Received: 9 February 2012 ■ *Revised: 24 May 2012* ■ *Accepted: 7 June 2012* ■ *Publication: 20 June 2012*

JNIRS JOURNAL OF NEAR INFRARED SPECTROSCOPY

The potential use of near infrared spectroscopy to discriminate between different pine species and their hybrids

Jesús A. Espinoza, Gary R. Hodge and William S. Dvorak

Department of Forestry and Environmental Resources, Camcore, North Carolina State University, Raleigh, NC 27695–8008, USA.
E-mail: jaespino@ncsu.edu

There is growing interest in the use of pine hybrids in commercial forestry plantations in the tropics and sub-tropics. However, the production of pine hybrid seeds can be difficult and is dependent on the presence of an adequate number of male and female strobili, timely application of the pollination bag, good pollination techniques and reasonable weather conditions. After pollination, a wait of two or more years is required for cones to mature and for seeds to be collected. The seeds collected from artificial hybrid crosses in an orchard are assumed to be true hybrids, but might also be the (female) pure species if pollen contamination has occurred prior to or during bagging of the male strobili. Confirming hybridity in pines is often very difficult in the seedling stage when only needle morphological characteristics are used. In this study, we examined ground oven-dried needle samples of 16 pine species from different geographic regions using near infrared (NIR) spectroscopy to determine if this method is effective in distinguishing between pine species. We also created three "simulated hybrids" by manually mixing needles from three sets of parental pure species. The raw near infrared reflectance spectroscopy data were transformed using standard normal variate and de-trending techniques and a model was developed to distinguish between pure pine species and their "hybrids" using discriminant analysis. A total of 120 paired-species models were developed (one for each potential hybrid of the 16 species). For each of the 120 paired-species models, there were 20 independent observations in a validation data set and the 2400 observations were classified with 94% accuracy. Models were also developed for each of six species-simulated hybrid data sets. A total of 120 independent validation observations were classified as either parental species or simulated hybrid with 90% accuracy. The results indicate that NIR spectroscopy can be used as an effective tool to distinguish between pure pine species and suggest that it will also distinguish hybrids from their parents. Using NIR spectroscopy to verify hybridity in pines might be quicker and less expensive and, in some cases, as accurate as using molecular techniques.

Keywords: hybridisation, introgression, hybrid verification, pine needles

Introduction

Tree breeders are always looking for ways to improve adaptability, productivity and wood quality of plantation species. One technique is to make interspecific crosses between two parental species that possess important complementary traits to create a well-adapted hybrid population for use in the next commercial generation. Deployment of commercial hybrids is common in *Populus* and *Eucalyptus* because of the ease in making crosses, the short waiting period to obtain seeds, the ability to mass-produce hybrid material vegetatively and the expression of hybrid vigour in the progeny. Use of pine hybrids, although commercially important in some locations,[1] has developed more slowly than with broadleaf species. Contributing factors include erratic flowering patterns in some parental species (especially when planted as exotics) and a long waiting period of at least two years to collect seeds following controlled pollination. However, the primary reason has been the difficulty in developing suitable multiplication techniques to mass produce the hybrid once created and field-tested.

Recently, there has been renewed interest in pine hybrid forestry[2,3] because the sites available for plantation

ISSN: 0967-0335
doi: 10.1255/jnirs.1006

establishment today are often slightly dry or cold and better adapted varieties are needed. A large pine hybrid testing programme, coordinated by Camcore (International Tree Breeding and Conservation Program) at North Carolina State University, which is sponsored by the forest industry in the southern hemisphere, was initiated in 2005. Since that time, more than 45 different hybrid combinations have been attempted by the industrial partners and a total of 60 hybrid trials have been established in the field in South America and southern Africa.[4,5] Furthermore, tree diseases are becoming more common and new pine hybrid varieties need to be developed[6] in order to take advantage of superior tolerance of one of the pure species. One of the best success stories in pine plantation forestry in recent years is the rapid development of the *P. patula* × *P. tecunumanii* hybrid in South Africa to deal with survival problems in the nursery and in the field associated with the Pitch canker fungus (*Fusarium circinatum*).[2,3,7] Vegetative propagation techniques are improving and are becoming increasingly common, particularly in the southern hemisphere,[8–11] giving foresters the ability to mass produce hybrids for commercial deployment once they are developed.

One of the challenges when conducting pine hybrid breeding is to confirm that the seeds produced from an artificial cross between two different species are true hybrids and not the result of the female parent being contaminated with local pollen from the same species before or during the bagging of female strobili. Therefore, verification techniques are needed to confirm that the putative hybrid seeds or seedlings are actually hybrids before attempting any costly commercial mass multiplication procedure. Using needle morphology to discern whether a pine seedling is a hybrid between two parental species is often difficult and impractical.

To address the problem of verifying hybridity of pine seeds or seedlings, researchers have used a number of chemical and molecular markers such as isozymes,[12,13] randomly amplified polymorphic DNA (RAPD)[14,15] and single nucleotide polymorphisms (SNPs).[16,17] Even though these molecular tools are sometimes effective, they have multiple disadvantages in that they require highly skilled labour, have expensive set-up costs that can be time-consuming, often involve hazardous chemicals and are sometimes not accurate because hybrids have a tendency to generate atypical banding patterns.[18]

Another approach to assessing hybridity in pines might be near infrared (NIR) reflectance spectroscopy. NIR spectroscopy has been used for a number of years for quantitative assessment of many constituents in agricultural and food items such as protein, fibre, fat etc.[19] The NIR technique consists of irradiating the samples with near infrared light, which is absorbed selectively depending on the presence and chemical arrangement of the bonds, mainly between carbon, hydrogen, oxygen and nitrogen, that give rise to a variety of functional groups that are NIR active.[20,21] The technique has been used successfully to determine nitrogen and lignin content of both fresh leaves and litter from a range of forest tree and grassland species.[22] Several authors have demonstrated that NIR spectroscopy could be used to determine wood chemistry and is an inexpensive and rapid alternative to traditional laboratory methods.[23–28] In tree breeding, NIR spectroscopy has been used to distinguish *Betula pendula* and *B. pubescens* from their hybrids[19] and, more recently, the same technique was used to successfully discriminate seedlings of *Eucalyptus globulus*, *E. nitens* and their hybrids.[20]

Based on these results, a study was conducted to determine if seedlings from 16 different pure pine species could be identified using NIR spectroscopy assessment of dried pine needles. The study included two phases. The first phase attempted to separate pure species from each other using NIR spectroscopy. The second phase consisted of creating "simulated" pine hybrids by manually mixing needles from two different pure species and scanning these to evaluate if NIR spectroscopy could consistently separate them from the parental species. The objective of the study was to demonstrate that NIR technology could be used to discriminate between different pine species and, thus, possibly between pure species and hybrids, offering an alternative to more costly molecular assessments.

Materials and methods

Genetic material and test design

The study was conducted using seedlings from 16 different pine species (Table 1) from tropical, sub-tropical and temperate regions in the western hemisphere. Sub-populations of *Pinus tecunumanii* from low- (LE) and high elevations (HE) in Central America and the varieties of *P. greggii* var. *greggii* and var. *australis* in northern (N) and southern (S) Mexico, clearly distinguishable in molecular analysis,[29–30] were treated as different species in this study.

The seeds of the 16 pine species were germinated in 200 cm^3 containers, Treepots (Stuewe & Sons, http://www.stuewe.com/products/treepots.php), in the greenhouse at NC State University. The nursery medium was a 1 : 1 : 1 mixture of sterilised sand, vermiculite and perlite. Seedlings were watered every two days. A low dosage NPK fertilizer was applied to the seedlings when needed. At 12 months, the seedlings were transplanted to larger containers of 1650 cm^3 Treepots and fertilized with osmocote. At the time of transplanting, the seedlings of each species were randomly grouped into five replicates of five seedlings each (25 seedlings per species) and allowed to develop for an additional five months before harvesting the needles.

Sample processing

Phase I. Pure species

Completely elongated needles from the first flush of each seedling were collected and treated as an independent sample. Every needle sample from each replicate and species was placed in a lunch-sized paper bag. Information about date, species ID and replicate number were recorded. This sampling scheme yielded a total of 400 samples (16 species × 5 replicates × 5 seedlings). Seven samples were eliminated due to the low amount of foliage available, leaving a total of 393 samples. The foliage samples were dried by placing them in a forced-air oven at 500°C until a constant weight was achieved.[31]

Table 1. Pinus species used in this study and their climatic zones.

Species code	Species name	Zone	Latitude	Longitude	Elevation range (m)	Average rainfall (mm)
PCB	*P. caribaea* var. *bahamensis*	Tropical/ Sub-topical	12–27° N	77–90° W	1–12	1250
PCH	*P. caribaea* var. *hondurensis*	Tropical	12–18° N	83–90° W	5–800	2115
PM	*P. maximinoi*	Tropical	13–19° N	86–100° W	750–2100	1403
PO	*P. oocarpa*	Tropical/ sub-topical	13–28° N	86–95° W	500–2900	149
PTHE	*P. tecunumanii* (HE)	Tropical	14–17° N	86–93° W	1170–2900	1602
PTLE	*P. tecunumanii* (LE)	Tropical	12–17° N	85–89° W	400–1650	1393
PGS	*P. greggii* (S)	Sub-topical	20–21° N	98–99° W	1150–2350	1367
PH	*P. herrerae*	Sub-topical	16–23° N	97–106° W	1600–2450	1102
PL	*P. leiophylla*	Sub-topical/ temperate	12–29° N	96–108° W	1600–3000	1150
PPP	*P. patula* var. *patula*	Sub-topical/ temperate	16–24° N	96–100° W	1490–3000	1489
PPL	*P. patula* var. *longipedunculata*	Sub-topical	17–18° N	92–97° W	1800–2800	1100
PS	*P. pseudostrobus*	Sub-topical	14–22° N	90–105° W	1600–3200	1150
PR	*P. radiata*	Sub-topical/ temperate	27–30° N	115–120° W	300–1100	560
PE	*P. elliottii*	Temperate	27–35° N	81–92° W	1–150	1270
PGN	*P. greggii* (N)	Temperate	24–25° N	100–101° W	1960–2590	681
PT	*P. taeda*	Temperate	30–39° N	78–92° W	0–610	1270

Once dried, the samples were ground to a fine powder in a cyclone mill and passed through a 1 mm screen to minimise the particle-size effect. The samples were processed using the methodologies suggested by Atkinson and collaborators[19] where each sample of 1 g was held in a rotating sample cup and scanned at visible and NIR wavelengths between 400 nm and 2500 nm using a scanning monochromator (Model 6500, Foss NIRSystems, Silver Spring, MD, USA). Spectra were acquired at every 2 nm intervals, giving a total of 1050 data points for each sample. After grinding, all samples were re-dried at 50°C for 24 hours prior to NIR scanning. Samples were removed from the oven and allowed to equilibrate to room temperature inside a laboratory desiccating chamber before measurement. The temperature in the laboratory was maintained at 24°C. The time from foliage collection to the completion of the NIR scanning was approximately one week.

Phase II. Simulated pine hybrid samples by mixing needles from two pure species

In order to get some indication of how well NIR spectroscopy might be able to distinguish between a pure species and a hybrid between and one pure species and another pure species, a second experiment using "simulated hybrids" was conducted. The "hybrid" foliage was simulated by mixing foliage of two pure species. Throughout the remainder of the manuscript, these mixtures will be referred to as "simulated hybrids".

The processing of samples for Phase II was very similar to Phase I. However, in the second phase, only six of the pure pine species (*Pinus caribaea* var. *hondurensis, P. tecunumanii* (LE), *P. patula* var. *patula, P. greggii* (S), *P. elliottii* and *P. taeda*) representing tropical, sub-topical and temperate environments, respectively, were used (Table 2). Each needle sample from the six species was split into two sub-samples. One sub-sample represented a parental pure species. The second sub-sample was mixed with a sub-sample of another parental species to create simulated pine hybrids of *P. tecunumanii* (LE) × *P. patula, P. taeda* × *P. greggii* (S) and *P. elliottii* × *P. caribaea* (Table 2). All samples were then scanned. A total of 450 samples were used in this phase. One hundred fifty samples from pure pine species (six species × five replicates × five seedlings) and 300 samples from the simulated hybrids (mixed needles) (three pairs of two species × ten samples × five replicates) were processed using the methodology described for Phase I.

Data processing

As a preliminary step, we tried a few different combinations of pre-processing on a limited number of data sets, including standard normal variate, de-trend, multiple scatter correction and 2nd derivative. We then chose a single approach to use for

Table 2. Pure pine species selected to create simulated pine hybrids by mixing needles from the two pure species.

Species code	Species name	Region
Pure		
PCH	*P. caribaea* var. *honurensis*	Tropical
PTLE	*P. tecunumanii* (LE)	Tropical
PGS	*P. greggii* (S)	Sub-topical
PPP	*P. patula* var. *patula*	Sub-topical/temperate
PE	*P. elliottii*	Temperate
PT	*P. taeda*	Temperate
Simulated hybrid		
PTLE×PP	*P. tecunumanii* (LE)×*P. patula* var. *patula*	
PT×PGS	*P. taeda*×*P. greggii* (S)	
PE×PCH	*P. elliottii*×*P. caribaea* var. *hondurensis*	

all 120 paired-species datasets and all nine pure species—"simulated hybrid" datasets. In all cases, the raw spectral data, log(1/reflectance) were processed using standard normal variate and de-trending transformations.[32] All the processing was carried out using The Unscrambler X ver 10.0 software (CAMO Software Inc., NJ, USA). The data were first analysed using the entire range of wavelengths between 400 nm and 2500 nm in order to determine which regions of the spectrum showed the highest discriminatory power. Subsequently, based on the results from the single-wavelength analysis, further analyses were carried out using combinations of wavelengths from these regions. The discriminant analyses of the transformed data were conducted using PROC DISCRIM of SAS software, version 9.2 (SAS Institute Inc. Cary, NC, USA). PROC DISCRIM develops a discriminant function using a measure of the generalised squared distance between classes, based on the pooled covariance matrix and accounting for the prior probabilities of the classes. In order to test the classification model on an independent data set, the first three of five replicates (15 seedlings per species) of phase I were used as a calibration data set (i.e. to build the classification model) and the last two replicates (ten trees per species) were used as the validation data set. A classification model was built for every pair of species, a total of 120 models, in order to examine every possible hybrid for the 16 species. In the second phase, the calibration data set consisted of the first three replicates including the three artificial hybrids (15 trees per pure species plus 30 trees per artificial hybrid) and the last two replicates (ten trees per species from pure species and 20 trees per artificial hybrid) were used as the validation data set. For each simulated hybrid, two classification models were built to separate the two parental species from the hybrid, a total of six models (three simulated hybrids versus two parent species). This would represent a typical situation in a hybrid breeding programme, where the maternal pure species parent is known and, thus, the options for putative hybrid progeny would be "hybrid" produced from controlled-pollination, or "pure species" produced by pollen contamination from the same species as the mother.

Finally, to examine which wavelengths might be important in distinguishing between pairs of species, three paired-species datasets were examined as case studies *(P. elliottii* and *P. caribaea* var. *hondurensis*, *P. patula* var. *patula* and *P. tecunumanii* (LE) and *P. taeda* and *P. greggii* var. *australis*). For these datasets, principal component analysis was conducted using The Unscrambler on each of the three paired species datasets and loadings on wavelengths in the first principal component were compared. Also, a stepwise discriminant analysis was carried out with SAS to identify the best discriminant models using only five wavelengths. The characteristic bond vibration and chemical structures associated with the selected wavelengths were identified, using literature reviews of studies using samples of agricultural forage[33] and wood.[34]

Results

Discriminant analysis models were built for each of the 120 species pairs using three sets of wavelengths: (a) all available wavelengths (400–2500 nm), (b) wavelengths in the near infrared region (1100–2500 nm) and (c) wavelengths 1850–2500 nm. Examinations of weights placed on specific wavelengths showed that the region of the spectrum from 1850 nm–2500 nm was particularly useful to help distinguish species but, in some cases, the model with wavelengths 1850–2500 nm substantially improved the separation of species over the other two wavelength sets. For example, the comparison of *P. greggii* (S) with *P. caribaea* var. *bahamensis* improved classification accuracy from 50% using the entire spectrum 400–2500 nm and 80% using the region 1100–2500 nm, to 90% using the region 1850 nm–2500 nm. Of the three possible wavelength sets, the model with the best classification accuracy was chosen for each of the 120 species pairs. The most accurate model came from 400–2500 nm for 54 of the species pairs, from 1100–2500 nm for 34 of the species pairs and from 1850–2500 nm for 32 of the species pairs.

Each species was used in a paired-species model with all other species so there were 120 different models in total.

For each paired species model, 15 samples of each species were used to develop the calibration model and, in all 120 models, all the calibration samples were correctly classified by species. Table 3 presents the results of species classification for the independent validation data sets (ten samples per species × two species = 20 independent samples) for all of the 120 models. As an example, the paired-species model to distinguish *P. caribaea* var. *hondurensis* and *P. elliotti* correctly classified 90% of the *P. elliottii* validation samples as *P. elliottii* (see Table 3, row *caribaea* var. *hondurensis,* column *elliottii*) and correctly classified 100% of the *P. caribaea* var. *hondurensis* validation samples (see Table 3, row *elliottii,* column *caribaea* var. *hondurensis*). Across all 16 species and all 120 species pairs, the mean for correct classification was 94%, with a range from 60% to 100% (Table 3). As might be expected, some of the lower accuracy values (i.e. lower percentages of correct classification) occurred when species pairs were closely related. For example, *P. tecunumanii* (LE) and *P. oocarpa* from Central America are very closely related genetically. Microsatellite molecular markers are able to distinguish between the two taxa[30] but RAPD molecular markers cannot.[35] The NIR model correctly assigned only 60% of the *P. oocarpa* and 80% of the *P. tecunumanii* (LE). Similarly, for the two varieties of *P. tecunumanii* (high-elevation and low-elevation), NIR spectroscopy was able to distinguish between them with about the same level of accuracy, 80% for HE and 70% for LE. The NIR models also had difficulty distinguishing *P. patula* var. *longipedunculata* from both of the varieties of *P. tecunumanii*, with accuracies of 60% and 70% (Table 3). Thorough botanical analysis of needles and cones from mature trees and RAPD molecular marker assessment have both easily separated *P. patula* var. *longipeduculata* from *P. tecunumanii* (HE) in natural stands.[36] However, their external needle morphology and colour are very similar, making them practically indistinguishable in the seedling stage.

There were, however, cases with some closely related pairs of species where the NIR models were able to distinguish the species with very high accuracy. For example, *P. patula* var. *patula* and *P. patula* var. *longipedunculata* were distinguished from each other with 100% accuracy for both taxa. Two species in the Pseudostrobus group, *P. pseudostrobus* and *P. maximinoi* were distinguished from each other with 100% and 90% accuracy, respectively.

Overall, the two species which were most easily classified using NIR models, were *P. patula* var. *patula* and *P. pseudostrobus.* Regardless of which species it was paired with, the independent *P. patula* var. *patula* samples were classified as *P. patula* var. *patula* with 100% accuracy, with the other species being correctly classified with 96% accuracy (Table 3). The corresponding values for *P. pseudostrobus* were 99% and 97%. *Pinus elliottii* was another species that was easily classified using NIR spectroscopy, with 97% accuracy for *P. elliottii* and, on average, 99% accuracy for the other species in the paired analysis (Table 3).

The species that was the most difficult to classify correctly was *P. herrerae.* Good results were found when *P. herrerae* was paired with most species, but there were problems distinguishing *P. herrerae* from *P. pseudostrobus* (60% accuracy) and in correctly classifying both varieties of *P. tecunumanii* when paired with *P. herrerae* (60% and 70% accuracy for HE and LE, respectively). *Pinus herrerae*, an open-cone pine, does have some external needle morphologic traits that are similar to the closed-cone pines, *P. patula* var. *longipedunculata* and *P. tecunumanii*.[36] At one time, research was conducted to determine if *P. tecunumanii* in Central America was possibly an extension of *P. herrerae* in western Mexico.[37] Results indicated that the two species could be separated by both botanical and molecular analyses. However, when we analysed the data from *P. herrerae* and *P. pseudostrobus* in more detail and developed a model using different range of wavelength (1890–2268 nm), NIR spectra was able to distinguish between them with 100% accuracy (data not shown). These results suggest that if a general model does not work to distinguish some species, a specific model could be developed to achieve higher levels of accuracy.

Similar to the trend found for the pure species in phase I, the spectral region from 1850–2500 nm was found to be useful for the discriminant analysis of the three simulated hybrids and the six different pure pine "parental" species. Interestingly, wavelengths 400–780 nm, in the visible region of the spectrum, also produced classification models with good accuracy. However, we focused on the same three sets of wavelengths (400–2500 nm, 1100–2500 nm and 1850–2500 nm) in order to directly compare them with the pure species analyses.

For the hybrid data sets, there were six case studies of a pure species compared to a simulated hybrid. As before, the calibration models gave 100% correct classification in all cases. Table 4 presents the results of validation data sets for these case studies and is read in the same manner as Table 3. The best results were obtained for the simulated *P. taeda* × *P. greggii* (or *P. greggii* × *P. taeda*) versus the maternal species. For the *P. taeda* × *P. greggii* versus *P. taeda* model, the independent *P. taeda* samples were classified as *P. taeda* with 100% accuracy, while 85% of the hybrid samples were classified as hybrids. In other words, no *P. taeda* were classified as hybrid, all samples that the NIR model identified as a (simulated) hybrid was indeed a (simulated) hybrid with 15% of the hybrids being incorrectly classified as *P. taeda*. For the *P. greggii* × *P. taeda* versus *P. greggii* model, 100% of the hybrids were classified correctly and 80% of the *P. greggii* were classified correctly; however, 20% of the *P. greggii* samples were classified as hybrid. The results with the *P. elliottii* × *P. caribaea* var. *hondurensis* hybrids were also good. For the *P. elliottii* × *P. caribaea* var. *hondurensis* versus *P. elliottii* model, 100% of the hybrids and 80% of the *P. elliottii* were classified correctly. However, 20% of the *P. elliottii* samples were classified as hybrid. For the *P. caribaea* var. *hondurensis* × *P. elliottii* versus *P. caribaea* var. *hondurensis* model, both the hybrid and the *P. caribaea* samples were classified with 100% accuracy. Finally, the *P. patula* × *P. tecunumanii* hybrid models were reasonable, with between 70% to 100% accuracy for hybrid and pure species (Table 4). For the two cases, 20% to 30% of the pure species samples were mistakenly identified as hybrids.

Table 3. Percentage of the validation data sets samples classified correctly by species using NIR spectroscopy data from foliage scans of 16 pine species. Samples were classified using paired-species models and discriminant analysis models for all possible species pairs*.

Species	*caribaea* var. *bah.*	*caribaea* var. *hond.*	*maximi-noi*	*oocarpa*	*pseudos-trobus*	*herrerae*	*leiophylla*	*radiata*	*elliotii*	*taeda*	*tecunu-manii* (HE)	*tecunu-manii* (LE)	*greggii* (N)	*greggii* (S)	*patula* var. *pat.*	*patula* var. *long.*	*AVERAGE*
caribaea var. *bah.*		100	100	100	100	90	100	90	90	100	100	100	90	90	100	100	**97**
caribaea var. *hond.*	70		90	90	100	85	100	85	90	100	90	95	100	100	100	100	**93**
maximinoi	100	100		90	90	80	80	90	100	90	90	70	100	90	90	90	**90**
oocarpa	100	100	80		100	70	90	100	100	100	90	60	100	100	100	80	**91**
pseudos-trobus	100	100	100	100		90	100	100	100	100	100	100	100	100	100	100	**99**
herrerae	100	100	90	90	60		100	100	100	90	100	80	90	100	90	100	**93**
leiophylla	90	100	80	100	100	90		90	100	90	100	100	90	90	80	100	**93**
radiata	100	100	90	90	100	90	80		100	100	80	100	100	80	90	90	**93**
elliotii	80	100	100	100	100	100	100	90		90	100	100	90	100	100	100	**97**
taeda	100	100	90	100	100	100	100	100	100		100	100	90	100	100	100	**97**
tecunu-manii (HE)	60	100	100	70	100	60	90	100	100	90		80	90	90	100	90	**88**
tecunu-manii (LE)	95	100	90	80	100	70	100	95	100	100	70		100	100	100	85	**92**
greggii (N)	100	90	100	100	100	100	100	100	100	100	100	100		80	100	100	**98**
greggii (S)	90	100	90	100	100	100	90	100	100	100	100	90	100		100	100	**97**
patula var. pat.	100	100	100	100	100	100	100	100	100	100	100	100	100	100		100	**100**
patula var. long.	100	100	90	100	100	80	90	100	100	100	60	70	100	100	100		**93**
AVERAGE	**92**	**99**	**93**	**94**	**97**	**87**	**95**	**96**	**99**	**97**	**92**	**89**	**95**	**95**	**96**	**94**	**94**

*Values in the table are the percentage of independent observations correctly classified as Species A (rows) in a paired-species model with Species B (columns). Results shown are from the best of three NIR spectroscopy discriminant analysis models using different sets of wavelengths: a) 400–2500 nm, 1100–2500 nm and 1852–2500 nm.

Table 4. Percentage of simulated hybrids and pure pine species classified correctly for six hybrid-species combinations* using NIR data from foliage scans and discriminant analysis.†

Species	*PT×PGS*	*PT*	Species	*PT×PGS*	*PGS*
PT×PGS	85	15	*PT×PGS*	100	0
PT	0	100	*PGS*	20	80
Species	***PE×PCH***	***PCH***	**Species**	**PE×PCH**	**PE**
PE×PCH	100	0	PE×PCH	100	0
PCH	0	100	PE	20	80
Species	**PPP×PTLE**	**PPP**	**Species**	***PPP×PTLE***	***PTLE***
PPP×PTLE	100	0	*PPP×PTLE*	100	0
PPP	30	70	*PTLE*	20	80

*Hybrid-species abbreviations are as follows:
PT×PGS=P. taeda×P. greggii (S), *PT=P. taeda*, *PGS=*P. *greggii* (S)
PE×PCH=P. elliottii×P. caribaea var. *hondurensis*, *PE=P. elliottii*, *PCH=P. caribaea* var. *hondurensis*
PPP×PTLE=P. patula var. *patula×P. tecunumani* LE, *PPP=P. patula* var. *patula*, *PTLE=P. tecunumanii* (LE)
†Values in the table are the percentage of independent observations correctly classified as Species A (rows) in a paired-species model with Species B (columns). Results shown are from the best of three NIR discriminant analysis models using different sets of wavelengths: a) 400–2500 nm, 1100–2500 nm and 1852×2500 nm.

Figure 1 presents the wavelength loadings for the first principal component of three different species pairs. The first principal component explained 75% of the variation in the spectral dataset for the *P. patula* var. *patula* and *P. tecunumanii* (LE) dataset, 70% of the variation for *P. taeda* and *P. greggii* var. *australis* and 47% for *P. elliottii* and *P. caribaea* var. *hondurensis*. The loadings were very similar for all three pairs of species, with peaks with high (absolute value) loadings around 1320 nm, 1460 nm, 1866 nm, 1932 nm, 2108 nm and 2490 nm (Table 5). Data relating band position and chemical bonds and structures have been compiled for a wide array of forage and agricultural products[33] and these may have relevance to the ground foliage samples used in this study. Wavelength 1460 nm is associated with N–H stretch and urea and 1860 nm is associated with C–Cl bonds. Wavelengths 1930 nm and 2100 nm have been associated with starch and cellulose and wavelength 2488 nm is associated with C–H and C–C stretching and cellulose.[33] In studies on wood samples, wavelengths 2110 nm and 2491 nm have been associated with cellulose.[34] There were some regions where the loadings were somewhat different for the three pairs of species, specifically around 1100 nm, 1730 nm and 2320 nm. However, the loadings were not particularly high in any of these regions for any of the three datasets, with the absolute values of the loading less than 0.05 for all wavelengths.

Stepwise discriminant analyses was used to identify the five best wavelength models for each of the three paired-species datasets and the wavelengths selected were rather different for the three datasets (Table 5). All five-wavelength discriminant models were nearly as accurate as the corresponding models using all wavelengths, in terms of correctly classifying independent samples (results not shown). For *P. elliottii* and *P. caribaea* var. *hondurensis*, four of the five selected wavelengths were in the region 1704 nm to 1746 nm (Table 5). In contrast, for *P. patula* var. *patula* and *P. tecunumanii*, none of the wavelengths chosen were in that region (1172 nm, 1206 nm, 1416 nm, 1960 nm and 2264 nm). The *P. taeda* and *P. greggii* var. *australis* model did have one wavelength similar to the *P. patula* var. *patula* and *P. tecunumanii* data set (1414 nm and 1416 nm, respectively), but the other four wavelengths chosen (2042 nm, 2198 nm, 2310 nm and 2312 nm) were not similar to either of the other two datasets. Interestingly, some of the wavelengths selected for the models (2042 nm and 2444 nm) do not yet have characteristic bonds or chemical components associated with them in the forage[33] or wood[34] literature.

Discussion

The results of this study indicate that discrimination among pure pine species is possible using ground-dried pine needles and near infrared spectroscopy techniques. There were 20 independent observations in a validation data set for each of the 120 paired-species models and these 2400 observations were classified with 94% accuracy. For the six species-simulated hybrid data sets, there were a total of 120 independent validation observations that were classified with 90% accuracy.

The difficulty in accurately classifying closely related species or varieties (for example *P. tecunumanii* LE vs *P. tecunumanii* HE, with only 75% accuracy) might suggest that it will be more difficult to build NIR models to clearly distinguish between a pure species and a hybrid. However, many hybrid combinations would presumably be between more distantly related pine species, making the foliar chemical profile of the hybrid relatively distinct from the pure species. Furthermore, if a tree breeder was working with a particular hybrid on an operational scale where an NIR verification model would be useful, this would justify more comprehensive efforts to build a specific

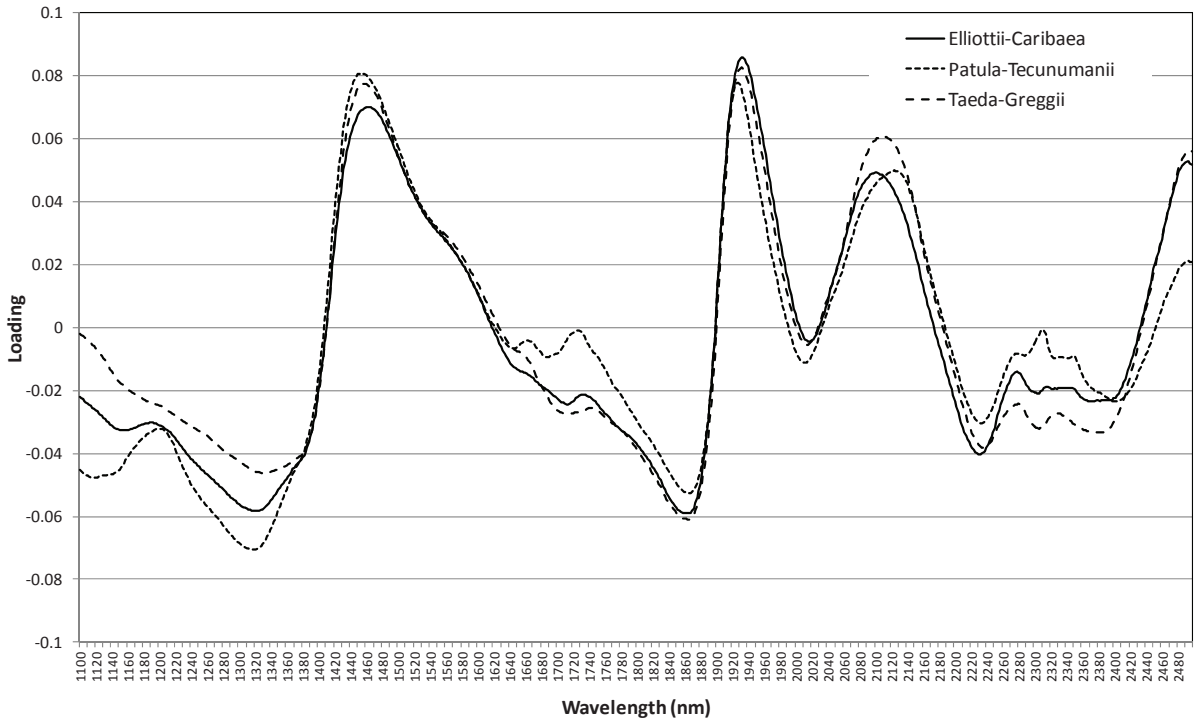

Figure 1. Wavelength loadings on the first principal component of NIR reflectance data from foliage scans of three paired-species datasets (*P. elliottii* and *P. caribaea* var. *hondurensis*, *P. patula* var. *patula* and *P. tecunumanii* (LE), *P. taeda* and *P. greggii* var. *australis*).

model rather than the one used in this study. Larger calibration sets (i.e. more samples) would likely produce a more robust classification model that would increase accuracy and would also allow a more comprehensive effort to optimise the wavelengths and/or data transformations used in the model.

There were 120 different paired-species models in this study and it was not feasible to examine all of them in detail. For the three paired-species datasets examined, the overall variation in the spectral data appeared similar, based on the wavelength loadings of the first principal component. However, the stepwise discriminant analyses identified very different sets of wavelengths as important. These results suggest that within-species variation in NIR spectra is rather large and that the between-species differences are relatively small and subtle. Overall, it is difficult to say exactly what chemical constituents and/or NIR wavelengths are most critical to distinguish among species, but it seems likely that these will differ depending on the species pair. It is known that pine species vary in chemical composition of their wood. For example, there is a large difference in the average lignin content of wood of *P. taeda* (29%) versus *P. maximinoi* (25%).[27] The types and quantities of hemicellulose sugars (for example, xylan, glucuronoxylan, arabinoxylan, glucomannan and xyloglucan) could also vary among species. If these kinds of differences occur in foliage, they should be detectable by NIR spectroscopy. In addition, taxonomists have long noted differences among pine species in monoterpene composition, specifically the presence or absence and relative amounts of various monocyclic and bicyclic terpenes, diterpenes, alkanes, aldehydes and esters.[38] For some of the paired-species data sets, these kinds of differences would discriminate between species and probably would be recognised by using NIR spectroscopy In addition to monoterpene composition, different species can have very distinct nutrient profiles in their foliage, even when grown in the same environment (i.e. common-garden type experiments). This has been seen for seedlings growing in nursery conditions[39,40] and in older trees growing in field conditions. For example, plantings of *P. maximinoi* and *P. tecunumanii* grown next to each other in the Mpumalanga province of South Africa had distinctly different nutrient profiles (Camcore, unpublished data).

Although the simulated hybrids created in this study using 50 : 50 mixtures of needles could differ greatly from natural hybrids (biochemically and genetically), the high accuracy values obtained in this initial study (94% overall in the pure species study and 90% overall in the simulated hybrid study) suggest that there might be sufficient spectral differences among pure pine species and hybrids that NIR spectroscopy could be used to verify putative hybrids. We are currently testing NIR techniques on putative hybrids created by controlled pollination from known parents in seed orchards. Prior to the NIR assessment, seed lots will be verified for hybridity using species-specific SNP markers.[16–17]

The Central American and Mexican pines are known to naturally hybridise in areas where their geographic ranges overlap with related species[41] and admixtures of two species can be

Table 5. First five wavelengths chosen in a stepwise discriminant analysis for three paired-species datasets, using NIR spectra from scans of ground foliage.

Dataset	Wavelength (nm)	Nearest wavelength and characterisation		
		Wavelength, bond vibration	Component	Reference
P. elliottii and *P. caribaea* var. *hondurensis*	1704	@ 1705, C–H stretch first overtone @1703, C–H stretch first overtone	•CH_3 •CH2, cellulose	33 34
	1718	@1720, C–H stretch first overtone	Lignin, hemi.	34
	1724	@1725, C–H stretch first overtone @1724, C–H stretch first overtone	•CH_2 hemicellulose	33 34
	1746	@1740, S–H stretch first overtone	–SH	33
	2444			
P. patula var. *patula* and *P. tecunumanii* (LE)	1172	@1170, C–H second overtone @1170, C–H, HC=CH stretch second overtone	•HC=CH lignin	33 34
	1206	@1215, C–H second overtone	•CH_2	33
	1416	@1415, C–H combination @1417, C–H combination @1417, C–H stretch end first overtone	•CH_2 aromatic lignin	33 33 34
	1960	@1960, O–H stretch/bend combination	cellulose	33
	2264	@2270, O–H stretch/C–H stretch combination @2267, O–H stretch	cellulose cellulose	33 34
P. taeda and *P. greggii* var. *australis*	1414	@1415, C–H combination	•CH_2	33
	2042			
	2198	@2200, C–H stretch, C=O stretch	lignin	34
	2310	@2310, C–H bend second overtone	oil	33
	2312	@2310, C–H bend second overtone	oil	33

as high as 20%.[30] On occasions, these putative hybrids are observed in field trials intermixed with the pure species. We would like to determine the reliability of NIR spectroscopy to classify trees in genetic field trials when crown architecture and needle morphology suggest that natural hybrids are intermixed with the pure species. NIR spectroscopy should be able to verify if trees with atypical morphology are true hybrids, assuming that a spectral database of both pure parental species exists. However, an important question is whether NIR assessments of parental species and putative hybrids must be based on needle samples taken from trees of approximately the same age growing in similar environments, or whether a more robust model based on samples from many environments and ages could be developed. In this study, all the needle samples were taken from seedlings of the same age growing in the same greenhouse. Finally, another application of NIR spectroscopy might be in the area of pine taxonomy and evolution, especially if it is able to discern differences between species closely related by descent more efficiently than molecular approaches.

Conclusions

It seems clear that NIR spectroscopy can be used to distinguish between pure pine species on the basis of dried foliage samples. NIR spectroscopy can also be used relatively successfully to distinguish between foliage samples from a pure species and a "simulated hybrid" (i.e. a mixture of foliage samples from the pure species and another species). These results suggest that NIR spectroscopy has the potential to distinguish between a pure species and a true hybrid. The ease of use and low cost of NIR spectroscopy makes this an attractive option for organisations working with a particular pine hybrid in an operational breeding and plantation program. Research is underway to further investigate this approach using foliage samples of true hybrids verified with molecular markers.

References

1. H.S. Dungey, "Pine hybrids—a review of their use performance and genetics", *Forest Ecol. Manage*. **148,** 243–258 (2001). doi: 10.1016/S0378-1127(00)00539-9
2. A. Kanzler, K. Payn and A. Nel, "Performance of two *Pinus patula* hybrids in southern Africa, *South. Forests: J. Forest Sci.* **74,** 19–25 (2012).
3. C. Snedden and K. Eatwell, "Addressing the challenges of a changing environment through a breeding programme for pine hybrids: A note on reported survival in

P. tecunumanii provenances and hybrid development", *HeartWood* **3,** 4–6 (2009).
4. Camcore, *2009 Annual Report*. Camcore, NC State University, Raleigh, NC. USA, pp. 44 (2009).
5. Camcore, *2010 Annual Report*. Camcore NC State University, Raleigh, NC, USA, pp. 46 (2010).
6. W.S. Dvorak, "The strategic importance of applied tree conservation programs to the forest industry in South Africa", *South. Forests: J. Forest Sci.* **74,** 1–6 (2012). doi: 10.2989/20702620.2012.683635
7. G.R. Mitchell, E.T. Steenkamp, T.A. Coutinho and M.J. Wingfield, "The pitch canker fungus, *Fusarium circinatum*: implications for South African forestry", *South. Forests* **73,** 1–13 (2011). doi: 10.2989/20702620.2011.574828
8. M. Hunt, S. Trueman and A. Rasmussen, "Indole-3-butyric acid accelerates adventitious root formation and impedes shoot growth of Pinus elliottii var. elliottii × P. *caribaea* var. *hondurensis* cuttings", *New Forests* **41,** 349–360 (2011). doi: 10.1007/s11056-010-9227-7
9. B.S. Baltunis, D.A. Huber, T.L. White, B. Goldfarb and H.E. Stelzer, "Genetic analysis of early field growth of loblolly pine clones and seedlings from the same full-sib families", *Can. J. Forest Res*. **37,** 195–205 (2007). doi: 10.1139/x06-203
10. C.L. Rosier, J. Frampton, B. Goldfarb, F.A. Blazich and F.C. Wise, "Improving the rooting capacity of stem cuttings of virginia pine by severe stumping of parent trees, *South. J. Appl. For.* **30,** 172–181 (2006).
11. J. Frampton, F. Isik and B. Goldfarb, "Effects of nursery characteristics on field survival and growth of loblolly pine rooted cuttings, *South. J. Appl. For.* **26,** 207–213 (2002).
12. M.T. Conkle, "Isozyme variation and linkage in six conifer species", in, *Proceedings of Symposium on Isozymes of North America Forest Tree and Forest Insects*. USDA Foresdt Services, Berkeley, CA, USA, pp. 11–17 (1981).
13. J. Chen, C.G. Tauer, B. Guihua, H. Yinghua, M.E. Payton and A.G. Holley, "idirectional introgression between *Pinus taeda* and *Pinus echinata*: evidence from morphological and molecular data", *Can. J. Forest Res*. **34,** 2508–2516 (2004). doi: 10.1139/x04-134
14. C. Plomion, N. Bahrman, C.E. Durel and D.M. O'Malley, "Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers, *Heredity* **74,** 661–668 (1995). doi: 10.1038/hdy.1995.90
15. B.J. Furman, D. Grattapaglia, W.S. Dvorak and D.M. O'Malley, "Analysis of genetic relationships of Central American and Mexican pines using RAPD markers that distinguish species", *Mol. Ecol*. **6,** 321–331 (1997). doi: 10.1046/j.1365-294X.1997.00183.x
16. G.F. Khan, *Single Nucleotide Polymorphism Discovery in Pines (Pinus) Species for Hybrid Analysis*. MSc. Thesis NC State University, Raleigh, USA, pp. 102 (2010).
17. J.P. Jimenez Madrigal, *Tropical Pine Hybrid Verification using Single Nucleotide Polymorphisms (SNPs) Marker Tecnology: Case Studies and Applications to the Forestry Industry*. MSc. Thesis, NC State University, Raleigh, NC. USA, pp. 52 (2011).
18. H. Budak, Y. Bölek, T. Dokuyucu and A. Akkaya, "Potential uses of molecular markers in crop improvement", *KSU J. Sci. Eng*. **7,** 75–79 (2004).
19. M.D. Atkinson, A.P. Jervis and R.S. Sangha, "Discrimination between *Betula pendula*, *Betula pubescens* and their hybrids using near-infrared reflectance spectroscopy", *Can. J. of For.Res*. **27,** 1896–1900 (1997).
20. J.R. Humphreys, J.M. O'Reilly-Wapstra, H.J.L. Harbard, N.W. Davies, A.R. Griffin, G.J. Jordan and B.M. Potts, "Discrimination between seedlings of *Eucalyptus globulus*, *E. nitens* and their F1 hybrid using near-infrared reflectance spectroscopy and foliar oil content", *Silvae Genetica* **57,** 262–269 (2008).
21. W.J. Foley, A. McIlwee, I. Lawler and L. Aragones, "Ecological applications of near infrared reflectance spectroscopy—a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance", *Oecologia* **116,** 293–305 (1998). doi: 10.1007/s004420050591
22. C.A. Wessman, J.D. Aber, D.L. Peterson and J.M. Melillo, "Foliar analysis using near infrared reflectance spectroscopy", *Can. J. For. Res*. **18,** 6–11 (1988). doi: 10.1139/x88-002
23. L.R. Schimleck, P.D. Kube, C.A. Raymond, A.J. Michell and J. French, "Extending near infrared reflectance (NIR) pulp yield calibrations to new sites and species", *J. Wood Chem. Technol*. **26,** 299–311 (2006). doi: 10.1080/02773810601076683
24. C.R. Mora, L.R. Schimleck, A. Clarke, III and R.F. Daniels, "Determination of basic density and moisture content of merchantable loblolly pine logs by near infrared spectroscopy", *J. Near Infrared Spectrosc*. **19,** 391–399 (2011). doi: 10.1255/jnirs.947
25. G.M. Downes, R. Meder and C. Harwood, "A multi-site, multi-species near infrared calibration for the prediction of cellulose content in eucalypt woodmeal", *J. Near Infrared Spectrosc.* **18,** 381–387 (2010). doi: 10.1255/jnirs.910
26. L.R. Schimleck, G.R. Hodge and W. Woodbridge, "Toward global calibrations for estimating the wood properties of tropical, sub-tropical and temperate pine species", *J. Near Infrared Spectrosc.* 1**8,** 355–365 (2010). doi: 10.1255/jnirs.900
27. G.R. Hodge and W.C. Woodbridge, "Global near infrared models to predict lignin and cellulose content of pine wood", *J. Near Infrared Spectrosc*. **18,** 367–380 (2010). doi: 10.1255/jnirs.902
28. R. Meder, D. Marston, N. Ebdon and R. Evans, Spatially-resolved radial scanning of tree increment cores for near infrared predicition of microfibril angle and chemical composition, *J. Near Infrared Spectrosc*. **18,** 499–505 (2010). doi: 10.1255/jnirs.903
29. W.S. Dvorak, A.P. Jordon, G.P. Hodge and J.L. Romero, "Assessing evolutionary relationships of pines in the Oocarpae and Australes subsections using RAPD

markers", *New Forests* **20,** 163–192 (2000). doi: 10.1023/A:1006763120982

30. W.S. Dvorak, K.M. Potter, V.D. Hipkins and G.R. Hodge, "Genetic diversity and gene exchange in *Pinus oocarpa*, a mesoamerican pine with resistance to the pitch canker fungus (*Fusarium circinatum*), *Int. J. Plant Sci.* **170,** 609–626 (2009). doi: 10.1086/597780

31. A.D. Richardson, J.B. Reeves, III and T.G. Gregoire, "Multivariate analyses of visible/near infrared (VIS/NIR) absorbance spectra reveal underlying spectral differences among dried, ground conifer needle samples from different growth environments", *New Phytol.* **161,** 291–301 (2004). doi: 10.1046/j.1469-8137.2003.00913.x

32. R.J. Barnes, M.S. Dhanoa and S.J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra", *Appl. Spectrosc.* **43,** 772–777 (1989). doi: 10.1366/0003702894202201

33. J.S. Schenk, J.J. Workman, Jr and M.O. Westerhaus, "Application of NIR spectroscopy to agricultural products", in *Handbook of Near-Infrared Analysis*, 3rd Edn, Ed by D.A. Burns and E.W. Ciurczak. CRC Press, Boca Raton, FL, USA, pp. 347–386 (2008).

34. M. Schwanninger, J.C. Rodrigues and K. Fackler. "A review of band assignments in near infrared spectra of wood and wood components", *J. Near Infrared Spectrosc.* **19,** 287–308 (2011). doi: 10.1255/jnirs.955

35. B.J. Furman and W.S. Dvorak, "Population level analysis to identify species diagnostic RAPD markers for classification of Central American and Mexican pines", *Forest Genetics* **12,** 67–78 (2005).

36. W.S. Dvorak, A.P. Jordan, J.L. Romero, G.R. Hodge and B.J. Furman, "Quantifying the geographic range of *Pinus patula* var. *longipenduculata* in southern Mexico using morphologic and RAPD marker data", *South African For. J.* 19–30 (2001).

37. W.S. Dvorak, 'Estamos más cerca de entender la ascendencia de las poblaciones de la "variante patula" en la Sierra Madre del Sur, México', *Foresta Veracruzana* **10,** 59–66 (2008).

38. N.T. Mirov, *The Genus Pinus*, 1st Edn. Ronald Press, New York, USA, pp. 602 (1967)

39. T.D. Landis, "Principles, procedures, and predictive abilities of major tests", in *Evaluating Seedling Quality,* Ed by M.L. Duryea. Forest Research Laboratory, Oregon State University, Corvallis, USA (1985).

40. E.A. Kirkby and D.J. Pilbeam, "Calcium as a plant nutrient", *Plant, Cell Environ.* **7,** 397–405 (1984). doi: 10.1111/j.1365-3040.1984.tb01429.x

41. J.P. Perry, *The pines of Mexico and Central America*. Timber Press, Portland, OR, USA, pp. 231 (1991).