**New Phytologist Supporting Information**

Article title: Reading light: Leaf optical properties capture fine-scale diversity of closely related, hybridizing alpine shrubs
Authors: Lance Stasinski, Dawson M. White, Peter R. Nelson, Richard H. Ree, José Eduardo Meireles
Article acceptance date: 6 September 2021

The following Supporting Information is available for this article:

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

**Fig. S2** STRUCTURE plots delimiting *D. ajaneneis* (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 1nm wavelength.

**Fig. S9** Leaf morphology by species.

**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

**Methods S1** DNA sequencing and Genetic Analysis

**Methods S2** Determining the effect of differing drying times on our classification accuracy
Figure S1 Proportion of ancestry from *Dryas alaskensis*. a) Ancestry by species. b) Ancestry for hybrids by site. In the box plots, the bold lines indicate the median, the top and bottom edges of the boxes indicate the first and third quartiles, the lines at the ends of the whiskers indicate the minimum (quartile 1 - 1.5 * interquartile range (IQR)) and maximum (quartile 3 + 1.5* IQR), and the dots indicate outliers.
Figure S2 STRUCTURE plots delimiting *D. ajaneneis* (DAJ) and *D. alaskensis* (DAK). a) DAK proportional genome assignments to two ($K=2$) or three ($K=3$) ancestral populations. Plots of posterior probability (red lines) and Evanno’s Delta $K$ (blue lines) indicate the most information content at $K=3$ for DAK (b) and $K=5$ for DO (c). d) DAJ proportional genome assignments to four ($K=4$) and five ($K=5$) ancestral populations.
Figure S3 Variable importance for classifying species. Red corresponds to *D. ajanensis*, blue corresponds to *D. alaskensis*, and black corresponds to hybrids. Variable importance is the contribution of each coefficient (wavelength) weighted proportionally to its reduction in the sum of squares. The y-axis shows the relative contribution of each wavelength scaled from 0-100. The dark lines indicate the mean variable importance and the shaded regions indicate the 95% confidence intervals.
**Figure S4** Classification matrix for collection site. The diagonal represents correct classifications. The size and shade of the orange square in each cell corresponds to the proportion of classifications within each row. White cells represent true zeros and zeros indicate proportions less than 0.001.
**Figure S5** Classification matrix for species and site without hybrids. The diagonal represents correct classifications. The size and shade of the orange square in each cell corresponds to the proportion of classifications within each row. White cells represent true zeros and zeros indicate proportions less than 0.01.
**Figure S6.** Classification of site with the visible spectrum and water absorption features removed from the reflectance spectra. The diagonal represents correct classifications. The size and shade of the orange square in each cell corresponds to the proportion of classifications within each row. White cells represent true zeros and zeros indicate proportions less than 0.01.

|      | BG  | ES  | MD  | TM  | WDA | WDB |
|------|-----|-----|-----|-----|-----|-----|
| BG   | 1   |     |     |     |     |     |
| ES   |     | 1   |     |     |     |     |
| MD   |     |     | 1   |     |     |     |
| TM   | 0   |     |     | 1   |     |     |
| WDA  |     |     |     |     | 1   |     |
| WDB  | 0.01|     |     |     |     | 0.99|
Figure S7 Variable importance for classifying site of collection. Variable importance is the contribution of each coefficient (wavelength) weighted proportionally to its reduction in the sum of squares. The y-axis shows the relative contribution of each wavelength scaled from 0-100. The dashed vertical lines indicate the primary water absorption features at 1450nm and 1940 nm (Carter, 1991). The dark lines indicate the mean variable importance and the shaded regions indicate the 95% confidence intervals.
Figure S8. Phylogenetic signal by 1nm wavelength. a) Blomberg’s K per individual 1 nm wavelength. The dashed gray line indicates the maximum K estimated from a tree with no phylogenetic covariance. b) p-values for the K estimates per wavelength.
**Figure S9.** Leaf morphology by species. a) Leaf length in millimeters. b) Presence of abaxial midvein glandular trichomes, feathery scales, both, or neither. c) Adaxial tomentum ranging from sparse to dense. For the box plot (a), the bold lines indicate the median, the top and bottom edges of the boxes indicate the first and third quartiles, the ends of the whiskers indicate the minimum (quartile 1 - 1.5 * interquartile range (IQR)) and maximum (quartile 3 + 1.5* IQR), and the dots indicate outliers.
Table S1 Distances between sampling sites (km)

|       | BG  | ES  | MD  | TM  | WDA |
|-------|-----|-----|-----|-----|-----|
| ES 252| 252 | -   | -   | -   | -   |
| MD 130| 130 | 153 | -   | -   | -   |
| TM 228| 228 | 28  | 126 | -   | -   |
| WDA 160| 160 | 128 | 32  | 100 | -   |
| WDB 163| 163 | 127 | 35  | 100 | 3   |

Abbreviations: BG = Bison Gulch, ES = Eagle Summit, MD = Murphy Dome, TM = Twelve Mile, WDA = Wickersham Dome A, WDB = Wickersham Dome B.

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

| Site                | D. ajanensis | D. alaskensis | Hybrids | Total   |
|---------------------|--------------|---------------|---------|---------|
| Eagle Summit        | 21 (124)     | 19 (114)      | 2 (12)  | 42 (250)|
| Twelve Mile         | 20 (124)     | 22 (124)      | 2 (12)  | 44 (260)|
| Wickersham Dome B   | 16 (79)      | 11 (66)       | 5 (30)  | 32 (175)|
| Bison Gulch         | 20 (120)     | -             | -       | 20 (120)|
| Murphy Dome         | 20 (120)     | -             | -       | 20 (120)|
| Wickersham Dome A   | 20 (120)     | -             | -       | 20 (120)|
| **Total**           | **117 (687)**| **51 (304)**  | **9 (54)**| **178 (1045)**|
Table S3 Accuracy and number of components for each classification unit

| Classification Unit                  | Accuracy (%) | Number of Components |
|--------------------------------------|--------------|----------------------|
| Species with hybrids                 | 92.85 ±1.83  | 27                   |
| Species without hybrids              | 99.65 ±0.44  | 33                   |
| Populations                          | 98.91 ±0.68  | 45                   |
| Site                                 | 99.80 ±0.31  | 59                   |
| Species + site                       | 92.05±1.25   | 30                   |
| Species + site without hybrids       | 98.92 ±0.82  | 55                   |

*The ‘±’ symbol indicates one standard deviation.*
**Methods S1** DNA sequencing and Genetic Analysis.

For DNA isolation, we first homogenized 30–40 mg of silica-dried leaves with a single 3.2 mm chrome-coated steel bead for 1 minute at 25 Hz in a TissueLyser® II (QIAGEN, Hilden, Germany). We then extracted with the Invisorb® Spin Plant Mini Kit (STRATEC Molecular, Birkenfeld, Germany) with the following modifications: 500 µl Lysis Buffer P, 25 µl Proteinase S centrifuged at 11,000 g for two minutes, 250 µl Binding Buffer P, 20 µl RNase, and eluting in 55 µl prewarmed (50º C) nuclease-free water with 10 mM Tris-HCl (pH 8.0), incubating elution buffer for 3 minutes, centrifuging, and eluting again with the flow-through. Genomic DNA (350 ng – 1 ug per sample) was plated onto one of two 96-well plates and sent to the University of Wisconsin Biotechnology Center DNA Sequencing Facility for genotyping-by-sequencing library preparation and sequencing (GBS; ApeKI endonuclease and size selected for 200–500 bp fragments). Each plate of 95 samples was pooled and paired-end sequenced (150 bp) on the NovaSeq 6000 (Illumina Inc., San Diego, CA, USA).

To generate our genetic dataset, we demultiplexed (perfect match to the barcode sequence), trimmed and filtered (removed reads less than 35 bp long or with 10 or more bp having a q-score <20; strict adapter trimming), and mapped our reads to the D. drummondii genome (GCA_003254865.1) using the ipyrad analysis toolkit (Eaton & Overcast, 2020). We removed GBS loci with >10% polymorphic sites (SNPs), more than five indels, or if they were missing any samples, and, finally, removed all multiallelic SNPs. Since Dryas are long-lived, mat-forming shrubs that can be connected underground, we identified clones as individuals with over 99% pairwise similarity and the individual with more missing data was dropped (vcf_clone_detect.py; github.com/pimbongaerts). Lastly, we removed SNPs with a minor allele count less than two and then removed all SNPs violating Hardy-Weinberg equilibrium at the species level using VCFtools (Danecek et al., 2011).
Methods S2 Determining the effect of differing drying times on our classification accuracy.

We suspected that measuring the reflectance of our specimens after differing periods of drying (36–60 hours) may have increased the differences between plants collected at different sites (plants collected from the same site were measured at relatively the same time), thus increasing the accuracy of our models when classifying collection sites and populations. We estimated the influence of water from the reflectance at 1450 nm and 1940 nm, which are known water absorption features (Carter, 1991), the Normalized Difference Water Index (NDWI; Gao, 1996), and variable importance values from PLS-DA for sites. We compared samples collected from different sites using pairwise t-tests with these variables. *Dryas ajanensis* and *D. alaskensis* samples were compared separately.

The pairwise t-tests of the water absorption features and NDWI, compared between the two species, only agreed that there was a significant difference between the water content of leaves collected from Eagle Summit and Wickersham Dome B. However, 2.8% of Wickersham Dome B *D. ajanensis* scans were still misclassified as Eagle Summit *D. ajanensis* in our PLS-DA (fig. 4) which indicates that factors other than water content may have influenced our classification accuracy. Furthermore, when we only accounted for differences between *D. ajanensis* (excluding *D. alaskensis*), our pairwise t-tests of water absorption features and NDWI only agreed on differences between Murphy Dome and Wickersham Dome B and between Wickersham Dome A and Wickersham Dome B. Despite differing water content, 2.0% and 2.4% Wickersham Dome B *D. ajanensis* scans were misclassified as belonging to Murphy Dome and Wickersham Dome A, respectively (fig. 4). These observations led us to conclude that differing drying times did not strongly influence accuracy of our classification models.

Finally, the variable importance values output by the PLS-DA for classifying sites did not show strong peaks at or near the primary water absorption features (1450 nm and 1940 nm) that were substantially different from peaks in other regions of the spectra. There was a peak near 1940 nm for classifying Twelve Mile specimens (Fig. S5); however, we cannot conclude that this peak indicates an importance of water content or a morphological feature that is more apparent when water is removed.
References

**Carter GA. 1991.** Primary and secondary effects of water content on the spectral reflectance of leaves. *American Journal of Botany* **78**: 916–924.

**Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011.** The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

**Eaton DAR, Overcast I. 2020.** ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**: 2592–2594.

**Gao B-C. 1996.** NDWI A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Remote Sensing of the Environment* **58**: 257–266.