NIFA Biomass Research and Development Initiative:
System for Advanced Biofuels Production from Woody Biomass in the Pacific Northwest

Agriculture and Food Research Initiative Competitive Grant
Award: 2011-68005-30407; Accession: 0225392; Project: WNZ-A61328

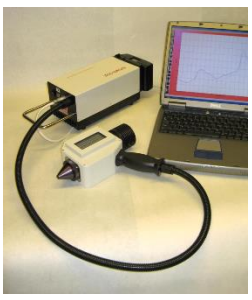
Final Technical Report
Feedstock Development – Task 1.2.5
Pest Management
Without Chemical Inputs

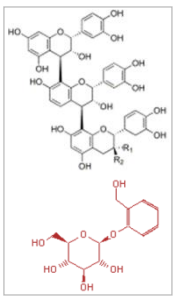
Can Near Infrared Spectroscopy Be Used to Identify Pest-Resistant Poplar Hybrids?

Submitted September 2016
Revised 2020

Jeffery C. Kallestad

WASHINGTON STATE UNIVERSITY
PUYALLUP RESEARCH AND EXTENSION CENTER
2606 W. Pioneer, Puyallup, WA 98371



Contents

Preface

In keeping with the thrust of the Advanced Hardwood Biofuels project; to advance years of theory and concept into planned development; to move many years of research out of academic laboratories and into a pilot plant and field demonstration projects; to discover and resolve questions about production and scale, this revised technical report is written primarily for tree production industry specialists who are contemplating the use of NIRS technology for 'high throughput screening' of new cultivars, and secondarily for the academic audience. The original version was never made publically available. After years of reviewing the data and findings that could not be published, and contemplating the failures, I felt there still was a story worth telling, if for no other reason than to serve as a cautionary tale. Failure gives us the ability to begin again more intelligently.

Executive Summary

Overall goal

Herbivory by mammalian and insect pests pose significant challenges for poplar plantation establishment and productivity. Our project goals were to find new ways to utilize near infrared reflectance spectroscopy (NIRS) to predict pest preferences for new genotypes of hybrid poplar. Our desired outcome was to provide NIR-based selection tools to feedstock breeders for the purpose of identifying genotypes that possess low palatability to herbivores, which when deployed will reduce pesticide use, reduce negative environmental impacts, and reduce risk of plantation failure.

Knowledge gaps and obstacles

NIRS has been used to rapidly and nondestructively measure the concentration of chemical constituents in agricultural commodities, food, and forest products for several decades (1), (2). More recently NIRS has also found applications in the field of chemical ecology to address the interaction of plant secondary metabolites with animal foraging strategy (3). Yet, few studies have explored the deployment of NIRS technology in tree improvement programs for identifying pest-resistant or low-palatability genotypes.

A common approach to identifying pest resistance has been to find differences in secondary compound composition in apparent resistant and susceptible genotypes in hopes of identifying a quantitative trait that can be used for genetic selection. Ample research emphasis has been placed on quantitatively and mechanistically describing the effects that 'defensive' phytochemical components have on feeding preferences and physiology of the herbivore (4), (5). Less research effort has been devoted to determining how multiple compounds; carbohydrates and nutrients in combination with 'deterrent' secondary compounds, act additively to simultaneously deter and attract pests. The implication of the holistic additive effects perspective is that individual genetic factors contributing to palatability become intractably complex. Characterizing these phytochemicals for screening purposes using standard laboratory methods can be a cost-limiting endeavor. In response to this analysis obstacle, developers of NIRS technologies have sought to increase speed and cut costs - yet with some tradeoffs in accuracy and precision. In this project we sought to extend NIRS screening technology, previously used for characterizing wood chemical composition, into this new arena of tree improvement. However, several limiting factors restrict further use and development:

- *Testing with animals is required*. Substantial costs and logistical difficulties are encountered when bringing plants and wild animals together in controlled settings with enough genetic representation and replication to account for natural variation and variability, and presented in a way that is relevant to the plantation pest problem.
- *Linking feeding response to phytochemistry is a multivariate problem*. The combination of defensive chemical compounds, nutrients, and structural carbohydrates in plant tissues that effect feeding preferences are unique to each hybrid genotype. The concentration of any one of these elements alone may not have a transferrable interpretation as the concentration responsible for an equivalent level of palatability in other poplar taxa. Each phytochemical constituent is perceived holistically in context with the others, integrated into the animal's assessment of tradeoffs between nutrient acquisition and digestive discomfort.
- *Phytochemistry is dynamic*. Levels of foliar defensive phytochemicals and animal nutrients are constantly changing throughout the growing season and from year to year. Phytochemicals are highly variable with respect to leaf ontogeny, position on the tree, and environmental conditions. Within a growing season of experimental testing a plant's chemical composition and the herbivore feeding responses will be different. While some of these variables can be normalized in standardized sampling protocols, all measured analyte concentrations (and their consumption) have a timestamp to their relevancy.
- *Herbivory and plant responses are mutually transitory.* Plants respond to damage by upregulating defensive compounds; herbivores respond to defensive compounds by modulating intake and upregulating neutralizing biochemicals. The responses of future progeny of both plant and herbivore operate in a context of potentially long-term epigenetic modifications. In essence, there are no 'fixed effects' that fit nicely into statistical models, nor random effects that can be confidently estimated.

These are the dynamic and inescapable obstacles we and others have encountered in attempting to develop NIR applications for quantifying pest resistance. At the beginning of the project we identified several knowledge gaps worthy of exploration, which could lay the foundation for future research in mitigating some of these barriers.

- *Discovery*. Variation of important 'defensive' phytochemicals and nutrients among and within hybrid poplar taxa being considered for commercial development needed further characterization.
- *Herbivory model development using known compound concentrations*. There were no known cases where NIRS-estimated concentrations of phytochemicals in poplars were used as predictor variables in a multiple linear regression model to predict herbivory. While NIR calibrations for estimating the concentration for single chemical constituents were feasible, how these estimations were to be used to predict herbivory needed further exploration.
- *Empirical herbivory model development*. Relatively recent research by others demonstrated that empirical NIR-based calibrations of foliage consumption were feasible (based on no-choice feeding studies or semi-quantitative determinations of plant damage in the field). No one had evaluated this method with herbivores of hybrid poplar.

Method overview

NIRS calibrations for an individual chemical constituent are commonly developed by associating laboratory measured concentrations of an analyte chromatographically extracted from plant tissue (the reference data) with covariate amplitudes of photon absorption across the near infrared spectrum using partial least squares (PLS) regression multivariate analysis (6). The resulting regression model predicts analyte concentration in future samples. The investigator must draw inferences about how model-estimated concentration impacts herbivory from other experimental data sources.

To be clear, these regression algorithms do not constitute a *calibration* in a strict sense. It is impossible to directly manipulate the concentration of individual analytes in wood, bark or foliar tissue in a manner similar to what one might use in generating a spectroscopic standard curve for a purified compound. In our case, analytes of interest are discovered in context with all other NIR-absorbing compounds of the material. Variation in laboratory-measured analyte concentration among different genotype samples facilitates the range of values needed for developing the regression algorithm. The algorithms are therefore generating a set of *correlation* coefficients that map NIR absorbance values that co-vary with the range of laboratory measured reference concentrations – but in a complex, multidimensional way. In this document we use the term *calibration* loosely to mean this quantitative (correlative and statistically validated) process.

NIRS calibrations can also be developed by associating NIR absorbance patterns with qualitative or quantitative attributes of the subject that result from variation in chemical composition which is unknown, or from chemical interactions that are not quantifiable. For example, NIRS has been used to 'bar-code' logs of different tree species arriving at a processing facility (7), predict the harvest-readiness of wine grapes and fruit (8) (9), and estimate the bending properties of lumber (10) (11). Using this second approach for our objective it was assumed that each poplar genotype has a unique assortment of NIR-absorbing compounds, which to the herbivore are simultaneously attractive, repulsive, or neutral, and act *additively* in the herbivore's decision to feed. By extension it is also assumed that each plant genotype has a unique NIR spectral fingerprint related to its phytochemical content. Experimentally, the critical measurement is the amount of foliar or bark consumption from each poplar genotype by the herbivore, but in a comparative way. The goal of this second calibration approach is to associate variation in each genotype's NIR spectral fingerprint to its empirically derived palatability ranking in order to predict the likelihood of future consumption. While this empirical approach cannot provide insight into a causal relationship between herbivory and a specific phytochemical, it is a potentially useful method to coarsely identify poplar genotypes having low palatability to pests. In addition, specific NIR wavelengths found to be highly correlated with herbivory can also provide information about specific NIR-absorbing chemical functional groups found in broad classes of compounds, which may support hypotheses about classes of phytochemicals that play a role in pest diet preferences (12).

Scientific questions explored

Our pest resistance investigations were predicated on knowledge that poplars and willows have naturally high concentrations of phenolic glycosides (PG) and condensed tannins (CT) (13) (14), but there were few specific reports documenting the ranges of concentrations found in the commercially important hybrid poplars being developed in the northwest and western regions of the US for bio-products feedstock. One of our primary objectives was to discover the range of variation in PG and CT concentration among these poplar taxa, and within individual genotype

leaves and bark as a function of environment and plant development. Investigations were also directed toward exploring the effect of interspecific hybridization on the levels of these compounds. Using these laboratory measurements, we sought to develop NIRS-based calibration algorithms to estimate CT and various PG concentrations for future screening operations.

A third objective was to determine whether and how NIRS calibrations could be used in identifying low palatability genotypes in progeny screening operations. Choosing genotypes from our sample population having high, low, and intermediate PG and CT concentrations we conducted controlled feeding studies with deer and voles in cafeteria-style multi-choice experiments in hopes of deriving a quantitative relationship between palatability and the concentration. Our hypothesis was that if a significant portion of variation in palatability could be explained with a linear or multiple linear regression model, that NIRS-estimates of CT and PG concentrations could be substituted into the models to predict palatability in new poplar samples with acceptable accuracy and precision.

Taking the empirical approach to NIRS-based modeling, and starting from a palatability study in which cottonwood leaf beetles were presented with 23 poplar genotypes in three hybrid taxa in pairwise tests, we sought to determine if a direct calibration of leaf consumption was feasible using the relative leaf area consumed as the palatability ranking variable for each genotype, and the NIR absorbance profile of the genotype as the set of independent variables. Additionally, by examining common absorbance patterns in the consumption calibrations we sought to determine which chemical functional groups were associated with consumption, and by extension, which compound classes might be important in diet choices.

Supplementary to the empirical calibration effort with cottonwood leaf beetle, we measured the proportion of total phenolics in foliar extracts that were oxidized in alkaline conditions, which were hypothesized to give rise to reactive oxygen species that can damage the insect midgut. We sought to determine if a NIRS calibration for proportion of total oxidized phenolics was feasible, and whether a significant correlation existed with observed herbivory.

Throughout our studies we addressed issues about standardizing sampling methods and approaches in context with the conventions used by our industry collaborator in evaluating new poplar hybrids in staged performance trials, and relative to the commonly encountered pest problems. We sought ways to simplify the acquisition of NIR spectral data, and provide training for non-expert industry personnel in the use of the instrument and calibrations in making progeny selections.

Results and implications of the research

Objective 1&2. Over the project period, NIR spectral data were gathered from more than one thousand leaf and bark samples collected from genotypes in 5 poplar species and 3 hybrid taxa. Nine hundred fifty two were analyzed for the concentration of condensed tannins and 5 prominent phenolic glycosides. Using these data, 34 single-constituent calibrations were made for predicting concentration in the foliage, 25 calibrations were made for predicting concentrations in bark tissue. All calibrations were considered minimally successful if they had estimated prediction R-square > 0.74. While this might sound impressive, we were only able to generate minimal accuracy calibrations for a third of the prominent foliar phenolic glycosides in all species and hybrid taxa. While sample sizes that were available to us in individual taxon were as large as or larger than some NIRS calibrations reported in the literature, in many cases they were not large enough to generate accurate and precise calibrations. Attempted calibrations for individual analytes generated from samples belonging to multiple taxa or from samples in hybrid

taxa were plagued by interference problems, and frequently had measured versus prediction R-square values < 0.7. Additionally, concentrations of salicin and tremulacin were found to be near or below the reporting limits of chromatographic separation for most genotypes. Calibrations for other phenolic glycosides such as salireposide and HCH-salicortin were infeasible in most taxa even though they were above chromatographic detection limits and sufficiently abundant. The reasons for this are not clear. Researchers at other institutions were also unsuccessful in generating NIRS calibrations for salireposide in *Populus tremuloides* (15).

We observed low precision from calibrations for PGs in which the sample population was comprised of hybrids or multiple taxa, but markedly improved precision if the sample population was large and comprised of a single taxon. One theoretical explanation for this phenomena is that for any taxon sample set there are different NIR-absorbing compounds having functional groups similar to the reference analyte, which may be collinear in abundance and have co-varying amplitudes of NIR Absorbance. Each taxon has its own idiosyncratic array of these unrelated absorption bands that co-vary with the reference analyte. When there are hybrids or several taxa represented in a calibration sample population, a complex mixture of idiosyncratic absorbance bands can present itself as interference or 'noise' that cripples the accuracy and precision of the calibration. These findings do not bode well for the using 'universal' multi-taxa calibrations for measuring PG compounds, especially for a genetically divergent population of poplars that are continuously interbred for new varietals. The prospect of acquiring adequate sample populations for developing individual calibrations for every PG in every poplar species and hybrid taxon seemed impractical and cost-limiting.

While single-taxon and multi-taxa calibrations for CT were consistently more precise ($R^2 \approx 0.93$) than the PG calibrations, the shortcoming these estimations are that they cannot estimate concentration of natural forms of CT which are a structurally heterogeneous class of flavonoid compounds whose secondary and tertiary structure imparts its biological activity (16) (17) (18). Rather, the calibration estimates the concentration of colored cyanidin subunits, products of acid-hydrolysis of extracted soluble proanthocyandin, spectroscopically measured at 550 nm to produce the reference data. These calibrations were incapable of discriminating short oligomer proanthocyanidins from biologically active CT comprised of long and often branched chains of cyanidin subunits per molecule. The user of these calibrations cannot infer that a poplar genotype with a high cyanidin subunit concentration would have low palatability or contribute to reduced protein assimilation. However, one might infer that genotypes with very low cyanidin concentration also have low levels of CT regardless of structure and activity.

Not surprising, we also observed spatial and temporal variability of analyte concentration with respect to tree age, leaf age, and environmental effects on plant development; making the attribution of concentration to be a genotypic quantitative trait problematic. Even with defined sample collection methods to normalize concentrations with respect to ontogeny, seasonal development, and tree age, the variability in concentration was greatest where the herbivory is often observed – at young, rapidly expanding shoot tips on young stems. The implication of these findings is that to account for this high variability, the sample size from each genotype would need to be quite large to detect significant differences in concentration among genotypes. In a progeny screening operation, breeders typically clonally propagate a genotype to large numbers for performance testing only after two previous selections for comparative growth were made. To obtain adequate sample size for using an NIRS concentration calibration, only a few of the initial progeny determined to have desirable high-growth attributes would be left for screening. This calls into question the utility of NIRS in facilitating "high throughput screening".

Objective 3. NIRS-calibration estimates of 'defensive' compound concentrations have little practical meaning apart from supporting evidence of their effect on herbivory derived from surveys of plantation damage or from controlled feeding studies. Multivariate regressions explaining consumption as a function of analyte concentration provide the meaningful link. However, palatability testing is very difficult to conduct at a scale that accommodates feeding variability among individual test subjects. This is especially true with wild mammalian herbivores, where the effects of the subject's experiential learning can result in changing diet preferences over the test period. Use of wild animals also requires capture permits, specialized facilities, and animal welfare oversight, which can limit how experiments are conducted. Our restricted-scope studies with deer and voles, characterized by having relatively few test subjects participating in short duration tests with no diet restrictions, *ad libitum* access to poplar test material, and no invasive procedures to examine physiological effects, aimed to generate analyte concentration-consumption regression models. If strong correlations were detected using the laboratory measured concentration data, we hypothesized that substituting NIR-estimated concentrations in the model might be accurate and precise enough for predicting herbivory.

Our controlled feeding tests with captive voles and captive deer demonstrated that there were no strong concentration-consumption relationships for CT or any individual PG analyte. Test results of bark consumption by voles using eighteen poplar genotypes presented in three groupings, repeated three times sequentially, indicated that as much as 39% and as little as 4% of the variation in consumption could be explained in a multiple linear regression by various combinations of analytes, depending on the genotype grouping. The feeding tests with two groups of confined deer, in which eighteen poplar genotypes were presented in two groupings of nine genotypes - each group repeated three times sequentially and analyzed with multiple linear regression, found that 79% to 98% of leaf area consumption could be explained in a multiple linear regression by combinations of analytes, which differed based on the genotype grouping. However, average genotype leaf area on its own explained 47% of consumption, suggesting that deer may be economical browsers by maximizing bite size. Leaf area was not strongly correlated with either CT, PGs, crude protein content, neutral detergent fiber, or total phenolic content, however, it was suspected that moisture content was greater (per bite) in the larger leaves.

A major shortcoming of the controlled feeding tests was that experiments were set up in a cafeteria-style design, where test subjects were presented with multiple genotype choices in groupings of 6 or 9 per test. At no time was a genotype presented in the first group combined with a genotype in the second or third group, which leaves to question whether the proportion of leaves or bark consumed from any genotype was influenced by the context of genotypes presented. This uncertainty was supported by the observation that the predictor variables in each multiple linear regression for each genotype grouping were unique. Adding to this uncertainty is the fact that some phenolic glycosides share common molecular precursors and biochemical pathways. Variation in abundance of any compound is likely to be collinear with the concentration of another in any particular genotype, making the predictor variables in any model interacting and potentially confounding.

In 2014 we were given an opportunity to conduct a deer browse damage survey at the GreenWood Resources tree farm in Boardman Oregon in which a growth performance trial comprised of 16 genotypes in 3 hybrid taxa in their second year of growth had been browsed by wild deer. Each genotype was replicated in 9-tree plots, repeated in 4 blocks. The browse damage appeared throughout the trial site. We rank-scored the damage on all 576 trees and collected leaf samples from all genotypes for laboratory PG analysis at the same lab that

produced the reference data from the NIR-based calibrations. Multiple linear regression analysis indicated that 68% of average browse score could be explained by the levels of total phenolics, HCH-salicortin, salireposide, and an unknown compound which was quantified chromatographically. Unexpectedly, CT and salicortin, both abundant and believed to be palatability deterrents, were positively correlated with increased browse damage. Like the captive deer study, genotypes with the largest leaves were browsed more heavily.

The controlled feeding studies and the deer browse damage survey left us with very little confidence that any common combination of phytochemical variables could sufficiently explain feeding preference in all taxa groups. That fact, coupled with our inability to generate accurate and precise NIR-based calibrations for many phenolic glycosides in hybrid taxa, and the compounding of errors when NIRS-estimated concentration values were substituted into a multiple linear regression model, meant that we would abandon this approach to using NIRS technology.

We then turned or efforts toward evaluating methods for testing the second NIR calibration approach in which NIR spectral absorption patterns were empirically associated with foliar consumption, which we did using an experimental system with cottonwood leaf beetles as test subjects. The poplar genotypes selected for the study were based on field observations of contrasting beetle damage at the GreenWood Resources poplar progeny trials and nursery at Boardman Oregon in the summer of 2014. Dormant branch cuttings were harvested in January 2015 for clonal propagation at the WSU Puyallup Research Center, and used for controlled feeding tests in the spring and summer of 2015. Twenty-three genotypes belonging to 3 hybrid taxa were used in the feeding tests and NIRS scans. For the purposes of verifying NIR-based consumption calibrations we also collected foliar samples from 68 genotypes based on their range of ranked damage from the same trials in 2014.

The consumption reference data was derived from averaged pairwise comparison tests in which we provided 3 beetles with freshly harvested shoots of two genotypes for 24 hours, then quantified the proportion of total leaf area consumed from each using image analysis. Each pairwise test was repeated three times. Because of growing season constraints and the limited availability of plant material needed to sustain the beetle colony, we measured average proportional consumption in each of the taxon group separately, then evaluated whether the palatability rankings aligned across taxa using separate feeding preference tests with select pairs of genotypes from different taxa. Finding that there was relatively good agreement, the reference data for all groups were combined into a ‘global’ NIRS consumption model, for which cross-validation predictions explained 72% of measured proportional consumption.

This study also revealed alarming findings about phytochemical variability among clones grown in identical conditions. We grew multiple clones of each genotype for feeding tests and for NIRS scans in identical conditions. When we used different samples of the same genotype for NIR scans the consumption model results varied widely. When we used all 3 scanned samples for each genotype in a global model (each associated with the same averaged proportional consumption reference) the precision was suboptimal. Before attempting to verify the global model using the 68 foliar samples from the progeny trail, we used principal component analysis to compare the NIR spectra of the verification samples with the model development samples. We found they were not sufficiently similar to be used for verification purposes. Even though they were same-aged leaves from the same taxa, and in some cases the same genotype, the spectral differences (chemical composition differences) resulting from tree age, apical versus lateral meristem leaves, environmental conditions of north central Oregon versus western Washington,

were too great. Perhaps this outcome is not surprising given the abundance of evidence in the research literature about foliar phytochemical variation in Populus by tree age and season (19) (20) (21) (22). The sensitivity of NIRS to these differences, and the down-side implications for using this instrument for screening phytochemical variables have not been discussed exhaustively in research articles that tend to over-emphasize speed and convenience of the technology.

To derive reliable information about the effects of specific 'defensive' phytochemicals on herbivory, or the combined effects of all unknown NIR-absorbing compounds in a leaf sample on palatability, requires large data sets derived from plant material possessing the maximum range of analyte variation, utilizing controlled feeding experiments with wild animals, all occurring within a narrow testing period restricted by seasonal variation in phytochemical abundance and relevant tree age. While it is quite possible to improve on our results by conducting more expansive tests over multiple years, what this study demonstrated is that NIRS cannot be recommended as a pest-resistance screening technology for hybrid poplar given the ephemeral, uncontrollable, and unreproducible nature of the variables.

~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~

Technical objective (*Abstract*)

We sought to extend the use of NIRS technology, used in a previous project for quantifying chemical constituents of poplar wood, to quantifying defensive phytochemicals in leaves and bark for the purpose of identifying naturally pest-resistant genotypes, improved for the production of biomass and bioenergy feedstock. The potential impacts of deploying pest-resistant genotypes is increased crop security, reduced input costs, and the reduction of negative environmental impacts resulting from pesticide application.

To develop this method we acquired chromatographically measured defensive compound concentrations in modest sized populations of poplar species and hybrids considered to be commercially important. From these data we developed numerous NIRS-based multivariate calibrations for estimating the concentration of condensed tannins subunits and several of the prominent phenolic glycosides, as well as for protein and total phenolics. To determine the practical value of these estimates, we conducted controlled feeding tests with voles and deer using rooted cuttings of selected genotypes from the same sample population. Our tests indicated that only a minor proportion of measured herbivory was explained by combinations of these compound concentrations in multiple linear regression models, and the modeled compound variables were not common among different taxa groupings used in the feeding tests. We then explored an alternative approach of empirically calibrating NIR spectral patterns to variation in leaf consumption by the cottonwood leaf beetle. From these tests we generated NIR-based consumption models, which were shown to have suboptimal precision due to uncontrollable variation in phytochemistry even among identically grown clones. The models could not be verified by beetle damage surveys in the field because of phytochemical variation due to leaf ontogeny, tree age, and environmental effects.

Our research demonstrated that while it is possible to generate passably precise calibrations for estimating concentrations of some of the abundant phenolic glycosides and condensed tannin, in our hands these quantities showed no generally predictable link to hybrid poplar palatability.

Specific Results and Accomplishments

Objective 1: Determine the range and variability of condensed tannins and phenolic glycosides in poplar hybrids

Milestone 1: Standardize a sampling protocol

Our first task was to quantify the extent of variation in defensive compound concentration as a function of leaf age and position on the tree to inform the development of a method for assigning quantitative phytochemical traits to any genotype in a coppiced tree plantation context. We focused our measurements on one and two-year-old trees for three reasons: breeders typically make early progeny selections for future performance trials in the first three years, apical leaves on 3-year old trees are not easily accessible, and coppice poplar feedstock for bioenergy was projected to be harvested in the second or third year of growth.

The initial CT and PG variability study begun in spring of 2011 was limited to four clonal replicates of five genotypes of *Populus* x *generosa* hybrids made available for sampling from a bioenergy performance trial at the GreenWood Resources nursery at Westport Oregon. We collected apical and lateral branch leaves from two replicates in late June, and from the other two replicates in mid-September. All trees were coppiced in the winter of 2011. We then sampled apical and lateral branch leaves from regrowth sprouts from the same trees on approximately the same dates in 2012. Leaves were sampled in four plastochron age groups, originating at the shoot apex, and beginning the first numbered leaf longer than 3cm. This was a less-refined method for representing morphological time scale than the leaf plastochron index (LPI) (23) – a leaf numbering system that computes the time interval between initiations of successive leaf primordia, and is based on accurate measurements of leaf length. While LPI (which can be expressed as rational numbers) more accurately represents the continuum of morphological and physiological development of each leaf, leaf number quantizes development by integers. We considered that for future sampling at larger scales by field technicians, collections would be expedited by using the simplified leaf numbering method.

We found that CT varied dramatically within the growing season and from year to year, dependent on leaf ontogeny and position on the tree. (The methods of CT and PG analyte extraction and quantification are described in the appendix). Within a growing-season, CT levels in same-aged apical meristem leaves of 2 year-old trees in spring were approximately 2-fold higher than lateral meristem leaves, but by summer, lateral leaves had comparable or greater levels as apical leaves. Within genotype clones CT concentrations were similar in leaf numbers older than 10 and 11, and slightly lower in leaf numbers 5 and 6 (Figure 2). Among different genotypes we observed a 5-fold range of concentration. In emerging shoots of coppiced regrowth, same-aged apical leaves have only 33-50% of the CT levels in spring as they have in summer, which then become comparable to levels found in 2 year-old trees (data not shown). In contrast, we observed that PG concentrations were generally higher in the youngest leaves numbered 5 and 6 compared to more mature leaves, but this was not reliably consistent (Figures 1 & 2). Salicortin, which is the most abundant of the phenolic glycosides in this hybrid taxa, is present in lower levels in young spring leaves, and increases as the growing season progresses. In contrast, salicin is highest in spring, low in summer, and is fairly consistent in all leaf age-classes.

The implication of these findings, bearing in mind the low replication number, is that CT and PG concentrations are dynamic, which requires a standardized date during the growing season to normalize temporal effects in sample collections year-to-year.

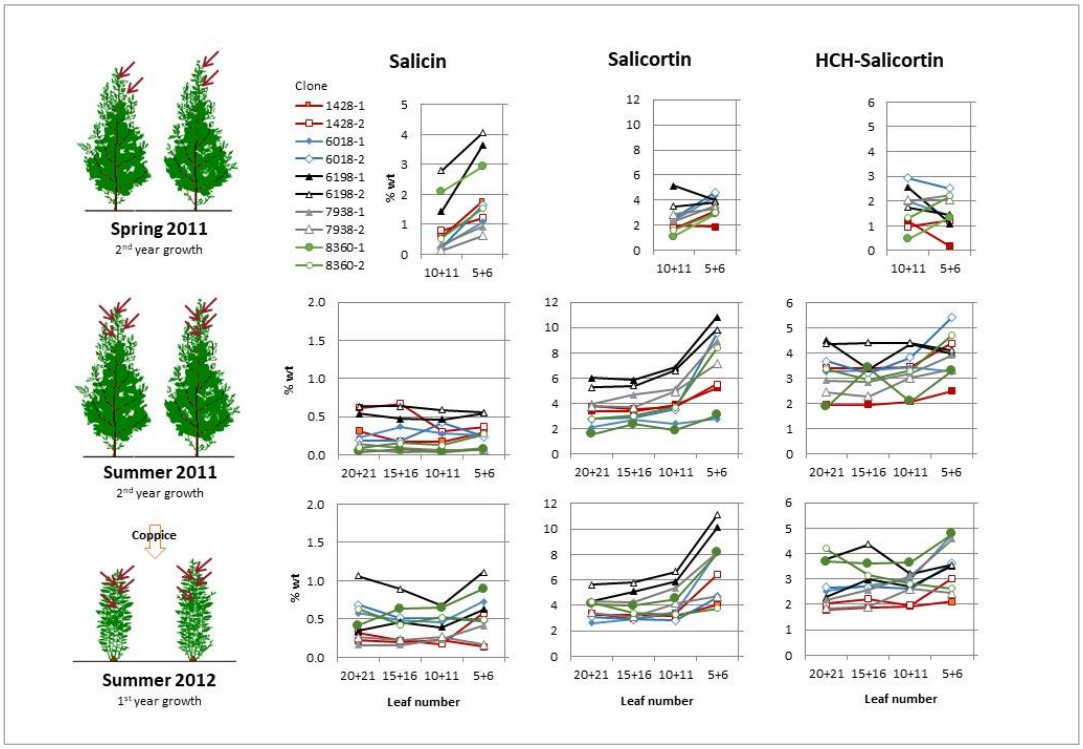


Figure 1. Concentration of 3 phenolic glycosides as percentage of dry sample weight, sampled from apical meristem leaves from 5 *Populus x generosa* genotypes in 4 plastochron age groups

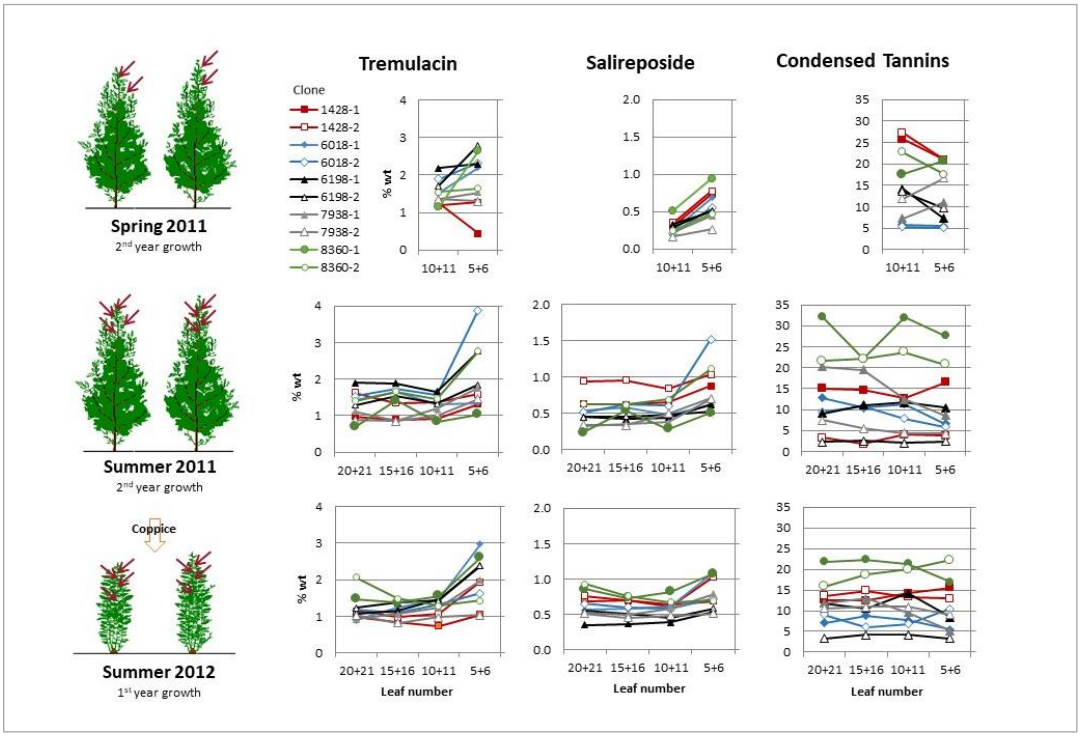


Figure 2. Concentration of 2 phenolic glycosides and condensed tannin as percentage of dry sample weight, sampled from apical meristem leaves of 5 *Populus x generosa* genotypes in 4 plastochron age groups.

While it may seem prudent to recommend that CT and PG concentrations in older-aged apical leaves be used in assigning a concentration as a discriminating genotypic characteristic because the values are more consistent, the sampling method should also be developed in context with the pest problem. From field observations we have found that the cottonwood leaf beetle prefers the third and fourth youngest leaves, and for some genotypes, deer prefer the youngest leaves and shoot tips at certain times of the season. If we used leaf numbers 5 & 6 as the herbivory-relevant leaf age-class, PG concentrations would be higher but have considerably more variability. Our data show that the relative concentrations of PGs in leaf numbers 5 & 6, where leaf lamina are still expanding, are not always proportionally higher than the older leaves from the same genotype. The differences in phytochemistry between leaf 5 and 6 are greater than the difference between leafs 10 and 22. Sampling younger leaves would require a greater sample population to account for greater concentration variability, and a more accurate method to normalize the developmental stage of the leaf samples – like the LPI system. This would be impractical because in conventional tree improvement practice clonal replication of genotypes in numbers needed to account for this variability only happens in later staged progeny performance trials. At that point more than 90% of the original progeny would have been culled based on growth parameters. But perhaps this may align with breeding priorities, i.e. pest resistance being less important than rapid growth and wood chemistry. Additionally, the LPI system would require time-consuming accurate measurements of each leaf, which still may not guarantee that the leaves are at the same developmental stage, especially among genotypes in hybrid families where growth and development phenotypes are more variable. As will be discussed below, this is an important source of imprecision in NIRS-based calibration models.

From 2011 to 2014 we collected and analyzed an additional 498 leaf and bark samples, primarily from the GreenWood Resources nursery in Westport Oregon, and at the GreenWood research site and tree farm in Boardman Oregon, with the goal of characterizing the range of variation of CT and PG analytes among a subset of polar species and hybrid varietals in development for bioenergy and other wood products (Table 1), and to use the measurements as reference data in developing NIRS calibrations. Leaf and bark collection methods, and methods used in NIR spectral data collection are described in the appendix.

Table 1. Initial sample populations of Populus species and hybrid taxa collected for characterizing foliar phytochemical variation and NIRS developing calibrations in 2012

Taxon Symbol	Species or hybrid	Tree age (yrs.)	Sample origin (meristem)	Coppiced	n
T	*P. trichocarpa*	1	Apical	+	43
K	*P. koreana*	1	Apical	+	20
C	*P. cathayana*	1	Apical	+	20
M	*P. maximowiczii*	1	Apical	+	20
G	*P. x generosa* (TxD, DxT)	2	Apical + lateral	+	64
DM	*P. deltoides x P. maximowiczii*	2	Apical + lateral	+	29
N	*P. nigra*	3	Lateral	-	44
					240

Concentrations of foliar CT and PGs were highly diverse among poplar taxa (Figure 3). Most notably, the Asian poplars *P. cathayana, P. koreana,* and *P maximowiczii*, had higher levels of HCH-salicortin and salireposide than the North American taxa. Tremulacin levels were highest in *Populus trichocarpa*, and intermediate in the *trichocarpa x deltoides* hybrids. CT levels in *cathayana* and *maximowiczii* were low, but *deltoides* x *maximowiczii* hybrids had much higher levels. Salicortin levels were generally lower in older trees, whereas CT levels were higher.

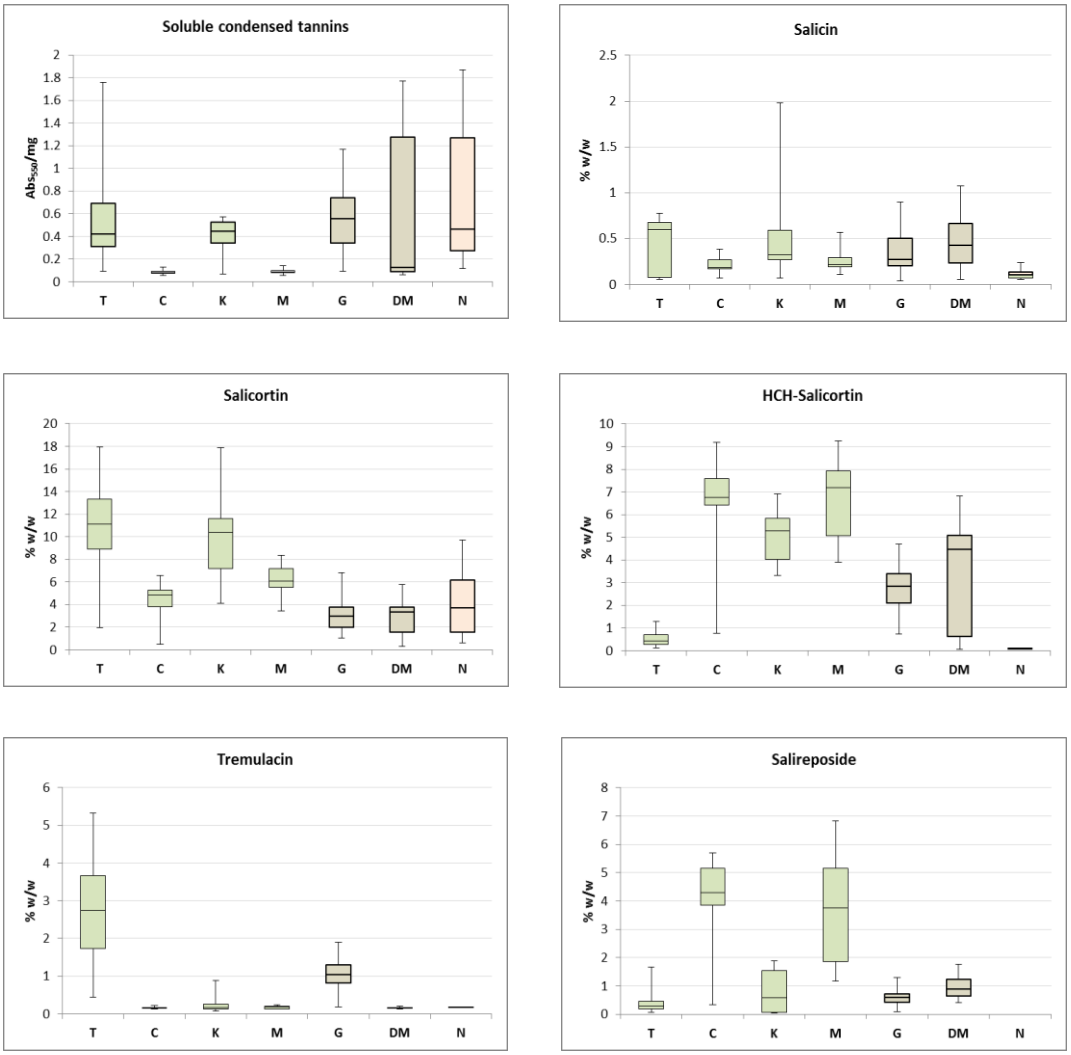


Figure 3. Quartile box plots of condensed tannins and phenolic glycoside concentrations in foliage of five *Populus* species and two hybrid taxa. Bar colors indicate tree age: green, 1 year-old; brown, 2 year-old; orange, 3-year-old.

In a side study comparing the CT and PG levels in summer apical leaves of one-year-old *P. trichocarpa* genotypes with lateral leaves from the same genotypes on trees aged greater than 15 years, we found that mature trees had a 6-fold increase in CT and a greater than 13-fold decrease in salicortin content (data not shown). These results highlight the dynamic nature of these phytochemicals and suggest that, in an ecological context, age-related expression of these

compounds have been economized through natural selection in response to different herbivores or pathogens.

In the same study we collected leaves from three clone replicates of 14 genotypes of one-year-old *P. trichocarpa* and found that the average percent coefficient of variation of CT and PG concentration values among replicates was 19% and 17% respectively. The trees were confined within a two acre plot with relatively uniform soil and environmental conditions. These findings suggest that to account for this kind of variability in a way that facilitates the distinction of a 5% difference in CT or PG concentration among genotypes with statistical confidence would require collecting large sample populations ($n > 48$) typically unavailable until final stage progeny performance trials.

Collection of most bark samples was restricted to one-year-old winter-dormant shoots scheduled for annual coppice at the GreenWood Resources Westport nursery. This strategy was adopted to minimize the impact of destructive sampling on progeny testing operations. We collected 5cm long shaved samples from the circumference of the stem at 10 cm from the base of the main leader and at 50cm from the base. We found diverse CT and PG concentrations in each poplar taxon (Figure 4), and relative abundance trends similar to those found in foliage. HCH-salicortin and salireposide contents were higher in the Asian poplars, while salicortin, tremulacin, and condensed tannins being more abundant in the North American poplars. We also found a curious abundance of tremulacin in the bark of *P. koreana*, which did not occur in foliage.

We found pronounced seasonal and positional effects on bark analyte concentration within the *P. nigra* and *P. x generosa* sample population where CT and salicortin decreased linearly as a function of distance from the shoot base in summer-harvested bark. In other experiments (discussed below) we noted large differences in CT and PG concentration in bark samples harvested in summer compared with those derived from winter dormant shoots from the same genotypes (correlation $r = 0.62$ to 0.95).

Overall findings for milestone 1:

- There is sufficient variation in CT and PG abundance to generate calibrations by combining levels from all poplar taxa, but calibrations within individual taxa would likely be infeasible for analytes having low abundance and variation.
- Leaf age, tree age, position on branch or shoot, and seasonal development status all contribute to variability, as has been well documented (19) (24) (25). While these variables can be normalized somewhat with a standardized sampling protocol, acquiring sufficient sample replication from individual genotypes to account for variability will be limited by the practices and priorities of the breeding operation.
- This survey of analyte abundance in leaves and bark was limited in scope and provenance: Nigra genotypes tested were derived from a few Italian provenances; Asian poplars came from only a few seed lots; most *P trichocarpa x P. deltoides* hybrids were derived from *trichocarpa* parents collected from the American Pacific Northwest.

- Sampling recommendation: leaves should be collected two weeks before the summer solstice from apical meristems on trees beginning their second year of growth. Approximately 40 samples, comprised of leaf numbers 6 and 7- pooled per sample, collected for each genotype, should provide sufficient material and account for non-genetic variability.

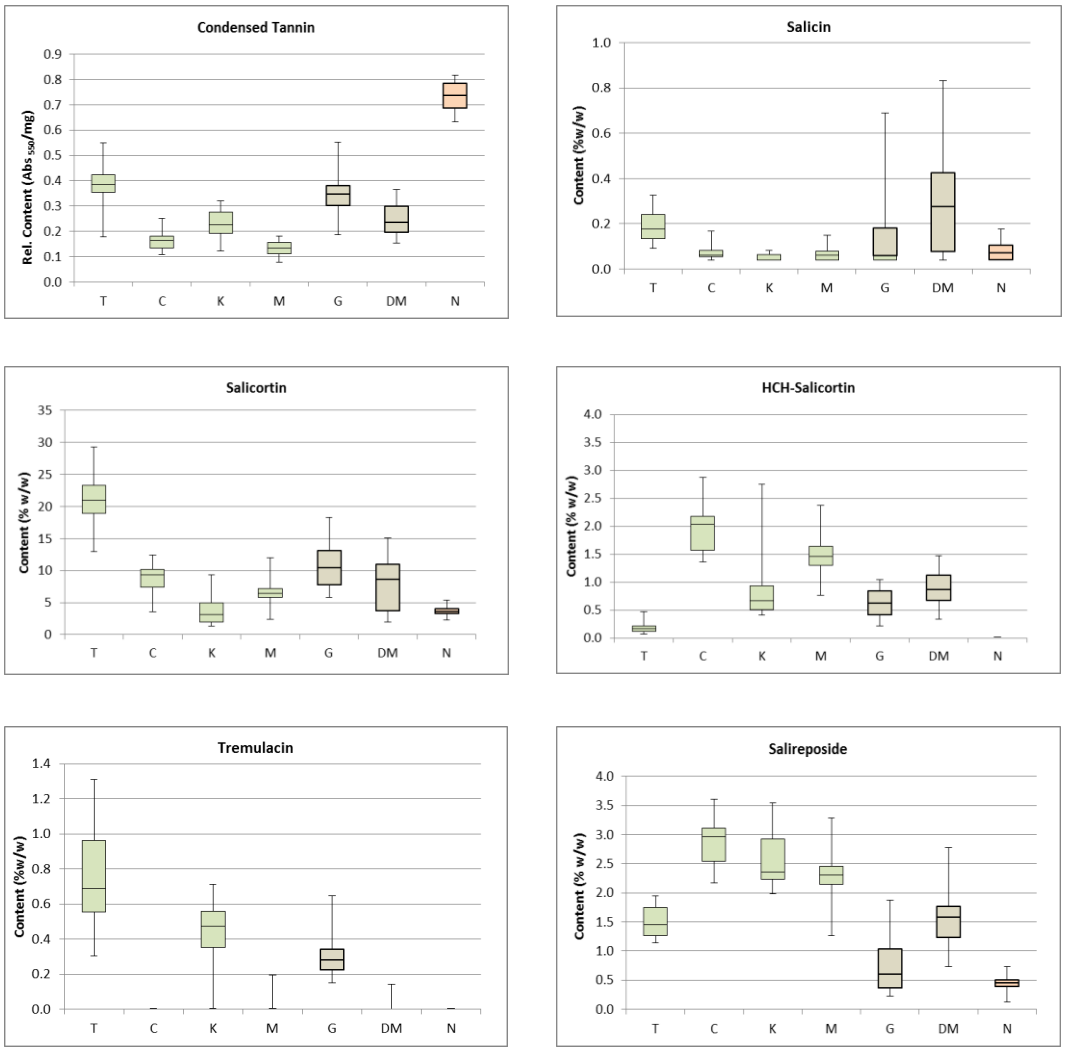


Figure 4. Quartile box plots of laboratory measured condensed tannins and phenolic glycosides concentration in bark of five *Populus* species and two hybrid taxa. Bar colors indicate tree age: green, 1 year-old; brown, 2 year-old; orange, 3-year-old. *P. nigra*, and some of the *P. x generosa* bark samples were harvested in summer 2011.

Milestone 2: Establish an NIR measurement protocol.

We determined that sample preparation and spectral data collection methods we previously developed for measuring structural carbohydrates in milled wood flour could be used for leaves and bark. These methods are summarized in the appendix. The major difference is that leaves and bark were flash frozen in the field, stored at < -40°C, and then lyophilized. All of the methods we developed were also described in detail in a NIR spectroscopy training manual we provided to GreenWood Resources (see objective 4).

One criticism of this method is that it is labor intensive, possibly to such an extent that it may discredit the notion that NIRS technology can be used as a "high-throughput" screening method.

While there are published examples of mobile NIR spectrometer measurements being used in the field in 'point-and-shoot' fashion, requiring no sample preparation; for example determination of eucalyptus species at a log receiving yard (7) or estimating wine grape harvest readiness (8), these calibrations utilize only a few select NIR wavelengths outside the range of the predominant absorption bands of water. In some cases, no specific compound concentration was estimated.

We conducted a study comparing the NIR absorbance profiles of live versus freeze-dried and milled poplar foliage, using the same leaves from the same plant, and found that the water bands around 1440nm and 1900nm overwhelmed absorbance variation that would be present if the samples were freeze-dried (Figure 5). Furthermore, live foliage scans resulted in considerably larger variability in NIR absorbance across larger portions of the spectrum compared to dried samples.

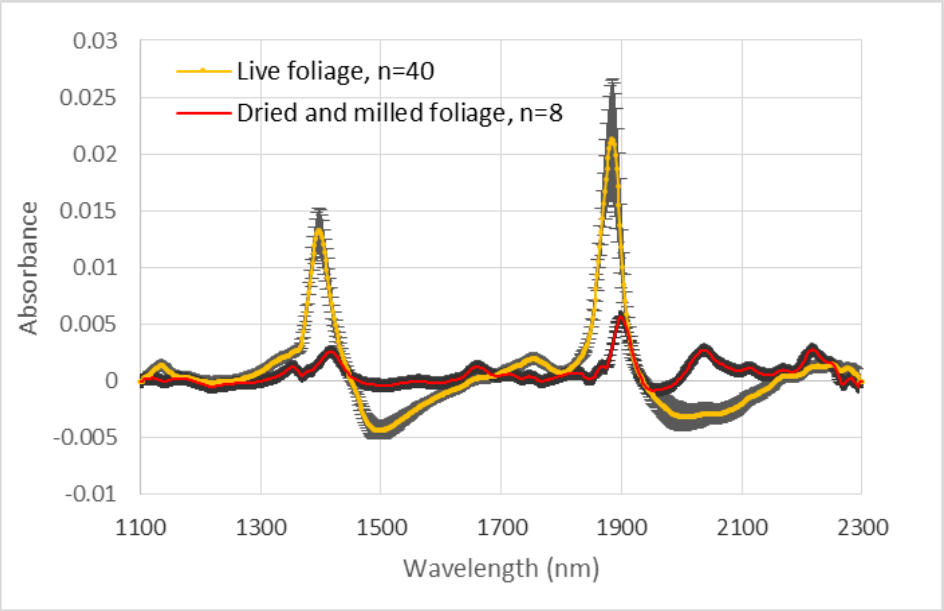


Figure 5. Average NIR absorbance and standard deviation of first-derivative transformed spectral data of the same leaf samples from the same plant measured live and then freeze-dried. n = number of NIR scans averaged.

The implications of these findings are that it is unlikely that accurate NIRS calibrations can be developed for estimating analyte concentrations using live foliage. It is also unlikely that acceptable empirical NIRS calibrations for palatability can be developed using live foliage samples. We estimate that it takes approximately 25 to 30 man-hours to collect, freeze-dry, mill, and scan a batch of 50 leaves. That does not include the 48 hour drying time in the lyophilizer.

Objective 2: Develop and validate NIR calibration models.

Introduction

Consider two general approaches to modeling and data analysis as it relates to this study. In the first approach an algorithm is proposed to capture the quantitative cause-effect relationship that is hypothesized to exist in some phenomenon based on *a priori* knowledge. The regression model is fit to the data by quantitatively adjusting parameters to optimize the fit. For example, to model the relationship between a foliar phenolic glycoside concentration and leaf palatability to a specific herbivore, one would regress experimentally measured leaf consumption data against laboratory measured phenolic glycoside concentration data in hopes of generating a linear or non-linear model. If the test subject and sample populations are sufficiently large and unbiased to account for natural variability, and if the data are strongly correlated, and if the statistics used to evaluate the degree of fit indicate the fit as better than expected by chance alone, there is high probability that the algorithm will be useful in making future predictions about herbivory by merely measuring (or estimating, using an indirect method such as NIR spectroscopy) the phenolic glycoside concentration.

In the second approach, data is analyzed to uncover hidden patterns related to the problem utilizing an inductive method, with less dependence on *a priori* knowledge, theory, and first principles. Take for example the use of NIRS chemometric methods to model the concentration of a specific phenolic glycoside compound in a dried and milled leaf sample based on the NIR absorption spectrum of all the samples collectively. Over a thousand NIR absorbance data points exists across the full spectrum for every leaf sample. Each sample likely contains hundreds of NIR-absorbing compounds, each with a unique NIR absorption profile. Multivariate data analysis, specifically, partial least-squares regression (PLS) enables us to estimate concentration of just one of the compounds within the mixture by identifying patterns of covariation between the NIR spectral absorbance values at every wavelength (the variables) and the laboratory measurements for analytes of interest for all samples (the objects).

Explaining how PLS regression is accomplished, either in geometric 'visual' terms, or in a higher-level statistical exposition in terms of matrix algebra, is fairly complex and is not appropriate for this report. In the interests of brevity the following overview highlighting the relationship between principal component analysis (PCA), principal component regression (PCR), and partial least squares regression may aid in understanding our modelling results. The reader is encouraged to refer to the text by Kim Esbensen (26) *Multivariate Data Analysis – in practice* for an introductory and comprehensive discussion of these topics; as well as these related monographs (27) (28) on the application of the PLS algorithm in NIR spectroscopy and chemometrics.

Principal component analysis is the workhorse of multivariate data analysis and a sub-algorithm of principal component regression and partial least squares regression. In PCA, the matrix of variable data (p) collected for a number of problem-related objects (n) is decomposed into a 'structure' part and a 'noise' part. The objective is to discover and model hidden or 'latent' phenomena. The operational assumption is that there is an underlying geometric coordinate system that describes the data, oriented along a projected axis of maximum variance or 'principal component', made possible by inter-variable covariance and correlation. The term 'matrix decomposition' means the PCA algorithm computes the original data matrix as a product of two matrices plus a residual matrix. One matrix, the 'score' matrix, contains all the information about the interrelationships among objects, the other 'loading' matrix contains all the information

about the interrelationships among the variables. The residual matrix is a measure of lack-of-fit relative to the total variance. The PCA algorithm, (more formally known as Non-linear Iterative Projections by Alternating Least Squares) is an iterative process where each subsequent principal component axis is, by definition, orthogonal to the preceding axis. For each iteration of the algorithm there is less and less variance to partition. To develop a PCA model, which is the set of variance-scaled orthogonal axes defining maximum variance directions, having a common origin defined by the 'average object', the optimal number of principal components (or algorithm iterations) is determined when the total residual variance is minimized relative to a preset threshold. As a result, there is a decrease in the "dimension" of variable space from 'p' (number of variables) to 'A' (the number of principal components) required to maximally explain the data structure.

Take for example a PCA model of NIR absorbance values for a group of poplar leaf samples. A plot of the scores of the objects projected onto the plane of the first two principal components (which are orthogonal) might show a distribution of clusters of objects defined by taxon groupings because of their highly correlated chemical constituents. One could further deduce from this plot which taxa are more chemically similar or dissimilar based on their 'distance' from each other. A plot of the third and fourth principal components may reveal object patterns much more difficult to interpret. For instance, sample characteristics such as leaf development stage, full-sun leaves versus shade leaves, and tree age might contribute important 'latent' factors in the distribution of objects. A plot of objects in the seventh and eighth principal components may not reveal any interpretable pattern - their distribution likely influenced by variability in instrument measurements and residual noise. While *a priori* knowledge cannot *prescribe* a quantitative relationship among objects projected in a plot of two principal components, it is certainly required for interpreting the PCA model results and making deductions about inter-object relationships. In our studies, PCA was used to assess whether samples used in our PLS calibration models were sufficiently chemically similar or whether they could be outliers that might result in model instability. Typically this was obvious with the first three principal components.

In a typical univariate calibration we relate a dependent property or response variable (Y), say, UV absorption, to an independent variable (X) corresponding to analyte concentration for each object using regression. In multiple linear regression several X variables are combined in linear combinations to find correlations to a single Y variable, with the required proviso that all X variables are independent. If there is collinearity among the X variables the regression is considered unstable. For example, the concentration of different phenolic glycosides in a single sample may be collinear because they are members of the same biochemical pathway, i.e. they are derived from a common molecular precursor, and, due to the reversibility of the biochemical reactions in the pathway, may be interconvertible. Including the concentrations of multiple phenolic glycosides from the same sample in a multiple linear regression may fortuitously result in a significant model, but such a model is not based on truly independent variables.

In principal component regression (PCR) the collinearity problem is addressed because PCA is first used to decompose the X data matrix into a matrix of score vectors that are orthogonal to each other (non-collinear). These are plugged into the standard multiple linear regression format in a second stage. Unlike PCA where modeling fit is optimized, PCR utilizes prediction error minimization methods to reduce the number of principal components. In this way collinear structured elements in the X matrix that are uncorrelated with Y are minimized or eliminated. The primary weakness of PCR is that there is no way of knowing if the information in the X data

matrix about the response variable Y sits in the first principal components of the model, that is, there is no guarantee that the decomposition of X necessarily produces the *only* structure correlated to Y. There may be other Y-correlated variance structures in the higher-order principal components that never make it into a model because other X-structure elements dominate the first principal components.

The partial least squares (PLS) algorithm addresses this weakness by modelling the X and Y space interdependently, actively connecting the X and Y spaces by interchanging (substituting) X and Y score vectors in the calculation of loadings. The Y matrix influences (directs) the decomposition of the X-matrix and vice versa, thereby maximizing the covariance between X and Y spaces. The 'components' in PLS are technically not the same as principal components used in PCA or PCR because they are derived differently, but both represent the latent dimension of the model. Like PCR, PLS regression uses prediction error minimization algorithms to derive the optimal number of PLS components.

There are a number of methods for validating a multivariate model, i.e. estimating its prediction performance. In an ideal situation where there is a large number of samples with accurate laboratory reference data, the samples can be split into a training set for calibration and a smaller test set for external validation. But as is often the case, economics restricts the number of laboratory measurements that can be made and there is insufficient sample size to comfortably allow splitting off a separate test set. In this situation internal cross-validation algorithms are employed. These algorithms prescribe an iterative process where one or a number of sample Y values are randomly left out of the error minimization computations, then uses the smaller proxy model to estimate the left-out Y values. When all of the samples have been left-out once, the collective differences between the estimated and measured Y values are calculated and expressed as the root mean square error of cross-validation. Goodness of fit of the predicted versus measured values are evaluated as usual in regression as the R-square of estimated Y values, where a R-square value of 1 represents perfect concurrence, and values less than 0.7 may suggest that such marginally correlated values are unreliable, and that the accuracy may be reduced to such a point that it may not be economic to deploy the model in operations; i.e. the cost of acting on the erroneous information exceeds the cost of measuring the sample directly by expensive chemical means. The data analyst is tasked with making these difficult, often vague, determinations.

We were restricted to using internal model validation methods in all of our developed models because of the relatively small sample sizes in each taxon group. Had our sample population exceeded 70 individuals in each species and hybrid taxon we could have utilized the more stringent external 'test set' validation method.

NIR calibration model results

Milestone 1. NIR spectral data collection and analysis, validate calibration models.

All attempted analyte calibrations were generated using PLS1 algorithm options in Unscrambler v.9.8 software. NIR spectral data was collected with a Brimrose 5030 Luminar AOTF spectrometer. Spectra were modified with various derivative and normalizing transformations, but primarily with Savitsky-Golay first derivative, seven or nine-point smoothing window, with center point estimated with a second-order polynomial. Models were validated with random-segmented or full cross-validation. Further details of our methods are described in the appendix.

The feasibility of developing some NIR calibrations for predicting phenolic glycoside concentration in foliage was initially encouraging. At the beginning of this project, PG

calibrations had not been widely demonstrated for hybrid poplar. In late 2012 our collaborators at the University of Wisconsin-Madison published their NIR calibrations for several phenolic glycosides found in *Populus tremuloides* and birch foliage (15). Most of our attempted calibrations were unstable, having calibration and cross-validation $R^2 < 0.7$. Only 44% of the attempted leaf phenolic glycoside calibrations with samples that included individual or combined taxa groups were successful, and only 35% of the bark calibrations (Table 2). Some of these models had artificially high validation R-squares due to bimodal distribution of concentration values and low sample number. Calibrations for salicin were not feasible in either leaves or bark of any taxon due to its low abundance in most samples. Only four calibrations for individual PGs and condensed tannins that included foliar samples from all taxa were feasible (Figure 6).

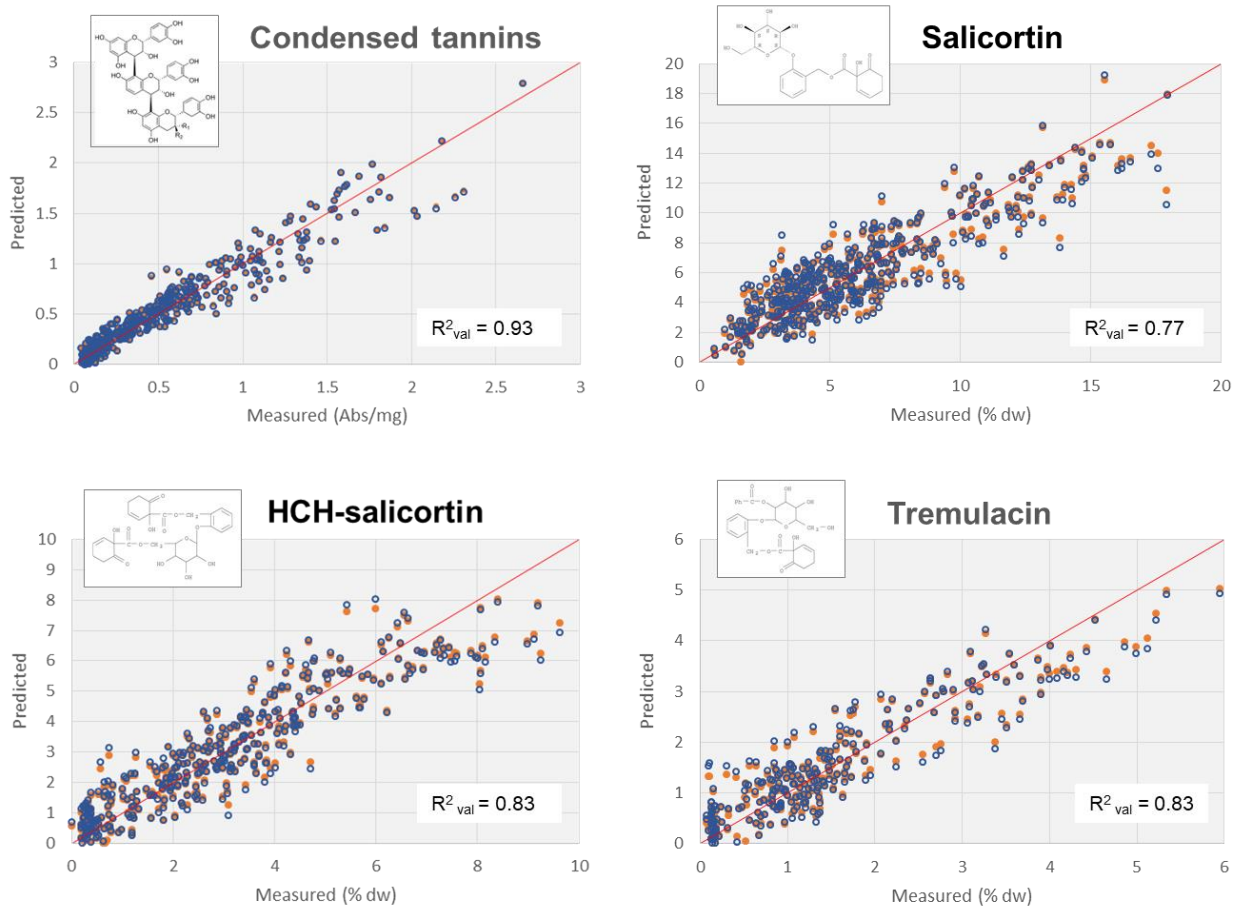


Figure 6. Calibration results plotted as laboratory measured data versus NIRS-predicted values for cyanidin (visible acid hydrolyzed product of condensed tannins), salicortin, HCH-salicortin, and tremulacin. Each of the calibrations utilize data from all hybrid poplar taxa and species in the study. The red line is the 1:1 correlation indicator. The precision of the calibrations are indicated with the R-square of the random-segmented cross-validation data points indicated in orange; the calibrated data are indicated with blue open circles.

Table 2. NIRS Calibration results for leaf and bark analytes

Cross-validation R^2 for leaf compound calibrations

Samples	Condensed tannin	Salicin	Salicortin	HCH-Salicortin	Tremulacin	Salireposide	Total PGs
All Taxa	0.925	*	0.765	0.825	0.826	*	0.809
T	0.941	*	0.872	*	0.849	*	0.874
G	0.948	*	0.770	*	*	*	0.815
T+G	0.914	*	0.867	0.813	0.819	*	0.888
K	0.879	*	0.759	*	*	0.802	*
C	*	*	0.757	*	*	*	.802
M	*	*	*	*	*	*	*
K+C+M	0.934	*	0.794	*	*	*	*
DM	0.980	*	0.875	0.758	0.799	0.765	0.819
DN	0.919	*	*	*	*	*	*
N	0.956	*	0.961	*	*	*	0.958

Cross-validation R^2 for bark compound calibrations

Samples	Condensed tannin	Salicin	Salicortin	HCH-Salicortin	Tremulacin	Salireposide	Total PGs
All Taxa	0.931	*	0.871	*	0.750	0.803	0.874
T	*	*	0.842	*	0.798	*	*
G	*	*	0.715	*	0.776	0.926	0.791
T+G	*	*	0.808	*	0.901	0.852	0.905
K	0.996	*	0.782	*	*	*	0.791
C	*	*	*	*	*	*	*
M	*	*	*	*	*	*	*
K+C+M	0.833	*	0.810	*	*	*	0.736
DM	0.789	*	0.805	0.810	*	*	0.748
N	*	*	*	*	*	*	*

Cross-validation R^2 is the fraction of variance accounted for by the cross-validation predictions; a rough measure of the correlation between measured values and those predicted by the calibration. Taxa represented: *P. trichocarpa* (T); *P.* x *generosa* (G); *P. koreana* (K); *P. cathayana* (C); *P. maximowiczii* (M), *P.* x *canadensis* (DN); *P. deltoides x P. maximowiczii* (DM); *P. nigra* (N). Calibration models were attempted with single analyte concentrations in individual taxon groups, or groups of genetically related genotypes (T+G) and (K+C+M), or with all samples combined to generate a 'global' model. Models were also attempted using the cumulative sum of all PGs concentrations for each genotype in a taxon group (Total PGs). Calibrations that were unstable or had cross-validation $R^2 < 0.7$ are indicated (*).

Except for a handful of models, our 'universal' (multi-taxa) single analyte calibration efforts were largely unproductive, and the results illuminated the limitations of the method. Modeling successes were realized chiefly for analytes found in highest concentration. Across species and hybrid taxa the average measured salicortin concentration was approximately 5.5% of dry sample weight with maximums between 18% and 23%, and for HCH-salicortin the average was

approximately 2.5% with maximums between 5% and 10%. In contrast, average concentration of salicin was approximately 0.3% of dry sample weight, with maxima less than 2%. We could not produce a single calibration for salicin in any species or hybrid sample grouping.

Measured concentrations of tremulacin were absent or below statistical reporting limits in *P. nigra* genotypes and most of the *P. deltoides x maximowiczii* hybrids, but highest in *P. trichocarpa* clones (1% to 5% of dry weight), and intermediate in *deltoides x trichocarpa* (G) hybrids. Thus, the multi-taxa tremulacin calibration for tremulacin was represented by a smaller sample population comprised mostly of North American genera. Tremulacin calibrations in Asian species were not feasible, and only one calibration was produced for the hybrid *P. Deltoides x maximowiczii* samples.

Concentrations of salireposide were not detected in *P. nigra* genotypes and *P. deltoides x P. nigra* hybrids. About a third of the *P. koreana* genotypes also lacked salireposide. We were unable to develop salireposide calibrations for most sample groups or a multi-taxon calibration. Our collaborators at the University of Wisconsin were also unable to independently produce a salireposide calibration for *P. tremuloides* samples, even though it had measurable abundance that should have permitted a calibration (15). One possibility for this failure is that the range of concentration values may have been too narrow, violating the rule of thumb that the range of expected values should be at least 10 to 20 times the error rate of the laboratory reference method. A narrow range of variation essentially reduces the scale of the X-data loadings producing an overall effect of diminishing the X-Y data correlation.

While concentration and range were important limiting factors to our modelling success, taxon-specific idiosyncrasies of the NIR absorbance data structure may also have played a role in reducing the precision of the calibrations. There are structured elements in the absorbance data matrix that may result from properties of compounds in the sample that are chemically similar to the analyte of interest, with similar extinction coefficients, and collinear in abundance with the analyte, for which the PLS algorithm is incapable of discriminating. These interferences, with an accompanying decrease in prediction performance, became apparent when we compared the pattern of cumulative loading weights at each NIR wavelength in attempted salicortin calibrations using separate taxon sample groupings.

Figure 7 shows two plots, each with seven graphs superimposed, derived from seven calibrations for salicortin and seven for total PG. Each calibration is made with samples grouped by species or hybrid taxon, or all samples combined. The data points are the cumulative sum of loading weights corresponding to each NIR wavelength that result from the covariation of salicortin or total PG concentrations with absorbance values. For example, the data point at 1698nm in the salicortin calibration for *P. trichocarpa* (T) has a cumulative loading weight of -1050. What does a cumulative loading weight mean? Each iteration of the PLS algorithm, corresponding to a ‘component’ – like a principal component - generates a vector **w** - the ‘loading weights vector’, which represents the direction that simultaneously maximizes the X-variance and Y-variance in a conventional least squares sense. Loading weights show how each absorbance value contributes to explaining the salicortin concentration or total PG concentration data for each iteration of the PLS algorithm.

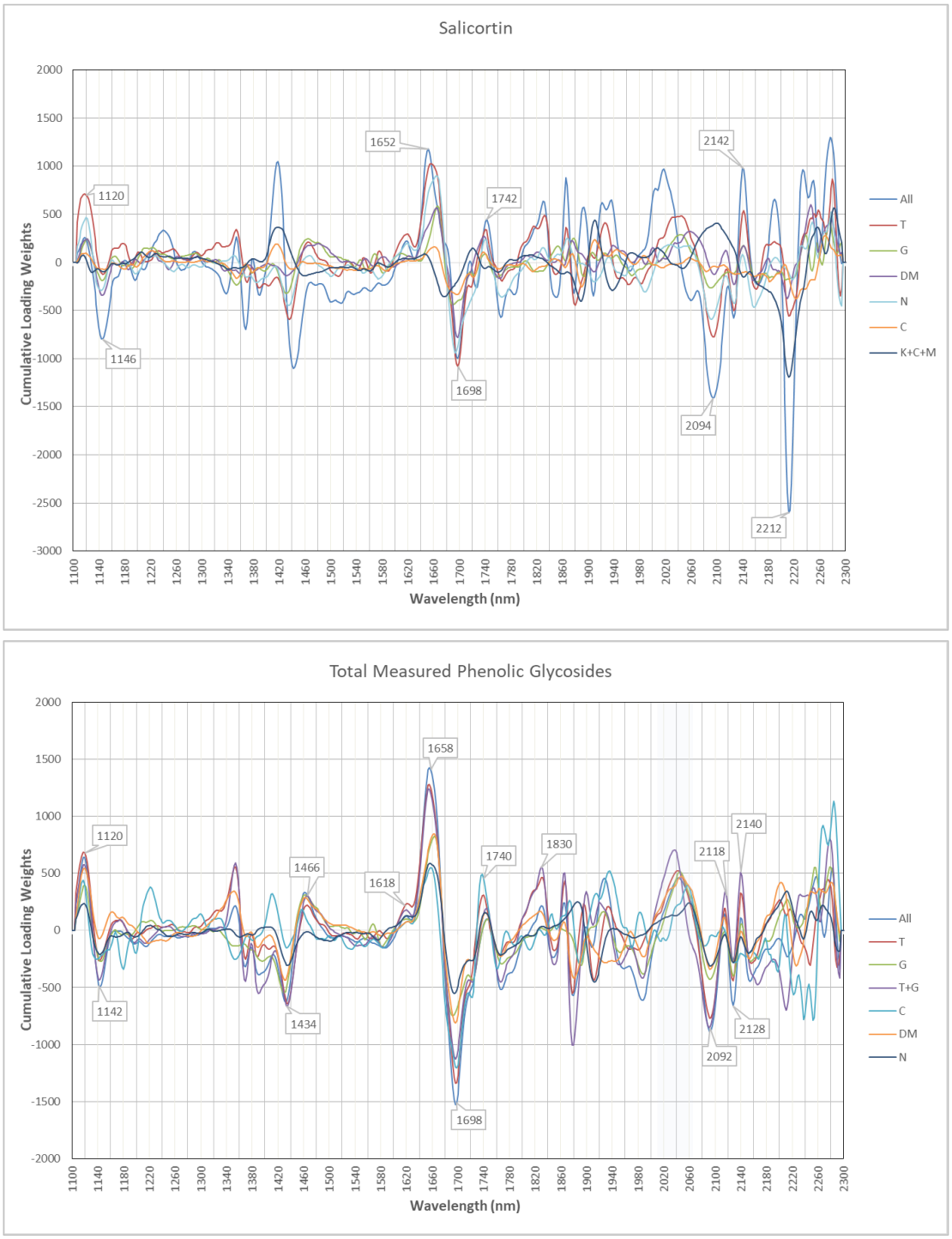


Figure 7. Cumulative loading weights at each NIR wavelength from individual calibrations for salicortin concentration and total phenolic glycosides developed with sample data sets restricted to poplar species and hybrid taxon. Taxa represented: *P. trichocarpa* (T); *P.* x *generosa* (G); *P. koreana* (K); *P. cathayana* (C); *P. maximowiczii* (M), *P.* x *canadensis* (DN); *P. deltoides x P. maximowiczii* (DM); *P. nigra* (N). The calibration using all samples indicated as (All)

The cumulative loading weight is a sum of all **w** vectors at each NIR wavelength. The graphs highlight which NIR wavelengths have absorbance values that are important in making a calibration.

There are some regions of the spectrum where the cumulative loading weights are similarly strong or trend in the same way in all the individual taxon calibrations for salicortin. For example the trends from 1120nm to 1146nm, and from1652nm to 1698nm are nearly the same for all calibrations except for the DM samples, where the trends are dissimilar for many of the common peaks. We might expect this throughout the spectrum if the NIRS/PLS algorithm method was sensitive and had highly specific detection. However, there are many regions of the spectrum where there doesn't appear to be any concurrence in peaks and trends among the calibrations. At some wavelengths the cumulative weights are positive for some taxa, and negative for others. The differences in cumulative loading weights between individual taxon calibrations and a calibration made with all samples are quite pronounced. These findings suggest that the covariance of NIR absorbance with salicortin concentration for each sample within a taxon group is either generally weak, and/or different among taxon groups. This does not mean that an individual calibration for salicortin can't be useful; it suggests that calibrations made with samples from multiple taxa may have reduced prediction performance.

For example, in comparing the salicortin model results shown in figure 6, derived from a sample population comprised of multiple species and hybrid taxa, with the model results shown in figure 8, where the sample population was relatively large but comprised of a single Populus *nigra* species from an Italian provenance, we see the precision of the calibration predictions was greatly improved when the samples came from a single species.

Curiously, when we used the mathematical sum of all the measured phenolic glycoside concentrations in each sample as the Y reference data, more calibration attempts were successful within taxon groups, and the modeling precision improved (Table 2). The rationale for combining the concentrations was to explore the possibility that certain NIR wavelength absorbance patterns might be common to all PGs since they share common chemical structures and functional groups. Furthermore, we knew that if NIRS-derived PG estimates were to be regressed against future assessments of palatability, we may not be able to use more than one PG estimate in a multiple linear regression model if we were to guarantee independence of variables. By combining the concentrations of all PGs we were accepting the biochemical interrelatedness of individual PG concentrations and putting emphasis on genotype differences in the bulk concentration of all PGs as a class of molecule.

Examining the cumulative loading weights for the total PG calibrations (figure 7, lower plot) we saw substantially more agreement in trends and peaks among the individual taxon calibrations than was seen for salicortin (figure 7, upper plot). The overall pattern between the two plots is similar because salicortin is the predominant phenolic glycoside. The fact that the calibration results for estimating total PGs improved may also point to the lack of sensitivity that NIRS has in discriminating fine chemical structure among related molecules in a non-purified sample. NIRS technology is commonly used to estimate overall protein content in plant material in commercial labs, but it is incapable of distinguishing complex tertiary structure of individual proteins. For example, distinguishing the concentration of a particular kinase enzyme from a structural protein.

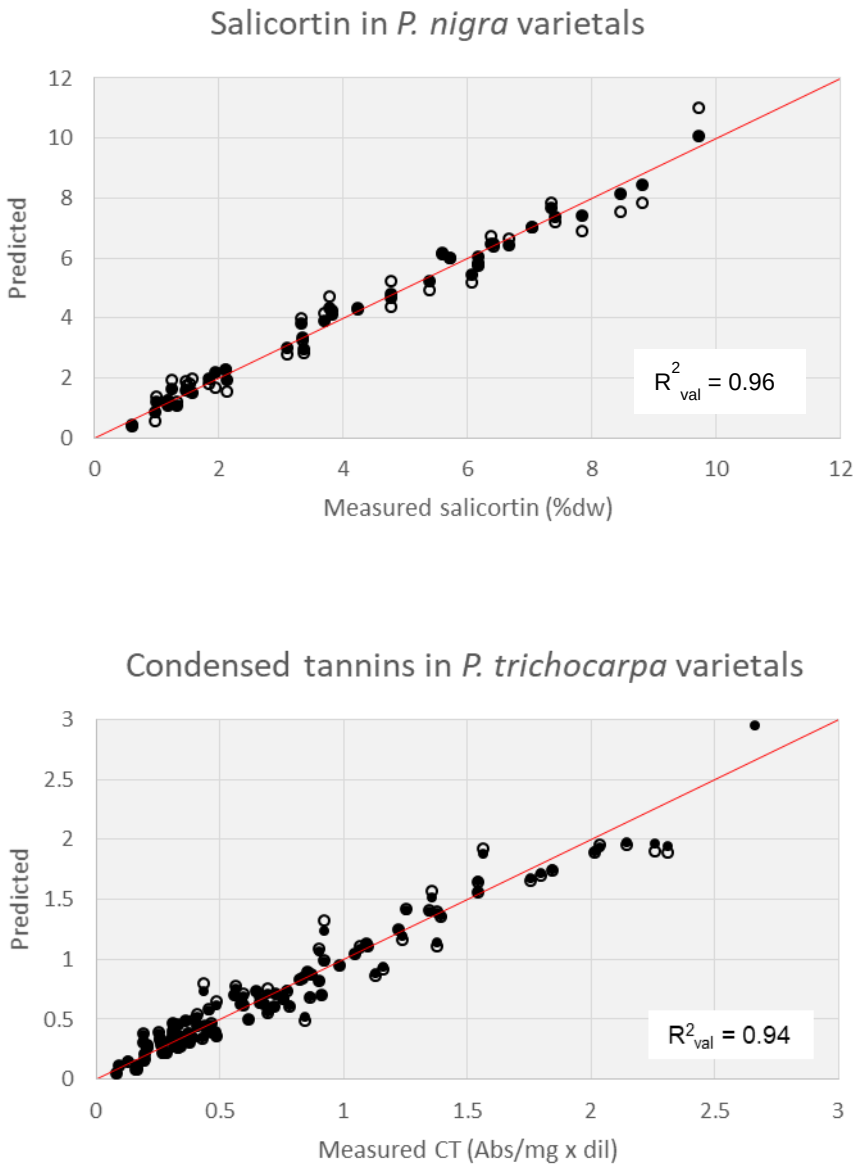


Figure 8. Calibration results for two poplar species, expressed as laboratory measured versus NIRS-predicted data values for salicortin in *Populus nigra* varietals and cyanidins (condensed tannins) in P. *trichocarpa* varietals. The precision of the calibrations are indicated with the R-square of the random-segmented cross-validation data points indicated as open circles; the calibrated data are indicated with solid circles.

Both of the analyte calibrations shown in figure 8 would be acceptable for clonal selection operations. However, they were generated with a sample population that came from a single poplar species. There were no sample idiosyncrasies in the absorbance data matrix coming from other taxa to confound the concentration-guided decomposition. While a single species calibration of any analyte might have a prediction performance that is useful, a calibration for a single hybrid taxon or multiple taxa would likely have poorer prediction performance. When our salicortin calibrations included samples from a hybrid taxon or multiple species the precision

suffered (see Table 2). For example, cross-validation R-square for salicortin in *trichocarpa* genotypes were 0.87, but models with samples from the hybrid taxon *P. generosa*, and for model comprised of all species and taxa combined had validation R-square of 0.77. Similar results were seen for total phenolic glycoside concentration. However, contradicting this observation was the finding that salicortin models for *P. maximowiczii* had a validation R-square less than 0.7, whereas salicortin models for the hybrid *P. deltoides x P. maximowiczii* (DM) samples had an R-square exceeding 0.8. These results were influenced by the fact that the two models were derived from very different sample sizes. The salicortin model for DM hybrids had 69 samples, whereas the model for *P. maximowiczii* had 20. Often, modelling attempts with 24 or fewer taxon-specific samples were infeasible or unstable. If the concentration data are distributed unevenly, being over-represented at the extremes, an artificially strong calibration can often result. Our general observation was that it was not practical to develop a calibration for each PG analyte using samples from each individual species or hybrid taxa.

Multi-taxon calibrations are feasible where the chemical structure of the analyte is fairly simple and is found in abundance. For example, condensed tannins are class of molecule comprised of flavan-3-ol subunits linked together to make structures of varying length and complexity (refer to the diagrammed chemical structure in figure 6). Specific types of condensed tannin are dependent on which functional groups are attached to the benzene ring, and how the flavan groups are linked, which varies among plant species. Each of the functional groups and structure of the flavan-3-ol can have a strong NIR absorbance signature, and based on the cumulative loading weights of our condensed tannin calibration (Figure 9), taxon-idiosyncratic influences on the absorbance data matrix resulting from differences in the functional group arrangements were relatively minor. All of the cumulative loading weight peaks and treads roughly concur in all of the individual taxon calibrations, and the cross-validation R-square of individual and combined taxa calibrations were consistently above 0.9 (Table 2).

To summarize our calibration findings and implications:

- The NIRS-PLS1 calibration method lacked the sensitivity to detect PG molecules in low abundance.
- Accurate and precise poplar species-specific calibrations were possible for analytes in high abundance (e.g. salicortin and total combined PGs in *P. nigra*), but it would be impractical to develop individual calibrations for each analyte in each hybrid taxa.
- Taxon-specific NIR absorbance idiosyncrasies compromised the calibration precision for PG analytes, making ‘global’ multi-taxon calibrations for estimating PGs problematic.
- Relatively simple molecules found in high abundance (e.g. condensed tannin) are well suited for calibrated estimation by NIRS-PLS1.

For the purposes of estimating analyte concentration in new hybrid poplar genotypes possessing unknown and potentially novel extremes of chemical variation that arise in a continuously interbred population of species, PLS calibrations, developed with a limited sample population and limited range of concentrations will be of limited use in spotting valuable extreme phenotypes in new progeny. Natural genetic variation and seasonal variability of tree phytochemicals will unavoidably plague the development of most if not all calibrations, resulting in low precision; perhaps too low to be used cost-effectively.

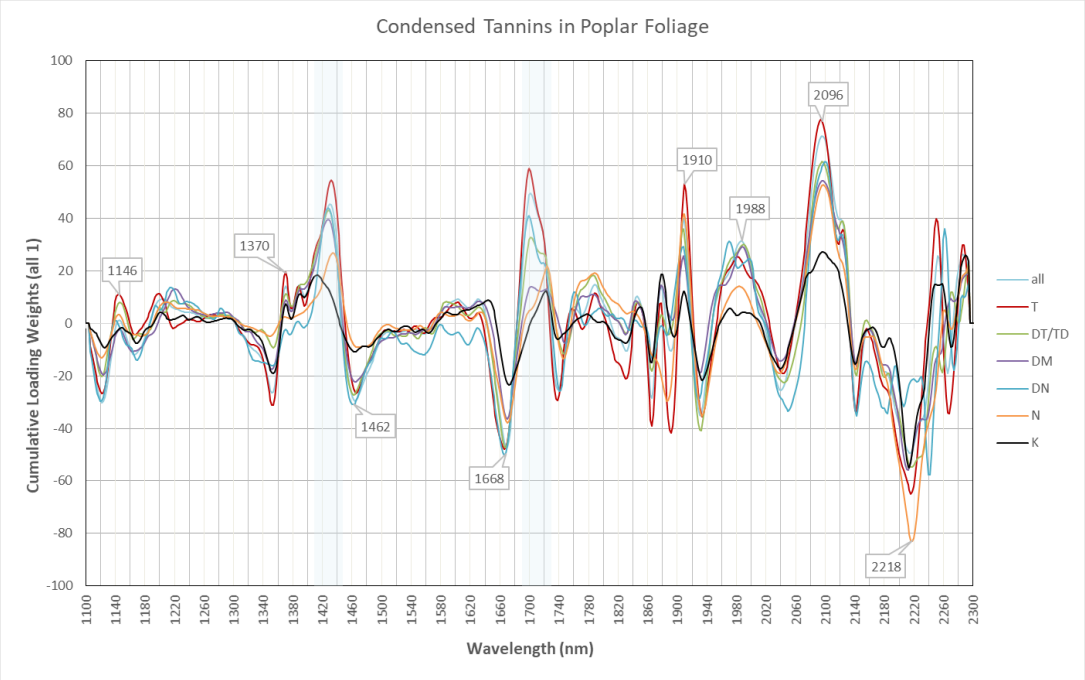


Figure 9. Cumulative loading weights at each NIR wavelength from individual calibrations for acid-solubilized cyanidins (proxy for condensed tannin content) developed with sample data sets restricted to poplar species and hybrid taxon. Taxa represented: *P. trichocarpa* (T); *P.* x *generosa* (G); *P. koreana* (K); *P. cathayana* (C); *P. maximowiczii* (M), *P.* x *canadensis* (DN); *P. deltoides x P. maximowiczii* (DM); *P. nigra* (N). The calibration using all samples indicated as (All)

These findings about the coarse sensitivity of NIRS set up a rationale for developing empirical calibrations, where the Y reference data result from the combined *effect* of all compounds in the sample on palatability to the herbivore and are mapped algorithmically to NIR absorbance patterns. Ultimately, for any NIRS single analyte calibration to have practical meaning, animal feeding tests must be conducted, which is precisely the starting point for developing the empirical 'palatability effect' calibration. These experiments are discussed in Objective 3.

Milestones 2 & 3. Standardize the NIR screening procedure, submit calibration findings for publication.

There are three distinct parts to developing a screening procedure: first, standardize the protocols for collecting leaf and bark samples in concert with the clonal propagation and testing regimes of a breeding program. Second, standardize protocols for processing samples to collect NIR spectral data, and prescribe spectral data transformations and optimizations to derive model estimations. Third; define procedures for using model estimations to identify new genotypes possessing pest-resistant traits.

The main challenge of the first part was described in detail in our findings under Objective 1. Natural variation in PG and CT concentration among same-aged leaves from different trees of the same genotype is substantial, such that, to ascribe a specific analyte concentration to a genotype with statistical confidence (normalizing for tree age, leaf age, and collection date) would require a number of samples that would typically not be available in early clone selection trials. It is only in the final staged clonal performance trial that genotypes are typically

propagated at sufficient scale to facilitate statistical sampling. At that point, the majority of the progeny resulting from a controlled breeding of two parent lines would have been culled in previous trials, and any genetic inheritance information to support proof-of-concept breeding with respect to pest resistance would be unavailable.

Methods for sample processing, spectral data collection, calibration development, and prediction analysis are found in the appendix. As discussed in the first milestone of Objective 2, none of the multi-taxon calibrations developed for phenolic glycosides were considered industry-deployable or noteworthy as a stand-alone publishable finding. While the multi-taxon calibration for condensed tannin is precise enough for estimating concentration, there is considerable debate in the research literature about the role of CT in pest resistance, and the uncertainty is largely attributable to the difficulty in accurately describing the tertiary structures of condensed tannin as they exist in any plant. Clearly, our calibration incapable of discriminating CT secondary and tertiary structures that likely contribute to its biological activity. As discussed below in Objective 3, we found that the concentrations of PGs and CT explained an insufficient proportion of observed herbivory in many controlled feeding tests with deer and voles using laboratory-measured concentrations.

Not having many precise calibrations for estimating PG concentrations in new genotypes undercut our prospects for generating peer-reviewed publishable findings related to the goal of using NIRS technology for identifying unpalatable genotypes.

Milestone 4, Measure the oxidation capacity of phenolic compounds in leaf and bark samples – determine if NIR calibrations are feasible

In 2014, extracts from 297 archived foliar samples from 5 poplar species and 3 hybrid taxa were used to determine the concentration of total phenolics, and the proportion of total phenolics that become oxidized in alkaline conditions. The hypothesis, initially proposed by Salminen and Karonen (29), suggested that oxidized phenolics and tannins may lead to oxidative stress in the digestive systems of insect herbivores. We sought to determine whether a NIRS calibration for oxidative capacity in extracted phenolics was feasible, and if so, determine whether a correlation exists between oxidative capacity and observed cottonwood beetle damage in a survey of a progeny trial and nursery at the GreenWood Resources Boardman Research Site.

We determined that the overall correlation between ranked beetle damage scores and the measured proportion of total phenolics oxidized in alkaline conditions using the Salminen and Karonen assay, was very poor (r = -0.006). While, we were able to develop a few acceptable NIR calibrations that estimate the concentration of pro-oxidant phenolics in individual taxon groups, calibrations for multiple-taxa or all taxa combined were infeasible, similar to our results for most PG calibrations. The best calibration for predicting pro-oxidant phenolics was developed from leaf samples harvested from *P. deltoides x P. nigra* genotypes at the GreenWood Resources Boardman Nursery (R^2 cross-validation = 0.863). In 2015 we deployed this model using the NIR spectra of DN genotypes used in a controlled study evaluating cottonwood leaf beetle feeding preferences. The regression of predicted pro-oxidant phenolic concentrations with measured leaf area consumption was poor (R^2 = 0.34). While generally there was a consistent negative correlation between calibration-predicted pro-oxidant phenolic content and observed herbivory, the strength of the correlation was not strong enough to be useful. Further explorations in this line of investigation were ended.

Objective 3: Verification - conduct controlled feeding tests with herbivores.

Milestone 1, Conduct controlled feeding preference tests with mammalian herbivores using presumed resistant and susceptible genotypes

Estimated concentrations of putative defensive compounds derived from NIRS calibrations by themselves are fairly meaningless numbers for the breeder without some reference to controlled studies that quantify level of herbivory as a function of analyte concentration found in relevant poplars. In 2013 we conducted controlled feeding tests with deer and voles, presenting them with poplar genotypes having high, low, and intermediate concentrations of total phenolic glycosides that we identified in the 2011-2012 calibration development data set. A flow diagram of our study with associated research questions is illustrated figure 10.

Introduction

Damage to hybrid poplar plantations by deer, elk, and moose have been observed and in some cases measured. The extent of damage has been reported to depend heavily on tree age and plantation location with respect to adjacent forests or woodlands (30) (31). Because large tracts of land are required to make chipped poplar feedstock production economically feasible, installing perimeter fencing to exclude ungulate herbivores could become a prohibitive input cost and significant management problem. An alternative approach might be to breed for varieties selected for less palatable foliage.

Multiple factors simultaneously influence diet preferences of wide ranging generalist herbivores like deer (32). No simple relationship between diet preferences and individual constituents like tannins, fiber, lignin, cutin, protein, non-tannin phenolics, minerals or any other extractive component have been found to consistently provide sufficient explanation. This may arise because of competing physiological benefits and costs associated with seasonally changing food mixtures, and due to the animal's adaptive capacity for detoxifying phytochemicals. Numerous physical and environmental factors also impinge on diet decisions as tradeoffs, such as feeding efficiency costs determined by leaf bite size and the spatial distribution of desirable foliage on individual plants and in landscape patches. Factors such as seasonal food availability, risk of predation, proximity to refuge and family, and social learning also impact the animal's decision to feed at a particular location or move elsewhere, and are not typically addressed in diet choice experiments.

Land area converted to short rotation woody crops for biomass feedstock can also result in landscape-scale changes in vegetation structure that can impact wild animal distributions and diversity (33) (34). Small mammal trapping studies suggest that short rotation plantations may provide more attractive habitats than other land-use types, especially where herbaceous understory is abundant soon after crop establishment (35). Voles (*Microtus* spp.) are seasonal polyestrous breeders producing several young per litter, and possibly several litters per year. In optimal conditions their populations can increase exponentially, but can also exhibit cyclical year-to-year patterns of growth and decline. When their density is high, voles can cause economic damage to plantations by girdling bark and phloem from the base of tree seedlings and shoots (36), especially in monoculture plantations.

Prior efforts to identify vole-unpalatable poplar or willow varieties have capitalized on findings demonstrating that voles prefer specific genotypes within a species or family (37) (38)

(39). Differences in palatability are thought to be correlated with the concentration of phenolic, alkaloid, salicylate, and flavonoid plant secondary compounds, but also balanced against nutritional value and digestibility (40). Quantitative associations of ingested secondary compounds and reduced preference or diminished physiologic performance have also been determined in non-Salicaceous tree species in controlled and natural settings (41).

Both approaches to using NIRS technology to achieve our objective; either as a method to rapidly quantify analyte concentration, which can then be substituted into a predetermined regression model to estimate a level of herbivory based on the concentration of analytes, or to estimate a genotype's palatability ranking resulting from the *effect* of a genotype's entire phytochemical profile on feeding preference, requires controlled feeding experiments with captive herbivores as test subjects. Confined animal feeding experiments automatically sets up a hierarchy of compromises, the most important being the restriction of diet choice to a level that rarely exists in nature. Another is the influence that each presented choice has on the herbivore's ranking of the other choices. We have no way of knowing whether the limitation of physical and environmental factors thought to impinge on diet tradeoff decisions in nature might also skew the experimental results to such a degree that they invalidate any extrapolation of findings to plantation settings. This uncertainty is the Achilles heel of the following experiments.

With the awareness that a determination of analyte concentration cannot be ascribed to a poplar genotype with any statistical confidence without drastically changing the way that staged progeny performance trials were being conducted (summarized above under Objective 1), and with the awareness that NIRS analyte calibration models using samples from multiple taxa were largely unsuccessful with the possible exception of salicortin, condensed tannin, and total phenolic glycosides (summarized above under Objective 2), we still proceeded with two confined animal (mammal) feeding experiments for several reasons:

- Washington State University was operating a wild ungulate facility in Pullman Washington to study deer feeding behavior in forested patches, as well as to serve as a rehabilitation center for injured or abandoned deer, and provide research and education opportunities for veterinary medicine.
- Farm and forested property at the Washington State University Research and Extension Center in Puyallup Washington had problems with vole populations periodically destroying test crops and poplar stool beds. Instead of exterminating these pests we procured Washington State Department of Wildlife's permission to capture and contain some of them for short periods to serve as test subjects in poplar bark feeding preference studies.
- While cafeteria-style feeding experiments using deer, and damage surveys conducted with voles using poplars have been reported, none of those studies utilized poplar species or hybrids that our collaborators were testing for commercial development. Our observations using these new genotypes would be novel.

We took advantage of these opportunities as exploratory endeavors with low expectations of developing usable regression models.

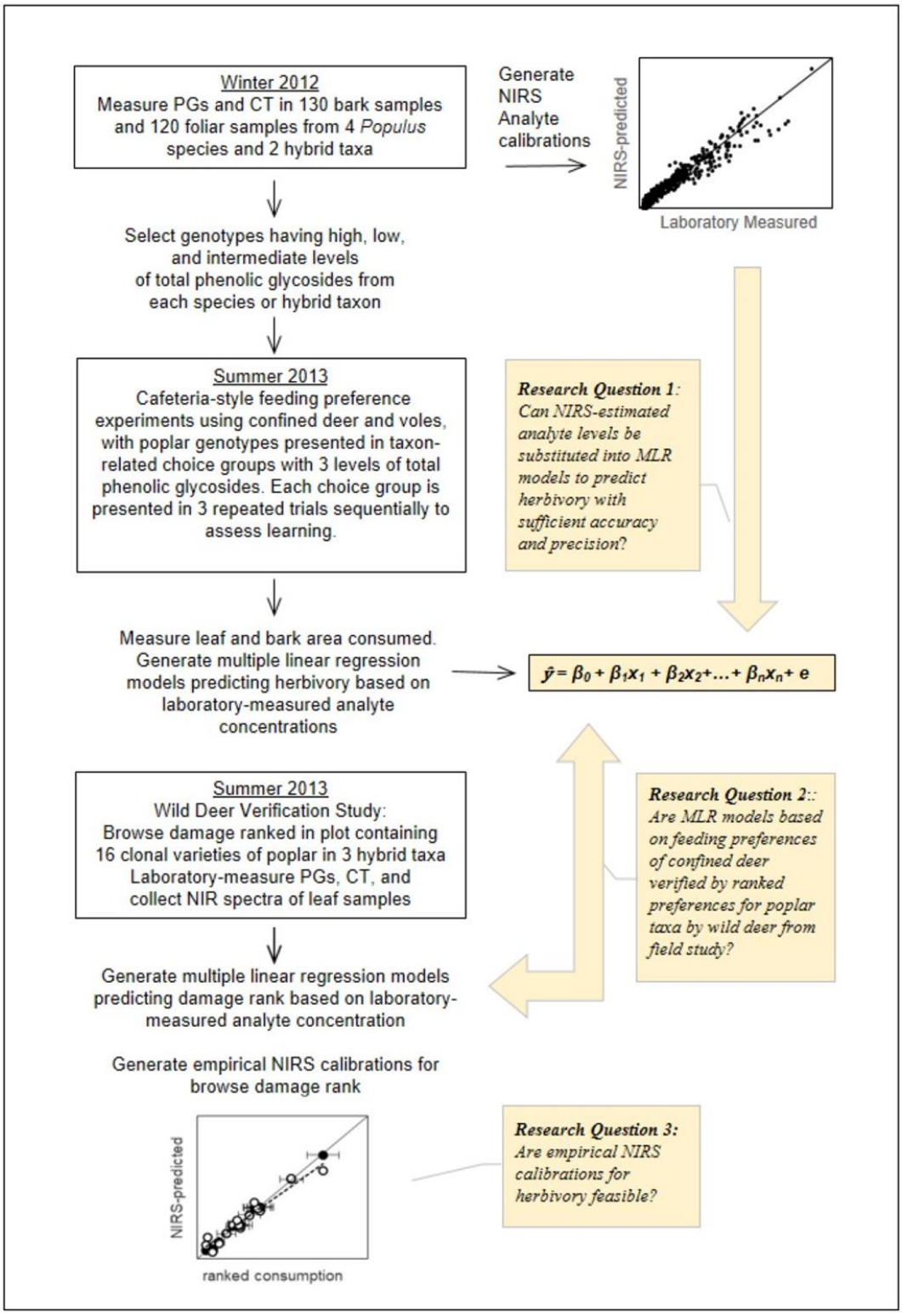


Figure 10. Research study flow, 2012 to 2013

Confined-deer feeding preference study

Our biggest obstacles to planning controlled foliage feeding experiments with confined deer was in presenting the test material in ways that resemble a plantation setting and deciding what metric best reflected feeding preference. Other food presentation options based on published experiments included foliage dried and milled into pellets presented in equal weights in cafeteria-style choice groups, then measuring weight consumed after a defined period; or, presenting excised branches with a fixed number of leaves and counting leaves consumed from each genotype; or presenting excised branches of approximately the same size weighed before and after presentation and estimating moisture weight loss during the experiment from non-presented controls (42) (43). Several observations lead us to our experimental method:

- Poplars react to physical damage by upregulating Kunitz trypsin inhibitors and secondary compounds (44). The act of cutting off branches or sapling stems, or harvesting foliage to dry and mill could alter the chemical profile to a degree that feeding preference might be altered.
- Our poplar test genotypes had widely divergent leaf sizes, which would make the presentation of equal leaf numbers appear quite different.
- Weighing approximately equal-sized branches and accounting for moisture weight loss would require many non-presented controls. We didn't have enough extra plants to accommodate this method. There was greater than a four-fold difference in average leaf area among genotypes, meaning there were huge evapotranspiration differences. If attractive volatiles or leaf moisture were factors in feeding preference, the large-leaved genotypes would wilt fast and that might impact preference.

We chose to present live potted saplings in arrays like in a plantation (see figure 11), which meant that our consumption metric was primarily leaf area, from which we could deduce dry weight consumed. The challenge was figuring out how to measure leaf area before presentation. Some of the saplings had over a hundred primary and secondary leaves. After presentation the remaining leaves were harvested and measured with a laser leaf area meter. The pre-presentation leaf area estimation method and validation is described in the appendix.

Figure 11. A 1-acre fenced paddock at the WSU Wild Ungulate Facility (WUF) with an array of potted poplar saplings. Pictured are open shelters in each paddock and a building housing veterinary examining rooms and feed storage. The WUF facility includes 5 pens spanning 8-acres.

We assumed that measuring diet preference would be time dependent, especially if all the choices were palatable but were preferred by degree, so we conducted one preliminary experiment to help estimate the presentation period. Potted poplar saplings were presented in two separate groups on different days to one male mule deer: 8 genotypes comprised of North American cultivars for approximately 24 hours on day one, and 9 genotypes comprised of Asian poplar species for approximately 24 hours on day two. In that period the buck consumed nearly all of the leaves from all of the clones. Based on this information we reduced the presentation period to 8 hours, decided to utilize smaller female deer, and also increased the number of deer to three or four to even-out the effects of individual preferences.

To keep the number of plant choices presented to the deer at a reasonable number, given the fairly short experimental season, 8 or 9 poplar plants were presented per trial in two separate taxonomic groupings; one comprised of domestic species and hybrids (*P. trichocarpa, P. x generosa,* and *P. deltoides x P. maximowiczii*), and one comprised of Asian species (*P. cathayana, P. koreana,* and *P. maximowiczii*). Each taxon in a group was represented by 3 genotypes that possessed either high, intermediate, or low levels of total phenolic glycoside concentration (cumulative abundance of salicin, salicortin, HCH-salicortin, tremulacin, and salireposide). *P. deltoides x P. maximowiczii* was represented by only two total phenolic glycoside levels due to plant mortality. Refer to table 3.

To assess the impact of learning in diet preference, trials with the same genotypes were sequentially repeated three times with new plants – one trial per day. The sequentially repeated trials were themselves replicated three times using three different groups of deer. It was thought that changes in preference over time might correlate with one or several of the phytochemical variables, especially since the Asian poplar species would be a unique food to these local deer. Between sequential trials of each taxon group the deer had a recovery day where no plants were presented and the deer were fed their normal daily ration of pelletized feed.

Saplings were placed in a 1-acre paddock with a group of 3 or 4 deer. The pots were configured in a 2 or 3 row array. Each pot was staked 2 meters apart in rows spaced 2 meters apart. The period during which trees were placed in the pen for group-1 deer lasted approximately 8 hours, from 0600 to 1400 hrs. The presentation period for group-2 and group-3 deer lasted 24 hours, from 0600 to 0600 hrs. See figure 12.

The three deer in group-1 ranged in age from 2 years to 11 years. The four deer in group-2 were all aged 6-years, and the four deer in group-3 were aged 6 and 7 years. All deer used in the trials were female. Each deer group received a completely balanced basal diet in pelletized herbivore chow (custom mule deer chow #9017 milled at Washington State University), as well as alfalfa hay, salt block, and free access to water. A variety of natural grasses and forbs growing in the 1 acre enclosure also supported some of their diet. The average daily pelletized chow provided to the deer grouped in each trial was 6.35 kg (1.59 kg per animal). The pelletized feed given to deer in group-1 during the first trial was initially reduced to 75% of their *ad libitum* level to encourage browsing of the poplar plants, but this was adjusted to 100% of *ad libitum* levels in subsequent trials. Plant browsing was encouraged in the morning by shifting their normal daily chow feeding from 0730 hrs to 1400 hrs. The objective was to present test plants in a way that was not compelled by hunger. Based on the remaining chow measured each day, we determined that consumed pelletized food inputs averaged 1.42 kg per animal. Between each group of trials the deer received 9 kg of pelletized chow and no poplar plants.

Figure 12. Deer feeding on potted poplar saplings in a feeding preference trial.

Quantifying leaf area *in situ* prior to trial presentation was the biggest challenge. We devised a method to estimate total leaf area by measuring every fourth primary leaf's elliptical area, then by a defined method of extrapolation, interpolation and use of empirically derived leaf shape correction factors, we estimated the total leaf area of the sapling (see appendix). We also used the elliptical leaf area of the median sized secondary leaf on branches multiplied by the number of leaves on the branch and leaf shape correction factors to estimate the secondary leaf area. Fortunately, some of the genotypes turned out to be quite unpalatable during the trial test period and were avoided for the most part. This provided an opportunity to validate whether our *in situ* leaf area estimation method was sufficiently accurate by comparing with the harvested leaf area determinations using a laser leaf area meter. The majority of our estimations were within 5% of measured values, but a minor portion of the estimates were ± 15% of measured leaf area, which resulted in poor models of herbivory using multiple linear regression.

Another limitation of the study was that the deer were quite tame. Most were born in captivity or were rehabilitated from injuries or abandonment at a young age and had become used to humans. Although a few of the deer may have participated in short-term forage studies in managed forest plots, most did not have natural foraging experience. All were sustained on a regimented diet of pelletized forage grasses and grains. So, while our research design attempted to account for learning in food preference determinations it could not account for the absence of learned preferences that may naturally occur among individuals in wild populations.

Our results indicated that while there were significant differences in preference among genotypes in each choice group (Figure 13, Table 3), these were not consistently correlated with the putative 'defensive' analyte concentration (Table 4).

Because we increased the presentation time in the deer group-2 and group-3 trials to 24 hours, we accounted for the differences in absolute leaf area consumed among different deer group trials by computing the proportion of individual genotype leaf area consumed with respect to the total leaf area consumed in the trial (PTAC) as the normalized metric of food preference. This works as long as no genotype is completely consumed or completely avoided during the trial period. As it turned out, there were a few instances where each of these conditions was nearly met. For example, DM-8002, with its large leaves, was a strong favorite and consumed early in the trials except for the highest leaves on the plant. In contrast, T-2683 having small leaves was consistently avoided except for a few leaves that were tasted.

What was remarkable in these experiments was the agreement in relative preference for each genotype among the different groups of deer, particularly between deer group-2 and deer group-3. Another observation of note was that, with exception to *P. koreana* genotype K-14628, the Asian poplar genotypes in choice group-2 were relatively less preferred compared to the North American genotype options presented in choice group-1.

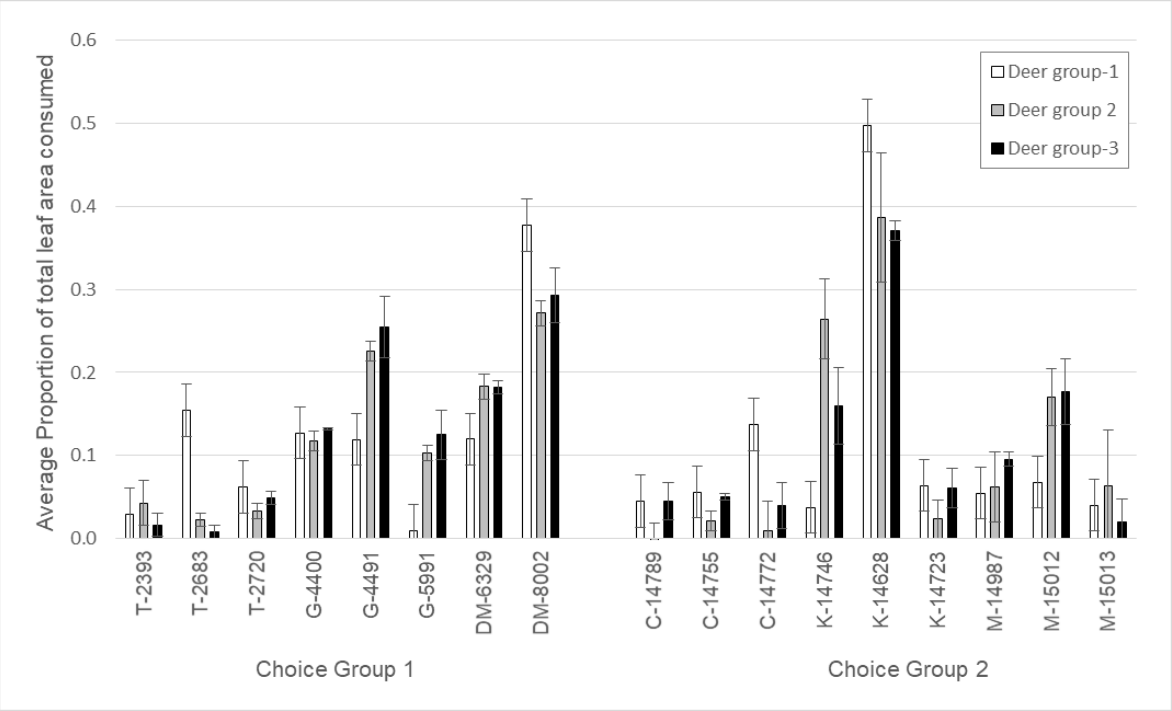


Figure 13. Normalized feeding preference expressed as the proportion of leaf area consumed from individual genotypes in each choice group with respect to the total leaf area consumed from all genotypes in a sequential trial, averaged over the tree replicate sequential trials. Deer group 1 (open bars) were presented potted trees for 8 hours; deer group 2 (shaded bars) and deer group 3 (filled bars) were presented trees for 24 hours. Error bars are the standard error of the mean.

Table 3. Foliar analyte concentrations and captive deer consumption measurements

			Analyte Concentration												Consumption Averages							
Choice group	Taxon-Clone	Avg. Leaf Area (cm^2)	Neutral Detergent Fiber (% dw)	Crude Protein (% dw)	BCA Protein (% dw)	Total Phenolic (µg/mg)	CT (Abs/mg)	Salicin (% dw)	Salicortin (% dw)	HCH-salicortin (% dw)	Trem (% dw)	Salirps (% dw)	Total PG (% dw)	Unknown RT 3.6 min m/z = 315 (rel)	Est. Leaf area consumed (cm^2)	Std. Error	Dry weight consumed (g)	Std. Error	Percent tree leaf area consumed	Std. Error	Proportion of total area consumed in trial	Std. Error
1	T-2393	23.5	33.9	11.8	17.4	100.9	0.92	*	14.70	0.65	5.94	1.34	22.68	0.89	1135	580	9.4	4.8	16.1	8.2	0.030	0.015
	T-2683	33.5	29.8	11.3	21.8	100.4	1.38	*	6.55	*	2.75	0.40	9.77	2.22	597	239	5.0	2.0	7.8	3.1	0.015	0.006
	T-2720	23.1	32.3	12.3	20.6	123.0	2.66	*	1.73	0.12	0.32	0.27	2.48	0.69	1496	289	10.7	2.1	25.1	6.2	0.041	0.006
	G-4400	64.5	31.9	12.0	12.7	109.4	0.84	0.41	3.30	3.91	1.37	1.41	10.41	0.70	4449	283	39.2	2.5	87.6	3.6	0.125	0.006
	G-4491	45.3	31.2	10.1	15.6	135.6	1.82	0.28	3.03	3.08	1.75	1.22	9.36	0.32	8460	394	73.5	3.4	92.3	4.2	0.240	0.018
	G-5991	44.0	32.8	9.6	19.7	96.3	0.69	0.43	4.30	3.81	2.24	1.53	12.31	0.57	4094	527	34.8	4.5	77.8	8.0	0.114	0.015
	DM-6329	54.5	38.8	11.7	16.6	86.2	0.13	0.73	4.51	6.26	*	2.56	14.06	1.70	6642	571	52.8	4.9	95.5	1.7	0.183	0.009
	DM-8002	97.7	29.9	10.7	12.5	146.6	2.18	*	0.82	0.28	*	*	1.18	0.36	9986	541	79.1	4.2	98.9	0.6	0.282	0.017
2	C-14789	26.0	36.1	15.0	13.7	85.1	0.13	0.53	6.12	8.06	*	4.21	18.93	2.98	394	236	3.4	2.1	9.9	5.7	0.022	0.017
	C-14755	48.0	34.4	12.8	14.2	88.8	0.14	0.38	7.55	3.30	*	4.44	15.68	3.06	594	168	4.8	1.5	7.5	2.2	0.036	0.009
	C-14772	25.3	36.2	14.3	14.3	75.9	0.09	0.43	6.98	9.63	*	5.67	22.70	3.69	559	361	4.6	3.1	15.8	10.8	0.025	0.021
	K-14746	50.4	31.6	10.4	13.7	118.9	0.62	0.11	20.32	*	*	0.20	20.64	0.85	3216	538	26.2	3.9	68.1	11.4	0.212	0.038
	K-14628	65.3	29.2	12.2	11.9	117.8	0.55	0.25	17.57	*	*	0.25	18.10	1.27	5677	306	48.8	2.1	94.9	4.1	0.379	0.035
	K-14723	41.7	37.2	14.8	14.5	76.6	0.09	0.38	5.85	5.99	*	3.52	15.74	2.30	516	262	4.6	2.3	12.7	5.8	0.038	0.017
	M-14987	29.3	34.4	13.6	15.5	92.4	0.10	0.57	8.33	6.55	*	3.81	19.26	3.50	1248	379	10.9	3.7	24.6	7.8	0.079	0.021
	M-15012	30.1	36.6	12.4	13.3	99.7	0.14	0.14	9.23	5.42	*	3.76	18.56	2.61	2783	492	22.9	4.5	56.5	11.1	0.174	0.023
	M-15013	26.2	36.3	13.30	16.6	85.4	0.10	0.31	8.94	6.09	*	3.88	19.23	2.87	911	613	6.8	5.1	12.0	7.8	0.042	0.034

Taxon: (T) *Populus trichocarpa* , (G) *P. x generosa* , (DM) *P. deltoides x P. maximowiczii* , (C) *P. cathayana* , (K) *P. koreana* , (M) *P. maximowiczii* .

Crude protein was determined as a percent of total nitrogen. BCA protein was determined using the micro-BCA method (Pierce). Total phenolic content was determined using a modified Folin-Ciocalteu reagent method. Relative condensed tannin (CT) is expressed as the cyanidin concentration in acid butanol (Porter method) measured as visible absorbance at 550 nm per milligram of dry sample weight. Concentration values below the statistical reporting limit or absent in salicin, HCH-salicortin, tremulacin (Trem), and salireposide (Salirps) are indicated (*). Total phenolic glycosides (Total PG) is the percent by weight sum of salicin, salicortin, HCH-salicortin, tremulacin, and salireposide. An unknown multicomponent chromatographic peak at retention time = 3.6 minutes, having a mass/charge ratio of 315, was normalized to the salicortin peak area and expressed as a relative quantity.

Leaf consumption values are the average of three sequential trials with deer group-2 and deer group-3 (n = 6), where the tree presentation time was 24 hours.

Table 4. Analyte and Leaf Consumption Correlation

	Choice Group 1		Choice Group 2	
	Avg. Leaf Mass Consumed	Avg. PTAC	Avg. Leaf Mass Consumed	Avg. PTAC
Avg. genotype Leaf Area	0.77	0.81	0.83	0.75
Crude Protein	0.76	-0.48	0.77	-0.68
Neutral Detergent Fiber	0.79	-0.01	0.79	-0.84
BCA Protein	0.72	-0.74	0.77	-0.72
Total Phenolics	0.78	0.56	0.89	0.90
Condensed tannin	0.63	0.08	0.86	0.83
Salicin	*	*	0.12	-0.62
Salicortin	-0.28	-0.57	0.88	0.84
HCH-salicortin	0.37	0.36	-0.78	-0.79
Tremulacin	-0.39	-0.56	*	*
Salireposide	0.24	0.08	-0.72	-0.86
Unknown m/z 315	-0.53	-0.49	-0.74	-0.77

Pearson correlation values relating the estimated milligram quantity of analyte in an average size leaf of each genotype with the averaged leaf mass consumed in three sequential trials for deer groups 2 and 3. Also, the correlation between the percent dry weight concentration of analyte in genotype foliage with the averaged proportion of individual genotype leaf area consumed with respect to the total leaf area consumed from all genotypes in a trial (PTAC), averaged in three sequential trails for deer groups 2 and 3. Missing values (*) indicate analyte concentrations below reporting limits for all or most of the genotypes in the choice group.

We examined the correlation between estimated leaf mass consumed from each genotype averaged over all sequential trials, and the corresponding estimated total milligram quantity of analyte in an average leaf because there was a three-fold range in leaf size among the genotypes (Table 4). This was supported by observations of deer 'taste-testing' single leaves. Each leaf consumed (experienced) was thought to influence preference, and its phenotypic size was a scalar of analyte content. In contrast, correlation values were very different when analyte concentrations expressed as percent by weight were associated with averaged proportion of leaf area consumed from each genotype relative to the total leaf area consumed in a trial. For example, crude protein, BCA protein, and neutral detergent fiber were negatively correlated with PTAC in both choice groups. These correlation differences highlight the importance of taking into consideration the average phenotypic leaf area of each genotype in analysis of the data, and supports the notion, for which there is evidence (45), that deer are economical browsers.

The strong positive correlation with total phenolic content suggests that the fraction of secondary compounds having a phenolic moiety do not appear to be a deterrent. This was also supported by the positive correlation with condensed tannins. While CT levels were high in genotypes consumed the most in each choice group, they were also high in genotypes that were avoided, suggesting that it did not impact preference strongly on its own.

Levels of individual phenolic glycosides did not have strong correlation with consumption in common with both choice groups, but some did have consistent impacts among trials within a

choice group. We observed a consistent decline in leaf area consumed in each sequential trial with the Asian poplars *P. cathayana* C-14789 and C-14772 (data not shown), for deer groups 2 and 3, possibly indicating learned avoidance. Leaf samples for these genotypes had higher concentrations of HCH-salicortin and salireposide than the other genotypes in that choice group, whereas the preferred genotypes K-14628 and K-14746 had the lowest levels of HCH-salicortin and salireposide. Similarly, the most favored genotype of choice group 1, DM-8004, also had low levels of HCH-salicortin and salireposide relative to the other genotypes in that group. One unknown compound whose HPLC chromatographic peak with a retention time of 3.6 minutes and having a mass to charge ratio of 315 (possibly an unidentified phenolic glycoside) demonstrated a consistent negative correlation with consumption in both choice groups.

The relative abundance of each phenolic glycoside can be viewed in aggregate as a phytochemical genotypic characteristic arising from unique differences in enzyme activities resulting from genetic variationone couldthere is variable PGthe ation ofof. Conceptually, biochemically linked and co-varying concentrations of PGs could be considered a static trait. Not to imply these concentrations are static. Rather, any seasonal differences and environmental influences on relative concentration could, in theory, be normalized with a standardized sampling protocol.

Utilizing the estimated leaf area-scaled milligram quantities of analyte per leaf as the independent variables, and the trial-averaged proportional consumption as the response variable, the following stepwise multiple linear regression models were derived:

- Choice Group 1: Average PTAC = 0.00196*Total Phenolics + 0.00397*HCH salicortin - 0.03592* Unknown m/z 315 + 0.03552. [Adj. $R^2 = 0.825$, RSE = 0.0405, $df = 12$, $F = 24.58$, $p < 0.0001$, variable selection method = both].
- Choice Group 2: Average PTAC = 0.01759*Total Phenolics -0.0706*Salicin – 0.00792*Salicortin – 0.0171*Salireposide + 0.07218. [Adj. $R^2 = 0.857$, RSE = 0.0454, $df = 13$, $F = 26.36$, $p < 0.0001$, variable elimination method = both].

While more than 80% of the consumption variation can be explained by these models, they are inherently weak because the degrees of freedom were greatly reduced by using trial-averaged consumption values. This was done because we did not have the resources to measure analyte concentrations from every plant used in a trial. Instead, we pooled leaf samples from several different individuals of the same genotype and assumed this to be the characteristic phenotypic concentration for all individuals of the genotype. In addition, we only considered average proportional consumption for deer groups 2 and 3, because for these groups the trees were presented to the deer for the same amount of time in each test.

Additional MLR models were generated based on analyte concentration (not scaled to leaf area), and limited to the analyte variables for which NIRS calibrations had been created.

- Choice Group 1: Average PTAC = 0.0202*BCA Protein + 0.0119*Total Phenolics – 0.2424*Condensed Tannin + 0.0189*Salicortin – 0.0612*Tremulacin – 0.0223*Unknown m/z 315 – 1.1955. [Adj. $R^2 = 0.981$, RSE = 0.0132, $df = 9$, $F = 132$, $p < 0.0001$. All variables significant at $p < 0.05$].
- Choice Group 2: Average PTAC = -0.0258*BCA Protein + 0.0051*Total Phenolics – 0.0092* Unknown m/z 315 + 0.0258. [Adj. $R^2 = 0.791$, RSE = 0.055, $df = 14$, $F = 22.38$, $p < 0.0001$].

While the model explained variance appeared impressive, the results likely reflect a fortuitous combination of laboratory-measured concentrations from pooled samples regressed against consumption measurements from too few tests.

The sobering takeaway from this study is that the extent that any genotype's foliage was consumed was relative to the array of choices presented – not merely the concentration of a few analytes. Said another way, the effect of concentration on consumption from one test array may not be applicable to a different array of choices. A thorough cafeteria style experiment designed to explore all possible combinations of genotypes in unique arrays of 9 from a field of 18 possibilities would have required 48,620 tests. Multiplied by 3 sequential repeats and 3 deer group replications would require 437,580 tests. The only way to make these combinatorics work is to reduce the number of choices and possibly conduct tests over several years.

Wild deer browse study

To investigate whether similar feeding preferences occurred in the wild, we conducted a deer browse damage survey at an experimental site located at the GreenWood Resources Tree Farm located 12km east of Boardman Oregon. The experiment was comprised of 16 genotypes belonging to 3 hybrid taxa; each genotype planted in 9-tree plots, each plot replicated randomly in four blocks oriented parallel to a frontage road. All trees received dripline fertigation. The trees had been coppiced once and were in their second year of regrowth. Deer browse was observed to be fairly uniform throughout the site, but the block adjacent to the road appeared to be damaged less and was excluded from our correlation and modeling analysis. The extent of browse was visually ranked at each of the 576 trees on a scale from 0 to 2 by two observers simultaneously. A score of 0 was assigned if no browse was observed; a score of 1 assigned if less than 20% of the branches were stripped or broken terminals were observed; a score of 2 was assigned if greater than 20% of the branches were browsed or entire stems were defoliated. Leaf samples were collected and pooled from three individuals of each genotype in each plot, and the dominant phenolic glycosides and condensed tannins were measured in a manner identical to the Pullman confined deer study except crude protein, neutral detergent fiber, average leaf size, and consumed leaf area variables were not determined.

The hybrid taxa were not equally represented in the experiment, comprised of 8 genotypes of *P. deltoides maximowiczii* (DM), 5 genotypes of *P. deltoides x trichocarpa* (DT), and 3 genotypes of *P. deltoides x nigra* (DN). Browse preferences for the larger-leaves of DT genotypes were observed, with the exception of clone 12812, which was a full sibling of the third-most severely browsed genotype DT-12813 (Figure 14). DN genotypes had the smallest sized leaves and were observed to have the least damage.

There were notable differences in phytochemical composition of the potted saplings used in the Pullman confined deer study compared and the Boardman genotypes grown in mineral soil. We measured lower levels of condensed tannin, salicortin, total phenolic glycosides, and total phenolics (Table 5). The extent of deer browse correlated with analyte concentration was similar to the confined deer study in that protein, total phenolics, and condensed tannin were positively correlated (Table 6). Overall there was little agreement of correlations of browse score with individual phenolic glycosides and unknown m/z 315 compound among the hybrid taxa.

Table 5. Wild Deer Browse Score and Analyte Data

Taxa	clone	BCA Protein (% dw.)	Total Phenolic (μg mg^{-1})	CT (Abs mg^{-1})	Salicin (% dw.)	Salicortin (% dw.)	HCH-salicortin (% dw.)	Trem. (% dw.)	Salirps. (% dw.)	Total PGs (% dw.)	Unknown m/z 315 (rel mg^{-1})	Avg. Browse Score (n = 9)
DM	16560	13.41	55.2	0.06	0.81	5.02	3.53	1.36	0.30	11.02	1.77	0.333
DM	16560	15.00	53.7	0.06	0.64	4.29	4.04	1.20	0.29	10.46	1.45	0.778
DM	16560	14.28	56.3	0.06	0.57	4.34	3.54	1.28	0.26	9.99	1.30	0.444
DM	16561	14.58	58.8	0.13	0.79	6.66	3.03	1.89	0.32	12.69	1.72	0.222
DM	16561	12.02	58.9	0.12	0.50	6.69	2.94	1.60	0.33	12.07	1.71	1.111
DM	16561	12.20	67.3	0.15	0.51	7.57	2.92	1.67	0.33	13.01	1.79	0.889
DM	16562	14.10	54.1	0.08	0.77	3.73	3.60	*	1.62	9.72	1.72	0.556
DM	16562	14.81	59.9	0.06	0.17	4.02	4.21	*	1.79	10.19	2.32	0.556
DM	16562	13.10	60.6	0.07	0.29	4.96	4.39	*	2.21	11.86	2.18	0.500
DM	16563	14.47	63.5	0.08	0.72	4.08	4.09	*	1.83	10.72	2.20	0.111
DM	16563	13.76	56.1	0.07	0.65	3.63	3.80	*	1.71	9.80	1.77	0.556
DM	16563	13.24	61.1	0.08	0.52	4.51	4.76	0.00	2.23	12.02	2.06	0.556
DM	16564	14.73	72.4	0.51	1.01	7.54	2.70	*	0.18	11.43	1.14	0.556
DM	16564	15.03	74.7	0.50	1.03	7.18	2.38	*	0.17	10.77	1.04	0.889
DM	16564	13.63	78.5	0.57	0.82	6.95	2.63	*	0.14	10.54	1.09	0.778
DM	16565	17.72	61.8	0.61	1.00	5.08	2.52	*	0.14	8.76	0.98	1.000
DM	16565	14.88	79.3	0.65	0.68	5.62	2.42	*	0.18	8.90	0.88	1.000
DM	16565	14.82	71.4	0.50	0.87	5.95	2.85	*	0.17	9.84	1.05	1.111
DM	16566	15.95	67.0	0.53	0.80	6.58	2.44	*	0.28	10.10	1.08	0.778
DM	16566	15.07	63.5	0.31	0.75	7.34	2.94	*	0.30	11.35	1.05	0.889
DM	16566	14.68	74.4	0.47	0.53	8.95	3.12	*	0.29	12.90	1.00	0.889
DM	16567	16.04	72.9	0.86	0.61	5.26	1.71	*	0.19	7.78	1.01	1.000
DM	16567	16.03	74.8	0.66	0.38	5.90	1.96	*	0.19	8.44	1.01	1.444
DM	16567	14.90	72.9	0.53	0.60	6.96	2.49	*	0.23	10.29	1.23	0.778
DN	10642	15.75	50.1	0.08	0.25	6.62	1.09	*	0.11	8.08	1.48	0.222
DN	10642	15.53	52.1	0.08	0.19	6.44	1.01	*	0.05	7.70	0.96	0.000
DN	10642	15.95	55.4	0.09	0.16	6.10	1.07	*	0.24	7.57	1.26	0.000
DN	10643	13.76	61.5	0.31	0.28	6.75	0.14	0.82	0.19	8.18	1.39	0.125
DN	10643	14.90	64.3	0.31	0.23	6.81	0.12	0.70	0.20	8.06	1.62	0.222
DN	10643	15.58	65.0	0.51	0.26	6.43	0.12	0.55	0.32	7.68	1.09	0.333
DN	10650	14.74	47.3	0.13	0.80	6.15	1.28	*	*	8.28	1.70	0.222
DN	10650	14.87	43.6	0.12	0.59	4.97	1.23	0.00	*	6.83	1.23	0.000
DN	10650	14.51	39.1	0.13	0.73	6.44	1.54	*	*	8.75	1.63	0.000
DT	12810	17.19	71.1	0.47	0.43	2.70	3.17	*	1.55	7.86	0.83	1.444
DT	12810	16.65	69.1	0.47	0.34	1.86	2.48	*	1.16	5.84	0.76	1.333
DT	12810	16.03	76.2	0.66	0.21	2.41	3.04	*	1.44	7.11	0.89	1.000
DT	12812	16.57	43.1	0.09	0.35	3.64	2.24	*	0.10	6.34	1.67	0.222
DT	12812	15.26	48.0	0.19	0.37	2.88	2.23	*	0.06	5.55	1.74	0.222
DT	12812	16.30	47.9	0.18	0.45	3.54	2.61	*	0.06	6.66	1.84	0.000
DT	12813	16.69	62.2	0.25	0.73	4.30	3.72	*	2.57	11.33	1.13	1.333
DT	12813	17.40	75.1	0.48	0.46	4.26	3.46	*	2.53	10.72	0.89	1.667
DT	12813	17.69	72.1	0.67	0.52	4.42	3.83	*	2.45	11.22	1.07	0.889
DT	12815	17.25	66.4	0.35	0.50	4.95	3.83	*	2.41	11.70	1.04	1.667
DT	12815	17.84	60.6	0.21	0.47	4.11	3.59	*	2.24	10.43	1.10	1.667
DT	12815	16.93	81.7	0.58	0.25	4.58	3.42	*	2.11	10.36	1.01	1.333
DT	12820	16.62	67.2	0.31	0.40	2.86	3.21	*	2.14	8.62	0.57	1.111
DT	12820	16.45	61.6	0.22	0.60	3.74	4.09	*	2.44	10.87	0.94	1.333
DT	12820	17.01	61.7	0.44	0.42	2.78	2.98	*	1.88	8.06	0.58	0.778

Measured analyte concentration in leaf samples of genotype clones replicated in three blocks. Tremulacin (Trem) and salireposide (Salirps) concentrations recorded as (*) were below statistical reporting limits.

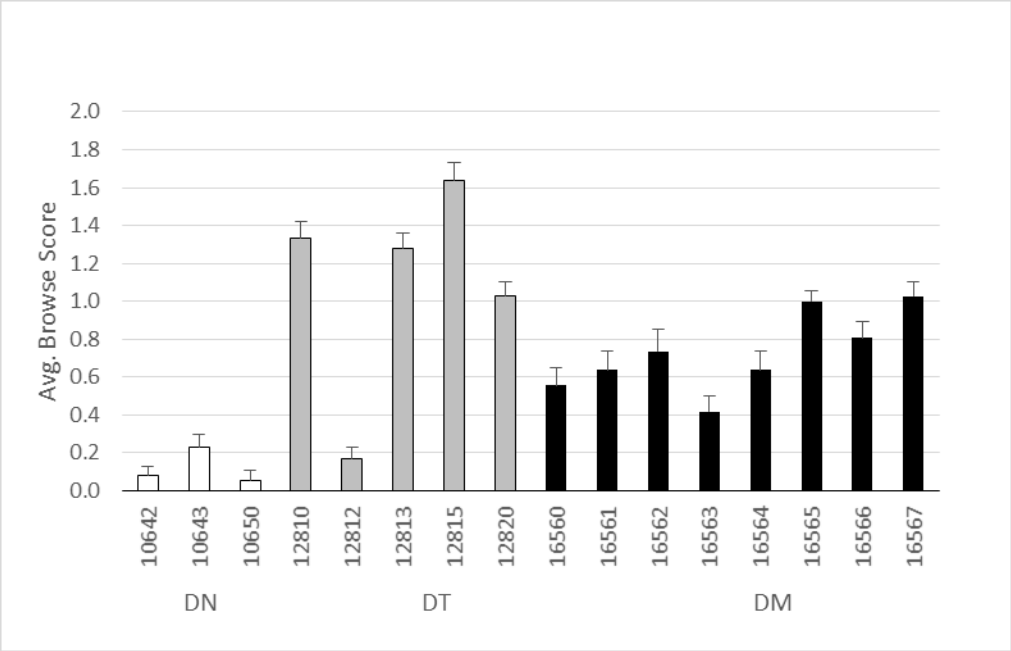


Figure 14. Averaged browse scores for clones of three hybrid poplar taxa: *P. deltoides x nigra* (DN) (open bars), *deltoides x trichocarpa* (DT) (shaded bars), and *deltoides x maximowiczii* (DM) (solid bars). Scores were averaged from 9-tree plots replicated in 4 blocks (n = 36). Error bars are the standard error. A score of 2 was assigned to trees with more than 20% of the leaves browsed.

Table 6. Analyte and Browse Score Correlation [a]

Analyte	Avg. Browse Score Overall	Avg. Browse Score DM Genotypes	Avg. Browse Score DN Genotypes	Avg. Browse Score DT Genotypes
BCA Protein	0.47	0.45	0.13	0.65
Total Phenolics	0.67	0.56	0.55	0.42
Condensed tannin	0.53	0.65	0.47	0.31
Salicin	0.09	-0.03	0.12	0.34
Salicortin	-0.19	0.43	0.33	0.33
HCH salicortin	0.39	-0.68	-0.43	0.63
Tremulacin	*	*	*	*
Salireposide	0.55	-0.53	0.43	0.62
Total Phenolic Glycosides	0.31	-0.3	0.17	0.58
Unknown m/z 315	-0.62	-0.72	0.21	-0.31

a.) Spearman ranked correlation rho values for analyte concentration and averaged browse scores for all genotypes combined and by separate taxon. Tremulacin was absent or low in the majority of samples.

The widely varying analyte correlation values by taxon suggests that individual phenolic glycoside concentrations alone are non-transferrable predictors of browse preference across taxa. Forcing generalized inferences from a relatively small data set where the response variable is based on visual impressions rather than measured extent of browse was evident in our stepwise multiple linear regression analysis where only 68% of the variation in browse preference was modelled: Average Browse Score = 0.015*Total phenolics + 0.151*HCH-salicortin + 0.095*Salireposide - 0.515*Unknown m/z315 – 0.0501, [Adj. R^2 = 0.68, RMSE = 0.279, Adj. R^2 = 0.683, F = 26.28, p <0.0001, variable elimination method = both].

One of our interests in analyzing the relationship between browse preference and analyte concentrations in this study was to determine whether some combination of analytes might be used in developing a common multiple linear regression model to sufficiently predict feeding preference in both the wild deer and captive deer studies. Using backward and forward variable elimination methods in multiple linear regression revealed four of the analytes in the Boardman study as significant predictors: total phenolics, HCH-salicortin, salireposide, and unknown m/z 315. Of these, total phenolics, and unknown m/z 315 were significant predictors in all three data sets. (Note: to make three data sets comparable we used models generated from the analyte concentrations not scaled to leaf size in the Pullman deer study because this scalar was not determined in the Boardman data set). A linear model using these variables and HCH-salicortin produced the parameter estimates listed in Table 7, which explain between 67% and 84% of variation in herbivory measured either as proportional leaf area consumption or ranked browse intensity.

Table 7. Common Multiple Linear Regression Model Parameter Estimates [a] & Summary

	Pullman Confined Deer Average PTAC[b]		**Boardman Wild Deer** Average Browse Score
Explanatory variable	Choice Group 1	Choice Group 2	All genotypes
Total Phenolics	0.01392 ***	0.01966 **	0.01537 **
HCH-salicortin	0.17027 ***	0.03797	0.20555 ***
Unknown m/z315	-0.00986	-0.10747	-0.55318 ***
Intercept	-1.31503 *	-1.41341 *	-0.07156
Residual Standard Error	0.1513	0.1563	0.2834
Degrees of freedom	12	14	44
Adj. R^2	0.841	0.756	0.673
F-statistic	27.5	18.6	33.2
p-value	<0.0001	0.0004	<0.0001

a) Significance level of the parameter estimates: (***) $p < 0.001$; (**) $p < 0.01$; (*) $p < 0.1$, ( ) $p > 0.1$
b) (PTAC) the proportion of leaf area consumed from a genotype relative to the total leaf area consumed from all trees during a trial

Clearly, the combination of these three analyte predictors do not provide reliable estimates of herbivory that could guide breeding selection. Both deer studies concur that condensed tannins and salicortin did not have consistent covariance with feeding preference, and levels of tremulacin and salicin were often low or too low to report. Leaf size played an important role in both studies. Large supple leaves of DT, DM, and some K genotypes may have been desired as a source of higher moisture content and lower fiber content. So, there is a reasonable likelihood that leaf properties other than 'defensive' secondary compounds are also driving consumption and confounded in the results. The weakness of both studies and models rests on poor precision in measuring the consumption response variables.

To follow through with the intended approach of substituting estimated analyte concentration derived from NIRS calibration models into a general MLR model as an exercise, we devised a test case using PLS1 cross-validation estimates of concentration for total phenolics, HCH-salicortin, and unknown m/z 315 in the DM genotypes of the Boardman deer study as input variables in combination with the MLR parameter estimate multipliers in Table 6 to predict average browse score. To accomplish this we generated more accurate calibrations for unknown m/z 315 (Figure 15), and total phenolics (Figure 16), utilizing analyte measurements in the samples collected at the Boardman study added to the original multi-taxon sample set collected in 2011 and 2012.

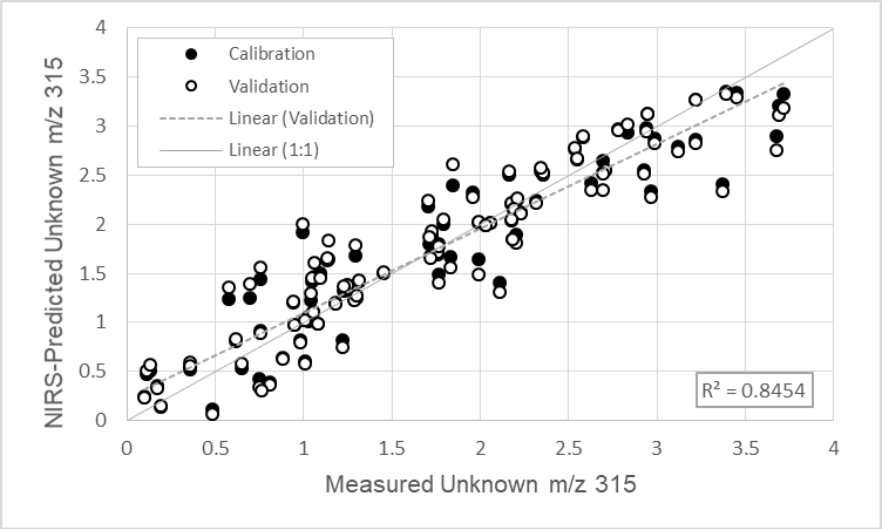


Figure 15. NIRS calibration model for estimating unknown m/z 315 in *P. deltoides x maximowiczii* genotypes in the Boardman deer study plus the Asian poplar species in the original calibration set, n = 95. The PLS1 model required 6 principal components to minimize residual variance. Calibrations for this analyte in North American poplar species and hybrids were not feasible.

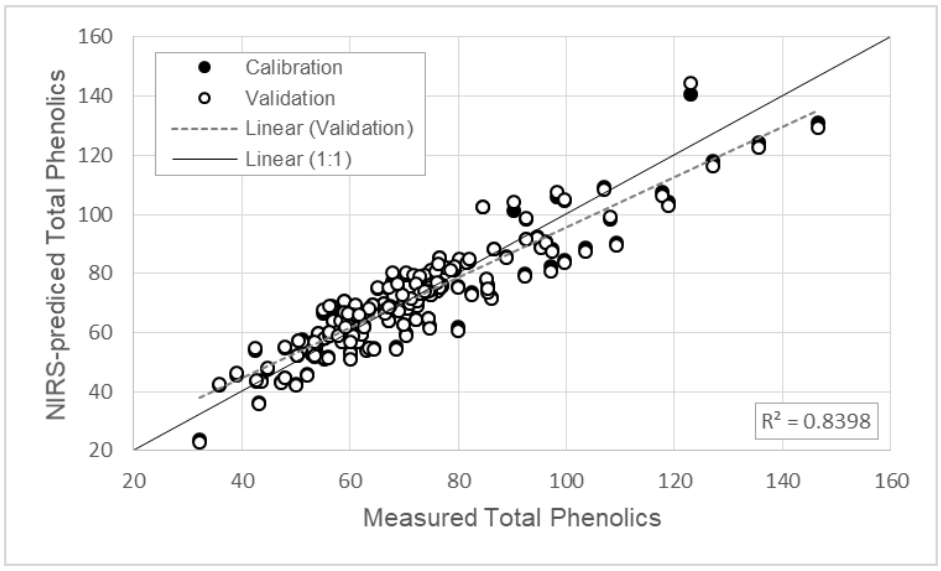


Figure 16. NIRS calibration model for estimating Total Phenolics in all study species and hybrid taxa *n* = 154. The PLS1 model required 4 principle components to minimize residual variance.

Not surprising, the MLR predictions compared to measured values were unsatisfactory (Figure 17). The poor result highlights the impact of compounding standard errors from each NIRS-estimated analyte concentration in combination with the residual standard error of the common MLR model to further reducing precision. Overall, the imprecision of visually ranked browse score measurement was the major contributing factor.

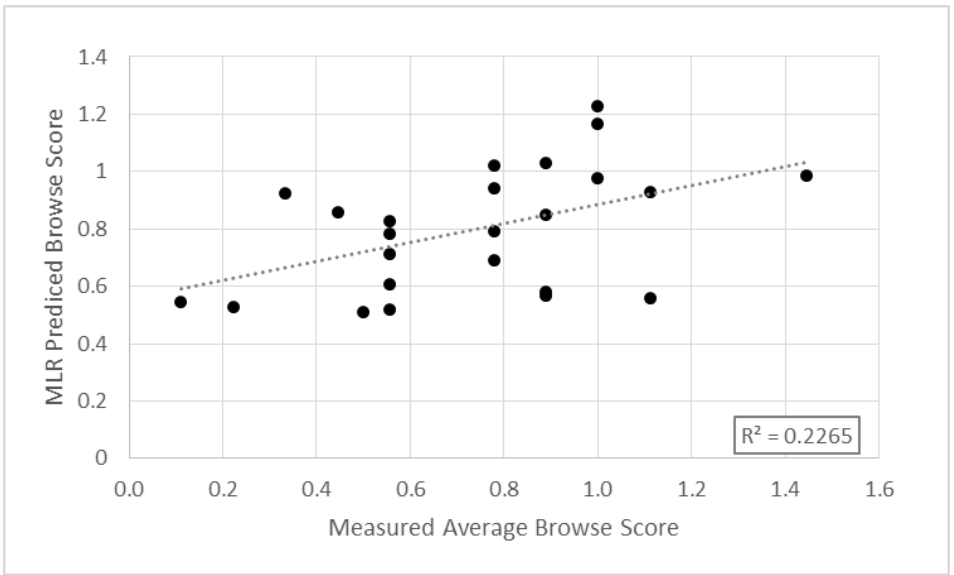


Figure 17. Regression of measured average browse score against the common MLR model-predicted browse score, where NIRS calibration model estimates for unknown m/z 315, HCH-salicortin, and total phenolics in DM genotypes of the Boardman deer study were used as input variables in combination with the MLR parameter estimates.

The key issues related to this approach for using NIRS technology in this application are:

- There are large negative impacts from compounding errors. NIRS-based analyte calibration models developed with multiple species and hybrid taxa are plagued by taxon-idiosyncratic covariance (summarized in Objective 2). Accounting for this variation in analyte concentration among same-aged leaves from the same genotype, grown at the same site, requires sample sizes not commonly available, logistically challenging to collect and process, (summarized in Objective 1), and therefore impractical and uneconomic as a general approach.
- Even if acceptably accurate and precise NIRS analyte calibrations were to be developed, the inescapable variability in measured herbivory in controlled experiments or from damage survey observations, overwhelms any accuracy and precision of the independent variable inputs.

We then took the alternate approach, predicated on the assumptions: (a) that many NIR-absorbing compounds exist in foliage that act additively to impact palatability and preference; (b) that these unknown compounds give rise to a direct covariance and correlation between NIR absorption patterns in the spectrum of each genotype and the feeding response of the herbivore; and (c) that it is not necessary to identify and quantify each compound individually. We thereby eliminate individual analyte calibration errors by incorporating them (virtually) into one empirical NIRS-based PLS1 model that predicts consumption directly. This does not eliminate the variability problems associated with measuring consumption, but the empirical approach does allow a greater number of explanatory variables to contribute in herbivory estimation, and eliminates the costs associated with generating a chemical reference data set.

To be clear, the algorithmic process for generating a single analyte calibration is no different than generating an empirical model of estimated herbivory. We've merely switched the reference data from measured analyte concentration to measured herbivory. By foregoing the intermediate step of identifying and modeling the concentrations of particular (known) phytochemical elements affecting herbivory we are essentially eliminating compounding errors and adding unknown chemical input variables. Our primary goal was to create a genotype selection tool for poplar breeders, not solely to further an understanding of the phytochemistry of palatability.

We utilized this alternate approach in generating a PLS1 model for predicting average browse score in the Boardman deer study (Figure 18). While the accuracy and precision of the model is greatly improved over the 3-analyte common MLR model using NIR calibration estimates (Figure 17), the low precision inherent in visually ranking browse damage could not be overcome by adding many more explanatory (absorbance) variables. A model with cross-validation R^2 of 0.53 is unusable for selecting low-palatability genotypes. If a genotype is equally likely to have a predicted browse score of either 0 or 1.33 then it is not worth spending money on man-hours and materials to collect samples to scan with NIR. However, the empirical model result is promising as a proof-of-concept. If the accuracy of herbivory measurements were improved and normalized to a scale that reflects preference (i.e. proportional consumption among all possible choices), it is conceivable that the empirical approach might produce usable models. We further explored this modeling approach in predicting genotype foliage preferences by the cottonwood leaf beetle, described below.

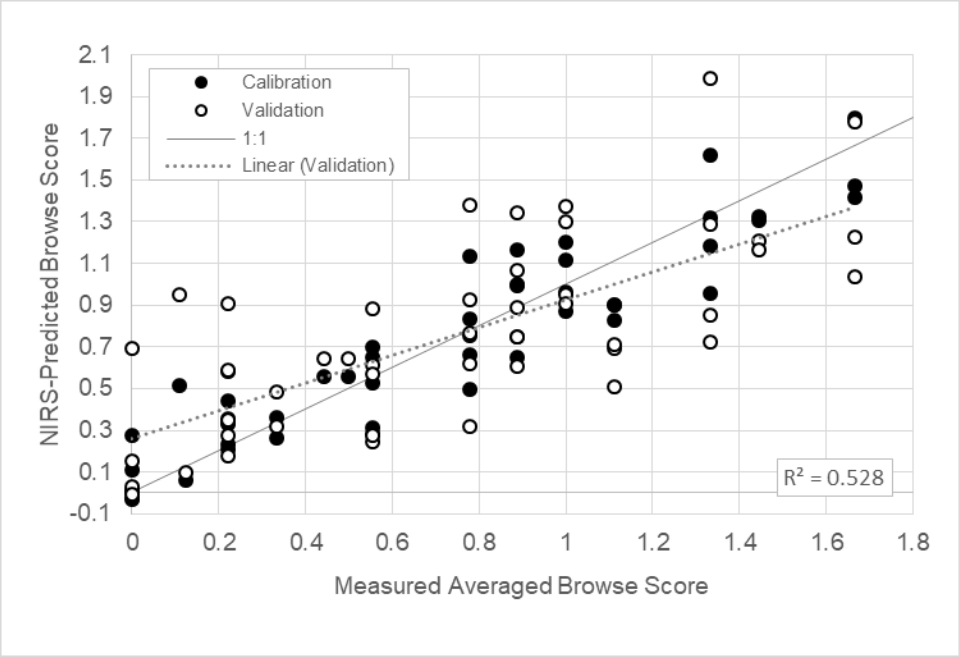


Figure 18. An empirical PLS1 model result for estimating average browse scores based on variation in NIR absorbance from leaf samples collected from the Boardman Experimental site.

Bark consumption by voles

In 2013 we conducted controlled feeding experiments with wild captive voles at the WSU-Puyallup Goss Research Farm using a selection of poplar species and hybrids from GreenWood Resources accessions we characterized in 2012 having extreme and intermediate levels of total phenolic glycosides in bark tissue.

Eleven wild voles (*Microtus townsendii*, and *M. oregoni*) were live-captured in Sherman traps from various locations on WSU research properties, 8 were used in tests; 3 were kept as standby in case of pregnancy or mortality. Each vole was individually housed outdoors in a 1.25 m diameter x 0.6 m high metal pen provisioned with silty-loam soil to a depth of 20 cm for burrowing, clumps of transplanted grass, sections of mature poplar bark and woody debris for shelter material, dried moss, grasses, and jute twine for nesting materials. The openings of the pens were covered with wire fabric and a wood-frame access door. Environmental conditions inside the pens were modulated by partially covering the top with burlap shade cloth or plastic sheet covering on rainy days (Figure 19). All voles received a daily variable diet regimen of apple wedges, rodent chow, bird seeds, rolled oats, and fresh clover totaling approximately 30-50% of their body weight. There were modest diet restrictions on the first day of each three-day feeding test period, and food in excess of the estimated *ad libitum* quantities were supplied for one recovery day following each test. Measured *ad libitum* food consumption could not be determined because voles normally store their food in their burrows. After completion of the tests, when the voles were released into a natural area, excavation of the pens revealed food middens stocked with seeds and rodent chow. *M. oregoni* voles, reported to be fungivores, had tunnel galleries where mature fungi were found growing on the food. Vole capture and pen

acclimation began July 17th, 2013; feeding trials began on August 21st, and tests were concluded on September 30th, 2013. Animal capture was approved by the Washington Department of Wildlife; animal housing, and welfare protocols were approved by WSU- IACUC.

For each feeding test, stem segments from 6 genotypes were presented together as a choice group. To keep the choices in a group related, and to coarsely explore the effect of genetic hybridization on palatability, choice group 1 was comprised of three *Populus trichocarpa* varieties together with three *P. deltoides* x *P. trichocarpa* hybrid genotypes; choice group 2 contained three *P. maximowiczii* varieties together with three *P. deltoides* x *P. maximowiczii* hybrid genotypes; and choice group 3 was comprised of three genotypes of *P. cathayana* and three from *P. koreana*. Similar to the Pullman deer study, tests with the same choice group were repeated three times sequentially to determine if there were learned avoidance responses over time.

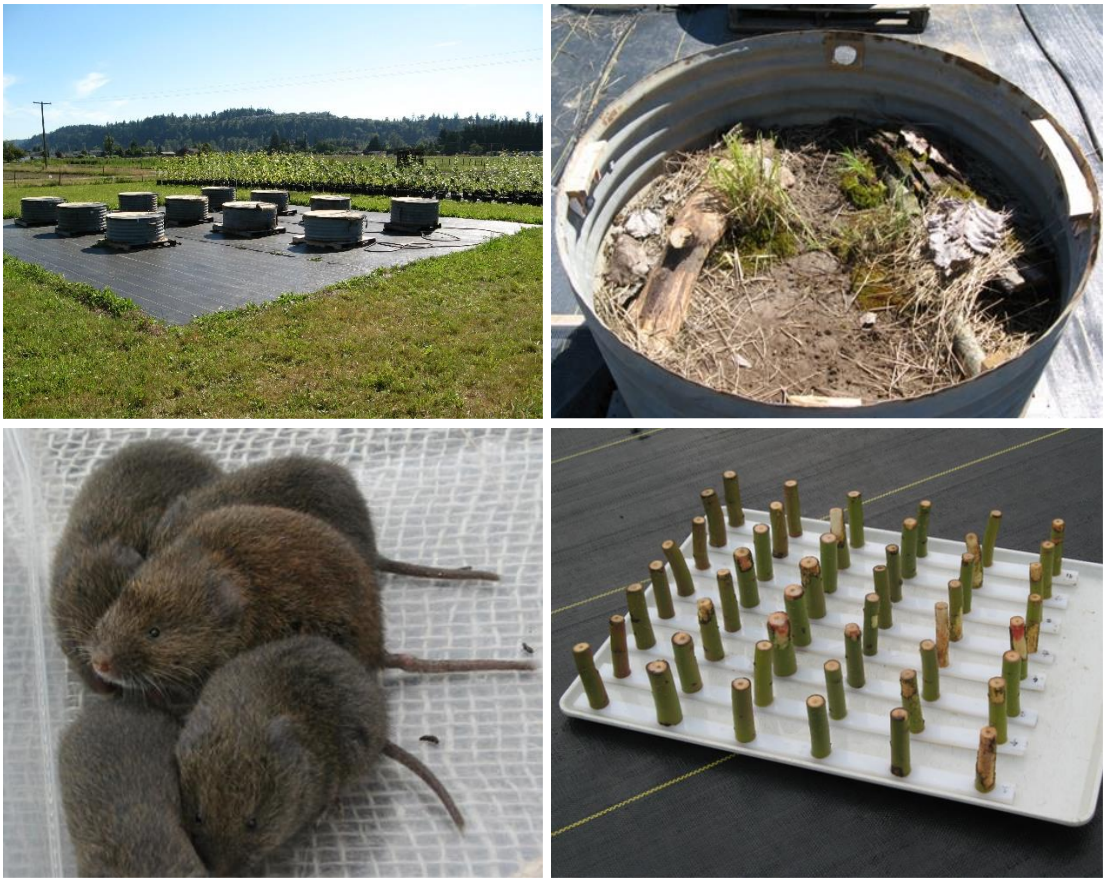


Figure 19. Top left: study site with metal pens, potted poplar trees; Top right: vole pen (wire fabric top removed) with soil, shelter and nesting materials, and planted vegetation; Bottom left: *Microtus* test subjects; Bottom right: racks of poplar stem segments screwed to plastic strips after three days in pens. Note the natural red stain in areas where bark is consumed.

Stem segments (7 cm long) were harvested 4 cm above the soil immediately before each test from potted trees, which were propagated from stem cuttings and grown outdoors for 4 months. The segments were screwed bottom-down to a 20 mm wide plastic rack (Figure 19). All vole test subjects received identical racks of stem segments from the same choice group on the same day.

After the three-day test period, consumed bark area was measured by wrapping a piece of clear acetate tightly around the segment, tracing the perimeter of the area chewed away, filling in the traced area with a black marking pen, then measuring the blackened area using a laser leaf area meter. Over the entire testing period, each vole was presented with 9 racks of stems segments (3 sequential tests of three genotype choice groups), each lasting three days, with one recovery day between each test. The racks of segments were temporarily removed and sprayed with water every day to prevent desiccation, and reoriented in the pens to randomize position.

Bark samples harvested for chemical analysis and NIRS scans, were pooled from four individual trees of each of the 18 genotypes, cut from stem material immediately above where the test segments were cut. The bark (including phloem tissue) was lyophilized, milled, chemically analyzed, and scanned for NIRS absorbance using methods identical to the initial 2012 leaf sample collection methods (see appendix). A summary of analyte concentrations and feeding test results are shown in Table 7.

There were significant differences in averaged values for bark consumption among genotypes, measured as the area of bark consumed from a genotype divided by the total bark area consumed from all genotype segments on the rack during a test, averaged over all voles (Figure 20). There were also significant changes in consumption between sequential tests in a choice group suggesting learned preference over time. The intent of normalizing consumption in terms of proportion of total area consumed in test, and averaging over all voles, was to account for differences in individual vole physiology, body mass, and feeding habits. The range of variation in consumption among voles was quite substantial, with a maximum value of 37.3 cm^2, minimum of 0 cm^2, and an overall percent coefficient of variation (%CV) = 153%. The most favored genotype (G-7175), having an average consumed area of 9.8 cm^2 over three tests, had a %CV = 117%. So, while there were significant differences among *averaged* consumption values, the number of test subjects needed to accommodate the extent of this response variability in order to detecting small effect sizes with reasonable statistical power should have been much larger than 8 animals. We did not have the facilities, animals, or time to scale these experiments appropriately.

Another shortfall of these experiments was the range of variation in analyte concentration among the potted saplings in a taxon group turned out to be much narrower than our initial 2011-2012 phytochemical characterizations indicated. This had a negative impact on analyte-consumption regression modelling.

There was no agreement in correlations between consumption and analyte concentration among the three choice groups, with exception to HCH-salicortin which was unexpectedly positive in all groups (Table 8). One unknown compound whose chromatographic peak measured at 8.2 minutes retention time (m/z ratio = 405) was included because it was negatively correlated with consumption for choice group 1, but its concentration is below method reporting limits for many Asian species, and hybrids of Asian species. If one includes the minimum detection limits for this compound in place of the missing values, the correlation with consumption was negative in all choice groups.

.

Table 7. Analyte Concentrations and Feeding Test Summary

												Bark Area Consumed (cm^2)			Averaged Proportion of Total Area Consumed in a Test							
												All Tests			Test 1		Test 2		Test 3		All Tests	
Choice Group	Taxon-Clone	BCA Protein (% dw)	Total Phenolics (µg/mg)	CT (Abs/mg)	Salicin (% dw)	Salicortin (% dw)	HCH-salicortin (% dw)	Trem (% dw)	Salirps (% dw)	Total PGs (% dw)	Unk 8.2 (% dw)	Sum	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error
1	T-1909	3.566	58.4	0.316	*	13.95	0.40	0.35	2.63	17.36	1.36	62.5	2.60	1.083	0.133	0.032	0.037	0.021	0.136	0.091	0.102	0.033
	T-6962	3.978	58.5	0.339	*	10.73	0.55	0.38	2.47	14.17	1.91	18.1	0.75	0.327	0.049	0.020	0.046	0.019	0.014	0.011	0.037	0.010
	T-2136	3.371	62.1	0.409	*	14.39	0.50	0.33	2.63	17.90	1.80	14.2	0.59	0.168	0.077	0.042	0.041	0.015	0.045	0.024	0.054	0.017
	G-4491	5.274	69.31	0.428	0.06	13.04	1.07	0.35	1.79	16.31	1.08	80.1	3.34	1.211	0.125	0.036	0.140	0.053	0.285	0.092	0.183	0.039
	G-7188	7.266	66.45	0.498	*	12.04	0.39	0.14	0.93	13.54	0.79	47.0	1.96	0.401	0.141	0.038	0.137	0.050	0.142	0.065	0.140	0.029
	G-7175	5.647	70.38	0.402	0.12	12.79	1.16	0.16	1.48	15.70	0.77	234.7	9.78	2.344	0.475	0.088	0.474	0.123	0.377	0.092	0.442	0.057
2	M-15031	6.806	42.2	0.175	*	10.47	1.39	*	2.70	14.60	*	109.1	6.03	1.704	0.250	0.081	0.115	0.052	0.268	0.114	0.211	0.050
	M-14949	6.361	34.6	0.108	0.08	6.70	1.58	0.00	3.21	11.58	*	109.6	5.53	1.378	0.103	0.050	0.300	0.065	0.246	0.076	0.217	0.040
	M-14951	7.125	31.4	0.076	0.08	1.81	0.63	0.11	3.01	5.64	*	38.6	1.77	0.454	0.023	0.014	0.110	0.015	0.097	0.044	0.077	0.018
	DM-7416	5.458	60.8	0.231	0.11	11.70	1.80	*	1.87	15.49	0.63	89.3	5.20	1.006	0.293	0.063	0.094	0.023	0.214	0.077	0.201	0.037
	DM-6294	3.954	60.4	0.345	0.20	16.18	1.72	*	2.14	20.25	1.03	54.3	2.56	0.605	0.155	0.093	0.227	0.079	0.072	0.023	0.151	0.042
	DM-8005	6.866	69.5	0.456	0.14	4.59	1.00	*	1.90	7.63	1.05	81.0	3.54	1.367	0.176	0.075	0.153	0.067	0.102	0.052	0.144	0.036
3	C-14811	4.943	40.3	0.105	0.06	7.23	1.12	0.00	2.11	10.52	*	148.6	6.19	1.489	0.317	0.104	0.098	0.037	0.256	0.075	0.224	0.047
	C-14832	5.363	32.5	0.078	0.13	5.13	1.87	*	3.41	10.55	*	253.0	10.54	1.745	0.281	0.062	0.376	0.058	0.335	0.090	0.331	0.040
	C-14772	4.089	31.6	0.075	0.22	3.85	1.42	*	3.19	8.68	*	109.6	4.56	0.999	0.045	0.013	0.160	0.037	0.246	0.062	0.151	0.029
	K-14723	4.446	32.6	0.143	0.18	7.24	2.36	0.00	3.82	13.61	*	94.9	3.96	0.980	0.162	0.058	0.118	0.035	0.096	0.033	0.125	0.025
	K-14607	4.457	27.6	0.099	*	1.86	0.36	0.11	1.74	4.10	1.06	108.9	4.69	1.059	0.148	0.043	0.224	0.046	0.044	0.025	0.139	0.026
	K-14629	4.560	49.9	0.209	0.08	1.22	0.15	0.22	2.56	4.23	1.26	22.6	0.94	0.223	0.047	0.014	0.024	0.009	0.022	0.008	0.031	0.006

Analyte concentrations are primarily expressed as percent of dry sample weight. Values below statistical reporting limits or method detection limits are indicated (*). Relative condensed tannin levels (CT) are expressed as visible absorbance of cyanidin at 550 nm in acid butanol per mg of dry sample weight. Total phenolic glycosides (Total PGs) are the sum of salicin, salicortin, HCH-salicortin, tremulacin (Trem), and salireposide (Salirps) weights. Unknown compound with mass/charge ratio = 405 (potentially phenolic glycoside) with retention time of 8.2 minutes (Unk 8.2) was included because of correlation with bark consumption. Feeding preference expressed as the proportion of total bark area consumed in a test, is the ratio of bark area consumed by a vole from a genotype divided by the total bark area consumed in a test averaged (either by test or all sequential tests) for all voles.

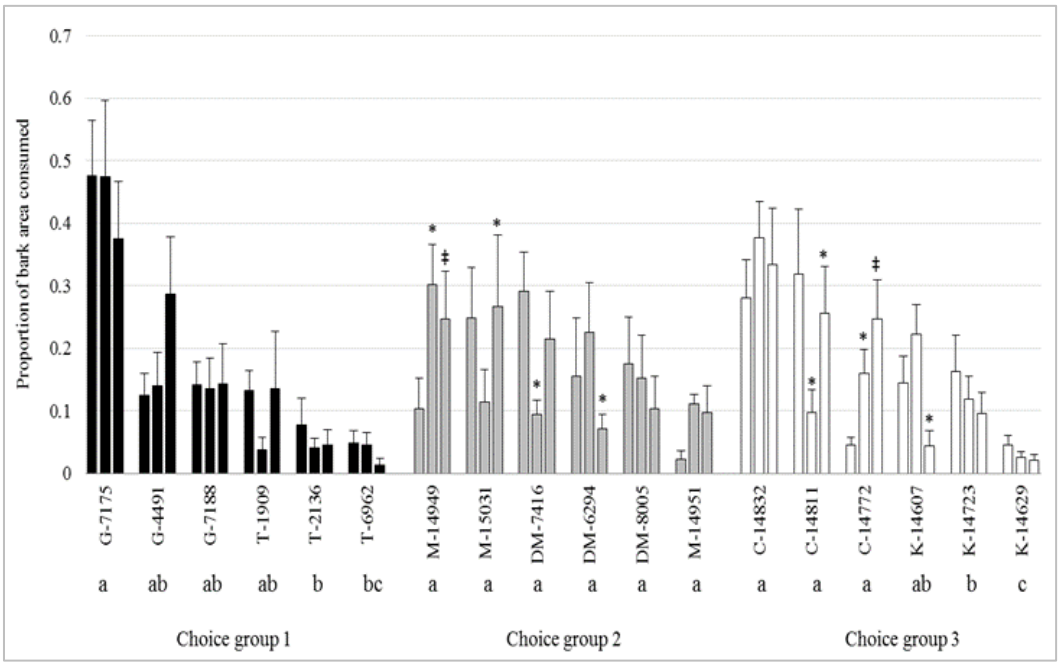


Figure. 20. Mean proportion of bark area consumed from genotypes in a test for each choice group, averaged from 8 vole test subjects. Error bars represent the standard error of the mean. Taxa in each choice group are indicated: *Populus* x *generosa*; (G), *P. trichocarpa* (T); *P. maximowiczii*, (M); *P. deltoides* x *P. maximowiczii*, (DM); *P. cathayana*; (C), and *P. koreana*, (K). Similarities and significant differences in averaged consumption in all trials combined are indicated by letter. Genotypes with the same letter were not consumed significantly different. Significant changes in averaged proportional consumption from test to test are indicated (*) if significantly different from the prior test, and (‡) if significantly different from the first test.

Table 8. Correlation of bark consumption and analyte concentration

Analyte	Choice Group 1	Choice Group 2	Choice Group 3
BCA Protein	0.47	-0.17	0.75
Total Phenolics	0.78	0.08	-0.43
Condensed tannins	0.22	-0.04	-0.76
Salicin	*	-0.34	0.08
Salicortin	0.04	0.48	0.52
HCH-salicortin	0.79	0.78	0.51
Tremulacin	-0.64	*	*
Salireposide	-0.56	-0.03	0.18
Unknown 8.2 min	-0.75	*	*

Missing values (*) indicate concentrations below statistical reporting limits

In choice group 3 tests we observed a progressive increase in consumption of *P. cathayana* genotype C-14772 and a decreasing consumption of *P. koreana* clones over time, suggesting learned preference. However, we found no quantitative association with any analyte concentration or combination of concentrations that could explain this change in preference.

Generally, stem segment diameter was slightly negatively correlated with consumption. We made no qualitative assessments of bark thickness or texture, but these attributes may have influenced choice, especially on the larger diameter segments. A greater frequency of smaller diameter segments were observed for the Asian genotypes.

Vole species effects on preference added a confounding layer to interpreting the data. We found significant differences in total area consumed attributable to vole species only in choice group 3 tests. Generally, *M. oregoni* voles exhibited strong preference for *P. cathayana* genotypes over *P. koreana,* and *P.* x *generosa* over *P. trichocarpa,* whereas the *M. townsendii* voles did not prefer any taxon in a choice group test.

Attempts at developing multiple linear regression models to explain proportional bark consumption for each of the choice group tests were thwarted by the extreme variability in consumption among too few test subjects. Using non-averaged consumption data, models explained only 4% to 39% of consumption, with no agreement in analyte parameters among choice groups. Using consumption data averaged for all voles in a test, models explained 0% to 69% of consumption, and frequently the analyte variables selected using forward and backward methods were different than the parameters identified using non-averaged data.

The upshot of our discovery questions: (1) whether putative 'defensive' analytes effect bark consumption by voles; we found that none of the defensive compound concentrations in these particular poplar genotypes, alone or in combinations, explain useful amount feeding preference. (2) Regarding the effects of genetic hybridization on analyte composition in bark we observed that hybridization of *P. trichocarpa* and *P. maximowiczii* with *P. deltoides* (in these particular genotypes) was associated with increased concentrations of total phenolics, condensed tannins, and HCH-salicortin, and decreased concentrations of tremulacin and salireposide. However, these trends were not consistent with the foliar concentrations found in the same species and hybrid counterparts used in the confined deer study.

To develop phytochemical criteria for selecting hybrid poplar genotypes that possess low herbivore palatability using NIRS requires a reference dataset arising from animal feeding experiments. One output of such experiments is the quantitative mapping of consumption to a combination of analyte concentrations using multiple linear regression. Another possible output is the placement of each genotype into a ranked feeding preference scale, and to map genotype-specific NIRS spectral patterns to that scale – working under the assumption that there are many unknown NIRS-absorbing phytochemicals, each with unique spectral patterns, which act additively to affect preference. Regardless of the approach, animal feeding experiments are challenged by variability and combinatorics. Clearly, plugging NIRS-estimated analyte concentrations into a MLR model was not achievable because the amount of consumption of any genotype depends on the limited array of choices presented, and because of compounding errors of each estimate in the model. In the following Cottonwood leaf beetle experiments we took the second approach; establishing a ranked feeding preference scale among multiple genotype choices, and mapping genotype-specific NIRS spectral patterns to that scale using partial least squares regression models.

Milestone 2 Identify insect-resistant and susceptible clones from plantation damage assessments and conduct controlled feeding preference tests.

Introduction

The cottonwood leaf beetle (*Chrysomela scripta*,) is one of the most important economic pests of poplar plantations in North America (47). Selection of poplar clones possessing natural insect-resistance phytochemicals is a potential approach to an integrated pest management strategy. In numerous studies, genotype-specific susceptibility of poplar and aspen genotypes to defoliation by *C. scripta* has been observed (48) (49) (50). To understand the chemical basis of this differential preference investigations have proceeded along several lines: identifying chemical attractants originating from the plant (51) (52); identifying attractants originating from the beetle experiencing a host (53); and identifying primary metabolites and deterrent plant secondary metabolites that affect feeding preference and larval performance (54). The problem of understanding *C. scripta* feeding preferences is complicated by the fact that it has coevolved with, and is exclusively adapted to host trees of *Salicaceae* (55). To repel their natural enemies *C. scripta* larvae present defensive secretions of salicylaldehyde from dorsal exocrine glands, which they convert from ingesting the phenolic glycosides salicin and salicortin. *C. scripta* has effectively co-opted the plant's anti-herbivore defense for its own protective use (56).

Each hybrid poplar genotype has a unique combination of primary and secondary metabolites, digestible and non-digestible structural components, and anti-nutritive factors that inform the adult beetle about food quality and impinge upon its decision to stay, feed, lay eggs, or move elsewhere. To our knowledge, no single naturally occurring phytochemical or anti-nutritive factor has been shown to be a sufficiently strong predictor of attraction or resistance in hybrid poplar to be considered a candidate for quantitative trait selection.

Modelling foliage preferences of the cottonwood leaf beetle

Our study capitalized on a widespread *C. scripta* infestation at the GreenWood Resources research nursery and plantation near Boardman Oregon in 2014. A progeny trial comprised of 468 *Populus deltoides x maximowiczii* genotypes in 18 families and 272 *deltoides x nigra* genotypes in 10 families, established as ramet plantings and replicated randomly in four blocks (2858 trees total), was the source of most of the genotype selections used in our study based on contrasting severity of beetle attack. We also made selections from their nursery comprised of 921 seedlings in 5 hybrid taxa and 8 maternal families. Our study program is diagramed in Figure 21.

Since each genotype was replicated only 4 times in the trial we attempted to visually rank the severity of beetle attack on a refined scale from 0 to 2.33: 0 = little or no damage; 0.33 = less than 5%; 0.66 = greater than 5% but less than 10%, 1 = approximately 10%; 1.33 = approximately 15%; 1.66 = approximately 20%; 2 = approximately 25%; and 2.33 = greater than 25% damage. The non-quantitative visual assessments were made by one observer, spending no more than 30 seconds at each tree. Ranking damage of the nursery seedlings in their second year of regrowth after coppice was coarser: 0 = no leaf damage; 1 = less than 10% damage; 2 = greater than 10% but less than 33% damage; 3 = greater than 33% damage. Fifteen of the 24 genotypes used in the controlled feeding experiments, and 69 genotypes used for model verification came from this progeny trial and nursery. Seven of the 24 genotypes used in the controlled feeding experiments came from GreenWood Resources Boardman tree farm trials for which we had phytochemical data. Two of the 24 genotypes used in the feeding study came from

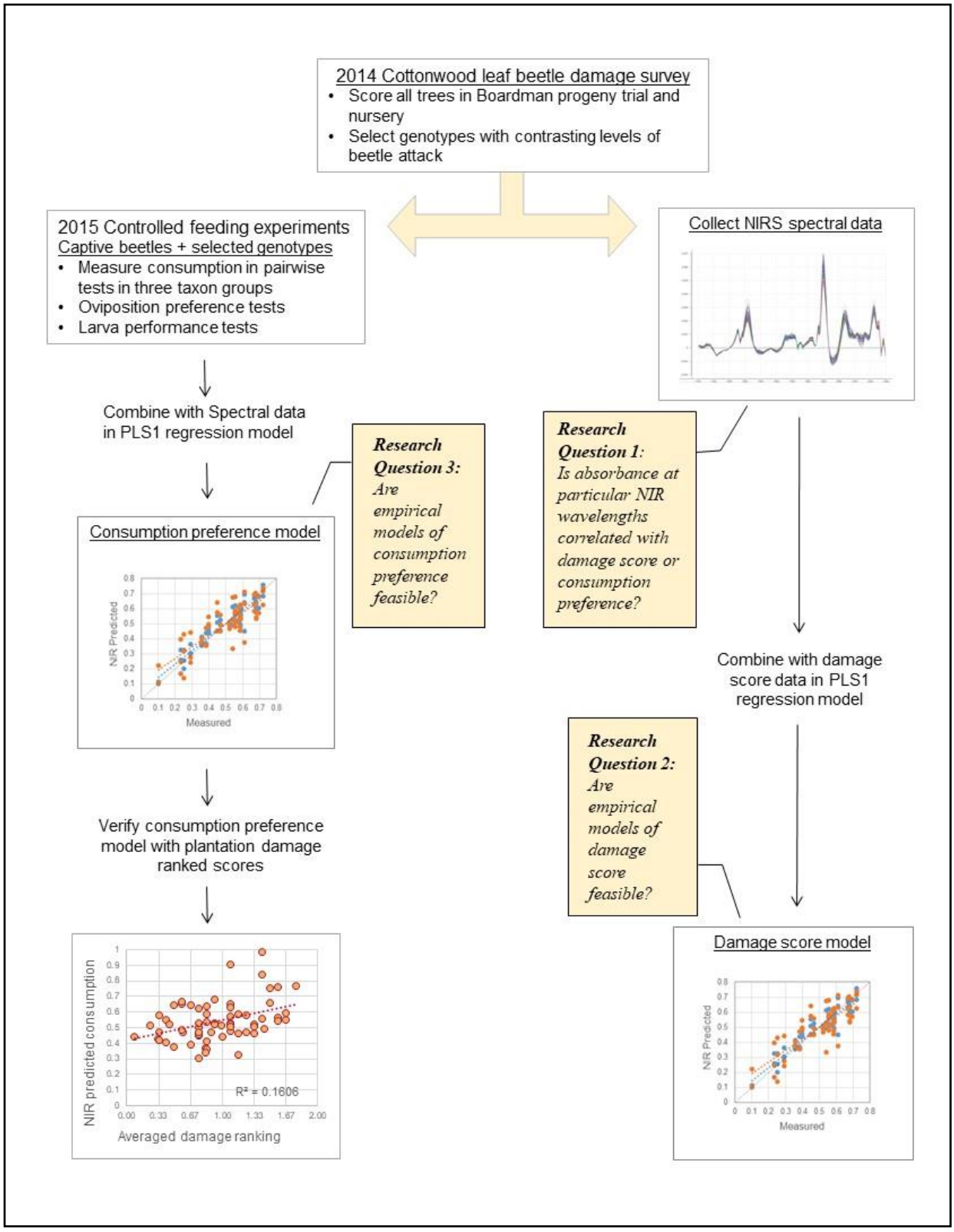


Figure 21. Study flow diagram

WSU poplar progeny trials in Puyallup for which there was no phytochemical reference data or beetle damage assessment. Multiple leaves from lateral branch terminals were harvested from each selected genotype, flash frozen in the field, lyophilized, milled, and scanned using NIRS (see methods in appendix).

NIRS spectral data analysis of selected genotypes

Our first investigation was to determine whether differences in NIR absorbance spectra might be apparent between genotypes scoring high in beetle attack versus genotypes scoring low. Such differences might be helpful in eliminating non-informative spectral variables in modelling efforts and indicate classes of chemical functional groups that play a role in deterrence or attraction of beetles. For comparison we selected five full-sibling pairs from four breeding families within the hybrid taxon *P. deltoides x maximowiczii* to eliminate any confounding spectral differences that could arise from comparisons of different hybrid taxa. The NIR foliage spectrum of the highly damaged sibling was subtracted from the spectrum of sibling with low damage (Figure 22). The damage rank difference between each full sib pair was approximately equal to 1, or three damage rank increments.

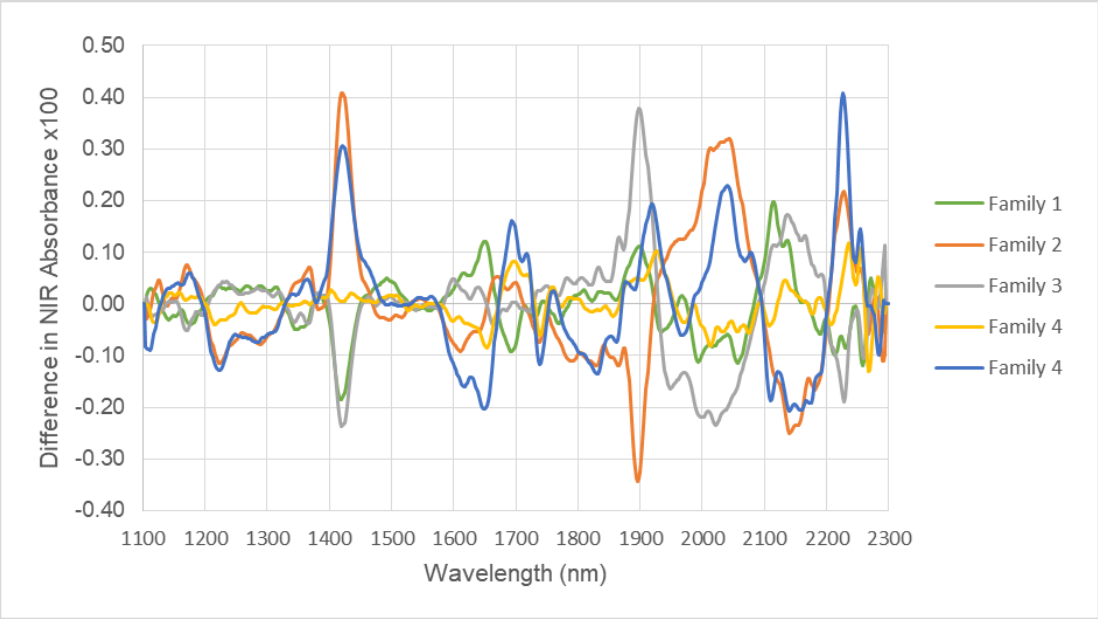


Figure 22. Each plot line is the NIR absorbance difference (x 100) from a genotype with high average beetle damage score minus its full-sibling having a low average beetle damage score. Five full-sib contrasting pairs were selected from 4 breeding families within the DM hybrid taxon. All NIR spectra were transformed with Savitsky-Golay first derivative before subtraction.

The peaks at 1420 nm and 1900 nm are primarily due to slight differences in residual water, but also 1420 is a common band for OH groups attached to phenolics. Absorbance at wavelengths from 1630 nm to1690 nm are associated with absorbing vibrations from C-H vinyl, C-H methyl, C-H nitro, and C-H aromatic functional groups; absorbance between 2000 nm and 2080 are associated with N-H functional groups in amides, amines, and proteins; absorbance around 2100 nm is associated with carbohydrate and polysaccharides; absorbance around 2150 nm is associated with C-H aromatic functional groups; and wavelengths around 2270 are associated with O-H stretching in functional groups common in glucose and cellulose.

A sobering take-away was that the differences were not aligned in the same direction, which we might expect if the compared genotypes had the same levels of attracting phytochemical factors and differed only in deterrent compounds. But a beetle damage rank difference of 1 can arise in several ways. Consider the possibilities in Table 9.

Table 9. Balanced factors in beetle damage rank

Attraction Factors	Deterrence Factors	Beetle attack
+	-	High rank
+	+	Intermediate rank
-	-	Low rank
-	+	Very low rank

There is no guarantee that either attraction or deterrence factors are held constant while the other factor varies in these comparisons. Almost all of the category pair combinations in Table 9 can result in a damage rank difference of 1. So, the spectral subtraction study merely highlighted regions of the spectrum where there was greater NIR absorbance variability, associated with multiple compounds, which may have given rise to differences in beetle attack.

Another possibility for explaining the lack of concurrence among full-sib spectral subtractions was inaccurate scoring. Aside from the crudeness of the visual ranking method, there was high uncertainty in the low damage ranks. We could not assume uniform distribution of beetles at the site. Undamaged trees may have possessed high levels of deterrent compounds, low levels of attraction compounds, the beetles may not have arrived at the tree, or there wasn't enough time for the beetles to cause damage. In contrast, there was high confidence in the rankings of highly damaged trees. Such scoring inaccuracy may be evident in the spectral subtraction of a full sibling pair in family 4 (yellow plot line, Figure 22) showing negligible absorbance differences even though they had a damage score difference of 1.1.

Empirical NIRS model of beetle damage rank

Given the uncertainty in ranking low-damaged trees, we utilized principle component analysis of scores and loading plots to identify and eliminate samples that may have been inaccurately ranked. Recall that in PCA analysis a dataset undergoes 'matrix decomposition' where the original matrix of variables (the NIR spectral data) becomes the product of a 'score' matrix and a 'loading' matrix. The 'score' matrix, contains all the information about the interrelationships among samples, the 'loading' matrix contains all the information about the interrelationships among the variables. By plotting PCA sample 'scores' we identified incorrectly ranked samples as outliers by their distance from the 'average object' (model center), and also by their distance from samples having the same damage rank. More specifically, in graphical analysis of 'scores' plots, objects that are similar tend to group in clusters or align in patterns in relation to the principal component axes. In our case, samples were clustered primarily into taxon groups in the lower-order principal component 'score' plots. In higher-order 'scores' plots the samples clustered and aligned differently. When a low beetle damage-ranked sample was consistently clustered with high damage-ranked samples in several 'scores' plots, that sample was flagged as an outlier for removal during model development. Conversely, when a high damage-ranked sample was not in a cluster with the others in its taxon

group it was flagged. Using this method, and also by flagging all samples with high leverage and high residual variance, we eliminated only 9 of the 69 collected samples in developing the NIRS-based damage rank model (Figure 23), suggesting that 86% our evaluations of damage were roughly correct. The optimized model required eight PLS1 principal component factors to minimize residual variance The model had a root mean square error of cross-validation = 0.232, or about 10% of the full range of damage ranks; and a standard error of prediction = 0.165.

We also sought to optimize the model by silencing absorbance variables that were uninformative. Wavelengths associated with absorbance variables having consistently small loading weights in the first four PLS1 principal components were identified and excluded from the model on the condition that the model's explained variance increased, root mean square error of cross-validation decreased, and fewer PLS1 components were necessary to minimize cumulative residual variance. Significant spectral variables were also identified using the Martens' uncertainty test – an algorithm embedded in the Unscrambler software package. The remaining informative absorbance variables selected at wavelengths for the optimized model are graphically presented in Figure 24.

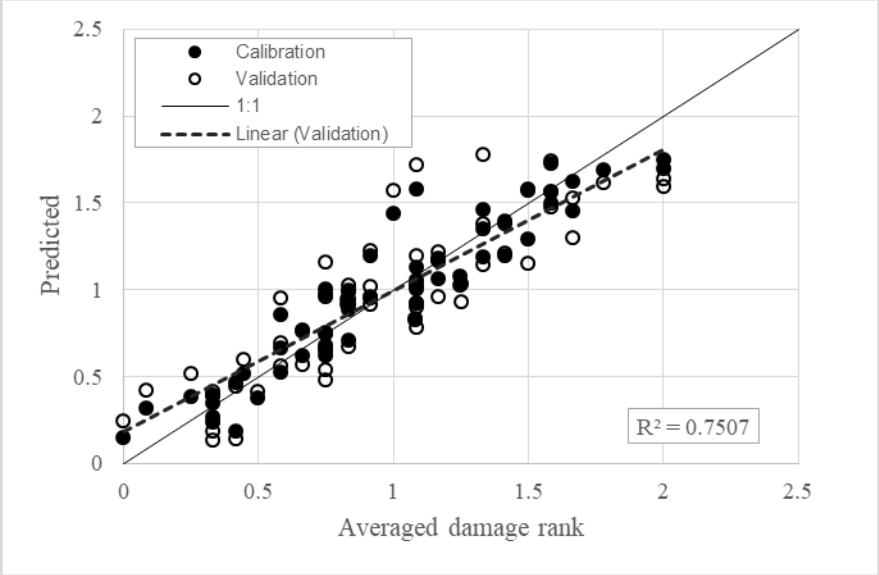


Figure 23. Optimized beetle damage rank regression model based on select NIR absorbance values of 60 dried and milled foliar samples and associated averaged damage rank for each genotype (n = 4). The R-square value is the explained variance of random segmented cross-validation predictions.

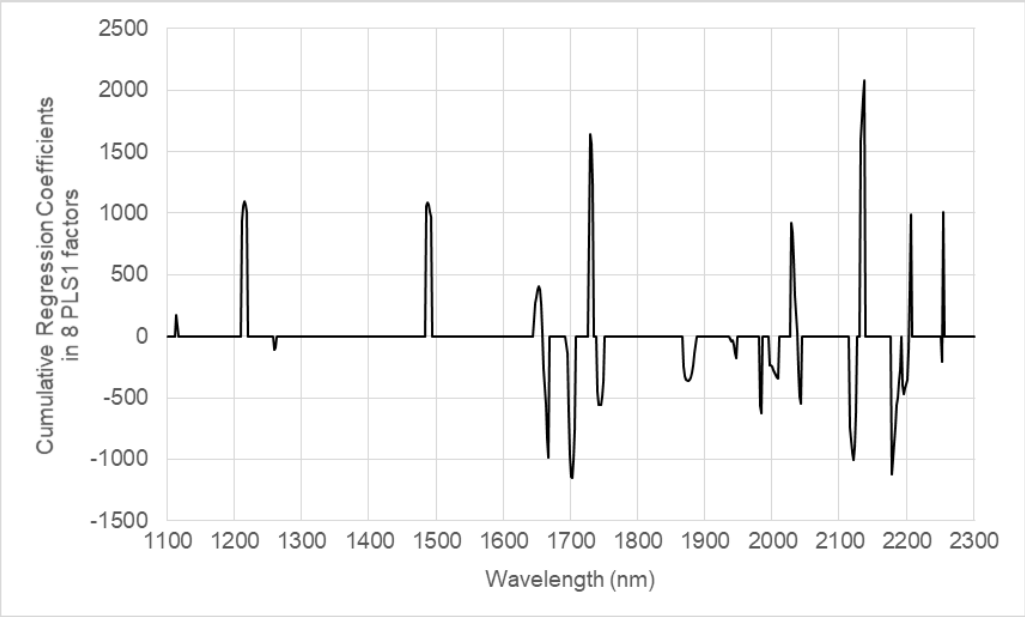


Figure 24. Cumulative regression coefficients for absorbance values at select wavelengths across the NIR spectrum from 1100 to 2300 nm used for the optimized beetle damage rank model (Figure 23). Silenced absorbance values have cumulative regression coefficients = 0.

While the optimized beetle damage-rank model provides further proof that empirical models of herbivory are feasible, the accuracy and precision of these particular predictions would be sub-optimal for general use in genotype selection unless the new test populations are confined to DN and DM hybrid genotypes, and, as will be demonstrated below, the tree age, leaf age, and growing conditions are identical. Overall, 75% explained variance in cross-validation predictions seems an unlikely result given such coarse field measurements of herbivory and little replication for computing genotype averages. The practice of silencing portions of the spectrum considered "noisy" or irrelevant to derive an improved and more parsimonious model comes with risks, although the technique is occasionally seen in published reports. The analyst is directly influencing the model result by choosing variables instead of letting the PLS1 regression algorithm compute vectors of maximal variance for each principal component based on full-spectrum loading weights and regression coefficients. While prediction errors are diminished with the reduced variable model, it is also possible to derive a good model by silencing important wavelengths, leading one to question the extent to which the result is artificial! Just because a model can be "calibrated" does not necessarily give it logical or actual validity. The practice of variable selection essentially turns up the volume on any variable having high covariance with the reference data while drowning out any subtleties that might be useful in making the calibration more robust (and less precise) to slight NIR absorbance band shifts that arise from variability in sample preparations and scanning temperature. A reduced band-width model is also more likely to be idiosyncratic to the calibration sample population, producing poor predictions for variants with slightly novel absorbance profiles.

Comparing the selected variable wavelengths of the optimized damage rank model (Figure 24), with the wavelengths corresponding to maximum differences in absorbance in paired spectral subtractions among full-siblings having contrasting damage scores (Figure 22), we observed some common bands (Table 10).

Table 10. Important absorbing chemical structures

Common Wavelengths (nm)	Reported NIR absorbing structure and compound type*
1428-1430	N-H, R-C=NH2, primary amides
1594-1606	C=O/N-H, polyamide
1638-1648	CH2=CH-, C-H-vinyl
1690-1698	CONH2 peptide bond
~1820	OH/CH combination, cellulose
~2056	N-H combination, amide and protein
2136-2140	C-H/C=O combination, lipids
2146-2148	N-H/C-N/C=O combination, lactams, C-H aromatic

* NIRS absorption bands, functional group structures, and compound types per *Practical Guide to Interpretive Near-Infrared Spectroscopy*, Workman and Weyer, 2008, CRC Press

There are many classes of compounds, both nutritive and potentially deterrent, that absorb NIR photons at these frequencies. Researchers who have described the phytochemical composition of young leaves in *Populus tremuloides* have shown that primary amides and polyamines like valine, asparagine, proline, glutamic acid, glutamine; and carbohydrates such as malate, citrate, and glutaric acid are very abundant in young foliage and that some of these rapidly decline in concentration as leaves age from LPI-0 to LPI-4 (57). (*C. scripta* beetles prefer to feed on leaf numbers 2 to 4). Five of the eight common bands listed in Table 10 are associated with nitrogen containing compounds that would include those described for young foliage, only one is aromatic, which could suggest that poplar genetic variation resulting in different gradients of nutrient compounds may play an important role in feeding preference as attractants, except, the other nine selected wavelengths of the damage rank model are associated with phenolics, thiols, lipids, aliphatic hydrocarbons, and ketones. The mélange of chemical factors impacting feeding preference defies obvious interpretation. Without precise and comprehensive chemical profiling in each age-class leaf in every genotype, coupled with feeding tests and principal component analysis, a phytochemical understanding of feeding preference will remain, at best, crudely explained. To that end, NIRS provides no shortcut.

Controlled feeding experiments with Cottonwood Leaf Beetle

For both the beetle damage-rank model above and the wild deer browse model, field assessment of herbivory produced unsuitable dependent variable data sets in the following ways:

- Estimates were based on non-quantitative visual impression; impossible to reproduce.
- Even with good replication of genotypes at the site for averaging damage, the range of damage in each rank increment was too large, although convenient for rapid assessment. The wide dispersion of estimates within the narrow range of rank variables contributed to imprecise regressions.
- The opportunity for herbivory was uncontrolled and the feeding decision process was unmonitored resulting in great uncertainty about trees with little or no damage.

For the controlled feeding study with confined deer, the accuracy and precision of the models were limited by the method for estimating leaf area of potted saplings prior to

presentation, but more importantly, limited by inferences about preference that could be made due to combinatorics. In the following study with leaf beetles we eliminated visual ranking of herbivory, and located each genotype on a preference scale by accurately measuring leaf area consumed in all permutated pairwise tests of an array of genotypes. By constructing a self-sustaining colony of beetles we would not be limited by test subjects. The NIRS application would be in mapping patterns of NIR absorption to the preference scale in an empirical consumption calibration. But there were stumbling blocks.

Perhaps with all designed experiments for deriving models of animal feeding behavior the researcher must balance the need for sufficient number of data points for the regression against the requirements of sufficient test replication to statistically account for natural variability – a quantity that may be unknown until the experiment is underway. When animals are feeding on plants that are growing for a limited amount of time, there will always be some element of the experiment that will be limiting, whether it's the number of test subjects, the number of testing spaces, the amount of feed required to propagate and sustain the test subjects, the amount of greenhouse space needed to propagate test material, the amount of land area required to grow the test material, etc. Typically there will be at least one of these elements for which one cannot purchase additional capacity, thereby setting the boundaries of the experiment. In our case, the critical limitation was time. Beetles prefer to feed on young foliage at shoot terminals. Poplar shoot growth processes end in August when changes in ambient light quality drives trees to set terminal buds. In nature this environmental change synchronizes with the beetle's life history - toward reduced procreation activity and exploring habitats for overwintering dormancy. So, we had at most a 75-day window to conduct all tests.

Presuming we would conduct pairwise tests of 24 genotypes in three hybrid taxa, the experimental design was determined by how many replications of each pairwise test we could do per day in a 75 day season, given a limitation of 22 cages for both rearing and testing; each rearing cage limited to 50 beetles fed with 4 or 5 shoots per day, and each cohort of beetle test subjects requiring 26 days to mature from egg to female breeding adult. As shown in Table 11, a minimum of three replicate tests of all permutated pairs of 24 genotypes would have exceeded our capacity in testing/rearing cages and potted test shoots (limited to ~900), and vastly exceeded our ability to produce material to sustain the colony.

Table 11. Requirements for pairwise testing of 24 genotypes

	Separated by taxa 3 x (8(C)2)		Not separated by taxa 24(C)2	
Replication of pairwise tests	3	5	3	5
Pairwise tests	252	420	828	1380
Tests per day in 75-day season	6	9	17	27
Potted test shoots	744	1192	2376	3912
Total female test beetles	504	840	4968	8280
Rearing cages (50/cage)	13	26	39	78
Total poplar rearing shoots	2250	4500	6750	13500

The only way to generate a data set was to do triplicate pairwise tests within each of the three hybrid taxon separately, then conduct a reduced number of strategic cross-taxon pairwise tests to confirm that each hybrid taxon's quantitative feeding preference scale was aligned with the others. That would still leave some uncertainty as to whether the three sets and their associated NIR spectral data could be combined into one global data set.

We selected 23 hybrid poplar genotypes in 3 hybrid taxa for controlled feeding experiment tests: *Populus trichocarpa* x *P. deltoides* (TD), *P. deltoides x P. maximowiczii* (DM), and *P. deltoides x P. nigra* (DN). These came from a variety of sources. Twelve of the genotypes were selected from the progeny trial and nursery located at the GreenWood Resources Research Site near Boardman Oregon. These selections were based on observations of contrasting beetle damage made in 2014. Six genotypes were also selected from a coppiced experimental planting at the GreenWood Resources Boardman Tree Farm, for which we had determined foliar phytochemical composition for the wild deer browse study one year previous. The other genotypes were selected because they were commercially important, or were chosen randomly from the WSU-Puyallup poplar plantation to replace some of the GreenWood Resources TD genotypes eliminated from the study because of disease. To verify the results of this empirical model, 68 foliar samples were collected from 55 DM genotypes and 13 DN genotypes of the Boardman progeny trial to represent a maximum range of expected beetle damage. These are the same samples used in the spectral subtraction study and the beetle damage-rank model described above.

Plant material used for pairwise feeding tests, NIR absorbance scans, and for beetle colony rearing was collected in January 2015, propagated from winter-dormant branch cuttings in greenhouse conditions, transplanted into 23.5 or 13.6 liter pots, and grown outdoors on drip irrigation until harvested for tests. Additional cuttings were propagated to provide fresh shoots for beetle colony rearing were planted in mineral soil at the WSU-Puyallup farm. Additional rearing shoots were also harvested from existing WSU poplar plantation trees.

A sustained colony of *Chrysomela scripta*, initially expanded from approximately 15 adults, 25 third-instar larvae, and 12 egg masses, collected at the Boardman research site, was established at WSU-Puyallup where the feeding tests were conducted. Female beetle test subjects, reared from eggs, were used for feeding tests 6 days after emergence from pupation.

For the bioassays, a shoot with five of the youngest leaves from each of two genotypes were presented to 3 gravid female beetles in tests lasting 24 hours (see Figure 25). After 24 hours, the shoots were removed, beetles euthanized, egg masses counted on each numbered leaf and genotype, and each leaf lamina was removed and photographically scanned in a flat-bed scanner. Consumed leaf area and remaining leaf area were measured using the image analysis software *Assess v2.0*. Some unconsumed leaf lamina were selected, dried and weighed to determine specific leaf area.

The quantitative scaling of feeding preference for an individual genotype relative to the all genotype pairings within the taxon group was made by computing the leaf area consumed from all 5 shoot leaves of the genotype of interest divided by the total leaf area consumed from both genotype shoots during the test, averaged over 3 replicate tests. The averaged proportional consumption of all permutated pairings with that genotype, (π_{ai}) shown in Table 12 and Figure 26, is the relative preference value.

Figure 25. **Top left**: 38 liter glass aquarium rearing cages housing cohorts of ~50 same-aged beetles, equipped with fresh poplar shoots in sealed flasks of water and dried poplar branches to facilitate movement between shoots. Sand floor covering was provided for moisture and feces absorption. The number of egg masses per day were recorded to monitor beetle development and fecundity. **Top right**: Aquarium cage divided in half to make two feeding test chambers, each with a pair of shoots in sealed flasks of water with leaves touching for easy migration of beetle test subjects. **Bottom left**: Three gravid female beetles feeding on paired genotype leaves. **Bottom right**: Image analysis of a consumed leaf using *Assess 2.0*. Remaining leaf area was determined by measuring green pixels; consumed area was determined by measuring inscribed areas not green (shown as red area).

Gravid beetles were used as test subjects to address ancillary questions about whether oviposition preference influenced feeding preference or vice versa, and whether growth performance of larvae was correlated with the ranked preference for oviposition. In our test conditions, beetle preferences for feeding and oviposition sites may have been influenced by a combination of factors linked to leaf age, such as phytochemical differences, leaf area, and competition. Averaged over all tests in each taxon group, the proportion of leaf area consumed was greatest from the third leaf and least from the first leaf, whereas the fourth leaf was the most favored for oviposition (Figure 26). The data suggest that females chose to oviposit on genotypes preferred for food, and on average, partitioned their own feeding activity from their offspring's food source by laying eggs on an older leaf.

To determine larvae growth performance we measured differences in the rate of weight gain and also the efficiency of assimilation (weigh gain per milligram of dry leaf matter consumed) among cohorts of 6 first-instar larvae feeding on an LPI-3 leaf of each genotype, with weights determined daily and tests replicated three times. Based on the rate of weight gain we found no significant correlation between cohort performance and the relative preference of the genotype by the adult females, except for the very least favored genotypes TD A11-2 and DM 6323, where the larvae growth rates were also the slowest.

Table 12. Averaged proportional consumption in pairwise feeding tests

TD Genotype	A27-1	12804	12813	12820	12812	A11-2	6600	7300
Designate	A	B	C	D	E	F	G	H
A		0.4360	0.0410	0.3238	***0.7582***	0.1843	0.2181	0.7600
B	0.5640		0.1510	0.2348	0.3833	0.0434	0.3851	0.3797
C	***0.9590***	0.8490		0.8710	***0.7903***	0.0789	0.4813	0.4621
D	0.6760	***0.7550***	0.1290		***0.8068***	0.1125	0.3540	0.4947
E	0.2420	0.6170	0.2097	0.1932		0.0323	0.3941	0.6466
F	0.8160	***0.9570***	***0.9211***	0.8875	***0.9677***		0.7844	***0.9561***
G	0.7820	0.6150	0.5187	0.6460	0.6059	0.2156		0.3354
H	0.2400	0.6200	0.5379	0.5053	0.3534	0.0439	0.6646	
π_{ai}	0.6113	0.6927	0.3583	0.5231	0.6665	0.1016	0.4688	0.5764

DM Genotype	6323	16561	17690	17679	17930	178473	17884	17847
Designate	A	B	C	D	E	F	G	H
A		***0.7860***	***0.8866***	0.6608	0.7329	0.5060	0.8393	***0.9562***
B	0.2140		0.5435	0.2987	0.4639	0.4177	0.4800	0.6845
C	0.1134	0.4565		0.5026	0.7314	0.3459	0.4194	0.3366
D	0.3392	0.7013	0.4974		0.7591	0.4496	0.6473	0.8106
E	0.2671	0.5361	0.2686	0.2409		0.6050	0.6026	0.6937
F	0.4940	0.5823	0.6541	0.5504	0.3950		0.4896	0.6997
G	0.1607	0.5200	0.5806	0.3527	0.3974	0.5104		0.5799
H	0.0438	0.3155	0.6634	0.1894	0.3063	0.3003	0.4201	
π_{ai}	0.2332	0.5568	0.5849	0.3994	0.5409	0.4478	0.5569	0.6802

DN Genotype	OP367	18126	Nur-C	Nur-D	18187	10650	10643
Designate	A	B	C	D	E	H	I
A		0.4208	0.6839	0.3070	0.1515	***0.7792***	0.1365
B	0.5792		0.9797	0.8147	0.5489	0.7423	0.8149
C	0.3161	0.0203		0.5274	0.1828	0.6027	0.2870
D	0.6930	0.1853	0.4726		0.6192	0.3778	0.1446
E	0.8485	0.4511	0.8172	0.3808		***0.9423***	0.8032
H	0.2208	0.2577	0.3973	0.6222	0.0577		0.1195
I	0.8635	0.1851	0.7130	0.8554	0.1968	***0.8805***	
π_{ai}	0.5868	0.2534	0.6772	0.5846	0.2928	0.7208	0.3843

Table values are the proportion of leaf area consumed, averaged over three replicate tests. The scaled relative preference is the averaged leaf area proportions of the column genotype (π_{ai}). Column genotypes significantly preferred over row genotypes in paired t-test: ($p < 0.05$, $n = 3$) are indicated in ***bold italic*** font.

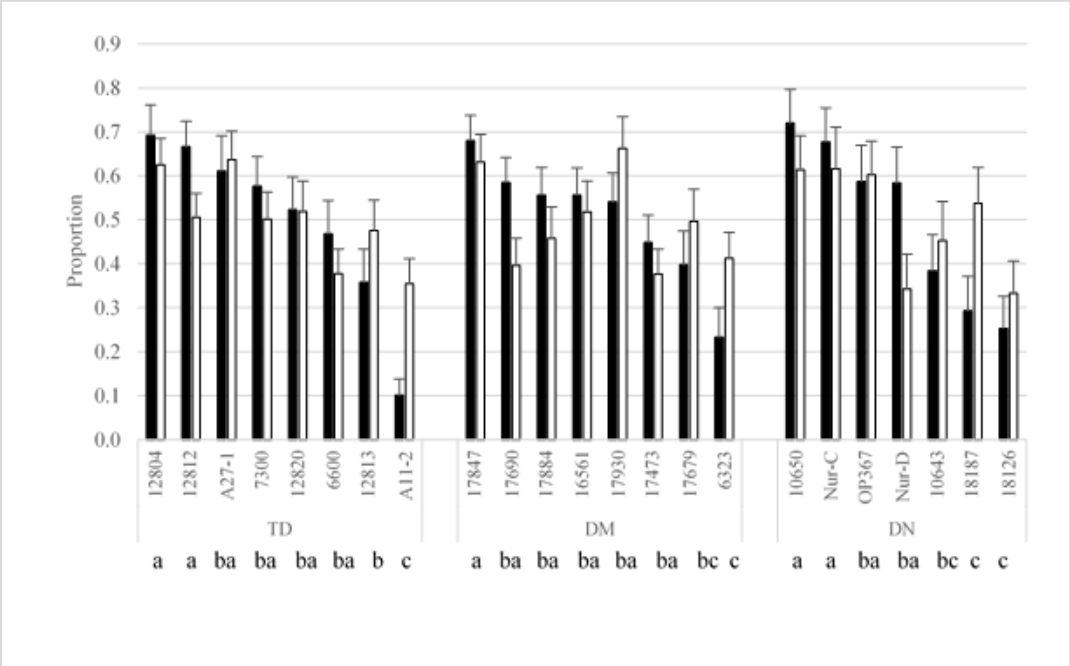


Figure 26. Proportion of leaf area consumed (filled bars) and proportion of egg masses laid (open bars) with standard error, averaged over all pairwise tests within taxon group. Significant differences in proportional consumption within hybrid taxon are indicated with Student-Newton-Kuehl grouping letters. Genotypes with the same letter(s) are not significantly different.

To address the hypothesis that genotype-specific differences in consumption might be negatively correlated with leaf toughness, which has been shown to be positively correlated with dry mas per unit area, or specific leaf area (SLA), our data show that there were significant differences in SLA among genotypes within each taxon, but no significant linear relationship between SLA and proportional consumption. In the 231 pairwise tests, only 51 had differences in SLA greater than 1mg/cm^2 at leaf number 3. Of these, the genotype with the lower SLA was preferred in only 14 tests (27%), suggesting that feeding preferences associated with leaf age were not driven by greater leaf toughness.

Our pairwise feeding test results indicate that there were few significant differences in scaled preference among genotypes in each taxon group. Overall, eleven of the 23 genotypes are scaled approximately the same. There were no genotypes that were strongly favored, and only two (TD A11-2, DM-6323) that were avoided somewhat consistently, which highlights how well the cottonwood leaf beetle is adapted to consuming its host, but also unpromising for deriving a multivariate regression model.

To determine if it was feasible to combine the scaled preference data from individual taxa into a single 'global' dataset we conducted 31 cross-taxon pairwise tests to assess whether the relative position of a genotype's preference value on a combined "global scale" held true (Figure 27, Table 13). Six of the 8 cross-taxon pairings confirmed their relative positions, given a typical 10% standard error for $\pi_{ai.}$ The poor agreement from pairings with clone TD-B left some uncertainty as to whether a combining preference values in a single data set would affect model precision.

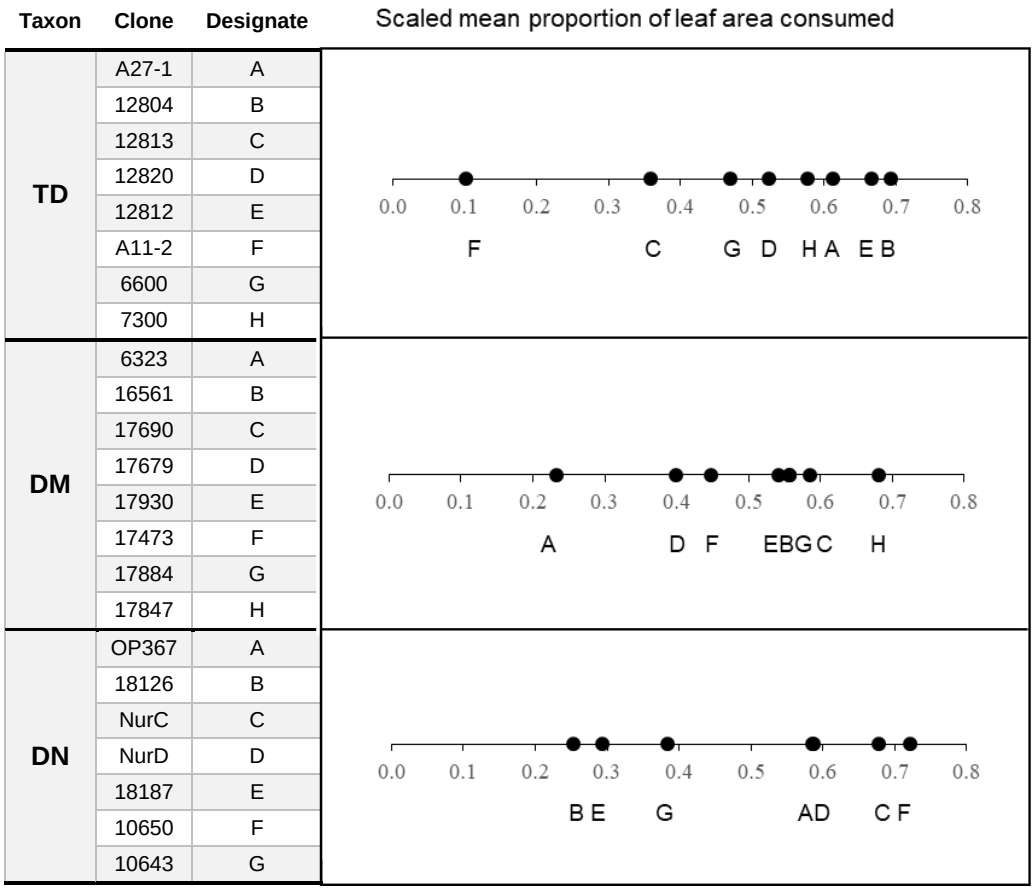


Figure 27. Scaled feeding preferences of genotypes on aligned number lines

Table 13. Cross-taxon feeding preference results

Genotype pair		Mean π_{ai}			Scaled-rank
i	*j*	*i*	*j*	*n* tests	agreement
DN-A	TD-A	0.544	0.456	3	+
DM-A	TD-C	0.443	0.557	4	+
DN-E	TD-C	0.418	0.582	4	+
DM-A	DN-E	0.329	0.671	4	+
TD-B	DM-H	0.344	0.656	4	-
TD-B	DN-C	0.274	0.726	4	-
DM-H	DN-C	0.586	0.414	4	+
TD-F	DM-A	0.015	0.985	4	+

One of the issues concerning the association of NIR absorbance values to herbivore preference scales (or relative consumption values) in a multivariate regression model is that it is not practicable to scan the same leaves that are used in the consumption test. One must rely on an assumption that a clone propagated from a cutting from the same tree, grown in the same conditions for the purposes of obtaining NIRS spectral data, has the same phytochemical profile as the test plant fed to the herbivore. To address any uncertainty about that assumption we grew extra clones of each genotype for the purposes of obtaining 3 unique NIRS scans from different plants of the same genotype. Each spectra was generated from a pooled sample of young leaves (leaf numbers 2 through 4). Each genotype's 3 NIR spectra were used in a PLS1 model for predicting feeding preference and associated with the 1 averaged value (π_{ai}) for the genotype as the reference datum (Figure 28).

The resulting model explained 72% of the variance in random segmented cross-validated predictions, requiring 9 principal components to minimize residual variance, and had a root mean square error of cross-validation of 0.084 and standard error of prediction of 0.056. This model was optimized by excluding 3 samples and eliminating uninformative spectral variables (Figure 29). Without optimization the model explained only 68% of the cross-validated predictions, and required 10 principal components to minimize residual variance after excluding 5 samples. A model requiring this many components to minimize residual variance is, by 'rule of thumb', considered over-fit, given the low number of samples and variables; meaning, the model is explaining too much unrelated 'noise'. The most unpromising outcome of this model was evident in the amount of predicted preference variance that resulted from variation in NIR absorbance among the 3 identically grown clones for each genotype, seen in in Figure 28 in the vertical distribution of predictions for the same measured preference value.

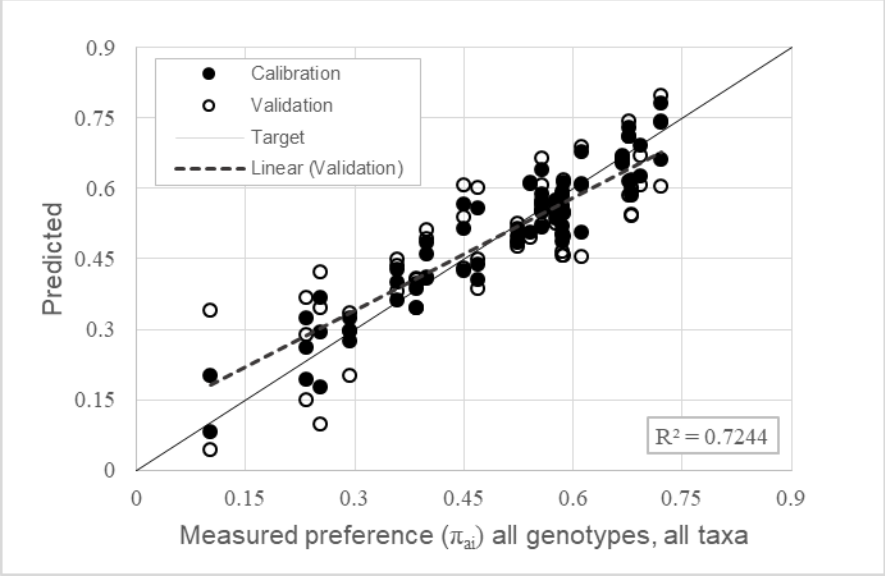


Figure 28. PLS1 regression model predicting feeding preference, based on three NIRS absorbance scans for each genotype.

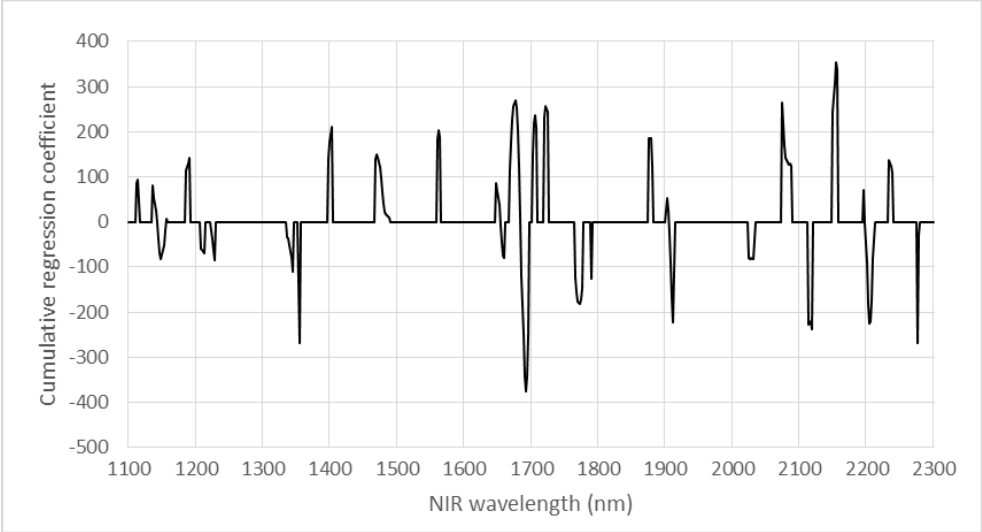


Figure 29. Cumulative regression coefficients corresponding to NIR wavelengths for the optimized PLS1 feeding preference model shown in figure 28. Silenced wavelengths are indicated with zero as the regression coefficient value.

A reasonable work-around for this within-genotype spectral variance issue would be to collect NIR spectral data from many more clones and average them. This would be a worthy endeavor if the scaled feeding preference data were characterized by less variability and there were more significant differences among genotypes.

Large variation in modeled predictions arising from what was expected to be minimal phytochemical variation among identically grown clones pointed to the larger issue of whether leaf samples used in model development would be similar enough to the leaves harvested from plantation trees to conduct a model verification study. To address this issue we conducted a principal component analysis of NIR spectra collected from the DM and DN genotype samples used in the development of the feeding preference model combined with the NIR spectra from DM and DN genotypes collected at the Boardman Research site used for the ranked beetle damage model. Separation of the two sample groups were apparent in the three dimensional PCA plot, and this was not attributable to taxon differences (Figure 30). Ideally, spectral characteristics of the model development samples and plantation-grown tree samples should be overlapping if the model were to generate meaningful predictions for the verification study, or trees grown in any condition. Leaf samples from both groups were approximately the same ontological age (leaf numbers 2 to 5). The Boardman Oregon leaf samples were harvested from lateral branches on two year-old trees growing in mineral soil, whereas the model development samples were harvested from apical shoots of eight month-old cutting saplings grown in peat-based potting soil in Western Washington (most were approximately 2 meters tall with emerging secondary branches). Quite surprisingly, six of the DM genotypes and two of the DN genotypes were present in both sample groups. Their spectral separation in PCA analysis suggests that tree age, leaf origin (apical versus lateral meristem), and growing environment resulted in dramatically altered phytochemistry.

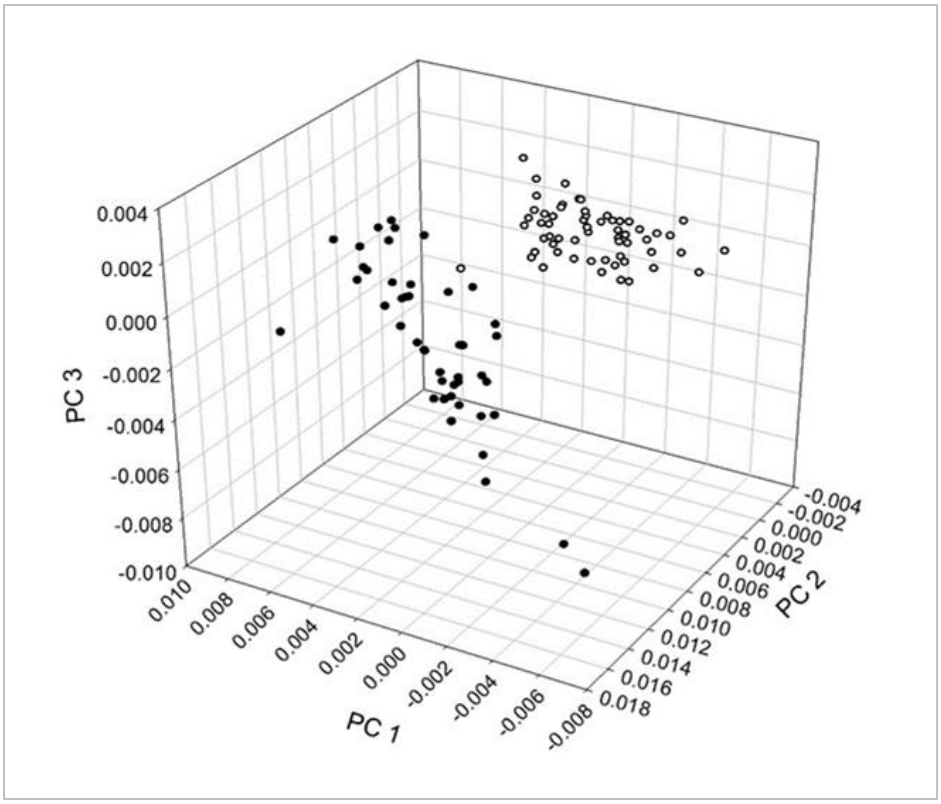


Figure 30. Results of principal component analysis of NIR spectra of leaf samples collected from DN and DM genotypes used in the development of the feeding preference model (filled circles), combined with spectra of leaf samples collected from DM and DN genotypes of the Boardman Oregon progeny trial (open circles), plotted in the space defined by the first three principal components (PC1, PC2, PC3).

One observation from this analysis is how powerful PCA analysis of NIR spectra is in discriminating aggregate chemical changes in a genotype grown in different conditions. *The most important implication of these findings is that all of the NIRS-based regression models described in this document have produced quantitative associations of absorbance patterns with measured reference data that were essentially fleeting snapshots in plant development time, relevant only to a particular set of test conditions that are unlikely to be reproduced in any commercial breeding settings or tree production operations. The ascription of an NIRS-modeled value as a fixed phenotypic trait, be it analyte concentration or scaled palatability to a particular herbivore, would be a false representation outside a precisely defined developmental and environmental context. While standardized sampling methods can normalize this context and highlight genotypic differences in analyte abundance, it may seem impractical to limit the herbivore experience of the plant to a defined set of developmental and environmental conditions, and as we have demonstrated, difficult to achieve experimentally.* Whether apical leaves are quantitatively preferred to lateral leaves of the same ontological age, are unexplored questions. Our field observations suggest that apical leaves are preferred by beetles and deer. But when the trees are older and taller, lateral branch shoots are exponentially greater in number than apical shoots, and, from a practical sampling standpoint, apical leaves become more difficult to access.

Objective 4: Technology transfer

Milestone 1. Provide training in sample preparation, NIR calibration development, and NIR screening procedure

Throughout this project, and in another federally funded project in which we collaborated with GreenWood Resources on developing the NIRS technology, WSU-Puyallup has worked closely with GreenWood on developing protocols for sample preparation and sample presentation using a spectrometer that they own. We have held training sessions on NIR spectral data acquisition, how to populate the multivariate data files for the software, how to generate calibrations, and how to generate calibration predictions. Many of the developed procedures are specific to their particular instrument, and laboratory resources. Throughout this project we have presented our research findings to GreenWood personnel, highlighting the challenges of sampling to address the pest resistance problem, and the challenges of developing NIR-calibrations.

We developed and provided GreenWood Resources an 89-page training manual with step-by-step procedures on sample collection, preparation, NIRS scan presentation, as well as detailed instructions on how to analyze the spectral data using the Unscrambler and Brimrose software. While many of the analytical methods outlined and discussed in the manual are available in other commercially available texts, the manual presents common methods with an emphasis on applications for tree improvement problems.

Milestone 2. Test the utility of the NIR tool in a practical screening scenario

This is a milestone that was not realized for the following reasons:

- *Limited scope* - Many of the empirical calibrations related to consumption that we created were developed with insufficient taxon representation and replication to be utilized broadly in screening operations. While our work demonstrated limited feasibility of the method, the limited scale of the data sets restricted the applicability of calibrations that were produced to specific taxon groups.
- *Limited reliability* - Concentrations estimated with the few single constituent calibrations we developed for foliar and bark phytochemicals were shown to explain only a minor portion of variation in herbivory. Additionally, there were some important palatability-related phytochemicals for which calibrations were not feasible.
- *Low priority* – GreenWood research and development personnel are involved in breeding and identifying fast growing, high quality wood in cultivars for production in different regions. As long as existing asset protections are available (i.e. commercial pesticides and mating disruption pheromones), efforts to select natural pest resistance traits will likely take a lower priority in the dedication of real estate, crop production inputs, and manpower.
- *Questionable cost effectiveness* – An early operational assumption was that NIR spectroscopy and the associated software would be easy to use – meaning that, with minimal training a technician could collect samples and generate prediction data. This perspective has changed, and the costs associated with employing skilled personnel is an issue. Furthermore, the application of NIRS predictions in an empirically determined progeny selection/ranking system, and efforts and costs associated with continuous calibration improvements, will require dedicated long-term effort from skilled personnel.

Project outputs

Research products

Oral and/or poster presentations at the Advanced Hardwood Biofuels meetings

- *"Development of NIR-based screening technology for detecting defensive chemical traits in hybrid poplar leaf and bark tissue",* AHB meeting, Seattle, WA, September 2011.
- *"Development of NIR-based clone screening technology for detecting defensive chemical traits*", Northwest Hardwood Association field day, Westport OR, August 2012.
- *"Development of NIR-based clone screening technology for detecting defensive chemical traits",* AHB meeting**,** Boardman, OR, September 2012.
- *"Development of NIR-based screening technology for detecting natural defensive chemical traits",* AHB meeting, Corvallis, OR, September 2013.
- *"Vole and deer preferences for varieties of poplar: a role for NIR spectroscopy in predicting consumption based on known and unknown compounds in foliage and bark",* AHB meeting, Davis, CA, September 2014.
- *"Prospects for Selecting Pest-Resistant Hybrid Poplar Using NIR Spectroscopy",* AHB meeting, Seattle, WA, September 2015.
- *"Using NIR Spectroscopy for Selecting Pest-Resistant Hybrid Poplar",* AHB meeting, Walla Walla, WA, September 2016.

Oral presentations with GreenWood Resources staff

- "*Vole and Deer Feeding Preferences for Clonal Variants of Poplar*" GreenWood Resources staff meeting, Westport, OR, January 2014.
- "*Hybrid Poplar Resistance to the Cottonwood Leaf Beetle: Development of Empirical Models of Leaf Consumption Predicted by NIR Spectroscopy"* GreenWood Resources staff meeting, Boardman, OR, December 2015
- "*Hybrid Poplar Resistance to the Cottonwood Leaf Beetle: Development of Empirical Models of Leaf Consumption Predicted by NIR Spectroscopy"* GreenWood Resources staff meeting, Westport, OR, February 2016.

Oral and poster presentations at symposia

- Oral presentation: "*Vole and deer preferences for varieties of poplar: a role for NIR spectroscopy in predicting consumption based on known and unknown compounds in foliage and bark"*, Short Rotation Woody Crops Conference, Seattle, July 2014.
- Poster presentation: *"Vole and deer preferences for varieties of poplar: a role for NIR spectroscopy in predicting consumption based on known and unknown compounds in foliage and bark"*, International Poplar Commission Symposium, Vancouver BC, July 2014.

Training manual: *Near Infrared Spectroscopy in Tree Improvement*: *A practical guide to sampling, NIR spectral data collection, calibration, and screening using the Brimrose Luminar 5030 Spectrometer", S*ubmitted to GreenWood Resources personnel electronically, June 2016.

Final Technical Report*: Feedstock Development - Task 4: Pest Management without Chemical Inputs*, submitted electronically to Project Administrator, August 2016. Revised August 2020 - this document.

Workforce trained and supported

Over the project period we have hired four undergraduate or recent college graduates as temporary technical research assistants. In many cases this opportunity was their first professional science-related job. Three students have capitalized on their experience in subsequent employment.

Our chemical analysis sub-contractor for the project was the Russel Labs at the University of Wisconsin-Madison. Fee-for-service payments partially supported several post-docs in the Department of Entomology.

Conclusions

- *Can NIRS technology be used to predict natural pest resistance in new hybrid poplar varietals?* Our research revealed that variability of phytochemicals and herbivory is exceedingly difficult to control and account for experimentally. This is the biggest challenge to modeling palatability. The process of selecting test material and measuring feeding preference frames any derived quantitative association to reflect artificial, time-dependent conditions that may have limited relevancy to the broader plantation pest problem. The main herbivores of hybrid poplars have evolved adaptations to exploit a range of phytochemical variation. Our experiments demonstrated that herbivores do have preferences for particular genotypes and avoided others, but this was not explained by concentration of the limited panel of 'deterrent' phytochemicals we measured with any useful precision.

  Because we did not find a strong quantitative relationship between palatability and concentration of any single phenolic glycoside or condensed tannins we explored two approaches to using NIRS: either as a method to estimate concentration of many analytes, using independent calibrations, which could then be substituted into an experimentally derived multiple linear regression model of herbivory as explanatory variables; or, as a method to estimate herbivory directly using quantitative association of NIR absorption patterns with measured palatability, assuming that many NIR absorbing compounds present in foliage may act additively to affect consumption. The first approach revealed that combinations of measured analytes did not predict consumption uniformly in all poplar taxa, or with sufficient precision. Further, multiple concentration estimates could not be successfully substituted into MLR models because of compounding errors. The second approach was limited by the combinatorics of bringing plants and herbivores together in all permutated groupings to derive a global feeding preference scale across all taxa. Our experiments also revealed that environmental differences in growing conditions and differences in developmental status of genetically identical sample material had large impacts on NIRS absorbance profiles. When uniform growing conditions and sampling were practiced we observed variation in NIR absorbance profiles among genetically identical, same-aged, identically grown samples, which had substantial effects on modeled herbivory estimates.

  Even with this apparent sensitivity to uncontrollable variability, the NIRS-based modelling method lacks the precision and accuracy to tease apart important chemical

structure details in molecules like condensed tannins. It cannot discriminate between small oligomers of flavonoids versus long, branched-chain polymers that may be more biologically active. It cannot discriminate functional group substitutions in salicinoid compounds like salireposide. The PLS algorithm is incapable of completely disentangling problem-related absorbance patterns from non-problem co-varying patterns, as we observed when attempting to generate multi-taxon calibrations for the prominent phenolic glycosides.

- *How accurate does a pest resistance model need to be to provide useful information for selection?* The answer to this question is linked to the difference between the estimated prediction errors derived from the model validation algorithm versus the deviation from those predictions measured in subsequent field trials for verification. The root mean square error of cross-validation computed for the model is not sufficient on its own to be a reliable estimate of precision in all situations. If the experimental conditions for deriving palatability models were the same or very close to conditions in the plantation, then verifying a model's accuracy in a subsequent field trial should theoretically match the validation-estimated prediction error. Some of our findings demonstrated that these conditions are quite different (e.g. the leaf beetle experiments). In other experiments we attempted to use observations of herbivory *in plantation* to construct empirical models of ranked browse level or ranked tree damage. While the conditions of validation and verification were theoretically the same or similar, the broad range of damage in each arbitrary ranking increment; the uncertain implications of undamaged trees; the impact of neighboring trees on herbivore choice, challenge any effort to replicate and verify a field-survey-derived model.

  If it were determined that predictions of genotype palatability using experimentally grown leaf samples collected from apical shoots of young trees were offset by a common amount from predictions of the same genotypes using samples from lateral branch leaves of plantation trees, then the relative differences among different genotypes might provide useful information to the breeder. We determined that there was no correlation ($r = -0.085$) or reliable offset in the modeled palatability predictions for the eight genotypes in the leaf beetle experiments that were represented in both model development samples and older field-collected samples for verification.

- *Are there ways to generate a controlled empirical NIRS-based model of herbivory in the plantation to minimize the disparity between model validation and verification?* This may be possible for insect herbivores, though logistically challenging. Given a breeding progeny field trial laid out in a complete randomized design with plenty of genotype replication and closely spaced trees, such that branches of neighboring trees overlap (not an uncommon situation), insect test subjects could be enclosed in tents with pairs of lateral shoots from neighboring trees. Leaf area consumed from each genotype in a defined period could form the basis of a measured genotype preference scale. Neighboring shoots of non-tented branches of the same genotype pair could be collected for NIR absorbance profiles.

- *Can NIRS technology be used to characterize phytochemicals in poplar breeding parents*? Yes; just as wood composition in poplar genotypes can be estimated with NIRS, concentrations of some phytochemicals in foliage and young bark can be estimated – but not all. Deriving models to estimate phytochemical concentrations that are commonly near the limits of chromatographic detection are not feasible.

We observed that NIRS calibration models derived from samples of a poplar species were more accurate and precise than models derived from samples of poplar hybrids or from multiple taxa. While one may be able to confidently identify breeding parents of a species having desirable levels of an analyte by using a separate analyte concentration model derived for each parent species, one would have to accept much less precise estimates of analyte levels in the resulting hybrid genotypes, using a model developed with hybrid samples. Current breeding strategy is to cross genetically diverse species in order to gain maximum growth heterosis. This is a recipe for generating persistent low-precision issues in NIRS calibration models.

From a business perspective, it takes 7 or 8 years starting from the hybridization of two parent species, through three multi-year stages of hybrid progeny trials, to the final selections for production suitability. During that time selected trees are clonally expanded or culled, requiring manpower, real estate, water, nutrient resources, and various asset protections in the form of fencing, pesticide, herbicide, mowing, and maintenance. A breeder possessing knowledge derived from NIRS-based models that the parents of the new hybrid genotypes had phytochemical attributes in their foliage that conferred low palatability to herbivores would likely *not* decrease the number of progeny trials required to evaluate growth rate, stem form, and environmental suitability. However a decision about comparative advantage among equally competing genotypes in the final stage could boil down to pest resistance traits. The pay-off for developing the NIRS technology; invested in the equipment, and spending resources to collect and process samples, can only come from future asset protection cost reductions, which is speculative at best. The same or better information can be obtained with the same technicians collecting pest damage information in all the staged trials along with planned growth biometrics. There is no need to use NIRS "high throughput screening" on parents of hybridizations to measure putative defensive phytochemicals because there aren't that many parents being crossed in any given year. The standard 'wet-lab' analyte determinations would, over all, be less expensive and much more accurate.

- *Is NIRS screening technology cost effective?* Responses to this question can addressed on several levels. One can compare operational costs of screening samples with NIRS relative to other methods of analyte quantification. One can compute the number of samples that would need to be screened using NIRS to recover the capital costs of the instrument, sample preparation machinery, facilities, and model development. And one might speculate whether the overall costs of using NIRS screening system to identify and deploy naturally resistant hybrids is competitive relative to the cost of conventional protection inputs for pest-susceptible genotypes, and traditional methods of detecting pest resistance in progeny.

  The popular appeal of NIRS technology is to cut phenotype screening costs. As an example, the cost (in 2012 dollars) of characterizing a panel of analytes in 100 samples might range from $5,000 to $12,500 depending on which analytes are being measured and the amount of laboratory preparation involved in extraction and sub-purification before chromatographic analysis. Not every analytical lab is set up for measuring the analytes of interest. More likely, a breeder would need to negotiate a contract with an academic lab that is currently supporting research in the relevant phytochemistry. The costs can vary widely based on the needs of the contractor in supporting student workers with stipends and benefits, costs for covering laboratory operations and instrument maintenance, and costs for administration and facilities paid to the university. The amount of time required to analyze

100 samples also varies widely, depending on the extraction and preparation procedures, chromatographic cycle time per sample, and the work schedule of student helpers. It would not be unreasonable to expect a turn-around time measured in multiples of weeks.

Our costs for screening 100 samples with NIRS is estimated to be $925 with a turn-around time of 50 hours, assuming 27 hours for sample milling, 20 hours for NIR scanning and spectral data upload, and 3 hours for data clean-up and analysis; paying a cumulative compensation of $18.50/hr for each of these tasks.

Accounting for capital costs directly related to NIRS screening: Brimrose Luminar spectrometer, Virtis benchtop lyophilizer and deep cycle vacuum pump, Unscrambler multivariate data analysis software, Wiley Mini-mill, plus various incidentals and replacement parts totaled approximately $50,000. This assumes the user possesses a computer with appropriate capacity, and has facilities for storing frozen samples.

Another capital cost is in development of the NIRS calibration model used in screening. Again, assuming 100 samples needed for generating a NIRS calibration, and taking our contract with the University of Wisconsin Russel Labs as an example of per unit cost, we would pay $8370 in direct and indirect costs for HPLC-mass spec analyte quantification. Add to this $2600 for sample collection and travel costs, sample prep and milling, NIR scanning and calibration algorithm development. So approximately $11,000 (underestimate) for development of a NIRS calibration model, not taking into account fringe benefits for the analyst(s) or any facilities and administration charges, and assuming the calibration is developed by the breeder or associates with no intellectual property fees and royalties paid to a third party.

If the breeder were to use that NIRS calibration to estimate the analyte concentration in an additional 630 samples, they would break-even on capital outlays relative to the costs for data chromatographically measured by an outside lab. However, this assumes that the NIRS method generates data of comparable quality. Whatever apparent advantage one gains in turn-around speed and long term cost reduction using the NIRS system can be diminished by inaccuracy and imprecision of the output. There is also a risk of failure; not being able to generate a calibration from 100 samples, or generating a prediction model that is unreliable. Our research demonstrated that this risk is not only quite high, it is guaranteed.

Estimating lost revenue to a breeding operation resulting from an inaccurate prediction entails many components; more than can be addressed here. But these risks need to be weighed relative to the other breeding priorities. Currently, pest-resistance is a third or fourth tier selection criteria behind volume growth increment, wood chemical composition, stem form, and drought resilience. The importance of natural pest-resistance may change if the costs of pesticides increase, or regulations prohibit or restrict their use.

Defining the value of a pest-resistant varietal – meaning, ‘cost savings’ relative to conventional asset protection costs, boils down to defining what ‘economic pest damage’ is. This depends on the scale of the damage relative to the scale of the production operation, and resulting yield loss from that damage. This is dependent on the price of the wood product, the cost of production, profit margins, and the cost of conventional pest management interventions. For hybrid poplar, the valuations of these variables are currently estimated from pilot projects and economic projections based on pulp markets or the engineered wood products markets. As there are no cellulosic biomass-to-biofuel conversion plants in operation in the US, there is no competition for chipped wood feedstock, and thus there is no

verifiable market-tested price discovery, and no way to establish a measurable "economic pest damage", or cost savings from deploying pest resistant varietals.

- *What do we need to know going forward to further this technology?*
  a) Until there are specific NIR absorption bands defined for each analyte of interest to aid in the elimination of non-informative spectral variables in a calibration process, the problem of disentangling concentration-covariant data structures from non-analyte taxon-idiosyncratic co-varying absorbance patterns will remain an intractable problem resulting in low calibration precision. A study to determine whether NIR absorption bands from purified analyte in solution can be correlated with important (high loading weight) absorption bands of a PLS1 concentration calibration for that analyte could provide a potentially useful information.
  b) Our studies demonstrated that mammalian herbivores strongly avoided some poplar genotypes, but the 'defensive' analyte concentrations did not provide strong explanatory evidence. A study in which various fractions of poplar foliage extract are infused into milled pelletized chow and presented to deer or voles might provide clues to discovering different classes of phytochemicals that are responsible for feeding avoidance or attraction. Not enough is known about attractive compounds.
  c) While there are important studies that have quantified the changes in abundance of a number of phytochemicals in young poplar foliage as leaves age, there needs to be a better understanding of how herbivores respond to the relative abundance of these foliar phytochemicals as trees age year over year and in different environments. This will inform future efforts to model herbivory.
  d) There is plenty of room to incorporate machine learning algorithms in the software to aid spectral variable selection and model optimization. These may now exist in other analytical software.

Summary

The goal of identifying and deploying naturally pest-resistant or low-palatability poplar varietals to reduce negative impacts to wildlife and the environment from pesticide application was a part of a larger effort to generate purpose-bred, regionally adapted hybrid poplar varieties for biofuel and bio-product feedstock. NIR spectroscopy in combination with multivariate data analysis is a technology that was considered to provide a method for identifying genotypes that are less palatable to herbivorous pests. However, the issues we faced related to experimental scale, time, environment-dependent phytochemical variation, and animal behavior variability, limited our ability to acquire quality reference data for creating deployable models.

Generating NIRS-based calibrations to estimate the phytochemical constituents in hybrid poplar, which arise from interbreeding genetically diverse poplar species to create desirable and extreme variants, is a challenge that pushes the limits of this technology. Analyte regression models are typically less reliable at the extremes of the reference data variation because extreme values are rare. Calibrations must therefore be continually upgraded and improved, which adds an additional expense challenge.

Appendix

Measurement of analytes

Phenolic glycosides were quantified in methanolic extracts of plant tissue by ultra high-performance liquid chromatography with photodiode array and single quadrupole mass spectrometer detection (Waters Acquity I-Class UPLC® System, Milford, MA) following a modified version of the method presented by Abreu et al. (58), and further details described by Rubert-Nason et al. (15). Briefly; 20-30 mg of dried milled leaf or bark sample was extracted into methanol at 4 °C, spiked with an internal standard (d_6-salicylic acid), filtered through a 0.45 µm fluoropolymer membrane, and introduced to the UHPLC (2µl). Analytes with available standards (salicin, salicortin, HCH salicortin, tremuloidin, tremulacin, salireposide, and 2-cinnamoyl-salicortin) were quantified by negative electrospray ionization (ESI^-) single quadrupole mass spectrometry in selective ion recording mode (Waters Acquity 3100 mass detector) using mass/charge ratios corresponding to formate adducts. Data quality was verified by monitoring peak areas of a process control standard (β-resorcylic acid) spiked into the methanol extractant, and by monitoring the phenolic glycoside content of a laboratory standard prepared from methanol-extracted *Populus tremuloides* foliage. The method reporting limit for all analytes was 0.1 % of dry weight.

Additional methanol-extractable phytochemicals were monitored in ESI^- total ion chromatograms (250-600 *m/z*) and in UV-VIS chromatograms (210 – 400 nm). Peaks appearing in the ESI^- total ion chromatograms were designated as "dominant" if they appeared in ≥ 3 chromatograms with peak areas ≥ 5 % of the total combined peak areas in those chromatograms. These "unknown" compounds were tabulated by retention time (RT, minutes), and concentrations were estimated as salicortin equivalents based on a calibration curve prepared using salicortin peak areas in the total ion chromatograms ratioed to the peak areas of the internal standard.

Relative proanthocyandin content (condensed tannins), was determined using the acid-butanol method described by Porter et al. (59), and modified by Hagerman (60). Subsamples of 20 or 30 mg of lyophilized and milled foliage or bark was sequentially extracted 4 times with 500µl of 10 mM ascorbic acid in 70% acetone using a sonicating ice water bath. Sequential extracts were pooled (2 ml total). A 50µl aliquot of extract was added to 3 ml of butanol-HCl solution plus 550 µl of a 2% solution of ferric ammonium sulfate in 2N HCl, and heated to 96 °C for 50 min. Anthocyanidin absorbance was measured at 550 nm at 22 °C with a UV/Vis spectrophotometer (Shimadzu, model 1201). Listed values were the average of 3 analytical replicates for each of 2 subsamples from each genotype. Because of the limitations of this assay in discriminating heterogeneous structural and chemical characteristics of native condensed tannins, and because of the inherent difficulty in isolating relevant standards among a population of interbred poplar taxa to convert spectral absorbance to mass-based concentration, anthocyanidin absorbance at 550 nm was normalized to the dry weight of the sample for comparing the relative abundance. For reference, the extent of normalized absorbance values ranged from 0.075 to 0.816. A previous determination of absorbance converted to mass-based proanthocyanidin concentration using a *P. trichocarpa* x *P. deltoides* standard extracted from foliage revealed a linear relationship up to $A_{550} = 1.2$.

Solubilized protein content was measured by extracting 5 ± 0.2 mg lyophilized milled foliage or bark three times in 500µl of a buffer comprised of 125 mM Tris-HCl pH 8.8, 1 % (w/v) SDS, 10 % (v/v) glycerol, 50 mM sodium meta-bisulfite, and 1 mM EDTA. From the pooled extracts 2-50µl subsamples were subjected to protein precipitation using the Compat-

Able® Protein Assay Preparation kit (Prod# 23215, Thermo-Scientific, Rockford, IL). Pellets were suspended in 0.5ml of distilled water, and protein was quantified using the BCA Protein Assay Kit (Pierce, Rockford IL), measuring absorbance at 562 nm, and calibrated to a bovine serum albumin standard curve. Four analytical samples were averaged for each genotype and expressed a percent protein per dry sample weight.

Total phenolic content was determined using a modified Folin-Ciocalteau method as described by Salminen and Karonen (29). Lyophilized and milled leaf or bark samples (10 mg) were extracted three times with 500µl of 70% acetone in a sonicating water bath at 4°C. Pooled sample extracts were dried under vacuum and resolublized in 500µl of distilled water. In 96 well round-bottom mixing plates, 20µl of plant extract were combined with 280µl of a buffer comprised of 9 parts of 50mM sodium carbonate pH 10, and 5 parts of 0.6% formic acid. Plates were mixed on a table shaker for 10min. Colorimetric detection of total phenolics was made by combining 50µl of the buffered extract from the mixing plate with 50µl 1N Folin-Ciocalteau reagent and 100µl of 20% sodium carbonate in a 96-well microtiter plate and incubated at room temperature for 1 hour on a table shaker. Concentration of total phenolics was determined by measuring absorbance at 750nm (BioTek ELx800 microplate reader) calibrated with a gallic acid standard curve and expressed as micrograms of gallic acid equivalents per mg of dry sample.

Leaf and bark sample collection, drying, and milling

For the 2011-2012 survey of seasonal changes in CT and PG analyte abundance in two-year-old *Populus x generosa* genotypes, leaves were harvested from the apical meristems of the main leader. The fifth and sixth oldest leaves with length greater than 3 cm were combined in one sample, and leaf numbers 10 and 11 were combined in another. Two samples for each leaf age-group in each genotype were collected in the spring. For the summer collection four leaf age-group samples were harvested from leaf numbers 5 and 6 combined, 10 and 11combined, 15 and 16 combined, and 20 and 21 combined, for two replicate trees of each genotype. All leaf age-group samples were placed in aluminum foil pouches and flash-frozen between blocks of dry ice in the field, then stored at -40° or -80°C until lyophilized.

For the multi-taxon survey of analyte abundance in genotypes in various field trials the majority of leaf samples (leaf number 5 and 6) were collected from one-year-old shoots arising from coppiced stools. Bark and phloem tissue collected as 5 cm shavings from around the shoot at 10cm and 80cm from the base were combined into a single sample for each genotype. For the Boardman wild deer browse study, and the cottonwood leaf beetle study verification set, leaf numbers 3 through 6 from lateral branches were combined. All samples were placed in aluminum foil pouches and flash-frozen between blocks of dry ice in the field, then stored at -40° or -80° C until lyophilized.

Samples in their opened foil pouches were placed in a VirTis benchtop lyophilizer connected to a deep-cycle vacuum pump, with condenser temperature operating at -80 °C and pressure at 4 millitorr for 48 hours. Lyophilized leaves were ground in a Wiley mini-mill (Thomas Scientific, Swedesboro NJ) to pass through a #40 mesh screen, then stored refrigerated in brown glass vials at 5 °C, or in coin envelopes kept in sealed containers with desiccant and stored refrigerated until needed for NIR scans or analyte determinations.

NIR spectral data collection and PLS1model development

Milled samples were compressed into an 8mm diameter metal cuvette to a uniform depth of 6 mm with 1.5 kg force using the apparatus shown in Figure 31 to compress the material to a uniform density.

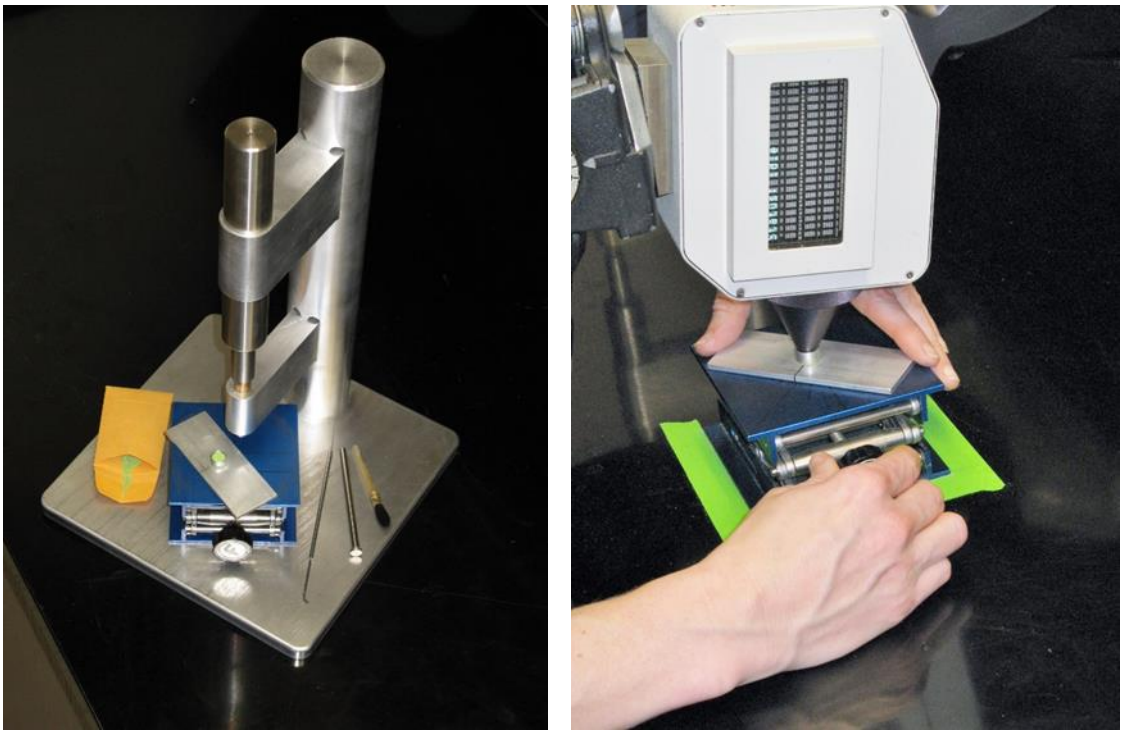


Figure 31. The cuvette packing device (left panel), in which milled leaf or bark material is packed with uniform force into the metal cuvette by raising the scissor jack until the 1.5 kg cylindrical weight is lifted from its resting position. The sliding weight support was locally fabricated; not commercially available. The cuvette, glued to a piece of aluminum flat bar, was raised to the cone-shaped aperture of the NIR spectrometer with the scissor jack (right panel) and scanned from 8 positions. For each captured scan the cuvette is rotated 11.25 degrees. The cuvette and the aperture cone piece were aligned and in perfect contact to exclude ambient light from interacting with the milled sample material.

Near infrared absorbance spectra were measured using a Brimrose Luminar 5030 AOTF spectrometer (Brimrose Corp., Sparks, MD) in wavelengths ranging from 1100 to 2300 nm (instrument limited). The spectrometer was programmed to compile and average 50 spectra in 2nm increments for each scan-set to minimize scatter-dependent variability. Using the Brimrose *Prospect* software to assess spectral profile and data quality of each scan-set, 8 additional scan-sets of the sample, acquired from different rotational positions, were averaged to produce the combined spectral data set for each sample, which was imported as 601 spectral variables into the Unscrambler multivariate analysis data table (version 9.8, CAMO, Woodbridge NJ). Aberrant scan profiles were eliminated in the *Prospect* data quality step. No fewer than 8 scan-sets were averaged in the combined spectral data set, meaning, no fewer than 400 absorbance data points acquired at each wavelength were averaged.

Using algorithms within the Unscrambler software package, spectral data was commonly transformed with Savitsky-Golay first derivative with a seven data-point kernel, and center point

estimated using a second order polynomial function. Wider kernels and second derivative transformations were also attempted but used less commonly.

Bilinear modeling of dependent variables (laboratory measured analyte concentration or herbivory) using NIR spectral data was accomplished with the Unscrambler PLS1 algorithms and internally validated with 'leave-one-out' full cross-validation, or random segmented cross-validation algorithms, with segment sizes approximately 3% to 5% of the sample set. PLS1 calibration models were optimized by: (a) eliminating sample outliers identified with principal component analysis, and (b) spectral variable selection techniques, in which wavelengths associated with consistently small loading weights were silenced, and wavelengths associated with consistently large loading weights in the 1st through 9th principal component factors were selected. The Martens' uncertainty test (Martens and Martens 2000) was also used to study the effect of influential samples and wavelengths on the model. Influential wavelengths of each attempted calibration were individually tested by temporary removal and recalculation, and added back into the model only if the cross-validation R-square increased, root mean square error of cross-validation (RMSECV) decreased, and fewer factors were necessary to minimize residual variance.

Potted tree leaf area estimation method

To minimize plant handling and expedite the non-destructive determination of total leaf area for each plant prior to each confined deer feeding trial, we devised the following system of estimating leaf area by interpolation and extrapolation. For primary leaves, we estimated the elliptical area of every 4th leaf on each shoot beginning with the 6th leaf; measuring leaf length from the base of the lamina, and maximum width, and using the formula:

$$EA_n = \pi(0.5\ length * 0.5\ width)$$

The combined elliptical area of the 3 intermediate leaves plus the next 4th leaf (EA_i) were interpolated using the expression

$$EA_i = 3 * EA_n \pm 1.5 * \|EA_n - EA_{n+4}\|$$

where EA_n is the elliptical area of the previous 4th leaf. The elliptical area in the second term of the equation is positive if the area of the 4th older leaf is larger than the previous 4th leaf and negative if smaller.

Since the interpolation method is sequential, the areas of the oldest leaves on the shoot were interpolated as follows:

$$EA = 4 * EA_n \pm 1.5 * \|EA_n - EA_{n+4}\| + EA_{n+4} \quad \text{for the last 4}$$

$$EA = 3 * EA_n \pm \|EA_n - EA_{n+3}\| + EA_{n+3} \quad \text{for the last 3}$$

$$EA = 2 * EA_n \pm 0.5 * \|EA_n - EA_{n+2}\| + EA_{n+2} \quad \text{for the last 2}$$

$$EA = EA_n + EA_{n+1} \quad \text{for the last 1}$$

The cumulative area of the youngest 5 leaves was estimated by an empirically derived coefficient specific to each genotype based on the elliptical area measured for the 5th leaf. The cumulative interpolated and extrapolated elliptical areas were then modified using an empirically derived taxon-specific shape correction factor to account for deviation from the elliptical area.

The secondary leaf area was estimated by measuring the elliptical area of the median sized leaf of each branch, multiplied by the number of leaves on that branch, then modified by the same shape coefficient used for the primary leaves.

Total leaf area estimations were further modified to account for daily growth and leaf expansion between the time the leaf area estimation measurements were made prior to the trial, and the time immediately after the trial in which the remaining leaves were stripped from the shoots and measured with the leaf area meter. This period averaged 4.9 days; was as short as 2 days, but as long as 9 days. The daily growth correction factor, specific to each clone, was derived by taking the difference between the average total leaf area of the first 3 plants in the group-1 tests, and the average total leaf area of last 3 plants in the group-3 tests, and dividing by the time period between (17 to 20 days). The leaf area estimation methods were validated by comparing the estimated area with the measured leaf area of plants that were not eaten by the deer.

References

1. *A review of the applications of portable near infrared spectrometers in the agro-food industry.* **Teixeira Dos Santos C.A., Lopo M., Pascoa R.M.N.J., Lopes J.A.** 2013, Applied Spectroscopy OA, Vol. 67, pp. 1215-1233.

2. *A review of recent near infrared research for wood and paper.* **Tsuchikawa, S.** 2007, Applied Spectroscopy Reviews, Vol. 42, pp. 43-71.

3. *Near infrared spectroscopy in wildlife and biodiversity.* **Vance C.K., Tollefson D.R., Kinoshita K., Rodriguez J., Foley W.J.** 2016, Journal of Near Infrared Spectroscopy, Vol. 24, pp. 1-25.

4. *Clonal variation in foliar chemistry of aspen: effects on gypsy moths and forest tent caterpillars.* **Hwang, S-Y., Lindroth, R.** 1997, Oecologia, Vol. 111, pp. 99-108.

5. *Phenolic metabolites in the resistance of northern forest trees to pathogens - past experiences and future prospects.* **Witzell, J., Martin, J.A.** 2008, Canadian Journal of Forest Resources, Vol. 38, pp. 2711-2727.

6. *PLS-regression: a basic tool of chemometrics.* **Wold, S., Sjostrom, M., Eriksson, L.** 2001, Chemometrics and Intellegent Laboratory Systems, pp. 109-130.

7. **John, Julia.** *Mongabay, News & Inspiration from Natures Frontline.* [Online] 2016. https://news.mongabay.com/2016/04/portable-scanners-identify-timbers-detect-illegal-logging/.

8. *The Determination of red grape quality parameters using the LOCAL algorithm.* **Dambergs, R.G., Cozzolino, D., Cynkar, W.U., Janik, L., Gishen, M.** 2006, Journal of Near Infrared Spectroscopy, Vol. 14, pp. 71-79.

9. *Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality.* **Cozzolino, D., Cynkar, W.U., Shah, N., Smith, P.** 2011, Food Research International, Vol. 44, pp. 1888-1896.

10. *Use of near infrared spectroscopy to measure chemical and mechanical properties of solid wood.* **Kelley, S.S., Rials, T.G., Snell, R., Groom, L.H., Sluiter, A.** 2004, Wood Science and Technology, Vol. 38, pp. 257-276.

11. *The relationship between near infrared spectra of radial wood surfaces and wood mechanical properties.* **Gindl, W., Teischinger, Schwanninger, M., Hintertoisser, B.** 2001, Journal of Near Infrared Spectroscopy, Vol. 9, pp. 255-261.

12. *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.* **Foley, W.J., McIllwee, A., Lawler, I., Aragones, L. Woolnough, A. P., Berding, N.** 1998, Oecologia, Vol. 116, pp. 293-305.

13. *Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses.* **Boeckler, A.G., Gershenzon, J., Unsicker, S.B.** Phytochemistry, Vol. 72, pp. 1497-1509.

14. *Tannins in plant-herbivore interactions.* **Barbehenn, R.V., Constabel, C.P.** 2011, Phytochemistry, Vol. 72, pp. 1551-1565.

15. *Rapid phytochemical analysis of birch (Betula) and poplar (Populus) foliage by near-infrared spectroscopy.* **Rubert-Nason, K.F., Holeski, L.M., Couture, J.J.,Gusse, A., Undersander, D.J., Lindroth, R.L.** 2013, Anal Bioanal Chem, Vol. 405, pp. 1333-1344.

16. *Diversity of structure and antiherbivore activity in condensed tannins.* **Ayers, M.P., Clausen, T.P., MacLean S.F., Redman, A.M., Reichardt, P.B.** 16, 1997, Ecology, Vol. 78, pp. 1696-1712.

17. *Qualitative variation in proanthocyanidin composition of Populus species and hybrids: genetics is the key.* **Scioneaux, A.N., Schmidt, M.A., Moore, M.A., Lindroth, R.L., Wooley, S.C., Hagerman, A.E.** 2011, Journal of Chemical Ecology, Vol. 37, pp. 57-70.

18. *Impact of condensed tannin size as individual and mixed polymers on bovine serum precipitation.* **Halbertson, J.F., Kilmister R.L., Kelm, M.A., Downey, M.O.** 2014, Food Chemistry, Vol. 160, pp. 16-21.

19. *Seasonal patterns in the phytochemistry of three Populus species.* **Lindroth, R.L., Hsia, M.T.S., Scriber, J.M.** 6, 1987, Biochemical Systematics and Ecology, Vol. 15, pp. 681-686.

20. *Within and between-year variation in early season phytochemistry of quaking aspen (Populus tremuloides Michx) clones.* **Osier, T.L., Hwang, S-Y., Lindroth R.L.** 2000, Biochemical Systematics and Ecology, Vol. 28, pp. 197-208.

21. *Age-related shifts in leaf chemistry of clonal aspen (Populus tremuloides).* **Donaldson, J.R., Stevens, M.T., Barnhill, H.R., Lindroth, R.L.** 2006, Journal of Chemical Ecology, Vol. 32, pp. 1415-1429.

22. *Phenological and chemical differences among hybrid poplar clones (Salicaceae) varying in resistence to Cryptorhynchus lapathi (L.) (Coleoptera: Curculionidae).* **Broberg C.L., Inkster, J.A.H., Borden, J.H.** 2010, Biochemical Systematics and Ecology, Vol. 38, pp. 29-48.

23. *The plastochron index as applied to studies of cottonwood.* **Larson, P.R., Isebrands, J.G.** 1, 1971, Canadian Journal of Forest Research, Vol. 1, pp. 1-11.

24. *Dynamic changes of phenolic compound contents in leaf and bark of poplar during autumn temperature drop.* **Li, C-M., Wang, Y., Yu, W-X.** 22, 2011, Journal of Forestry Research, Vol. 3, pp. 481-485.

25. *Carbohydrate translocation determines the phenolic content of Populus foliage: a test of the sink-source model of plant defense.* **Arnold, T., Appel, H., Patel, V., Stocum, E., Kavalier, A., Schultz, J.** 2004, New Phytologist, Vol. 164, pp. 157-164.

26. **Esbensen, K.H.** *Multivariate Data Analysis - In Practice; An Introduction to Multivariate Data Analysis and Experimental Design.* 5th Edition. Oslo : CAMO Process AS, 2004. ISBN 82-993330-3-2.

27. *PLS-regression: a basic tool of chemometrics.* **Wold, S., Sjostrom, M., Ericksson, L.** 2001, Chemometrics and Intellegent Laboratory Systems, Vol. 58, pp. 109-130.

28. *Prediction of chemical characteristics of fibrous plant biomasses from their near infrared spectrum: comparing local versus partial least square models and cross-calidation versus independent validations.* **Godin, B., Agneessens, R., Delcarte, J., Dardenne, P.** 2015, Journal of Near Infrared Spectroscopy, Vol. 23, pp. 1-14.

29. *Chemical ecology of tannins and other phenolics: we need a change in approach.* **Salminen, J.P., Karonen, M.** 2011, Functional Ecology, Vol. 25, pp. 325-338.

30. *Wintertime use of hybrid poplar paltations by deer adn medium-sized mammals in the midwestern U.S.* **Christian, D.P.** 1, 1997, Biomass and Bioenergy, Vol. 12, pp. 35-40.

31. **Netzer, D.A.** *Hybrid poplar plantations outgrow deer browsing effects.* North Central Forest Experiment Station, USDA-Forest Service. s.l. : U.S. Government Printing Office, 1984. pp. NC-325, Research Note. 1985-565-329-10098.

32. *Behavioral strategies of mammalian herbivores against plant secondary metabiolites: The avoidance-tolerance continuum.* **Iason, G.R., Villalba, J.J.** 2006, Journal of Chemical Ecology, Vol. 32, pp. 1115-1132.

33. *A meta-analysis of bird and mammal response to short-rotation woody crops.* **Riffell, S., Verschuyl, J., Miller, D, Wigley, T.B.** 2011, Global Change Biology Bioenergy, Vol. 3, pp. 313-321.

34. *Is energy cropping in Europe compatible with biodiversity? Opportunities and threats to biodiversity from land-based production for bioenergy purposes.* **Pedroli, B., Elbersen, B., Frederiksen, P. Grandin, U., Heikkila, R., Krough, P.H.** 2013, Biomass and Bioenergy, Vol. 55, pp. 73-86.

35. *Small animal use of hybrid poplar plantations relative to stand age.* **Moser, B.W., Pipas, P.J., Whitmer, G.W., Engeman, R.M.** 2002, Northwest Science, Vol. 76, pp. 158-165.

36. *Density-dependent vole damage in silviculture and associated economic losses at a nationwide scale.* **Huitu, O., Kiljunen, N., Korpima, E., Koselka, E., Mappes, T., Petiainen, H., Poysa, H., Henttonen, H.** 2009, Forest Ecology and Management, Vol. 258, pp. 1219-1224.

37. *Small animals can damage young hybrid poplar plantations.* **Bowersox, T.W.** 1973, Tree Planter's Notes, Vol. 24, pp. 10-11.

38. *Productivity, drough tolerance and pest status of hybrid Populus: tree improvement and silvicultural implications.* **Robison, D.J., Raffa, K.F.** 1998, Biomass and Bioenergy, Vol. 14, pp. 1-20.

39. *Vole Preference for Salix caprea, S. repens and their F1, F2 and backcross hybrids.* **Hallgren, P., Hjalten, J.** 2004, Oikos, Vol. 106, pp. 61-66.

40. *Bark salicylates and condensed tannins reduce vole browsing amongst cultivated dark-leaved willows (Salix myrisinifolia).* **Heiska, S., Tikkanen, O.P., Rousi, O.P., Julkunen-Tiitto, R.** Chemoecology, Vol. 17, pp. 245-253.

41. *Lezpedza phenolics and pensetmon alkaloids: effects on digestion efficiencies and growth of voles.* **Lindroth, R.L., Batzli, G.O., Alvidsen, S.I.** 1986, Journal of Chemical Ecology, Vol. 12, pp. 713-728. Errata in 12:1977-1978.

42. **Verch, R.L.** *Feeding preferences of penned white tailed deer for hybrid poplar clones.* USDA-Forest Service, North Central Experiment Station. 1979. Research Note, NC-248. 1980-667-861/92.

43. *Aspen decline, aspen chemistry, and elk herbivory: Are they linked?* **Wooley, S.C., Walker, S., Vernon, J., Lindroth, R.L.** 1, 2008, Rangelands, Vol. 30, pp. 17-21.

44. *Systematically wound-responsive genes in poplar trees encode proteins similar to sweet potato and legume Kunitz trypsin inhibitors.* **Bradshaw, H.D., Hollick, J.B., Parsons, T.J., Clarke, H.R.G., Gordon, M.P.** 1989, Plant Molecular Biology, Vol. 14, pp. 51-59.

45. *Predicting bite size selection of mammalian herbivores: a test of a general model of diet optimization.* **Shipley, L.A., Illius A.W., Danell, K., Hobbs, N.T., Spalinger, D.E.** 1999, Oikos, Vol. 84, pp. 55-68.

46. *Diet selection by a ruminant generalst browser in relation to plant chemistry.* **McArthur, C., Robbins, C.T., Hagerman, A.E., Hanley, T.A.** 1993, Canadian Journal of Zoology, Vol. 71, pp. 2236-2243.

47. *Biology of insect pests in North American intensively managed hardwood forest systems.* **Coyle, D.R., Nebeker, T.E., Hart, E.R., Mattson W.** 2005, Annual Review of Entomology, Vol. 50, pp. 1-29.

48. *Evaluation of adult cottonwood leaf beetle Chrysomela scripta (Coleoptera:Chrysomelidae) feeding preference for hybrid poplars.* **Harrell, M.O., Benjamin, D.M., Berbee, J.G., Burkot, T.R.** 4, 1981, Great Lakes Entomologist, Vol. 14, pp. 181-184.

49. *Poplar clonal preferences of the Cottonwood Leaf Beetle.* **Caldbeck, E.S., McNabb, H.S., Hart, E.R.** 1978, Entomological Society of America, Vol. 71, pp. 518-520.

50. *Clonal and leaf age variation in Populus phenolic glycosides: implications for host selection by Chrysomela scripta (Coleoptera: Chrysomelidae).* **Bingamin, B.R., Hart, E.R.** Population Ecology, Vol. 22, pp. 397-403.

51. *Insect feeding stimulants from the leaf surface of Populus.* **Lin, S., Binder, B.F., Hart, E.R.** 11, 1998, Journal of Chemical Ecology, Vol. 24, pp. 1781-1790.

52. *Constitutive and herbivore-induced monoterpenes emitted by Populus x euroamericana leaves are key volatiles that prient Chrysomela populi beetles.* **Brilli, F., Ciccioli, P., Frattoni, M., Prestininzi, M., Spanedda, A.F., Loreto, F.** 2009, Plant, Cell and Environment, Vol. 32, pp. 542-552.

53. *Sources of insect and plant volatiles attractive to Cottonwood Leaf Beetles feeding on hybrid poplar.* **Kendrick, A.P., Raffa, K.F.** 2006, Vol. 32, pp. 2585-2594.

54. *Feeding and oviposiion preferences of adult Cottonwood Leaf Beetles (Coleoptera: Chrysomelidae) amonf Populus clones and leaf age classes.* **Bingaman, B.R., Hart, E.R.** 1992, Environmental Entomology, Vol. 21, pp. 508-517.

55. *Sequestration of plant-derived glycosides by leaf beetles: A model system for evolutino and adaptation.* **Boland, W.** 2015, Perspectives in Science, Vol. 6, pp. 34-48.

56. *Salicin from host plant as precursor of salicyl-aldehyde in defensive secretion of chrysomine larvae.* **Pasteels, J.M., Rowell-Rahier, M., Braekman, J.C., Dupont, A.** 1983, Physiological Entomology, Vol. 8, pp. 307-314.

57. *Metabolic profiling of sink-to-source transition in developing leaves of Quaking Aspen.* **Jeong, M.L., Jiang, H, Chen, H-S., Tsi, C-J., Harding, S.A.** 2004, Plant Physiology, Vol. 136, pp. 3364-3375.

58. *UHPLC-ESI/TOFMS determination of salicylate-like phenolic glycosides in Populus tremula leaves.* **Abreu, I.N., Ahnlund, M., Moritz, T., Albrectsen, B.R.** 2010, Journal of Chemical Ecology, Vol. 37, pp. 857–870.

59. *The conversion of proanthocyanidins to cyanidin and delphinidin.* **Porter, L.J., Hrstich, L.N., Chan, B.G.** 1986, Phytochemistry , Vol. 25, pp. 223-230.

60. **Hagerman, A.E.** *The Tannin Handbook.* [Online] 2011. http://www.users.muohio.edu/hagermae/

Acknowledgements

USDA/NIFA for generously supporting this research. Project guidance was provided by Dr. John J. Brown, Dr. John Stark, Dr. Lisa Shipley, and Dr. Robert Wielgus, Washington State University. Material and technical support was provided by members of GreenWood Resources: Brian Stanton, Rich Shuren, Rick Stonex, Andrew Rodstrom, Austin Himes, Grant Beauchamp, Jason Mack, Kathy Haiby, and Jesus Espinoza. The sample chemical analysis contract was administered by Dr. Rick Lindroth at Russel Labs, University of Wisconsin Madison, and conducted by Dr. Kennedy Rubert-Nason. We gratefully acknowledge technical and analytical support from Dr. Carri LeRoy at The Evergreen State College, and from technicians Evan Shields, Jennifer Vittetoe, Daniel Nyquist, and Michael Galvan.