

Université de Montréal

**Identifier les arbres du Québec grâce à la spectroscopie foliaire :
différenciation fonctionnelle et phylogénétique des espèces**

par
Florence Blanchard

Département de sciences biologiques, Faculté des arts et sciences

mémoire présenté en vue de l'obtention du grade de Maître ès Science (M. Sc)
en sciences biologiques, option recherche en biologie

Avril 2022

© Florence Blanchard, 2022

Université de Montréal
Département de sciences biologiques, Faculté des arts et sciences

Ce mémoire intitulé

**Identifier les arbres du Québec grâce à la spectroscopie foliaire:
différenciation fonctionnelle et phylogénétique des espèces**

Présenté par
Florence Blanchard

A été évalué par un jury composé des personnes suivantes

Simon Joly
Président-rapporteur

Etienne Laliberté
Directeur

Anne Bruneau
Codirectrice

Étienne Léveillé-Bourret
Membre du jury

Résumé

La spectroscopie représente un puissant outil en conservation grâce à la possibilité d'effectuer le suivi de la diversité végétale à travers de larges étendues géographiques. La réflectance spectrale montre un potentiel certain pour l'identification des espèces d'arbres et même des taxons inférieurs, mais ceci a rarement été testé sur un grand nombre d'espèces. J'examine la qualité de la classification de 45 espèces d'arbres des forêts tempérées du Québec à partir de plus de 3500 spectres de réflectance foliaires (400-2400 nm). Nous évaluons cette classification sur la base de la variation spectrale des espèces, de même qu'à partir des distances fonctionnelles et phylogénétiques mesurées. Nos résultats indiquent un taux de classification très satisfaisant ($\kappa = 0.736, \pm 0.005$). Nous observons des erreurs de classification plus fréquentes entre les espèces évolutivement proches, alors qu'il semble que la distance fonctionnelle établisse un seuil voulant qu'au-delà d'une certaine distinction fonctionnelle globale, il soit peu probable que deux espèces soient confondues. Ces résultats viennent renforcer le lien entre la diversité spectrale et l'organisation taxonomique des espèces, ajoutant à la valeur de substitution de la première pour la diversité phylogénétique. Cela suggère par contre que de fortes convergences fonctionnelles peuvent faire obstacle à l'identification des espèces à partir de la réflectance spectrale. Cette étude est prometteuse pour la classification de spectres foliaires non préalablement identifiés, et améliore notre compréhension du lien entre les données spectrales et la différenciation des espèces, d'une grande importance pour assurer la validité des estimations de la biodiversité à partir de données de télédétection.

Mots-clés : Identification des arbres, spectroscopie hyperspectrale, phylogénétique, diversité fonctionnelle, feuille, GLM hurdle

Abstract

Imaging Spectroscopy is a powerful tool for conservation due to its ability to monitor plant diversity over broad geographic areas. Increasing evidence suggests that spectral reflectance can be used to identify trees at the species level, and even below. However, most studies focus on only a few species. Here, we use foliar reflectance (400-2400 nm) to discriminate among 45 temperate forest tree species from southern Quebec, using over 3500 leaf-level spectra. Furthermore, we connect those classification results to functional and phylogenetic distinctiveness, as well as to intraspecific variation. We find that spectral reflectance shows a very good discriminatory power even with an extensive set of species ($\kappa = 0.736, \pm 0.005$). We find that close phylogenetic species get mistaken for one another more frequently than distantly related species, while functional variation acts as a threshold, beyond which misclassifications are unlikely. These results reinforce the link between spectral diversity and taxonomic organization or phylogenetic diversity, but also reiterate the potential confounding effects of functional convergences on species identification from hyperspectral reflectance. We believe these findings hold promise for the classification of unknown spectra and further improve the link between ground truth and remotely sensed data for biodiversity assessments.

Keywords: Hyperspectral classification; Imaging spectroscopy; Leaf-level; Plant functional traits; Tree species; Phylogenetic diversity.

Table des matières

Résumé.....	iii
Abstract.....	iv
Liste des tableaux.....	vii
Liste des figures.....	viii
Liste des sigles et des abréviations.....	ix
Remerciements.....	xi
Chapitre 1 Mise en contexte.....	12
Problématique : Pressions sur la biodiversité forestière.....	12
Solution : la télédétection des plantes.....	13
Assurer le suivi de la biodiversité par la détection des espèces.....	16
Objectifs de l'étude.....	20
Méthodologie.....	21
Chapitre 2 Foliar spectra accurately distinguish temperate tree species and show strong phylogenetic signal.....	23
Introduction.....	24
Methods.....	27
Study sites and species sampling.....	27
Leaf collection and preparation.....	27
Spectral data acquisition and processing.....	28
Functional traits.....	29
Phylogenetic reconstruction.....	29
Statistical analyses.....	30
Results.....	33
Distinguishing species with partial least squares discriminant analysis (PLS-DA).....	34
Explaining classification mistakes: patterns of phylogenetic and functional differentiation.....	39
Discussion.....	45
Classification accuracy.....	45
Effect of species relatedness on species classification.....	46
Effect of functional differences on species classification.....	48
Conclusions.....	50
Chapitre 3 Discussion.....	52

Retour sur les objectifs et hypothèses.....	52
Perspectives.....	55
Références bibliographiques.....	58
ANNEXES.....	70

Liste des tableaux

Tableau 1 Summary statistics from species classification by PLS-DA.....	37
Tableau 2 Summary of the fitted hurdle regression model on misclassification data	43
Tableau Supplémentaire S1 Additionnal species taxonomic and sampling information	68
Tableau Supplémentaire S2 Model comparison for the hurdle analysis	76

Liste des figures

Figure 1 Représentation d'un spectre foliaire typique entre 400 et 2400 nm	14
Figure 2 Spectral data overview for major clades.....	34
Figure 3 Confusion matrix from species classification by PLS-DA.....	36
Figure 4 Phylogenetic signal across reflectance spectra.....	40
Figure 5 Coefficient of variation across the spectrum for selected species.....	41
Figure 6 Misclassification rate and probability according to phylogenetic and functional distances.....	44
Figure Supplémentaire S1 Map of sampling sites.....	71
Figure Supplémentaire S2 Phylogeny of the 45 species used for classification.....	72
Figure Supplémentaire S3 Wavelength loading from PLS-DA model.....	73
Figure Supplémentaire S4 Coefficient of variation across the spectrum for every species.....	74
Figure Supplémentaire S5 Principal component analysis of spectral reflectance.....	77
Figure Supplémentaire S6 Principal component analysis of species functional traits.....	78

Liste des sigles et des abréviations

AIC : Akaike information criterion ; Critère d'information d'Akaike

BM : Brownien motion ; Mouvement brownien

c. : *circa*

CABO : Canadian Airborne Biodiversity Observatory ; Observatoire Canadien Aérien de la Biodiversité

CV : Coefficient de variation

e.g : *exempli gratia*

e.i : *Id est*

et al. *Et allia*

EWT : Equivalent water thickness ; Épaisseur équivalente en eau

FD : Functional distance ; Distance fonctionnelle

FRQNT : Fond de recherche du Québec – Nature et technologies

GLM : Generalized linear model ; Modèle linéaire généralisé

IRBV : Institut de Recherche en Biologie Végétale

κ : Kappa de Cohen

LiDAR : Laser imaging detection and ranging

MFFP : Ministère des Forêts, de la Faune et des Parcs

MSE : Mean square error ; Carré moyen des erreurs

NDVI : Normalized vegetation index ; Indice de végétation normalisé

NIR : Near infrared ; Infrarouge proche

NSERC : National Science and Engineering Research Council of Canada ; Conseil de Recherche en Sciences naturelles et en Génie du Canada

PC : Principal component ; Composante principale

PLS-DA : Partial least-squares discriminant analysis ; Analyse discriminante par moindres carrés partiels

RE : red edge

SÉPAQ : Société des établissements de plein air du Québec

SD : Standard deviation ; Écart type

Sp. : Espèce

Spp : Espèces

SWIR : Shortwave infrared ; Infrarouge court

UAV : Unmanned aerial vehicle ; Véhicule aérien téléguisé

VIS : Visible

Remerciements

Tout d'abord, j'aimerais remercier mon directeur ainsi que ma codirectrice de recherche, Etienne Laliberté et Anne Bruneau, qui m'ont offert l'occasion de participer aux travaux d'un groupe de recherche stimulant et innovateur. La mission de CABO répond à des préoccupations majeures sur la sauvegarde de notre diversité dans un contexte dominé par l'incertitude, préoccupations qui ne peuvent laisser insensible aucun·e aspirant·e biologiste. Je vous remercie de votre confiance dans ma capacité à accomplir ce projet qui sortait des sentiers battus, et ce malgré quelques changements de direction en cours de route. Je vous remercie également de votre disponibilité et de votre écoute, surtout dans le contexte particulier des deux dernières années. Merci également aux membres de mon comité-conseil, Simon Joly et Pierre-Luc Chagnon, pour leur accompagnement et leurs commentaires.

Pour ses conseils avertis et sa générosité, je remercie Rosalie Beauchamp-Rioux, dont la rigueur et l'excellence du travail ont été pour moi un modèle tout au long de mon projet. Merci aussi à Anna K. Schweiger, Johanna Jantzen et José E. Meireles pour leur expertise respective en spectroscopie et en analyse phylogénétique, qui ont pu me guider à travers certains défis méthodologiques. Pour leur aide précieuse dans la collecte des données, je remercie Clément Robert-Bigras, Isabelle Gareau, Myriam Cloutier, Fabien Cichonski, Alexandra Massé, Jocelyne Ayotte, Vincent Fournier et Rime Néron. Je tiens à souligner la contribution inestimable d'Antoine Mathieu, Maria Juliana Pardo Losada, ainsi que de Madeleine Trickey-Massé, dont les efforts ont permis la réalisation de cette recherche. Votre support et votre compagnie m'ont été précieux. Un merci particulier également à Sabrina Demers-Thibeault pour son souci des détails et son efficacité qui ont rendu possible le bon déroulement du projet.

Un énorme merci à mes parents pour leur soutien indéfectible. Merci pour vos encouragements et votre confiance dans cette entreprise comme dans toutes les autres. Finalement, merci Fred. J'ose à peine imaginer terminer ce projet sans ta présence, ton écoute et ton réconfort, et je t'en suis extrêmement reconnaissante.

Chapitre 1

Mise en contexte

Problématique : Pressions sur la biodiversité forestière

Les pressions grandissantes exercées par les changements climatiques et les perturbations anthropiques menacent l'intégrité et la diversité des communautés végétales (Larsen et al., 2005). Par le biais de transformations directes, comme la déforestation et la fragmentation des habitats, ou indirectes, tels que l'incidence et l'intensité accrue des feux de forêt et des épidémies d'insectes, les changements globaux fragilisent la qualité des milieux naturels et mettent en jeu la survie des espèces (Millar et Stephenson, 2015 ; Scheffers et al., 2016). Dans l'est de l'Amérique du Nord, la variabilité climatique, en particulier les périodes de sécheresse de plus en plus fréquentes et de longue durée, provoque des déclinés marqués de l'abondance de certaines espèces d'arbres ainsi que des modifications dans leur distribution géographique (Aitken et al., 2008 ; Charney et al., 2016). Sur la base des changements du climat uniquement (température, précipitations), une majorité d'espèces végétales devraient voir leur aire de répartition se déplacer vers le nord et se restreindre en matière d'étendue globale d'ici la fin du siècle (McKenney et al., 2007).

Au cœur des paysages forestiers, les arbres fournissent des services écosystémiques inestimables, incluant la séquestration du carbone atmosphérique, la régulation du climat et de l'hydrologie, leur qualité d'habitat essentiel pour de nombreuses espèces animales, sans oublier leur importance économique et culturelle (Alkama and Cescatti, 2016). Le maintien de la plupart de ces services dépend souvent de l'espèce dominante, qui définit la structure des communautés en modulant les caractéristiques environnementales, par exemple la chimie du sol par le biais du taux d'acidité de sa litière (Ellison et al., 2005; Angst et al., 2019). Ainsi, des transitions dans la richesse et la composition des assemblages forestiers menacent l'intégrité d'écosystèmes entiers, incluant la variété d'organismes associée à ces peuplements (Hooper et al., 2012; Lindenmeyer and Laurance, 2017). Pour cette raison, la dégradation sévère à laquelle font face les écosystèmes forestiers représente

une préoccupation de taille pour le suivi de la biodiversité à l'échelle régionale et mondiale (Scheffers et al., 2016. Watson et al., 2018).

Solution : la télédétection des plantes

Un des principaux défis de l'évaluation des changements en milieux forestiers consiste en l'ampleur géographique de ces écosystèmes et des perturbations qui les affectent, alors que les méthodes traditionnelles de suivi de la biodiversité demeurent presque inchangées depuis près d'un siècle (Wang et Gamon, 2019). En effet, plus de la moitié du territoire québécois est recouvert de forêts, dont 92 % se situent sur des terres publiques (MFFP, 2022). Heureusement, de nouvelles approches basées sur les données de télédétection ouvrent des perspectives inédites pour le suivi et la conservation de la biodiversité végétale (Asner et al., 2017 ; Cavender-Bares et al., 2017). La télédétection permet notamment un suivi rapide et reproductible sur de larges territoires, incluant des sites inaccessibles. De plus, il s'agit d'un mode d'inventaire non destructif et beaucoup moins exigeant sur le plan du temps, de la main-d'œuvre et des coûts, particulièrement dans des écosystèmes forestiers, où les parcelles atteignent généralement de grandes tailles (Dalponte et al., 2008 ; Jetz et al., 2016).

Toutefois, l'avantage central de la télédétection repose sur la capacité de lier les produits de celle-ci aux caractéristiques biophysiques de la diversité végétale. Par le biais d'indices mesurés à partir des propriétés spectrales, tel que l'indice de différence de végétation normalisé (NDVI) qui permet d'estimer la productivité primaire et l'activité photosynthétique, il est possible d'évaluer des attributs reflétant la composition, la structure et le fonctionnement des communautés végétales (Kerr et Ostrovsky, 2003 ; Roberts et al., 2004). Plus encore, la spectranomique élève la spectroscopie au niveau supérieur en établissant un rapport direct entre les propriétés optiques des plantes et leurs traits fonctionnels, par exemple la quantité d'eau, de carbone, ou de pigments présents dans les feuilles (Asner et Martin, 2009 ; Asner et Martin, 2016a). Ces traits, liés à la phénologie et la physiologie des plantes, constituent des indicateurs essentiels pour améliorer notre compréhension de la biodiversité et des changements qui l'affectent (Pereira et al., 2013). Leur adéquation avec les données de spectroscopie revêt donc une importance capitale pour le suivi à long terme des changements globaux (Shiklomanov et al., 2019).

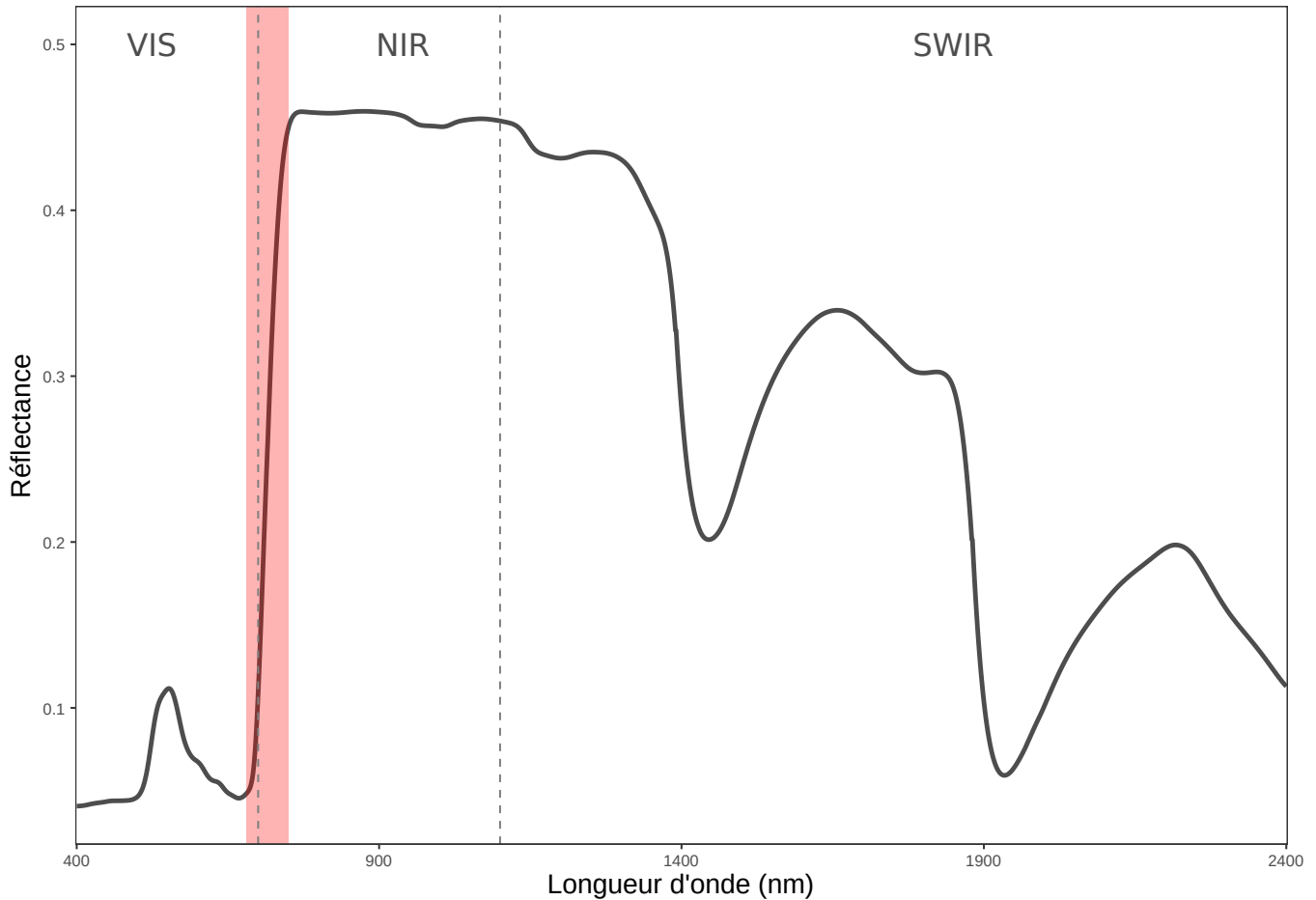


Figure 1 Représentation d'un spectre foliaire typique entre 400 et 2400 nm

Les régions du visible (VIS), de l'infrarouge proche (NIR) et de l'infrarouge court (SWIR) du spectre sont délimitées par les lignes pointillées, et l'étendue du *red-edge* est illustrée en rouge

La télédétection des plantes, plus particulièrement la spectranomique, se fonde sur la relation biophysique de la lumière avec les plantes, principalement avec les feuilles. Du point de vue écologique, ce lien est informatif puisque ces dernières contrôlent l'acquisition et l'allocation des ressources à la plante par le biais de l'activité photosynthétique (Wyka et al., 2012 ; Williams et al., 2020). La coordination des traits foliaires chez une espèce reflète donc une stratégie écologique particulière et informe sur les processus environnementaux, tels que la productivité (Reich et al., 2003 ; Wright et al., 2004). Du point de vue pratique, les propriétés optiques reviennent à mesurer l'intensité du rayonnement électromagnétique à la suite de son interaction avec des tissus végétaux grâce à un spectroradiomètre (Cavender-Bares et al., 2017). Typiquement, ces appareils permettent d'effectuer des mesures sur les longueurs d'ondes allant du visible à

l'infrarouge, soit de 400 à 2500 nm, ce qui inclut la majorité de la radiation solaire (Ollinger et al., 2011). L'interaction entre la lumière et la végétation se solde soit par l'absorption de la radiation par les différents constituants de la plante, soit par sa diffusion, qui se décompose en réflexion et en transmission selon la surface foliaire impliquée (Jacquemoud et Ustin, 2019). La variation dans les composés biochimiques ainsi que la structure des feuilles affectent donc la proportion de réflectance foliaire, de loin la plus étudiée à ce jour (Curran et al., 1992 ; Asner, 1998, Homolová et al., 2013).

Lumière visible (VIS)

Les spectres foliaires se divisent en trois régions principales (Figure 1). La première comprend le rayonnement associé à la lumière visible (VIS), de 400 à 700 nm. L'absorption par les pigments photosynthétiques et accessoires, soit par la chlorophylle *a* (Chl *a*), la chlorophylle *b* (Chl *b*), les caroténoïdes ainsi que par les anthocyanes, influence les propriétés spectrales de cette région (Knipling, 1970 ; Blackburn, 2007 ; Ustin et al., 2009 ; Ollinger et al., 2011). Grâce à son lien avec les pigments, la portion visible du spectre nous informe sur la capacité photosynthétique des plantes (Gamon et al., 1997, Yan et al., 2021), de même que sur l'état physiologique des individus, puisque le stress hydrique ou les dommages causés par les épidémies d'insectes peuvent être détectés dans le spectre visible (Carter et Knapp, 1991 ; Huang et al., 2019). Par exemple, la télédétection permettrait de repérer les zones affectées par la tordeuse des bourgeons de l'épinette à un stade précoce grâce de l'augmentation de la réflectance dans la région du visible occasionnée par la dégradation de la chlorophylle suite à l'interruption du flux de sève (Ahern, 1988 ; Abdullah, 2018).

L'infrarouge proche (NIR-RE)

La seconde région du spectre, l'infrarouge proche (NIR), correspond aux longueurs d'onde de 700 nm à 1000 nm. Cette région est influencée par la structure des feuilles, plus précisément la disposition des cellules ainsi que le ratio d'espaces intercellulaires par unité de surface (Slaton et al., 2001 ; Castro-Esau et al., 2006). Une plus grande homogénéité anatomique réduit la dispersion ainsi que l'absorption de la radiation à travers la feuille, ce qui se traduit par de fortes valeurs de transmittance et une plus faible réflectance à ces

longueurs d'onde. La structure homogène des aiguilles de conifères explique leur réflectance plus faible ainsi que le lien entre l'épaisseur des aiguilles et leur réflectance dans cette région (de Tomás Marín et al., 2016). À plus large échelle, des éléments architecturaux tels le regroupement des feuilles sur la branche ainsi que l'angle des feuilles influencent la réflectance de la canopée (Sullivan et al., 2013). Entre le visible et l'infrarouge proche se trouve le *red-edge*, zone située entre 680 nm et 750 nm. La pente et la position de cette forte inflexion du spectre sont influencées par la quantité de chlorophylle, et ces caractéristiques permettent de détecter différents stress environnementaux qui modifient la quantité de pigments photosynthétiques, par exemple l'herbivorie (Horler et al., 1983 ; Noomen et al., 2006).

L'infrarouge court (SWIR)

Finale­ment, la troisième région, l'infrarouge court (SWIR) comprend les longueurs d'onde de 1000 nm à 3000 nm, qui contiennent l'information d'une majorité de constituants biochimiques, notamment la lignine, la cellulose et l'azote, lié aux protéines (Kokaly et al., 2009). C'est aussi la portion où se manifestent en majorité les pics d'absorption de l'eau, dans les bandes caractéristiques à 1450 nm et 1930 nm (Ollinger et al., 2011 ; Jacquemoud et Ustin, 2019).

Assurer le suivi de la biodiversité par la détection des espèces

Les propriétés spectrales des plantes suscitent donc un intérêt en écologie depuis de nombreuses années en raison de leur accord avec les traits fonctionnels, puisque ces derniers permettent d'améliorer notre compréhension de l'interaction entre les espèces végétales et leur environnement, ainsi que des mécanismes régissant la diversité et la distribution de celles-ci (Billings & Morris, 1951 ; Gates et al., 1965 ; Zakharova et al., 2019). Grâce à la caractérisation rigoureuse de la variation fonctionnelle et de son effet sur le spectre foliaire (Asner, 1998 ; Kokaly et al., 2009 ; Serbin et al., 2014 ; Hovi et al., 2017), la réflectance foliaire permet de détecter des changements de la végétation en réponse à différents stress et maladies (Zagajewski et al., 2017 ; Huang et al., 2019 ; Fallon et al., 2020), de prédire les types fonctionnels de la végétation, de détecter des gradients environnementaux (Asner and Martin, 2016b; Schweiger et al., 2017; Sousa et al., 2021)

ou encore d'évaluer la diversité taxonomique et fonctionnelle d'une région (Laliberté et al., 2020 ; Frye et al., 2021).

Toutefois, ce lien entre les données spectrales et la diversité biologique et fonctionnelle d'une région n'est informatif que si la variabilité spectrale représente réellement la biodiversité végétale d'une région, ou permet au moins de classer les espèces en types fonctionnels (Schweiger et al., 2017 ; Wang et al., 2018 ; Rossi et al., 2021). Ce constat a mené au développement de la théorie des types optiques, aussi nommée « espèce spectrale », qui se base sur l'émergence dans le spectre d'une série de traits foliaires en tant que syndrome permettant d'identifier les patrons de diversité fonctionnelle (Ustin et Gamon, 2010, Féret et Asner, 2014). Alors que des patrons de diversité, incluant la richesse spécifique et l'unicité, peuvent être dérivés de la variation spectrale (Laliberté et al., 2020), le potentiel des données hyperspectrales pour la reconnaissance précise des espèces repose d'abord sur la différenciation de celles-ci au niveau de la feuille (Clark et al., 2005 ; Durgante et al., 2013).

Or, la variabilité intraspécifique de la réflectance foliaire représente un enjeu majeur qui limite notre capacité à identifier les espèces (Castro-Esau et al., 2006 ; Hycza et al., 2018 ; Rossi et al., 2021). Cette variabilité intraspécifique peut être attribuée à différents facteurs tels que les changements phénologiques saisonniers (Yang et al., 2016), l'âge des feuilles (Hovi et al., 2017 ; Wu et al., 2017), de même que par les conditions environnementales, particulièrement du sol (Asner et al., 2016), qui influencent les traits des végétaux. Dans certains cas, la variation spectrale intraspécifique des feuilles est bénéfique, puisqu'elle est essentielle à la détection de la végétation atteinte par la sécheresse ou des pathogènes par rapport à des individus sains (Zagajewski et al., 2017 ; Ramoelo et al., 2015). Dans d'autres cas, elle peut entraver l'identification des espèces (Wang et al., 2018). Le chevauchement interspécifique de la variation spectrale dû à la variation intraspécifique réduit notre capacité à distinguer autant les types fonctionnels (Schweiger et al., 2017) que les espèces elles-mêmes (Leckie et al., 2005).

Un aspect encourageant pour l'identification des plantes repose sur le fait que, malgré l'influence de facteurs environnementaux locaux, la majorité de la variation spectrale peut être attribuée à l'identité taxonomique des espèces, indiquant une plus grande variance entre les espèces (variation interspécifique) qu'au sein de celles-ci (Asner et al., 2014, Hovi et al., 2017). Dans certains cas, la spécificité des signatures spectrales se montre même assez élevée pour permettre la reconnaissance, bien que faible, des populations (Cavender-Bares et al, 2016) et même de génotypes particuliers (Cui et al., 2012 ; Madritch et al, 2014 ; Stasinski et al., 2021). De plus, le lien entre la phylogénie et certains axes principaux de la variation spectrale, de même qu'avec certaines longueurs d'onde et certains traits fonctionnels sous-jacents renforce notre capacité à discerner l'organisation taxonomique des espèces à travers les patrons de variabilité spectrale (Cavender-Bares et al., 2016 ; McManus et al., 2016 ; Meireles et al., 2020). Ce lien, nommé signal phylogénétique, se calcule à l'aide de différentes statistiques, mais toutes réfèrent à l'estimation de la tendance d'espèces proches du point de vue évolutif à se ressembler davantage que des espèces partageant une histoire évolutive plus ancienne (Münkemüller et al., 2012). Cette ressemblance se traduit par une relation positive entre la diversité spectrale des communautés et leur diversité fonctionnelle et phylogénétique (Schweiger et al., 2018 ; Frye et al., 2021). La présence de signal justifie donc l'utilité du spectre comme valeur de procuration à la diversité phylogénétique, cette dernière étant un critère important pour la conservation de la diversité biologique (Verón et al., 2019). Ainsi, l'identification des arbres au niveau de l'espèce par la spectroscopie semble un objectif réaliste, et même souhaitable pour assurer le suivi de la diversité des assemblages forestiers.

On constate par ailleurs des résultats prometteurs à cet effet. Bien qu'il soit difficile de comparer directement les résultats en raison des différentes techniques de classification utilisées ainsi que de l'identité et du nombre différent d'espèces étudiées, Fassnacht et al. (2016) rapporte un taux de succès de classification variant entre 75 et 90 % dans l'identification d'espèces végétales à l'aide de données hyperspectrales. La distinction de groupes taxonomiques plus larges, comme les familles, se révèle également possible (Meireles et al, 2020). Par contre, les études qui se concentrent sur la classification des

arbres se focalisent souvent sur un petit nombre d'espèces, souvent les espèces dominantes. Or, les espèces plus rares contribuent beaucoup à la valeur écologique d'une région, mais sont souvent exclues parce que leur faible abondance diminue leur détectabilité (Cerrejón et al. 2021). De plus, un paradigme fréquent en écologie montre que l'addition d'information peut mener à la saturation des relations entre deux variables, dans ce cas-ci entre la diversité spectrale et la diversité soit fonctionnelle ou phylogénétique, plus le nombre d'espèces considéré est grand (Schweiger, 2020). En effet, il semble exister un seuil taxonomique au-delà duquel des espèces phylogénétiquement plus éloignées ne sont pas nécessairement plus différentes au niveau de leurs traits fonctionnels en raison de convergences écologiques ou simplement du hasard (Kelly et al., 2014 ; Paquette et al., 2015). Cadotte et al. (2017) recensent d'ailleurs plusieurs raisons pour lesquelles il peut y avoir une divergence entre les différences phylogénétiques et écologiques des espèces. Parmi celles-ci, on trouve les stress environnementaux, la variation intraspécifique, ainsi que la résolution et la portée des phylogénies considérées. À ce propos, alors que Kim et al. (2004) trouvent une forte corrélation entre la diversité spectrale et taxonomique, bien que Durgante et al. (2013) mettent en garde que cette relation est influencée par la résolution phylogénétique du groupe observé, et que celle-ci n'a jamais été testée à grande échelle. Malgré une identification satisfaisante des grands clades et familles, Meireles et al. (2020) relèvent d'ailleurs ne pas avoir exploré la raison pour laquelle certaines lignées étaient plus difficilement distinguables.

Ainsi, les facteurs pouvant influencer la qualité des inférences à partir des données spectrales sont nombreux, surtout à de plus grandes échelles. Alors que la moitié des études portant sur la classification des arbres se concentrent sur l'optimisation des algorithmes de classification, les différences écologiques qui sous-tendent la capacité de distinguer les espèces à partir de la réflectance spectrale demeurent peu connues (Fassnacht et al., 2016). C'est pourquoi, dans le but d'optimiser l'identification des espèces à partir de la réflectance spectrale et assurer un suivi précis de celles-ci, il est nécessaire de tester la faisabilité de classifications supervisées sur une plus grande diversité d'espèces. Plus spécifiquement, il faut évaluer la qualité de leur identification en tenant compte de leur distinction

fonctionnelle et phylogénétique, puisque ces différences informent par la suite nos estimations de la biodiversité.

Objectifs de l'étude

L'objectif principal de mon étude consiste donc à évaluer la qualité de la classification de l'ensemble des espèces d'arbres d'une région à partir de leur réflectance spectrale foliaire. Plus spécifiquement, je vais évaluer le potentiel de celle-ci dans l'identification de 45 espèces d'arbres présentes dans le sud du Québec échantillonnées dans diverses conditions environnementales, ce qui constitue une analyse d'une rare ampleur. Afin de valider le lien entre la réflectance foliaire et l'organisation taxonomique de mon ensemble d'espèces, je testerai la présence et mesurerai la force du signal phylogénétique pour chaque longueur d'onde. De plus, j'examinerai le rôle de la différenciation fonctionnelle, de la distance phylogénétique, de même que de la variabilité spectrale des espèces considérées sur la qualité de la classification. À cet effet, je construirai une phylogénie incluant toutes ces espèces, et une série de traits fonctionnels foliaires seront mesurés indépendamment des signatures spectrales.

Puisque la plupart des études montrent que la variation interspécifique de la réflectance est plus élevée que la variation intraspécifique (Asner et al., 2014 ; Hovi et al., 2017), et que d'autres encore parviennent à des taux de succès très élevés dans leur classification (Durgante et al., 2013, Girard et al., 2020 ; Paiva et al., 2021), je m'attends à ce que la réflectance permette d'identifier correctement toutes les espèces considérées dans mon étude. Toutefois, même si je m'attends à une identification adéquate, celle-ci est rarement parfaite. Par conséquent, j'avance l'hypothèse que les espèces proches au point de vue phylogénétique auront davantage de chance d'être confondues, i.e. seront plus difficiles à distinguer avec précision. Cette hypothèse s'appuie sur le lien trouvé entre la diversité spectrale, fonctionnelle et phylogénétique (Schweiger et al., 2017, Frye et al., 2021) ; selon celui-ci, des espèces proches devraient être plus similaires, tant au niveau de leurs traits et de leurs spectres, et donc devraient être plus difficiles à reconnaître.

Suivant cette hypothèse, je m'attends également à trouver un certain signal phylogénétique dans les spectres, mais que celui-ci sera inégalement distribué à travers les

clades. Cette prédiction se base sur les résultats de Meireles et al., (2020) qui démontrent notamment une disparité dans la répartition et la force du signal entre les conifères et les clades de plantes à fleurs, puisque les premiers présentent un signal très fort sur l'entièreté du l'infrarouge court, alors que les seconds montrent un signal plus faible, mais constant dans le visible et l'infrarouge proche.

Méthodologie

Afin de réaliser mes objectifs, j'ai d'abord localisé et identifié des arbres qui convenaient aux mesures spectrales. J'ai visité 15 sites sur l'île de Montréal et autour de celle-ci, majoritairement des parcs-nature, et récolté des données de réflectance foliaire pour 50 espèces, ce qui représente près de 72 % de la diversité arborescente de la province (Brouillet et al., 2010). Parmi ces espèces, nous avons inclus des espèces rares au statut vulnérable, tels l'érable noir (*Acer nigrum*) et le pin rigide (*Pinus rigida*), dont l'aire de distribution se limite à quelques occurrences autour de Montréal et en Montérégie. Des espèces naturalisées, soit l'épinette de Norvège (*Picea abies*) et l'érable de Norvège (*Acer platanoides*), ont également été incluses, puisque leur suivi représente un enjeu pour la biodiversité québécoise (Lapointe and Brisson, 2011). Afin de mesurer la distance phylogénétique des espèces, j'ai construit une phylogénie à partir de données moléculaires. Pour la distance fonctionnelle, 11 traits foliaires, chacun possédant une influence documentée sur les propriétés spectrales foliaires, ont été mesurés à partir des feuilles récoltées dans le cadre de la prise de données spectrales. Afin de classifier les espèces, j'ai utilisé une analyse discriminante par moindres carrés partiels, une méthode dont l'efficacité a été prouvée pour les données spectrales (Richter et al., 2016). En effet, cet algorithme est adapté à l'analyse de données fortement multivariées et autocorrélées, et croît en popularité pour la classification de taxons par les données spectrales (Girard et al., 2020; Meireles et al., 2020; Paiva, Perdiz and Almeida, 2021; Stasinski et al., 2021). Afin d'évaluer la ressemblance spectrale des espèces proches, le signal phylogénétique a été calculé avec le K de Blomberg (Blomberg et al., 2003). Celui-ci évalue la ressemblance réelle des taxons proches par rapport à ce qui serait attendu suivant une évolution stochastique des traits (modèle brownien). Finalement, afin de tester notre hypothèse, nous avons utilisé une

forme de modèle linéaire généralisé pour modéliser l'effet de la différenciation fonctionnelle et phylogénétique des espèces sur l'incidence et l'intensité des erreurs de classifications entre chaque paire d'espèces.

Les résultats obtenus dans le cadre du projet de recherche sont présentés sous la forme d'un article scientifique préparé en vue d'une soumission à la revue *American Journal of Botany*, qui forme le corps du second chapitre. Les autrices.eurs ayant participé à l'élaboration du manuscrit sont Florence Blanchard, Etienne Laliberté et Anne Bruneau. La première autrice a effectué le travail de terrain, l'analyse des données, ainsi que la rédaction du manuscrit. Etienne Laliberté et Anne Bruneau ont supervisé le projet, commenté et corrigé le manuscrit.

Chapitre 2

Foliar spectra accurately distinguish temperate tree species and show strong phylogenetic signal

Florence Blanchard^{1*}, Anne Bruneau¹, Etienne Laliberté^{1,2},

¹Institut de recherche en biologie végétale and Département de sciences biologiques, Université de Montréal, Université de Montréal, 4101 Sherbrooke Est, Montréal, Québec, Canada H1X 2B2

²Centre de la science de la biodiversité du Québec, McGill University, Stewart Biology Building Office S3/18, 1205, Avenue du Dr. Penfield Montréal, Québec, Canada H3A 1B1

*Corresponding author: contact florence.blanchard@umontreal.ca

Introduction

Plant communities are being altered by global drivers such as habitat loss and fragmentation, changing climate and increased disturbances from fires or pest outbreaks (Larsen et al., 2005; Scheffers et al., 2016). Because of their rapid pace and spatial extent, changes in species composition and diversity, which threaten ecosystem integrity (Hooper et al., 2012), are increasingly difficult to evaluate using traditional field-based approaches. New biodiversity monitoring approaches are needed to keep track of, explain, and predict these changes. One such approach involves the use of remotely sensed foliar reflectance spectra, an increasingly employed method in the Earth observation field globally (Ustin and Middleton, 2021). Spectral properties of plants attracted interest in ecological sciences as early as the 1950s, when research focused on how spectral differences pertained to specific plant adaptations and their link to physiological processes (Billings and Morris, 1951; Gates et al., 1965). Since then, refined characterization of the link between functional traits and leaf-level spectra has been undertaken (Asner, 1998; Curran, 1989; Kiang et al., 2007; Kokaly et al., 2009; Hovi et al., 2017), and has led to gradually more reliable models for estimating these traits from spectra in a variety of ecological context (Serbin et al., 2014; Feilhauer et al., 2015; Singh et al., 2015; Serbin et al., 2019).

The close relationship between spectral reflectance and foliar traits stems from the biochemical and physiological adaptations of vegetation to their environment. As such, these differences highlight how functional traits impact ecosystem processes as well as the trade-offs that underlie plant strategies (Reich, 2014; Cavender-Bares et al., 2017). Leaf trait variability arises both from the structural and biochemical composition of leaves. These affect different portions of the electromagnetic spectrum due to scattering effects or the absorption of chemical constituents at specific bands, for example pigments, water or cellulose (Ollinger et al., 2011). As such, even subtle differences in leaf constituents modify light absorptance and reflectance, allowing spectral data to capture effectively taxonomic and functional diversity across large areas, as well as broad environmental gradients (Ustin and Gamon, 2010; Asner and Martin, 2016b, Schweiger and Laliberté, 2022). Moreover, vegetation changes in response to stress such as drought or damage due to insect outbreaks affect detectable spectral properties at the leaf and landscape scale

(Carter and Knapp, 1991; Ramoelo et al., 2015; Huang et al., 2019). As a result, spectroscopy holds high potential for monitoring plant biodiversity and improving our understanding of ecosystem function and species response to global change (Asner and Martin, 2016a; Cavender-Bares et al., 2017; Kothari and Schweiger, 2022).

While efforts have been made to understand the link between functional traits and leaf-level reflectance variation across ecological (Asner et al., 2014; Wu et al., 2017; Schweiger et al., 2020) and temporal (Yang et al., 2016) gradients, the impact of this variation on species distinction from spectra has received much less attention. Classifying plants at the species level is important for conservation because it constitutes an essential requirement for the reproducible estimation of distribution shifts, especially considering highly contrasted responses of different tree species to climate change (Fei et al., 2017; Taheri et al., 2021). Therefore, exploring the opportunities and limits of species distinction from reflectance spectra is necessary to monitor tree species distribution and diversity over large areas using imaging spectroscopy (Ustin and Gamon, 2010).

Many methods have been employed to classify tree species from hyperspectral data, as reviewed in Fassnacht et al. (2016), with levels of species classification accuracy varying between 75 and 90%. For example, Racko and Zagajewski (2017) compared three non-parametric methods to taxonomically identify five European tree species, yielding classification accuracies between 62 and 77%, while Shi et al. (2018) combined hyperspectral data with LiDAR and leaf functional traits estimates to identify five tree species, with different data combinations resulting in overall accuracies of 65 to 83%. Partial least-squares discriminant analysis has been shown to accurately classify plant species when applied to spectral data (Richter and al., 2016), and has become a popular approach to distinguish either plant species (Girard et al., 2020; Paiva, Perdiz and Almeida, 2021), broader taxonomic groups (Meirless et al., 2020) or even hybrid species complexes (Stasinski et al., 2021).

The promising results from the high levels of accuracy obtained in species classification are often attributed to the taxonomic organization of spectral variation (Meireles et al., 2020). For instance, Asner and Martin (2011) and Asner et al. (2014) both find that most

of the spectral variation, among respectively 328 and 1449 tropical rainforest plants, can be attributed to taxonomic identity. In addition, spectral data integrates genetic and phylogenetic information (Madritch et al., 2014; Cavender-Bares et al., 2016; McManus et al., 2016; Meireles et al., 2020). This helps to explain the positive relationships found between spectral, functional and phylogenetic distance (Schweiger et al., 2018), a link which justifies the use of spectral diversity as a proxy for taxonomic diversity. However, while closely related species tend to be more similar, there exists a certain threshold, generally within the family level, after which more phylogenetically distant species are not necessarily more functionally dissimilar (Kelly et al., 2014, Paquette et al., 2015). This may hinder species identification from spectral data when multiple species from distant clades are involved. Estimates of the relation between functional and phylogenetic diversity might also be impaired by species richness (Pavoine et al., 2013). Moreover, studies that have tried to classify plant species from spectra have, for the most part, focused on just a few species, from three to a little under 20. They also concentrate on local areas, while species classification at larger scales remain rare (Fassnacht et al., 2016). Due to the effects of stand diversity and heterogeneity, as well as high intraspecific variability from environmental or phenological variation, both of which can impede species classification accuracy (Modzelewska et al., 2020; Clark et al., 2005), it is necessary to challenge the discriminatory power of spectral reflectance across wider sets of species. In addition, it is worth investigating patterns of misclassifications based on species phylogenetic and functional distinctiveness, to further the value of spectral data as proxy for these as well as to help correct identification estimates (Fassnacht et al., 2016).

In this study, we explore the relationship between phylogenetic relatedness, functional similarity in foliar traits and spectral data, by focusing on a regional tree species pool. We aim to distinguish close to 50 eastern North American tree species (conifers and angiosperms) from their foliar reflectance spectra, which were sampled across different ecological conditions, and examine how reflectance spectra are phylogenetically structured at different levels of the taxonomic hierarchy. We further investigate whether tree species are more likely to be confused with others that are phylogenetically or functionally similar. We hypothesize that evolutionary patterns correlate with leaf functional differences,

resulting in significant phylogenetic signal across the spectrum. Specifically, we predict that tree species will be mistaken more often for close relatives; alternatively, the opposite could indicate functional convergence in foliar traits. We suggest that integrating phylogenetic relatedness patterns with spectral data, which has not received much attention (but see Cavender-Bares et al., 2016; McManus et al., 2016, Meireless et al., 2020, Diniz et al., 2020), will help us to better understand the drivers of spectral distinctiveness among species. Furthermore, the potential to link species relatedness patterns with spectra could allow us to assess plant diversity from spectral data in the case where spectral signatures from unknown species are recovered.

Methods

Study sites and species sampling

In order to sample tree species across environmental gradients that cover as much as possible the environmental niche of each species, multiple sites were included in this study spanning ecological reserves to public parks. In total, 15 locations were visited, mostly on and around the island of Montréal (Quebec, Canada), as well as a few more distant sites in order to sample certain tree species with a more restricted range (Supplementary Figure S1). Our criteria for site selection were to: (1) sample species in their natural context, (2) include species with very limited distributions in Quebec (e.g. *Pinus rigida* P. Miller, *Acer nigrum* F. Michaux), (3) include environmental variation in order to account for the potential effects of abiotic factors (e.g. soil fertility) on spectral properties (Asner et al., 2014). To satisfy this last criterion, whenever possible, we sampled individuals from the same species from at least two sites, with a minimum of three individuals per site. Our final dataset includes 819 trees representing 51 species, distributed across 12 taxonomic families (Annexe 1).

Leaf collection and preparation

Field sampling took place from June to August in 2018, 2019 and 2020. After identifying suitable trees to the species level, we pruned a small branch from the side of the tree crown using an extendable pole or an arborist throw-line launcher (Youngentob et al., 2016). We selected branches that were exposed to direct sunlight for at least four hours per day. We

selected six leaves per branch for measurements of spectral reflectance and transmittance, targeting the most intact leaves since herbivory or fungal/insect based diseases can alter chemical composition of the leaf and influence spectral properties (Fallon et al., 2020). The remaining leaf material was conserved as bulk samples for functional trait analyses. Pressed specimens of each individual are also available, and a representative sample for each species per site will be deposited at the Marie-Victorin Herbarium (MT), Université de Montréal, Quebec.

Spectral data acquisition and processing

Foliar spectral properties were measured with a HR-1024i field spectroradiometer fitted with a DC-R/T integration sphere (Spectra Vista Corporation, Poughkeepsie (NY) USA). We followed the protocol of Laliberté and Soffer (2018a) for broad leaves and Laliberté and Soffer (2018b) for conifer needles and scales. As newly formed coniferous needles are very different from their older counterparts in evergreens (Jensen et al., 2015), we used needles from the preceding growing season, with the exception of *Larix laricina* (Du Roi) K. Koch which is deciduous. Processing of spectral data and all subsequent statistical analysis were conducted in R v. ≥ 3.6 (R Development Core Team 2019). Specifically, we used the *spectrolab* package (Meireles et al., 2017) to interpolate spectral data to match 1 nm intervals and remove detector overlap. We applied a Savitzky-Golay filter from package *signal* (Signal Developers, 2013) to smooth out high-frequency noise with varying order and length according to the spectral region: order 3 and length 21 from 350-715 nm, order 3 and length 35 from 715-1390 nm, order 3 and length 75 from 1390-1880, and order 5 and length 175 from 1880-2500 nm. When working with spectra from individual leaves (for PLS-DA), reflectance was brightness-normalized to emphasize differences in spectral shape rather than amplitude (i.e. albedo), otherwise the mean reflectance per individual, that is the mean of all six leaves, was used. Wavelengths from 400-2400 nm were used in all analyses.

Functional traits

Pigments and total phenolic concentrations were measured from leaf disks recovered directly on leaves selected for spectral analysis. Other functional traits were measured from the dried and ground remaining bulk material collected from the field. We measured leaf carbon fractions (Ayotte and Laliberté, 2019), total carbon and nitrogen total concentrations (Ayotte et al., 2019) and photosynthetic pigments, including chlorophyll *a*, chlorophyll *b* and bulk carotenoids (Girard et al., 2019). Water content and leaf mass per area (LMA) were measured by weighing fresh, rehydrated and dried leaves selected from the bulk sample. We scanned fresh leaves using a flatbed scanner (CanScan LIDE 220, Canon, Brampton, Canada) and the software WinFOLIA Reg. 2016b (Régent Instruments Inc., Sainte-Foy, Canada) (Laliberté, 2018).

Phylogenetic reconstruction

A concatenated DNA matrix for all species of interest was created using the PYPHLAWD pipeline of Smith and Walker (2018). We used a clustering analysis, which searches GenBank for all available sequences pertaining to a given species list (here, our 51 tree species). It then creates aligned clusters, loosely equivalent to individual genes, identified by BLAST and Markov clustering analysis. We then ran the prebuilt scripts to select clusters and construct the concatenated supermatrix, which grouped 68 clusters for a total of 78,205 characters. Default settings were used. From this supermatrix, we constructed a maximum likelihood topology with RAxML v. 8.2.10 (Stamatakis et al., 2014), using the GTRCAT substitution model, and ran 100 bootstrap replicates. Following Paquette et al. (2015), who studied a similar set of taxa, we added four species (*Euonymus americanus* L., *Catalpa bignonioides* Walter, *Liriodendron tulipifera* L., *Viola palustris* L.) in the PYPHLAWD search and subsequent RAxML reconstruction for the purpose of cutting long branches and ensuring topological conformity. Indeed, our aim was to use an accurate phylogeny for all our target species, not to gain phylogenetic insight on alternative taxon placement in the topology.

Branch lengths were time-calibrated in BEAST v.1.10.2 (Suchard et al. 2018) applying the penalized likelihood ages from Magallón et al. (2015) to the coinciding nodes in our topology as priors with a lognormal distribution. We fit a relaxed lognormal clock model

across partitions. We used the RaxML topology as a starting tree, which was linked between clusters, while the substitution model was unlinked. We fit a birth-death model as the tree prior, with birth and death rate set to a gamma (2,100) model. We ran the Markov-Chain Monte Carlo algorithm for 30 million generations, sampling every 5,000 generations. Chain convergence was examined in Tracer (Rambault et al., 2014), after which we discarded the first 25% trees as burn-in and extracted the Maximum clade credibility (MCC; Supplementary Figure S2) from the remaining tree distribution. We find one instance of conflict: *Larix laricina* was placed as a sister group to the remainder of Pinaceae, while other molecular evidence and fossil data place *Larix* as a part of the pinoid clade, with *Pinus* and *Picea* (Lin et al., 2010; Ran et al., 2018). Despite this minor conflict, our method yields results consistent with prior phylogenetic knowledge across the angiosperms (Hawkins et al., 2014; Chase et al., 2016). The four species added for topological conformity were dropped for all statistical analyses.

Statistical analyses

Distinguishing species with partial least squares discriminant analysis (PLS-DA)

We used partial least-squares discriminant analysis, using the R package *caret* (Kuhn, 2009) to predict the species-level classification accuracy from spectral reflectance. This method is designed for using with a large number of predictive components showing a high degree of collinearity (Barker et Rayens, 2003; Chevallier et al. 2006), as is the case spectral data. We used the individual leaf-level, full spectrum data ranging from 400-2400 nm inclusively and resampled to 10 nm intervals. To avoid heavily unbalanced classes, we randomly resampled smaller subsets of the most abundant species in the dataset and restricted species sampling to between five and 25 individuals, pooling them from different sites to retain as much environmental variation as possible. We also removed species with spectral measurements on fewer than ten leaves (*Aesculus hippocastanum* L., *Juglans nigra* L., *Pinus banksiana* Lambert, *Prunus nigra* Aiton) and pooled doubtful species level identifications to the genus level (*Sorbus* spp., *Ulmus* spp.). This yielded a final dataset of 3654 leaf-level observations from 609 individual trees distributed across 45 species from 21 genera (Supplementary Table S1).

Our spectral dataset was separated into training and testing subsets comprising respectively 75% and 25% of the data, with samples randomly assigned to a partition in each iteration. We first trained the model over 50 iterations with bootstraps and extracted Cohen's Kappa (κ) to assess the accuracy of our model according to the number of components conserved. To avoid overfitting, we performed a Tukey test on Kappa values in an initial model fitting phase following Beauchamp-Rioux et al. (in preparation). For our final model, we chose 45 components, the smallest number of components yielding a model statistically equivalent to the most accurate model. The final model then ran for 100 iterations, allowing us to build an average confusion matrix across iterations to show classification results, used to extract a final model Kappa, specificity, precision and sensitivity accuracy measures.

Explaining classification mistakes: patterns of phylogenetic and functional differentiation

Phylogenetic signal and intraspecific variability

We assess similarity in spectral reflectance among related species by measuring phylogenetic signal strength. For this we use Blomberg's K (Blomberg et al., 2003), implemented in the *phylosig* function of the R package *phytools* (Revell, 2012). Blomberg's K measures the ratio between a clade's observed tip mean square error and its expected mean squared errors (MSE). The expected MSE follows a stochastic model of evolution in which disparity between taxa is strictly proportional to time since divergence, known as Brownian motion (BM). Values of K usually vary between 0 and 1, with 0 signifying a trait value distribution independent from the phylogeny, while a value of 1 expresses trait distribution perfectly concordant to what should be expected under BM. However, values can exceed 1 when similar species show very strong trait clustering within the observed clade (Münkemüller et al., 2012). We chose Blomberg's K for its ability to easily account for intraspecific variation. We measure the phylogenetic signal at two distinct taxonomic levels: first by comparing the two broad clades represented by all conifers and all angiosperms, and then at the family level. Because Blomberg's K compares the similarity of within-group traits, taxa with a single representative at the family partition were left out of the analyses. In addition, we measure the coefficient of variation (CV) for each spectral wavelength to compare species intraspecific variation to the mean variation across and within species. Finally, We perform as well a Kolmogorov-Smirnov Importance index test to distinguish between significantly distinct clades for spectral reflectance

(Cornwell et al 2014). This identifies lineages that alter the global trait distribution and ranks nodes from those which contribute the most to functional diversity to those which contribute least (i.e. their trait values do not form a distinct distribution). This test accounts for species richness in a clade, but does not consider branch lengths. We perform the analysis using each wavelength as a trait. We order the most distinct clades based on the sum of wavelength scoring the respective rank. For example, if a node is identified as the most distinct (rank 1) for 15 of the wavelengths, while a second node is considered the most distinct for 1300 of the wavelengths, then we consider this second node as the most distinct across the entire spectrum, and so on for each rank.

Phylogenetic and functional patterns between misclassified species

To test our hypothesis regarding misclassification frequency, we discarded diagonal elements of the confusion matrix, which represent accurate classifications. Here, our analyses focus on the off-diagonal elements from our PLS-DA results, which represent all possible classification mistakes (or misclassifications) between different species pairs. Since our classification analysis from reflectance spectra was largely successful, it yielded few or no misclassification events between most species pairs. However, since this over-representation of zeros arises from successful differentiation of two species by our models, we consider them “true zeros” (for discussion on true and false zeros see Zuur et al. (2009)). As such, we further explored misclassification frequency and their relationship with either phylogenetic or functional distances as two separate processes using a hurdle model, designed to accommodate the large number of zeros (Mullahy, 1996; Cameron and Trivedi, 1998). Using misclassification frequency (in percent) from the PLS-DA, we consider the two following processes in our hurdle model: 1) processes that allow two species to be distinguished or not, represented by the zero-hurdle model; 2) processes that control misclassification frequency, when they arise, between a particular species pair, represented by the count model. The first process is described by a binary “misclassification or not” output between a given species pairs, while the second determines the number of subsequent misclassification events.

To test for the initial correlation and develop our hurdle model, we first extracted the mean phylogenetic and functional distances among all species pairs from our data. Phylogenetic

distance was measured as the sum of branch lengths, from the *cophenetic.phylo* function from the package *ape* (Paradis and Schliep, 2019), while functional distance was measured using the Gower distance from the *FD* package (Laliberté and Legendre, 2010) from the following leaf traits: leaf mass per area (LMA), equivalent water thickness (EWT), lignin (%), cellulose (%), hemicellulose (%), soluble carbon (%), total carbon (%), total nitrogen (%), bulk carotenoids, and chlorophyll *a* and *b*. Both distances were rescaled between 0 and 1 to allow comparison of model coefficients. We then used the R package *pscl* (Jackman et al., 2020) to fit the hurdle model using functional and phylogenetic distances as predictors in both the zero and count parts of the model. Models with a Poisson and negative binomial specification for the count model were compared by AIC and by a maximum likelihood test from *lmtest* (Zeileis and Hothorn, 2002) to test for overdispersion. Models were also tested with and without interaction between functional and phylogenetic distance.

Results

We find spectral differences across the different clades (Figure 2). Distinct lineages identified by the Kolmogorov-Smirnov test are presented on the MCC (Supplementary Figure S2).

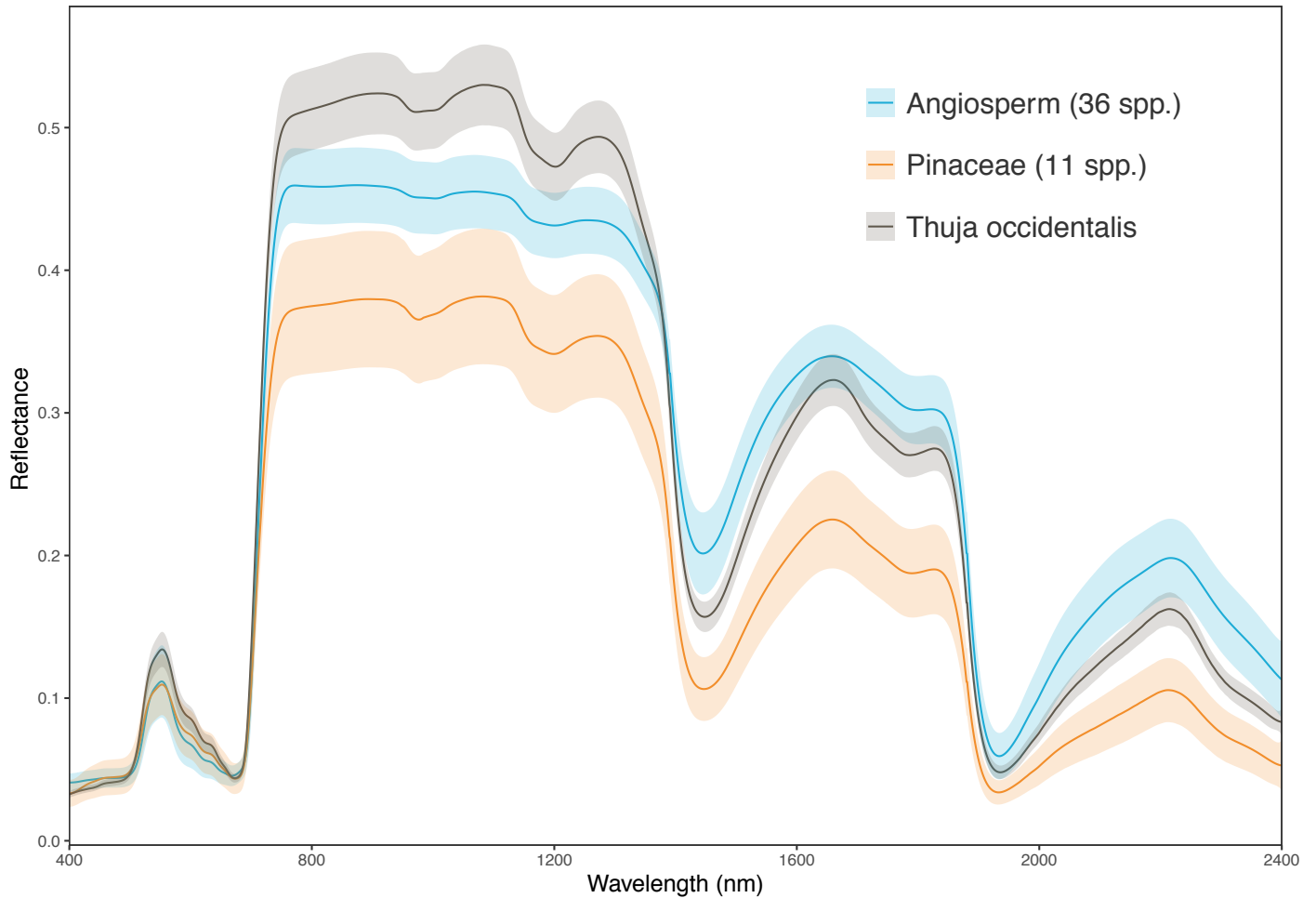


Figure 2 Spectral data overview for major clades

Mean reflectance with the ribbon representing \pm one standard deviation for broad clades. We separate *Thuja occidentalis* from the Pinaceae based on a Kolmogorov-Smirnov test which confirmed a significant difference between these two lineages.

Distinguishing species with partial least squares discriminant analysis (PLS-DA)

We were able to accurately classify almost all 45 southern Quebec regional tree species using PLS-DA ($\kappa = 0.736, \pm 0.005$) but there were substantial differences in classification performance across species (Figure 3; Table 1). This large variance can be attributed to some poorly classified species, notably *Picea abies* L., *Fraxinus pennsylvanica* Marshall and particularly *Quercus bicolor* Willdenow, for which our model failed to accurately classify a single sample. The median classification accuracy of 86% better represents the overall model performance. Indeed, 20 of our 45 species (44%) could be classified with a mean accuracy $\geq 90\%$, and seven more reached a mean classification accuracy of $\geq 70\%$. Moreover, while the model lacks sensitivity for some species, it still shows good precision,

whereas more successful identifications are counterbalanced by lower precision (Table 1). Our model's precision ranges from 57% (*Picea rubens* Sargent) to 100% (*Pinus resinosa* Aiton). Important wavelengths for species discrimination are clustered in the visible and red edge, as well as in wavelengths greater than 1880 nm (Supplementary Figure S3)

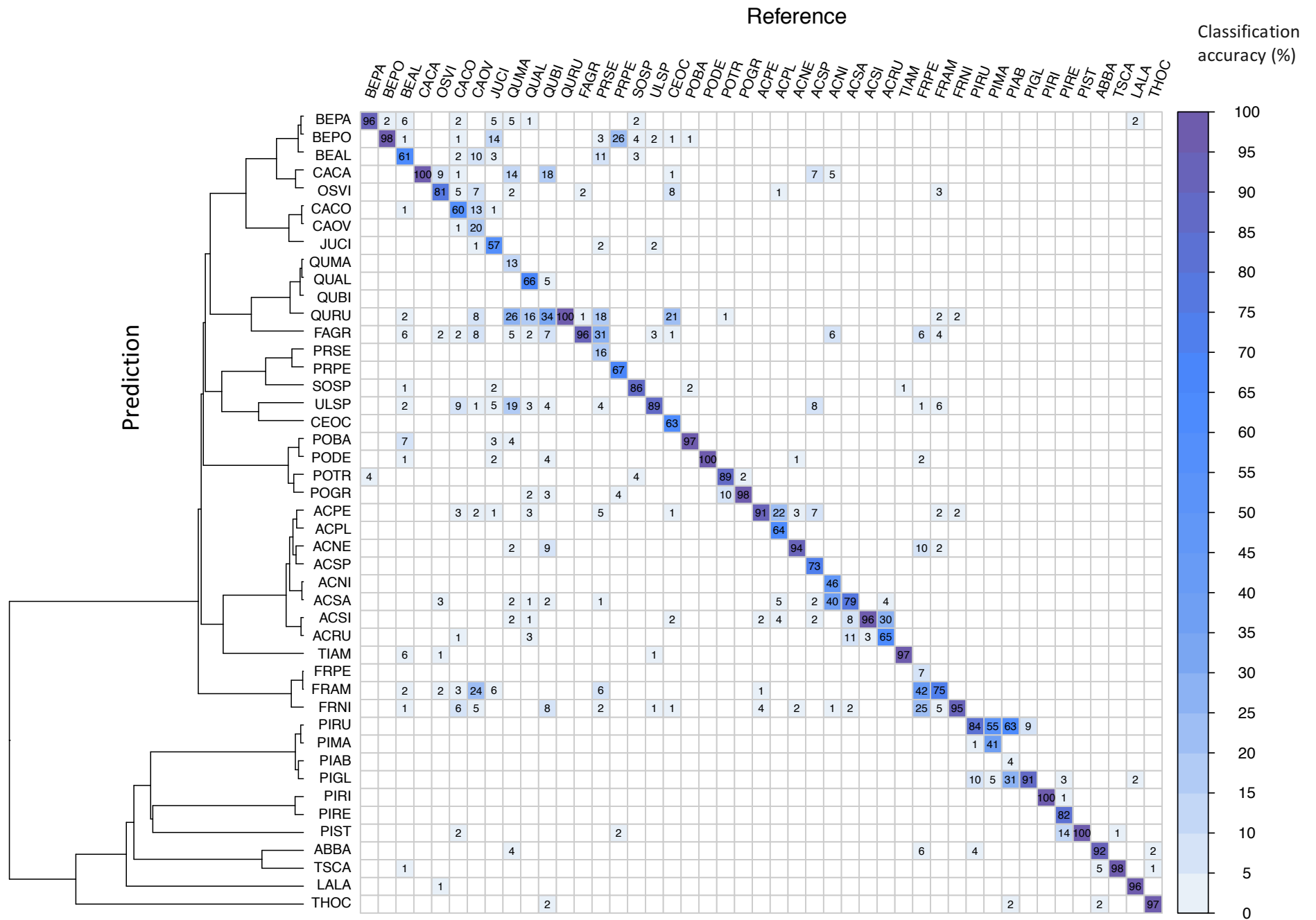


Figure 3 Confusion matrix from species classification by PLS-DA

Diagonal shows correctly classified spectra, while off-diagonal elements represent misclassifications. The topology is joined to facilitate the visualization of species evolutionary relationships. Abbreviations for species name are listed in Table 1

Tableau 1 Summary statistics from species classification by PLS-DA

Sensitivity measures the chances of correctly identifying a species (true positive rate), while precision measures the model's capacity to avoid classifying spectra that do not belong to the species in question. We recover as well a mean specificity 99.5 % (0.5 SD), and a Kappa of 0.736 (0.005 SD). Specificity is a measure of the true negative rate while Kappa is a measure of accuracy corrected for agreements that may have occurred by chance. Species are listed in decreasing order of classification accuracy (sensitivity), then by decreasing order of precision. The PLS-DA was performed with 45 components.

Species	Code	Sensitivity (%)		Precision (%)	
		Value	SD	Value	SD
<i>Populus deltoides</i> W. Bartram ex Marshall	PODE	100.0	(0.0)	91.7	(5.6)
<i>Pinus strobus</i> Linnaeus	PIST	100.0	(0.0)	83.3	(6.4)
<i>Carpinus caroliniana</i> Walter	CACA	100.0	(0.0)	70.3	(7.8)
<i>Quercus rubra</i> Linnaeus	QURU	100.0	(0.0)	63.4	(5.0)
<i>Pinus rigida</i> P. Miller	PIRI	99.9	(0.9)	98.5	(2.7)
<i>Populus grandidentata</i> Michaux	POGR	98.5	(2.1)	87.9	(5.5)
<i>Betula populifolia</i> Marshall	BEPO	98.1	(2.4)	73.8	(5.9)
<i>Tsuga canadensis</i> (Linnaeus) Carrière	TSCA	97.9	(3.1)	90.7	(5.8)
<i>Thuja occidentalis</i> Linnaeus	THOC	96.9	(3.4)	95.9	(4.4)
<i>Populus balsamifera</i> Linnaeus	POBA	96.7	(3.4)	83.4	(7.2)
<i>Tilia americana</i> Linnaeus	TIAM	96.5	(4.2)	87.5	(7.9)
<i>Acer saccharinum</i> Linnaeus	ACSI	96.3	(3.5)	70.9	(5.5)
<i>Larix laricina</i> (Du Roi) K. Koch	LALA	96.0	(4.4)	98.3	(3.0)
<i>Betula papyrifera</i> Marshall	BEPA	95.9	(3.7)	82.4	(4.9)
<i>Fagus grandifolia</i> Ehrhart	FAGR	95.6	(4.1)	64.9	(5.4)
<i>Fraxinus nigra</i> Marshall	FRNI	95.2	(3.8)	74.9	(5.8)
<i>Acer negundo</i> Linnaeus	ACNE	94.1	(5.2)	87.9	(6.6)
<i>Abies balsamea</i> (Linnaeus) Miller	ABBA	92.4	(5.2)	91.4	(4.9)
<i>Acer pensylvanicum</i> Linnaeus	ACPE	91.2	(5.2)	76.7	(5.4)
<i>Picea glauca</i> (Moench) Voss	PIGL	90.9	(6.9)	72.8	(8.5)
<i>Ulmus sp</i> Linnaeus	ULSP	89.2	(6.6)	71.3	(6.1)
<i>Populus tremuloides</i> Michaux	POTR	88.9	(6.2)	92.0	(5.1)
<i>Sorbus sp</i> Linnaeus	SOSP	86.1	(7.0)	91.7	(5.3)
<i>Picea rubens</i> Sargent	PIRU	83.8	(7.5)	57.2	(6.1)
<i>Pinus resinosa</i> Aiton	PIRE	81.8	(8.2)	99.9	(1.0)
<i>Ostrya virginiana</i> (Miller) K. Koch	OSVI	80.7	(7.3)	76.9	(7.4)

<i>Acer saccharum</i> Marshall	ACSA	78.7	(7.6)	68.6	(6.9)
<i>Fraxinus americana</i> Linnaeus	FRAM	75.3	(7.1)	63.2	(6.6)
<i>Acer spicatum</i> Lamarek	ACSP	73.3	(10.7)	98.7	(4.2)
<i>Prunus pensylvanica</i> Linnaeus f.	PRPE	66.6	(9.8)	98.8	(3.2)
<i>Quercus alba</i> Linnaeus	QUAL	66.4	(10.3)	97.3	(4.5)
<i>Acer rubrum</i> Linnaeus	ACRU	65.4	(8.1)	79.3	(7.7)
<i>Acer platanooides</i> Linnaeus	ACPL	64.3	(12.0)	96.9	(5.4)
<i>Celtis occidentalis</i> Linnaeus	CEOC	62.5	(10.6)	97.6	(4.5)
<i>Betula alleghaniensis</i> Britton	BEAL	61.3	(7.0)	80.6	(6.4)
<i>Carya cordiformis</i> (Wangenheim) K. Koch	CACO	59.9	(10.2)	81.1	(9.8)
<i>Juglans cinerea</i> Linnaeus	JUCI	56.6	(9.7)	91.2	(9.3)
<i>Acer nigrum</i> F. Michaux	ACNI	46.5	(12.7)	99.0	(3.4)
<i>Picea mariana</i> (Miller) Britton. Sterns & Poggenburgh	PIMA	40.7	(13.0)	95.6	(9.5)
<i>Carya ovata</i> (Miller) K. Koch	CAOV	20.1	(8.2)	89.3	(19.0)
<i>Prunus serotina</i> Ehrhart	PRSE	16.0	(6.9)	93.9	(19.6)
<i>Quercus macrocarpa</i> Michaux	QUMA	13.0	(7.9)	88.7	(31.5)
<i>Fraxinus pennsylvanica</i> Marshall	FRPE	7.4	(8.2)	50.8	(49.3)
<i>Picea abies</i> (Linnaeus) H. Karsten	PIAB	3.7	(6.6)	24.0	(42.3)
<i>Quercus bicolor</i> Willdenow	QUBI	0.0	(0.0)	0.0	(0.0)
Mean		73.8	(30.2)	80.7	(2.0)

Explaining classification mistakes: patterns of phylogenetic and functional differentiation

Phylogenetic signal

We find significant phylogenetic signal ($p < 0.05$, 999 iterations) at many wavelengths, but patterns vary strongly across taxonomic divisions (Figure 4). First, we find a striking contrast between conifers and broadleaf species. The conifers show very strong phylogenetic signal, with values of $K \gg 1$ across all spectral regions except the visible range. By contrast, the broadleaf species show some moderate to strong phylogenetic signal only in the visible range and around a minor and major water feature, which are located at 1381 nm and from 1940 nm to 1960 nm, respectively. At the family level, we also observe strong phylogenetic signal in the Pinaceae, although less in the NIR and more in the visible range. Most families show phylogenetic signal towards the end of the SWIR range (2051–2400 nm), a region associated with cellulose, lignin and protein content (Homolová et al., 2013; Jacquemoud and Ustin, 2019). A strong and consistent phylogenetic signal in the NIR distinguishes the Betulaceae from other families. Additionally, the Sapindaceae shows restricted phylogenetic signal, with only 17 wavelengths between 1200 and 1365 nm presenting significant but small K values, indicating a relatively poor relationship between reflectance variation and phylogenetic structure in maples. We find no phylogenetic signal in the Rosaceae.

Intraspecific variation

We show that in most cases, interspecific variation is far superior to species-specific coefficients of variation (Figure 5; Supplementary Figure S4). Exceptions occur almost exclusively in the VIS between 550–600 nm and 700 nm. Variation in conifers usually exceeds mean intraspecific variation in the NIR, and even exceeds interspecific variation in the case of the white pine (*Pinus strobus* L.). American beech (*Fagus grandifolia* Erhart) however, displays the most notable variation in the SWIR.

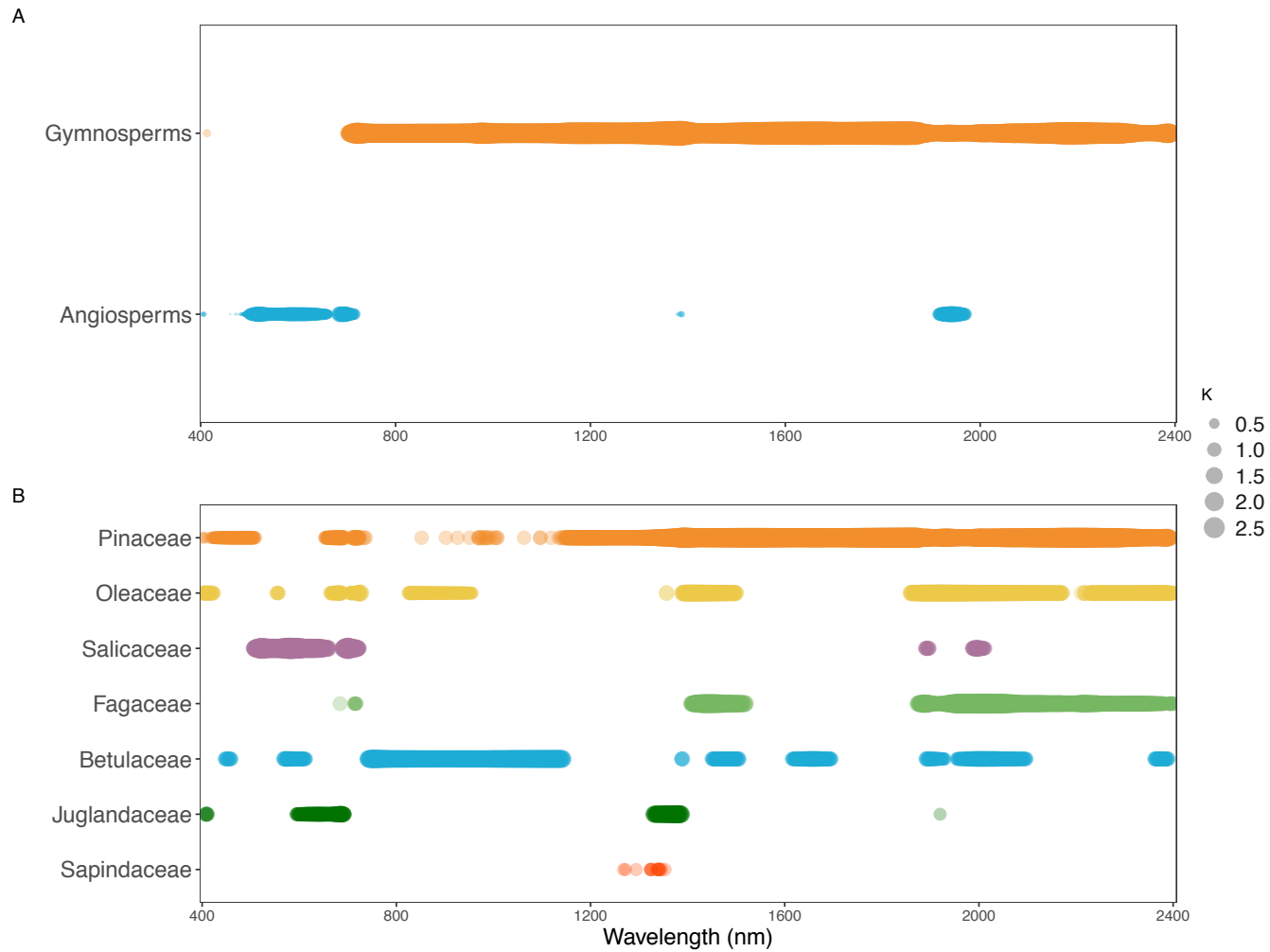


Figure 4 Phylogenetic signal across reflectance spectra

Phylogenetic signal measured for each wavelength for A) Broad clades and B) Families. Families with only one representative are excluded from the analysis since Blomberg's K measures variability within a clade to variability across clades

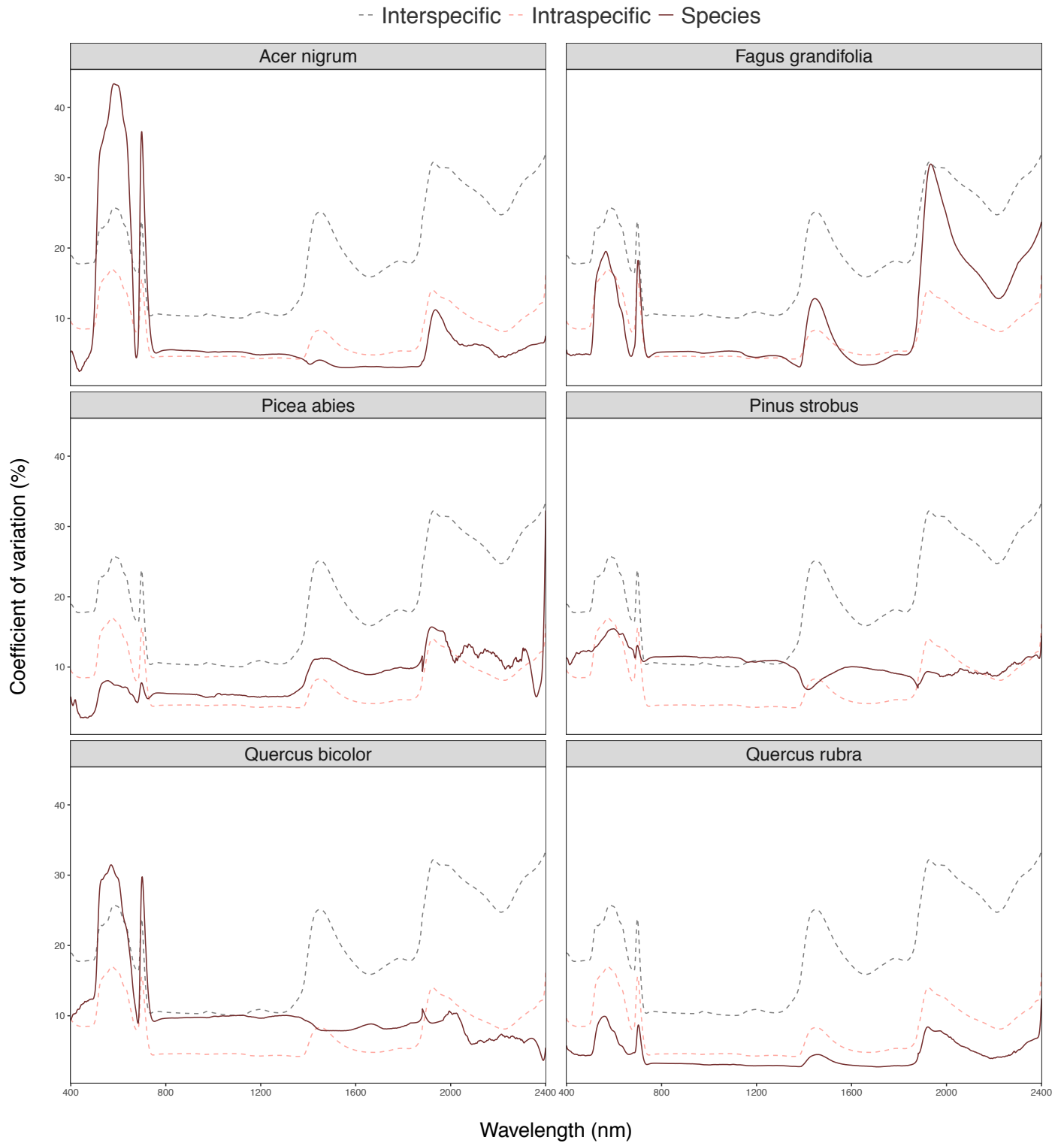


Figure 5 Coefficient of variation across the spectrum for selected species
 Dotted lines represent mean interspecific and intraspecific variation.

Phylogenetic and functional patterns between misclassified species pairs

Results from the hurdle model show that the interaction between functional and phylogenetic normalized distances did not yield a significantly different model (Supplementary table S2), and therefore we present results for the most parsimonious model without this interaction term (Table 2). Increases in both phylogenetic or functional distance significantly decreases the probability of mistaking two species based on their reflectance spectra (zero-hurdle model), as shown by negative coefficients. Although these predictors both affect misclassification likelihood in a similar way, a slightly higher marginal effect for functional distance indicates a steeper decline in the likelihood of misclassification occurrence with increasing functional differentiation between two species. While coefficients linked to misclassification frequency (count model) were similar between functional and phylogenetic distances, only the latter showed a significant effect. For one unit increase in phylogenetic distance between two species, misclassification frequency is expected to drop by 59 percentage points. This significant drop is further illustrated by results that show that extremely close relatives are expected to be misclassified twice as frequently as distant species (Figure 6).

	Coefficients (SD)		Marginal effects (SD)		P-value
Zero hurdle model coefficients					
Misclassification occurrence					
Intercept	-0.268	(0.185)			0.146
Normalized functional distance	-3.312	(0.775)	-0.202	(0.056)	1.92e-05 ***
Normalized phylogenetic distance	-2.294	(0.362)	-0.140	(0.057)	2.28e-10 ***
Count model coefficients:					
Misclassification frequency					
Intercept	1.440	(0.733)	-		0.0494 *
Normalized functional distance	-2.148	(1.410)	-0.459	(0.271)	0.1277
Normalized phylogenetic distance	-2.753	(0.560)	-0.588	(0.617)	8.79e-07 ***
Log(theta)	-2.051	(0.862)			0.0174 *
Log likelihood	-1089				
Number of estimated parameters	7				
AIC	2192.361				

Tableau 2 Summary of the fitted hurdle regression model on misclassification data

Coefficients for zero-hurdle model, count model (with both standard errors in parenthesis), marginal effects (mean slope), number of estimated parameters, log likelihood, and AIC. The zero hurdle part refers to misclassification occurrences (i.e chances of being mistaken or not) while the count model refers to the number of times a species was mistaken for another. Marginal effects (Mean effect slope) represent the change in outcome probability, in percentage points, due to an additional unit of the predictor, all other factors being held equal to a mean value.

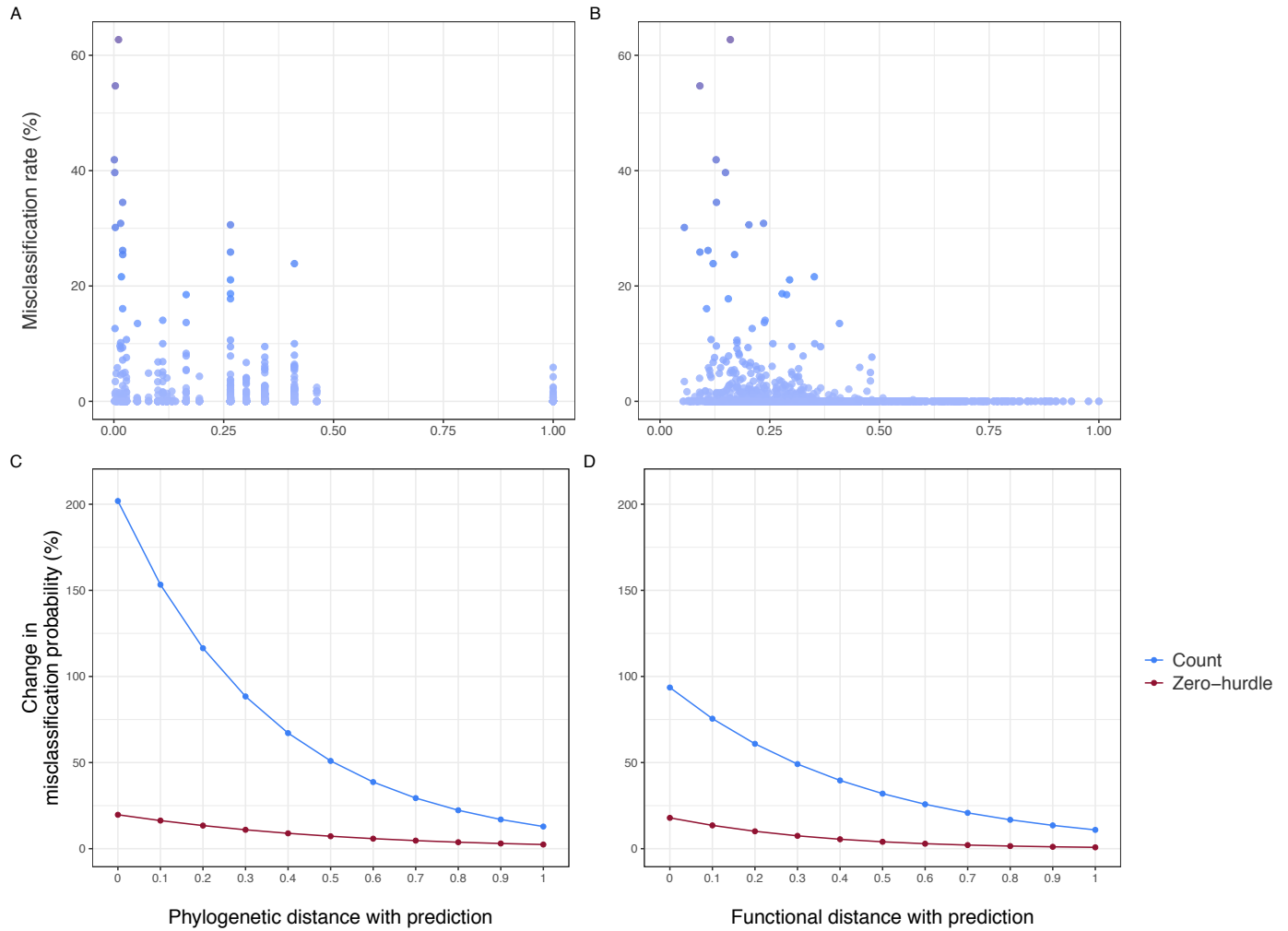


Figure 6 Misclassification rate and probability according to phylogenetic and functional distances
 Scatterplots of misclassification occurrences and predictor effects according to A) phylogenetic or B) functional distances between the reference and predicted species. Effects of increasing C) phylogenetic and D) functional distance on misclassification occurrence (zero - hurdle model) and frequency (count model) are expressed in terms of changes in misclassification probability (slope).

Discussion

In our study, we assess the potential of spectral reflectance for species classification and monitoring, a crucial step to follow shifts in the geographic distribution of species. To better understand and validate the species diversity patterns we derive from spectral data, it is also important that we establish links between spectral data and the macroevolutionary patterns of leaf functional traits. This study represents a first attempt at classifying tree species from hyperspectral data from a large species pool. We also seek to find explanations for the misclassification results in terms of phylogenetic and functional distances. We were able to accurately identify most of the species, but we find some frequently mistaken taxa, most of which are closely related phylogenetically. Indeed, we find that functional differentiation is crucial for distinguishing species and that phylogenetic relatedness explains misclassification frequency. This pattern is consistent with the trend linking functional, spectral and phylogenetic diversity (Schweiger et al., 2018; Frye et al., 2021).

Classification accuracy

We show that hyperspectral reflectance can be used to accurately distinguish most of our 45 tree species, sampled over varying environmental conditions at a regional scale. The median classification accuracy of 86% is comparable to that found in other studies (Cavender-Bares et al., 2016; Richter et al., 2016; Shi et al., 2018), and higher than those recovered for broader taxonomic clades (Asner et al., 2011; Meireles et al., 2020). Although some studies using PLS-DA as their classification algorithm report overall species classification accuracy of 90% and higher (Girard et al., 2020, Paiva, Perdiz and Almeida, 2021, Beauchamp-Rioux et al., in preparation), these studies focused on a much lower number of species. In contrast to these, we focus on 45 species and find extensive variation in the classification accuracy of the 45 species in our study, ranging from 100% for four species to 0% for *Quercus bicolor*. However, if we exclude *Picea mariana*, which dominates Quebec's extensive boreal forests, the poorly identified species are not the dominant species across Quebec's landscape. More common species in southern Quebec such as *Betula papyrifera*, *Q. rubra*, *Fagus grandifolia*, *Abies balsamea* and all poplars were all classified with *c.* 90% accuracy and higher, with *Acer saccharum* being on the lower end with a mean classification accuracy of 79%. Moreover, the distinction of conifers

was near perfect for all species except for the genus *Picea*. We also find that important wavelengths for classification are similar to other classification studies, most of them being situated in the VIS and close to water features in the SWIR (Fassnacht et al., 2016).

Some of the most misclassified species, including *Q. bicolor* and *P. abies*, or *Fraxinus pennsylvanica* were also less well represented in our dataset as a whole, and this reduced amount of training data probably affected classification accuracy for these species. However, there is clearly an influence of species spectral characteristics themselves, as *Carya ovata* and *Carpinus caroliniana* Walter were trained using the same number of spectra and yet their classification accuracies, 20% and 100% respectively, differ strongly.

Effect of species relatedness on species classification

As expected, more closely related species (e.g. congeners) get mistaken for one another more often than distantly related ones. We find that increases in phylogenetic distances as well as functional distances between two species lead to significant declines in misclassification occurrence, i.e. the probability of a species being identified as another. This constitutes further evidence that spectral variation is taxonomically organized (Asner et al., 2011; Asner et al., 2014; Cavender-Bares et al., 2016; Meireles et al., 2020). In addition to exerting a broad control on misclassification occurrence, phylogenetic distance is also strongly correlated to misclassification frequency, allowing us to confirm our hypothesis that tree species will be mistaken more often for close relatives. Thus, despite some evidence for functional convergence in certain cases, and the fact foliar traits have a much more direct impact on spectral separability, for the most part spectral differences reflect broad evolutionary differences. Indeed, phylogeny encompasses species differences on a larger scale than a single set of traits (Cadotte et al., 2017). This might explain the fact that while we still observe some misclassification between distant species, including broadleaf species and conifers, these mistakes are rare.

In addition, we find significant phylogenetic signal in spectral data, at least for some parts of the spectrum for most plant families, meaning most of the variation lies among families rather than within (Blomberg et al., 2003). Exceptions are found in the Rosaceae, for which no wavelength presents significant phylogenetic signal, and Sapindaceae, with near

absence of phylogenetic signal, which suggests a poor coordination between phylogenetic distances and spectral variation in both these clades. Although phylogenetic signal is widespread in various types of ecological data, this measure is nevertheless sensitive to the number of taxa included (Blomberg et al., 2003). In addition, adjacent wavelengths influence one another and cannot be considered independent observations. For this reason, statistical testing on K values must be interpreted with caution, and K strength is mostly a relative indicator of phylogenetic signal. Moreover, because we tested phylogenetic signal on a subset of taxa to fit our regional species pool, our analyses do not represent the extent of trait variation across the clade, which can lead to either disproportionately high signal or to its absence depending on the ecological processes dominating community assembly in the region (Cadotte et al., 2019). This poor coordination might be augmented by the exclusion of non-arborescent species, as trees occupy a distinct portion of the whole-plant functional traits spectrum (Díaz et al., 2016). For the maples, our sampling includes only a few species of this extensive temperate genus for which phylogenetic relationships are poorly resolved and in which morphological variation is considerable (Grimm et al., 2007; Li et al., 2019). It is therefore possible these taxa might be less similar amongst themselves than expected from our topology. Phylogenetic signal as used here also assumes a Brownian mode of evolution, which does not always correspond to the true trait evolution rate (Blomberg et al., 2019).

Nevertheless, the influence of phylogeny on species classification from spectral data was hinted at in studies that find positive relationships between spectral, functional and phylogenetic diversity (Schweiger et al., 2018, Frye et al., 2021), which can be extended to signify closer species have a more similar spectral signature. This pattern also provides further evidence that spectra carry evolutionary relatedness information across taxonomic groups (McManus et al., 2016, Diniz et al., 2020; Meireles et al., 2020). However, different regions of the reflectance spectra are not equally correlated to phylogenetic diversity, and trait conservatism varies across lineages, as alluded to by Meireles et al. (2020). As such, species relatedness might track misclassification frequency more or less closely depending on the taxonomic resolution and depth recovered from remotely sensed plant assemblages, and this relationship might also vary depending on the level of intraspecific variation.

In concordance with the strong link throughout the spectrum between reflectance variation and evolutionary distances in the gymnosperms, our ability to differentiate conifer species proves to be excellent. This is explained by the larger trait differences among conifer species than among broadleaf species, which is mirrored in spectral space (Supplementary Figure S5). This clear separation between conifer species is found despite their high coefficient of variation in the NIR region compared to the mean intraspecific variation. This supports the notion that structural characteristics are major drivers of spectral variability in needles (de Tomás Marín et al., 2016), and suggests that the NIR region differences have an important role in the differentiation of coniferous species. This would explain the spectral distinctiveness of *Thuja occidentalis* L., whose imbricate, scale-like leaves set it apart from the Pinaceae spectrally in this region, but not trait-wise (Supplementary Figure S5; Supplementary Figure S6). These classification results support the finding by Hovi et al. (2017) that species account for a major part of the spectral variation in conifers. However, it must be noted that, we controlled for leaf side and age, factors that also influence needle spectral variation, so these sources of variation could nonetheless hinder coniferous species detection at the canopy and stand level.

Effect of functional differences on species classification

According to our model, functional differentiation is a prerequisite for accurate species classification and acts as a threshold, whereby increasing functional distinctiveness diminishes the probability of misclassification occurring between any two species. However, it has little to no effect on misclassification frequency if the “hurdle” has been crossed (i.e. if the two species are very distinct functionally). This threshold effect is consistent with the strong linkage between spectral properties and foliar attributes of plants (Serbin et al., 2015, Feilhauer et al., 2015), in the sense that functionally very different species will also be different in spectral shape, and thus more easily differentiated using spectral reflectance. It does, however, reinforce the view that convergent properties of plant traits, which tend to organize around integrated strategies given similar environmental conditions (Reich et al., 2013; Díaz et al., 2016), pose a particular challenge for the remote sensing of species (Ollinger et al., 2011; Gamon et al., 2019).

The importance of functional distinctiveness can explain the lower classification accuracy we observe for some species, notably within the spruces or the oaks, in which hybridization is known to be frequent. For example, *Picea mariana* and *P. rubens* have been described as an example of incomplete speciation (De Lafontaine et al., 2015), while the phylogenetic history of oaks is fraught with evidence of introgressions, especially in the lobed-leaf White Oak clades (Rushton, 1993; McVay et al., 2017). This applies to *Acer nigrum* Michaux and *A. saccharum* Marshall as well (Skepnner et al., 1998). Although spectral reflectance data have been shown to successfully capture population-level structure (Cavender-Bares et al., 2016) and even intraspecific genetic diversity (Madritch and al, 2014), as shown by Stasinski et al. (2021), species in hybrid complexes may be more difficult to differentiate using spectral data. In a *Dryas* species complex, hybridization leads to a decrease in overall classification accuracy because the hybrids possessed intermediate traits to those of their parents, and this was captured in the spectral reflectance values as well (Stasinski et al., 2021).

While we cannot identify hybrids directly, we note the large variation displayed by *Acer nigrum* in the visible and red edge, particularly around wavelengths of great importance for the classification. This could indicate the presence of hybrid individuals which could have restricted accurate species determination by lowering spectral distinctiveness. We also note high intraspecific variation in *A. saccharum*, *Quercus bicolor*, *Q. macrocarpa* as well as in *Picea mariana* and *P. glauca*, almost exclusively in the visible portion of the spectrum. Of course, the likelihood of having hybrids is unequally distributed throughout a species range. For example, in Quebec, *A. nigrum* occurs only in areas where *A. saccharum* is also present, while the reverse is not true. This could account for the asymmetry in the misclassification between some close species pairs. A more thorough analysis of hyperspectral differentiation coupled with genetic information within genera with known hybrids, for example *Salix* spp, would be necessary to better assess remote sensing potential for monitoring these complex species groups. Considering that forests are not homogenous across the landscape, there might be some concern that ecosystems dominated by a single or a few genera with frequently hybridizing species, such as the

boreal forest which is dominated by spruces, may be more difficult to monitor at the species level.

Despite the importance of functional distance to differentiate spectra, we find that functional distinctiveness in some traits is not always sufficient to ensure high classification accuracy from spectral data. For example, *Celtis occidentalis*, the sole representative of the Cannabaceae in our sampling, shows a particularly low C:N ratio but shows a classification accuracy of only 63%. Similarly, *Acer platanoides* presents very fibrous leaves, but still gets often misclassified as *A. pensylvanicum*. As spectra constitute an integrative measure of plant traits (Cavender-Bares et al., 2017; Kothari and Schwieger, 2022), it is difficult to tease apart the relative contribution of each trait to spectral differentiability. Extreme values for one trait does not guarantee accurate discrimination because other traits may be very similar or because a combination of different traits might lead to a similar spectral pattern (Ollinger et al., 2011). For example, with regard to *A. platanoides*, the effect of high hemicellulose on spectral reflectance might be masked by water in fresh leaves (Jiménez and Díaz-Delgado, 2011; Kothari et al., 2021). Patterns of foliar traits influence on the SWIR region are complex because of their overlapping absorption (Curran, 1989), but this region seems the most important to distinguish multiple species in our study, and others also find strong phylogenetic structure in this region (McManus et al., 2016). Leaf-level studies examining functional differentiation in relation to phylogenetic diversity are thus still needed to understand what makes a spectral signature reliable for species identification.

Conclusions

Our results confirm the potential for hyperspectral distinction of a regional tree species pool. Although intraspecific variation and close resemblance due to potential hybridization seem to affect classification, functional differentiation of trees is sufficient to produce recognizable spectral signatures in most cases. Detecting individual species is key in the integration of this remotely sensed data at higher scale to develop a global biodiversity observatory, as proposed by Jetz et al. (2016). A clear advantage of hyperspectral data is our ability to derive plant traits from them, and as such, to bridge the severe gap in trait

data availability (Cornwell et al., 2019). For this, it is important that we be able to associate remotely sensed spectra to species identity in the most precise way possible. The highly significant relationship we find between phylogenetic distance and misclassification frequency constitutes a novel and exciting insight, which can help calibrate remote sensing classification models by correcting species abundance estimates. Work is still needed to explore this relationship, especially in hyperdiverse environments such as tropical forests, where the high number of species per clade might challenge our ability to relate spectral diversity to phylogenetic diversity. Our analyses still show the potential contribution of spectral data for species differentiation in what constitutes a representative and important part of the eastern North American forests.

Acknowledgments

The authors thank Rosalie Beauchamp-Rioux, Antoine Mathieu, Maria Juliana Pardo Losada, Madeleine Trickey-Massé, Sabrina Demers-Thibault, Clément Robert-Bigras, Isabelle Gareau, Myriam Cloutier, Fabien Cichonski, Alexandra Massé, Jocelyne Ayotte, Vincent Fournier and Rime Néron whose contribution to field work and laboratory analyses were essential to this study. We are grateful for the precious advice from Rosalie Beauchamp-Rioux, Anna K. Schweiger, José Eduardo Meireles, Johanna Jantzen and Simon Joly during the study. Furthermore, the authors thank the Société des établissements de plein air du Québec (SÉPAQ); the City of Montréal; CANOPÉE; the ministère de l'Environnement et de la Lutte contre les Changements climatiques (MELCC); as well as the staff from the Station de biologie des Laurentides for granting us access and sampling authorization at our various study sites. Finally, the authors thank IRBV colleagues for their enriching discussions.

Funding

This study is part of the Canadian Airborne Biodiversity observatory (CABO) project, which was funded by a Discovery Frontiers grant from the Natural Sciences and Engineering Research Council of Canada to E.L. and A.B. (NSERC; Grant number 509190-2017). F.B. received scholarships from NSERC and from Fond de recherche du Québec- Nature et Technologies (FRQNT) and has financial support from CABO.

Chapitre 3

Discussion

Retour sur les objectifs et hypothèses

L'identification des arbres revêt une importance capitale pour déterminer leur distribution, ainsi que pour assurer le suivi de l'abondance et de la composition des forêts, particulièrement dans le contexte des changements globaux (Hycza et al, 2018). D'autres utilités en conservation incluent la gestion durable des forêts (Durgante et al, 2013) ou encore la détection d'espèces envahissantes (He et al, 2011). L'objectif de cette étude était d'évaluer le potentiel de la réflectance spectrale pour l'identification des espèces arborescentes du sud du Québec. En incluant la majorité des espèces d'arbres d'une région, nous souhaitons représenter l'ensemble des confusions potentielles à travers un territoire. De plus, nous souhaitons examiner les raisons sous-jacentes à ces erreurs en fonction de la différence taxonomique et fonctionnelle des espèces. En effet, la tendance observée veut que la distance spectrale entre les espèces reflète leurs différences fonctionnelles ainsi que leur histoire évolutive (Schweiger et al, 2018 ; Frye et al, 2021). Bien que le lien entre les spectres et les traits foliaires ne soit plus à prouver, l'influence du niveau de différence fonctionnelle sur notre capacité à distinguer les espèces restait essentiellement à tester, sans compter que la variation intraspécifique a tendance à diminuer ces différences. De la même façon, plusieurs facteurs peuvent venir entraver le lien entre la relation évolutive des taxons et leur ressemblance écologique (Cadotte, et al., 2017). Il est donc impératif d'évaluer comment ces facteurs déterminent l'identification des espèces au niveau de la feuille, dans le but de valider leur lien avec la diversité spectrale et guider l'interprétation des résultats de classification par télédétection à plus grande échelle.

Afin de répondre à cette problématique, mon premier objectif était d'évaluer la sensibilité et la précision de la classification des espèces d'arbres du Québec à partir de leur réflectance spectrale. Mon second objectif portait quant à lui sur l'explication des cas de mauvaises identifications grâce aux distances effectives entre les espèces. Pour ce faire, j'ai mesuré les signatures spectrales et les traits fonctionnels de 45 espèces d'arbres, ainsi que reconstruit leurs relations évolutives en produisant une phylogénie basée sur des

données moléculaires. Par la suite, j'ai évalué la sensibilité et la précision de la classification à partir de la réflectance spectrale de toutes les espèces. Ensuite, j'ai examiné la relation entre les erreurs de classification et les distances phylogénétiques et fonctionnelles des espèces, en m'appuyant également sur le signal phylogénétique de la réflectance foliaire ainsi que sur la variation intraspécifique de chaque espèce.

Tout d'abord, les résultats obtenus démontrent que la réflectance foliaire permet d'identifier avec justesse un assemblage diversifié d'espèces provenant d'une région donnée. Cela est en accord avec les excellents résultats de classification obtenus avec un nombre plus restreint d'espèces, ainsi qu'à des niveaux d'organisation taxonomique supérieurs (Meireles et al, 2020). Les longueurs d'onde importantes pour la distinction de ces espèces se concentrent dans le visible et dans le *red edge*, autour des bandes d'absorption de l'eau, ainsi que dans la région de l'infrarouge court entre 1900 et 2400 nm qui présente une influence généralement élevée sur la différenciation des espèces. Toutefois, il faut noter que mon analyse de classification souffre d'un déséquilibre dans le nombre de spectres recueillis par espèce, ce qui semble avoir affecté notre capacité à identifier avec précision certaines d'entre elles moins bien représentées dans notre échantillonnage. De plus, il faut noter que malgré une bonne sensibilité dans la classification de plusieurs espèces, la précision du modèle n'est pas toujours excellente, c'est-à-dire que les prédictions peuvent être faussées par les erreurs de classification provenant d'autres espèces. Par exemple, le modèle identifie avec succès tous les spectres correspondant au chêne rouge, ce qui correspond à une sensibilité de 100 %, mais 11 espèces ont vu une partie de leurs spectres classifiés comme celui-ci, à une fréquence allant jusqu'à 34 %. Ainsi, seuls les deux tiers des prédictions de chêne rouge correspondent véritablement aux signatures spectrales de cette espèce, ce qui se solde par une précision beaucoup plus faible de 63 %. Cet effet a tendance à favoriser davantage les espèces déjà plus nombreuses dans notre échantillonnage, au détriment d'espèces plus rares. Ceci peut entraîner à plus large échelle une sous-estimation des espèces peu abondantes, particulièrement si elles ne présentent pas de caractéristiques foliaires assez distinctives ou encore si ces dernières sont trop variables (Cerrejón et al, 2021).

Les résultats montrent l'importance de la différence des traits fonctionnels foliaires pour la distinction des espèces selon leur signature spectrale. Il semble y avoir un seuil de distance fonctionnelle au-delà duquel les erreurs de classification deviennent improbables. Je montre aussi que les espèces proches se mélangent davantage entre elles, ce qui valide mon hypothèse et consolide le lien entre la diversité spectrale, la diversité fonctionnelle et la diversité taxonomique (Schweiger et al, 2018). Cette tendance à confondre les espèces proches est sûrement renforcée par la présence supposée d'individus issus de l'hybridation dans l'échantillonnage, dont les caractères foliaires intermédiaires ou plus variables semblent avoir affecté négativement l'identification. D'autres erreurs peuvent être attribuées à la ressemblance fonctionnelle des espèces. Par exemple, le caryer ovale (*Carya ovata* (Miller) K. Koch) partage des valeurs similairement élevées de lignine et de cellulose avec le frêne américain (*Fraxinus americana.*), pour lequel il se méprend à une fréquence d'environ 24 %. Certaines méprises, par contre, restent plus difficiles à expliquer. Par exemple, la fréquente méprise du cerisier tardif (*Prunus serotina*) pour le hêtre à grande feuille (*Fagus grandifolia*) ne semble pas découler d'une forte ressemblance pour un trait particulier. Il reste donc encore à évaluer les contributions relatives de chaque trait dans notre capacité à distinguer les espèces à partir de la réflectance foliaire. À cet effet, les modèles de transferts radiatifs comme PROSPECT (Jacquemoud et Baret, 1990) représentent des outils d'une grande utilité. Ces modèles se basent sur des principes physiques du transfert de la radiation, et permettent donc de moduler certains traits impliqués dans l'estimation du spectre foliaire tout en gardant d'autres constants, ce qui permet d'évaluer l'effet relatif de ceux-ci sur la différenciation des spectres (Jaquemoud et al, 2009 ; Féret et al, 2017 ; Kattenborn et al, 2019). Il est toutefois important de mentionner que le modèle décrivant l'impact des distances phylogénétiques et fonctionnelles se base sur les résultats de classification. Un échantillonnage plus important aurait peut-être permis une meilleure sensibilité du modèle pour certaines espèces, et par le fait même aurait réduit les erreurs sur lesquelles j'observe ces tendances.

Perspectives

Ces résultats restent particulièrement intéressants pour l'optimisation de modèles de classification à partir de la réflectance spectrale. Grâce au lien entre la distance phylogénétique et les erreurs de classification, il serait possible de calibrer les données de télédétection à large échelle, et potentiellement d'identifier des spectres dont l'espèce est préalablement inconnue, une proposition centrale de la spectranomique et nécessaire à l'estimation de la biodiversité par celle-ci (Cavender-Bares et al, 2017). À cette fin, la distinction des espèces au niveau de la feuille reste la fondation sur laquelle la faisabilité de la classification à grande échelle peut être entreprise, en fournissant les bases biologiques des différences spectrales (Jetz et al, 2016). La consolidation du lien entre la diversité spectrale et l'organisation taxonomique permettrait aussi, lorsque couplée à des données génétiques d'identifier et délimiter des complexes d'espèces très proches et leurs hybrides (Abasolo et al, 2013 ; Stasinski et al, 2021). La présence de signal phylogénétique dans les signatures spectrales, indiquant une variation des spectres en lien avec l'organisation taxonomique des espèces, constitue la base pour présumer la distinction des espèces grâce aux spectres. Cependant, la mesure du signal est sensible aux taxons et à l'identité des espèces impliquées, de même que l'échelle taxonomique considérée (Losos, 2008). En outre, le modèle d'évolution implicite dans la mesure de signal phylogénétique, soit un modèle de divergence stochastique, ne représente pas toujours le mode d'évolution réel des traits (Blomberg et al, 2019). Ainsi, l'étude de l'évolution des traits suivant différents processus évolutifs et de leur émergence dans la variabilité spectrale, tel que proposé par Meireles et al. (2020) constitue la prochaine étape dans la détermination de l'influence de la phylogénie sur la production de signatures spectrales reconnaissables ainsi que pour l'identification des convergences évolutives dans les traits foliaires.

Cette étape prend d'autant plus d'importance avec l'avènement de capteurs hyperspectraux destinés à être intégrés à des missions satellitaires, qui couvriront de larges territoires et un grand nombre d'espèces. Certains ont commencé leurs activités, par exemple HISUI, du Japon, lancé en 2019. D'autres sont attendus dans les prochaines années, comme HypIRI de la NASA, prévu en 2023, ou encore CHIME, de l'Union européenne, qui fait partie de la famille des satellites Sentinel. L'obtention prochaine d'un grand volume de données

hyperspectrales, récoltées non seulement sur de larges superficies, mais aussi de façon régulière, nécessite d'évaluer d'abord la qualité de la distinction entre les espèces dans des contextes régionaux (Hill et al., 2019 ; Pu, 2021). En effet, un des défis majeurs de la spectroscopie concerne le transfert du lien entre les spectres et les propriétés fonctionnelles des plantes à de plus grandes échelles spatiales. Cela provient du fait que ce lien peut se montrer variable selon le trait observé ou le type de végétation (Asner et Martin, 2008 ; Knyazikhin et al., 2013, Anderson, 2018 ; Rossi et al., 2021). Afin de remédier au compromis entre la résolution spatiale et la résolution spectrale ou temporelle lors de l'acquisition des données par satellite, certains proposent l'intégration de plusieurs sources de données récoltées à des échelles intermédiaires, par le biais de véhicules aériens téléguidés (UAV) ou d'avions (Alvarez-Vanhard, 2021). Une base de données spectrales telle que celle élaborée dans le cadre de ce projet constitue alors une ressource précieuse pour l'entraînement des algorithmes d'intelligence artificielle lors de la classification des espèces à moyenne échelle. Celles-ci permettent d'obtenir des interprétations fiables, particulièrement en conjonction avec les données fonctionnelles, nécessaires pour estimer l'état de la végétation, notamment en contexte agricole (Jiménez et Díaz-Delgado, 2015 ; Zhang et al., 2019). Une bonne compréhension de la variabilité spectrale des espèces et de leurs traits foliaires représente donc la fondation sur laquelle doit se construire notre analyse des données spectrales aux échelles supérieures (Gamon et al., 2019 ; Wang et Gamon, 2019), et cette étude y contribue.

En définitive, l'identification des espèces végétales par spectroscopie répond à de multiples besoins pour assurer le suivi de la biodiversité, autant pour cartographier les changements dans la distribution et l'abondance des espèces que pour évaluer leur réponse aux changements globaux par le biais de leurs traits fonctionnels. Mes résultats confirment le potentiel de la réflectance spectrale pour la classification des arbres à travers le Québec. De plus, je cible certaines espèces qui méritent une considération particulière dans l'élaboration et la généralisation de modèles de classification. La différenciation des épinettes et des chênes par exemple, nécessiterait davantage d'études en vue d'améliorer leur détection. Pour ce faire, la combinaison de la télédétection hyperspectrale avec d'autres types d'approches, telles que l'imagerie thermique ou par micro-ondes, reçoit de

plus en plus d'attention (Gerhards, 2019). L'intégration du LiDAR en particulier, une technologie laser qui renseigne sur la structure tridimensionnelle des végétaux, améliore la qualité de la classification des arbres à l'échelle de la canopée (Michałowska et al., 2021 ; Pu, 2021). Finalement, le développement rapide de méthodes alternatives de classification dans les dernières années, par exemple les réseaux neuronaux convolutifs, développés pour la classification d'images, pourrait permettre de mieux quantifier la séparabilité des spectres grâce à leur capacité d'extraire des paramètres distinctifs extrêmement précis. Leur application aux données spectrales foliaires permettrait entre autres d'estimer le contenu en chlorophylle et en xanthophylle en dépit du chevauchement de leur valeur d'absorption dans le spectre (Shi et al., 2022). Considérant cela, le potentiel de la télédétection pour le suivi de la biodiversité végétale ne fait aucun doute. Il est cependant nécessaire de comprendre les mécanismes qui gouvernent notre capacité de distinguer les espèces au niveau de la feuille, ainsi que ceux qui ont le potentiel de biaiser nos estimations, afin d'obtenir des relevés fiables et généralisables à plus grande échelle.

Références bibliographiques

- Ayotte, J. Guilbault-Meyers, X., Laliberté, E. (2019). Measuring CN content in leaf samples using Elementar Vario MICRO Cube V.1. protocols.io
- Ayotte, J., Laliberté, E. (2019). Measuring leaf carbon fractions with the ANKOM2000 Fiber Analyzer. protocols.io
- Ahern, F. J. (1988). The effects of bark beetle stress on the foliar spectral reflectance of lodgepole pine. *International Journal of Remote Sensing*, **9**(9), 1451-1468.
- Abdullah, H., Darvishzadeh, R., Skidmore, A. K., Groen, T. A., & Heurich, M. (2018). European spruce bark beetle (*Ips typographus*, L.) green attack affects foliar reflectance and biochemical properties. *International journal of applied earth observation and geoinformation*, **64**, 199-209.
- Alkama, R. et Cescatti, A. (2016). Biophysical climate impacts of recent changes in global forest cover. *Science*, **351**(6273), 600-604.
- Alvarez-Vanhard, E., Corpetti, T. et Houet, T. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing*, **3**(January), 100019.
- Angst, G., Mueller, K. E., Eissenstat, D. M., Trumbore, S., Freeman, K. H., Hobbie, S. E., ... Mueller, C. W. (2019). Soil organic carbon stability in forests: Distinct effects of tree species identity and traits. *Global Change Biology*, **25**(4), 1529-1546.
- Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, **64**(3), 234-253.
- Asner, G. P. et Martin, R. E. (2016a). Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. *Global Ecology and Conservation*, **8**, 212-219.
- Asner, G. P. et Martin, R. E. (2016b). Convergent elevation trends in canopy chemical traits of tropical forests. *Global Change Biology*, **22**(6), 2216-2227.
- Asner, G. P. et Martin, R. E. (2011). Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian forest. *New Phytologist*, **189**(4), 999-1012

- Asner, G. P. et Martin, R. E. (2009). Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, **7**(5), 269-276
- Asner, G. P. et Martin, R. E. (2008). Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing of Environment*, **112**(10), 3958-3970.
- Asner, G. P., Martin, R. E., Carranza-Jiménez, L., Sinca, F., Tupayachi, R., Anderson, C. B. et Martinez, P. (2014). Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. *New Phytologist*, **204**(1), 127-139.
- Asner, G. P., Martin, R. E., Knapp, D. E., Tupayachi, R., Anderson, C. B., Sinca, F., ... Llactayo, W. (2017). Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science*, **355**(6323), 385-389.
- Barker, M. et Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, **17**(3), 166-173.
- Beauchamp-Rioux, R., Schweiger, A.K, Kalacska, M., Laliberté, E. (2022). Leaf spectra distinguish temperate tree species despite environmental and phenological variation [Document en préparation], Département de sciences biologiques, Université de Montréal.
- Billings, W. D. et Morris, R. J. (1951). Reflection of Visible and Infrared Radiation from Leaves of Different Ecological Groups. *American Journal of Botany*, **38**(5), 327.
- Blomberg, S. P., Garland, T. et Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, **57**(4), 717-745.
- Blomberg, S. P., Rathnayake, S. I. et Moreau, C. M. (2019). Beyond brownian motion and the ornstein-uhlenbeck process: Stochastic diffusion models for the evolution of quantitative characters. *American Naturalist*, **195**(2).
- Brouillet, L., F. Coursol, S.J. Meades, M. Favreau, M. Anions, P. Bélisle & P. Desmet. 2010+. VASCAN, the Database of Vascular Plants of Canada. <http://data.canadensys.net/vascan/>(consulted on 2022-04-25)
- Cadotte, M. W., Carboni, M., Si, X. et Tatsumi, S. (2019). Do traits and phylogeny support congruent community diversity patterns and assembly inferences? *Journal of Ecology*, **107**(5), 2065-2077.

- Cadotte, M. W., Davies, T. J. et Peres-Neto, P. R. (2017). Why phylogenies do not always predict ecological differences. *Ecological Monographs*, **87**(4), 535-551.
- Castro-Esau, K. L., Sánchez-Azofeifa, G. A., Rivard, B., Wright, S. J. et Quesada, M. (2006). Variability in leaf optical properties of mesoamerican trees and the potential for species classification. *American Journal of Botany*, **93**(4), 517-530.
- Cavender-Bares, J., Gamon, J. A., Hobbie, S. E., Madritch, M. D., Meireles, J. E., Schweiger, A. K. et Townsend, P. A. (2017). Harnessing plant spectra to integrate the biodiversity sciences across biological and spatial scales. *American Journal of Botany*, **104**(7), 966-969.
- Cavender-Bares, J., Meireles, J. E., Couture, J. J., Kaproth, M. A., Kingdon, C. C., Singh, A., ... Townsend, P. A. (2016). Associations of leaf spectra with genetic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity. *Remote Sensing*, **8**(3), 1-17.
- Cerrejón, C., Valeria, O., Marchand, P., Caners, R. T. et Fenton, N. J. (2021). No place to hide: Rare plant detection through remote sensing. *Diversity and Distributions*, **27**(6), 948-961.
- Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., ... Díaz, S. (2000). Consequences of changing biodiversity. *Nature*, **405**(6783), 234-242.
- Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., ... Weber, A. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, **181**(1), 1-20.
- Chevallier, S., Bertrand, D., Kohler, A. et Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, **20**(5), 221-229.
- Clark, M. L., Roberts, D. A. et Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, **96**(3-4), 375-398.
- Cornwell, W. K., Pearse, W. D., Dalrymple, R. L. et Zanne, A. E. (2019). What we (don't) know about global plant diversity. *Ecography*, **42**(11), 1819-1831.

- Cui, H. F., Ye, Z. H., Xu, L., Fu, X. S., Fan, C. W., & Yu, X. P. (2012). Automatic and rapid discrimination of cotton genotypes by near infrared spectroscopy and chemometrics. *Journal of Analytical Methods in Chemistry*, 2012, 1-
- Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D. et Næsset, E. (2013). Tree species classification in boreal forests with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, **51**(5), 2632-2645.
- De Lafontaine, G., Prunier, J., Gérardi, S. et Bousquet, J. (2015). Tracking the progression of speciation: Variable patterns of introgression across the genome provide insights on the species delimitation between progenitor-derivative spruces (*Picea mariana* × *P. rubens*). *Molecular Ecology*, **24**(20), 5229-5247.
- De Tomás Marín, S., Novák, M., Klančnik, K. et Gaberščik, A. (2016). Spectral signatures of conifer needles mainly depend on their physical traits. *Polish Journal of Ecology*, **64**(1), 1-13.
- Díaz, S. et Cabido, M. (2001). Vive la différence: Plant functional diversity matters to ecosystem processes. *Trends in Ecology and Evolution*, **16**(11), 646-655.
- Diniz, É. S., Amaral, C. H., Sardinha, S. T., Thiele, J. et Aleves Meira-Neto, J. A. (2020). Phylogenetic signatures in reflected foliar spectra of regenerating plants in Neotropical forest gaps. *Remote Sensing of Environment*, (May).
- Dumais, D. et Prévost, M. (2007). Management for red spruce conservation in Québec: The importance of some physiological and ecological characteristics - A review. *Forestry Chronicle*, **83**(3), 378-392.
- Durgante, F. M., Higuchi, N., Almeida, A. et 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.
- Ellison, A. M., Bank, M. S., Clinton, B. D., Colburn, E. A., Elliott, K., Ford, C. R., ... Webster, J. R. (2005). Loss of foundation species: Consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment*, **3**(9), 479-486.
- Elmqvist, T., Folke, C., Nyström, M., Peterson, G., Bengtsson, J., Walker, B. et Norberg, J. (2003). Response diversity, ecosystem change, and resilience. *Frontiers in Ecology and the Environment*, **1**(9), 488-494.

- Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., ... Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, **186**, 64-87.
- Frye, H. A., Aiello-Lammens, M. E., Euston-Brown, D., Jones, C. S., Kilroy, H., Cory, M., ... Silander, J. A. (2021). Plant spectral diversity as a surrogate for species, functional and phylogenetic diversity across a hyper- diverse biogeographic region. *Global Ecology and Biogeography*, (March), 1-15.
- Gamon, J. A., Somers, B., Malenovský, Z., Middleton, E. M., Rascher, U. et Schaepman, M. E. (2019). Assessing Vegetation Function with Imaging Spectroscopy. *Surveys in Geophysics*, **40**(3), 489-513.
- Gates, D. M., Keegan, H. J., Schleter, J. C. et Weidner, V. R. (1965). Spectral Properties of Plants. *Applied Optics*, **4**(1), 11.
- Gerhards, M., Schlerf, M., Mallick, K. et Udelhoven, T. (2019). Challenges and future perspectives of multi-/Hyperspectral thermal infrared remote sensing for crop water-stress detection: A review. *Remote Sensing*, **11**(10).
- Girard, A., Schweiger, A. K., Carteron, A., Kalacska, M. et Lalibert, E. (2020). Foliar Spectra and Traits of Bog Plants across Nitrogen Deposition Gradients. *Remote Sensing*, **12**(2448).
- Grimm GW, Denk T, Hemleben V. 2007. Evolutionary history and systematics of Acer section Acer—A case study of low-level phylogenetics. *Plant Systematics and Evolution* 267: 215–253.
- Hawkins, B. A., Rueda, M., Rangel, T. F., Field, R. et Diniz-Filho, J. A. F. (2014). Community phylogenetics at the biogeographical scale: Cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography*, **41**(1), 23-38.
- Homolová, L., Malenovský, Z., Clevers, J. G. P. W., García-Santos, G. et Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity*, **15**, 1-16.
- Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., ... Connor, M. I. O. (2012). a global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, **486**(7401), 105-108.

- Hovi, A., Raitio, P. et Rautiainen, M. (2017). A spectral analysis of 25 boreal tree species. *Silva Fennica*, **51**(4), 1-16.
- Hycza, T., Stereńczak, K. et Bałazy, R. (2018). Potential use of hyperspectral data to classify forest tree species. *New Zealand Journal of Forestry Science*, **48**(1).
- Jackman, S., Tahk, A., Zeileis, A., Maimone, C., Fearon, J. et Meers, Z. (2020). pscl: Political Science Computational Laboratory, R package version 1.5.5, 100.
- Jacquemoud, S. et Ustin, S. (2019). *Leaf Optical Properties in Different Wavelength Domains Chapter 5. Leaf Optical Properties*.
- Jiménez, M. et Díaz-Delgado, R. (2015). Towards a standard plant species spectral library protocol for vegetation mapping: A case study in the shrubland of Doñana National Park. *ISPRS International Journal of Geo-Information*, **4**(4), 2472-2495.
- Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F. W., Asner, G. P., ... Ustin, S. L. (2016). Monitoring plant functional diversity from space. *Nature Plants*, **2**(3), 16024.
- Kattenborn, T., Fassnacht, F. E. et Schmidlein, S. (2019). Differentiating plant functional types using reflectance: which traits make the difference? *Remote Sensing in Ecology and Conservation*, **5**(1), 5-19.
- Kerr, J. T. et Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology and Evolution*, **18**(6), 299-305.
- Kim, S. W., Ban, S. H., Chung, H., Cho, S., Chung, H. J., Choi, P. S., ... Liu, J. R. (2004). Taxonomic discrimination of flowering plants by multivariate analysis of Fourier transform infrared spectroscopy data. *Plant Cell Reports*, **23**(4), 246-250.
- Kothari, S. A. (2021). Reflectance spectroscopy allows rapid , accurate , and non-destructive estimates of functional traits from pressed leaves.
- Kothari, S. et Schweiger, A. K. (2022). Plant spectra as integrative measures of plant phenotypes. *EcoEvoRxiv*, <https://ecoevorxiv.org/bfc5t/>
- Laliberté, E., Schweiger, A. K. et Legendre, P. (2020). Partitioning plant spectral diversity into alpha and beta components. *Ecology Letters*, **23**(2), 370-380.

- Laliberté, E. & Soffer, R. (2018a). Measuring spectral reflectance and transmittance (350-2500 nm) of large leaves using the Spectra Vista Corporation (SVC) DC-R/T Integrating Sphere. *protocols.io*.
- Laliberté, E. & Soffer, R. (2018b). Measuring spectral reflectance and transmittance (350-2500 nm) of small and/or narrow leaves using the Spectra Vista Corporation (SVC) DC-R/T Integrating Sphere. *protocols.io*.
- Laliberté, E. (2018). Measuring specific leaf area and water content V.1 *protocols.io*
- Leckie, D. G., Tinis, S., Nelson, T., Burnett, C., Gougeon, F. A., Cloney, E. et Paradine, D. (2005). Issues in species classification of trees in old growth conifer stands. *Canadian Journal of Remote Sensing*, **31**(2), 175-190.
- Li, J., Stukel, M., Bussies, P., Skinner, K., Lemmon, A. R., Lemmon, E. M., ... Swenson, N. G. (2019). Maple phylogeny and biogeography inferred from phylogenomic data. *Journal of Systematics and Evolution*.
- Lindenmayer, D. B. et Laurance, W. F. (2017). The ecology, distribution, conservation and management of large old trees. *Biological Reviews*, **92**(3), 1434-1458.
- Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**(10), 995-1003
- Madritch, M. D., Kingdon, C. C., Singh, A., Mock, K. E., Lindroth, R. L. et Townsend, P. A. (2014). Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1643).
- McManus, K. M., Asner, G. P., Martin, R. E., Dexter, K. G., Kress, W. J. et Field, C. B. (2016). Phylogenetic structure of foliar spectral traits in tropical forest canopies. *Remote Sensing*, **8**(3), 1-16.
- Meireles, J. E., Cavender-Bares, J., Townsend, P. A., Ustin, S., Gamon, J. A., Schweiger, A. K., ... O'Meara, B. C. (2020). Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytologist*, **228**(2), 485-493.

- Meireles JE, Schweiger A, Cavender-Bares J (2017). *_spectrolab*: Class and Methods for Hyperspectral Data in R. R package version 0.0.9, URL: <https://CRAN.R-project.org/package=spectrolab>.
- MFFP, (2022). Forêts, <https://mffp.gouv.qc.ca/les-forets/>
- Michałowska, M. et Rapiński, J. (2021). A review of tree species classification based on airborne lidar data and applied classifiers. *Remote Sensing*, **13**(3), 1-27.
- Millar, C. I. et Stephenson, N. L. (2015). Temperate forest health in an era of emerging megadisturbance. *Science*, **349**(6250), 823-826.
- Modzelewska, A., Fassnacht, F. E. et Stereńczak, K. (2020). Tree species identification within an extensive forest area with diverse management regimes using airborne hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, **84**(August 2019), 101960.
- Moreau, G., Achim, A. et Pothier, D. (2020). An accumulation of climatic stress events has led to years of reduced growth for sugar maple in southern Quebec, Canada. *Ecosphere*, **11**(7).
- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K. et Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, **3**(4), 743-756.
- Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, **189**, 375-394.
- Paiva, D. N. A., Perdiz, R. de O. et Almeida, T. E. (2021). Using near-infrared spectroscopy to discriminate closely related species: a case study of neotropical ferns. *Journal of Plant Research*, **134**(3), 509-520.
- Paquette, A., Joly, S. et Messier, C. (2015). Explaining forest productivity using tree functional traits and phylogenetic information: Two sides of the same coin over evolutionary scale? *Ecology and Evolution*, **5**(9), 1774-1783.
- Paradis E. & Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., ... Dulloo, E. (2013). Essential Biodiversity Variables. *Science*, **339**(6117), 277-278.

- Pu, R. (2021). Mapping Tree Species Using Advanced Remote Sensing Technologies: A State-of-the-Art Review and Perspective. *Journal of Remote Sensing*, **2021**, 1-26.
- Pu, R. et Liu, D. (2011). Segmented canonical discriminant analysis of in situ hyperspectral data for identifying 13 urban tree species. *International Journal of Remote Sensing*, **32**(8), 2207-2226.
- Ramoelo, A., Dzikiti, S., Van Deventer, H., Maherry, A., Cho, M. A. et Gush, M. (2015). Potential to monitor plant stress using remote sensing tools. *Journal of Arid Environments*, **113**, 134-144.
- Ran, J. H., Shen, T. T., Wu, H., Gong, X. et Wang, X. Q. (2018). Phylogeny and evolutionary history of Pinaceae updated by transcriptomic analysis. *Molecular Phylogenetics and Evolution*, **129**(August), 106-116.
- Reich, P. B. (2014). The world-wide « fast-slow » plant economics spectrum: A traits manifesto. *Journal of Ecology*, **102**(2), 275-301.
- Reich, P. B., Wright, I. J., Cavender-Bares, J., Craine, J. M., Oleksyn, J., Westoby, M. et Walters, M. B. (2003). The evolution of plant functional variation: Traits, spectra, and strategies. *International Journal of Plant Sciences*, **164**(SUPPL. 3).
- Richter, R., Reu, B., Wirth, C., Doktor, D. et Vohland, M. (2016). The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. *International Journal of Applied Earth Observation and Geoinformation*, **52**, 464-474.
- Rossi, C., Kneubühler, M., Schütz, M., Schaeppman, M. E., Haller, R. M. et Risch, A. C. (2021). Spatial resolution, spectral metrics and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands. *Remote Sensing in Ecology and Conservation*, 1-18.
- Scheffers, B. R., De Meester, L., Bridge, T. C. L., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. T., ... Watson, J. E. M. (2016). The broad footprint of climate change from genes to biomes to people. *Science*, **354**(6313).
- Schweiger, A.K. (2020). Spectral field campaigns: planning and data collection. In: Cavender-Bares, J., Gamon, J.A., Townsend, P.A. (eds) *Remote Sensing of Plant Biodiversity*. Springer, Cham. https://doi.org/10.1007/978-3-030-33157-3_15

- Schweiger, A. K., Cavender-Bares, J., Townsend, P. A., Hobbie, S. E., Madritch, M. D., Wang, R., ... Gamon, J. A. (2018). Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nature Ecology and Evolution*, **2**(6), 976-982.
- Schweiger, A. K., Schütz, M., Risch, A. C., Kneubühler, M., Haller, R. et Schaepman, M. E. (2017). How to predict plant functional types using imaging spectroscopy: linking vegetation community traits, plant functional types and spectral response. *Methods in Ecology and Evolution*, **8**(1), 86-95.
- Shi, Y., Skidmore, A. K., Wang, T., Holzwarth, S., Heiden, U., Pinnel, N., ... Heurich, M. (2018). Tree species classification using plant functional traits from LiDAR and hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, **73**(May), 207-219.
- Signal Developers (2013). signal: Signal processing. <http://r-forge.r-project.org/projects/signal/>.
- Slaton, M. R., Hunt, E. R. et Smith, W. K. (2001). Estimating near-infrared leaf reflectance from leaf structural characteristics. *American Journal of Botany*, **88**(2), 278-284.
- Stasinski, L., White, D. M., Nelson, P. R., Ree, R. H. et Meireles, J. E. (2021). *Reading light: Leaf spectra capture fine-scale diversity of closely related, hybridizing arctic shrubs*. *New Phytologist*.
- Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M. et Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. *Science*, **277**(5330), 1300-1302.
- Townsend, P. A., Serbin, S. P., Kruger, E. L. et Gamon, J. A. (2013). Disentangling the contribution of biological and physical properties of leaves and canopies in imaging spectroscopy data. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(12), 2013.
- Ustin, S. L. et Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*, **186**(4), 795-816.

- Véron, S., Saito, V., Padilla-García, N., Forest, F. et Bertheau, Y. (2019). The Use of Phylogenetic Diversity in Conservation Biology and Community Ecology: A Common Base but Different Approaches. *The Quarterly Review of Biology*, **94**(2), 123-148.
- Wang, R. et Gamon, J. A. (2019). Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, **231**(June), 111218.
- Wang, R., Gamon, J. A., Schweiger, A. K., Cavender-Bares, J., Townsend, P. A., Zyguelbaum, A. I. et Kothari, S. (2018). Influence of species richness, evenness, and composition on optical diversity: A simulation study. *Remote Sensing of Environment*, **211**(March), 218-228.
- Watson, J. E. M., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., ... Lindenmayer, D. (2018). The exceptional value of intact forest ecosystems. *Nature Ecology and Evolution*, **2**(4), 599-610.
- Wright, I. J., Westoby, M., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., ... Cornelissen, J. H. . (2004). The worldwide leaf economics spectrum. *Nature*, **428**(6985), 821–827.
- Wu, J., Chavana-Bryant, C., Prohaska, N., Serbin, S. P., Guan, K., Albert, L. P., ... Saleska, S. R. (2017). Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. *New Phytologist*, **214**(3), 1033-1048.
- Yang, X., Tang, J., Mustard, J. F., Wu, J., Zhao, K., Serbin, S. et Lee, J. E. (2016). Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. *Remote Sensing of Environment*, **179**, 1-12.
- Youngentob, K. N., Zdenek, C., & Gorsel, E. (2016). A simple and effective method to collect leaves and seeds from tall trees. *Methods in Ecology and Evolution*, **7**(9), 1119-1123.
- Zagajewski, B., Tømmervik, H., Bjerke, J. W., Raczko, E., Bochenek, Z., Klos, A., ... Ziólkowski, D. (2017). Intraspecific differences in spectral reflectance curves as indicators of reduced vitality in high-arctic plants. *Remote Sensing*, **9**(12), 1-19.

- Zhang, T., Niinemets, Ü., Sheffield, J. et Lichstein, J. W. (2018). Shifts in tree functional composition amplify the response of forest biomass to climate. *Nature*, **556**(7699), 99-102.
- Zhang, J., He, Y., Yuan, L., Liu, P., Zhou, X. et Huang, Y. (2019). Machine learning-based spectral library for crop classification and status monitoring. *Agronomy*, **9**(9).
- Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R* (Vol. 574). New York: Springer.

ANNEXES

Tableau Supplémentaire S1 Additional species taxonomic and sampling information

Alphabetical list of species studied with vernacular (French and English) names (taken from VASCAN; <https://data.canadensys.net/vascan/>), species code and plant family, number of sampling sites for each species, total number of trees sampled, subset of trees used to avoid heavily unbalanced classes in the PLS-DA and corresponding number of foliar spectra.

Species	Vernacular (français)	Vernacular	Code	Taxonomic family	N° of sites	N° of trees sampled	N° of trees for PLS-DA	N° of leaves for PLS-DA
<i>Abies balsamea</i> (Linnaeus) Miller	Sapin baumier	Balsam fir	ABBA	Pinaceae	3	19	19	114
<i>Acer negundo</i> Linnaeus	Érable à Gigère	Manitoba maple	ACNE	Sapindaceae	3	12	12	72
<i>Acer nigrum</i> F. Michaux	Érable noir	Black maple	ACNI	Sapindaceae	2	10	10	60
<i>Acer pensylvanicum</i> Linnaeus	Érable de Pennsylvanie	Striped maple	ACPE	Sapindaceae	3	20	20	120
<i>Acer platanoides</i> Linnaeus	Érable de Norvège	Norway maple	ACPL	Sapindaceae	1	10	10	60
<i>Acer rubrum</i> Linnaeus	Érable rouge	Red maple	ACRU	Sapindaceae	5	58	18	108
<i>Acer saccharinum</i> Linnaeus	Érable argenté	Silver maple	ACSI	Sapindaceae	4	30	20	120
<i>Acer saccharum</i> Marshall	Érable à sucre	Sugar maple	ACSA	Sapindaceae	4	30	18	108
<i>Acer spicatum</i> Lamarck	Érable à épis	Mountain maple	ACSP	Sapindaceae	2	8	8	48
<i>Aesculus hippocastanum</i> Linnaeus	Marronnier d'Inde	Common horsechestnut	AEHI	Sapindaceae	1	3	0	0
<i>Betula alleghaniensis</i> Britton	Bouleau jaune	Yellow birch	BEAL	Betulaceae	4	24	24	144
<i>Betula papyrifera</i> Marshall	Bouleau blanc (à papier)	Paper birch	BEPA	Betulaceae	3	32	18	108
<i>Betula populifolia</i> Marshall	Bouleau gris	Grey birch	BEPO	Betulaceae	5	60	18	108
<i>Carpinus caroliniana</i> Walter	Charme de Caroline	Blue-beech	CACA	Betulaceae	2	11	11	66

<i>Carya cordiformis</i> (Wangenheim) K. Koch	Caryer cordiforme	Bitternut hickory	CACO	Juglandaceae	4	13	13	78
<i>Carya ovata</i> (Miller) K. Koch	Caryer ovale	Shagbark hickory	CAOV	Juglandaceae	2	11	11	66
<i>Celtis occidentalis</i> Linnaeus	Micocoulier occidental	Common hackberry	CEOC	Cannabaceae	3	10	10	60
<i>Fagus grandifolia</i> Ehrhart	Hêtre à grande feuille	American beech	FAGR	Fagaceae	5	42	19	114
<i>Fraxinus americana</i> Linnaeus	Frêne blanc	White ash	FRAM	Oleaceae	5	18	18	108
<i>Fraxinus nigra</i> Marshall	Frêne noir	Black ash	FRNI	Oleaceae	3	21	21	126
<i>Fraxinus pennsylvanica</i> Marshall	Frêne rouge	Red ash	FRPE	Oleaceae	1	6	6	36
<i>Juglans cinerea</i> Linnaeus	Noyer cendré	Butternut	JUCI	Juglandaceae	4	13	13	78
<i>Juglans nigra</i> Linnaeus	Noyer noir	Black walnut	JUNI	Juglandaceae	1	3	0	0
<i>Larix laricina</i> (Du Roi) K. Koch	Mélèze larcin	Tamarack	LALA	Pinaceae	2	11	11	66
<i>Ostrya virginiana</i> (Miller) K. Koch	Ostryer de Virginie	Eastern hop-hornbeam	OSVI	Betulaceae	3	15	15	90
<i>Picea abies</i> (Linnaeus) H. Karsten	Épinette de Norvège	Norway spruce	PIAB	Pinaceae	1	5	5	30
<i>Picea glauca</i> (Moench) Voss	Épinette blanche	White spruce	PIGL	Pinaceae	1	10	10	60
<i>Picea mariana</i> (Miller) Britton. Sterns & Poggenburgh	Épinette noire	Black spruce	PIMA	Pinaceae	2	7	7	42
<i>Picea rubens</i> Sargent	Épinette rouge	Red spruce	PIRU	Pinaceae	2	12	12	72
<i>Pinus banksiana</i> Lambert	Pin gris	Jack pine	PIBA	Pinaceae	1	1	0	0
<i>Pinus resinosa</i> Aiton	Pin rouge	Red pine	PIRE	Pinaceae	2	12	12	72
<i>Pinus rigida</i> P. Miller	Pin rigide	Pitch pine	PIRI	Pinaceae	1	10	10	60
<i>Pinus strobus</i> Linnaeus	Pin blanc	White pine	PIST	Pinaceae	3	12	12	72
<i>Populus balsamifera</i> Linnaeus	Peuplier baumier	Balsam poplar	POBA	Salicaceae	2	13	13	78

<i>Populus deltoides</i> W. Bartram ex Marshall	Peuplier deltoïde	Eastern cottonwood	PODE	Salicaceae	3	13	13	78
<i>Populus grandidentata</i> Michaux	Peuplier à grande dents	Large-toothed aspen	POGR	Salicaceae	3	30	18	108
<i>Populus tremuloides</i> Michaux	Peuplier faux-tremble	Trembling aspen	POTR	Salicaceae	6	60	18	108
<i>Prunus nigra</i> Aiton	Cerisier noir	Canada plum	PRNI	Rosaceae	1	1	0	0
<i>Prunus pensylvanica</i> Linnaeus f.	Cerisier de Pennsylvanie	Pin cherry	PRPE	Rosaceae	2	11	11	66
<i>Prunus serotina</i> Ehrhart	Cerisier tardif	Black cherry	PRSE	Rosaceae	3	12	12	72
<i>Quercus alba</i> Linnaeus	Chêne blanc	White oak	QUAL	Fagaceae	1	10	10	60
<i>Quercus bicolor</i> Willdenow	Chêne bicolor	Swamp white oak	QUBI	Fagaceae	2	4	4	24
<i>Quercus macrocarpa</i> Michaux	Chêne à gros fruit	Bur oak	QUMA	Fagaceae	4	8	8	48
<i>Quercus rubra</i> Linnaeus	Chêne rouge	Red oak	QURU	Fagaceae	2	21	21	126
<i>Quercus sp.</i> Linnaeus	Chêne spp.	Oak spp.	QUSP	Fagaceae	2	7	0	0
<i>Sorbus sp</i> Linnaeus	Sorbier sp.	Mountain ash spp.	SOSP	Rosaceae	3	12	12	72
<i>Thuja occidentalis</i> Linnaeus	Thuja occidental	Eastern white cedar	THOC	Cupressaceae	2	13	13	78
<i>Tilia americana</i> Linnaeus	Tilleul d'Amérique	Basswood	TIAM	Malvaceae	3	13	13	78
<i>Tsuga canadensis</i> (Linnaeus) Carrière	Pruche du Canada	Eastern hemlock	TSCA	Pinaceae	2	14	14	84
<i>Ulmus sp</i> Linnaeus	Orme spp.	Elm spp.	ULSP	Ulmaceae	4	18	18	108
Total						819	609	3654

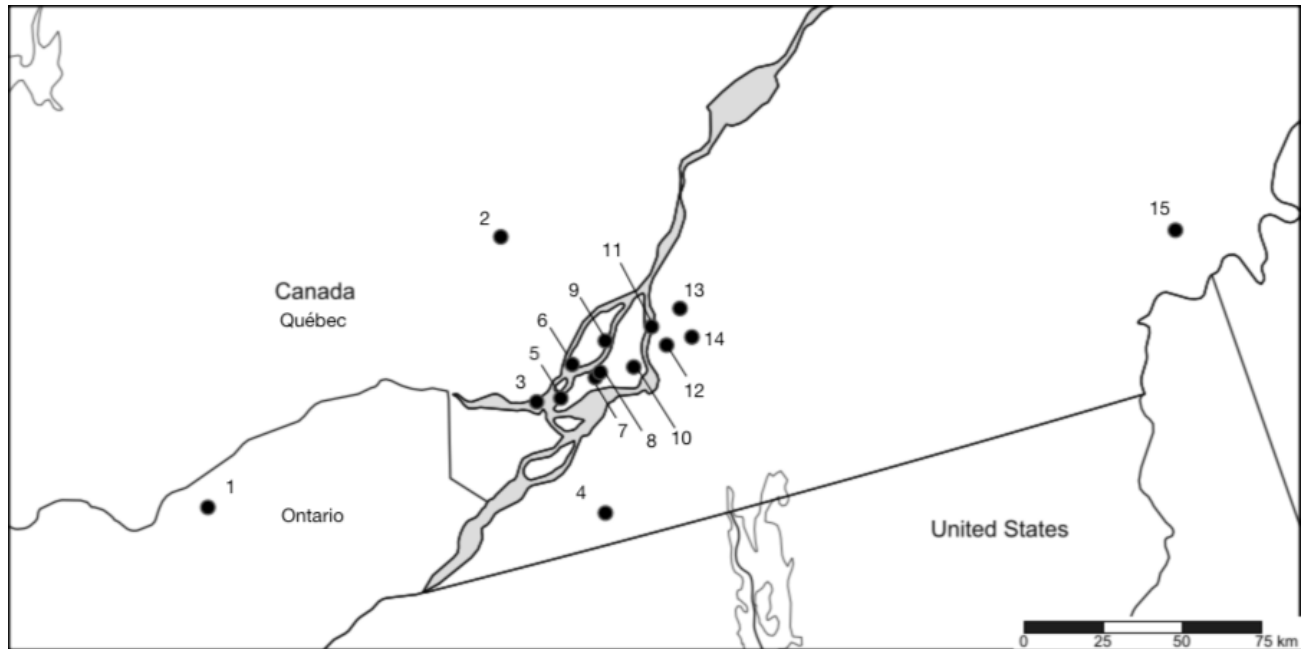


Figure Supplémentaire S1 Map of sampling sites

1- Mer Bleue Peatland ; 2- Station de biologie des Laurentides ; 3- Parc national d'Oka ; 4- Réserve écologique du Pin rigide ; 5- Parc-nature du Cap St-Jacques ; 6- Bois de la Source ; 7- Parc-nature du Bois-de-Liesse ; 8- Parc-nature du Bois de Saraguay ; 9- Bois Papineau ; 10- Parc du Mont-Royal ; 11- Parc national des îles de Boucherville ; 12- Boisé du Tremblay ; 13- Parc des étangs Antoine-Charlebois ; Parc national du Mont Saint-Bruno ; 15- Parc national du Mont-Mégantic.

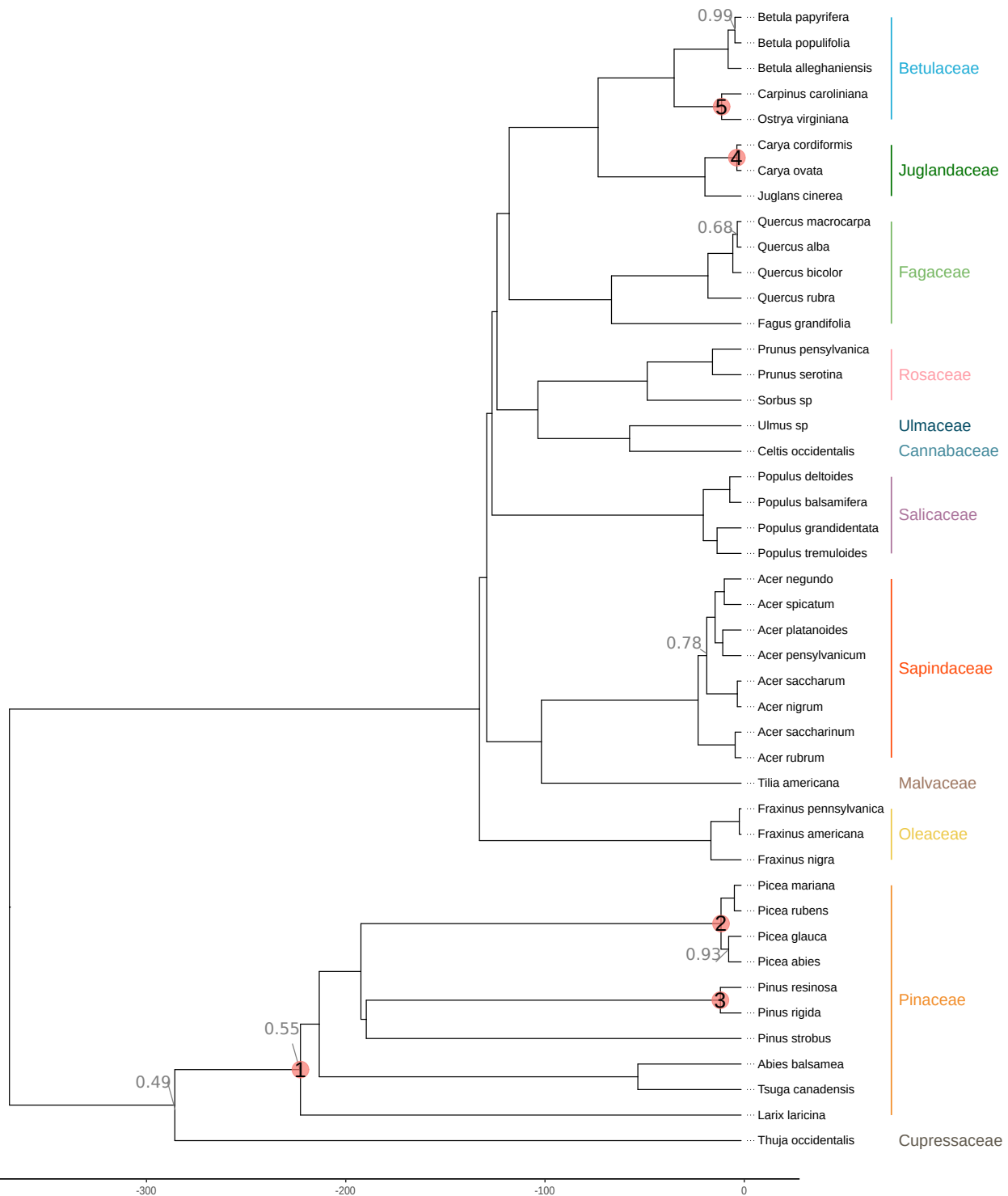


Figure Supplémentaire S2 Phylogeny of the 45 species used for classification

Posterior probabilities less than one are indicated in grey at the corresponding nodes. The most distinct clades in terms of spectral reflectance are indicated by red dots with the corresponding rank. The scale is in million years.

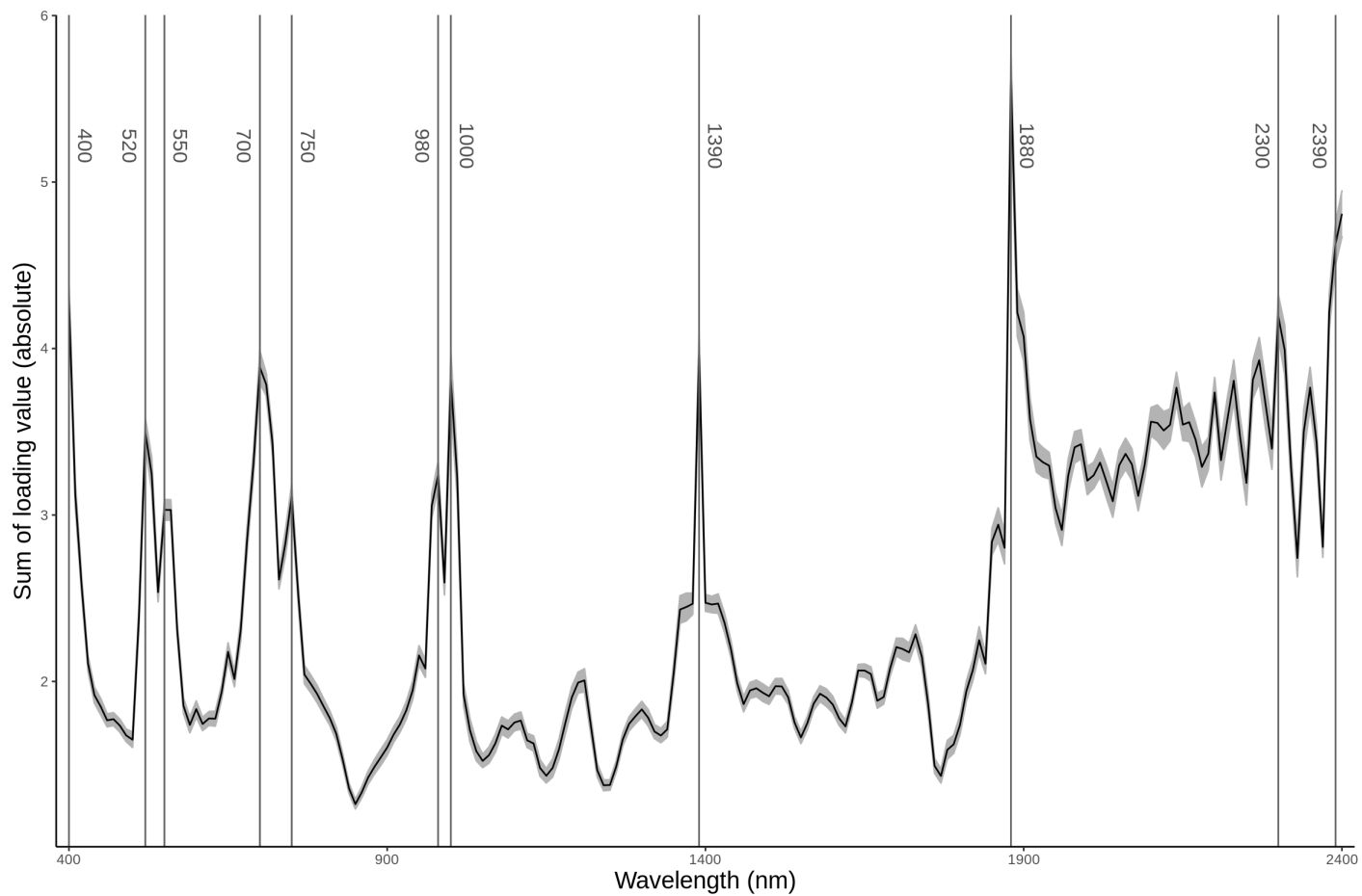
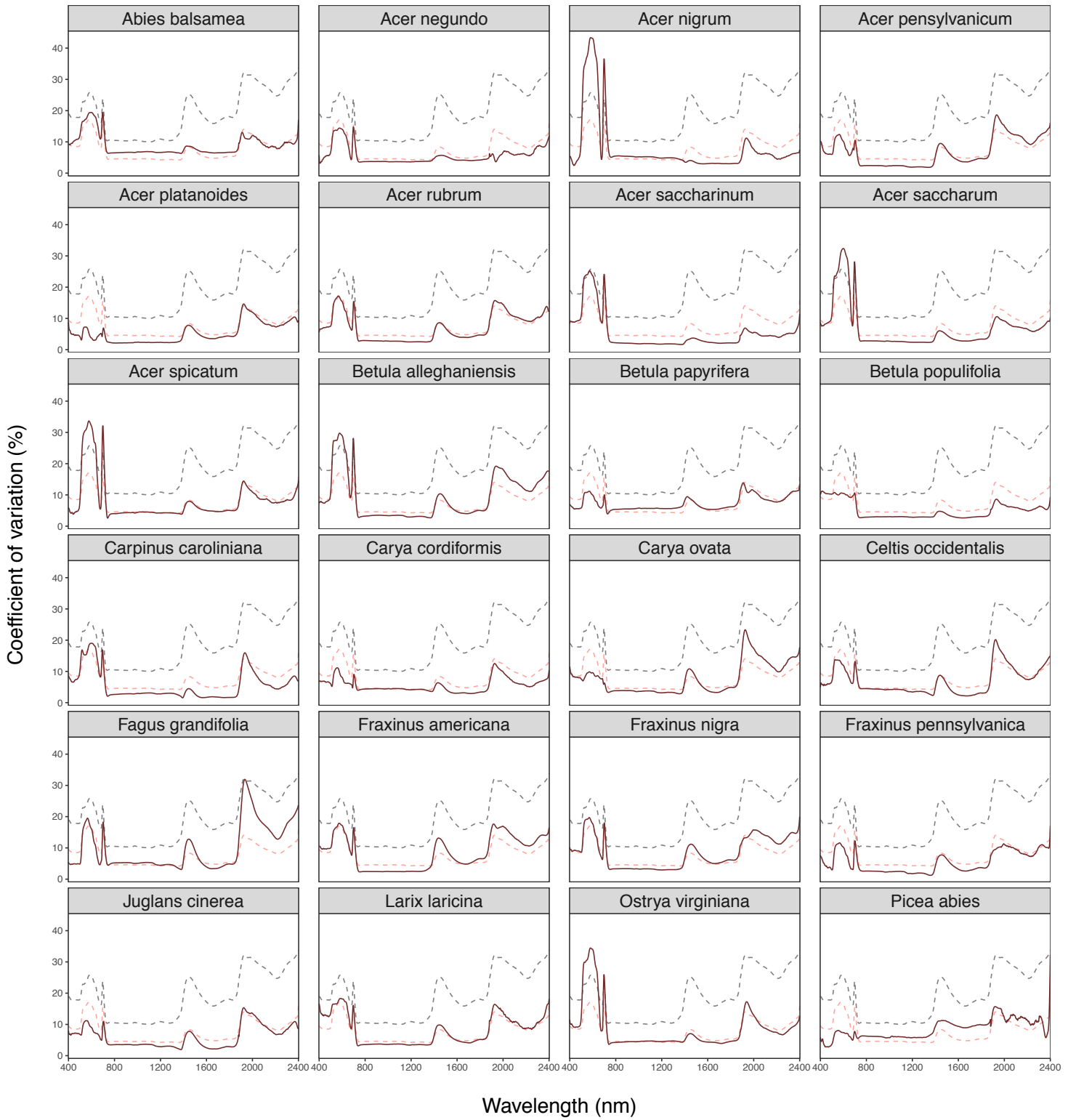


Figure Supplémentaire S3 Wavelength loading from PLS-DA model
 Local maxima are indicated by vertical lines and corresponding band number

-- Interspecific - - Intraspecific — Species



-- Interspecific - - Intraspecific — Species

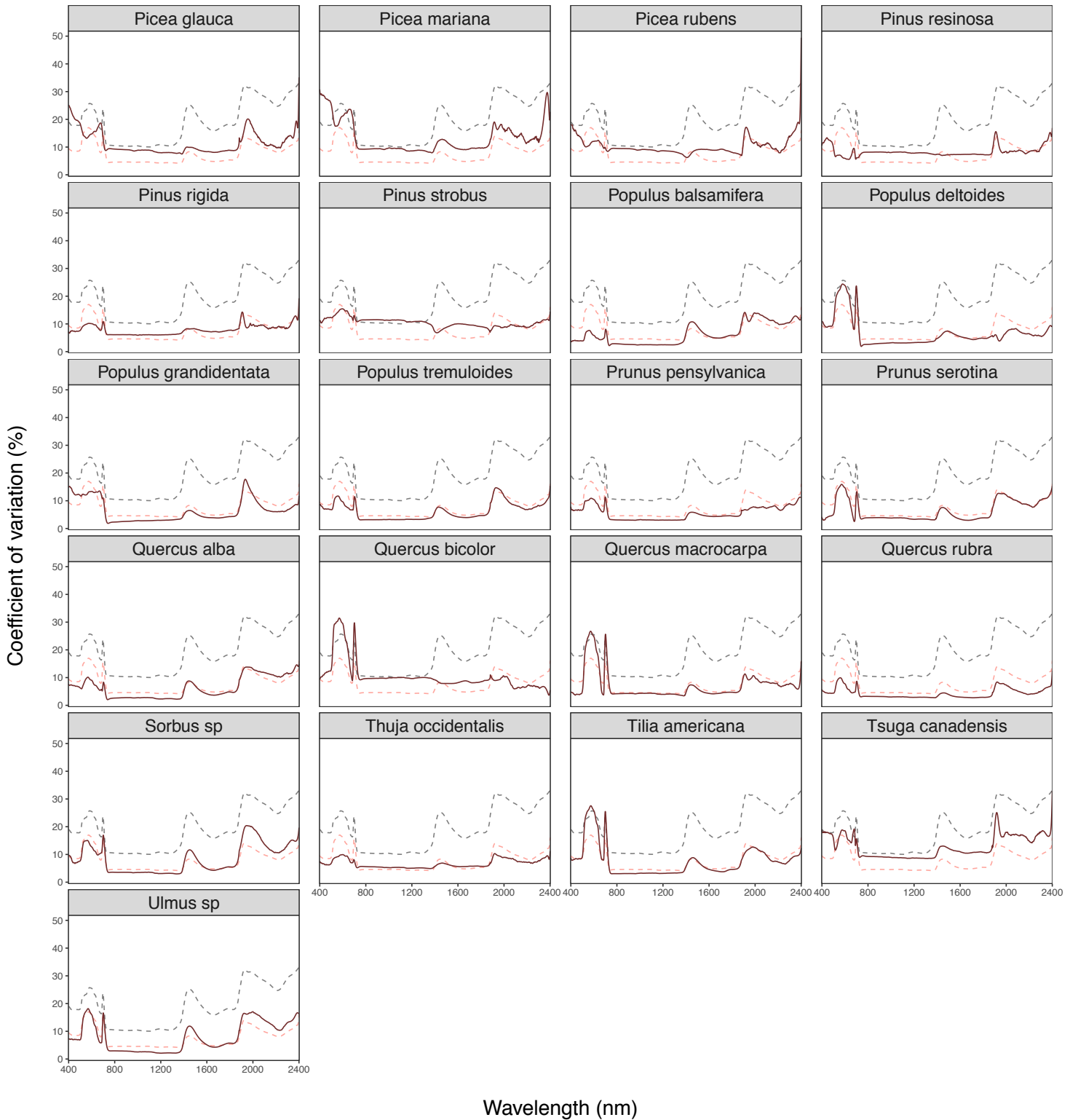


Figure Supplémentaire S4 Coefficient of variation across the spectrum for all species

Dotted lines represent mean interspecific and intraspecific variation Species code are listed in supplementary table S1

Tableau Supplémentaire S2 Model comparison for the Hurdle Analyses
 Model comparison between a Poisson model, negative binomial with and without interaction. The Poisson model does not take overdispersion into account

	AIC	Log likelihood ratio test			
		Comparison	Log-likelihood	Chisq.	P val.
M1-Hurdle model. poisson (ZAP)	3103.606		-1545.8		
M2-Hurdle model. no interaction (ZANB)	2192.361	M1-M2	-1089.2	913.24	< 2.2e-16 ***
M3-Hurdle model. interaction (both in zero and count parts)	2193.320	M2-M3	-1087.7	3.0408	0.2186

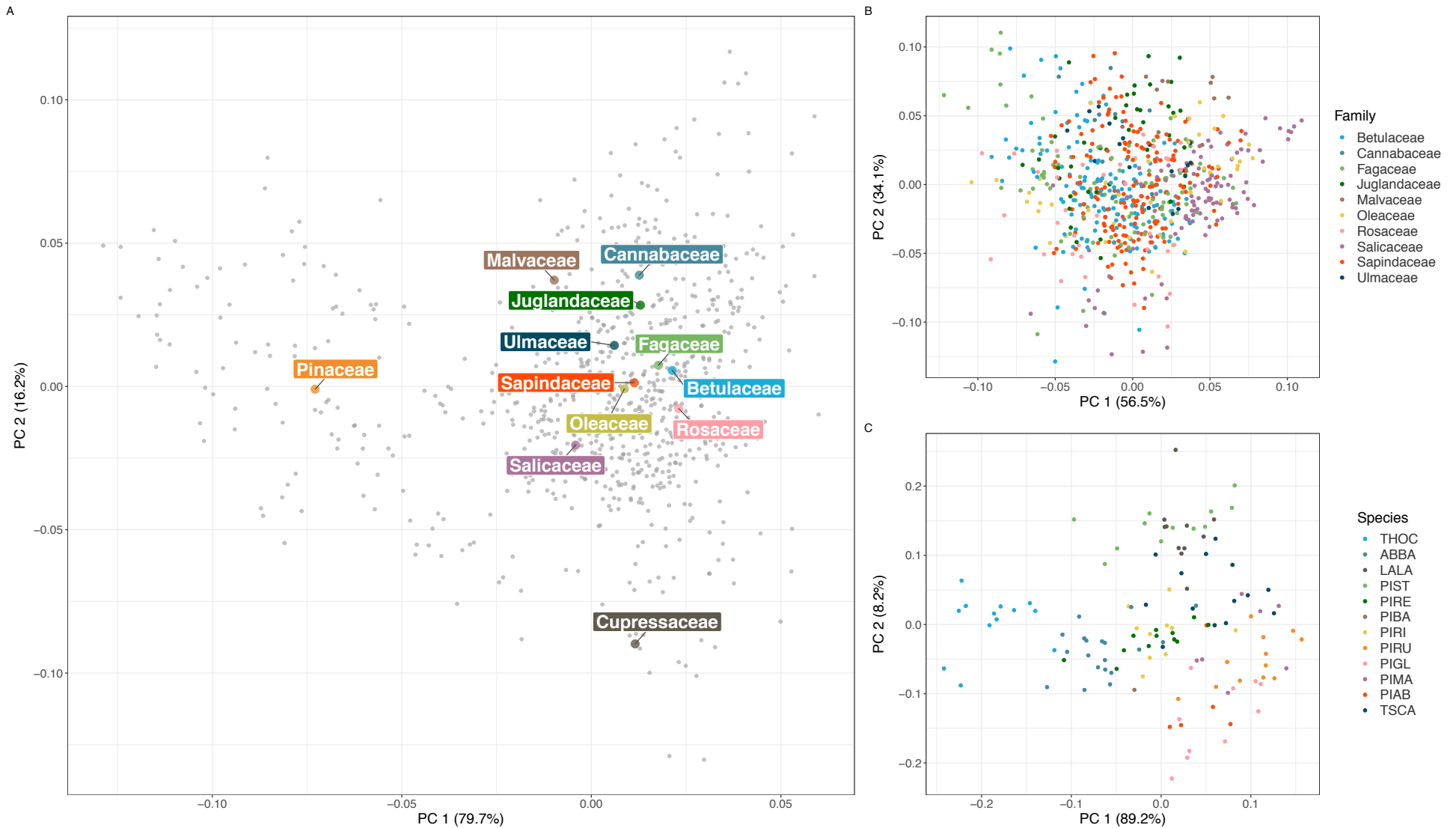


Figure Supplémentaire S5 Principal component analysis of spectral reflectance

A) PCA of the full set of species. The coloured dots correspond to the centroid of the corresponding plant family in spectral space; B) PCA of angiosperm clades (36 species) ; C) PCA of conifer species.(12 species)

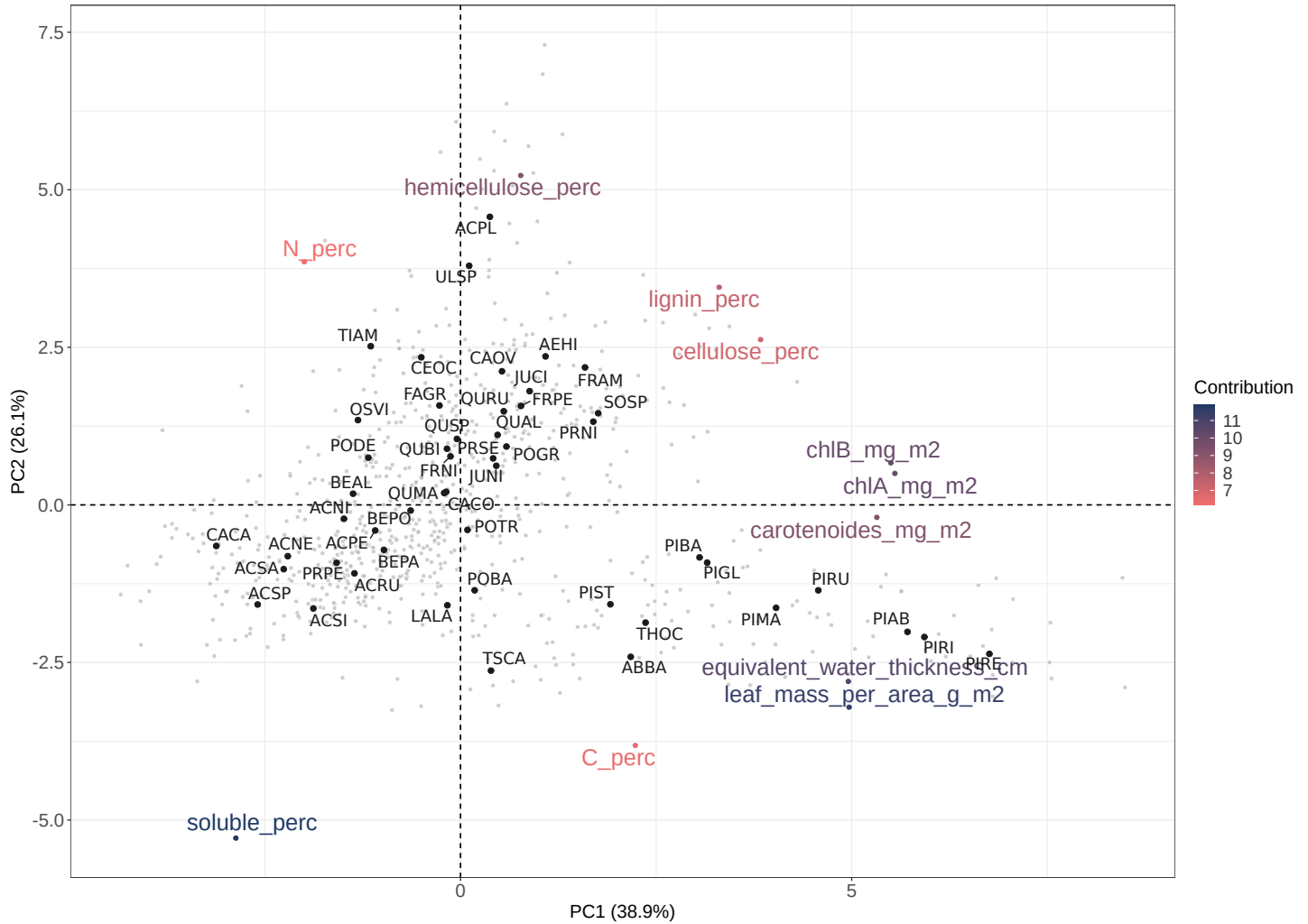


Figure Supplémentaire S6 Principal component analysis of species functional traits
 Species centroid in trait space are shown as black dots. The color axis refers to strength of trait contribution in the construction of trait space determined by principal axis one and two.