Enhancing NIR Calibration Robustness for Pennycress by Adding Data from Related Species Using Honigs Regression

David E. Honigs(1), M. David Marks(2), Ratan Chopra(2), Peter Aberg(1), Martin Lagerholm(1)
(1) PerkinElmer. Inc 940 Winter Street, Waltham, MA – applications@perten.com
(2) Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108

Introduction

Pennycress (*Thlaspi arvense)* is a flowering plant in the cabbage family. This ancient plant can be planted in the fall, overwinter and then harvested in spring. This makes it a candidate for a protective crop to cover the fields between plantings; especially between plantings of corn (maize) and soy. The pennycress seeds are oilseeds and can be used for fuel oil.

The oil from pennycress would be improved if the fatty acid mix of the oil could be changed by modern breeding techniques.[1] Breeding programs require large number of rapid measurements for the protein, oil and oil types on small amounts of seed so decisions can be made on which lines to propagate. These measurements must be non-destructive and rapid. NIR is well suited for breeding program testing.

Unfortunately, several seasons of a crop can be necessary to make a good NIR calibration[2]. In the case of pennycress, historical NIR data is simply unavailable. Furthermore, a major purpose of the breeding program is to develop changes from prior year yields. This works directly against the NIR calibration needs. To attempt to make robust NIR calibrations for pennycress breeding we combined the spectra of pennycress alongside those of the related *Brassica* family oilseeds, canola and rapeseed, to create calibrations that cover wider ranges of variability. Honigs Regression[3] (HR) was used to bridge the non-linear spectral differences in the data. The resulting calibrations were used to test a new crop of pennycress.[4]

Results

Figure 1 shows the combined calibration data for eicosenoate. The addition of the other species data expands the range significantly. The eicosenoate data alone do not have enough variability to make a robust NIR calibration. The statistics for the individual groups in the combined calibration and the penny cress alone calibration are presented in Table 1. Figure 2 and Table 2 show how the oleate measurement fared with the pennycress alone and expanded data set

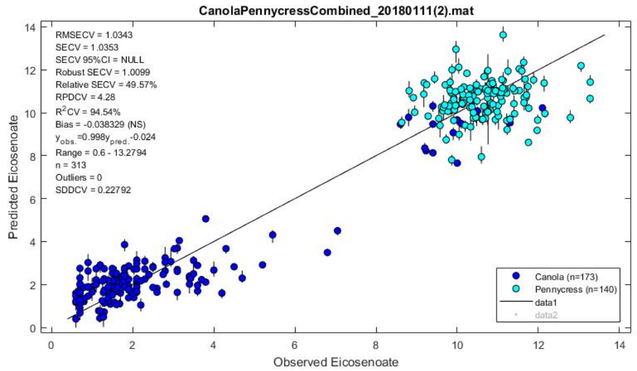

Figure 1 - Eicosenoate Lab v NIR

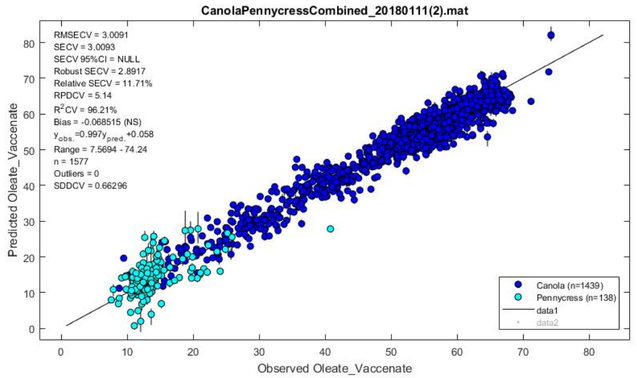

Figure 2- Oleate Lab v NIR Results

The data from this calibration was used for a mutant seed line screening[4]. The authors reported that the correlation between the NIR and Lab for oleic acid for this prediction of this new population was $R^2 = 0.46$ which is better than the initial calibration statistics for pennycress samples alone. These mutant population samples had an oleate level ranging from 8.6 – 44.1 according to the wet chemistry. Predicting the mutant plants accurately would have required the pennycress alone NIR calibration to predict well beyond its previous range of data experience and concentration. The authors[4] further reported that the NIR values were useful for screening for traits as the higher and lower concentrations corresponded well between NIR and lab for the samples tested.

Eicosenoate	Samples	Range	SECV	R^2
Canola/Rapeseed	173	0.6 – 12.1	0.96	0.86
Pennycress	140	8.6 – 13.3	1.16	0.14
Combined	313	0.6 – 13.3	1.0	0.945
Penny Cress Only	141	8.0 – 13.4	0.86	0.07

Table 1 - Eicosenoate Statistics

Oleic Acid	Samples	Range	SECV	R^2
Canola/Rapeseed	1439	8.7 – 74.2	2.8	0.938
Pennycress	138	7.6 – 40.8	4.6	0.328
Combined	1577	7.6 – 73,2	3.0	0.96
Penny Cress Only	145	7.6 – 25.9	2.9	0.37

Table 2- Oleate Statistics

Methods

Spectra of the oilseeds were collected using a PerkinElmer DA7250 measuring sample reflection from 950nm to 1650nm at approximately 12nm resolution and interpolated onto a 5nm data point spacing. Because only small amounts of individual pennycress lines are available, a mirrored cup (Perten Instruments) was used to present these samples to the instrument. The mirror returns the stray and specular light back towards the light source and allows the diffusely reflected light to pass to the detector even when the sample size is smaller than the illumination beam diameter.

The calibration was performed using a proprietary Perten Calibration Tool software program. The data were pretreated using detrending (2nd order) followed by Standard Normal Variate (SNV) conversion. The resulting spectra were then converted individually via a KNN algorithm to use local means in the form of the HR algorithm. The relationship between the analyte and the data was then calculated using Partial Least Squares.

PerkinElmer DA7250

Conclusions

Without the addition of data from other oilseeds it would not have been possible to create calibrations for eicosinate and the oleate that would be expected to predict future breeding results. The number of samples was limited and the diversity of fatty acid concentrations was limited as well. This study showed the HR technique can create a useable calibration from multiple related species of oilseeds.

References

1. Chopra, R., Johnson, E.B., Erin, D., Michaela, M., Dorn, K., Esfahanian, M., Folstad, N., Amundson, K., Altendorf, K., Betts, K., Frels, K., Anderson, J.A., Wyse, D.L., Sedbrook, J.C., Marks, M.D., (2018) Translational genomics using Arabidopsis as a model enables the characterization of pennycress genes through forward and reverse genetics. *The Plant Journal*, 96(6):1093-1105.
2. Inge, B., Gibson, L.R., Rippke, G., Hurburgh, C., (2007) Influence of Yearly Variability of Agricultural Products on Calibration Process: A Triticale Example. *Cereal Chemistry,* 84(6):576-581.
3. Åberg, P., Honigs, D (2012 Global calibrations for animal feed. Perten Science World 6(Nov) ISSN- 2002-245X Retrieved from: www.perten.com/Global/PSW/Perten%20Science%20World%20No.%206.pdf
4. Chopra, R., Folstad, N., Lyons, J., Ulmasov, T., Gallaher, C., Sullivan, L., McGovern, A., Mitacek, R., Frels, K., Altendorf, K., Killam, A., Ismail, B., Anderson, J.A., Wyse, D.L., Marks, M.D., (2019) The adaptable use of Brassica NIRS calibration equations to identify pennycress variants to facilitate the rapid domestication of a new winter oilseed . Industrial Crops & Products 128 55-61.