Protein, weight, and oil prediction by single-seed near-infrared spectroscopy for selection of seed quality and yield traits in pea (*Pisum sativum*)

Running Title: Pea seed quality prediction by single-seed NIRS.

Gokhan Hacisalihoglu[†], Jelani Freeman[†], Paul R. Armstrong[‡], Brad W. Seabourn[§], Lyndon D. Porter[⊥], A. Mark Settles[‖], and Jeffery L. Gustin[‖*]

[†]Biological Sciences Department, Florida A&M University, Tallahassee, FL 32307, USA

[‡] USDA-ARS Grain Marketing & Production Research Center, Manhattan, KS 66502, USA

[§]USDA-ARS Grain Quality & Structure Research Unit, Manhattan, KS 66502, USA

[‖]Horticultural Sciences Department, University of Florida, Gainesville, FL 32611, USA

[⊥]USDA-ARS Grain Legume Genetics and Physiology Research Unit, Prosser, WA 99350, USA

[*]jgustin@ufl.edu; 1-239-324-4817

ABSTRACT

Background: Pea (*Pisum sativum*) is a prevalent cool season crop that produces seeds valued for high protein content. Modern cultivars have incorporated several traits that improved harvested yield. However, progress toward improving seed quality has received less emphasis, in part due to the lack of tools for easily and rapidly measuring seed traits. In this study we evaluated the accuracy of single-seed near-infrared spectroscopy (NIRS) for measuring pea seed weight, protein, and oil content. A total of 96 diverse pea accessions were analyzed using both single-seed NIRS and wet chemistry methods. To demonstrate field relevance, the single-seed NIRS protein prediction model was used to determine the impact of seed treatments and foliar fungicides on protein content of harvested dry peas in a field trial.

Results: External validation of Partial Least Squares (PLS) regression models showed high prediction accuracy for protein and weight (R^2 = 0.94 for both) and less accuracy for oil (R^2 = 0.75). Single seed weight was not significantly correlated with protein or oil content in contrast to previous reports. In the field study, the single-seed NIRS predicted protein values were within 1% of an independent analytical reference measurement and were sufficiently precise to detect small treatment effects.

Conclusion: The high accuracy of protein and weight estimation show that single-seed NIRS could be used in the dual selection of high protein, high weight peas early in the breeding cycle allowing for faster genetic advancement toward improved pea nutritional quality.

KEYWORDS

High-throughput phenotyping; single-seed phenotyping; nutritional quality; seed protein; near-infrared spectroscopy; *Pisum sativum*

INTRODUCTION

Pea (*Pisum sativum* L.) is the third most cultivated legume crop after soybean (*Glycine max*) and peanut (*Arachis hypogaea*) with approximately 37 million tons of dry and fresh peas harvested worldwide in 2017[1]. Peas are a nutritious food for both human and animals and are particularly valued for high protein content, which typically ranges from 18% to 29% in diverse accessions[2-4]. Despite the importance of protein in determining quality, modern cultivars tend to fall on the lower end of the protein content range[3, 5]. Increasing the protein content has proven difficult due to unfavorable linkages with important traits including yield[2-6], seed weight[3], and essential amino acid composition[4]. Protein content is also positively correlated with days to maturity[5], which causes an additional challenge for maintaining protein content in temperate climes where quick maturing types are favored. Selection for any one of these unfavorably correlated traits is likely to reduce protein content unless there is counter selection for high protein.

Selecting new pea breeding lines that have improved seed characteristics, such as high protein, requires high-throughput measurement methods that can analyze large number of lines per breeding cycle. Near-infrared reflectance spectroscopy (NIRS) has been the method of choice because it can accurately predict multiple seed composition traits simultaneously[7]. In peas, bulk NIRS analyzers are used to estimate protein, fiber, and carbohydrates from pea flour[7-11]. Even minor components such as moisture, lignin and hemicellulose could be predicted accurately from flour[8]. A significant time savings was achieved by Arganosa et al (2006) who showed that NIRS could be used to report crude protein content of whole seeds[7], thus saving the destructive process of grinding to flour. To our knowledge, estimation of other seed components using whole seed NIRS has not be been reported for pea.

While bulk NIRS grain analyzers are a powerful and widely used tool for rapidly determining the chemical composition of seeds, they require a pool of seeds for a single measurement. Bulk NIRS cannot be used when seed number is limited in early stage selections of new breeding lines, when evaluating a segregating population, or when sorting seeds for desired traits. In such cases, single-seed NIRS provides an attractive alternative. Armstrong (2006) designed a NIRS platform that collects a spectrum from a seed as it tumbles down an illuminated light tube[12]. This device can accurately predict composition and quality traits of numerous seeds and kernels including maize kernels[13-15], soybean seeds[16], and common beans[17]. Single-seed NIRS has been used to study segregation patterns of kernel composition among kernels from individual maize ears[14] and to conduct QTL analysis of soybean seed composition where seed number per line was limited[18].

The aim of this study was to assess the feasibility of single-seed NIRS as a method of predicting protein, oil, and weight of whole pea seeds. In addition, the potential of single-seed NIRS for breeding and research was evaluated in a field study that tested the effects of multiple seed treatments on harvested pea protein content.

MATERIALS AND METHODS

Pea Seeds. A wide diversity of pea (*Pisum sativum* L.) accessions from USDA National Germplasm Center were selected to ensure NIRS prediction models would be applicable to most mature pea seeds (Figure 1, Supplemental Table 1). Curated phenotypic data associated with each accession was used to select a subset of 96 accessions with the maximum range of protein

content and weight along with selections for diversity by country origin. Three seeds were used per accession.

NIRS and Seed Weight Collection. A custom-built single-seed analyzer was used to acquire seed weight and near infrared reflectance (NIR) spectra from individual pea seeds. The NIR spectrometer (NIR-256-1.7T1, Control Development, South Bend, IN) collected reflectance measurements from 907 to 1688 nm at 1 nm intervals as described previously[12]. The spectral integration time was 24 ms. All spectra were centered to a mean of 1 and absorbance was calculated as log[1/reflectance]. Individual seed weights were measured with a MK4 microbalance (CI Electronics, Salisbury, U.K.) mounted in-line with the NIR spectral acquisition light tube. Four replicates of NIR spectra and weights were collected for each pea seed. The four spectra were averaged for each seed and the averaged spectrum was used for model construction and evaluation. Seed weight NIRS calibration models were on a fresh weight basis.

Seed Composition Analysis. Oil content was measured from individual seeds using a single-seed nuclear magnetic resonance analyzer (NMR, MiniSpec mq20, Bruker BioSpin, Billerica, MA). Seeds were placed in 20 mm test tubes, warmed to 40°C and placed into the instrument for measurement. Oil content was adjusted to a dry weight basis using an initial moisture content of 6.24%.

Total protein content was measured from individual whole seeds with a Leco FP-628 combustion analyzer (St. Joseph, MI, U.S.A.) according to AACC International Approved Method 46-30.01. The reported protein content was 6.25 x N and adjusted to a dry weight basis using an initial moisture content of 6.24%.

Prediction Model Development and Evaluation. Partial Least Squares (PLS) regression was used to model and predict the wet chemistry reference data from the NIR spectra using 'pls' statistical package of R[19, 20]. Calibration and validation datasets were created by randomly splitting the pea accessions into two groups. The calibration dataset contained 192 seeds from 64 randomly selected accessions and the validation set contained 96 seeds from the remaining 32 accessions. The optimum number of PLS factors was selected based on minimization of root mean square error of prediction (RMSEP) of leave-one-out (LOO) cross-validation within the calibration set. The coefficient of determination (R^2) of prediction and calibration along with RMSEP were used to evaluate model performance.

Field Trial for Seed Treatments to Improve Protein Content. Seeds for two green pea cultivars, Ariel and Banner, were obtained from the Northwest Plant Introduction Station, Pullman, WA and treated with combinations of three seed and plant variables as well as untreated seed to test 18 genotype by seed and field conditions for increased protein content (Table 1). Except for a single untreated control condition, all seeds had a standard commercial treatment of mefenoxam, fludioxonil, and molybdenum trioxide using the following products: Apron XL (Syngenta AG, Greensboro, NC) at a rate of 4.73 mL/45.4 kg of seed, Maxim XL (Syngenta AG, Greensboro, NC) at a rate of 2.37 mL/45.4 kg of seed, and molybdenum trioxide at a rate of 9.4 g/45.4 kg of seed.

The first seed treatment variable compared TagTeam Peat Pea and Lentil (Monsanto BioAg, St. Louis, MO), Exceeds Peat Pea and Lentil (Visjon Biologics, Wichita Falls, TX), and no inoculation with *Rhizobium leguminosarum*. Rhizobial inoculates can stimulate nodulation and increase nitrogen fixation in fields that lack sufficient quantities or appropriate strains of nitrogen fixing bacteria[21]. TagTeam Peat Pea and Lentil was applied at a rate of 1,361 kg of

seed/2.2 kg of product. Exceeds Peat Pea and Lentil was applied at a rate of 680.4 kg of seed/2.1 kg of product. The second seed treatment variable was Micnelf DCT micronutrients (Micnelf USA Inc., Hillsboro, OR) applied at a rate of 7.82×10^{-5} g product/g seed. Addition of micronutrients such as zinc, iron, and boron as seed treatments have been shown to improve early plant vigor and final yield in some crops[22]. The third variable compared phosphorous acid salts applied at the third or fourth node growth stage using the commercial fungicide, Phostrol (Nufarm Americas Inc., Alsip, IL) applied at a rate of 4.5 L/hectare. Treatments were replicated four times in plots arranged in a randomized complete block design within the commercial field sites at Milton-Freewater, OR and Kendrick, ID. Each plot was 5.8 m with row spacing of 28 cm, seed spacing of 5 cm, and seeding depth of 5 cm. Standard commercial dry pea production practices were followed under dryland conditions. Plots were harvested in July 2016 and August 2017 in Milton-Freewater and Kendrick, respectively.

From each replicated plot, the harvested seeds were pooled and three sets of 15 seeds were evaluated using single-seed NIRS. Predicted protein was corrected for the bias observed in the PLS regression model using the equation, bias corrected protein = (PLS predicted protein-2.3)/0.9. Each 15-seed sample was then ground into a fine powder using a Retsch Ultra Centrifugal Mill ZM 200 (Verder Scientific Inc., Newtown, PA), homogenized, and 0.1 to 0.19 g of flour was analyzed for N levels in a LECO-FP 520 Combustion Analyzer (LECO Corp., St. Joseph, MI) using the AACC International Cereals and Grains Association Method 46-30.01 Crude Protein-Combustion Method[23]. A 6.25 nitrogen conversion factor was used to determine protein content[24].

Seed and plant treatment effects on protein content were determined by a linear mixed effects analysis using the lme4[25] package in R[19]. The treatment variables were included as fixed

effects without interaction. Planting location, cultivar, and plot were included as random effects with plot nested within location. The calculation of p-values were based on conditional F-tests with Kenward-Roger approximation for the degrees of freedom implemented in the sjPlot package[26]. Separate treatment effect models were built using the protein values from single-seed NIRS and from the bulk analytical protein measurement from the LECO-FP 520 Combustion Analyzer.

RESULTS

Variation among Seed Traits of Diverse Pea Accessions. The analytical values of individual seeds of 96 pea accessions used to calibrate the NIRS prediction models showed wide variation in weight (102-394 mg), protein content (14.2-31.8%), and oil content (0.57-2.14%) (Figure 2a-c; Table 2). The means and ranges of the three traits were similar between the calibration and validation datasets. No significant correlations were detected between seed weight, oil content, or protein content (Figure 2d; Table 3).

PLS Predictive Models. Table 4 shows the statistics for PLS model calibration and validation. Protein was best predicted when the NIR spectra were pre-treated with multiplicative scatter correction (MSC) and six factors were used for the PLS model. Seed weight and oil content required no spectral pretreatment and 12 factors for optimal calibration model performance. The coefficient of determination (R^2) of the calibration model for all three modeled traits was above 0.89 indicating good to excellent model fit. The prediction of the external validation set was likewise excellent for both weight and protein content with R^2 values of 0.94 and root mean square error of predictions (RMSEP) more than three-fold lower than the

population standard deviations shown in Table 2. Close correspondence between the error values for the calibration and validation datasets shows that the calibration models were not overfit for protein and weight. Also, since the validation dataset contained accessions that were not included in the model construction, the validation statistics should reflect the expected error associated with trait predictions of new accessions provided that the trait values fall within the ranges in Table 2. The R^2 for oil prediction in the external validation set fell to 0.76 and the RMSEP rose to 0.16% indicating a loss of predictive accuracy as compared to the calibration model. Reducing the PLS factors did not improve the external validation predictions. However, the RMSEP remained nearly two-fold lower than the population standard deviation indicating that the model would be useful for screening for high and low oil accessions.

Figure 3a-c shows scatterplots of predicted and analytically determined values for the external validation set for each trait. The slope of the linear regression line was over 0.9 for both seed weight and protein content predictions indicating that the prediction error was similar at both the high and low ends of the phenotypic range. There was a slight bias in oil content prediction with a slight underestimate of high oil seeds. This bias would tend to degrade the accuracy of prediction of extremely high oil peas. Our population contained only 11 seeds with oil content higher than 1.7%. Supplementing the calibration population with additional high oil genotypes would likely reduce bias and improve the prediction accuracy for higher oil seeds.

Field Study. The utility of NIRS predictions in legume production research was tested for comparing dry pea protein content in green peas. Two cultivars, Ariel and Banner, were given pre-planting seed treatments and a foliar fungicide treatment to test if the treatments altered protein content of the harvested peas seeds (Table 1). A linear mixed effects (LME)

model was used to evaluate the effects of the treatments. Independent LME models were constructed for the NIRS predicted protein values and bulk analytical values.

The two protein measurement techniques found very similar treatment effects (Table 5). The largest treatment effect was from the standard seed treatment, which significantly increased protein content by 1.54% and 1.12% in the NIRS predictions and bulk analytical measurements, respectively. The positive effect was consistent across both locations and cultivars (Table 6). The only substantial difference between the two protein measurement techniques was the effect of the rhizobial inoculant, Tag Team, which reduced protein by 0.25% and was significant in the bulk analytical method. The protein reduction estimated by NIRS was 0.13% and not significant in the mixed effect model. Neither phosphorous acid nor micronutrient treatment altered protein levels of the harvested peas using either protein measurement technique. Among the random effects, the cultivar Ariel had higher protein content than Banner and the protein difference between the cultivars was greater in the NIRS predicted protein data. There was little difference in protein content between trial locations. The conditional R^2, which takes into account both fixed and random effects, was slightly higher for the NIRS predicted protein values indicating that the fixed and random effects explained more of the variation in the predicted protein values than the analytical protein values.

The overall mean of NIRS protein values was 1% higher than bulk analytical values (19.3% and 18.3%, respectively) ($t(428)=16.2$, $p<0.001$, Paired t-test). However, there was as strong cultivar effect between the two methods (Figure 3d). Ariel had a larger deviation between protein assays with a mean difference of 1.7% between NIRS and C/N analysis. By contrast, Banner had a mean difference of 0.3% between NIRS and C/N analysis (Table 6).

DISCUSSION

The PLS models developed in this study showed that NIRS can accurately predict the weight and protein content of individual, intact pea seeds. The error associated with protein prediction (RMSEP) of the external validation set was 1.16%, which was slightly higher than the standard error of prediction of 0.93% reported by Arganosa et al (2006)[7]. That study collected NIRS from samples of pooled seeds and then ground the pooled seeds for a crude protein measurement, while our study used single-seed spectra and single-seed analytical values.

The minor difference in accuracy between the two studies could be caused by experimental design differences. Our study included a diverse array of 96 accessions, which were divided into 64 accessions to develop the PLS protein prediction model and test the calibration on a set of 32 additional accessions. Arganosa et al (2006) used a smaller set of only 36 accessions and subsampled the same lines for both the calibration and validation steps. Sampling the same lines for calibration and validation will tend to underestimate error for predicting new genotypes. For example, if we split the 96 accessions in this study with 2 of the 3 seeds from each genotype for the calibration data set and the remaining seed used for validation, the error of protein prediction drops to 0.94%. Therefore, the protein prediction error associated with single-seed NIRS appears to be at least as robust as whole seed bulk NIRS.

Peas have very low oil content at <4% when compared to other legumes such as soybean, which has an average oil content of 20%[27, 28]. Several studies have discussed the value in developing pea as a low cost oilseed crop for cultivation in colder climates, such as Canada and Northern Europe[27, 29]. We found that single-seed NIRS could estimate pea oil content to within 0.16%. The R^2 value for oil prediction was lower than seed protein or weight prediction models. However, the phenotypic range of the population was relatively narrow, 0.61% to 1.89%. By

comparison, a pea population used in Ahmad et al. (2015) ranged from 0.87% to 3.67%[29]. It is likely the lower R^2 is due, at least in part, to the narrow phenotypic range. The GRIN germplasm database did not contain information on oil content of peas, which limited our ability to choose accessions with maximal oil variation. Even so, the error of prediction (0.16%, Table 4) was approximately half of the standard deviation of the population (0.30%, Table 2), which would make single-seed NIRS an effective tool for separating high and low oil peas within the phenotypic range of our population. To apply this method to predict oil content of pea varieties far outside the oil range used in this study, the calibration would need to incorporate peas with higher oil content. We expect a broader range of oil content would improve the R^2, and that single-seed NIRS could be used for selection of high oil pea varieties.

We found no significant correlation between pea seed weight and protein or oil content (Table 3). There was a slight negative trend for both comparisons, but the coefficient did not surpass a p-value threshold of 0.05. This is consistent with several other studies that found little evidence of a relationship between seed weight and protein or oil content in pea[4, 6, 27], with a noted exception[3]. Larger seeds tended to produce higher yields in some populations[3, 6] although other studies found no relationship[2]. The lack of a correlation between seed weight, oil, and protein should allow for selection of larger seeds with improved nutritional content. For example, two accessions within our dataset, ILCA5075 and Agassiz, had high weight (>300 mg) and high protein content (>25%) and additionally Agassiz contained relatively high oil (>1.6%) for our population (Supplemental Table 1). The ability of single seed NIRS to simultaneously determine seed oil, protein, and weight would allow for faster screening of these potentially high value lines early in a breeding population where large quantities of seed are not available.

To test the practical application of single-seed NIRS, we applied the calibrations to a production field study with a replicated design incorporating multiple treatments and cultivars. This design allowed for further independent evaluation of NIRS prediction accuracy and precision. NIRS protein values were approximately 1% higher than the analytical measurements with a trend for over-predicting values in both cultivars. It is possible that differences in C/N measurements between the Leco FP-628 combustion analyzer used for NIRS calibration and the Leco FP-520 analyzer used for bulk pea flour in the field study contributed to this small discrepancy. In addition, the Ariel cultivar had a larger discrepancy between NIRS and C/N analysis than the Banner cultivar (1.7% versus 0.3%, respectively). The relatively high prediction error associated with Ariel, but not Banner, suggests that Ariel may be a less accurately predicted cultivar, which would impact our overall estimate of NIRS accuracy. Coincidentally, Ariel was among the 100 cultivars selected for building the NIRS prediction model (Supplemental table 1) and was used for external validation. Consistent with the field study observations, Ariel protein content was overpredicted by 0.7%.

The standard seed treatment application had the largest effect on protein content of the harvested peas and was detectable with both NIRS and C/N methods. This increase was estimated to be >1% with both protein methods, which is similar to the 1.16% RMSEP of the NIRS calibration. The much smaller effect of rhizobia inoculation with TagTeam, which lowered protein by 0.25%, was only significant with C/N analysis. These results are consistent with the relative error associated with NIRS and C/N analysis and suggest that single-seed NIRS is able to accurately detect 1-1.5% differences in protein for new samples. This error is approximately 5% of the protein content range we observed in our diverse accessions.

ACKNOWLEDGEMENTS

The authors thank USDA National Plant Germplasm System for the seeds and John Baier for technical assistance. This work was supported by the National Science Foundation Plant Genome Research Program awards IOS-103141 and IOS-1444456, National Institute of Food and Agriculture award 2018-51181-28419, and the Vasil-Monsanto Endowment. The mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval for the exclusion of other products that may also be suitable. USDA is an equal opportunity provider and employer.

REFERENCES

1. Food and Agriculture Organization of the United, N., FAOSTAT statistics database. [Rome, Italy] : FAO, 1998-: 2018.

2. Gali, K. K.; Liu, Y.; Sindhu, A.; Diapari, M.; Shunmugam, A. S. K.; Arganosa, G.; Daba, K.; Caron, C.; Lachagari, R. V. B.; Tar'an, B.; Warkentin, T. D., Construction of high-density linkage maps for mapping quantitative trait loci for multiple traits in field pea (Pisum sativum L.). *BMC Plant Biology* **2018,** *18* (1), 172.

3. Jha, A. B.; Arganosa, G.; Tar'an, B.; Diederichsen, A.; Warkentin, T. D., Characterization of 169 diverse pea germplasm accessions for agronomic performance, Mycosphaerella blight resistance and nutritional profile. *Genetic Resources and Crop Evolution* **2013,** *60* (2), 747-761.

4. Solberg, S. Ø.; Yndgaard, F.; Poulsen, G.; von Bothmer, R., Seed yield and protein content in the Weibullsholm Pisum collection. *Genetic Resources and Crop Evolution* **2017,** *64* (8), 2035-2047.

5. Tar'an, B.; Warkentin, T.; Somers, D. J.; Miranda, D.; Vandenberg, A.; Blade, S.; Bing, D., Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (Pisum sativum L.). *Euphytica* **2004,** *136* (3), 297-306.

6. Krajewski, P.; Bocianowski, J.; Gawłowska, M.; Kaczmarek, Z.; Pniewski, T.; Święcicki, W.; Wolko, B., QTL for yield components and protein content: a multienvironment study of two pea (Pisum sativum L.) populations. *Euphytica* **2012,** *183* (3), 323-336.

7. Arganosa, G. C.; Warkentin, T. D.; Racz, V. J.; Blade, S.; Phillips, C.; Hsu, H., Prediction of crude protein content in field peas using near infrared reflectance spectroscopy. *Canadian Journal of Plant Science* **2006,** *86* (1), 157-159.

8. Saha, U.; Vann, R. A.; Chris Reberg-Horton, S.; Castillo, M. S.; Mirsky, S. B.; McGee, R. J.; Sonon, L., Near-infrared spectroscopic models for analysis of winter pea (Pisum sativum L.) quality constituents. *Journal of the Science of Food and Agriculture* **2018,** *98* (11), 4253-4267.

9. Jung, C. S.; Kim, B. J.; Kwon, Y. C.; Han, W. Y.; Kwack, Y. H., Analysis of protein and moisture contents in pea (Pisum Sativum L.) using near-infrared reflectance spectroscopy. *Korean Journal of Crop Science* **1998,** *43*, 101-104.

10. Davies, A. M. C.; Coxon, D. T.; Gavrel, G. M.; Wright, D. J., Determination of starch and lipid in pea flour by near infrared reflectance analysis. the effect of pea genotype on starch and lipid content. *Journal of the Science of Food and Agriculture* **1985,** *36* (1), 49-54.

11. Davies, A. M. C.; Wright, D. J., Determination of protein in pea flour by near infrared analysis. *Journal of the Science of Food and Agriculture* **1984,** *35* (9), 1034-1039.

12. Armstrong, P. R., Rapid single-kernel NIR measurment of grain and oil-seed attributes. *Applied Engineering in Agriculture* **2006,** *22* (5), 767-772.

13. Gustin, J.; Jackson, S.; Williams, C.; Patel, A.; Armstrong, P.; Peter, G.; Settles, A., Analysis of Maize (Zea mays) Kernel Density and Volume Using Microcomputed Tomography and Single-Kernel Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry* **2013,** *61* (46), 10872-10880.

14. Spielbauer, G.; Armstrong, P.; Baier, J.; Allen, W.; Richardson, K.; Shen, B.; Settles, A., High-Throughput Near-Infrared Reflectance Spectroscopy for Predicting Quantitative and Qualitative Composition Phenotypes of Individual Maize Kernels. *Cereal Chemistry* **2009,** *86* (5), 556-564.

15. Tallada, J.; Palacios-Rojas, N.; Armstrong, P., Prediction of maize seed attributes using a rapid single kernel near infrared instrument. *Journal of Cereal Science* **2009,** *50* (3), 381-387.

16. Hacisalihoglu, G.; Gustin, J. L.; Louisma, J.; Armstrong, P.; Peter, G. F.; Walker, A. R.; Settles, A. M., Enhanced Single Seed Trait Predictions in Soybean (Glycine max) and Robust Calibration Model Transfer with Near-Infrared Reflectance Spectroscopy. *Journal of Agricultural and Food Chemistry* **2016,** *64* (5), 1079-1086.

17. Hacisalihoglu, G.; Larbi, B.; Settles, A. M., Near-Infrared Reflectance Spectroscopy Predicts Protein, Starch, and Seed Weight in Intact Seeds of Common Bean (Phaseolus vulgaris L.). *Journal of Agricultural and Food Chemistry* **2010,** *58* (2), 702-706.

18. Hacisalihoglu, G.; Burton, A.; Gustin, J.; Eker, S.; Asikli, S.; Heybet, E.; Ozturk, L.; Cakmak, I.; Yazici, A.; Burkey, K.; Orf, J.; Settles, A., Quantitative trait loci associated with soybean seed weight and composition under different phosphorus levels. *Journal of Integrative Plant Biology* **2018,** *60* (3), 232-241.

19. Team, R. C. *R: A language and environment for statistical computing.*, 2017.

20. Mavik, B.; Wehrens, R.; Liland, K. H. *pls: partial least squares and principal component regression.*, 2.3-0; http:// CRAN.R-project.org/package=pls, 2011.

21. Catroux, G.; Hartmann, A.; Revellin, C., Trends in rhizobial inoculant production and use. *Plant and Soil* **2001,** *230* (1), 21-30.

22. Mondal, S.; Bose, B., Impact of micronutrient seed priming on germination, growth, development, nutritional status and yield aspects of plants. *Journla of Plant Nutrition* **2019,** *42* (19), 2577-2599.

23. Analysis., A.I.A.M.., Crude Protein-Combustion Method. 11 t^h Ed. ed.; St. Paul, MN, U.S.A, 2019.

24. Boisen, S.; Bech-Andersen, S.; Eggum, B. r. O., A Critical View on the Conversion Factor 6.25 from Total Nitrogen to Protein. *Acta Agriculturae Scandinavica* **1987,** *37* (3), 299-304.

25. Bates, D.; Maechler, M.; B, B.; Walker, S. *Fitting Linear Mixed-Effects Models Using lme4.*, 2015.

26. Lüdecke, D. *sjPlot: Data Visualization for Statistics in Social Science.*, 2.7.2; 2019.

27. Khodapanahi, E.; Lefsrud, M.; Orsat, V.; Singh, J.; Warkentin, T. D., Study of Pea Accessions for Development of an Oilseed Pea. *Energies* **2012,** *5* (10).

28. Li, D.; Zhao, X.; Han, Y.; Li, W.; Xie, F., Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* **2019,** *111* (1), 90-95.

29. Ahmad, S.; Kaur, S.; Lamb-Palmer, N. D.; Lefsrud, M.; Singh, J., Genetic diversity and population structure of Pisum sativum accessions for marker-trait association of lipid content. *The Crop Journal* **2015,** *3* (3), 238-245.

FIGURE LEGENDS

Figure 1. Example seeds for the 96 pea accessions used for development of NIRS prediction models.

Figure 2. Distribution of (a) weight, (b) protein, and (c) oil for the 96 pea accessions. (d) Scatterplot of analytically determined weight and protein content for all seeds in the study including and a linear regression trend line (solid line) and the coefficient of determination (R^2) for the linear regression model.

Figure 3. Scatterplot of NIRS predicted and analytical pea seed (a) weight, (b) protein, and (c) oil. Each plot shows values for the external validation samples and a linear regression trend line (solid line) with the associated linear model terms. Dotted line shows one to one correspondence. (d) Bar chart showing average protein content and standard errors for harvested pea seed composition. NIRS (white) and C/N (grey) measures are given for each cultivar at each location. All treatments were included.

SUPPLEMENTAL MATERIAL

Supplemental Table 1. Average and standard deviation (sd) of seed composition for the pea accessions used in the study. Each accession is identified by USDA National Germplasm Center information.

Table 1. Description of seed treatments and foliar applications of fungicides associated with dry pea field trials.

treatment #	cultivar	Inoculant †	micronutrient	phos. acid	seed treatment
1	Ariel	no	yes	yes	standard
2	Ariel	TagTeam	no	no	standard
3	Ariel	no	no	no	standard
4	Ariel	no	no	yes	standard
5	Ariel	TagTeam	yes	yes	standard
6	Ariel	TagTeam	no	yes	standard
7	Ariel	Exceeds	no	no	standard
8	Ariel	Exceeds	no	yes	standard
9	Ariel	no	no	no	no
10	Banner	no	no	yes	standard
11	Banner	TagTeam	yes	yes	standard
12	Banner	TagTeam	no	no	standard
13	Banner	Exceeds	no	no	standard
14	Banner	no	no	no	standard
15	Banner	TagTeam	no	yes	standard
16	Banner	no	yes	yes	standard
17	Banner	Exceeds	no	yes	standard
18	Banner	no	no	no	no

†The word “no” signifies no inoculant was applied.

Table 2. Descriptive statistics of pea germplasm for calibration set and validation set

	calibration				validation			
seed trait	**N**	**mean**	**sd**	**range**	**N**	**mean**	**sd**	**range**
protein (%)	192	22.8	4.31	12.6-33.1	96	21.6	4.0	12.4-31.9
weight (mg)	192	228	64.0	86.5-453	96	216	66.7	87.6-379
oil (%)	192	1.03	0.31	0.70-2.20	96	1.14	0.30	0.61-1.89

Table 3. Pearson correlations coefficient (*r*) between seed composition traits of 96 pea accessions

	protein	weight	oil
protein (%)	1		
weight (mg)	-0.20	1	
oil (%)	0.02	-0.21	1

Table 4. PLS Calibration and Validation statistics for predicting pea seed traits with single seed NIRS.

			calibration		validation	
seed trait	**pre-treatment** †	**factors**	**R^2**	**RMSEC** ‡	**R^2**	**RMSEP**
protein (%)	MSC	6	0.97	1.02	0.94	1.16
weight (mg)	none	12	0.97	17.0	0.94	16.7
oil (%)	none	12	0.89	0.10	0.76	0.16

†MSC – multiplicative scatter correction, ‡ RMSEC/P – root mean square error of calibration/prediction.

Table 5. Comparison of mixed effect model effect sizes and significance of seed and foliar treatments on single-seed NIRS predicted and analytical protein content of cultivars Ariel and Banner grown in two commercial field trials.

	NIRS predicted protein content (% fw)		*analytical protein content (% fw)*	
fixed effects	**estimates**	**std error**	**estimates**	**std error**
intercept	17.99 ***	1.26	17.37 ***	0.53
inoculant: Exceeds	-0.24	0.15	-0.18	0.10
inoculant: TagTeam	-0.13	0.13	-0.25 **	0.08
micro-nutrient: yes	-0.04	0.15	-0.04	0.10
phos. acid: yes	0.06	0.13	0.14	0.08
standard: yes †	1.54 ***	0.20	1.12 ***	0.13
random effects				
σ^2	1.22		0.49	
plot:location	0.27		0.08	
cultivar	1.55		0.27	
ICC ‡	0.60		0.41	
$N_{location}$	2		2	
N_{Plot}	4		4	
$N_{Cultivar}$	2		2	
marginal R^2/ conditional R^2	0.07/0.63		0.13/0.49	

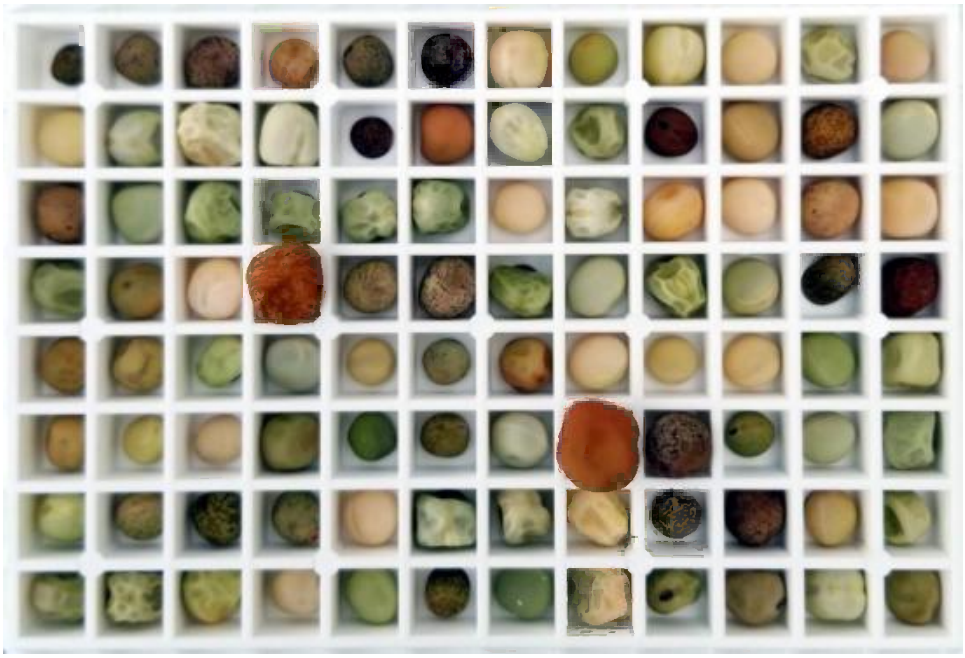
† Standard seed treatment, ‡ CC – interclass correlation, p<0.05=*, p<0.01=**, p>0.001=***

Table 6. NIRS predicted and analytical percent protein content of dry pea cultivars harvested from field trials in Kendrick, ID and Milton-Freewater, OR.

		Idaho			Oregon		
treatment †	cultivar	% NIRS protein ‡	% C/N protein §	NIRS-C/N ¶	% NIRS protein	% C/N protein	NIRS-C/N
1	Ariel	21.7	19.4	-2.3*	20.5	18.8	-1.7*
2	Ariel	21.5	19.0	-2.5*	19.9	18.5	-1.4*
3	Ariel	21.1	19.5	-1.6*	19.9	18.5	-1.4*
4	Ariel	21.1	19.6	-1.5*	20.9	18.9	-2*
5	Ariel	20.9	19.3	-1.6*	20.5	18.7	-1.8*
6	Ariel	20.7	19.0	-1.7*	20.2	18.7	-1.5*
7	Ariel	20.7	18.8	-1.9*	20.6	18.8	-1.8*
8	Ariel	20.5	19.0	-1.5*	20.3	19.0	-1.3*
9	Ariel	19.0	17.3	-1.7*	19.4	18.2	-1.2*
10	Banner	18.5	18.3	-0.2	18.5	17.9	-0.6
11	Banner	18.4	17.9	-0.5	18.0	17.9	-0.1
12	Banner	18.3	17.6	-0.7*	18.2	17.9	-0.3
13	Banner	18.2	17.6	-0.6	18.4	17.9	-0.5
14	Banner	18.2	18.2	-0.1	18.4	17.9	-0.5
15	Banner	18.0	17.5	-0.5	18.6	17.9	-0.7
16	Banner	18.0	17.7	-0.2	17.9	18.1	0.2
17	Banner	17.8	17.7	-0.1	18.5	18.3	-0.2
18	Banner	16.8	16.5	-0.2	17.8	17.5	-0.3

† See Table 2 for treatment descriptions. ‡ Mean predicted percent protein content of pea seeds based on Near infrared Spectroscopy of individual seeds. § Mean percent protein content of ground pea flour. ¶ The subtraction of the mean predicted percent protein content from the mean actual percent protein. * Significant ($p<0.01$) difference between the mean predicted percent protein content and the mean analytical percent protein content using a paired Student's t-test.

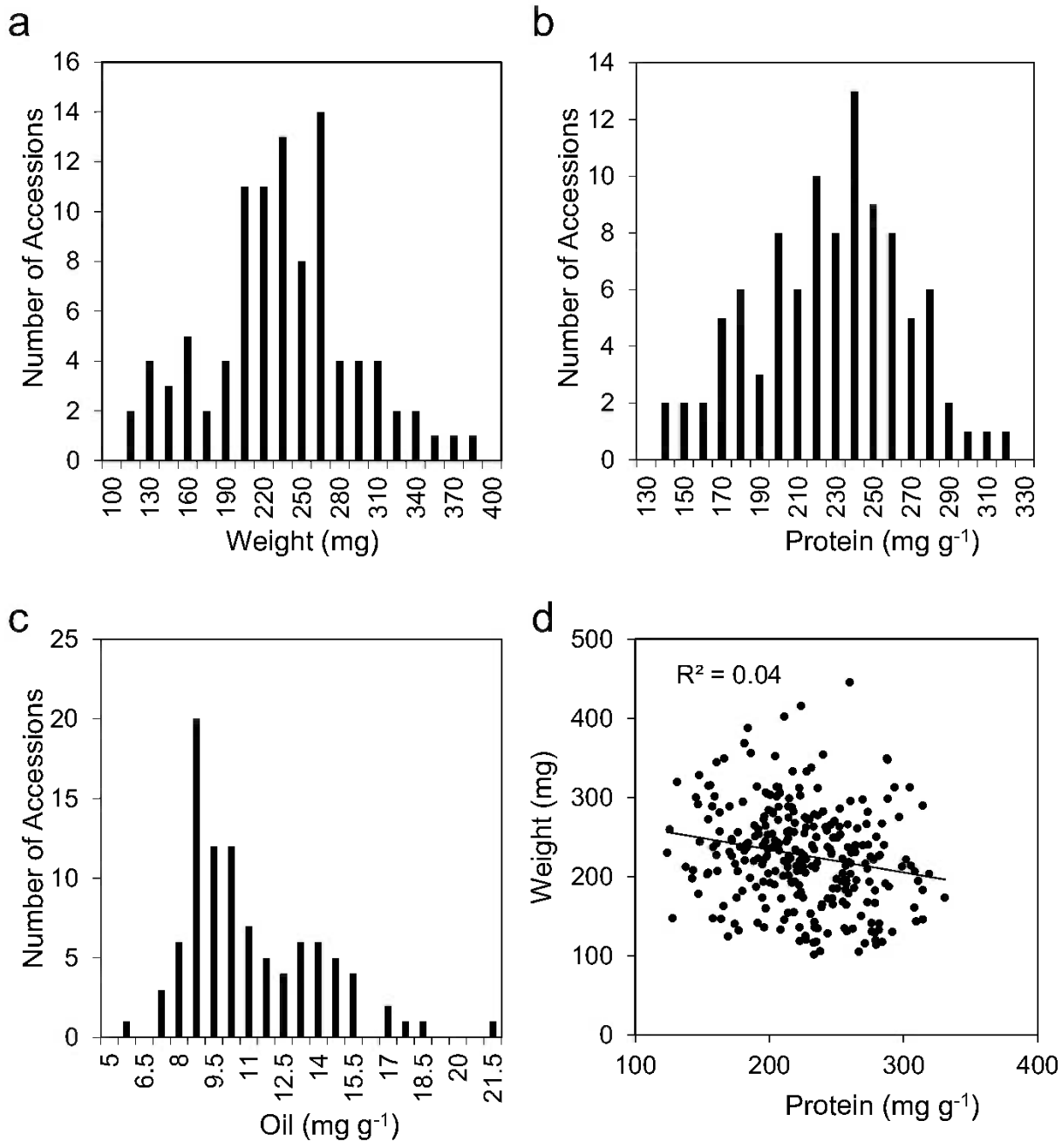
443 Figure 1.



444

445

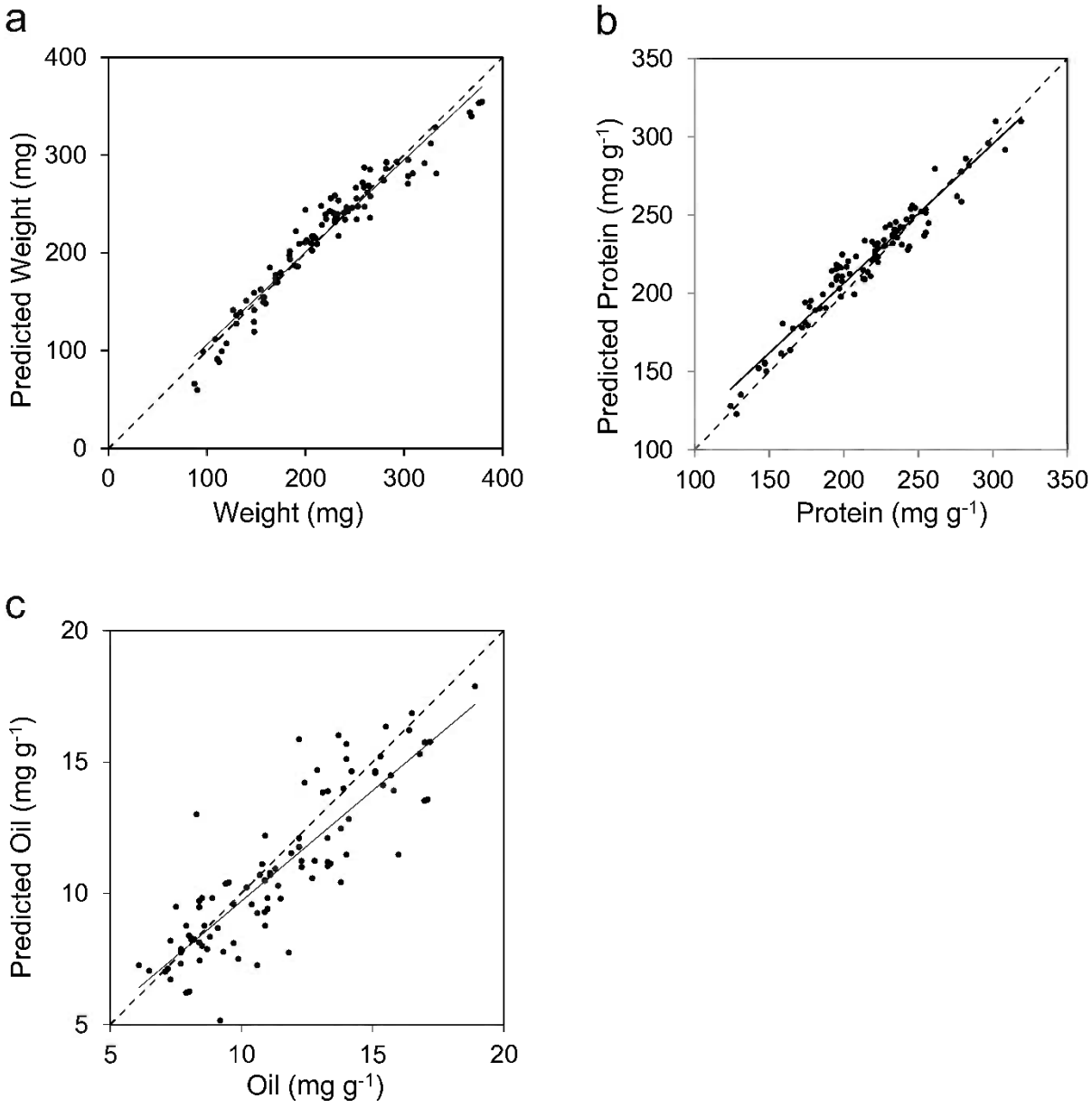
446 Figure 2.



447

448

449 Figure 3.



450

451

452

Supplemental Table 1. Average and standard deviation (sd) of seed composition for the pea accessions used in the study. Each accession is identified by USDA National Germplasm Center information.

Accession Name	Plant Name	PlantID	Accession ID	Taxonomy	Origin	Weight (mg)	Weight sd	Protein (mg g-dw^{-1})	Protein sd	Oil (mg g-dw^{-1})	Oil sd
CDC-Sonata	250	PI 383730	1287912	Pisum sativum	Turkey	158	32.3	278	31.5	10.7	2.4
CDC-Striker	251	PI 383731	1287913	Pisum sativum	Turkey	153	20.8	240	20.0	11.2	0.9
CDC-Bronco	22719	PI 343988	1256643	Pisum sativum subsp. sativum	Turkey	231	12.2	194	31.7	9.0	1.2
ILCA5052	27b	PI 358300 PSP	1823038	Pisum sativum	United States, Washington	129	10.5	254	30.9	11.3	1.0
Canstar	9006/60	PI 273207	1822941	Pisum sativum	United States, Washington	111	5.7	272	6.6	12.2	0.7
Amigo	Aa87	PI 269777	1202942	Pisum sativum	United Kingdom, England	188	17.3	318	11.6	9.0	0.8
Carneval	ACACIA	PI 275824	1206544	Pisum sativum	Sweden	281	19.2	269	18.1	12.7	0.6
Ucero	AI ZI WAN DOU	W6 44796	1902193	Pisum sativum	China, Guangxi	295	6.4	185	18.1	9.8	2.0
Cruiser	ALFETTA	PI 596334	1536467	Pisum sativum	Netherlands	263	44.1	178	11.5	9.1	1.3
Carousel	ARTHUR	PI 278777	1208109	Pisum sativum	Canada, Ontario	232	36.9	224	50.5	9.1	1.8
Champagne	AWARD	PI 570649	1465620	Pisum sativum var. sativum	United States	222	15.6	271	1.7	11.0	0.5
Teresa	BAI WAN	W6 44781	1902178	Pisum sativum	China, Guangxi	246	41.8	172	36.3	10.6	1.8
Snowbird	BAI WAN DOU	W6 44762	1902159	Pisum sativum	China, Nei Monggol	251	41.8	198	16.7	8.3	1.0
22719	BON AMI	PI 244101	1188703	Pisum sativum	Netherlands	237	34.2	229	3.5	7.6	0.4
ILCA5075	BR 1-49-9	PI 365419 PSP	1823043	Pisum sativum	United States, Washington	305	13.9	304	10.4	21.4	0.8
HL67	BRZAK	PI 357290 PSP	1823036	Pisum sativum	United States, Washington	301	34.5	201	14.8	8.7	0.6
VavD265	CAI WAN	W6 44797	1902194	Pisum sativum	China, Guangxi	198	22.3	219	13.4	9.3	0.8
Camry	CAPUCINORUM	PI 272212	1204560	Pisum sativum	Ethiopia	225	75.8	231	44.0	7.9	1.4
22722	CELSIOR	PI 244106	1188708	Pisum sativum	Netherlands	227	5.3	267	8.9	13.1	0.8
CDC-Bronco	CENNIA	PI 341888	1254996	Pisum sativum	Netherlands	220	29.6	248	35.6	18.2	2.5
Cameor	CENTRALI-SIBIRICUM	PI 272207	1204555	Pisum sativum	Greece	186	11.2	250	34.9	8.7	1.3
CDC-Acer	CHINESE SNOW PEA	PI 279933	1208541	Pisum sativum	United States, New York	318	73.2	225	12.1	5.7	0.9
Agra	Col. No. 431	PI 269543	1202819	Pisum sativum	Pakistan	265	35.6	213	33.8	7.0	0.8
29579	COLUMBA	PI 244116	1188718	Pisum sativum	Netherlands	281	38.7	165	15.4	8.7	1.1
Bolero	CONCOLON	PI 272195	1204543	Pisum sativum	Germany	238	16.2	212	25.1	8.3	0.6
45760	CONCORDIA	PI 244117	1188719	Pisum sativum	Netherlands	256	12.0	260	17.3	8.3	0.9
Wt11238	cri-1	PI 669379	1913342	Pisum sativum	United States, California	328	29.6	232	27.9	13.5	1.7
Wt1238	cri-1 af	PI 669381	1913344	Pisum sativum	United States, California	245	58.8	219	18.4	14.5	1.0
Wt3557	cri-1 st	PI 669384	1913347	Pisum sativum	United States, California	213	23.3	210	13.1	16.3	0.9
Wt10245	Ctirad	W6 45205	1905439	Pisum sativum	United States, Ohio	241	47.0	206	10.7	17.6	3.1
Alfetta	DE HAAN 204.1	PI 269766	1202931	Pisum sativum	Sweden	214	29.5	231	41.2	8.9	0.8
1-150-81	DELI	PI 244126	1188728	Pisum sativum	Netherlands	259	3.1	218	7.7	13.7	1.8
Cream40	DELTA	PI 594358	1852355	Pisum sativum	United States, Washington	257	24.1	169	10.6	9.7	1.0
Parafield	DS ADMIRAL	PI 633698 PSP	1852356	Pisum sativum	United States, Washington	195	23.5	216	13.5	8.5	0.9
Melrose	ROSAKRONE	PI 477371 PSP	1823069	Pisum sativum	United States, Washington	312	1.1	216	17.1	8.5	0.7
Cerise	ERBI	PI 568227	1463198	Pisum sativum subsp. sativum	Czechoslovakia	258	54.3	166	30.7	10.0	2.5
1-2347-144	FOLI	PI 244142	1188744	Pisum sativum	Netherlands	252	21.9	235	10.1	13.4	1.6
Dakota	FORAGER	PI 634508	1650792	Pisum sativum	United States, Wyoming	207	37.4	234	35.5	9.6	0.4
CDC-Golden	G 22442	PI 371796	1278678	Pisum sativum	New Zealand	336	18.4	161	5.1	8.9	0.5
CDC-Sage	GASTRO MAPLE	PI 381333	1285870	Pisum sativum	Netherlands	368	58.7	210	22.9	8.6	0.6
Badminton	GRISEO COLORATUM	PI 272158	1204507	Pisum sativum	Germany	220	33.4	166	11.9	8.5	0.7
Ballet	GRISEO COLORATUM	PI 272160	1204509	Pisum sativum	Ethiopia	272	24.5	149	25.8	10.1	1.7
JI281	GRUNE PERLE	PI 413698 PSP	1823058	Pisum sativum	United States, Washington	223	2.5	235	34.9	14.4	1.3
37A	GRUNO	PI 244155	1188757	Pisum sativum	Netherlands	223	35.3	250	13.0	8.0	0.8
Agassiz	HADA	PI 244157	1188759	Pisum sativum	Netherlands	381	56.0	278	16.3	9.9	2.2
Cascadia	HERALDA	PI 279826	1208522	Pisum sativum	Germany	260	5.4	245	27.0	8.0	0.7
B10-10	HIEMALE	PI 272155	1204504	Pisum sativum	Peru	161	45.8	173	36.9	12.6	0.8
Isard	HOHENHEIMER ROSABLUEHENDE	PI 409031 PSP	1823050	Pisum sativum	United States, Washington	264	10.9	246	17.0	8.3	0.9
Sugar-Anne	HONG WAN DOU	W6 44769	1902166	Pisum sativum	China, Nei Monggol	205	19.8	172	18.3	9.8	1.1
Solara	HUI HU LU	W6 44764	1902161	Pisum sativum	China, Nei Monggol	200	55.0	181	16.8	11.0	1.5
Kaspa	I.P. 2	PI 413703 PSP	1823059	Pisum sativum	United States, Washington	297	50.3	209	26.7	7.9	0.3
Delta	IG 122954	PI 654542	1738087	Pisum sativum var. sativum	Turkey	231	11.0	251	16.7	8.4	0.3
DS-Admiral	IG 122965	PI 654544	1738098	Pisum sativum var. sativum	Kazakhstan	192	31.6	241	23.1	8.3	0.8
Dullwhitepea	IG 122966	PI 654545	1738099	Pisum sativum var. sativum	Turkmenistan	141	10.0	290	20.6	10.6	0.5
Early-Alaska	IG 122968	PI 654547	1738101	Pisum sativum var. sativum	Turkey	201	11.5	297	19.6	8.2	0.0
Eclipse	IG 122991	PI 654563	1738123	Pisum sativum var. sativum	China	204	7.8	270	25.7	12.4	0.6

Accession Name	Plant Name	PlantID	Accession ID	Taxonomy	Origin	Weight (mg)	Weight sd	Protein (mg g-dw^{-1})	Protein sd	Oil (mg g-dw^{-1})	Oil sd
Eiffel	IG 122992	PI 654564	1738124	Pisum sativum	Kazakhstan	189	53.9	277	11.0	8.9	2.0
Esla	IG 123002	PI 654573	1738133	Pisum sativum var. sativum	Italy	218	24.8	222	2.1	8.6	0.7
Espace	IG 123003	PI 654574	1738134	Pisum sativum var. sativum	Tunisia	283	13.7	264	13.6	6.9	0.6
Fortune	IG 123005	PI 654576	1738136	Pisum sativum var. sativum	Italy	214	24.2	251	17.1	14.7	2.3
Frosty	IG 123006	PI 654577	1738137	Pisum sativum var. sativum	Ethiopia	160	21.7	234	33.4	9.9	1.0
G611764	IG 123010	PI 654581	1738141	Pisum sativum var. sativum	Ethiopia	161	17.9	260	30.8	9.5	0.6
Daytona	IG 52553	PI 654538	1738083	Pisum sativum	Russian Federation	178	14.1	258	26.9	10.5	0.4
CDC-Vienna	ILCA 5075	PI 505111	1400047	Pisum sativum var. sativum	Syria	229	44.0	196	17.8	8.0	0.8
MPG87	ILCA 5076	PI 505059 PSP	1823073	Pisum sativum var. sativum	United States, Washington	129	4.7	235	8.4	11.4	0.8
MPG87	ILCA 5089	PI 505122 PSP	1823078	Pisum sativum var. sativum	United States, Washington	125	9.4	228	5.6	12.4	1.0
Orb	ILCA 5115	PI 505144 PSP	1823080	Pisum sativum var. sativum	United States, Washington	238	23.3	253	7.6	7.3	0.2
ILCA-5094	IMPOSANT BROWN	PI 381334 PSP	1823046	Pisum sativum	United States, Washington	343	33.9	189	8.9	8.3	0.3
Sugar-Snap	KE KE WAN DOU	W6 44772	1902169	Pisum sativum	China, Nei Monggol	266	47.6	142	25.9	10.8	1.5
ILCA5115	KLATOVSKA OZIMA	PI 393488 PSP	1823047	Pisum sativum	United States, Washington	150	6.5	226	10.6	11.2	0.4
Green-Arrow	KOMIDORI	PI 355906 PSP	1823028	Pisum sativum	United States, Washington	235	35.1	205	8.4	8.1	0.8
J12605	LANCET	PI 413683 PSP	1823055	Pisum sativum	United States, Washington	203	55.9	248	32.4	13.4	4.6
Rambo	Lifter	PI 628276 PSP	1882459	Pisum sativum	United States, Washington	244	26.9	275	26.1	6.8	0.8
Treasure	LU CAI WAN	W6 44791	1902188	Pisum sativum	China, Guangxi	199	35.2	196	51.6	13.6	2.5
Sugar-Daddy	MA LI WAN	W6 44770	1902167	Pisum sativum	China, Nei Monggol	156	21.5	173	6.1	11.8	0.6
Wando	MA WAN	W6 44804	1902201	Pisum sativum	China, Guangxi	145	3.8	153	23.6	14.1	1.3
P651	SW Carousel	PI 638516 PSP	1852353	Pisum sativum	United States, Washington	201	37.5	226	25.2	9.9	0.7
Ariel	WILLIAM MASSEY, LINE	PI 269820	1202982	Pisum sativum	United Kingdom, England	263	62.0	221	7.5	16.3	0.8
CDC-April	WANDO	PI 285744	1214156	Pisum sativum	Poland, Warszawa	233	27.7	202	10.0	15.4	1.8
Carrera	TINY TIM	PI 279742	1208477	Pisum sativum	Canada, Manitoba	266	16.6	212	8.4	13.1	1.0
Kelvedon	MUSUS	PI 429839 PSP	1823060	Pisum sativum	United States, Washington	108	8.5	235	2.4	13.1	0.8
SW-Midas	NA QIAN DONG DOU	W6 44779	1902176	Pisum sativum	China, Guangxi	307	18.8	151	7.0	9.2	1.4
Austrian winterpea	NAVALE ARTTURI	PI 272148	1822930	Pisum sativum	United States, Washington	202	49.7	174	10.1	10.4	2.1
Jetset	NZ 51	PI 413685 PSP	1823056	Pisum sativum	United States, Washington	140	17.9	288	17.8	13.5	1.0
J11831	PANIA	PI 411141 PSP	1823051	Pisum sativum	United States, Washington	211	33.8	231	61.7	15.2	4.0
J12376	PATEA	PI 411142 PSP	1823052	Pisum sativum	United States, Washington	253	42.6	241	13.3	15.5	1.6
Attika	PERFECTION	PI 269822	1822920	Pisum sativum	United States, Washington	266	24.4	232	13.0	14.1	0.9
Garden-Sweet	PLP 10	PI 347281 PSP	1823022	Pisum sativum	United States, Washington	193	14.6	195	42.0	9.2	1.4
CDC-Centennial	PLP 150	PI 356979	1266363	Pisum sativum	India, Punjab	213	9.5	265	50.3	7.8	0.6
HL60	PLP 70	PI 356984 PSP	1823032	Pisum sativum	United States, Washington	123	8.9	240	36.1	13.0	0.9
Sommette	QING WAN DOU	W6 44766	1902163	Pisum sativum	China, Nei Monggol	214	11.4	200	12.3	9.9	1.3
11340	RECO	PI 601680	1187923	Pisum sativum	Switzerland	205	10.1	245	18.6	15.4	2.2
PI137118	RIL 846-07	PI 660729	1862747	Pisum sativum	United States, Washington	259	34.4	194	5.5	9.3	1.4
PI175226	RIL 847-08	PI 660730	1862748	Pisum sativum	United States, Washington	231	43.7	238	12.7	9.6	1.3
PI222071	RIL 847-22	PI 660731	1862749	Pisum sativum	United States, Washington	253	0.6	198	5.1	8.0	0.1
PI227258	RIL 847-28	PI 660732	1862750	Pisum sativum	United States, Washington	226	11.1	200	22.6	8.9	0.7