UC Davis

UC Davis Electronic Theses and Dissertations

Title

Dissecting the genotypic and environmental basis of kernel compositional traits in pistachio using near-infrared spectroscopy

Permalink

https://escholarship.org/uc/item/3fr9p73s

Author

Kane, Elton

Publication Date

2024

Supplemental Material

https://escholarship.org/uc/item/3fr9p73s#supplemental

Peer reviewed|Thesis/dissertation

By

Elton Alexander Kane

Dissecting the genotypic and environmental basis of kernel compositional traits in pistachio

using near-infrared spectroscopy

Submitted in partial satisfaction of the requirements for the degree of
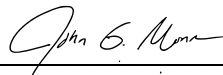
Master's of Science

in

Plant Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Co-Chair

_____

Co-Chair

_____

Committee Member

Committee in Charge

2024

By

Elton Alexander Kane

Dissecting the genotypic and environmental basis of kernel compositional traits in pistachio

using near-infrared spectroscopy

Submitted in partial satisfaction of the requirements for the degree of

Master's of Science

in

Plant Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

2024

Abstract:

Understanding the intricate interplay between genetic, environmental, and management factors shaping kernel compositional traits in pistachio (*Pistacia vera* L.) is pivotal for enhancing crop quality. This study examines variation in genetic, environmental, and management factors underlying fat, protein, and moisture content across 24 diverse pistachio cultivars grown in 7 environments, and with some grown under different management conditions: a pruning trial with three treatment levels, and an irrigation trial with four. Development of a Near-Infrared Spectroscopy (NIRS) calibration tailored to pistachio could enable rapid assessment of kernel composition in breeding programs and in post-harvest quality control. Furthermore, when integrated with multi-environment field trials, the impacts of genotype, environment, and agronomic practices on kernel quality could be elucidated. Our findings indicate that NIRS-based predictions were successful in accurately determining total fat, protein, and moisture content (with r of 0.43, 0.93, and 0.70, respectively) in ground pistachio kernels. Total mean fat percent was 45.94%, protein was 21.28%, and moisture was 3.23%. Significant differences were detected between pistachio compositional traits across cultivars and locations (within harvest years), but not in their interaction effects. The elite cultivars Golden Hills and Lost Hills recorded higher average fat content than Kerman, which recorded higher average protein than those two cultivars. Additionally, samples from the 2022 harvest year had higher average fat and moisture percent than samples from the 2023 harvest year. Through the integration of spectral analytical approaches with genetic and multi-environment trials, this research provides an understanding of the complex interactions governing pistachio kernel composition. Findings from this study offer valuable insights for breeding programs and agronomic strategies aimed at optimizing pistachio quality in diverse agricultural contexts.

Introduction:

The *Pistacia* genus is part of the Anacardiaceae (cashew) family and comprises at least 11

species with a native range spanning from tropical to subtropical throughout the dry regions of

the Mediterranean (Choulak et al. 2023, Nezami & Gallego 2023, Sheikhi et al. 2019). Native to

Central-Western Asia, pistachio was first domesticated approximately 3000 years ago in this

region. Cultivation of domesticated pistachios (*Pistachia vera* L.) spread throughout the

Mediterranean region, and by the late 19th century would later be introduced to Australia

(Choulak et al 2023, Nezami & Gallego 2023). In the early 20th century, *P. vera* L. was

introduced by American botanist William E. Whitehouse from Iran to the United States (Sheikhi

et al. 2019). Pistachio is currently a major economic crop export for semi-arid regions in Central

Asia, the Middle East, along the Mediterranean Sea, and California (Nezami & Gallego 2023). Together, the USA and Iran are the two leading producers of pistachio, comprising 82% of the world market, with Turkey as a third major producer (Nezami & Gallego 2023, Sheikhi et al. 2019). Before 2005, Iran was ranked first before droughts and subsequent water shortages reduced production, and an increase in U.S. production matched then overtook them in production rankings (Nezami & Gallego 2023, Shakerardekani et al. 2022, Sheikhi et al. 2019). In 2020, global pistachio production was estimated at approximately 1.1 million tonnes (Choulak et al. 2023) with 42% of global production being in the U.S., making it the largest producer (FAO. 2021)—primarily in California. In the U.S. industry, Kerman has been the primary cultivar. However, there has been interest in the industry to expand to more than a single cultivar to increase the overall diversity in U.S. orchards (Kallsen et al. 2009 & Sheikhi et al. 2019). Some of these cultivars of interest include Golden Hills, Lost Hills, Kahlegouchi, Aria and Gumdrop (Kallsen et al. 2009, Choulak et al. 2023, Nezami & Gallego. 2023). Cultivars Golden Hills and Lost Hills were introduced after 2002 and have since made up about 95% production in orchards over the last 10 years (Mandalari et al. 2021).

Tree nuts are the most carbon-efficient protein source of any food (Poore & Nemecek 2018). *P. vera* L. has gained increased attention for its high nutritional value and subsequent potential health benefits compared to other tree nuts, such as almonds, hazelnuts, and walnuts (Dreher 2012 & Nezami et al. 2023). In 2015, Bullo et al. conducted one of the most comprehensive investigations of pistachio nut compositional traits with the following findings: Protein comprises roughly twenty percent of pistachio weight and have been reported to have a higher ratio of essential amino acids compared to other conventionally consumed tree nuts; and pistachios possess a low to moderate carbohydrate content as with most nuts, but are rich in

soluble and insoluble forms of fiber, rivaled only by almonds. In addition, this study also compared macronutrient contents of raw and roasted almonds, hazelnuts, macadamia nuts, peanuts, pecans, walnuts, and pistachio; and found that despite pistachios being low in fat (45.4 g of total fat per 100 g of raw nuts) compared to other tree nuts they are richer in unsaturated fats. Pistachios are also rich in vitamins (A, B1-3, B5, B9, C, D, E, & K), minerals (calcium, iron, magnesium, manganese, phosphorus, potassium, & zinc), chlorophylls, and other phytochemicals (such as phenolic and carotenoid compounds) (Bullo et al. 2015 & Nezami et al. 2023). The suite of nutrients for which pistachio is nutritionally dense makes it an interesting study organism for characterizing variation in nut compositional traits and understanding the main and interaction effects of G, E, and M that underlie this variation. Cultivar genetic differences or disruption in nutrient cycling as a consequence of environmental stressors or management practices could be potential sources of variation in pistachio kernel compositional traits. Okay (2002) compared fatty acid and protein content between five domestic pistachio cultivars (Uzun, Kırmızı, Halebi, Siirt, and Ohadi) and found no significant difference in fatty acid content outside of linoleic acid, but significant difference in protein content across cultivars examined. Shakerardekani et al. (2022) found that fat, protein, and carbohydrate levels increased, and moisture levels decreased, with maturity. Another study that looked at the effect of harvest time on pistachio cultivars, similarly saw significantly different decreasing trends in moisture over increased maturity across the four cultivars they examined (Panahi & Khezri 2011). Another study, Tsantili et al (2010), examined protein and fat traits in eight pistachio varieties (Aegina, Pontikis, Bronte, Cerasola, Joley, Kerman, Mumtaz and Sirora), found fat content significantly varied most between cultivars Cerasola (67.86%) and Kerman (51.6%); and protein significantly varied most between cultivars Cersola (21.87%) and Joley (18.99%). Taken

together, these studies suggest there is variation dependent on genetic factors. There are few studies looking at pistachio nutritional variation across environments, but a study from Bellomo and Fallico (2007) examined geographic variation in anthocyanin, chlorophyll, and carotenoid content; and found anthocyanin content increased over ripening time, whereas carotenoid (lutein) and chlorophyll levels decreased, and had a greater influence on metabolite content than both cultivar and origin effects. There were also significant differences across locations: where samples from Italy (Agrigento) had significantly more anthocyanin than samples from Italy (Bronte) and Turkey; samples from both Italian locations had significantly more chlorophyll (a, b, & a+b) and carotenoid lutein than samples from Turkey. Although not primary compositional traits, these findings give insight into how environmental or management factors could impact nutritional traits in pistachio.

A genotype's performance can change across different growing environments or management practices (Cooper et al. 2022 & Langridge at al. 2021). Traditionally, studies have focused on parsing the effects of genotype and environment on priority traits, and interactions with management effects (GxE(xM). In the coming future, changes in climate bring with it new or worsening climate pressures that will challenge conventional crop growth practices forcing agricultural practices to modify in response (Mahmood et al. 2022, Smith et al. 2021, Hammer et al. 2020, & Raza et al. 2019). Examples of management effects include seedbed management, pruning, irrigation, nutrient treatments, crop protective measures, harvest practices, and post-harvest storage and distribution (Mahmood et al. 2022, Madson 1995, & Messina et al. 2020). These GxExM studies have classically focused on productivity, and GxExM studies pertaining to pistachio (or nut tree crops more generally) are scarce, with much of the literature focused on other crops such as wheat, cassava, barley, maize, and groundnut (Roberts et al. 2022, Ibrahim et

al. 2019, Hajjarpoor et al. 2021, Ciampitti et al. 2021, Ceballos et al. 2021, & San Martin et al. 2013). GxExM examinations in pistachio for productivity and/or quality traits could inform breeding and agronomic efforts. Subsequently, developing fast and accurate methods to analyze these traits also becomes critical.

Near-infrared spectroscopy (NIRS) has been used for high-throughput phenotyping (HTP) of compositional traits, particularly macronutrients, in plant (among other) materials (Alamu et al. 2020, Jasinski et al. 2016, Font et al. 2006, & Buning-Pfaue 2003). Suitable calibrations (or predictive models) must be developed for the traits and crop tissue(s) being phenotyped. Calibrations are developed by cross-analyzing NIR spectral data with wet-chemistry reference data on a subset of samples, using statistical methods such as Partial Least Squares regression (Jasinski et al. 2016 & Buning-Pfaue 2003). The developed predictive model, once validated in cross-validation and on a held-out test set, can then be applied to NIR spectral data for the larger, remaining set of samples (from the same location-year combinations) that have not been phenotyped via wet chemistry. NIRS analysis has been used in some crops such as tubers, *Brassica* species, cereal crops, and legumes to name a few (Jasinski et al. 2016, Alamu et al 2020, Font et al. 2006, Li et al. 2020, Wang et al. 2021, Tang et al. 2023). NIRS instruments record spectral data in the near-infrared portion (780-2500 nm) of the electromagnetic spectrum, the most energetic of the infrared regions (Alamu et al 2020, Agelet & Hurburgh Jr. 2010, & Osborne 2006); some NIRS instruments also include visible wavelengths (e.g., also reporting spectral data in the 400-700 nm range). Global and Neighborhood H are two performance metrics for NIRS calibrations, where Global H measures the spectral distance of a given sample to the mean of the samples in the database that was used to build the calibration, and Neighborhood H measures the spectral distance of a given sample to the next closest sample in

the database (Kaur et al 2016 & Westerhaus 2014). In case helpful for contextualization, Global H values less than 3 and a Neighbourhood H value greater than 0.6 are usually considered favorable in the NIRS realm (Kaur et al 2016). Further adaptation of a pre-existing calibration for another nut tissue (in the case of this study, a ground peanut calibration) and/or the development of a custom calibration for pistachio will be needed before the predicted trait values are considered for use in decision-making. To date, a NIRS calibration for kernel compositional traits in pistachios has not yet been published or made publicly available.

The first aim of this study was to test the adaptation of a pre-existing ground peanut NIRS calibration–and to develop a custom NIRS calibration–for use on pistachio kernels. The second aim was to deploy this calibration to assess kernel compositional traits (protein, fat, moisture, and ash; total mineral content) in pistachios, and dissect the effects of genotype, environment, and management on these traits in three contrasting sample sets: one comprising diverse cultivars and a half-sib family, and two focused on commercial cultivars of high relevance to pistachio production in California (Kerman, Golden Hills, and Lost Hills).

Materials & Methods:

Plant materials:

A total of 174 ground pistachio samples from seven locations (Campus Farm Trial, Jebs Cousin's Farm, Jebs Dad's Farm, Maricopa Orchards, Wolfskill Orchards, Tucker's Farm, and Wonderful Orchards) were analyzed in this study. Samples were provided as mature kernels with shells by three academic laboratories (one lab provided samples deshelled) and one commercial entity, with known genotype, environment, and (when applicable) pruning or irrigation treatment. These samples were collected over two harvest seasons (74 samples for harvest year

2022 and 100 samples for harvest season 2023) and from seven farms in four counties: including Yolo, Fresno, King's, and Kern County on a north-south gradient in California. Only a few of the samples represented the same trees sampled in both harvest years (ten samples). Therefore, samples were examined as three additional subsets: a 113-sample subset representing a half-sibling population (Half-sib), a 40-sample subset representing three commercially relevant elite cultivars (Golden Hills, Lost Hills, and Kerman), and a 23-sample subset representing a diverse group of cultivars. These sample sets represented diverse cultivars and a half-sib family; two commercially relevant cultivars for which three clusters were sampled from each of eight trees from two genotypes (Golden Hills & Kerman) at one location; and pruning trial samples with known cultivars from three farms (Tucker's Farm and Jebs' Dad and Cousin's Farms), and irrigation samples from a University of California-Davis on-campus trial location combining 18 trees into one sample from four different rows (See Supplemental Table 2). Samples from the diverse genotypes and half-sib family were stored at 4°C, the same temperature at which they were stored prior to receipt for use in this study, whereas the samples from the other two labs and a commercial entity were stored at room temperature.

Data collection & sample processing:

Samples were shelled (if needed), then weighed in grams and scanned as whole kernels on a FOSS DS2500 NIRS instrument using the small sample cup. This instrument provides reflectance data at each 0.5 nm step from 400 to 2500 nm. Trait values for fat, protein, and moisture percentage were predicted from the resulting spectra using a ground peanut calibration (which was the best-performing pre-existing calibration available from the instrument manufacturer in a pilot study conducted on pistachio samples). Samples were placed into coin envelopes and freeze-dried in a HarvestRight freeze dryer. After freeze-drying, the kernels were

re-weighed and scanned as whole kernels again before being ground using an IKA tube mill control with stainless steel blade and scanned for a third time. The ground samples were stored back into the coin envelopes and placed at -4°C for storage. Data from the NIRS instrument was then exported and saved in .csv or Excel format. NIRS-predicted trait values on ground samples are reported and analyzed herein, as a percentage of kernel composition on an as-is basis (i.e., not corrected for moisture content).

Collection of wet-chemistry reference values and comparison with NIRS-predicted trait values:
A total of 35 freeze dried and ground samples harvested in 2022 were randomly sampled using sample() function in R and sent for wet-lab reference analysis. Two grams of sample was weighed and stored in 50 mL tubes or Whirl-Pak sample bags for later in-house analysis and the remaining sample mass (4-12g) was stored in 50 mL tubes and sent for wet-lab reference analysis at the UC Davis Analytical Lab for total fat, protein, and moisture. After reference values were received (on an as-is basis; i.e., without correction for moisture content), a Pearson correlation was calculated and wet-chemistry reference vs. NIRS-predicted values were plotted using the ggscatter cor.method = "pearson" function to determine the accuracy of the NIRS-predicted values.

Statistical analysis:
Across the 174 total samples between the two harvest seasons, multiple subsets of samples were examined in addition to the whole sample set. These subsets included a subset of 113 half-sib samples, 40 samples representing all elite cultivars from all sources (Golden Hills, Lost Hills, and Kerman), and a subset of 23 samples representing the diverse cultivars were compiled to analyze genotypic and environmental (location and harvest year) effects on pistachio

compositional traits. Statistical analyses were conducted in RStudio using R version 4.3.2 (R Core Team 2023). Correlations and corresponding P-values were calculated using the corr.test() function from the *psych* R package. Plot matrices containing X-Y scatterplots and histograms were created with the ggpairs() function from the *ggally* R package. Mixed ANOVA was conducted using the aov() function from the *stats* package (which is included in base R installations). If a given term (genotype, treatment, or genotype-by-treatment interaction) was significant in mixed ANOVA (at $\alpha = 0.05$) for a given trait, Tukey's Honestly Significant Difference was conducted as a post-hoc comparison of means using the HSD.test() function from the *agricolae* R package. Model term harvest year was assigned to all sample subsets (half-sib, elite cultivars, & diverse cultivars), however only the elite cultivar sample subset had the location and cultivar model terms included in addition. If an interaction effect was significant, we only conducted Tukey's HSD for that interaction, and did not conduct Tukey's HSD for either of the main effects (so as to not interpret main effects in the presence of a significant interaction). Principal component analysis was conducted using the prcomp() function from the *stats* package, with scale set to TRUE. Last, box plots were made using the geom_plot() function from the *ggplot2* package.

<u>Results:</u>

<u>Evaluating NIRS (comparison between NIRS and wet chemistry):</u>

Pistachio NIRS-predicted values were determined using a pre-existing ground peanut calibration as peanut was closest in compositional make up to pistachio than all other existing NIRS calibration models. A significant positive correlation ($r = $ **0.43**, **p = 0.0097**) was detected between the NIRS-predicted and wet-chemistry (observed) values for fat percent (Figure 1a).

Protein percent had the most significant positive correlation (*r* = **0.93**, p = **6.6e-16** respectively) between NIRS-predicted traits and wet-chemistry (observed) values (Figure 1b). Last, moisture percent had the second highest positive correlation between NIRS- predicted and wet-chemistry (observed) values (r = **0.70**, p = **2.9e-06**) (Figure 1c).
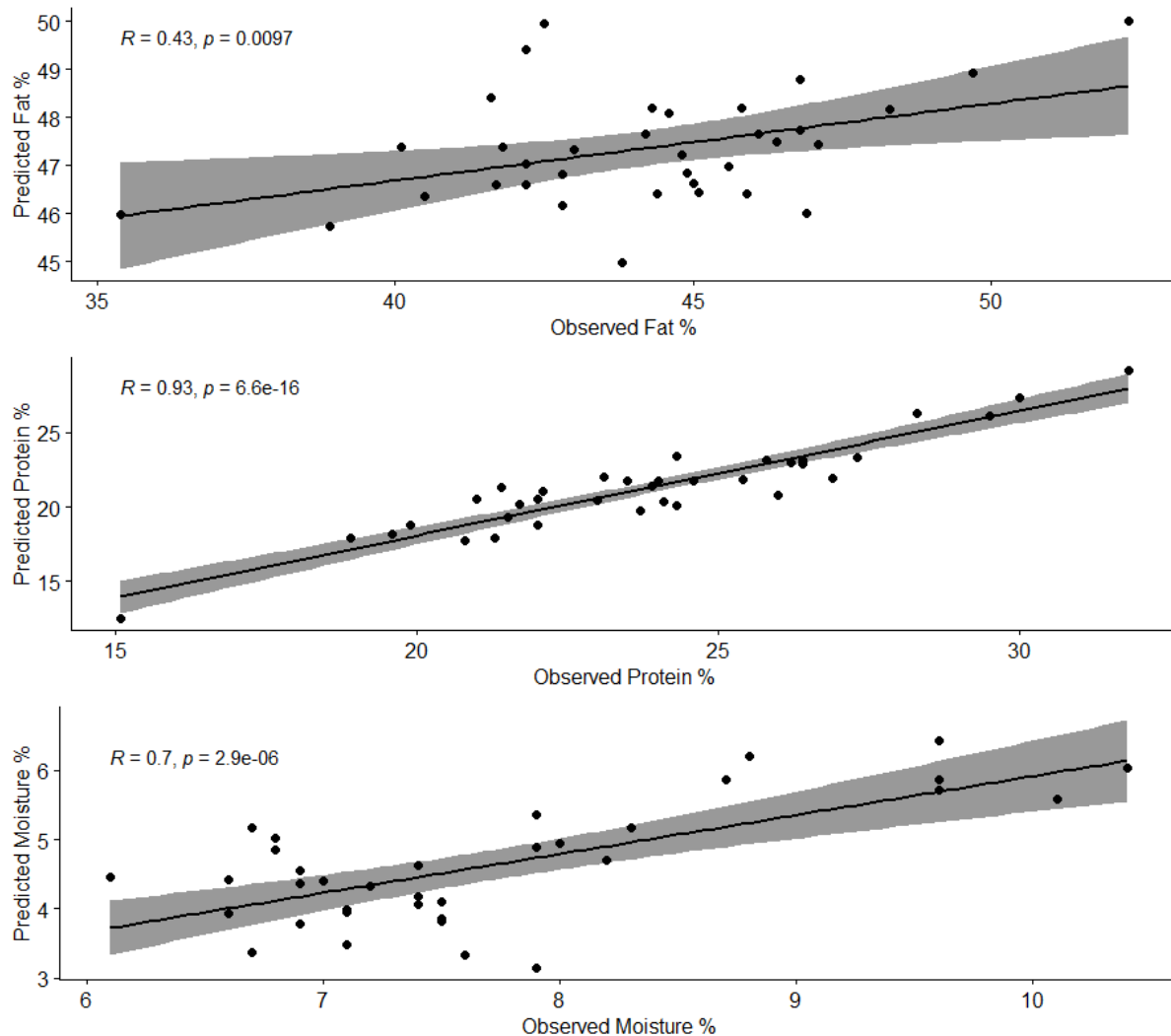

Figure 1: Pearson correlation between wet-chemistry reference analyses (observed values) vs. NIRS-predicted values for a) fat, b) protein, and c) moisture percent (n = 35).

Assessing differences across environments (Location x Harvest Years):

PCA results across harvest year for all 174 samples indicated PC1 explained 54.9% of the variance and PC2 explained 38.7% of the variance (Figure 2). Two distinct groups formed for harvest year with minor overlap and there was partially distinct clustering by the nine locations. PC1 negatively corresponded to moisture and fat percent and PC2 positively corresponded to protein percent (Figure 2). Location PCA results were examined first across orchard or farm locations, and results showed distinct clusters forming for each of the seven locations. However, there is overlap between all clustered except those from the Maricopa Orchards and Campus Trial samples (Figure 3a). A second location PCA across counties shows clusters form for all four of the represented counties, but only Fresno County was distinct and showed no overlap and are significantly associated with higher moisture percent (Figure 3b).

Correlation analysis for all 174 samples indicated significant positive correlations between fat and moisture percent (r = **0.647**) and protein and moisture percent (r = **0.269**) (Figure 4a). Global and Neighbourhood H performance metrics were strongly positively correlated with each other both across (r = **0.974**) and within locations (Supplemental Figure 2a). Correlation analysis for two harvest seasons (2022 & 2023) indicated a significant positive correlation between fat and moisture percent (r = **0.647**) represented by samples from harvest season 2023. Protein and moisture percent had a weakly significant positive correlation (r = **0.268**), but protein and fat percent were weakly negatively correlated (r = **-0.241**) (Figure 4b).

The main effects of location and harvest year were significant for both mean fat percent (location: p = **1.5e-13,** harvest year: p = **3.24e-14**) and mean moisture percent (location & harvest year p = **< 2e-16**) (Figure 5 & Supplemental Table 2a&c). The main effect of location for mean protein percent was significant (p = **0.000269**), and the interaction effect of location and

harvest year for mean protein percent was also significant (p = **0.020705**) (Figure 5 & Supplemental Table 2b). Samples from Jeb Dad's Farm location in the 2022 harvest set had the highest fat percent (48.01%), samples from the Campus Farm location from the 2023 harvest year had the lowest mean fat percent (41.36%), and samples from these two locations are most significantly different from each other (Figure 5 & Supplemental Table 2a). Samples from Tucker's Farm in harvest year 2022 had the highest mean protein percent (24.41%), and samples from Tucker's Farm of the same year recorded the lowest mean protein percent (19.84%) (Figure 5 & Supplemental Table 2b). Samples from the 2022 harvest year overall had significantly higher average fat and moisture percent than samples from the 2023 harvest year (Figure 5 & Supplemental Table 2a&c). Samples from Maricopa Orchards recorded the highest moisture % (5.76%) and were most significantly different from the Campus Farm location which recorded the lowest percentage (0.397%) (Figure 5 & Supplemental Table 2c). However, it should be noted samples from Maricopa Orchards experienced a 2.3 ± 0.2% loss (mean ± SD) in weight before vs. after freeze-drying (measured on the same day as the samples were taken out of the freeze dryer) vs. a 2.0 ± 0.2% (mean ± SD) re-increase in weight from the day they were taken out of the freeze dryer to the next day.
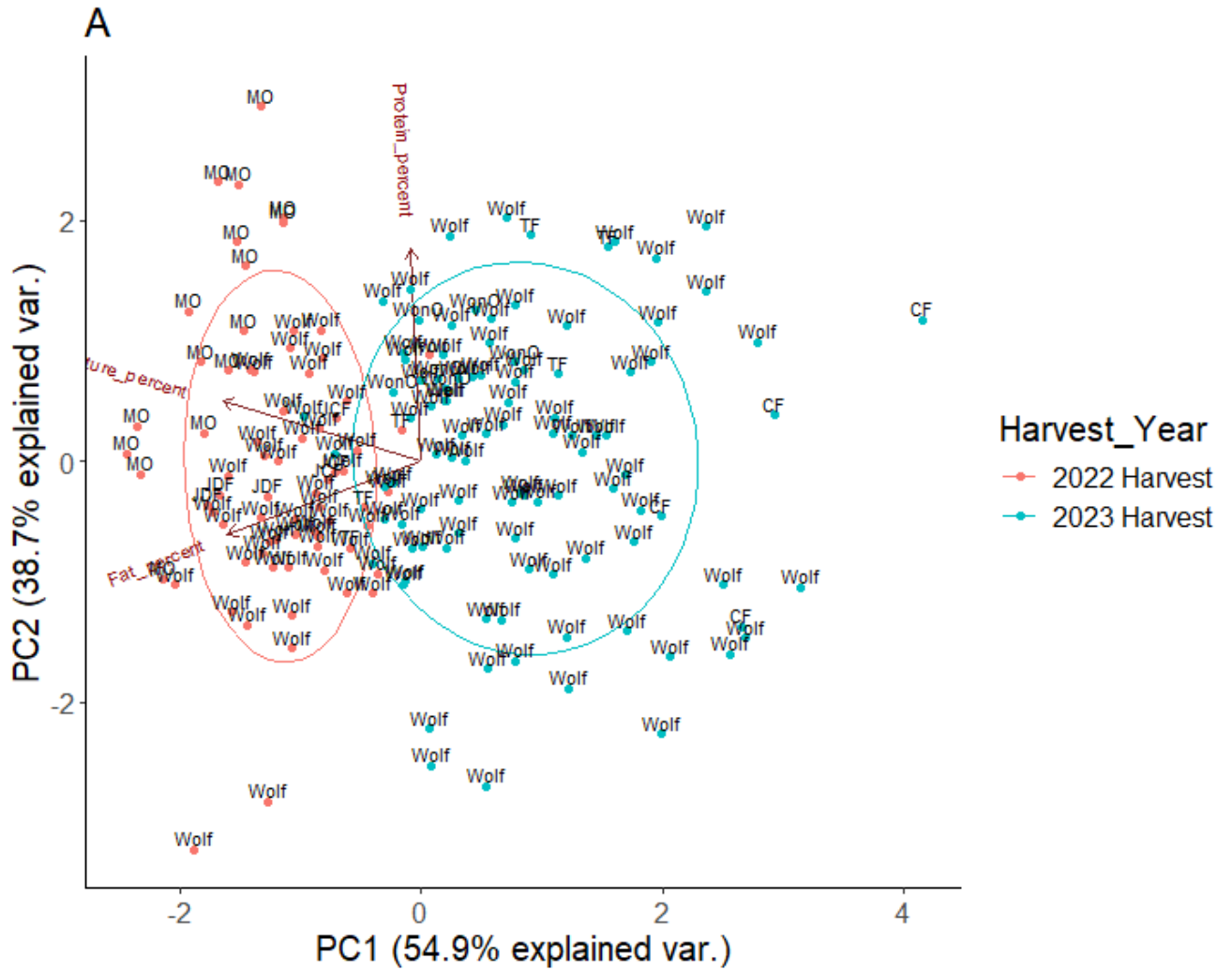
Figure 2: Principal component analysis biplot (depicting PC1 vs. PC2) for NIRS-predicted traits, color-coded by harvest year with points labeled by locations: Campus Farm (CF), Jebs Cousin's Farm (JCF), Jebs Dad's Farm (JDF), Maricopa Orchards (MO), Wolfskill Orchards (Wolf), Tucker's Farm (TF), Wonderful Orchards (WonO). N = 174.
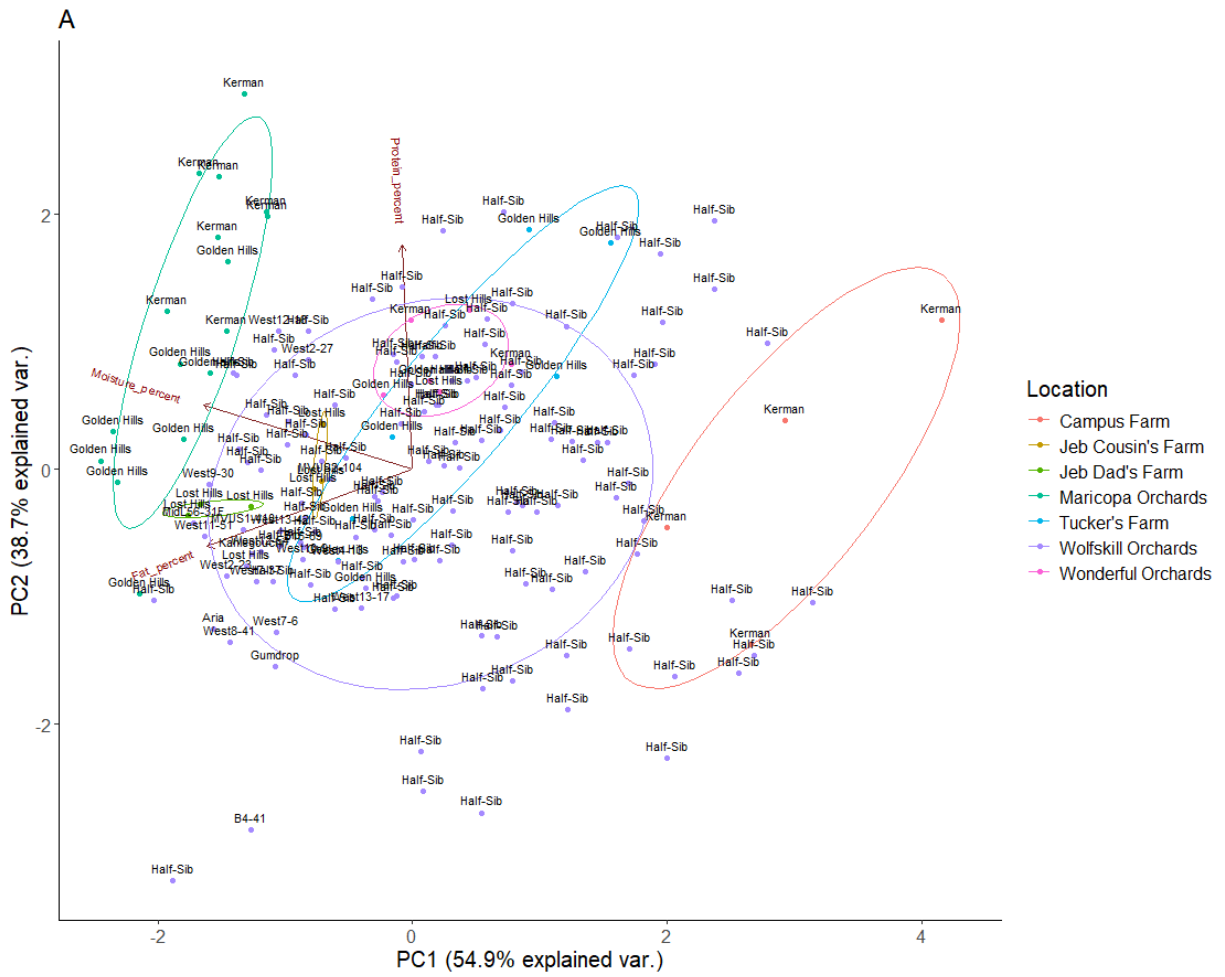
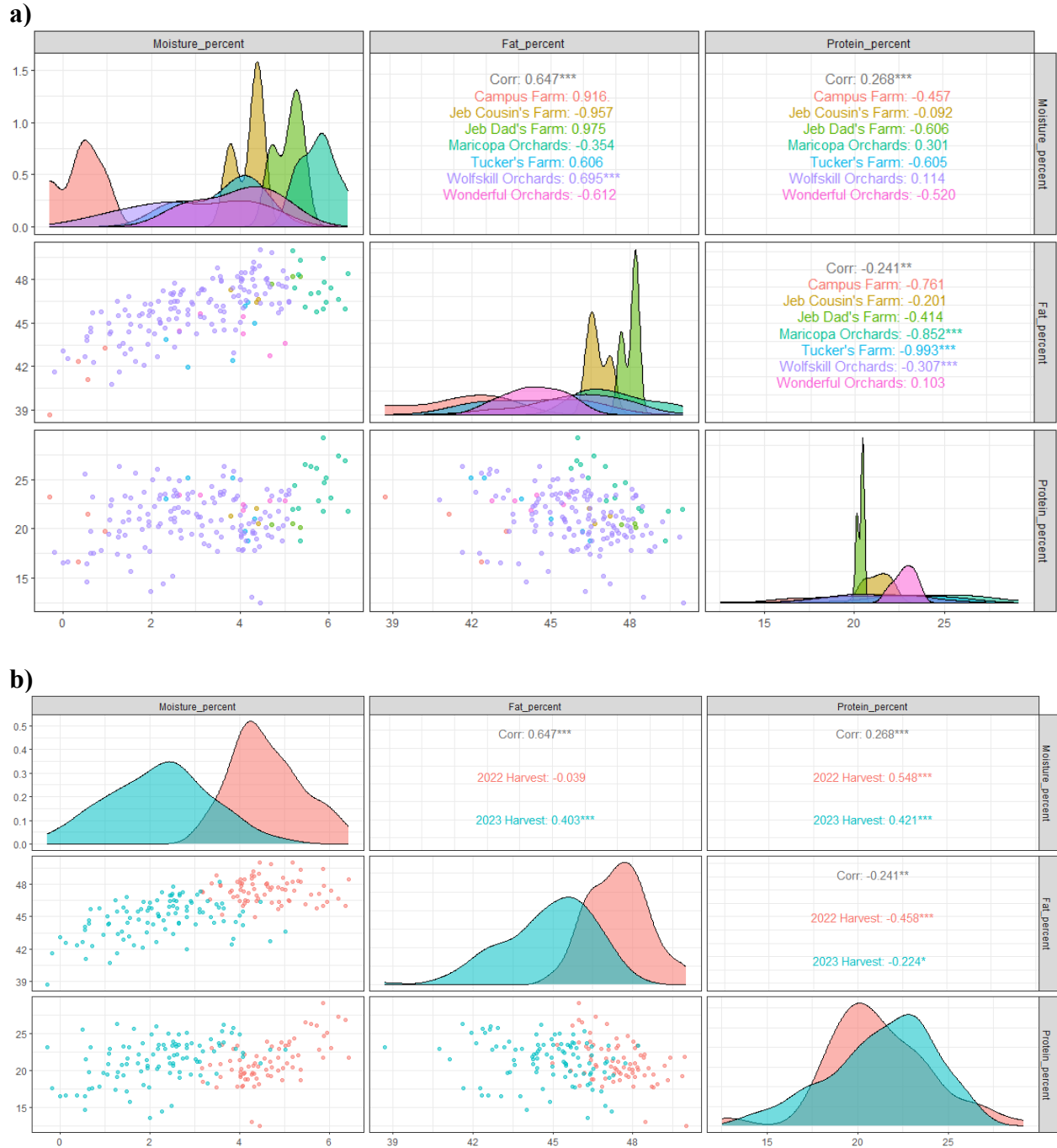**Figure 3:** PCA of NIRS-predicted traits, color-coded by location (Farm/Orchard). N = 174.

Figure 4: Correlation plot matrix of histograms (on the diagonal) and X-Y scatter plots depicting trait relationships (in the lower triangle) with corresponding Pearson r-values (in the upper triangle) for a) NIRS-predicted traits by location and b) plot matrix of histograms by harvest year. N = 174.

Figure 5: Boxplots of compositional traits by harvest year across location (from left to right: Campus Farm, Jebs Cousin's Farm, Jebs Dad's Farm, Maricopa Orchards, Wolfskill Orchards, Tucker's Farm, Wonderful Orchards). Gray data points that appear horizontally adjacent to a given colored data point indicate that that point was an outlier (more than 1.5 * interquartile range beyond its nearest—25% (for outliers on low side of distribution) or 75% (for outliers on high side of distribution)—quartile). N = 174.

Differences across harvest year for Half-Sib Samples:

Samples for this set of analyses focus on the half-sib family and diverse cultivars, as this group had the greatest number of genotypes (113). PCA results for the half-sib family show that PC1 explained 55.1% of the variance and PC2 explained 38.3% of the variance (Figure 6). Two

distinct clusters formed by harvest year, similar in shape and size as the clusters seen in Figure 2. PC1 negatively corresponded to fat and moisture percent and PC2 positively corresponded to protein percent (Figure 6).

Correlation results for the half-sib family indicated a significant positive correlation between fat and moisture percent between harvest seasons (r = **0.652**) (Supplemental Figure 4). Protein and moisture percent were weakly correlated (r = **0.273**) and protein and fat percent were weakly negatively correlated (r = **-0.222**) (Supplemental Figure 4).

The main effect of harvest year was significantly different for fat percent (p = **5.84e-11**) in 2022 (47.45%) than 2023 (45.09%) for the half-sib samples (Figure 7 & Supplemental Table 6a). The main effect of harvest year was significantly different for moisture percent (p = **<2e-16**), with year 2022 having nearly twice (4.18%) the moisture percent of year 2023 (2.14%) (Figure 7 & Supplemental Table 6c). However, protein percent showed no significant difference between the two harvest years (Figure 7 & Supplemental Table 6b).
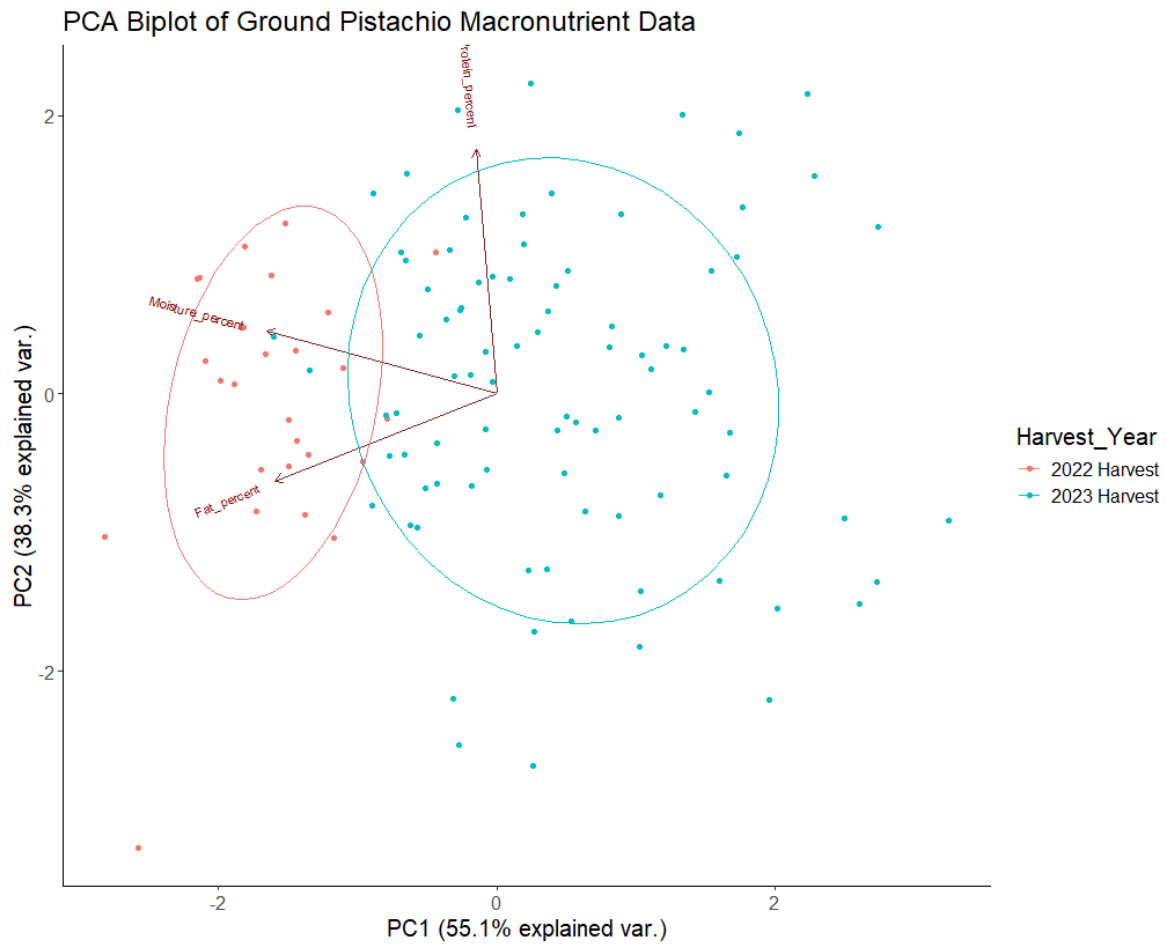
Figure 6: Principal component analysis biplot (depicting PC1 vs. PC2) for NIRS-predicted traits in the half-sib family, color-coded by harvest year. N = 113.
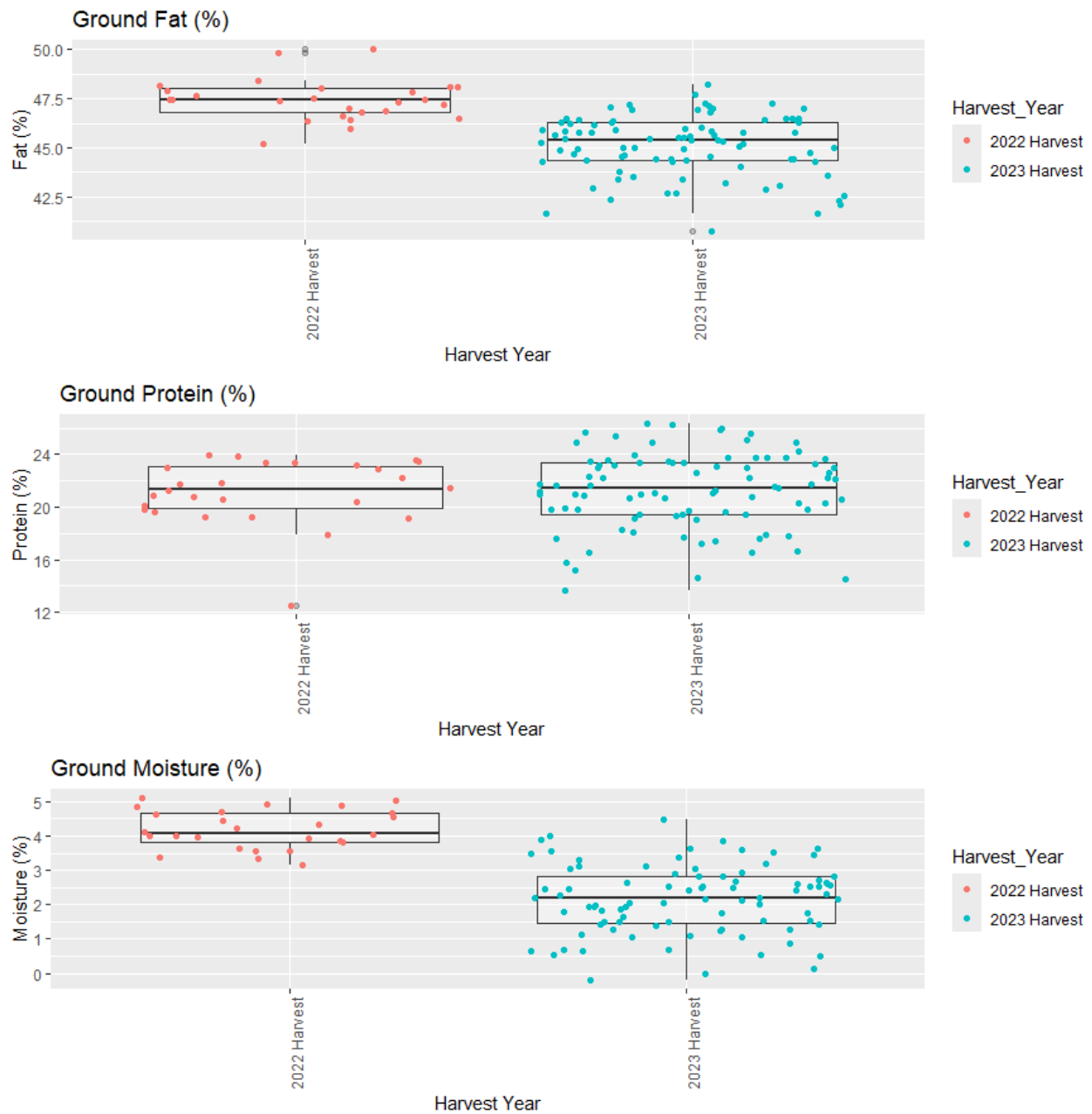
\

Figure 7: Boxplot for NIRS-predicted traits in the half-sib family, color-coded by harvest year. N = 113.

Results for this next subset focus on 40 samples representing elite cultivars: Golden Hills, Lost Hills and Kerman. PCA results showed two distinct clusters by harvest year (Figure 8). PC1 explained 60.8% of the variance, and PC2 explained 35.5% of the variance. Fat and moisture percent positively corresponded to PC1, and protein percent positively corresponded to PC2 (Figure 8). PCA results for location showed 6 distinct clusters for each location. It should be noted there are only two samples from Wolfskill Orchard, but they are grouped near each other.

Correlation analysis revealed a significant positive correlation between fat and moisture percent (r = **0.774**) and protein and moisture percent (r = **0.359**). The correlation between fat and protein percent was negligible (r = -0.085) (Supplemental Figure 7a).

Within this 40-sample subset of elite cultivars, the main effects of cultivar and harvest year were significant for fat percent (p = **0.00317** and **1.69e-09,** respectively). Post hoc analysis reveals, samples from harvest year 2022 had significantly higher fat percent than samples from 2023, with each cultivar recording significantly less fat % than the year prior (Figure 9a & Supplemental Table 8a). Only the main effect of cultivar was significantly different for protein percent (p = **0.000358**), but the interaction effect was also significant across cultivars and harvest years (p = **5.72e-06**). Kerman 2022 had significantly higher protein percent than the other two cultivars in 2022 and Kerman in 2023. Golden Hills in 2023 was significantly higher than Lost Hills in 2022 (Figure 9b & Supplemental Table 8b). Main effects for harvest year were significantly different (p = **4.7e-09**) whereas the main effect of cultivar was not, but their interaction effect was significant (p = **0.000394**) (Figure 9c & Supplemental Table 8c). Kerman

and Golden Hills in 2022 were both significantly higher than Kerman and Golden Hills in 2023, and Kerman 2023 was significantly lower than Lost Hills 2022 and 2023 (Figure 9c & Supplemental Table 8c). When examining variance in fat, protein, and moisture percent for the 40-sample subset of elite cultivars across location we see the main effects of cultivar and location are significantly different (fat: p = **0.0012**, p = **6.18e-08**; protein: p = **0.000637**, p = **0.000531**; moisture: p = **0.00984**, p = **2.09e-13**, respectively), however interaction effects for all three traits weren't significant (Figure 11 & Supplemental Table 10a,b,c). Post hoc analysis indicates cultivar Golden Hills from Maricopa Orchards recorded the highest fat % (48.28%) out of all the elite cultivars, just barely beating Lost Hills from Jeb's Dad's farm (48.01%) as they were not significantly different from one another, but they are significantly greater than Golden Hills from Tucker's Farm (44.31%), Kerman from Wonderful Orchards, Lost Hills from Wonderful Orchards, and Kerman from the Campus Trial (41.36%)-which recorded the lowest of all the elite cultivars (Figure 11 & Supplemental Table 10a).
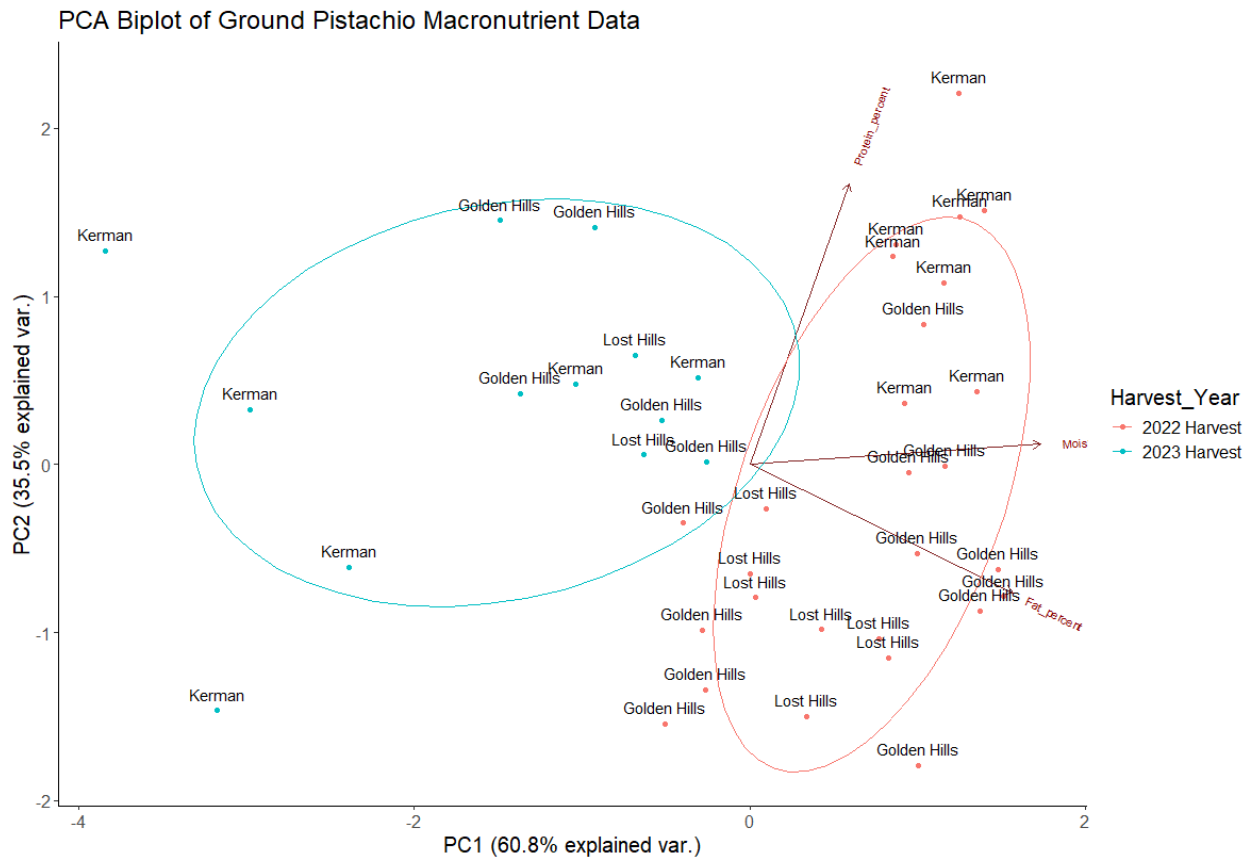
Figure 8: PCA of compositional traits for elite cultivars per harvest year and b) location. N = 40.
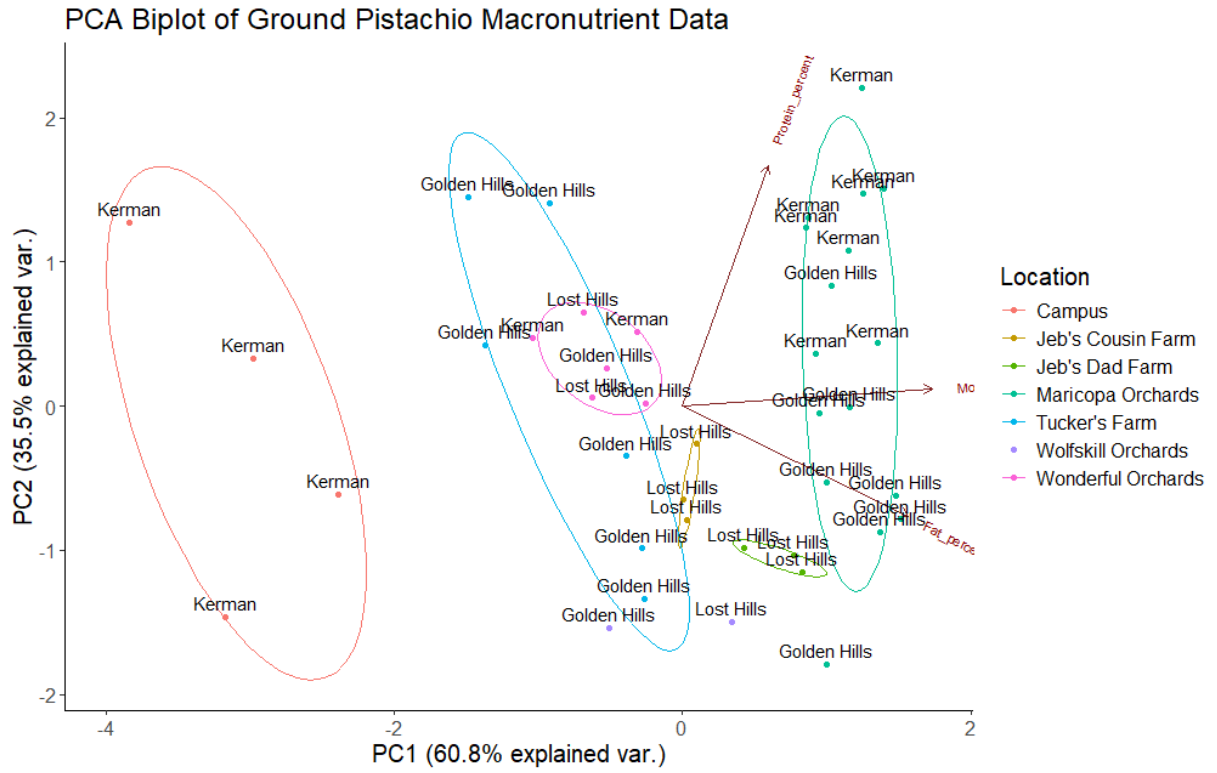
Figure 9: PCA of compositional traits for elite cultivars per location. N = 40.

Figure 10: Boxplots of compositional traits: a) fat %, b) protein %, and c) moisture %, for three elite cultivars (Golden Hills, Lost Hills, and Kerman) by harvest year. N = 40
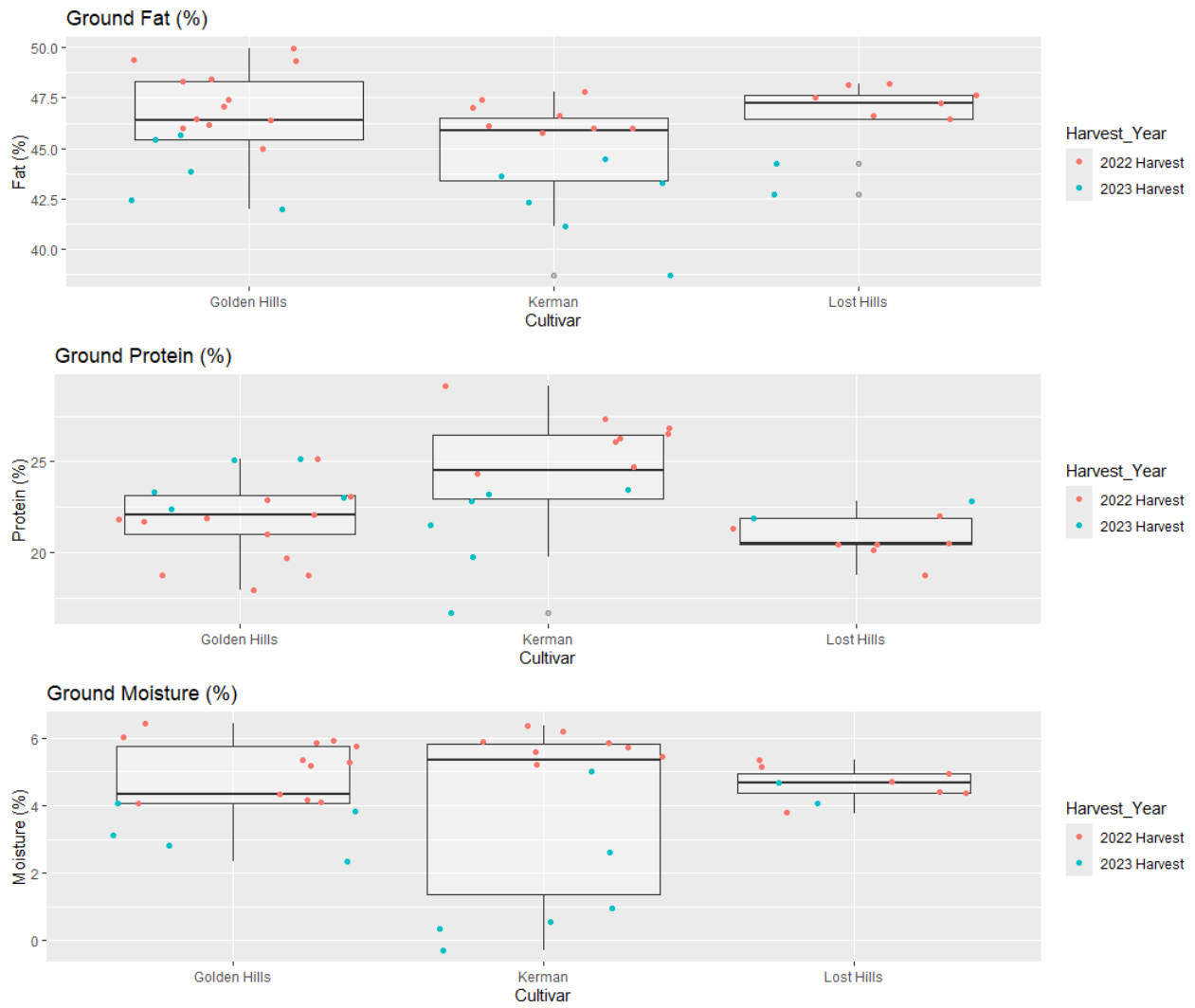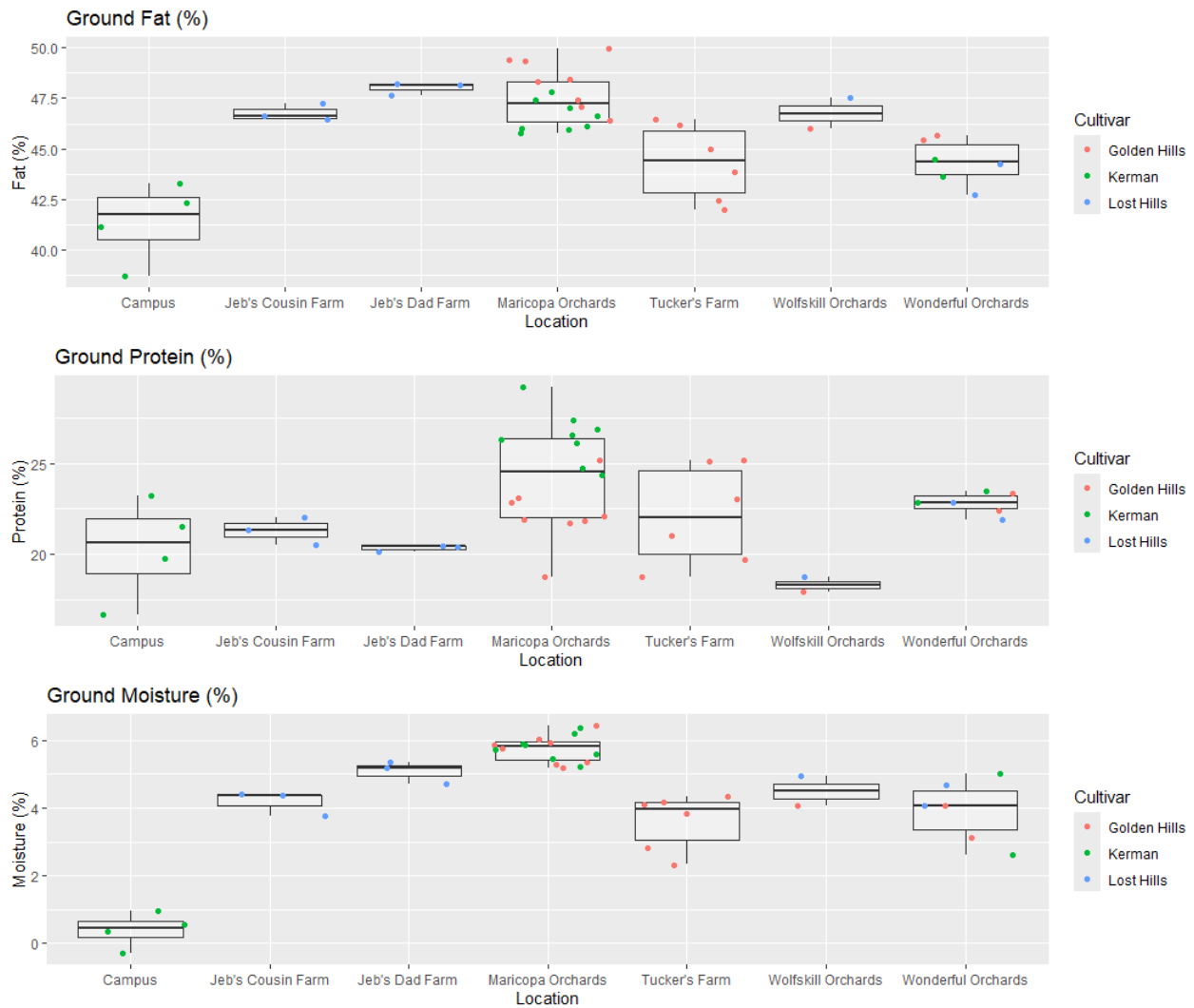
Figure 11: Boxplots of compositional traits: a) fat %, b) protein %, and c) moisture %, for three elite cultivars (Golden Hills, Lost Hills, and Kerman) by location. N = 40

Differences Across 2022 Harvest Year for Diverse Cultivars Sample Subset:

The next set of results are for a 23 diverse cultivar subset (the second largest cultivar sample set) from harvest year 2022 with one sample per cultivar. Cultivar West11-15 recorded the highest fat percent (49.2%), and West12-16 recorded the highest protein and moisture percent (23.92% and 5.1%, respectively). In contrast, MVUS2-104 recorded the lowest fat percent (45.68%), B4-41 recorded the lowest protein percent (13.11%), and West13-17 recorded the lowest moisture percent (3.18%) (Figure 12 & Supplemental Table 11).
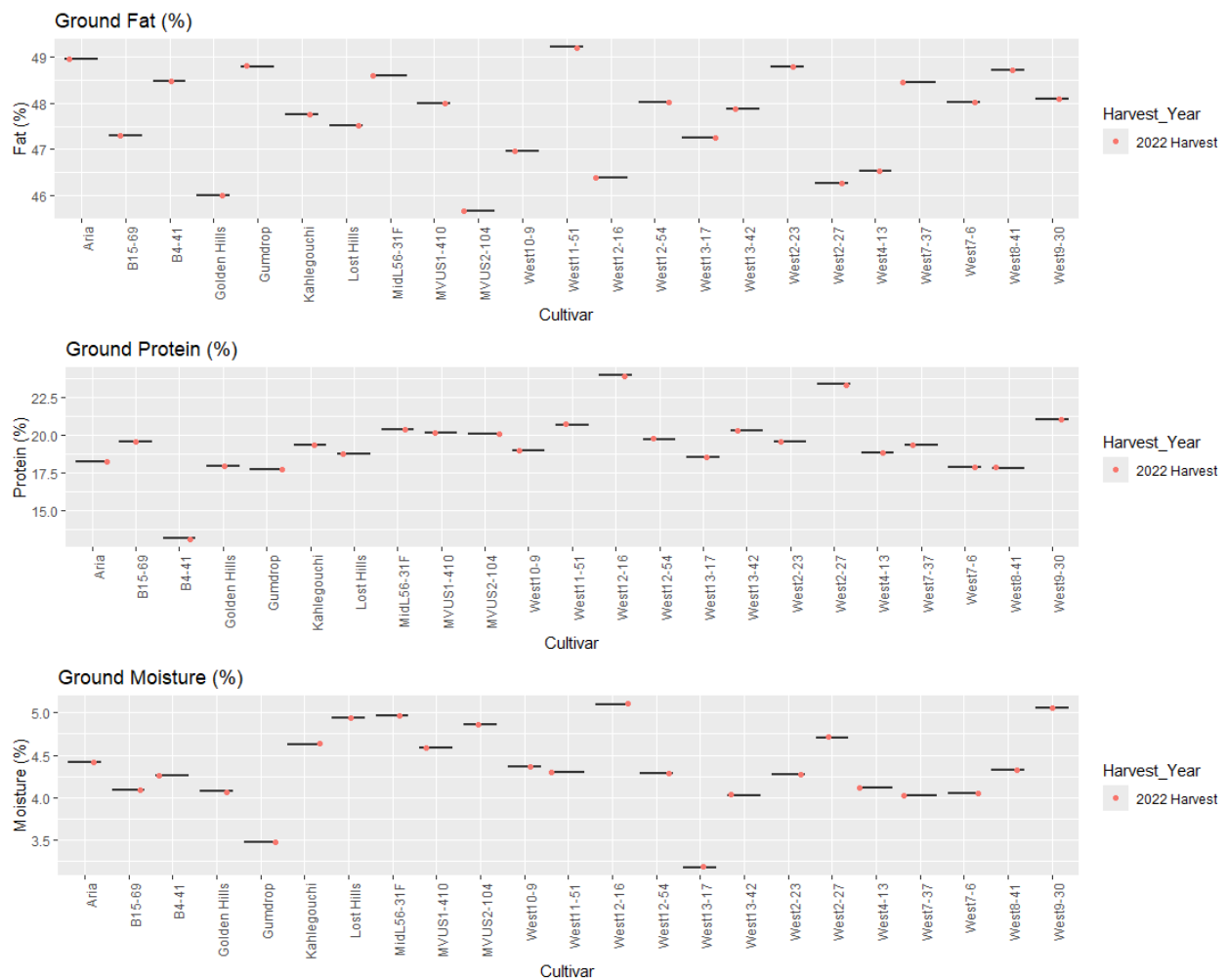
Figure 12: Boxplots of compositional traits for the diverse cultivars subset for harvest season 2022. N = 23

Discussion:

The overall goal of this study was to develop the first NIRS calibrations for kernel compositional traits in whole and ground pistachio and assess the main and interaction effects of genotype, environment, and management on these traits. While increased yield remains a foremost priority trait (Smith et al. 2021), improvement of kernel compositional traits is important in improving the nutritional value of our foods in the face of problems such as low nutrient bioavailability,

changes in rain patterns, and soil fertility, brought about by climate change and growing population pressures (Owino 2022). Developing the capacity to identify cultivars that can produce nutrient rich nuts in dry nutrient depleting soils are some important traits for adaptation to climate changes. However, before that comes developing the tools to differentiate nutritional trait variation in crops to identify which cultivars are richer than others in one or more compositional traits. We believe NIRS can be a useful tool for high-throughput characterization of pistachio traits, through the development of a spectral calibration for pistachio based on peanut, the closest crop in compositional make-up with a NIRS calibration. Afterwards, we conducted various analytical tests to determine if genetic or environmental main effects, interaction effects, or a combination were responsible for any significant variation in compositional traits detected.

Total mean NIRS-predicted values and wet-chemistry reference values for fat, protein, and moisture percent recorded in this study (Fat%: 45.94%, Protein%: 21.28%, and 3.23%), are comparable to trait values reported in prior studies (Supplemental Figure 12, Nezami et al. 2023, Mateos et al. 2022, Mandalari et al. 2021, Bullo et al. 2015). Samples from only the Maricopa Orchards experienced moisture re-absorption from an error during sample processing and thus need to be re-examined. Examining samples from the randomly selected subset group, correlations between NIRS-predicted and wet-lab reference values were shown to be significantly correlated for all compositional traits measured. Total protein had the highest NIRS-prediction accuracy followed by total moisture and fat percent. Our NIRS-based predictions were successful in accurately determining total fat, protein, and moisture content (fat % $r = 0.43$, protein % $r = 0.93$, and moisture % $r = 0.70$) in the pistachio nut samples analyzed

herein as determined by the significantly high correlation between our NIRS-predicted values and the analytical lab's wet-chemistry reference values (Figure 1).

Due to unforeseen challenges (low yield due to a combination of pests and alternate bearing), in addition to analyzing the complete 174-sample set, additional sample subsets were created to examine specific environmental effects. These include a 113-half sib subset, a 40-elite cultivar subset, and a 23-diverse cultivar subset. It should be noted, the sample subset number represents the total sample number in the set, not the number of cultivars. There weren't enough samples from both harvest years to analyze management effects, thus limiting us to examining genotype by environment effects to a few sample sets. When examining samples from the total 174 samples set, samples from harvest year 2022 recorded significantly higher total mean fat and moisture percentage than harvest year 2023 (Figure 5a & Table 2a&c). We see this trend continue with all subsequent sample subsets examined. There were significant differences in composition across cultivars suggesting potential genetic factors responsible for these differences (Figure 10 & 11). Amongst the elite cultivars examined, Kerman recorded significantly greater protein and moisture content than the other two genotypes (Protein: 26.41%, Moisture: 5.78%), and Lost Hills recorded the least (Protein: 20.52%, Moisture: 4.67%), however cultivars Golden and Lost Hills recorded significantly higher fat content (47.49%, 47.40%) than Kerman (46.58%), though not from one another (Figure 11, Supplemental Table 10a,b,c, Supplemental Table 11). If this distinction proves to be consistent, this could give us insight into knowing which cultivars are better for certain culinary or agro-economic purposes and could lead to potential experimental crosses between protein rich and fat rich cultivar. There were also significant differences in compositional traits due to environmental factors. For both the 174 and 40-sample sets, distinct clustering formed for environmental effects of location and harvest year

(Figures: 2, 3, 8, & 9). Taken together, traits controlling fat content may be more influenced by a combination of genetic and environmental factors, whereas protein is influenced primarily by genetic factors.

The development of this custom pistachio calibration will be finalized using Partial Least Squares regression after wet-lab reference values are received for the 2023 harvest year samples. Then further adaptation of the existing ground pistachio calibration and inclusion of a whole nut calibration can proceed for use in decision-making of pistachio breeding programs. Once the compositional traits data is acquired, further examination of main and interaction effects of GxExM can be assessed. Using a peanut calibration as the base calibration framework to build off from, we were successful in developing an initial predictive model for pistachio. Although all three compositional traits measured (total fat, protein, and moisture) recorded significant correlations between NIRS spectral data and wet-chemistry data, total protein percent had the most significant correlation between the three. Despite unforeseen challenges, alternate bearing and pest issues, we were successful in the initial development of the NIRS pistachio calibration, but the final PLS regression remains to be completed. Due to these limitations, we were further limited in cultivar collection as we were only able to secure sufficient sample size for three of the four commercially relevant cultivars of initial interest. The cultivar Gumdrop was also to be included, but was largely left out of the analysis due to collection of only a single sample. The perennial nature of pistachios (and time from planting to production of nuts) meant that samples for this study needed to be taken from existing nut-bearing orchards, which limited which samples were available. Preliminary data of GxExM effects on pistachio compositional traits suggest some influence from all three effects, however, additional samples are needed to make a more thorough assessment of GxExM effects on the compositional traits of pistachio nuts. It

should be noted that the high moisture content seen in samples from Maricopa Orchards could have been due to those samples having been scanned one day after they were removed from the freeze dryer rather than the same day (and having been stored at room temperature), which resulted in moisture reabsorption after freeze drying.

Future directions for this work could include the incorporation of NIRS with other compositional trait analytical tools such as gas chromatography-mass spectral analysis (GC-MS), liquid chromatography-mass spectral analysis (LC-MS), and high-performance liquid chromatography (HPLC) techniques for more detailed characterization of compositional traits, which has already been successfully integrated with similar techniques such as MIR (Mid-infrared) spectroscopy and NMR (nuclear magnetic resonance) (Ozaki 2021, Belleon-Maurel & McBratney 2011, & Fang et al 2013). There is also potential application of NIRS and GxExM analysis in combination with machine learning genomics-assisted breeding tools such as genomic selection (GS) or phenomic selection (PS), to more efficiently identify ideal traits in crops. This study would also benefit from additional sample sizes for all the diverse cultivars, elite cultivars, and locations sampled to verify the initial trends and results analyzed. Lastly, two immediate next steps for this project would be to scale up the study to include more samples from each location, cultivar, and management treatment presently examined in this study; and the other next step would be to incorporate in-house analysis for total phenolic, carotenoid, and chlorophyll content in pistachio as a downstream step. A study conducted by Mandalari et al. (2021), indicated pistachio seed coat is richest in flavonoid content, but the seed + coat has recorded total phenolic content over 1600 mg GAE. Further work can corroborate to determine how phytochemicals vary across genotypes, environments, and management effects.

References:

1. R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

2. Choulak, S., Chatti, K. & Rhouma, S. Recent advances in genomics, conservation, and breeding of pistachio. *Tree Genetics & Genomes* 19, 40 (2023). https://doi.org/10.1007/s11295-023-01615-9

3. Nezami, E. & Gallego, P.P. History, Phylogeny, Biodiversity, and New Computer-Based Tools for Efficient Micropropagation and Conservation of Pistachio (Pistacia spp.) Germplasm. Plants 2023, 12, 323. https://doi.org/10.3390/ plants12020323

4. Sheikhi, Abdollatif, et al. "Pistachio (Pistacia spp.) breeding." *Advances in Plant Breeding Strategies: Nut and Beverage Crops*, 2019, pp. 353–400, https://doi.org/10.1007/978-3-030-23112-5_10.

5. Shakerardekani, Ahmad, et al. "A Review of Pistachio Kernel Composition during Fruit Ripening." *Journal of Nuts*, vol. 13, no. 4, 11 Oct. 2022. *301-311*, https://doi.org/DOI: 10.22034/jon.2022.1963848.1189.

6. Kallsen, Craig E., et al. "New pistachio varieties show promise for California cultivation." *California Agriculture*, vol. 63, no. 1, May 2008, pp. 18–23, https://doi.org/10.3733/ca.v063n01p18.

7. Mahmood, Tariq, et al. "Genotype X environment X management (GEM) reciprocity and crop productivity." *Frontiers in Agronomy*, vol. 4, 28 June 2022, https://doi.org/10.3389/fagro.2022.800365.

8. Smith, Daniel T., et al. "Scaling up high-throughput phenotyping for abiotic stress selection in the field." *Theoretical and Applied Genetics*, vol. 134, no. 6, June 2021, pp. 1845–1866, https://doi.org/10.1007/s00122-021-03864-5.

9. Hammer, Graeme. L., et al. "Designing crops for adaptation to the drought and high-temperature risks anticipated in future climates." *Crop Science*, vol. 60, no. 2, Mar. 2020, pp. 605–621, https://doi.org/10.1002/csc2.20110.

10.   Raza, Ali, et al. "Impact of climate change on crops adaptation and strategies to tackle its outcome: A Review." *Plants*, vol. 8, no. 2, 30 Jan. 2019, p. 34, https://doi.org/10.3390/plants8020034.

11.   Cooper, Mark, and Carlos D Messina. "Breeding crops for drought-affected environments and improved climate resilience." *The Plant Cell*, vol. 35, no. 1, 12 Nov. 2022, pp. 162–186, https://doi.org/10.1093/plcell/koac321.

12.   Langridge, Peter, et al. "Breeding crops for climate resilience." *Theoretical and Applied Genetics*, vol. 134, no. 6, 27 May 2021, pp. 1607–1611, https://doi.org/10.1007/s00122-021-03854-7.

13.   Madsen, Eugene L. "Impacts of agricultural practices on subsurface microbial ecology." *Advances in Agronomy*, 1995, pp. 1–67, https://doi.org/10.1016/s0065-2113(08)60897-4.

14.   Messina, Carlos D., et al. "Crop science: A foundation for advancing predictive agriculture." *Crop Science*, vol. 60, no. 2, Mar. 2020, pp. 544–546, https://doi.org/10.1002/csc2.20116.

15.   Robert, Pauline, et al. "Phenomic selection in wheat breeding: Prediction of the genotype-by-environment interaction in multi-environment breeding trials." *Theoretical and Applied Genetics*, vol. 135, no. 10, 8 Aug. 2022, pp. 3337–3356, https://doi.org/10.1007/s00122-022-04170-4.

16.   Ibrahim, Ahmed, et al. "Examining the yield potential of barley near-isogenic lines using a genotype by environment by management analysis." *European Journal of Agronomy*, vol. 105, Apr. 2019, pp. 41–51, https://doi.org/10.1016/j.eja.2019.02.003.

17.   Hajjarpoor, Amir, et al. "Environmental characterization and yield gap analysis to tackle genotype-by-environment-by-management interactions and map region-specific agronomic and breeding targets in groundnut." *Field Crops Research*, vol. 267, June 2021, p. 108160, https://doi.org/10.1016/j.fcr.2021.108160.

18.   Ciampitti, Ignacio A., et al. "Does the critical n dilution curve for maize crop vary across genotype x environment x management scenarios? - A bayesian analysis." *European Journal of Agronomy*, vol. 123, Feb. 2021, p. 126202, https://doi.org/10.1016/j.eja.2020.126202.

19.   Ceballos, Hernán, et al. "Fifty years of a public cassava breeding program: Evolution of breeding objectives, methods, and decision-making processes." *Theoretical and Applied Genetics*, vol. 134, no. 8, 4 June 2021, pp. 2335–2353, https://doi.org/10.1007/s00122-021-03852-9.

20.   Martín, Manuel Montesino-San, et al. "A genotype, Environment and Management (GxExM) analysis of adaptation in winter wheat to climate change in Denmark." *Agricultural and Forest Meteorology*, vol. 187, Apr. 2014, pp. 1–13, https://doi.org/10.1016/j.agrformet.2013.11.009.

21.   Dreher, Mark L. "Pistachio nuts: Composition and potential health benefits." *Nutrition Reviews*, vol. 70, no. 4, 27 Mar. 2012, pp. 234–240, https://doi.org/10.1111/j.1753-4887.2011.00467.x.

22.   Bulló, M., et al. "Nutrition attributes and health effects of Pistachio Nuts." *British Journal of Nutrition*, vol. 113, no. S2, Apr. 2015, https://doi.org/10.1017/s0007114514003250.

23.   Ojeda-Amador, Rosa M., et al. "Composition and properties of Virgin Pistachio oils and their by-products from different cultivars." *Food Chemistry*, vol. 240, Feb. 2018, pp. 123–130, https://doi.org/10.1016/j.foodchem.2017.07.087.

24.   Alamu, Emmanuel Oladeji, et al. "Near-infrared spectroscopy applications for high-throughput phenotyping for cassava and Yam: A Review." *International Journal of Food Science & Technology*, vol. 56, no. 3, 5 Sept. 2020, pp. 1491–1501, https://doi.org/10.1111/ijfs.14773.

25.   Jasinski, Sophie, et al. "Arabidopsis seed content qtl mapping using high-throughput phenotyping: The assets of Near Infrared Spectroscopy." *Frontiers in Plant Science*, vol. 7, 10 Nov. 2016, https://doi.org/10.3389/fpls.2016.01682.

26.   Font, Rafael, et al. "The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs." *Industrial Crops and Products*, vol. 24, no. 3, Nov. 2006, pp. 307–313, https://doi.org/10.1016/j.indcrop.2006.06.012.

27.   Büning-Pfaue, Hans. "Analysis of water in food by near infrared spectroscopy." *Food Chemistry*, vol. 82, no. 1, July 2003, pp. 107–115, https://doi.org/10.1016/s0308-8146(02)00583-6.

28.   Li, Hongbo, et al. "Near-infrared spectroscopy coupled chemometric algorithms for rapid origin identification and lipid content detection of pinus koraiensis seeds." *Sensors*, vol. 20, no. 17, 30 Aug. 2020, p. 4905, https://doi.org/10.3390/s20174905.

29.   Wang, Maoyao, et al. "A systematic high-throughput phenotyping assay for sugarcane stalk quality characterization by near-infrared spectroscopy." *Plant Methods*, vol. 17, no. 1, 13 July 2021, https://doi.org/10.1186/s13007-021-00777-8.

30.   Tang, Chaochen, et al. "High-throughput phenotyping of nutritional quality components in sweet potato roots by near-infrared spectroscopy and chemometrics methods." *Food Chemistry: X*, vol. 20, Dec. 2023, p. 100916, https://doi.org/10.1016/j.fochx.2023.100916

31.   Agelet, Lidia Esteve, and Charles R. Hurburgh. "A tutorial on near infrared spectroscopy and its calibration." *Critical Reviews in Analytical Chemistry*, vol. 40, no. 4, 29 Oct. 2010, pp. 246–260, https://doi.org/10.1080/10408347.2010.515468.

32.   Osborne, Brian G. "Near-infrared spectroscopy in food analysis." *Encyclopedia of Analytical Chemistry*, 30 Oct. 2000, https://doi.org/10.1002/9780470027318.a1018.

33.   Kaur, Balraj, et al. "Calibration of nirs for the estimation of fatty acids in *brassica juncea*." *Journal of the American Oil Chemists' Society*, vol. 93, no. 5, 2 Mar. 2016, pp. 673–680, https://doi.org/10.1007/s11746-016-2802-0.

34.   Westerhaus, Mark. "Eastern Analytical Symposium Award for Outstanding Achievements In near infrared spectroscopy: My contributions to near infrared spectroscopy." *NIR News*, vol. 25, no. 8, Dec. 2014, pp. 16–20, https://doi.org/10.1255/nirn.1492.

35.   Bellon-Maurel et al. "Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – critical review and research perspectives." *Soil Biology and Biochemistry*, vol. 43, no. 7, July 2011, pp. 1398–1410, https://doi.org/10.1016/j.soilbio.2011.02.019.

36.   Mandalari et al. Pistachio Nuts (Pistacia vera L.): Production, Nutrients, Bioactives and Novel Health Effects. Plants 2022, 11, 18. https://doi.org/10.3390/plants11010018.

37.   Okay, Yesim. "The comparison of some pistachio cultivars regarding their fat, fatty acids and protein content." *Gartenbauwissenschaft* 67.3 (2002): 107-113.

38.   Guihua Fang et al. Characterization of oils and fats by 1H NMR and GC/MS fingerprinting: Classification, prediction and detection of adulteration, Food Chemistry, Volume 138, Issues 2–3, 2013, Pages 1461-1469, ISSN 0308-8146, https://doi.org/10.1016/j.foodchem.2012.09.136.

39.   FAO. Food and Agriculture Organization of the United Nations. Retrieved September 22, 2021, http://www.fao.org/common-pages/search/en/?q=pistachio

40.   Tsantili et al. Physical, compositional and sensory differences in nuts among pistachio (Pistachia vera L.) varieties, Scientia Horticulturae, Volume 125, Issue 4, 2010, Pages 562-568, ISSN 0304-4238, https://doi.org/10.1016/j.scienta.2010.04.039.

41.   Bellomo & Fallico. Anthocyanins, chlorophylls and xanthophylls in pistachio nuts (Pistacia vera) of different geographic origin, Journal of Food Composition and Analysis, Volume 20, Issues 3–4, 2007, Pages 352-359, ISSN 0889-1575, https://doi.org/10.1016/j.jfca.2006.04.002.

42.   Mateos et al. Why Should Pistachio Be a Regular Food in Our Diet? *Nutrients* **2022**, *14*, 3207. https://doi.org/10.3390/nu14153207