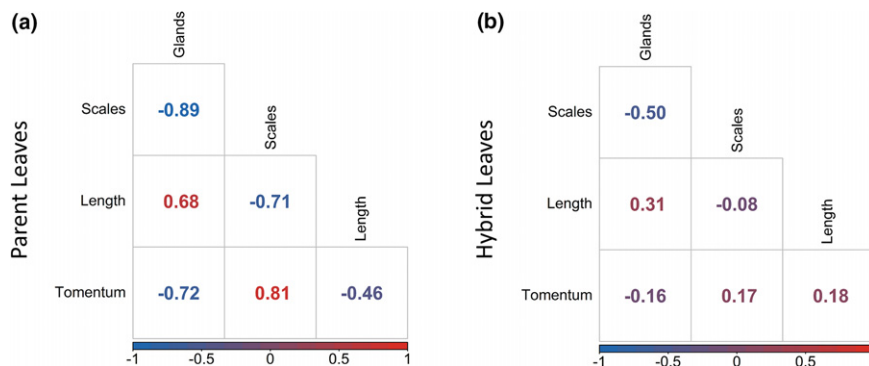


Fig. 6 Pairwise correlations between taxonomically informative leaf traits. (a) Trait correlations among leaves of the parent species, *Dryas ajanensis* and *D. alaskensis*. (b) Trait correlations among leaves of hybrids. More detailed descriptions of traits are found in the Materials and Methods section. Blue, negative correlations; red, positive correlations; shading corresponds to the magnitude of the correlations.



conductive to such an analysis. Nonetheless, we provide multiple lines of evidence that fine-scale genetic diversity is an identifiable signal in the spectra.

Our analysis of trait correlations was an effort to understand why hybrids failed to achieve classification accuracies as high as parent taxa (80% vs 92–96%; Fig. 3b). This analysis indeed showed that hybrid leaves overall had very weak correlations among several taxonomically informative traits compared to leaves of parent species, as would be expected from the differential inheritance and expression of alleles in hybrid individuals (Guo *et al.*, 2004; Cheng *et al.*, 2011). However, we did not classify hybrids poorly at every site, and the average classification accuracy of 80.3% is most likely pulled down by 41.3% of hybrids from ES being classified as *D. alaskensis* from ES (Fig. 4). This may be because the genetic admixture of the hybrid sample

ES03_DX was 31.5% (Fig. S1b) – the least balanced hybrid we sampled – and the leaf morphology of both ES hybrids leaned toward *D. alaskensis* (glandular trichomes, sparse tomentum, little to no midvein scales). In contrast to the ES hybrids, the TM and WDB hybrids were classified quite accurately (88.4–99.5% accuracy; Fig. 4). The two samples from TM had genetic admixture close to 50% (47.2% and 47.7%), and the five samples from WDB had admixtures ranging from 36.2% to 48.1%; WDB hybrids showed a range of admixture but benefitted from a larger sample size.

In a similar study, three hybrid aspen individuals with intermediate leaf morphology between parent species were classified using spectra with 94–99% accuracy (Deacon *et al.*, 2017). However, > 900 individuals representing several *Corymbia* hybrid taxa could be classified using 30 morphological traits with only 9.5–

76% accuracy, but these morphologically challenging and inconsistent hybrid taxa could be classified using spectra with 72–100% accuracy (Abasolo *et al.*, 2012, 2013). In summary, we should expect classification accuracy of genetically or morphologically inconsistent hybrids to improve when more individuals are scanned – a standard effect of sample size in statistical classification (Foody, 2009). Our study design differs from the others because we were not focused primarily on hybrid taxon classification, so our hybrid sample sizes were determined by their relative abundance along a transect.

We have shown that closely related, co-occurring plant species, their hybrids and their populations can be distinguished by the way light is reflected from their leaves. Our study extends the body of evidence on the utility of leaf-level spectral profiles by showing that they can successfully detect fine-scale genetic variation and, thus, can be applied to all levels of biological diversity above the level of the individual (Bickford *et al.*, 2007; Madritch *et al.*, 2014; Cavender-Bares *et al.*, 2016; Meireles *et al.*, 2020). Scaling biodiversity detection from leaf-level spectra to remotely sensed imagery comes with challenges that should not be understated, such as the presence of pixels that contain multiple individuals of different species (Rocchini, 2007; Wang & Gamon, 2019). However, the fact that fine-scale evolutionary diversity is captured by the spectrum of a leaf – a fundamental biological unit – suggests that assessing genetic variation using remote sensing should be possible if scaling methods improve and higher spatial resolution spectral images become more widespread. Likewise, the fact that hybrids can be accurately classified from spectra indicates that spectroscopy could be leveraged for studying plants along hybrid zones and accelerate the study of speciation, functional ecology and community interactions (Evans *et al.*, 2012; Abbott, 2017; Campbell *et al.*, 2018).






Acknowledgements

We thank K. Smith and K. Chan for their help with the collection of the data, M. Chaiken for assistance with DNA extraction, and B. J. McGill for statistical advice. We also would like to thank S. Kothari and L. Williams for their supportive review of the manuscript. Dawson M. White was supported by the Grainger Bioinformatics Center at the Field Museum. This work was also funded by the Tawani Foundation through the Field Museum, NASA's Arctic Boreal Vulnerability Experiment (ABOVE) (grant no. NNX15AU05A), USDA NIFA (Hatch project ME0-22022) and the National Science Foundation (DBI-2021898).

Author contributions

All authors contributed to the design of this study; LS, DMW and PRN collected data; LS performed the spectral and morphological analyses; DMW performed the genetic analyses; JEM, PRN and RHR helped guide analyses; LS, DMW and JEM wrote the manuscript, and all authors edited it. LS and DW contributed equally to this work.

ORCID

José Eduardo Meireles  <https://orcid.org/0000-0002-2267-6074>
 Peter R. Nelson  <https://orcid.org/0000-0001-5539-4914>
 Richard H. Ree  <https://orcid.org/0000-0003-1898-3335>
 Lance Stasinski  <https://orcid.org/0000-0002-3221-9416>
 Dawson M. White  <https://orcid.org/0000-0002-0670-9390>

Data availability

Raw sequence reads are available in the NCBI BioProject PRJNA670097. Leaf spectra are available at https://figshare.com/projects/Spectra_from_Dryas_alaskensis_Dryas_ajanensis_and_their_hybrids/96251. The R code for spectral analysis is available at <https://github.com/LanceStasinski/Dryas2>.

References

- Abasolo M, Lee DJ, Raymond C, Meder R, Shepherd M. 2013. Deviant near-infrared spectra identifies *Corymbia* hybrids. *Forest Ecology and Management* 304: 121–131.
- Abasolo M, Lee DJ, Shepherd M. 2012. Identification of intersectional *Corymbia* hybrids based on seedling morphology improves with parental divergence. *Forest Ecology and Management* 279: 189–202.
- Abbott RJ. 2017. Plant speciation across environmental gradients and the occurrence and nature of hybrid zones: hybrid zones and plant speciation. *Journal of Systematics and Evolution* 55: 238–258.
- Atwater DZ, Callaway RM. 2015. Testing the mechanisms of diversity-dependent overyielding in a grass species. *Ecology* 96: 3332–3342.
- Barker M, Rayens W. 2003. Partial least squares for discrimination. *Journal of Chemometrics* 17: 166–173.
- Baum DA. 2009. Species as ranked taxa. *Systematic Biology* 58: 74–86.
- Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters* 15: 365–377.
- Bertrand F, Meyer N, Beau-Faller M, El Bayed K, Namer I-J, Maumy-Bertrand M. 2013. Regression beta PLS. *Journal de la Societe Francaise de Statistique* 154: 143–159.
- Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* 22: 148–155.
- Blomberg SP, Garland T Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology* 15: 899–910.
- Blomberg SP, Garland T Jr, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57: 717–745.
- Campbell DR, Faidiga A, Trujillo G. 2018. Clines in traits compared over two decades in a plant hybrid zone. *Annals of Botany* 122: 315–324.
- Carrió E, Rosselló JA. 2014. Salt drying: a low-cost, simple and efficient method for storing plants in the field and preserving biological repositories for DNA diversity research. *Molecular Ecology Resources* 14: 344–351.
- Carter GA. 1991. Primary and secondary effects of water content on the spectral reflectance of leaves. *American Journal of Botany* 78: 916–924.
- Cavender-Bares J, Gamon JA, Hobbie SE, Madritch MD, Meireles JE, Schweiger AK, Townsend PA. 2017. Harnessing plant spectra to integrate the biodiversity sciences across biological and spatial scales. *American Journal of Botany* 104: 966–969.
- Cavender-Bares J, Meireles J, Couture J, Kaproth M, Kingdon C, Singh A, Serbin S, Center A, Zuniga E, Pilz G *et al.* 2016. Associations of leaf spectra with genetic and phylogenetic variation in oaks: prospects for remote detection of biodiversity. *Remote Sensing* 8: 221.

- Chen K, Zhang F, Kan J. 2012. Characterization of chlorophyll breakdown in green prickleyashes (*Zanthoxylum schinifolium* Zucc.) during slow drying. *European Food Research and Technology* 234: 1023–1031.
- Cheng D, Vrieling K, Klinkhamer PGL. 2011. The effect of hybridization on secondary metabolites and herbivore resistance: implications for the evolution of chemical diversity in plants. *Phytochemistry Reviews* 10: 107–117.
- Chevallier S, Bertrand D, Kohler A, Courcoux P. 2006. Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics* 20: 221–229.
- 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.
- Crawford KM, Rudgers JA. 2013. Genetic diversity within a dominant plant outweighs plant species diversity in structuring an arthropod community. *Ecology* 94: 1025–1035.
- Curran PJ. 1989. Remote sensing of foliar chemistry. *Remote Sensing of the Environment* 30: 271–278.
- Deacon NJ, Grossman JJ, Schweiger AK, Armour I, Cavender-Bares J. 2017. Genetic, morphological, and spectral characterization of relictual Niobrara River hybrid aspens (*Populus × smithii*). *American Journal of Botany* 104: 1878–1890.
- Dittrich-Reed DR, Fitzpatrick BM. 2013. Transgressive hybrids as hopeful monsters. *Evolutionary Biology* 40: 310–315.
- 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.
- Eaton DAR, Overcast I. 2020. ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36: 2592–2594.
- Elshire RJ, Glaubit JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379.
- Elven R, Murray DF, Razzhivin VY, Yurtsev BA. 2011. *Annotated checklist of the panarctic flora (PAF) vascular plants*. [WWW document] URL <http://panarcticflora.org/> [accessed 7 January 2021].
- Evans LM, Allan GJ, Whitham TG. 2012. Populus hybrid hosts drive divergence in the herbivorous mite, *Aceria parapopuli*: implications for conservation of plant hybrid zones as essential habitat. *Conservation Genetics* 13: 1601–1609.
- Fang M, Ju W, Zhan W, Cheng T, Qiu F, Wang J. 2017. A new spectral similarity water index for the estimation of leaf water content from hyperspectral data of leaves. *Remote Sensing of Environment* 196: 13–27.
- Féret J-B, Asner GP. 2012. Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and LiDAR. *Remote Sensing* 4: 2457–2476.
- Field DL, Ayre DJ, Whelan RJ, Young AG. 2009. Molecular and morphological evidence of natural interspecific hybridization between the uncommon *Eucalyptus aggregata* and the widespread *E. rubida* and *E. viminalis*. *Conservation Genetics* 10: 881–896.
- Foody GM. 2009. Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing* 30: 5273–5291.
- Gallego-Tévar B, Rubio-Casal AE, de Cires A, Figueroa E, Grewell BJ, Castillo JM. 2018. Phenotypic plasticity of polyploid plant species promotes transgressive behaviour in their hybrids. *AoB Plants* 10: ly055.
- Gates DM, Keegan HJ, Schleter JC, Weidner VR. 1965. Spectral properties of plants. *Applied Optics* 4: 11.
- Grundt HH, Kjølner S, Borgen L, Rieseberg LH, Brochmann C. 2006. High biological species diversity in the arctic flora. *Proceedings of the National Academy of Sciences, USA* 103: 972–975.
- Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS. 2004. Allelic variation of gene expression in maize hybrids. *Plant Cell* 16: 1707–1716.
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.
- Hoffmann AA, Sgrò CM, Kristensen TN. 2017. Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution* 32: 506–517.
- Hultén E. 1959. Studies in the genus *Dryas*. *Svensk Botanisk Tidskrift* 53: 507–542.
- Hultén E. 1968. *Flora of Alaska and neighboring territories: a manual of the vascular plants*. Stanford, CA, USA: Stanford University Press.
- Ibsey F, Craven D, Connolly J, Loreau M, Schmid B, Beierkuhnlein C, Bezemer TM, Bonin C, Bruehlheide H, de Luca E *et al.* 2015. Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature* 526: 574–577.
- Jetz W, Cavender-Bares J, Pavlick R, Schimel D, Davis FW, Asner GP, Guralnick R, Kattge J, Latimer AM, Moorcroft P *et al.* 2016. Monitoring plant functional diversity from space. *Nature Plants* 2: 16024.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
- Kokaly RF, Skidmore AK. 2015. Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 µm. *International Journal of Applied Earth Observation and Geoinformation* 43: 55–83.
- Kuhn M. 2021. caret: classification and regression training. R package. [WWW document] URL <https://cran.r-project.org/web/packages/caret/index.html> [accessed 2 July 2021].
- Lang C, Almeida DRA, Costa FRC. 2017. Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). *Forest Ecology and Management* 406: 219–227.
- Madritch MD, Kingdon CC, Singh A, Mock KE, Lindroth RL, Townsend PA. 2014. Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 369: 20130194.
- Max KN, Mouchaty SK, Schwaegerle KE. 1999. Allozyme and morphological variation in two subspecies of *Dryas octopetala* (Rosaceae) in Alaska. *American Journal of Botany* 86: 1637–1644.
- McGraw JB, Antonovics J. 1983. Experimental ecology of *Dryas Octopetala* ecotypes: I. Ecotypic differentiation and life-cycle stages of selection. *Journal of Ecology* 71: 879–897.
- Meireles JE, Cavender-Bares J, Townsend PA, Ustin S, Gamon JA, Schweiger AK, Schaepman ME, Asner GP, Martin RE, Singh A *et al.* 2020. Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytologist* 228: 485–493.
- Meireles J, Schweiger A, Cavender-Bares J. 2017. *Spectrolab: class and methods for hyperspectral data*. R package v.0.0.15 [WWW document] URL <https://cran.r-project.org/web/packages/spectrolab/> [accessed 7 September 2021].
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.
- Neuwirthová E, Lhotáková Z, Albrechtová J. 2017. The effect of leaf stacking on leaf reflectance and vegetation indices measured by contact probe during the Season. *Sensors* 17: 1202.
- Oksanen J, Blanchet G, Friendly M, Kindt R, Legendre P, McGinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P *et al.* 2019. *VEGAN: community ecology package*. R package v.2.5-7 [WWW document] URL <https://cran.r-project.org/web/packages/vegan/index.html>.
- Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.
- Páscoa RN, Moreira S, Lopes JA, Sousa C. 2018. Citrus species and hybrids depicted by near- and mid-infrared spectroscopy. *Journal of the Science of Food and Agriculture* 98: 3953–3961.
- Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344: 1246752.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- R Core Team. 2021. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raffard A, Santoul F, Cucherousset J, Blanchet S. 2019. The community and ecosystem consequences of intraspecific diversity: a meta-analysis. *Biological Reviews of the Cambridge Philosophical Society* 94: 648–661.
- Revell LJ. 2012. PHYTOOLS: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217–223.

- Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129: 149–165.
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–1216.
- Rocchini D. 2007. Effects of spatial and spectral resolution in estimating ecosystem α -diversity by satellite imagery. *Remote Sensing of Environment* 111: 423–434.
- Schweiger AK, Cavender-Bares J, Townsend PA, Hobbie SE, Madritch MD, Wang R, Tilman D, Gamon JA. 2018. Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nature Ecology & Evolution* 2: 976–982.
- Serbin SP, Wu J, Ely KS, Kruger EL, Townsend PA, Meng R, Wolfe BT, Chlus A, Wang Z, Rogers A. 2019. From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. *New Phytologist* 224: 1557–1568.
- Skrede I, Eidesen PB, Portela RP, Brochmann C. 2006. Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Molecular Ecology* 15: 1827–1840.
- Springer JC, Parfitt BD. 2014. Dryas. In: Flora of North America Editorial Committee, eds. *Flora of North America North of Mexico, vol. 9: Magnoliophyta: Picramniaceae to Rosaceae*. New York, NY, USA: Oxford University Press, 326–331.
- Thenkabail PS, Lyon JG. 2016. *Hyperspectral remote sensing of vegetation*. Boca Raton, FL, USA: CRC Press.
- Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, Heredia SM, Hahn MA, Caseys C, Bock DG *et al.* 2016. Hybridization and extinction. *Evolutionary Applications* 9: 892–908.
- Turner W. 2014. Conservation. Sensing biodiversity. *Science* 346: 301–302.
- Ustin SL, Gamon JA. 2010. Remote sensing of plant functional types. *New Phytologist* 186: 795–816.
- Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M, Asner GP, Gamon JA, Zarco-Tejada P. 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of the Environment* 113: S67–S77.
- Wang R, Gamon JA. 2019. Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment* 231: 111218.

Supporting Information

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

Fig. S1 Proportion of ancestry from *Dryas alaskensis*.

Fig. S2 STRUCTURE plots delimiting *D. ajanensis* (DAJ) and *D. alaskensis* (DAK).

Fig. S3 Variable importance for classifying species.

Fig. S4 Classification matrix for collection site.

Fig. S5 Classification matrix for species and site without hybrids.

Fig. S6 Classification of site with the visible spectrum and water absorption features removed from the reflectance spectra.

Fig. S7 Variable importance for classifying site of collection.

Fig. S8 Phylogenetic signal by 1-nm wavelength.

Fig. S9 Leaf morphology by species.

Methods S1 DNA sequencing and genetic analysis.

Methods S2 Determining the effect of differing drying times on our classification accuracy.

Table S1 Distances between sampling site (km).

Table S2 Number of individuals and scans (in parentheses) per taxon and site.

Table S3 Accuracy and number of components for each classification unit.

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