IMPROVEMENT OF NUTRITIONAL CHARACTERISTICS AND OTHER HEALTH BENEFITS OF PROCESSED MAIZE FOOD PRODUCTS

BY

CAROLYN J BUTTS-WILMSMEYER

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Crop Sciences in the Graduate College of the University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Associate Professor Martin Bohn, Chair
Professor Jack Juvik
Professor Fred Kolb
Assistant Professor Alex Lipka
Professor Emerita Rita Mumm
ABSTRACT

Maize is one of the world’s most important and abundant cereal crops. It also contains phytochemicals which are beneficial to human health. This presents an opportunity to breed for maize food products which possess high levels of these beneficial phytochemicals. However, maize grain is not consumed directly. Before consumption, the grain must be processed. Due to the physical, chemical, and thermal stresses encountered during processing, the fate of various phytochemicals is unknown. Similarly, it is unknown if all maize genotypes respond to processing stresses in the same manner or if some genotypes are more resistant to changes in the concentrations of beneficial phytochemicals than others. Should some maize genotypes exhibit an ability to resist changes in phytochemical content during processing, it may be possible to breed for healthier processed maize food products. However, if beneficial phytochemicals are lost during processing, then it may be more effective to breed for all-natural food additives which can be extracted from the whole grain and used to fortify processed food products.

From a breeding perspective, variability must exist for a trait to be improved. Also, it is helpful if the germplasm evaluated is representative of that which is most likely to be processed into maize food products so that direct conclusions can be made. Twelve inbreds consisting of ten ex-PVPs and two public lines were selected for this study. These inbreds are genetically representative of the maize germplasm grown in the US Cornbelt and, consequently, maize which is used in the making of processed food products. These inbreds were crossed using a half-diallel design to create 66 F₁ hybrids. The parental inbreds and their hybrids (N = 78 entries) were evaluated together and grown in a Resolvable Incomplete Block Design with three
replications for three years. One pound (454 g) of grain was harvested from each plot and set aside for phytochemical analysis.

Most beneficial phytochemicals were relatively easy to measure using wet-lab chemistry or Fourier Transform Infrared Spectroscopy (FTIR), but the insoluble-bound hydroxycinnamic acids were difficult to quantify using standard extraction protocols. Therefore, at the onset of the project, a high-throughput method for the extraction of insoluble-bound hydroxycinnamic acids was developed (Chapter 1). The overall concentration of the insoluble-bound and soluble hydroxycinnamic acids, tocopherols, unsaturated fatty acids, and protein were measured at the whole kernel level in all inbreds and a subset of the hybrids to determine which phytochemicals in which genotypic variability was present. Only the insoluble-bound hydroxycinnamic acids and soluble cinnamic acid showed considerable genotypic variability and high concentrations (Chapter 2). The hydroxycinnamic acids were selected as candidate phytochemicals for improvement, and both the insoluble-bound and soluble hydroxycinnamic acids were measured in subsequent portions of the project.

All inbreds and a subset of the hybrids (N = 7) used in this study were then processed into ready-to-eat breakfast cereals using laboratory scale techniques. The processing stages at which the hydroxycinnamic acids significantly changed concentration were identified. The concentrations of all candidate phytochemicals were analyzed at each of these processing stages (Chapter 3). Most phytochemicals showed a large and significant change in concentration due to processing. Also, there was a significant interaction between the genotype and the processing stage that resulted in a change of rank among the genotypes tested. Therefore, it would be
impossible to use the phytochemical concentration at the whole kernel level to predict the phytochemical concentration in the final processed food product.

Next, the possibility of extracting hydroxycinnamic acids from maize and using the extracts as all-natural food additives was examined (Chapter 4). The insoluble-bound hydroxycinnamic acid content of all hybrids and inbreds used in this study was measured. Genetic variance components were calculated, and the broad- and narrow-sense heritabilities were estimated. Two particular insoluble-bound hydroxycinnamic acids, namely ferulic acid and p-coumaric acid, were identified as having both relatively high concentrations in maize and high heritabilities. As such, the overall finding of this study was that insoluble-bound ferulic acid and p-coumaric acid concentrations in maize grain could be improved through selection. Also, the most efficient method of increasing the concentration of beneficial phytochemicals in processed food products is through the use of all-natural food additives.

Another consideration in regards to breeding is high-throughput phenotyping. In addition to the different wet-lab techniques that were optimized in this study, the potential of measuring ferulic acid content and p-coumaric acid content using NIR was examined (Chapter 5). Correlation coefficients among phytochemical traits and agronomic, ear, and cob traits were calculated to determine if indirect selection might be possible. NIR methods proved to be inaccurate, and the correlations between the beneficial phytochemicals and the agronomic, ear, and cob traits were not strong enough to build a reliable prediction model for indirect selection. Therefore, the phenotyping approach which will most likely be successful is high-throughput wet-lab chemistry to measure the overall concentration of ferulic acid and p-coumaric acid in
maize. Furthermore, this phenotyping should be conducted at the whole kernel processing stage because the most feasible means of increasing the ferulic acid and p-coumaric acid content of maize food products is via the extraction of hydroxycinnamic acids from the whole kernel for later use as all-natural food additives.
ACKNOWLEDGEMENTS

Much like it takes a community to raise a child, it takes a community to raise a scholar.

To my parents, my first teachers and arguably two of my best teachers, thank you for your love and support, your prayers and counseling, and everything from sharing “divide and conquer” to being an example of what it truly means to give everything for the betterment of those you love.

To my grandfather, Kenneth, for teaching me how to never give up. To my grandparents, Wayne and Carolyn, for passing their love of agriculture on to their children, grandchildren, and great-grandchildren. To the communities of Caldwell County, Princeton, and Lewistown, for always giving me a place to call home and for being ever supportive.

To all of those whom I consider a teacher, including my advisor, Dr. Martin Bohn, and my committee members, who have all passed along gifts of knowledge. To Mrs. Wetzel, Mr. Gary Goodaker, Mr. York, Mr. Baker, Mr. Hays, Mrs. T. Scott, Mrs. Joanie Kindrick, and Ms. Pat Scott, without whom I would not be where I am today. To Dr. Bill and Dr. Mary Witt, who are phenomenal teachers and wonderful role models to young scientists. To Dr. Don Bullock, who taught me I really didn’t know that much about statistics…and then proceeded to do his best to correct that. To Dr. Alex Lipka and Dr. Germán Bollero, who continued my education in what I consider to be the most fun component of my graduate training. To Dr. Tom Patterson, Rhee Holley, Mary Evenson, Amy Latham, and all of the wonderful scientists that I met at Dow AgroSciences, thank you for your mentorship, your friendship, your help, and for being amazing people to work with. To Dr. Donn Cummings, for sharing his devotion to plant breeding and doing everything he could to give a young scientist with limited connections a chance.
To Ms. Nicole Yana, for not just being an amazing friend but also doing everything she could to make this project a success. I am convinced that it would have failed without you. To Mary, Alexandra, and Matt, who devoted more than a few hours to this project and its success. To all of my amazing in-laws, especially Doug and Tracy, for your support and for always making me feel at home. To both Dad and Doug, for helping me formulate a major component of my research project. To Dr. Ed Settle, for giving me the courage to pursue a field of study that I love.

Last but most certainly not least, thank you to my husband, Marcus, for your love and support, for your friendship, for doing everything from laboratory work to organizing seven tons of seed in a coldroom, for moving me six times in the course of three years as my research took me back and forth from Champaign to Indianapolis, and for being someone to laugh with or vent to, no matter what. Thank you for encouraging me to stay at the University of Illinois, even when I was so discouraged that I wanted to quit. I would not have stayed and completed my PhD without your help.

In Loving Memory of Those Who Fought and Have Passed On
Kenneth Ray Hammonds

In Honor of Those Who Have Fought and Won or Suffered with Their Loved Ones
David Wayne and Dixie Gail Butts

And Dedicated to the Future, That They Might Not Know the Pain of Aging-Related Diseases
Matthew David Butts
Kenneth Marcus Wilsmeyer

Ad Majorem Dei Gloriam, in nomine Patris, et Filii, et Spiritus Sancti.
# TABLE OF CONTENTS

LIST OF TABLES ....................................................................................... xi

LIST OF FIGURES ................................................................................... xiii

LITERATURE REVIEW ............................................................................. 1

Medical and Societal Importance ................................................................. 1

Types of Beneficial Phytochemicals in Maize ............................................. 3

Maize Food Product Processing ................................................................. 7

High-Throughput Phenotyping Via NIR ..................................................... 8

Objectives .............................................................................................. 9

Works Cited .......................................................................................... 10

CHAPTER 1: DEVELOPMENT OF A HIGH-THROUGHPUT PROTOCOL FOR
THE EXTRACTION OF INSOLUBLE-BOUND FERULIC ACID IN MAIZE ......... 14

Abstract ............................................................................................. 14

Introduction .......................................................................................... 15

Materials and Methods ......................................................................... 17

Results and Discussion ....................................................................... 23

Tables and Figures ............................................................................... 29

Works Cited ........................................................................................ 33
CHAPTER 2: IDENTIFICATION OF NUTRITIONAL COMPOUNDS OF INTEREST IN TYPICAL YELLOW DENT MAIZE.................................37

Abstract.............................................................................................................37
Introduction........................................................................................................38
Materials and Methods....................................................................................41
Results and Discussion....................................................................................47
Tables................................................................................................................53
Works Cited......................................................................................................56

CHAPTER 3: CHANGES IN PHENOLIC ACID CONTENT IN MAIZE DURING FOOD PRODUCT PROCESSING.............................................60

Abstract.............................................................................................................60
Introduction........................................................................................................61
Materials and Methods....................................................................................64
Results and Discussion....................................................................................73
Tables and Figures............................................................................................80
Works Cited......................................................................................................95
### CHAPTER 4: POTENTIAL FOR BREEDING MAIZE FOR USE AS ALL-NATURAL FOOD ADDITIVES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>98</td>
</tr>
<tr>
<td>Introduction</td>
<td>99</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>101</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>108</td>
</tr>
<tr>
<td>Tables and Figures</td>
<td>119</td>
</tr>
<tr>
<td>Works Cited</td>
<td>124</td>
</tr>
</tbody>
</table>

### CHAPTER 5: PREDICTING PHENOLIC ACID CONTENT IN MAIZE USING NIR SPECTROSCOPY AND INDIRECT SELECTION APPROACHES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>128</td>
</tr>
<tr>
<td>Introduction</td>
<td>129</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>131</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>135</td>
</tr>
<tr>
<td>Tables and Figures</td>
<td>138</td>
</tr>
<tr>
<td>Works Cited</td>
<td>144</td>
</tr>
</tbody>
</table>

Appendix A: GLOSSARY

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>146</td>
</tr>
</tbody>
</table>
LIST OF TABLES

CHAPTER 1

Table 1.1. Mean ferulic acid content and percent recovery obtained for 13 maize hybrids using two extraction protocols ...............................................................29

CHAPTER 2

Table 2.1. Inbred maize parents used and their background ..................................53
Table 2.2. Summary statistics of compounds of interest by generation ..................54
Table 2.3. Mean Squares and P-values of model effects for maize hybrid traits which exhibited appreciable means and ranges .........................................................55
Table 2.4. Correlations between traits which exhibited appreciable means and ranges ...................55

CHAPTER 3

Table 3.1. Means of phenotypic traits by cluster ......................................................80
Table 3.2. Multi-degree of freedom contrasts testing the difference in insoluble-bound hydroxycinnamic acid content in cooked grits, baked grits, and toasted cornflakes ....81
Table 3.3. Means and standard deviations of phenolic acids and cinnamic acid at different processing stages ....................................................................................82
Table 3.4. Significance of genotype-by-processing stage interaction ..........................82
CHAPTER 4

Table 4.1. Significance of ANOVA model effects for ferulic acid and p-coumaric acid and their variance component estimates.................................................................119

Table 4.2. Summary statistics of extractable ferulic acid and p-coumaric acid content........119

Table 4.3. Ferulic acid content of other grains and food sources..........................................120

Table 4.4. Correlations between key agronomic and phytochemical traits........................120

CHAPTER 5

Table 5.1. Description of traits used in stepwise linear regression analysis and discriminant analysis........................................................................................................138

Table 5.2. Description of classification variables used in discriminant analysis..................139

Table 5.3. R² values and error rates associated with indirect selection models......................139

Table 5.4. Correlations between phenolic acid traits and prediction traits..........................140
LIST OF FIGURES

CHAPTER 1
Figure 1.1. Photos of samples at various stages during extraction in Protocol B..........................30
Figure 1.2. Interaction plot between hybrids and protocol............................................................31
Figure 1.3. Scatterplot between values obtained using protocols A and B....................................31
Figure 1.4. Boxplot of percent recovery by hybrid.................................................................32
Figure 1.5. Cost efficiencies for different scenarios........................................................................32

CHAPTER 3
Figure 3.1. Processed maize food product production pathways................................................83
Figure 3.2. Cornflake production pipeline.........................................................................................84
Figure 3.3. Overview of laboratory scale processing used in the production of cornflakes........85
Figure 3.4. Description of materials used at different phases in this study......................................86
Figure 3.5. Custom fabricated nitrogen evaporator........................................................................87
Figure 3.6. Cluster analysis of the phenotypic diversity of the 13 hybrids used in this study...........88
Figure 3.7. Cluster analysis of the genetic diversity in the 12 inbred parents.................................89
Figure 3.8. Change in phytochemical content by processing stage..............................................90
Figure 3.9. Change in insoluble-bound ferulic acid content throughout processing using
            two years and seven hybrids.......................................................................................................93
Figure 3.10. Change in insoluble-bound p-coumaric acid content throughout processing using
              two years and seven hybrids...............................................................................................93
Figure 3.11. Interaction between processing stage and hybrid for phytochemical content............94
CHAPTER 4

Figure 4.1. Profile plot of the interaction between maize genotypes and their environments........................121
Figure 4.2. Selection response for ferulic acid content.............................................................................122
Figure 4.3. Selection response for p-coumaric acid content......................................................................122
Figure 4.4. Three-dimensional scatterplot of ferulic acid per hectare, p-coumaric acid

per hectare, and grain yield.....................................................................................................................123

CHAPTER 5

Figure 5.1. Validation of NIR predictions of ferulic acid content..............................................................141
Figure 5.2. Validation of NIR predictions of p-coumaric acid content.......................................................141
Figure 5.3. Chemical structure of ferulic acid, 4-hydroxy-3-methoxycinnamic acid...............................142
Figure 5.4. Chemical structure of p-coumaric acid, 4-hydroxycinnamic acid........................................142
Figure 5.5. NIR spectra following vector normalization..............................................................................143
LITERATURE REVIEW

MEDICAL AND SOCIETAL IMPORTANCE

Over the last 100 years, great advances have been made in medicine and healthcare. This is clearly evidenced in the drastic changes in the average life expectancy over time. The CDC reports that, in 1900, the average life expectancy of an individual in the United States was 49.24 years (Arias, 2014). In contrast, the average life expectancy of an individual in the United States today is about 78.7 years (CDC, 2015). This general trend has been seen worldwide, as well (United Nations, Department of Economic and Social Affairs, Population Division, 2015).

These advances, while outstanding, have created a new issue. As a result of the world’s aging population, especially in developed nations, the incidence of aging-related diseases has increased (World Health Organization, 2011). While people may be living longer, the quality of their lives has not necessarily improved concurrently with their life expectancy.

Current research suggests that there may be several ways of addressing these issues, one of which is a life-long diet filled with beneficial phytochemicals (Adom and Liu, 2002; Kanski et al., 2002; Liu, 2007; Rose et al., 2009; Rumpagapom, 2011; Wong et al., 2006). For this study, the main interest was in prebiotics, antioxidants, chemopreventive agents, total protein content, and unsaturated fatty acids. Foods higher in dietary protein lead to increased satiety and a lower chance of developing diseases such as diabetes, heart disease, and other obesity-related diseases (Westerterp-Plantenga et al., 2009). As many obesity-related diseases tend to occur later in life, several of these diseases, such as diabetes and heart disease, are also considered aging-related diseases. Likewise, the consumption of unsaturated fatty acids (e.g. healthy plant oils) in a
balanced diet is associated with the maintenance of cell lipids and a lower incidence of heart disease and high cholesterol.

Antioxidants serve the general purpose of capturing reactive oxygen species (ROS), such as superoxide \( (O_2^-) \), hydrogen peroxide \( (H_2O_2) \), and hydroxyl radicals \( (\cdot OH) \), before the ROS can cause oxidative damage to DNA, proteins, lipids, and other essential biological molecules (Ames et al., 1993). Thus, antioxidants prevent the accumulation of DNA lesions and mutations as well as the formation of neurodegenerative plaques that are associated with diseases such as Alzheimer’s Disease. However, some beneficial phytochemicals that were once thought to prevent aging-related illnesses via their antioxidant properties may instead do so via prebiotic or other chemopreventive mechanisms (Fresco et al., 2006). Prebiotic compounds are food ingredients or components which, when consumed, selectively stimulate the growth and/or activity of beneficial bacteria in the colon. This results in improved systemic health (Gibson and Roberfroid, 1995). Additionally, improved systemic health has been tied to a decrease in the incidence of some aging-related diseases (Buttriss and Stokes, 2008).

Unfortunately, not all people have access to the phytochemicals that could improve their quality of life. Traditionally, doctors and researchers have prescribed the consumption of fresh produce and whole-grain carbohydrates. However, diet quality, particularly the consumption of fresh produce and non-processed foods, follows a socioeconomic gradient. People from lower socioeconomic classes tend to consume more processed, calorie-dense foods due to the relatively higher price of fresh produce (Darmon and Drewnowski, 2008; Drewnowski and Darmon, 2005). Therefore, beneficial phytochemicals are not readily accessible to all people.
TYPES OF BENEFICIAL PHYTOCHEMICALS IN MAIZE

Protein

Maize is primarily composed of starch, but protein and oil are minor components. Typically, maize contains approximately 10% protein. The protein is distributed between the endosperm and the germ (Earle et al., 1946), with the endosperm containing a greater quantity but a lower concentration of protein than the germ (Wilson, 1987). As most processed maize food products require the removal of the germ during food product processing, the small amount of protein in the whole kernel will be partially removed and lost during production.

Diets with a relatively high protein intake tend to result in satiety and weight maintenance. Specifically, as satiety is maintained for longer periods of time, people tend to eat less ad libitum (Westerterp-Plantenga et al., 2009). This contributes to a lower overall calorie intake and a decreased chance of developing an obesity-related disease. Additionally, while maize protein is often criticized for being low quality and lacking the amino acids tryptophan and lysine, the quality of protein does not influence satiety maintenance (Westerterp-Plantenga et al., 2009).

Unsaturated Fatty Acids

Unsaturated fatty acids are carboxylic acids characterized by a polar head and a long, nonpolar aliphatic tail in which there is at least one carbon-carbon double bond. Most of the fatty acids in plants occur as oils and tend to be located in the germ. In maize, almost all of the fatty acids present are either monounsaturated fatty acids (a fatty acid containing only one
carbon-carbon double bond) or polyunsaturated fatty acids (a fatty acid containing two or more carbon-carbon double bonds).

Oleic acid and linoleic acid are the two most common unsaturated fatty acids in maize (Baur Jr and Brown, 1945). Oleic acid is a monounsaturated fatty acid which is reported to decrease blood pressure levels (Teres et al., 2008). Linoleic acid is a polyunsaturated acid which is an essential fatty acid and cannot be synthesized by humans. However, it is an important component of cell membranes and must be consumed for good health (Simopoulos, 1999). Maize is not typically recognized as an oil seed crop due to its minor concentration of oil. Furthermore, fatty acids are found almost exclusively in the germ and will most likely be removed during food product processing.

**Tocopherols**

Eight different compounds comprise Vitamin E, those compounds being four different tocopherols and four different tocotrienols. The four different tocopherols are alpha, beta, gamma, and delta tocopherol. In maize, typically only the alpha, gamma, and delta tocopherols are present (Weber, 1987). Besides their role in Vitamin E, tocopherols are the primary antioxidants in lipid membranes and prevent oxidation damage caused by peroxyl radicals and nitrogen oxide species (Christen et al., 1997). Of the three tocopherols present in maize, alpha tocopherol tends to be the most efficient antioxidant (Sies and Stahl, 1995). Specifically, alpha tocopherol is noted for reducing peroxyl radicals, such as those from polyunsaturated fatty acids or lipoproteins (Sies and Stahl, 1995).
Hydroxycinnamic Acids

The hydroxycinnamic acids are a class of aromatic compounds which are hydroxy derivatives of cinnamic acid. The most abundant of these compounds in cereals is ferulic acid (Adom and Liu, 2002). Other types of hydroxycinnamic acids which may be present in grains are p-coumaric acid and sinapic acid. Cinnamic acid, a carboxylic acid, may also be present, although ferulic acid and p-coumaric acid are typically the only hydroxycinnamic acids with appreciable concentrations in maize (Yadav et al., 2007). Most of the hydroxycinnamic acids in cereals tend to be in the insoluble-bound form, meaning that they are esterfied to the arabinoxylans of the cell wall. Hydroxycinnamic acids in their soluble form may also appear in the vacuoles (Pande and Rizvi, 2009). Hydroxycinnamic acids are primarily found in cereal bran (Rumpagapom, 2011). Since the bran is removed during the production of many maize-based snack foods and breakfast cereals, a major percentage of the hydroxycinnamic acids are expected to be lost during processing.

In their insoluble-bound form, hydroxycinnamic acids act as cross-linkages which slow the fermentation of the arabinoxylan fibers. Together, the insoluble-bound hydroxycinnamic acids and the arabinoxylans to which they are bound act as prebiotics by creating an environment in which beneficial gut bacteria (including Lactobacillus and Bifidobacterium species) can thrive. Furthermore, due to the extensive cross-linking of ferulic acid in maize fiber, the fermentation slows such that the beneficial properties of the phytochemicals can be recognized without the negative effects of extensive gas production (Rumpagapom, 2011). While these properties are beneficial, the insoluble-bound form of hydroxycinnamic acids is largely
unabsorbed in vivo (Adam et al., 2002), meaning that the other beneficial properties of the hydroxycinnamic acids are negligible in the insoluble-bound form.

Initially, the hydroxycinnamic acids were thought to act as antioxidants in vivo if present in their soluble form. In vitro, the hydroxycinnamic acids scavenge ROS species directly. Ferulic acid is particularly adept in its scavenging ability, followed by p-coumaric acid (Kansi et al., 2002). Ferulic acid’s high antioxidant potential is due to its resonance-stabilized phenoxy radical structure (Adam et al., 2002). Furthermore, a large pool of research studies found that the consumption of hydroxycinnamic acids resulted in the reduced incidence of aging-related diseases. Additionally, oxidative stress (caused by ROS) has been implicated in the formation of several aging-related diseases, especially neurodegenerative conditions. Consequently, it was thought that ferulic acid and the other hydroxycinnamic acids directly prevented the incidence of aging-related diseases, particularly neurodegenerative diseases, by directly acting as antioxidants in vivo.

More recent studies have offered an alternative mechanism of prevention that is not directly tied to the hydroxycinnamic acids’ ability as antioxidants. While the ability of hydroxycinnamic acids to scavenge ROS directly is certainly still a possibility, a more likely method of in vivo protection is via the activation and regulation of different gene pathways. For instance, the regulation of the heterodimers of NF-E2-related factors 2 (Nrf2)/antioxidant responsive element (ARE) pathway is implicated in the prevention of neurodegenerative diseases (Scapagnini et al., 2011). The regulation of other signaling pathways, including the nuclear
factor-κB (NF-κB), activator protein-1 (AP-1), or mitogen-activated protein kinases (MAPK), are hypothesized to be involved in the chemoprevention of various cancers (Fresco et al., 2006).

MAIZE FOOD PRODUCT PROCESSING

During the production of ready-to-eat breakfast cereals (also known as cold cereals) and several maize-based snack food products, harvested maize kernels encounter numerous physical, thermal, chemical, and shearing forces which might influence the nutritional content of the final food product. For instance, during the production of cornflakes, maize kernels are dry milled into large flaking grits, steam-pressure cooked in a salt, water, and sugar solution, baked, rolled using an apparatus which can generate a shearing force, and toasted (Kandhola, 2015; Macke, et al., 2016). During the dry milling process, physical changes can take place due to the removal of the bran and germ. Therefore, beneficial phytochemicals that are predominantly located in the bran or germ (i.e. unsaturated fatty acids, tocopherols, and hydroxycinnamic acids) are expected to be lost during processing. In regards to protein, it is possible that the thermal stresses encountered during processing may also degrade the protein which survived the dry milling process (i.e. the removal of the germ).

To date, at least to the knowledge of the researcher, there have been no nutritional studies which have i) examined the effect of processing in maize while ii) also examining the effect of different hybrids iii) with regard to the phytochemicals of interest in this study. Most nutritional studies which examine the effect of processing on end product nutritional quality use samples which are larger than the quantity of grain typically harvested from a research plot. As such, whether different genotypes are more resistant to nutritional changes during food product
processing has scarcely been examined. Kandhola (2015) examined the changes in resistant starch concentration among hybrids during processing using a small-scale laboratory processing technique. This work was a continuation of the small-scale laboratory dry milling technique outlined in Macke, et al. (2016). In total, these two processing techniques use less than 1 kg of harvested maize, making such a procedure ideal for plant breeding studies in this area.

HIGH-THROUGHPUT PHENOTYPING VIA NIR

While total protein content in maize can rapidly be measured using Fourier Transform Infrared Spectroscopy (FTIR) and total oil content can easily be measured using Nuclear Magnetic Resonance Spectroscopy (NMR) (Thomas Patterson and Daniel Garcia, Personal Communication, 2014), the measurement of other beneficial phytochemicals in this study typically must be conducted using wet-lab techniques that may reduce throughput. The implementation of Near Infrared Spectroscopy (NIR) could potentially increase the throughput at which certain nutritional compounds are quantified. Ferrer-Gallego, et al. (2011) reported success in using NIR to predict the concentration of phenolic compounds. Considering that hydroxycinnamic acids are phenolic acids and that, in regards to phenolic acids, Ferrer-Gallego, et al. (2011) reported a Ratio Performance Deviation (RPD) of 6.3 as well as an $R^2$ of 94.7%, it is possible that high-throughput NIR techniques may be adapted for use in maize. Furthermore, some NIR methods are non-destructive. Should non-destructive NIR techniques be adapted for use in predicting the concentration of some beneficial phytochemicals in maize, it may be possible to use the seed in future plant breeding endeavors.
OBJECTIVES

As the age of the world population continues to increase, so does the incidence of aging-related diseases. Several of these aging-related diseases have been shown to be prevented by diets high in certain beneficial phytochemicals; however, these beneficial phytochemicals are currently not available to all socioeconomic classes. Therefore, research is needed to identify inexpensive means of incorporating these beneficial phytochemicals that are often found in whole grains or fresh produce into food products which are affordable to all socioeconomic classes. The overall goal of this study is to provide information which will be instrumental in developing more nutritious processed maize food products. The sub-objectives of this research are to:

1) Develop high-throughput wet-lab techniques for quantifying the concentration of various beneficial phytochemicals at the whole kernel and all subsequent processing stages.

2) Determine which beneficial phytochemicals exhibit the most variability and highest average contents.

3) Quantify the concentrations of the selected beneficial phytochemicals at all processing stages involved in producing ready-to-eat breakfast cereals.

4) Quantify the concentrations of compounds of interest in a diverse and genetically representative set of maize hybrids and inbreds at the whole kernel level.

5) Determine the usefulness of NIR and other high-throughput screening methods (such as indirect selection) in evaluating the concentration of beneficial phytochemicals in whole maize kernels.
WORKS CITED


Kandhola, G. 2015. Processing and genetic effects on resistant starch in corn flakes. MS. University of Illinios, Urbana, IL.


Rumpagapom, P. 2011. Structural features of cereal bran arabinoxylans related to colon fermentation rate. PhD. Purdue University, West Lafeyette, IN.


CHAPTER 1: DEVELOPMENT OF A HIGH-THROUGHPUT PROTOCOL FOR THE 
EXTRACTION OF INSOLUBLE-BOUND FERULIC ACID IN MAIZE

ABSTRACT

Ferulic acid is a hydroxycinnamic acid that possesses positive health benefits and plays a role in both pathogen and insect resistance in maize (Zea mays L.). Many protocols for the extraction of insoluble-bound ferulic acid in maize use methods that are low-throughput and are not conducive to analyzing the large number of samples used in field research or breeding programs. The examination of two protocols for the extraction of insoluble-bound ferulic acid that use small sample sizes typical of experiments where a large number of samples need to be analyzed were studied. The main distinguishing feature between the two protocols is the inclusion of a starch digestion step in Protocol A to inhibit the swelling of intact starch when introduced to an alkali reagent. The results indicate that the inclusion of a starch digestion step during the extraction of insoluble-bound ferulic acid from maize is necessary if small sample sizes are used. The omission of the starch digestion step in Protocol B resulted in an average percent recovery of only 79.0% of the insoluble-bound ferulic acid. Conversely, the inclusion of the starch digestion step in Protocol A allowed for the accurate and high-throughput extraction of insoluble-bound ferulic acid at lower cost than Protocol B.
INTRODUCTION

Phenolics in maize (*Zea mays* L.) are of interest to researchers in plant pathology, entomology, and human nutrition. A better understanding of the biosynthesis and genetic basis of phenolics will help plant breeders select lines for hybrid production with higher phenolics content and subsequently improved resistance to insects and diseases as well as healthier grain. The majority of the phenolics in maize are in the insoluble-bound form, and most of these are hydroxycinnamic acids (Adom and Liu, 2002). Of the insoluble-bound phenolics, ferulic acid is the most common in maize (Adom and Liu, 2002; Saulnier et al., 1995; Yadav et al., 2007). Insoluble-bound ferulic acid is also implicated to be the insoluble-bound phenolic with the greatest antioxidant potential in maize (Kanski et al., 2002) and possesses positive prebiotic effects (Fresco et al., 2006; Rose et al., 2009; Rumpagapom, 2011; Saulnier et al., 1999; Wong et al., 2006). Ferulic acid’s ability to form extensive fiber networks makes pathogen infection and insect feeding more difficult in the grain (Arnason et al., 1992; Assabgui et al., 1993; Bily et al., 2003; Classen et al., 1990; Dowd et al., 1997; García-Lara et al., 2004). There is also evidence suggesting that ester-linked diferulates deter insect feeding in vegetative tissue as well as the grain (Barros-Rios et al., 2015). Likewise, ferulic acid is associated with pathogen resistance in vegetative tissue (Bennett and Wallsgrove, 1994). Ferulic acid in harvested grain was the focus of this study.

Most protocols for the extraction of insoluble-bound phenolics from grains, including insoluble-bound ferulic acid, usually include an alkali hydrolysis step to break the ester bonds that bind the insoluble-bound phenolics to the arabinoxylan backbone of hemicellulose (Adom and Liu, 2002; Naczk and Shahidi, 2006). Exposure to some alkali reagents, such as sodium
hydroxide, causes starch to swell (Sivasankar, 2002). The swelling of starch does not appear to be an issue if the size of the individual samples is large enough to disperse the gelatinous mass that forms due to the swelling of the starch. Therefore, many standard protocols use large sample sizes to avoid the swelling of starch. However, if using small sample sizes, the swelling becomes more pronounced (i.e. is more noticeable because the swollen starch is somewhat diluted in large samples) and may result in a significantly lower recovery of insoluble-bound phenolics (Thomas Patterson, Personal Communication, 2015).

In modern plant breeding programs, high-throughput, highly accurate, and cost-efficient phenotyping methods must be used to accommodate large quantities of samples. The most direct way of increasing laboratory throughput is to scale down laboratory protocols, thereby allowing more samples to be analyzed concurrently. Yet, the swelling of starch, if not addressed, may inhibit the extraction of phenolics and thus may lead to inaccurate results. One method of addressing this is to pre-digest the starch before adding an alkali reagent.

The overall objective was to develop an accurate, high-throughput, and cost-efficient protocol for the extraction of insoluble-bound ferulic acid from maize kernels. The throughput and recovery of insoluble-bound ferulic acid in two “small sample” extraction protocols were studied, as were the genotype-by-protocol interaction effects. Additionally, one of the benefits of both protocols is that both could easily be modified to scan for the other hydroxycinnamic acids in maize.
MATERIALS AND METHODS

Plant Materials

Thirteen maize hybrids (Table 1.1) were developed from a set of genetically diverse inbreds that are typical of the germplasm grown in the U.S. Cornbelt (Hauck et al., 2014). These hybrids were evaluated using an RCBD with three replications at the University of Illinois Crop Sciences Research and Education Center (Champaign, IL) in 2009. After harvest, approximately 0.1 kg of maize seed per hybrid was ground to a fine powder using a Foss Cyclone Mill. A 0.5 g subsample was used in each of the extraction protocols described below. The remainder of the ground maize was set aside as a reserve sample and for other chemical analyzes which will be discussed in later chapters.

Extraction of Insoluble-bound Ferulic Acid

In this chapter, small sample size is defined as a sample with a small mass of ground maize and small volumes of reagents, not as a small number of observations. In both Protocols A and B, 0.5 g of ground maize underwent three hexane washes to remove fatty acids and nonpolar compounds. The defatted maize was dried using a centrifugal evaporator. The defatted maize was then washed three times with 70% acetone to remove all soluble compounds, particularly the soluble phenolics.

In Protocol A, 8 mL of pH 7.5 buffer and 0.1 mL of α-amylase were added to each vial of defatted meal following the removal of the soluble phenolics and other soluble compounds. Samples were shaken and placed in a horizontal incubator/shaker at 65°C for at least 8 hours. Samples were removed from the incubator/shaker, cooled, shaken, and then centrifuged at 1,500
rpm for 5 minutes. The liquid was aspirated out of each vial. These digestion and incubation steps were not included in Protocol B.

From this point on, Protocols A and B only varied in their centrifugation times. In both cases, 2 mL of 2 M NaOH was added to each sample vial. The vial and cap were flushed with nitrogen gas. The vial was capped, and each sample was placed in a hot water bath at 50 ºC for two hours. Each sample was shaken approximately every 15 minutes.

Samples were removed from the hot water bath, and 0.7 mL 6 M HCl, 3 mL NaCl, and 3 mL ethyl acetate were added to each vial. Vials were again capped, vortexed, and placed in a horizontal shaker for 20 minutes. Vials were then centrifuged at 1,500 rpm for 7 minutes if Protocol A was used. If Protocol B was used, the starch gelatinization made it much harder for the layers to separate (Fig. 1.1). Therefore, the centrifugation time was increased to 15 minutes to allow for a better separation of the different solvent layers. The ethyl acetate fraction, which is the top fraction after centrifugation takes place, was removed and placed in a clean vial. Ethyl acetate extraction was repeated twice for a total of three ethyl acetate extractions. The total extract was then evaporated to complete dryness using a centrifugal evaporator.

Identification and Determination of Retention Time

A standard sample of ferulic acid was analyzed on a gas chromatography-mass spectroscopy (GC-MS) system to determine its retention time in this study. The retention time at which the ferulic acid peak occurred was recorded and used to calculate the concentration of insoluble-bound ferulic acid present in each of the test samples. Randomly selected samples
from among the 39 harvested samples used in this study were also analyzed on the GC-MS system to ensure that ferulic acid was present in test samples.

**Determination of Insoluble-Bound Ferulic Acid Content**

Samples were derivatized by the addition of 0.15 mL of pyridine and 0.15 mL of N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA). Capped tubes were placed in a dry bath incubator for 15 minutes and then analyzed by gas chromatography (GC). Briefly, the insoluble-bound ferulic acid content was determined using a gas chromatograph with a flame ionization detector (Agilent 6890). A DB-200 megabore column (ID = 0.53 mm, length = 30 m) was used. Hydrogen gas was used as the carrier gas with a split inlet and a split ratio of 1:10 (split vent flow = 10 mL/min, column flow = 1 mL/min). The initial temperature of the GC was held at 160°C for one minute and then increased to 225°C at a rate of 3°C per minute. The temperature was then held at 225°C for one minute before increasing to 270°C at a rate of 25°C per minute. The temperature was held at 270°C for 1.53 minutes. The area under the curve corresponding to the ferulic acid peak was integrated to calculate the insoluble-bound ferulic acid content. A standard curve for ferulic acid content was calculated using seven standards of ferulic acid ranging in concentration from approximately 50 µg to 1,925 µg of ferulic acid. This allowed maize samples in the range of 100 µg/g to 3,850 µg/g to be analyzed because ferulic acid was extracted from 0.5 g of maize.

Chemicals and reagents including Hexane, acetone, α-amylase, ethyl acetate, and methanol were purchased from Sigma-Aldrich (St. Louis, MO). Sodium hydroxide, sodium
chloride, hydrochloric acid, and pyridine were purchased from Fisher Scientific (Pittsburgh, PA). BSTFA was obtained from Thermo Scientific (Waltham, MA).

**Statistical Analyses**

The overall model used was a split-plot in an RCBD where the whole-plot unit was the field plot from which each hybrid was harvested and the subplot was the 0.5 g of ground maize used for the laboratory tests.

\[ y_{ijk} = \mu + R_i + \alpha_j + \varepsilon_{1ij} + \tau_k + \alpha\tau_{jk} + \varepsilon_{2ijk} \]

Where \( y_{ijk} \) is the insoluble-bound ferulic acid content recorded for the \( i \)th replication, the \( j \)th hybrid genotype, and the \( k \)th protocol, \( i = 1, 2, 3; j = 1, 2, \ldots, 13; k = 1, 2 \)

\( \mu \) is the grand population mean

\( R_i \) is the random effect of the \( i \)th replication, NID(0, \( \sigma_R^2 \))

\( \alpha_j \) is the fixed effect of the \( j \)th hybrid genotype

\( \varepsilon_{1ij} \) is the whole-plot random error term, NID(0, \( \sigma_{\varepsilon_1}^2 \))

\( \tau_k \) is the fixed effect of the \( k \)th protocol used

\( \alpha\tau_{jk} \) is the fixed interaction between the \( j \)th hybrid genotype and the \( k \)th protocol

\( \varepsilon_{2ijk} \) is the subplot random error term, NID(0, \( \sigma_{\varepsilon_2}^2 \))

All statistical analyses were conducted in SAS 9.3 using the MIXED, UNIVARIATE, GPLOT, MEANS, and GLM procedures. The ANOVA as described above was conducted in PROC MIXED. Residuals were obtained from the MIXED procedure. Residuals were then analyzed using the UNIVARIATE procedure to check the assumption of normality. A Brown-
Forsythe modification of the Levene test was used to check the assumption of homogeneous variances among the protocol and genotype treatment combinations.

The interaction between the protocol and the genotypes was examined using slice statements and profile plots. The slice option of the LSMEANS statement in PROC MIXED was used to examine the differences in the protocols for each hybrid genotype. PROC MEANS was used to calculate the mean of each protocol and hybrid genotype combination. These means were then used to produce profile plots in PROC GPLOT to graphically examine the interaction between protocols and hybrid genotypes.

Protocol A lacks the protease digestion step of the protocol reported by Rose et al. (2010). A preliminary study (data not shown) found that this digestion was not needed for typical yellow dent maize. Protocol A yielded consistent results with an average CV of approximately five percent. Therefore, “standard” Protocol B’s results were compared to Protocol A’s results. The expected insoluble-bound ferulic acid content is that obtained by Protocol A. The percent recovery using Protocol B was calculated as

\[
PR_i = 100\% - \left[ \frac{FER_B - FER_A}{FER_A} \times 100\% \right],
\]

where \( FER_B \) is the average insoluble-bound ferulic acid content of the \( i^{th} \) hybrid obtained using Protocol B and \( FER_A \) is the average insoluble-bound ferulic acid content of the \( i^{th} \) hybrid obtained using Protocol A.
Throughput and Cost Efficiency Evaluation

The throughput recorded was the typical number of samples completed each week using Protocol A or Protocol B. Since only 39 harvested grain samples were used in this study but both protocols had a throughput greater than 39 samples per week, technical replicates and ground seed from genetically similar maize material were also analyzed to compute the throughput of both protocols. The throughput of each protocol was recorded for ten weeks. The relative throughput was calculated as

\[
Relative \ Throughput = \frac{n_A}{n_B},
\]

where \( n_A \) is the typical number of samples analyzed per person per week using Protocol A and \( n_B \) is the typical number of samples analyzed per person per week using Protocol B. Weeks eliminated (WE) is the number of additional weeks needed to analyze \( n_A \) samples if Protocol B is being used. WE was calculated as

\[
WE = Relative \ Throughput - 1.
\]

The cost per week incurred due to the use of \( \alpha \)-amylase (CA) was calculated as

\[
CA = \frac{\text{cost}}{\text{volume \ in \ mL}} \times \frac{0.1 \text{mL}}{\text{sample}} \times n_A.
\]

The weekly cost efficiency (WCE) was calculated as

\[
WCE = WE \times Weekly \ Employee \ Salary - CA,
\]

where WCE is the weekly cost efficiency of Protocol A in comparison to Protocol B. Lastly, the overall cost efficiency (CE) of Protocol A in comparison to Protocol B was calculated as

\[
CE = WCE \times \frac{n_T}{n_A},
\]

where \( n_T \) is the total number of samples needed to be analyzed.
RESULTS AND DISCUSSION

Characteristics of a High-Throughput Protocol

The accuracy of measurements, the relative throughput, and cost efficiency were the criteria used to compare the usefulness of two protocols for high-throughput application. A high-throughput protocol must produce accurate results. As there was no swollen starch to inhibit the extraction of the insoluble-bound ferulic acid in Protocol A, the interest was whether “standard” Protocol B could recover as much insoluble-bound ferulic acid as Protocol A when using small sample sizes. A concern was that the swelling of starch when NaOH was introduced could interfere with the extraction of insoluble-bound ferulic acid if small sample sizes were used in Protocol B. The next criteria considered was the relative throughput of both protocols. Lastly, a high-throughput protocol should be cost-efficient. Due to the cost of the $\alpha$-amylase, the reagent costs are greater for Protocol A than they are for Protocol B. Therefore, Protocol A would be a cost-effective alternative only if Protocol B were inaccurate or if its throughput was much higher than that of Protocol B.

Percent Recovery and Predictability of Insoluble-Bound Ferulic Acid When Forgoing Starch Digestion

In all cases, Protocol A recovered more insoluble-bound ferulic acid than Protocol B. In most cases, the difference between the protocols was significant (Table 1.1). Therefore, Protocol B is inaccurate in a high-throughput setting. Even if Protocol B does not extract as much insoluble-bound ferulic acid as Protocol A or other protocols that address the swelling of starch, it might still be useful in some instances. For laboratories lacking incubators or paid labor, Protocol B could still be a high-throughput alternative to using large sized samples if this method
reliably extracts a certain percentage of the insoluble-bound ferulic acid extracted using Protocol A. However, the significant interaction between protocol and hybrid (P < 0.0001) indicates that the amount of insoluble-bound ferulic acid extracted using Protocol B is not indicative of the total amount of insoluble-bound ferulic acid (Fig. 1.2). The rank of several hybrids with regard to ferulic acid concentration changed in Protocol A versus Protocol B (Fig. 1.2). The significance of the protocol effect was hybrid dependent (Table 1.1).

Regressing the amount of insoluble-bound ferulic acid extracted using Protocol A against the amount of insoluble-bound ferulic acid extracted using Protocol B resulted in an $R^2$ value of 0.27 and a correlation coefficient of 0.52 (Fig. 1.3). Since both protocols used subsamples of the same homogenous maize sample to measure the insoluble-bound ferulic acid content, a very large $R^2$ value and a strong correlation coefficient would be needed before Protocol B could reliably be used to predict the actual insoluble-bound ferulic acid concentration.

This lack of predictive ability is further exemplified in Table 1.1. Three hybrids, all with PH207 as a parent, showed nonsignificant differences in the amount of insoluble-bound ferulic acid extracted using Protocols A and B. For the other ten hybrids, the differences between Protocols A and B were significant, but the degree of significance varied. Therefore, genetic factors might influence the swelling of starch when NaOH is introduced. These differences among hybrids can be explained by differences in starch properties, e.g. different starch granule morphology or starch matrix structure among hybrids. This provides an opportunity for further research. Also, inbred PH207 is a member of the Iodent heterotic pool. This raises the question
of whether the starch and its matrix in Iodents have specific properties that restrict the swelling of starch when exposed to NaOH.

The $PR_i$ of insoluble-bound ferulic acid using Protocol B is worrisome if small sample sizes are used. The $PR_i$ of insoluble-bound ferulic acid in Protocol B ranged from 61.9% to 92.6% with an average of 79% (Table 1.1). All replications within a hybrid tended to exhibit similar percent recoveries, although there were a few exceptions (Fig. 1.4). The varying percent recoveries seen in hybrids B73×PHG47, LH1×PHG47, and PHJ40×LH123HT may be attributed to variation in starch gelatinization between field replicates. Starch properties are highly influenced by environmental factors (Tester and Karkalas, 2001), and there may have been environmental variability between the field replications which caused the replicates within these three hybrids to have variable percent recoveries. Therefore, Protocol B or any small-scale protocol that does not contain a starch pre-digestion step cannot be used to reliably determine the insoluble-bound ferulic acid content in maize grain.

**Throughput of Extraction Protocols**

Upon the introduction of NaOH into the vial, the starch visibly swelled and inhibited the extraction of insoluble-bound ferulic acid (Fig. 1.1A). This swelling is a phenomenon that has been reported by others in the literature and is not unique to this study (Sivasankar, 2002; Sosulski et al., 1982). The use of small sample sizes (0.5 g, 2 mL of 2 M NaOH) made it possible to increase the number of samples a person can handle per week. However, this was only true if the starch swelling could be inhibited. If the starch swelled, not only was the total amount of extracted insoluble-bound ferulic acid considerably lower, both the mixing and
extraction of the ethyl acetate layer was physically very difficult (Fig. 1.1 B and C). Also, centrifugation times had to be increased to separate the layers in Protocol B. Due to these limitations, typically 72 samples per person per week could be analyzed using Protocol B. Thus, the ability to handle more samples concurrently was almost negated by the consequences (i.e. more difficult extraction and difficulty in mixing and separating layers) of using small sample sizes in Protocol B.

Protocol A enabled the extraction and analysis of about 180 samples per person per week. The key features of Protocol A that contributed to its usefulness in a high-throughput setting were 1. the ability to work with many small samples concurrently, 2. the pre-digestion of starch, and 3. the possibility to digest the starch during non-working hours (e.g. overnight). Pre-digestion of the starch was a key feature of Protocol A because it inhibited starch gelatinization from occurring. Pre-digestion allowed for relatively easy mixing of the ethyl acetate, separation of the layers, and successful extraction of more insoluble-bound ferulic acid because swollen starch was no longer interfering with any of these procedures.

Since the starch digestion step of Protocol A can be completed during non-working hours, the digestion did not add any additional man-hours to the overall extraction procedure. The only additional time spent during the extraction in Protocol A in comparison to the extraction in Protocol B was preparatory work for the starch digestion (approximately 15 minutes per 72 samples) and a cool-down period of approximately one hour following the digestion. During cool-down, other laboratory duties including the preparation of reagents and the labeling of vials could be completed. Since the extractions were much easier and quicker to
complete, fewer working man-hours are required for Protocol A than Protocol B. However, if
the starch digestion step had not been relatively easy to implement or had required more man-
hours to conduct, this could have been detrimental to the throughput and cost efficiency of
Protocol A. Protocols using an acidification technique, such as that outlined in Classen et al.
(1990) and Sosulski et al. (1982), are effective in reducing the interference of swollen,
gelatinized starch during the extraction of insoluble-bound hydroxycinnamic acids like ferulic
acid. However, these procedures are also more laborious and time consuming than the α-
amylase digestion applied in this study.

Cost Efficiency

Typical breeding programs require the analysis of numerous samples. Therefore, high-
throughput methods are necessary to provide results in a timely fashion, to reduce labor costs per
sample, and to increase cost efficiency. One of the major foreseeable criticisms of Protocol A is
the additional cost of α-amylase. To justify the use of α-amylase, the cost of the α-amylase
must be outweighed by the savings in labor costs. By using an α-amylase digestion in Protocol
A, the throughput increased 2.5 fold while using the same amount of effort as in Protocol B. At
$79.10 per 50mL bottle (Sigma-Aldrich, 2015), the cost of using α-amylase is about $0.16 per
sample, or $28.80 per person per week (assuming 180 samples per week). Furthermore, the
employee hourly salary at which the use of α-amylase becomes cost-efficient is $0.48 per hour.
By this logic, the cost of α-amylase needed is offset by the reduction in paid labor hours. Also,
in preliminary work, Protocol A yielded consistent results with an average CV of approximately
five percent. Thus, Protocol A is a consistent, accurate, cost-efficient, and high-throughput
method.
The relative throughput, WE, WCE, and CE are dependent upon the researcher and available laboratory space. For instance, using Protocol A, up to 216 samples per person per week could be extracted and derivatized in this study. However, it was not realistic to extract this many samples because the available facilities could only feasibly analyze 200 samples, including standards, on the GC each week. There were also many aspects of the laboratory space which made the observed \( n_A \) larger than the \( n_A \) other laboratories might observe. Figure 1.5 shows several cost efficiency outcomes given different employee salaries as well as \( n_A \) and \( n_T \) values. Interestingly, for a yearly employee salary of $15,000 and \( n_A \) equal to 80, the WCE and CE are both still positive. As the yearly employee salary and \( n_A \) increase, so does CE.

Given that Protocol A is accurate, consistent, high-throughput, and cost-efficient, even when the yearly employee salary and \( n_A \) are relatively small, it is recommended that Protocol A or a similar protocol utilizing a starch digestion step be used for the extraction of insoluble-bound ferulic acid. Protocol A is versatile and can easily be adapted to study the effect of ferulic acid on the nutritional value of maize based foods as well as on the resistance to maize diseases and pests. Also, given the inaccuracy of Protocol B and its inability to determine the actual concentration of insoluble-bound ferulic acid, Protocol B and similar protocols lacking a starch digestion step should only be used if \( n_T \) is small and the sample size can be increased so as to ameliorate the negative effects of gelatinized starch.
### Table 1.1. Mean ferulic acid content and percent recovery obtained for 13 maize hybrids using two extraction protocols

<table>
<thead>
<tr>
<th>Hybrid</th>
<th>Protocol A</th>
<th>Protocol B</th>
<th>p-Value†</th>
<th>PR‡ %</th>
</tr>
</thead>
<tbody>
<tr>
<td>B73 × MO17</td>
<td>2330.4 µg/g</td>
<td>1798.8 µg/g</td>
<td>&lt;0.001</td>
<td>77.2</td>
</tr>
<tr>
<td>B73 × PHG47</td>
<td>1824.5 µg/g</td>
<td>1130.1 µg/g</td>
<td>&lt;0.001</td>
<td>61.9</td>
</tr>
<tr>
<td>B73 × PHG84</td>
<td>2007.7 µg/g</td>
<td>1773.6 µg/g</td>
<td>0.045</td>
<td>88.3</td>
</tr>
<tr>
<td>LH1 × MO17</td>
<td>2519.0 µg/g</td>
<td>1581.6 µg/g</td>
<td>&lt;0.001</td>
<td>62.8</td>
</tr>
<tr>
<td>LH1 × PHG47</td>
<td>2230.9 µg/g</td>
<td>1616.9 µg/g</td>
<td>&lt;0.001</td>
<td>72.5</td>
</tr>
<tr>
<td>PH207 × PHG47</td>
<td>1606.1 µg/g</td>
<td>1407.1 µg/g</td>
<td>0.086</td>
<td>87.6</td>
</tr>
<tr>
<td>PH207 × PHG84</td>
<td>2020.5 µg/g</td>
<td>1871.2 µg/g</td>
<td>0.192</td>
<td>92.6</td>
</tr>
<tr>
<td>PHG39 × PHZ51</td>
<td>2117.9 µg/g</td>
<td>1725.7 µg/g</td>
<td>0.002</td>
<td>81.5</td>
</tr>
<tr>
<td>PHJ40 × LH123HT</td>
<td>2140.5 µg/g</td>
<td>1701.0 µg/g</td>
<td>&lt;0.001</td>
<td>79.5</td>
</tr>
<tr>
<td>PHJ40 × LH82</td>
<td>1794.5 µg/g</td>
<td>1345.7 µg/g</td>
<td>&lt;0.001</td>
<td>75.0</td>
</tr>
<tr>
<td>PHJ40 × MO17</td>
<td>1768.9 µg/g</td>
<td>1335.2 µg/g</td>
<td>&lt;0.001</td>
<td>75.5</td>
</tr>
<tr>
<td>PHJ40 × PH207</td>
<td>1619.4 µg/g</td>
<td>1448.8 µg/g</td>
<td>0.138</td>
<td>89.5</td>
</tr>
<tr>
<td>PHJ40 × PHG84</td>
<td>1657.8 µg/g</td>
<td>1370.4 µg/g</td>
<td>0.016</td>
<td>82.7</td>
</tr>
<tr>
<td>Mean</td>
<td>1972.1 µg/g</td>
<td>1546.6 µg/g</td>
<td></td>
<td>79.0</td>
</tr>
</tbody>
</table>

LSD (Between protocols within hybrid)§ 228.8

LSD (Between hybrids with protocol) § 242.1

† The p-value associated with the protocol effect within each particular hybrid.
‡ The average percent recovery for each hybrid.
§ LSD values were calculated using $\alpha = 0.05$. 
Figure 1.1. Photos of samples at various stages during extraction in Protocol B. Two samples from Protocol B following alkali hydrolysis for two hours are shown in (A). As the reaction takes place, the sample should turn orange or an orange-brown. As indicated by the arrows, portions of the sample could not be mixed with the NaOH, thereby preventing the reaction from taking place. The difficulty with which HCl, NaCl, and ethyl acetate were mixed following the alkali hydrolysis is exemplified in (B). When neutralized by HCl, the sample should turn from dark orange (or orange-brown) to white. In Protocol A, this took place with ease with little mixing required. However, in Protocol B, the swollen starch was so thick in some samples that only the top portion of the gelatinized mass would be exposed to HCl, even after vortexing. Occasionally, the separation of the ethyl acetate layer after centrifugation was acceptable (C). However, the gelatinous mass in the bottom of the vial was unable to be mixed well (as indicated by the yellow-orange color which should be white). Furthermore, there should only be a small white maize pellet at the bottom of the tube, but the starch was swollen to varying degrees in the two vials shown.
Figure 1.2 Interaction plot between hybrids and protocol. Plot showing the interaction between 13 maize hybrids and two high-throughput protocols (Protocols A and B) applied to extract insoluble-bound ferulic acid from ground grain samples.

Figure 1.3 Scatterplot between values obtained using protocols A and B. Scatter plot showing the relationship between the amount of insoluble-bound ferulic acid extracted applying two high-throughput protocols (Protocols A and B).
Figure 1.4 **Boxplot of percent recovery by hybrid.** The boxplots of percent recovery of insoluble-bound ferulic acid using Protocol B for each of the 13 maize hybrids used in this study.

Figure 1.5. **Cost efficiencies for different scenarios.** The cost efficiency will differ on a case by case basis, but Protocol A is cost efficient in all scenarios plotted above. Two different values of \( n_T \) were evaluated, those being 1,000 and 5,000 samples. The salaries shown are potential annual employee salaries. To calculate the values above, \( n_B \) was assumed to be 72 and the additional cost per sample due to the use of \( \alpha \)-amylase was assumed to be $0.16. In the most extreme scenario, the use of Protocol A saved approximately $50,000.
WORKS CITED


Rumpagapom, P. 2011. Structural features of cereal bran arabinoxylans related to colon fermentation rate. PhD. Purdue University, West Lafeyette, IN.


CHAPTER 2: IDENTIFICATION OF NUTRITIONAL COMPOUNDS OF INTEREST IN TYPICAL YELLOW DENT MAIZE

ABSTRACT

Numerous studies have examined the concentration of various nutritional compounds in maize. However, these studies have typically focused upon maize inbreds or hybrids which are not commonly grown by farmers in the United States. Furthermore, due to the overall lack of high-throughput methods, many of these studies have examined too few genotypes to draw conclusive answers about the typical variability of nutritional compounds in maize grain, such as phenolic acids. In this study, a genetically and phenotypically diverse set of maize hybrids and inbreds was evaluated for its variability and phytochemical content. Insoluble-bound phenolics, soluble phenolics, total protein, unsaturated fatty acids, and tocopherols were evaluated in this study. Of these compounds, the insoluble-bound and soluble phenolic acids, a subclass of the phenolics, were the only compounds which exhibited both high variability and were present at substantial concentrations. Therefore, only the phenolic acids were selected for further study in regards to their usefulness in future plant breeding programs which focus on value-added nutritional traits.
INTRODUCTION

Over the last decade, consumers have become increasingly interested in the natural nutritional value of their food products (Helbert Almeida, Kellogg’s Company, Personal Communication, 2013). Given the high obesity rates in affluent nations, an increase in aging-related diseases in developed nations, and the concern that processed foods are less nutritious, this increased interest is not surprising. Therefore, it has become important for food companies to consider not only their food products’ caloric and vitamin content but also the content of beneficial phytochemicals which occur naturally in their food products.

In discussions with industry leaders, a set of phytochemicals that are marketable and that exhibit positive health benefits were identified for study. These phytochemicals are phenolics, total protein content, and tocopherols. Also, since the aliquot needed for the analysis of the unsaturated fatty acids was the same aliquot used for the tocopherols and unsaturated fatty acids also have health properties, unsaturated fatty acids were also analyzed. The ultimate goal of this research was to identify which of these phytochemicals might be useful in future plant breeding programs. This study serves as a preliminary analysis to that endeavor by characterizing the variability and average content of these phytochemicals in typical Midwestern maize germplasm.

All of these phytochemicals are beneficial to human health. Diets high in protein, for instance, tend to result in satiety and weight maintenance. Specifically, as satiety is maintained for longer periods of time, people tend to eat less ad libitum (Westerterp-Plantenga et al., 2009). This contributes to a lower overall calorie intake and a decreased chance of developing an obesity-related disease. Unsaturated fatty acids play important roles in blood pressure
regulation, cell membrane integrity, and improving cholesterol levels. Some monounsaturated fatty acids, like oleic acid, are reported to decrease blood pressure levels (Teres et al., 2008). Linoleic acid, a polyunsaturated fatty acid, is another healthy fatty acid which is an important component of cell membranes (Simopoulos, 1999). As linoleic cannot be synthesized by humans but must be consumed for good health, it is considered an essential fatty acid. Given that the predominant fatty acids in maize grain oil are oleic acid and linoleic acid (Baur Jr and Brown, 1945), there may be some potential for creating maize-based food products with healthy levels of oleic acid and linoleic acid.

In contrast to the health benefits of protein and unsaturated fatty acids, phenolics and tocopherols exhibit antioxidant activity, prebiotic activity, and other chemopreventive properties. Along with the tocotrienols, the tocopherols comprise Vitamin E. Alpha, gamma, and delta tocopherol are typically found in maize. Beta tocopherol is rarely present (Weber, 1987). Most studies have focused on alpha tocopherol because it is the most efficient antioxidant of the three tocopherols present in maize (Sies and Stahl, 1995). Specifically, alpha tocopherol is noted for reducing peroxyl radicals, such as those from polyunsaturated fatty acids or lipoproteins (Sies and Stahl, 1995).

Like the tocopherols, many of the soluble phenolics exhibit antioxidant activity in vitro, but they are more likely to act as prebiotics in vivo by activating gene cascades involved in immunoprotection and chemoprevention (Fresco et al., 2006; Kanski et al., 2002; Scapagnini et al., 2011). The predominant phenolics in maize are the hydroxycinnamic acids, namely ferulic acid and p-coumaric acid (Adam et al., 2002; Adom and Liu, 2002). In cereals, these are almost
exclusively found in the insoluble-bound form, bound to the cell wall via ester bonds. In their insoluble-bound form, hydroxycinnamic acids cross-link arabinobxylan fibers. This complex network of cross-linked arabinobxylan fibers slows the fermentation of the arabinobxylans in the human gut. This enables them to act as prebiotics by creating an environment which is beneficial for gut bacteria (including *Lactobacillus* and *Bifidobacterium* species). Furthermore, due to the extensive cross-linking of ferulic acid in maize fiber, the fermentation of the arabinobxylans slows such that the beneficial properties of the arabinobxylans can be recognized without the negative effects of extensive gas production (Rumpagapom, 2011).

The overall objective of this study was to evaluate a diverse set of U.S. maize inbreds and hybrids for their phytochemical content and variability. Specifically, sample means, standard errors, and ranges were calculated for the phenolics, tocopherols, total protein, and unsaturated fatty acids in maize. Correlations between phytochemicals which exhibited appreciable means and ranges in content were also calculated.
MATERIALS AND METHODS

Plant Materials

Twelve maize inbreds (Table 2.1) which were identified as being genetically representative of the germplasm grown in the United States Midwest were crossed using a diallel design to create 66 F_1 hybrids (Hauck et al., 2014). Five of these inbreds (B73, LH1, PHG39, PHJ40, and 4676A) were derived from various cycles of the Iowa Stiff Stalk Synthetic (SSS) heterotic pool. The other seven inbreds (LH123HT, LH82, Mo17, PH207, PHG47, PHG84, and PHZ51) are all members of the Non-Stiff Stalk (NSS) heterotic pool.

All inbreds and hybrids (N=78 entries) were grown at the University of Illinois Crop Sciences Research and Education Center in Urbana, IL for three years (2009-2011). The 78 entries were grown in a resolvable incomplete block design with three replications for each of the three years. Bags containing 454 g (one pound) of harvested grain from each plot were placed in cold storage for nutritional analysis. When nutritional analyses began, 100 g from each of these bags was ground to a fine powder using a Foss Cyclone Mill (1 mm mesh screen) for use in the wet-lab chemistry and Fourier Transform Infrared Spectroscopy (FTIR) analyses.

Subset Determination

All 12 parental inbreds were used in this study. Since the available resources allowed for the analysis of 25 entries total, (25 entries × 3 years × 3 reps = 225 observations), 13 hybrids were selected. The 35 SSS×NSS hybrids in Macke, et al. (2016) were ranked by their dry milling efficiency. The top five and bottom five hybrids of this ranking were selected. Also, three additional hybrids (PH207×PHG47, PH207×PHG84, and PHJ40×PH207) were added to
this germplasm set so as to include hybrids in which PH207 was a parent and also to include two intra-heterotic pool crosses.

**Determination of Fatty Acid Content**

Ten random subsamples from among the 225 whole kernel maize samples and their corresponding ground samples were selected for calibration of the Nuclear Magnetic Resonance (NMR) machine in order to determine the total oil content. To determine the total oil content, ground samples underwent a hexane wash. The hexane layer was extracted, placed in a separate vial, and evaporated, leaving only the oil extracted for a particular sample in the bottom of the vial. All vials were weighed prior to extraction and after evaporation. The difference in the mass is equal to the mass of oil extracted for that particular sample. Dividing this mass by the mass of ground maize used in this extraction gives the total oil concentration in g oil / g maize. Three technical replicates were used for each of the ten samples, and the average concentration was calculated for each of the technical replicates. These values and their corresponding whole kernel samples were used to calibrate a prediction curve using the NMR. The calibration was performed by Dow AgroSciences. NMR was used to measure the total oil concentration of each of the 225 samples in this study.

Approximately 0.5 g of each sample was placed in a vial, and 3 mL of hexane was added. The vials were capped, placed in a sonicator at 75°C for 10 min, and then placed on a horizontal shaker for 10 min. Each vial was then vortexed before being centrifuged at 1,500 rpm for 5 min. The top hexane layer was extracted and placed in a clean vial. The hexane wash was repeated two more times. The hexane extract was dried using a centrifugal evaporator. The oil remaining
in the vial was reconstituted in 1 mL of heptane, and 300 μL of the heptane/oil mixture was placed in a Gas Chromatography (GC) vial and derivitized using 20 μL of sodium methoxide. The samples then underwent a Fatty Acid Methyl Ester (FAME) analysis.

The percentage of total oil content was recorded for each fatty acid. In order to calculate the concentration of each fatty acid, the following equation was used.

\[ FA_{ij} = \frac{PTO_{ij} \times TO_j}{m_j} \times \frac{1000 \text{ mg}}{1 \text{ g}} \]

where \( FA_{ij} \) is the concentration of a particular fatty acid in mg / g of maize

\( PTO_{ij} \) is the percentage of total oil content for the \( i^{th} \) fatty acid and the \( j^{th} \) observation,

\( TO_j \) is the total oil content for the \( j^{th} \) observation,

and \( m_j \) is the recorded mass of the \( j^{th} \) observation.

**Determination of Tocopherol Content**

Following the removal of 300 μL from the heptane aliquot for FAME analysis, the heptane/oil mixture was then evaporated to dryness. Approximately 70% of the originally extracted tocopherols remained in the oil extract. These extracts were provided to Dow AgroSciences’ analytical scientists for the analysis of tocopherols. Gas chromatography (GC) was used for quantification.

The total mass of tocopherols (μg) remaining was reported by the GC system. In order to calculate the concentration in the original sample, the following equation was used:

\[ T_{ij} = \frac{T'_{ij}}{0.7 \times m_j} \times \frac{1 \times 10^6 \text{ μg}}{g} \]
where $T_{ij}$ is the actual tocopherol content of the $i^{th}$ tocopherol and the $j^{th}$ sample and $T_{ij}'$ is the reported content of the $i^{th}$ tocopherol in the $j^{th}$ sample before the correction.

_Determination of Soluble Phenolic Acid Content_

Following the hexane wash, the maize pellet was dried using a centrifugal evaporator. Then, 3 mL of 70% acetone was added to each vial. Vials were placed in a sonicator for 10 min at 75°C and then placed on a horizontal shaker for 10 min. Vials were then vortexed and centrifuged at 1,500 rpm for 5 min. The top acetone layer was removed and placed in a separate vial. The acetone wash was completed two more times for a total of three acetone washes. The extracted acetone aliquot was then evaporated to dryness using a centrifugal evaporator. Soluble extracts were reconstituted in 2 mL of methanol, vortexed, and filtered using syringe filters. One mL of the filtered solution was then placed in an Ultra Performance Liquid Chromatography (UPLC) for analysis.

Standards of the typical hydroxycinnamic acids in grains (ferulic acid, p-coumaric acid, and sinapic acid) as well as cinnamic acid were first analyzed using UPLC to record the retention time of each of these compounds and to build a standard curve for the calculation of concentrations in the maize samples. The samples were then analyzed using UPLC, and the standard curve was used to calculate the concentration in each of the samples using the equation

$$S_{ij} = \frac{2 \times S_{ij}'}{m_j}$$

where $S_{ij}$ is the content of the $i^{th}$ soluble phenolic acid in the $j^{th}$ observation and
$S_{ij}'$ is the recorded mass of the $i^{th}$ soluble phenolic acid in the $j^{th}$ observation as calculated from the standard curve.

**Determination of Insoluble-Bound Phenolic Acid Content**

The determination of the insoluble-bound phenolic acid content took place using Protocol A as described in Butts-Wilmsmeyer and Bohn (2016) with the exception that standards of p-coumaric acid, cinnamic acid, and sinapic acid were also used in creating the standard curve. This protocol is an adaptation of the protocol outlined in Rose et al. (2009). However, preliminary work indicated that the additional protease digestion step used by Rose et al. (2009) was not needed for these materials (data not shown). All laboratory chemicals and standards were provided by Dow AgroSciences.

**Statistical Analyses**

The overall model used for the experiment was a Randomized Complete Block Design (RCBD).

$$y_{ijk} = \mu + Y_i + R_{(i)j} + G_k + YG_{ik} + \varepsilon_{ijk}$$

Where $y_{ijk}$ is the observed phenotypic value corresponding to the $i^{th}$ year, the $j^{th}$ replication within the $i^{th}$ year, and the $k^{th}$ genotype,

$\mu$ is the grand population mean

$Y_i$ is the random effect of the $i^{th}$ year

$R_{(i)j}$ is the random effect of the field $j^{th}$ field replication nested within the $i^{th}$ year

$G_k$ is the fixed effect of the $k^{th}$ genotype

$YG_{ik}$ is the random interaction between the $i^{th}$ year and the $k^{th}$ genotype
\( \varepsilon_{ijk} \) is the random error term associated with the \( i^{th} \) year, the \( j^{th} \) replication within the \( i^{th} \) year, and the \( k^{th} \) genotype, \( \text{NID}(0, \sigma^2) \).

This model was run separately for the hybrids and inbreds in SAS PROC MIXED (version 9.3) due to confounding of the generation effect with the physical location of the hybrid and inbred blocks in the field. Only phytochemicals which had appreciable means and variances were analyzed using the statistical model to verify that there was a significant genotypic effect for these phytochemicals. The assumptions of normality and homogenous variances were checked by conducting Shapiro-Wilk’s test of normality on the residuals in PROC UNIVARIATE and by using the Brown-Forsythe modification of the Levene test in the MEANS option of PROC GLM. If a highly significant interaction existed and the average concentration was high enough to warrant further investigation of the compound, multi-degree of freedom contrast statements were used to examine the genotype effect in each year.
RESULTS AND DISCUSSION

Precision of Analytical Chemistry Analyses

The standard errors reported in Table 2.2 were always very small in comparison to the average of the trait being measured. This is indicative of very precise analytical chemistry techniques. However, this also renders the P-value associated with the genotype term in the model virtually useless because even the slightest differences among hybrid means could result in a significant P-value. These “significant differences” may not be useful to a breeder, however. Therefore, better indicators of the usefulness of a trait are the average concentration and the range in concentration for that particular trait.

Unsaturated Fatty Acid Content

Neither the oleic acid content nor the linoleic acid content displayed a large mean or range. This held true for both the hybrids and inbreds (Table 2.2). Linoleic acid was more prevalent than oleic acid, and these results are similar to that reported by Baur and Brown (1945). Hybrids tended to accumulate more oleic acid and linoleic acid than the inbreds, but the difference in hybrid and inbred concentrations was negligible. Given that corn yield is chiefly governed by starch content, it is not surprising that there is little variability in the unsaturated fatty acid content and composition among the hybrids and inbreds used in this study. Breeders have traditionally prioritized yield, and the yield of high oil lines tends to be reduced (Alexander, 1988).

Furthermore, fatty acids are almost exclusively found in the germ. The germ is removed during dry milling, meaning the small amount of oleic acid and linoleic acid present in the whole
kernel will not be present in the final processed food product. Additionally, because plant oils can become rancid, oil-containing foods tend to have relatively short shelf lives (Egesel et al., 2003). These characteristics make the unsaturated fatty acid content impractical for further study.

**Tocopherols**

As expected, gamma tocopherol was the most prevalent tocopherol, followed by alpha tocopherol and then delta tocopherol. The incredibly small ranges observed in both alpha and delta tocopherol and the fact that the concentrations of these two phytochemicals was so small that they were barely detectable using GC make these two compounds unlikely candidates for continued study. However, gamma tocopherol did exhibit some variability which might have potential for continued study and selection. Of the three tocopherols found in maize, gamma tocopherol was the only tocopherol which exhibited a constantly detectable mean concentration, as well. The issue is that the tocopherols are also located almost exclusively in the germ (Egesel et al., 2003), and these will be removed by future processing. Additionally, gamma and delta tocopherol have much lower antioxidant activity than alpha tocopherol. Since gamma tocopherol is by far the most prevalent of the tocopherols and alpha tocopherol was only found in small quantities in the materials examined, this set of germplasm is not well suited for improvement of the tocopherols. Other sets of germplasm, such as those described in later paragraphs, display more variability and a higher concentration of alpha tocopherol than typical yellow dent maize from the Midwestern United States.
Interestingly, the tocopherol concentration reported here, the tocopherol concentration reported in Egesel, et al. (2003), and the tocopherol concentration reported in Lipka, et al. (2013) were considerably different. This can be explained by the different germplasm sets used in each study. Egesel, et al. (2003) chose high-oil lines, high tocopherol-containing lines, or inbreds which were known to have either a high or low ratio of alpha to gamma tocopherol. In Lipka, et al. (2013), less extreme phenotypes were selected for study, but more lines from different regions of the country were also included in that study. Specifically, 281 maize lines which collectively represented a significant portion of the variation in both temperate and tropical maize breeding programs were used. This most likely accounts for why the results of this study more closely matched those of Lipka, et al. (2013) than Egesel, et al. (2003), but the range observed in Lipka, et al. (2013) was considerably greater than that presented here. It also indicates that it may be possible to improve the tocopherol content of some maize lines, but the maize lines which are grown in the U.S. Midwest do not appear to be well suited for use in such a breeding program. This claim can be made with confidence because this is the largest study to the author’s knowledge which used materials solely from the U.S. Midwest.

Protein

The total protein content of hybrids and inbreds were found to be approximately 8.8% and 11.6%, respectively. These figures correspond very closely to the typical average of 8-11% given in Alexander (1988). Also, the range in the average protein content is quite small (Table 2.2). Grain yield decreases if protein content is increased beyond 11 to 12% (Alexander, 1988). As in the case of the unsaturated fatty acids, most breeding efforts have focused primarily on the improvement of grain yield. It also appears that breeders have selected for a lower protein
content and a higher starch content as they have selected for high-yielding maize crops. One portion of the DuPont Pioneer Hi-Bred Era study found that protein content decreased from approximately 13.4% in 1920 to the lower percentages seen today (Scott et al., 2006). Therefore, the protein content and lack of range in protein content observed in this study are not surprising. Due to the extremely small range in phenotypes (Table 2.2), protein was not selected for continued study in this dissertation research.

**Soluble and Insoluble-Bound Phenolics**

Two soluble phenolics were consistently detected using UPLC: ferulic acid and p-coumaric acid. Soluble cinnamic acid, a carboxylic acid, was also commonly detected using UPLC. Soluble ferulic acid and p-coumaric acid were barely detectable, the average concentration of these compounds in hybrids being 1.44 μg/g and 1.65 μg/g, respectively. In addition to the low mean concentration of both soluble ferulic acid and soluble p-coumaric acid, these compounds exhibited very little range in concentrations among either inbreds or hybrids (Table 2.2). Conversely, soluble cinnamic acid was consistently more prevalent than either of the detected phenolic acids (Table 2.2). The average soluble cinnamic acid concentration was 67.09 μg/g in the hybrids and 78.89 μg/g in the inbreds. Additionally, the range in soluble cinnamic acid content was relatively large (Table 2.2).

Two insoluble-bound hydroxycinnamic acids were consistently detected using GC, those being ferulic acid and p-coumaric acid. Insoluble-bound ferulic acid was the most prevalent of all the phenolics. The average concentration of insoluble-bound ferulic acid in hybrids and inbreds was 1,950.29 μg/g and 1,952.49 μg/g, respectively. In regards to insoluble-bound p-
coumaric acid, the average concentration was 176.19 μg / g in the hybrids and 222.94 μg / g in the inbreds. Additionally, the range observed in the insoluble-bound phenolics content was the highest of all compounds examined in this research (Table 2.2).

These results correspond closely to those reported in other studies (Yadav et al., 2007). Of the phenolic acids detected, all were hydroxycinnamic acids. Of these, ferulic acid was the most prevalent, followed by p-coumaric acid. Most of the hydroxycinnamic acids were in the insoluble-bound state, as in Adom and Liu (2002). Cinnamic acid was primarily found in the soluble state, however. The reason for this could be that cinnamic acid is a precursor to p-coumaric acid, which is itself a precursor to ferulic acid. Cinnamic acid-4 hydroxylase acts on cinnamic acid, producing p-coumaric acid. p-Coumaric acid is then acted upon by p-coumaric hydroxylase and then o-methyltransferase to produce ferulic acid. p-Coumaric acid and ferulic acid most likely are synthesized prior to esterification to arabinoxylans (Vance, et al., 1980). Cinnamic acid is stored in the vacuoles of plants (Pandey and Rizvi, 2009) prior to being modified in, most likely, the Golgi bodies. It also appears that hydroxycinnamic acids are esterified to arabinoxylans in the Golgi bodies (Fry, et al. 2000). It is possible that the soluble cinnamic acid found in this study was the precursor to more prevalent hydroxycinnamic acids.

Due to their high concentrations and large ranges, the insoluble-bound hydroxycinnamic acids were selected for further study. Furthermore, there is some evidence that thermal stresses encountered in processing may free insoluble-bound hydroxycinnamic acids from the arabinoxylans of the cell wall (Dewanto et al., 2002). If this observation holds true during the production of ready-to-eat breakfast cereals, then it is possible that the concentration of the
soluble ferulic acid and p-coumaric acid may increase while the concentration of the insoluble-bound hydroxycinnamic acids decreases. It is also possible that processing and thermal stresses may degrade the hydroxycinnamic acids, leading to higher concentrations of cinnamic acid. Therefore, both soluble and insoluble-bound hydroxycinnamic acids as well as cinnamic acid were selected for continued study.
### Table 2.1. Inbred maize parents used and their background

<table>
<thead>
<tr>
<th>Line</th>
<th>Group</th>
<th>Assignee</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>B73</td>
<td>SSS</td>
<td>None (Public)</td>
<td>Iowa Stiff Stalk Synthetic</td>
</tr>
<tr>
<td>LH1</td>
<td>SSS</td>
<td>Holden Foundation Seeds</td>
<td>Iowa Stiff Stalk Synthetic; B37 type</td>
</tr>
<tr>
<td>PHG39</td>
<td>SSS</td>
<td>Pioneer Hi-Bred International</td>
<td>Maiz Amargo/Iowa Stiff Stalk Synthetic; B37/B14 type</td>
</tr>
<tr>
<td>PHJ40</td>
<td>SSS</td>
<td>Pioneer Hi-Bred International</td>
<td>Iowa Stiff Stalk Synthetic</td>
</tr>
<tr>
<td>4676A</td>
<td>SSS</td>
<td>Dekalb Genetics Corporation</td>
<td>1067-1 / B-Line Composite</td>
</tr>
<tr>
<td>LH123HT</td>
<td>M</td>
<td>Holden Foundation Seeds</td>
<td>Pioneer Hybrid 3535</td>
</tr>
<tr>
<td>LH82</td>
<td>M</td>
<td>Holden Foundation Seeds</td>
<td>Krug /W153</td>
</tr>
<tr>
<td>Mo17</td>
<td>M</td>
<td>None (Public)</td>
<td>Lancaster</td>
</tr>
<tr>
<td>PH207</td>
<td>M</td>
<td>Pioneer Hi-Bred International</td>
<td>Iodent/Long Ear OPV/Minn13</td>
</tr>
<tr>
<td>PHG47</td>
<td>M</td>
<td>Pioneer Hi-Bred International</td>
<td>Oh43/Iodent*Wf9/MKSDTA C10 Synthetic</td>
</tr>
<tr>
<td>PHG84</td>
<td>M</td>
<td>Pioneer Hi-Bred International</td>
<td>Oh07-Midland/Minn13/Iodent/Reid YD/Osterland   YD/Lancaster/Pioneer Female Composite OPV</td>
</tr>
<tr>
<td>PHZ51</td>
<td>M</td>
<td>Pioneer Hi-Bred International</td>
<td>Minn13/Iodent/Reid YD/Osterland   YD/Lancaster/South US Land Race Synthetic/Funks G4949/ Midland</td>
</tr>
</tbody>
</table>
Table 2.2. Summary statistics of compounds of interest by generation. The minimum, maximum, and average values reported are based off of the LSMEANS of the hybrids and inbreds.

<table>
<thead>
<tr>
<th></th>
<th>Unsaturated Fatty Acids</th>
<th>Tocopherols</th>
<th>Protein</th>
<th>Soluble Phenolics and Cinnamic Acid</th>
<th>Insoluble-Bound Phenolics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mg/g</td>
<td>µg/g</td>
<td>µg/g</td>
<td>%</td>
<td>µg/g</td>
</tr>
<tr>
<td>HYBRIDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>4.78</td>
<td>11.57</td>
<td>5.42</td>
<td>21.74</td>
<td>8.77</td>
</tr>
<tr>
<td>Min</td>
<td>4.17</td>
<td>10.44</td>
<td>3.35</td>
<td>10.70</td>
<td>8.03</td>
</tr>
<tr>
<td>Max</td>
<td>5.35</td>
<td>12.92</td>
<td>6.12</td>
<td>38.18</td>
<td>9.50</td>
</tr>
<tr>
<td>Std. Error†</td>
<td>0.25</td>
<td>0.38</td>
<td>0.87</td>
<td>3.53</td>
<td>0.28</td>
</tr>
<tr>
<td>P-Value‡</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>INBREDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>4.00</td>
<td>10.90</td>
<td>5.70</td>
<td>18.68</td>
<td>11.60</td>
</tr>
<tr>
<td>Min</td>
<td>2.39</td>
<td>8.84</td>
<td>2.57</td>
<td>4.13</td>
<td>9.92</td>
</tr>
<tr>
<td>Max</td>
<td>5.58</td>
<td>13.01</td>
<td>10.03</td>
<td>34.01</td>
<td>12.57</td>
</tr>
<tr>
<td>Std. Error†</td>
<td>0.17</td>
<td>0.30</td>
<td>1.05</td>
<td>4.45</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>-0.36</td>
<td>-0.72</td>
<td>-2.46</td>
<td>-12.39</td>
<td>-1.23</td>
</tr>
</tbody>
</table>

†The standard error reported is the standard error of the difference between two genotypic means. Due to unequal replication, the range of standard errors between the means is reported for the inbreds.

‡The P-values reported are those associated with the genotype effect in the mixed model. Note that P-values were not calculated for inbreds. The reason for this is that numerous inbred whole kernel samples were missing in 2010 and 2011 due to herbicide carryover damage in 2010 and drought conditions during flowering in 2011. Since these whole kernel samples could not be analyzed, this seriously affected model results.
Table 2.3. Mean Squares and P-Values of model effects for maize hybrid traits which exhibited appreciable means and ranges

<table>
<thead>
<tr>
<th>Effect</th>
<th>Insoluble-Bound Ferulic Acid</th>
<th>Insoluble-Bound p-Coumaric Acid</th>
<th>Soluble Cinnamic Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Square</td>
<td>P-Value</td>
<td>Mean Square</td>
</tr>
<tr>
<td>Year</td>
<td>91,088</td>
<td>0.31</td>
<td>6,994</td>
</tr>
<tr>
<td>Rep(Year)</td>
<td>46,837</td>
<td>&lt;0.01</td>
<td>784</td>
</tr>
<tr>
<td>Genotype</td>
<td>491,501</td>
<td>&lt;0.01</td>
<td>24,381</td>
</tr>
<tr>
<td>Year×Genotype</td>
<td>35,513</td>
<td>&lt;0.01</td>
<td>1,143</td>
</tr>
</tbody>
</table>

Table 2.4. Correlations between traits which exhibited appreciable means and ranges.

Correlations observed in maize hybrid materials are shown in the blue cells, whereas correlations observed in inbred materials are shown in the green cells.

<table>
<thead>
<tr>
<th></th>
<th>FER</th>
<th>PCO</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FER</td>
<td>0.42</td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td>PCO</td>
<td>0.06</td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td>CA</td>
<td>0.77</td>
<td></td>
<td>0.00</td>
</tr>
</tbody>
</table>

†The abbreviations above represent the following phytochemicals: FER = Insoluble-bound ferulic acid; PCO = Insoluble bound p-coumaric acid; CA = Soluble cinnamic acid.


Rumpagapom, P. 2011. Structural features of cereal bran arabinoxylans related to colon fermentation rate. PhD. Purdue University, West Lafeyette, IN.


CHAPTER 3: CHANGES IN PHENOLIC ACID CONTENT IN MAIZE DURING FOOD PRODUCT PROCESSING

ABSTRACT

The notion that many nutrients and beneficial phytochemicals are lost due to food product processing is common. However, whether these changes actually occur has not been studied in great detail for the phenolic acids, some of which are chemopreventives of aging-related diseases. Therefore, it is unknown when and why these changes might occur, if some maize genotypes might be more resistant to changes in phenolic acid concentration than other genotypes, or if processing affects the bioavailability of phenolic acids in maize-based food products. Laboratory-scale processing techniques were developed for this study and used to process maize whole kernels into three intermediate products and then ultimately toasted cornflakes. High-throughput wet-lab analyses were used to determine the concentrations of soluble and insoluble-bound phenolic acids at each of the five processing stages in twelve ex-PVP inbreds and thirteen of their hybrids. Insoluble-bound ferulic acid was the most common phenolic acid, followed by insoluble-bound p-coumaric acid and then by soluble cinnamic acid, a precursor to the phenolic acids. Regardless of genotype, most of the phenolic acids were lost due to processing. Most of these changes occurred during dry milling and the removal of the bran. The concentration of bioavailable soluble ferulic and p-coumaric acid increased negligibly due to thermal stresses. Therefore, current processing techniques for the manufacturing of many maize-based foods, including breakfast cereals, are not conducive for increasing the content of bioavailable phenolics in processed maize food products.
INTRODUCTION

Diets rich in phenolic acids have been linked to the prevention of various aging-related diseases such as cancer and neurodegenerative and cardiovascular diseases (Adam et al., 2002; Ames et al., 1993; Fresco et al., 2006; Scapagnini et al., 2011). Although grains, especially maize, tend to possess high concentrations of phenolic acids, these are primarily found in the biologically unavailable, insoluble-bound form (Adom and Liu, 2002). It is possible that the insoluble-bound phenolic acids may be released into their bioavailable, soluble state during processing (Dewanto et al., 2002), but it is also possible that they may be lost. Given the health benefits and potential marketing advantages of products which contain bioavailable chemopreventive compounds, such as soluble phenolic acids, food companies are interested in identifying maize hybrids which maintain favorable phytochemical concentrations throughout food product processing (Helbert Almeida, Kellogg’s Company, Personal Communication, 2008).

Little is known about the variability of phenolic acid content in commercial maize germplasm or how the fate of the phenolic acids during food product processing may differ among commercial germplasm. Most of the phenolic acids in maize are hydroxycinnamic acids, and these are especially known for their chemopreventive properties (Fresco et al., 2006; Kanski et al., 2002). Maize is also one of the most concentrated sources of hydroxycinnamic acids (Adom and Liu, 2002). Since many of the same processing stresses encountered in the processing of toasted cornflakes are involved in the processing of other maize-based food products (Fig. 3.1), a better understanding of the genetic characteristics which could affect phenotypic responses to processing stresses and the general phenotypic response to processing
stresses could lead to the improvement of maize-based processed food products in regard to their antioxidant and chemopreventive phytochemical content. Specifically, processed food products could be improved directly by selecting hybrids which maintain favorable phenolic acid concentrations throughout processing. However, if most of the phenolic acids are lost during food processing or are not bioavailable, then alternative methods for increasing the phenolic acid content of processed maize food products should be explored.

The effects of chemical, thermal, and physical stresses that are typically encountered in the production of grain-based processed foods were examined, namely the removal of certain seed components, steaming, cooking, shearing forces, and toasting (Fig. 3.2, 3.3). As the majority of the phenolic acids are located in the bran (Rumpagapom, 2011) and the bran is removed during dry milling (Slavin et al., 2000), it is possible that a considerable portion of the phenolic acids is lost during processing. Conversely, some studies have found that thermal processing releases insoluble-bound, biologically unavailable phenolic acids from the cell walls to which they are esterified. In these instances, the phenolic acids were then present in their soluble, bioavailable state (Dewanto et al., 2002).

The main goal of this study was to determine if it is feasible to breed for maize hybrids which can be used in the production of healthier processed maize-based food products, particularly breakfast cereals. Specifically, the change in phenolic acid content throughout the production of cornflakes was monitored to determine which processing stresses impacted the phenolic acid content most and if there were some hybrids that showed a greater resistance, or susceptibility, to these changes than others. Both the means and variances of the phenolic acids
were recorded throughout processing because these statistical moments are indicative of the potential success of breeding for these traits. In addition, the state of the phenolic acids, which is indicative of their relative bioavailability, throughout processing was also determined.
MATERIALS AND METHODS

Plant Materials

All plant materials used in this study were acquired from the work of Macke, et al. (2016). These included 454 g (1 lb) samples of harvested grain material and the large flaking grit constituents which were produced as part of Macke et al. (2016). After combining the two subsamples of the large flaking grits for each field plot, approximately 500 g was available for each plot. In Macke et al. (2016), twelve genetically diverse inbreds that represent the parentage of important heterotic groups in current commercial maize breeding programs in the United States (Hauck et al., 2014) were crossed using a diallel design to create 66 F₁ hybrids. Five of these inbreds (B73, LH1, PHG39, PHJ40, and 4676A) are derived from various cycles of the Iowa Stiff Stalk Synthetic (SSS) heterotic pool. The other seven inbreds (LH123HT, LH82, Mo17, PH207, PHG47, PHG84, and PHZ51) are all members of the Non-Stiff Stalk (NSS) heterotic pool.

All inbreds and hybrids (N=78 entries) were grown at the University of Illinois Crop Sciences Research and Education Center in Urbana, IL for three years (2009-2011). The 78 entries and 6 commercial hybrid checks were grown in a resolvable incomplete block design with three replications for each of the three years. Due to time limitations and the laboriousness of the processing and analytical chemistry work (approximately 40 total hours per sample), not all samples were used. The rationale for this decision was that if the genotypes selected for use displayed considerable variability in their resistance to changes in phenolic acid concentration throughout processing, then more time could be devoted to studying this method for the improvement of phenolic acid content in maize. However, if this diverse subset did not display
much variability or if the phenolic acids were primarily biologically unavailable after processing, then more time could be devoted to alternative methods of improving the bioavailable phenolic acid content in processed maize food products. The multi-step process used and the specific plant materials used at each step in that process are detailed in Figure 3.4.

**Preliminary Soluble Phenolics Extraction and Quantification**

A modification of the protocol outlined in Adom and Liu (2002) was used. Since these preliminary analyses were to be used in determining the subset of hybrids to be analyzed, all 13 hybrids and 12 inbreds used in Chapter 2 were analyzed at the whole kernel and flaking grit processing stages. Specifically, approximately 4 g of maize was ground in a coffee grinder and dried at 65°C for eight hours. The maize sample was weighed before and after drying to determine the moisture content. The following formula was used to calculate the percent moisture content:

\[
\% \text{ Moisture} = \frac{\text{Wet weight} - \text{Dry weight}}{\text{Dry weight}} \times 100%
\]

2.5 g of the dried maize material was placed in a 30 mL centrifuge tube with 10 mL of 80% ethanol. The maize and ethanol solution was homogenized with a Polytron homogenizer for two minutes. Samples were centrifuged at 4,000 rpm for 17 minutes. The soluble fraction was removed following centrifugation and placed in a newly labeled test tube. The extraction was repeated one more time for a total of approximately 20 mL of extracted sample solution.

The extracted soluble fraction was evaporated using a custom fabricated Nitrogen evaporator (Fig. 3.5). PVC pipe was mounted onto a steel frame. Valves were attached to the PVC pipe so that up to six samples could be evaporated at once. Brass nozzles were fitted to the
PVC pipe, and dialysis tubing was attached to the nozzles so as to deliver \( \text{N}_2 \) gas to the samples. \( \text{N}_2 \) gas flow to the apparatus from a gas tank was regulated so as not to exceed 30 PSI. This design ensured that samples would not splash out of the vials and onto the dialysis tubing or elsewhere. A hot water bath was placed beneath the samples to aid in evaporation. Samples were evaporated to 5 mL.

Following evaporation, samples were reconstituted to a total of 10 mL using double deionized water. After the addition of water, samples were usually cloudy due to unknown lipids, and this affected the quantification of the samples’ soluble phenolics concentration. Therefore, these lipids were removed using NaCl. Five mL of 4 M NaCl was added to each reconstituted sample. Samples were vortexed for two minutes and then centrifuged for 10 minutes. The clear portion of the sample was placed in a new test tube.

For the colorimetric reactions, 0.2 mL of each sample was added to a new tube, as was 1.8 mL of methanol to create a 10% dilution, and 250 \( \mu \)L of sample was added to a new test tube. This was followed by the addition of 200 \( \mu \)L of Folin-Ciocalteu reagent and 550 \( \mu \)L \( \text{Na}_2\text{CO}_3 \) the new test tube. Two new test tubes were created for each sample so as to have laboratory replication at this step. All test tubes were placed in an incubator for 9 minutes at 42°C. Following incubation, samples were placed in a cool and dark location until they had reached room temperature (approximately 30 minutes).

Quantification took place using a spectrometer. The absorbance was read at 765 nm. Six standards of gallic acid were used to calculate a standard curve. The standard curve and sample
absorbance values were used to calculate the concentration of soluble phenolic acids in gallic acid equivalents (GAE).

**Starch, Oil, and Protein Content**

The starch, oil, and protein content of the hybrids and inbreds used in preliminary soluble phenolics analyses (note that the starch, oil, and protein content reported here were not analyzed in the manner of the more thorough analyses reported in Chapter 2) were conducted using non-destructive NIR. Calibration curves were built based on known standards. The remainder of the procedure was conducted following the DA 7200 Diode Array Analyzer Operation Manual (2007).

**Subset Determination**

In determining which entries to examine, it was noted that 2009 and 2010 had similar weather conditions. Therefore, to maximize the diversity due to environmental effects, only samples from 2009 and 2011 were used. All replications within both of these years were used. All 12 inbreds were used. Of the 13 hybrids listed in Chapter 2, seven of these were selected for use based on not only their phenotypic diversity (Fig. 3.6) but also their genetic diversity (Fig. 3.7). Cluster analyses were conducted as described in the statistical analyses section.

**Processing**

Whole kernels and large flaking grits were available at the beginning of the study. These materials are described in detail in Macke et al. (2016). The large flaking grits which were selected for further nutritional analysis were processed using the protocol described by Kandhola (2015). Small samples were taken after the pressure cooking (cooked grits), baking (baked
grits), and toasting (toasted cornflakes) portions of this processing work (Fig. 3.2). All samples were dried at 65°C for at least eight hours and then ground to a fine powder using a coffee grinder.

**Extraction and Quantification of Soluble and Insoluble-Bound Phenolic Acids**

The soluble phenolic acids were extracted and quantified using the procedure outlined in Chapter 2. Following the extraction of the soluble phenolic acids, the insoluble-bound phenolic acids were extracted and quantified using Protocol A of Butts-Wilmsmeyer and Bohn (2016) with the exception that standards of ferulic acid, p-coumaric acid, sinapic acid, and cinnamic acid were used rather than ferulic acid alone. This was so all four of these acids could be quantified, not just ferulic acid.

**Statistical Analysis**

**Cluster Analysis**

To determine the phenotypic dissimilarity between the 13 hybrids used in Chapter 2, hierarchical clustering was conducted using Ward’s Minimum Variance Approach in SAS 9.3. Phenotypic traits used in hierarchical clustering were test weight, dry milling efficiency, whole kernel soluble phenolics content, flaking grit soluble phenolics content, moisture, oil, starch, and protein content. Test weight and dry milling efficiency measures for each of the entries were generously provided by Macke et al. (2016). Dry milling efficiency, expressed as a percentage, is the mass of large flaking grits produced from a given mass of whole kernels which are dry milled. All of these traits are tied to grain composition and may influence the concentration of other beneficial phytochemicals. SAS (version 9.3) PROC MEANS was used to calculate the
mean of each phenotypic trait for each genotype. Raw phenotypic data were used in the calculations of the means. These means were used in the construction of a dendrogram in PROC CLUSTER. As Ward’s Minimum Variance Approach was used, the distance between two clusters can be defined as “the ANOVA sum of squares between the two clusters added up over all the variables” (SAS Institute, 2016). The arithmetic means of each trait for each cluster was calculated using PROC TREE and PROC MEANS. Genotypic data reported in van Heerwaarden et al. (2012) and hierarchical clustering were used to create a dendrogram of the genetic distance of the twelve original parents. A total of 45,997 SNPs were used to calculate the genetic distance between individuals in SIMGEND. A dendrogram was created using SAHN and the UPGA clustering method and was provided courtesy of Dr. Mark Mikel, University of Illinois.

*Identification of Most Important Processing Stages for Phytochemical Analysis*

Initially, plots were created in R (version 3.1) using the plot command. Phenolic acid means of each hybrid pedigree were plotted against the processing stage. An apparent lack of change in phenolic acid content was further investigated using the statistical model below.

The statistical model used was a split-plot in an RCBD where the whole plot unit was the field plot from which grain was harvested, and the subplot unit was processing stage. The same mixed linear model was used for both insoluble-bound ferulic acid and p-coumaric acid. Soluble phenolics were not of immediate concern due to their relatively low concentrations and plots which indicated that there was relatively little change in their concentrations after steaming (cooking). Only 200 samples (40 whole plots and 200 subplots) were analyzed as part of this
portion of the study, and all of these came from 2009. Therefore, unlike the other models in this study, there is no year term in the model. The model was run separately for hybrids and inbreds due to confounding of the generation effect with the physical location of the hybrid and inbred blocks in the field trial.

\[ y_{ijk} = \mu + R_i + \alpha_j + \varepsilon_{1ij} + \tau_k + \alpha \tau_{jk} + \varepsilon_{2ijk} \]

Where \( y_{ijk} \) is the observed concentration of the phenolic compound under observation,
\( \mu \) is the grand population mean,
\( R_i \) is the random effect of the \( i^{th} \) field replication, NID(0, \( \sigma_R^2 \)),
\( \alpha_j \) is the fixed effect of the \( j^{th} \) genotype,
\( \varepsilon_{1ij} \) is the random whole plot error term, NID(0, \( \sigma_{\varepsilon_1}^2 \)),
\( \tau_k \) is the fixed effect of the \( k^{th} \) processing stage,
\( \alpha \tau_{jk} \) is the fixed interaction between the \( j^{th} \) genotype and the \( k^{th} \) processing stage,
\( \varepsilon_{2ijk} \) is the random subplot error term, NID(0, \( \sigma_{\varepsilon_2}^2 \))

These statistical analyses were conducted in SAS 9.3 using the MIXED, UNIVARIATE, and GLM procedures. The ANOVA as described above was conducted in PROC MIXED. The assumptions of the ANOVA were verified as described below.
**Phytochemical Persistence Through Processing**

The statistical model used was a split-plot in an RCBD where the whole plot unit was the field plot from which grain was harvested, and the subplot unit was processing stage. The same mixed linear model was used for all traits measured as part of the evaluation of the changes in phytochemical content through processing. The model was run separately for hybrids and inbreds.

\[ y_{ijkl} = \mu + Y_i + R_{(i)j} + \alpha_k + Y\alpha_{ik} + \varepsilon_{1ijkl} + \tau_l + Y\tau_{il} + \alpha\tau_{kl} + Y\alpha\tau_{ikl} + \varepsilon_{2ijkl} \]

Where \( y_{ijkl} \) is the observed concentration of the phenolic compound under observation, 
\( \mu \) is the grand population mean, 
\( Y_i \) is the random effect of the \( i^{th} \) year, \( \text{NID}(0, \sigma_Y^2) \) 
\( R_{(i)j} \) is the random effect of the \( j^{th} \) rep nested within the \( i^{th} \) year, \( \text{NID}(0, \sigma_R^2) \) 
\( \alpha_k \) is the fixed effect of the \( k^{th} \) genotype, 
\( \varepsilon_{1ijkl} \) is the whole plot error term, \( \text{NID}(0, \sigma_{\varepsilon_1}^2) \) 
\( \tau_l \) is the fixed effect of the \( l^{th} \) processing stage, 
\( Y\tau_{il} \) is the random interaction between the \( i^{th} \) year and the \( l^{th} \) processing stage, \( \text{NID}(0, \sigma_{Y\tau}^2) \), 
\( \alpha\tau_{kl} \) is the fixed interaction between the \( k^{th} \) genotype and the \( l^{th} \) processing stage, 
\( Y\alpha\tau_{ikl} \) is the random three-factor interaction between the \( i^{th} \) year, the \( k^{th} \) genotype, and the \( l^{th} \) processing stage, and 
\( \varepsilon_{2ijkl} \) is the subplot error term, \( \text{NID}(0, \sigma_{\varepsilon_2}^2) \).
In this model, the effect of genotype is considered fixed because the genotypes used in the processing portion of this study were selected specifically.

These statistical analyses were conducted in SAS 9.3 using the MIXED, UNIVARIATE, MEANS, and GLM procedures. The ANOVA as described above was conducted in PROC MIXED. Additionally, R was used to create an interaction plot (profile plot) of the interaction between hybrid genotype and processing stage using the interaction.plot command.

**Verification of ANOVA Assumptions**

Residuals were obtained from the MIXED procedure and then analyzed using the UNIVARIATE procedure to check the assumption of normality. The Brown-Forsythe modification of the Levene test was used to check the assumption of homogenous variances among the stage and genotype treatment combinations. If the assumptions of normality or homogenous variances were violated, then the appropriate data transformation or outlier removal was performed. The QQ-plots and histograms of the residuals were used in the evaluation of whether outlying observations should be removed. Observations whose residuals significantly deviated from the normal line in the QQ-plot and who appeared to be drastically skewing the distribution of the residuals in the histogram were deleted. Multi-degree of freedom contrasts and slice statements were used to test the difference in processing stages if a significant interaction between genotype and processing stage occurred.
RESULTS AND DISCUSSION

Preliminary Analyses and Subset Determination

Preliminary analyses of the soluble phenolics content indicated that difference among genotypes were significant. Among the 13 hybrids used during the preliminary analyses, LSMEANS values ranged from 308.18 µg GAE / g maize in the whole kernels to 473.83 µg GAE / g maize. In the flaking grits, the soluble phenolic acid content of the hybrids ranged from 221.81 µg GAE / g maize to 435.72 µg GAE / g maize.

Hybrids could be grouped into three clusters, as indicated by a large jump in the RMSE score when comparing the RMSE values for two clusters versus three clusters. These clusters separated hybrids primarily based upon their soluble phenolics content and dry milling efficiency (Table 3.1). These clusters could be characterized as Cluster 1: moderate test weight, whole kernel soluble phenolics content, and flaking grit soluble phenolics content; Cluster 2: high test weight and low soluble phenolics content at both the whole kernel and flaking grit processing stages; and Cluster 3: low test weight and high soluble phenolics content at both the whole kernel and flaking grit processing stages. It was possible to select seven hybrids which represented each major cluster in the dendrogram of the phenotypes of the hybrids (Figure 3.6) and in the dendrogram of the genotypes of the parental inbreds (Figure 3.7).

Identification of the Most Important Processing Stages for Studying Change in Phenolic Acid Content Throughout Processing

To ensure the most informative data were collected, the most important processing stages which would aid in the study of the changes in phenolic acid content throughout processing were
examined. Little additional time is needed for the collection of cooked grits and baked grits during processing. However, increasing the number of samples undergoing analytical chemistry analyses would have greatly increased the amount of time spent on this project. Therefore, the elimination of unnecessary processing stages would enable more informative data to be collected in a timely fashion. While measuring the first 200 processed samples (material from 40 harvested plots at their five processing stages) to determine if there was a significant change due to processing between each of the adjacent processing stages, it was noted that not all processing stages needed to be analyzed. Since the two predominant phenolic acids were insoluble-bound ferulic acid and p-coumaric acid, the processing analyses focused primarily on these two compounds. However, plots of the mean soluble phenolic acid content for each genotype versus processing stage were made (Fig. 3.8.b, 3.8.d) to ensure that no additional processing stages were needed in order to analyze the soluble phenolics content. All plots showed that the soluble ferulic acid, p-coumaric acid, and cinnamic acid content remained relatively stable between the cooked grit and the final toasted cornflake stage. There was a slight decrease in the concentration between baked grits and the final toasted cornflakes, but the overall change was negligible.

Plots of insoluble-bound ferulic acid and p-coumaric acid content through the five different processing stages showed a drastic decrease between the whole kernel and flaking grit stage and then a small decrease after the flaking grit stage before stabilizing for the remainder of processing (Fig. 3.8.a, 3.8.c). There was also a decrease in the concentration of soluble cinnamic acid following dry milling before nearly stabilizing between the cooked grit and toasted cornflake processing stages (Fig. 3.8.e). These results indicated that whole kernels and flaking
grits should be analyzed for their phenolic acid content. Multi-degree of freedom contrasts indicated that the change in insoluble-bound phenolic acid content between the cooked grit and final toasted cornflake is non-significant, regardless of the genotype under study (Table 3.2). Therefore, only the whole kernel, large flaking grit, and toasted cornflake stages were deemed necessary for subsequent phytochemical analyses.

These results are in agreement with the available literature. Adom and Liu (2002) and Yadav, et al. (2007) noted that most of the phenolic acids in maize are in their insoluble-bound form. Yadav, et al. (2007) also found that the most predominant phenolic acids in maize were insoluble-bound ferulic acid and insoluble-bound p-coumaric acid. Considering that most of the phenolic acids are located in the bran and that the bran is removed during dry milling, it is not surprising that the overall concentration of the phenolic acids decreased after dry milling. Furthermore, while some of the insoluble-bound phenolic acids were released into the soluble state due to thermal stresses during cooking and baking, it does not appear that the concentration of the insoluble-bound phenolic acids fluctuated much following dry milling. These findings suggest that the structure of the phenolic acids, particularly when they are esterified to arabinoxylans, inhibits degradation of the phenolic acids when exposed to thermal stresses. However, during rolling and toasting, it appears that some of the soluble phenolic acids were degraded. The degradation of the soluble phenolic acids is possibly because, unlike their insoluble-bound counterparts, the soluble phenolic acids are not bound to arabinoxylans and are not as protected from thermal stress. However, this change in the soluble phenolic content is slight.
**Phenolic Acid Content Throughout Processing**

Having identified the most important processing stages at which changes in phenolic acids occur during the production of toasted cornflakes, 19 entries × 2 years × 3 reps × 3 processing stages = 342 samples were analyzed for their phenolic acid content. The phenolic acids were almost exclusively composed of hydroxycinnamic acids, as expected. Insoluble-bound ferulic acid was the most prevalent (average concentration of 1,948.4 μg / g in the hybrid whole kernels), followed by insoluble-bound p-coumaric acid (176.4 μg / g) and then soluble cinnamic acid (63.7 μg / g). These results are in agreement with the available literature (Arnason et al., 1992; Serratos et al., 1987; Yadav et al., 2007). However, this study analyzed a much broader germplasm base than previous studies. For instance, Serratos, et al. (1987) analyzed only 4 maize entries (3 local populations from Belize and one double-cross hybrid), and Yadav, et al. (2007) and Adom and Liu (2002) did not analyze individual genotypes in their studies. Therefore, this research extends the findings in the literature to a broader germplasm base.

Dry milling resulted in the largest loss of phenolic acids, particularly insoluble-bound ferulic acid and insoluble-bound p-coumaric acid (Fig. 3.9, 3.10). The large loss in phenolic acid content at this stage is likely because the majority of phenolic acids in maize kernels are located in the bran (Rumpagapom, 2011), and the bran and germ are removed during dry milling (Slavin et al., 2000). While hybrid maize kernels had an average of 1,948.4 μg / g of insoluble-bound ferulic acid and 176.4 μg / g of insoluble-bound p-coumaric acid at the whole kernel processing stage, the average concentration of insoluble-bound ferulic acid and insoluble-bound p-coumaric acid dropped to 980.0 μg / g and 74.5 μg / g at the flaking grit stage before stabilizing at
approximately 900 μg/g and 55 μg/g for the remainder of processing, respectively. In comparison, the average soluble ferulic acid and p-coumaric acid content in the hybrids increased from only 1.5 μg/g and 1.8 μg/g to 7.3 μg/g and 7.0 μg/g during processing, respectively. Additionally, the soluble cinnamic acid content decreased from 71.2 μg/g to 26.7 μg/g. A similar observation can be seen among the inbreds (Table 3.3).

The variance in phenolic acid content among hybrids and inbreds decreased throughout processing. In regards to insoluble-bound ferulic acid, the phenotypic variance among hybrid genotypes dropped from 82,811.6 \( \left( \frac{\mu g}{g} \right)^2 \) to 54,447.6 \( \left( \frac{\mu g}{g} \right)^2 \) between the whole kernel and final toasted cornflake. Similarly, the phenotypic variance among hybrid genotypes for their p-coumaric acid content dropped from 1,899.2 \( \left( \frac{\mu g}{g} \right)^2 \) to 636.0 \( \left( \frac{\mu g}{g} \right)^2 \) between the whole kernel and final toasted cornflake. Similar trends were recorded for the inbreds used in this study (Table 3.3). These results indicate that not only are most of the insoluble-bound hydroxycinnamic acids located in the bran, but most of the variability in hydroxycinnamic acid content is also localized in the bran.

Furthermore, the rank of a particular hybrid at the whole kernel processing stage in regards to its phenolic acid content was not indicative of the rank of that hybrid in later processing stages (Fig. 3.11.a, 3.11.b). Since the rank of the hybrids changed throughout processing, the concentration of insoluble-bound hydroxycinnamic acids in the whole kernels is not indicative of the concentration at other processing stages. This conclusion is supported by the significant interaction between genotype and processing stage (Table 3.4). Therefore, to improve the concentration of hydroxycinnamic acids in toasted cornflakes and other maize-based
processed foods, the final processed product must be analyzed. In a plant breeding setting, this is laborious, time-consuming, and requires large laboratory space for processing.

This observation then warrants examination of whether high-throughput technologies might make plant breeding for insoluble-bound phenolic acid content that persists through processing more feasible. The most likely candidate is NIR, as this is a high-throughput phenotyping technique. However, wet-lab chemistry still must be conducted so that NIR prediction models can be built. This also means that all of the laboratory-scale processing work must be completed. Furthermore, as will be seen in Chapter 5, this research found that NIR models could not successfully predict the insoluble-bound ferulic acid or insoluble-bound p-coumaric acid content in maize, even when approximately 700 samples were used to build the models. Therefore, the construction of an NIR prediction model would still require the analysis of many processed materials. Similarly, genomic prediction and other high-throughput techniques would still require a considerable amount of phenotyping before accurate prediction models could be built.

Regardless of the genotype, most of the phenolic acids present in the whole kernel were lost during processing. Most importantly, those phenolic acids that remained were primarily present in their least bioavailable form as insoluble-bound hydroxycinnamic acids. Adam, et al. (2002) found that the insoluble-bound hydroxycinnamic acids tended not to be absorbed. Rather, they serve as cross-linkages between arabinoxylan fibers and exhibit negligible chemopreventive properties in regards to the prevention of cancer or neurodegenerative diseases. These results, taken in conjunction with the knowledge that the lab-scale processing involved in this project
was very time-consuming, indicates that improving the concentration of bioavailable hydroxycinnamic acids by selecting for hybrids which maintain positive phytochemical attributes throughout processing is not feasible.
### Table 3.1. Means of phenotypic traits by cluster

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Test Weight†</th>
<th>Dry Milling Efficiency</th>
<th>Whole Kernel Soluble Phenolics</th>
<th>Flaking Grit Soluble Phenolics</th>
<th>Moisture</th>
<th>Starch</th>
<th>Protein</th>
<th>Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kg/hL</td>
<td>%</td>
<td>µg/g GAE</td>
<td>µg/g GAE</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>1</td>
<td>71.04</td>
<td>27.88</td>
<td>418.64</td>
<td>300.41</td>
<td>8.12</td>
<td>65.78</td>
<td>9.01</td>
<td>3.72</td>
</tr>
<tr>
<td>2</td>
<td>72.35</td>
<td>32.91</td>
<td>361.64</td>
<td>237.23</td>
<td>7.89</td>
<td>66.16</td>
<td>9.32</td>
<td>3.45</td>
</tr>
<tr>
<td>3</td>
<td>70.21</td>
<td>25.02</td>
<td>466.24</td>
<td>402.91</td>
<td>7.81</td>
<td>66.48</td>
<td>9.10</td>
<td>3.43</td>
</tr>
</tbody>
</table>

† The conversion factor for test weight is kg/hL = 1.25(lb/bu). Therefore, the test weights could also be reported as 56.83, 57.88, and 56.17 lb/bu for clusters 1, 2, and 3, respectively.

‡ For the pictorial representation of this cluster analysis, please refer to Fig. 3.6.
Table 3.2. Multi-degree of freedom contrasts testing the difference in insoluble-bound hydroxycinnamic acid content in cooked grits, baked grits, and toasted cornflakes

<table>
<thead>
<tr>
<th></th>
<th>Ferulic Acid</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F Value</td>
<td>P-Value</td>
<td></td>
</tr>
<tr>
<td>Difference in B73×MO17</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Difference in B73×PHG47</td>
<td>0.02</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Difference in LH1×MO17</td>
<td>0.08</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Difference in PHJ40×LH123HT</td>
<td>0.32</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Difference in PH207×PHG47</td>
<td>0.15</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Difference in PHJ40×MO17</td>
<td>0.01</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Difference in PHG39×PHZ51</td>
<td>0.06</td>
<td>0.94</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>p-Coumaric Acid</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F Value</td>
<td>P-Value</td>
<td></td>
</tr>
<tr>
<td>Difference in B73×MO17</td>
<td>0.34</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Difference in B73×PHG47</td>
<td>0.61</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Difference in LH1×MO17</td>
<td>0.14</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Difference in PHJ40×LH123HT</td>
<td>0.74</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Difference in PH207×PHG47</td>
<td>0.24</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>Difference in PHJ40×MO17</td>
<td>0.31</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Difference in PHG39×PHZ51</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3. Means and standard deviations of phenolic acids and cinnamic acid at different processing stages

<table>
<thead>
<tr>
<th>Processing Stage</th>
<th>Generation</th>
<th>Soluble Cinnamic Acid µg/g</th>
<th>Soluble Ferulic Acid µg/g</th>
<th>Soluble p-Coumaric Acid µg/g</th>
<th>Insoluble-Bound Ferulic Acid µg/g</th>
<th>Insoluble-Bound p-Coumaric Acid µg/g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Kernels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYBRIDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>63.67</td>
<td>1.50</td>
<td>1.81</td>
<td>1,948.41</td>
<td>176.40</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>43.34</td>
<td>0.51</td>
<td>0.44</td>
<td>287.77</td>
<td>43.58</td>
<td></td>
</tr>
<tr>
<td>INBREDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>90.87</td>
<td>1.78</td>
<td>2.42</td>
<td>2,004.42</td>
<td>223.08</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>68.15</td>
<td>0.55</td>
<td>0.81</td>
<td>374.13</td>
<td>81.62</td>
<td></td>
</tr>
<tr>
<td>Large Flaking Grits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYBRIDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>34.77</td>
<td>2.17</td>
<td>2.13</td>
<td>980.03</td>
<td>74.54</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>28.42</td>
<td>0.46</td>
<td>0.66</td>
<td>273.79</td>
<td>34.99</td>
<td></td>
</tr>
<tr>
<td>INBREDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>31.86</td>
<td>2.35</td>
<td>2.21</td>
<td>787.05</td>
<td>63.81</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>41.38</td>
<td>0.57</td>
<td>0.75</td>
<td>234.30</td>
<td>38.49</td>
<td></td>
</tr>
<tr>
<td>Toasted Cornflakes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYBRIDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>26.70</td>
<td>7.28</td>
<td>7.03</td>
<td>902.83</td>
<td>56.48</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>21.73</td>
<td>1.98</td>
<td>2.52</td>
<td>233.34</td>
<td>25.22</td>
<td></td>
</tr>
<tr>
<td>INBREDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>27.32</td>
<td>7.13</td>
<td>8.10</td>
<td>786.92</td>
<td>50.50</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>30.31</td>
<td>2.32</td>
<td>3.03</td>
<td>273.04</td>
<td>32.81</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4. Significance of genotype-by-processing stage interaction

<table>
<thead>
<tr>
<th></th>
<th>Ferulic Acid</th>
<th>p-Coumaric Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-Value</td>
<td>P-Value</td>
</tr>
<tr>
<td>HYBRIDS</td>
<td>7.15</td>
<td>0.001</td>
</tr>
<tr>
<td>INBREDS</td>
<td>4.07</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Note: All year-by-genotype-by-processing stage interactions were non-significant at α = 0.05
Figure 3.1. Processed maize food product production pathways. Processing steps involved in the production of many maize-based food products. The three major categories of maize food product processing are alkaline hydrolysis, wet milling, and dry milling. The toasted cornflakes production pipeline is shown in red. During the production of most food products in the figure, various seed components are removed, thermal stresses are encountered, or other physical stresses may be seen. The production of toasted cornflakes involves a number of the same physical, chemical, and thermal stresses seen in the production of other maize food products, not just those that also are produced via dry milling.
Figure 3.2. Cornflake production pipeline. The plant breeding and food processing steps involved in the production of toasted cornflakes and other maize-based snack foods. All processes involved are indicated with a blue rectangle, green circles indicate non-grain based maize products, large yellow circles indicate marketable grain-based products, and small circles indicate intermediate, non-marketable products.
Figure 3.3. Overview of laboratory scale processing used in the production of cornflakes. Photos of the different processes used in the laboratory scale production of toasted cornflakes and intermediate processing stages are shown above.
Subset 1: All years, all reps, all 12 inbreds, 13 hybrids
Subset 2: Two years, 3 reps per year, all 12 inbreds, 7 hybrids
Subset 3: First 200 processed samples (Subset 3) from 2009 (40 maize entries, 5 processing stages)
Subset 4: Two years, 3 reps per year, all 12 inbreds, 7 hybrids, 3 processing stages

Figure 3.4. Description of materials used at different phases in this study.
Figure 3.5. Custom fabricated nitrogen evaporator. Six samples sit in a hot water bath while evaporating under N₂ gas. Nitrogen evaporator constructed courtesy of David Butts.
Figure 3.6. Cluster analysis of the phenotypic diversity of the 13 hybrids used in this study.
The seven hybrids selected for study during processing are shown in bold font. The remaining six hybrids which were not selected are shown in red font. Hybrids clustered into three major groups, and at least two hybrids were selected from each of the three groups.
Figure 3.7. Cluster analysis of the genetic diversity in the 12 inbred parents. The parental inbreds of the seven hybrids used in the processing work are shown in bold font. The remaining hybrids are shown in red font. All major clusters of the parental inbreds were represented in the processing analyses. The dendrogram was constructed using genetic information published in van Heerwaarden et al. (2012). Dendrogram provided courtesy of Dr. Mark Mikel, University of Illinois.
Figure 3.8. Change in phytochemical content by processing stage.

a. Change in insoluble-bound ferulic acid content by processing stage. Means from the same hybrid are represented by the same color in each processing stage. All five processing stages, whole kernels (WK), large flaking grits (FG), cooked grits (CG), baked grits (BG), and toasted cornflakes (TO) were evaluated for their nutritional content. Hybrids are color-coded as follows: Red = B73×Mo17, Green = B73×PHG47, Blue = LH1×Mo17, Black = PHJ40×LH123HT, Gray = PHJ40×Mo17, Purple = PH207×PHG47, and Tan = PHG39×PHZ51.

Figure 3.8 (cont.) b. Change in soluble ferulic acid content by processing stage. Means from the same hybrid are represented by the same color in each processing stage. Hybrids are color-coded as described in a.
Figure 3.8 (cont.) c. Change in insoluble-bound p-coumaric acid content throughout processing. Means from the same hybrid are represented by the same color in each processing stage. Hybrids are color-coded as described in a.

Figure 3.8 (cont.) d. Change in soluble p-coumaric acid content throughout processing. Means from the same hybrid are represented by the same color in each processing stage. Hybrids are color-coded as described in a.
Figure 3.8 (cont.) e. Change in soluble cinnamic acid content throughout processing. Means from the same hybrid are represented by the same color in each processing stage. Hybrids are color-coded as described in a.
Figure 3.9. Change in insoluble-bound ferulic acid content throughout processing using two years and seven hybrids. Means from the same hybrid are shown in the same color in each of the processing stages. Hybrids are color-coded as follows: Green = LH1×Mo17, Blue = B73×Mo17, Brown = PHG39×PHZ51, Black = PHJ40×LH123HT, Red = B73×PHG47, Orange = PHJ40×Mo17, and Purple = PH207×PHG47.

Figure 3.10. Change in insoluble-bound p-coumaric acid content throughout processing using two years and seven hybrids. Means from the same hybrid are shown in the same color in each of the processing stages.
Figure 3.11. Interaction between processing stage and hybrid for phytochemical content.  

*a. Interaction between processing stage and hybrid for ferulic acid content.* The plot indicates that there is a change-of-rank interaction.

---

Figure 3.11 (cont.).  

*b. Interaction plot between processing stage and hybrid for the natural log of p-coumaric acid content.* The plot indicates that there is a change-of-rank interaction.
WORKS CITED


Kandhola, G. 2015. Processing and genetic effects on resistant starch in corn flakes. MS. University of Illinios, Urbana, IL.


Rumpagapom, P. 2011. Structural features of cereal bran arabinoxylans related to colon fermentation rate. PhD. Purdue University, West Lafeyette, IN.


CHAPTER 4: POTENTIAL FOR BREEDING MAIZE FOR USE AS ALL-NATURAL FOOD ADDITIVES

ABSTRACT

Hydroxycinnamic acids are chemopreventive agents of many aging-related diseases. However, not all people have equal access to these potentially life-saving compounds. This study examines the potential of breeding maize for increased hydroxycinnamic acid content so that the hydroxycinnamic acids could be extracted and used as food additives. These food additives could then be added to a variety of inexpensive, processed food products. Both ferulic acid and p-coumaric acid content were found to be highly heritable in maize. The lack of a strong negative correlation between important agronomic characteristics and hydroxycinnamic acid content suggests that both ferulic acid and p-coumaric acid content can be improved without incurring yield or test weight penalties. A three-dimensional plot of total ferulic acid per hectare, total p-coumaric acid per hectare, and grain yield by hybrid verify that it is possible to increase the total amount of ferulic acid per hectare as well as grain yield quite easily, and it might also be possible to increase the total amount of p-coumaric acid per hectare with little additional effort. Lastly, both ferulic acid and p-coumaric acid had high mean contents and large variances among hybrids. This indicates that breeding maize for use in all-natural food additives is feasible.
INTRODUCTION

Diets rich in phenolic acids have been linked to the prevention of various aging-related diseases such as cancer, neurodegenerative diseases, and cardiovascular disease (Adam et al., 2002; Ames et al., 1993; Fresco et al., 2006; Scapagnini et al., 2011). Although grains, especially maize, tend to possess high concentrations of phenolic acids, these are primarily found in the biologically unavailable, insoluble-bound form in maize and other grains (Adom and Liu, 2002). In contrast, fresh fruits and vegetables contain smaller quantities of phenolic acids, but these tend to be soluble and bioavailable when consumed by humans (Adam et al., 2002; Adom and Liu, 2002). However, the consumption of fresh produce follows a socioeconomic trend. Individuals from poorer socioeconomic areas are more likely to purchase processed, starchy, calorie-dense foods than fresh produce (Darmon and Drewnowski, 2008; Drewnowski and Darmon, 2005). For this reason, these chemopreventive phytochemicals and antioxidants may not be readily available to all socioeconomic groups. Another point of consideration is that the incidence of aging-related diseases tends to be more prevalent in socioeconomically poor groups (House et al., 1990). One potential means of addressing the lack of access to dietary chemopreventive phytochemicals in some socioeconomic groups is by increasing the content of bioavailable phenolic acids in processed food products.

Most of the phenolic acids in maize are hydroxycinnamic acids. Of these, ferulic acid and p-coumaric acid are the most prevalent (Yadav et al., 2007). In their soluble forms, ferulic acid and p-coumaric acid are renowned for being very potent chemopreventive agents (Fresco et al., 2006; Kanski et al., 2002; Scapagnini et al., 2011). Although they may act as antioxidants in vivo, a more likely method of in vivo protection is via the activation and regulation of different
gene pathways. For instance, the regulation of the *heterodimers of NF-E2-related factors 2* (Nrf2)/antioxidant responsive element (ARE) pathway is implicated in the prevention of neurodegenerative diseases (Scapagnini et al., 2011). The regulation of other signaling pathways, including the nuclear factor-κB (NF-κB), activator protein-1 (AB-1), or mitogen-activated protein kinases (MAPK), are hypothesized to be involved in the chemoprevention of various cancers (Fresco et al., 2006).

The research conducted in Chapter 3 showed that most of the hydroxycinnamic acids in maize are either lost during food product processing or remain in their insoluble-bound state. Therefore, for the chemopreventive properties of the hydroxycinnamic acids in maize to be available, the hydroxycinnamic acids must be extracted prior to processing. Acosta-Estrada, et al. (2014) was successful in extracting hydroxycinnamic acids from maize and using these as food additives. However, the germplasm in that study was not representative of the germplasm grown in the United States.

In this study, the prospect of breeding for hydroxycinnamic acid food additives which can be extracted from maize is explored in more detail. The specific objectives of this study were to examine the variability and mean content of hydroxycinnamic acids in maize, to conduct quantitative genetics analyses in regard to the hydroxycinnamic acid content in maize, and to determine if the hydroxycinnamic acid content and grain yield could be improved concurrently.
MATERIALS AND METHODS

Plant Materials

Twelve maize inbreds (Table 2.1) which were identified as being genetically representative of the germplasm grown in the United States Midwest were crossed using a diallel design to create 66 $F_1$ hybrids (Hauck et al., 2014; Macke et al., 2016). Five of these inbreds (B73, LH1, PHG39, PHJ40, and 4676A) were derived from various cycles of the Iowa Stiff Stalk Synthetic (SSS) heterotic pool. The other seven inbreds (LH123HT, LH82, Mo17, PH207, PHG47, PHG84, and PHZ51) are all members of the Non-Stiff Stalk (NSS) heterotic pool.

All inbreds and hybrids ($N = 78$ entries) were grown at the University of Illinois Crop Sciences Research and Education Center in Urbana, IL for three years (2009-2011) (Macke, et al., 2016). The 78 entries were grown in a resolvable incomplete block design with three replications for each of the three years. All entries were grown in four-row plots, and the center two plots were harvested so as to minimize the effects of the neighboring plots. Bags containing 454 g (one pound) of harvested grain from each plot were placed in cold storage for phytochemical analysis. Approximately 50 g of whole kernels from each plot were ground to a fine powder using a Foss Cyclone Mill ($< 1$ mm diameter particle size) in preparation for phytochemical analysis.

Extraction and Quantification of Hydroxycinnamic Acid Content

Since the soluble hydroxycinnamic acid content in maize is negligible, as shown in Chapter 3, only the insoluble-bound ferulic acid and insoluble-bound p-coumaric acid content were measured. Protocol A of Butts-Wilsmeyer and Bohn (2016) was used with the exception
that standards of ferulic acid and p-coumaric were used in the creation of the standard curve.

Since Chapters 1-3 verified that these hydroxycinnamic acids were present in maize, mass spectroscopy was eliminated in this chapter.

Statistical Analyses

Effect of Genotype and Genotype-by-Environment, Calculation of Midparent Heterosis

The statistical model used was a resolvable incomplete block design. The same mixed linear model was used for both ferulic acid and p-coumaric acid. The model was run separately for hybrids and inbreds.

\[ y_{ijkl} = \mu + Y_i + R_{(i)j} + B_{(ij)k} + G_l + YG_{il} + \varepsilon_{ijkl} \]

Where \( y_{ijkl} \) is the observed concentration of the phenolic compound under observation,

\( \mu \) is the grand population mean,

\( Y_i \) is the random effect of the \( i^{th} \) year, NID(0, \( \sigma_Y^2 \))

\( R_{(i)j} \) is the random effect of the \( j^{th} \) rep nested within the \( i^{th} \) year, NID(0, \( \sigma_R^2 \)),

\( B_{(ij)k} \) is the random effect of the \( k^{th} \) incomplete block, NID(0, \( \sigma_B^2 \)),

\( G_l \) is the random effect of the \( l^{th} \) genotype, NID(0, \( \sigma_G^2 \)),

\( YG_{il} \) is the random interaction between the \( i^{th} \) year and the \( l^{th} \) genotype, NID(0, \( \sigma_{YG}^2 \)),

and \( \varepsilon_{ijkl} \) is the random error term, NID(0, \( \sigma_\varepsilon^2 \))

In this model, the genotype term was considered random because all possible hybrid combinations of the parental inbreds which are representative of the germplasm grown in the
Midwest were studied and because the inference space was to all Midwestern hybrids and inbreds. Analyses were conducted in PROC MIXED of SAS (version 9.3). The assumptions of homogenous variances and normality were verified. The main goal of this statistical analysis was to determine if the effect of genotype was significant. If a significant year-by-genotype interaction (also known as a genotype-by-environment interaction, as each year was regarded as a separate environment) was present, PROC GPLOT was used to plot the BLUPs of each hybrid for each year. If the lines, which represented different years, displayed the same pattern, there were noticeable troughs and peaks in the plot of the BLUPs which corresponded to different hybrids, and the effect of genotype was found to be highly significant, then the effect of hybrid was also considered to be significant.

Midparent heterosis was calculated by extracting the BLUPs by entry from the statistical model as described above. The midparent value was calculated as the average of the BLUPs of the two parental inbreds. Midparent heterosis was then calculated as

\[
MP_i = \frac{HCA_{ij} - MV_{ij}}{MV_{ij}} \times 100\%
\]

Where \( MP_i \) is the midparent heterosis of the \( i^{th} \) hydroxycinnamic acid, \( HCA_{ij} \) is the BLUP associated with the \( i^{th} \) hydroxycinnamic acid and the \( j^{th} \) hybrid, and \( MV_{ij} \) is the midparent value associated with the \( i^{th} \) hydroxycinnamic acid and the \( j^{th} \) hybrid.
Calculation of Heritability

Diallel crossing designs allow for the decomposition of the genotypic effect into general combining ability and specific combining ability. Consequently, both narrow-sense and broad-sense heritability calculations can be calculated. The model used was

\[ y_{ijklm} = \mu + Y_i + R_{(ij)} + B_{(ij)k} + GCA_l + GCA_m + SCA_{lm} + YG_{il} + YG_{lm} + YS_{ilm} + \varepsilon_{ijklm} \]

Where \( y_{ijkl} \) is the observed concentration of the phenolic compound under observation,

\( \mu \) is the grand population mean,

\( Y_i \) is the random effect of the \( i^{th} \) year, NID(0, \( \sigma^2_Y \))

\( R_{(ij)} \) is the random effect of the \( j^{th} \) rep nested within the \( i^{th} \) year, NID(0, \( \sigma^2_R \)),

\( B_{(ij)k} \) is the random effect of the \( k^{th} \) incomplete block, NID(0, \( \sigma^2_B \)),

\( GCA_l \) is the random effect of the general combining ability of the \( l^{th} \) inbred parent, NID(0, \( \sigma^2_GCA \))

\( GCA_m \) is the random effect of the general combining ability of the \( m^{th} \) parent, NID(0, \( \sigma^2_GCA \))

\( SCA_{lm} \) is the random effect of the specific combining ability of the \( l^{th} \) and \( m^{th} \) inbred parents, NID(0, \( \sigma^2_SCA \))

\( YG_{il} \) is the random interaction between year and the general combining ability of the first parent, NID(0, \( \sigma^2_{YGCA} \))

\( YG_{lm} \) is the random interaction between year and the general combining ability of the second parent, NID(0, \( \sigma^2_{YG} \))

\( YS_{ilm} \) is the random interaction between year and the specific combining ability of the parents used in a hybrid cross, NID(0, \( \sigma^2_{YS} \))

\( \varepsilon_{ijklm} \) is the random error term, NID(0, \( \sigma^2_\varepsilon \))
This model was analyzed in PROC MIXED of SAS in order to calculate variance components. Narrow-sense ($h^2$) and broad-sense ($H^2$) heritabilities were estimated using the following equations:

$$h^2 = \frac{2\sigma_{GCA}^2}{2\sigma_{GCA}^2 + \sigma_{SCA}^2 + \frac{2\sigma_{YG}^2}{env} + \frac{\sigma_{YS}^2}{env} + \frac{\sigma_e^2}{env \times rep}}$$

$$H^2 = \frac{2\sigma_{GCA}^2 + \sigma_{SCA}^2}{2\sigma_{GCA}^2 + \sigma_{SCA}^2 + \frac{2\sigma_{YG}^2}{env} + \frac{\sigma_{YS}^2}{env} + \frac{\sigma_e^2}{env \times rep}}$$

Where env is the number of environments (env = 3 because the experiment was conducted for 3 years) and rep is the number of replications used in this experiment (rep = 3).

**Calculation of BLUPs of Total Ferulic Acid per Acre, Total p-Coumaric Acid per Acre, and Grain Yield**

Total ferulic acid per acre and total p-coumaric acid per acre can be defined as the total amount of these phenolic acids that can be extracted from an acre of harvestable maize. Therefore,

$$FERAC_{ij} = FA_{ij} \times \frac{454g}{lb} \times TW_{ij} \times GY_{ij} \times \frac{1mg FA}{10^3 \mu g FA} \times \frac{2.471ac}{1ha}$$
Where $F_{ERA}C_{ij}$ is the total amount of ferulic acid per hectare for the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid,

$FA_{ij}$ is the ferulic acid content of the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid,

$TW_{ij}$ is the test weight of the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid, and

$GY_{ij}$ is the grain yield of the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid.

\[
PCOAC_{ij} = PA_{ij} \times \frac{454g}{lb} \times TW_{ij} \times \frac{1mg FA}{10^3 \mu g FA} \times \frac{2.471ac}{1ha}
\]

Where $PCOAC_{ij}$ is the total amount of p-coumaric acid per hectare for the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid,

$PA_{ij}$ is the p-coumaric acid content of the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid,

$TW_{ij}$ is the test weight of the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid, and

$GY_{ij}$ is the grain yield of the $i^{th}$ hybrid and the $j^{th}$ observation from that hybrid.

As for the analysis of the significance of the genotype-by-environment and genotype effects, the statistical model used was

\[
y_{ijkl} = \mu + Y_l + R_{(i)j} + B_{(ij)k} + G_t + YG_{tl} + \epsilon_{ijkl}
\]

The description of these effects is the same as before. Residuals were output from PROC MIXED and analyzed in PROC UNIVARIATE to check the assumption of normality. Observations which appeared to be extreme outliers that were skewing the data were removed. Transformations of the data were performed accordingly so as to meet the assumptions of the
ANOVA. The assumption of homogenous variances was tested using the Brown-Forsythe modification of the Levene in PROC GLM. After these assumptions had been met, BLUPs of the ferulic acid content per hectare, p-coumaric acid content per hectare, and grain yield (in Mg/ha) were calculated in SAS. These BLUPs were then plotted against one another using the scatterplot3d function in R to determine if all three of these traits might be improved concurrently.

**Correlations Between Key Traits of Interest**

The correlations between ferulic acid content, p-coumaric acid content, grain yield, and test weight were calculated using PROC CORR in SAS. Correlations were not calculated between these traits and total ferulic acid per hectare or total p-coumaric acid per hectare because the latter two traits are functions of ferulic acid content and p-coumaric acid content as well as grain yield. Furthermore, the interest in calculating these correlations was to determine if the concentrations of these phenolic acids could be improved concurrently with yield and test weight. If so, then the total amount of ferulic acid and p-coumaric acid per hectare could inherently be increased.
RESULTS AND DISCUSSION

**Significance of Model Effects**

Table 4.1 includes a list of all the model effects, F-values, P-values, and variance estimates of the Mixed model terms. While the genotype-by-year (i.e. genotype-by-environment) interaction is significant, the profile plot of the BLUPs (Fig. 4.1) and the high F-values associated with the genotype effect indicated that the genotypic effect is highly significant. The profile plot of the BLUPs (Fig. 4.1) shows that while there may be crossovers between the years, the overall trend in the plots is the same. Relevant to itself, hybrid performance tends to be similar, regardless of environment (i.e. the “best” hybrids perform the best each year).

**Mean, Variance, and Heritability of Hydroxycinnamic Acid Extracts**

The total ferulic acid and p-coumaric acid content varied greatly among hybrids and inbreds (Table 4.2). Since hybrids are typically sold commercially due to their yield advantage over inbreds, this discussion focuses solely upon the hybrids. However, the summary statistics regarding both hybrids and inbreds can be found in Table 4.2. Whole kernels from hybrids had average concentrations of 1,902.4 µg/g of insoluble-bound ferulic acid and 178.6 µg/g insoluble-bound p-coumaric acid which could be recovered in their soluble forms after extraction. Also, ferulic acid content ranged from 1,382.7 µg/g to 2,759.9 µg/g, and p-coumaric acid content ranged from 80.2 µg/g to 278.9 µg/g. These figures are based on the best linear unbiased predictors (BLUPs) computed for each hybrid.
From a breeding standpoint, both an appreciable mean and a large variance must be present if plant improvement via plant breeding is to be successful. The mean hydroxycinnamic acid content must be relatively large in comparison to other plant sources, or at least present at considerably large concentrations, to make the extraction of hydroxycinnamic acids from maize economical. A large variance is needed to maximize the expected selection response as highlighted by the following equation:

\[ R = h^2 \times S = h^2 \times i_\alpha \times \sigma_p, \]

where \( R = \mu_{n+1} - \mu_n \) (i.e., the difference between the mean of the generation after selection and the mean of the generation in which selection was performed), \( h^2 \) is the narrow-sense heritability, \( S = \mu_S - \mu_n \) (i.e. the difference between the mean of the selected individuals and the mean of all individuals in the generation in which selection is performed), \( i_\alpha \) is a constant which represents the selection intensity, and \( \sigma_p \) is the phenotypic standard deviation of the generation in which selection is performed (or of the original population) and is also equal to \( \sqrt{\sigma_p^2} \). Therefore, as the phenotypic variance increases, so does the potential response to selection provided the other values in the model are held constant.

This study demonstrated that the amount of ferulic acid and p-coumaric acid extractable in their soluble, potentially bioavailable forms from maize is both variable and appreciable. Also, heritability estimates indicated that much of the phenotypic variation observed in this study was due to genetic variation rather than variation due to environmental effects or genotype-by-environment interactions. Ferulic acid had a narrow-sense heritability of 69.0% and a broad-
sense heritability of 85.4%. p-Coumaric acid had a narrow-sense heritability of 79.2% and a broad-sense heritability of 87.2%.

The similarity in narrow-sense heritability and broad-sense heritability of both ferulic acid content and p-coumaric acid content also indicates that the gene action controlling these two traits is mostly additive. Calculations of midparent heterosis from the BLUPs also demonstrates the importance of additive effects in both of these traits. Ferulic acid content has a midparent heterosis of -2.86%, whereas p-coumaric acid has a midparent heterosis of -18.52%. The quantitative and additive nature of the ferulic acid and p-coumaric acid content make these traits relatively simple for plant breeders to improve and may even aid in the identification of the gene pathways involved in the expression of these traits in future studies. This, in turn, would enable the use of marker assisted selection (MAS) in breeding for improved hydroxycinnamic acid content in maize. However, the largely additive nature of these traits also implies that both heterotic pools must be improved in maize breeding programs.

The high broad-sense heritability seen in ferulic acid and p-coumaric acid has implications regarding the use of open-pollinated seed in the evaluation of hydroxycinnamic acids. Seed quality traits are sometimes influenced by xenia effects. For instance, Letchworth and Lambert (1998) found that xenia effects were significant in determining oil concentration but not starch or protein concentrations. Therefore, the authors concluded that open-pollinated ears could be used to evaluate starch and protein content, but not oil content. The seed analyzed in this study was open-pollinated and harvested from either an F₁ hybrid or an inbred (i.e. the maternal plant was either an F₁ hybrid or an inbred). The maternal plant’s genotype was used in
the calculation of the broad- and narrow-sense heritability. Most of the hydroxycinnamic acids are localized in the bran (Rumpagapom, 2011), and the bran (or the seed coat) is predominantly maternal tissue. One possible hypothesis, therefore, is that the concentration of hydroxycinnamic acids is predominantly maternally influenced. Given the high broad-sense heritability seen here, it appears that the hydroxycinnamic acid content of the seed is predominantly maternally influenced, i.e., the hydroxycinnamic acid content of the harvested seed is not significantly influenced by the pollen source. Also, due to the apparent lack of xenia effects, it appears that maize breeders would be able to evaluate open-pollinated maize hybrids or inbreds in their breeding programs.

Since ferulic acid content has a moderately-high narrow-sense heritability and p-coumaric acid content has a high narrow-sense heritability, this means that much of the genetic variation seen is directly heritable from one generation of plant advancement to the next in a plant breeding program. However, the response to selection (R, see equation above) is influenced by both the heritability and the phenotypic variability of a trait. Figures 4.2 and 4.3 show the expected selection responses of ferulic acid and p-coumaric acid at different selection intensities. As can be seen in Figure 4.2, a selection intensity of 10% corresponds to a selection response of 335.02 μg / g for ferulic acid, and a selection intensity of 5% corresponds to a selection response of 393.76 μg / g for ferulic acid. These calculated values for selection response indicate that ferulic acid could very feasibly be improved. In regards to p-coumaric acid, selection intensities of 10% and 5% correspond to selection responses of 1.65 μg / g and 1.80 μg / g, respectively (Fig. 4.3). This indicates that while the narrow-sense heritability of p-coumaric acid may be somewhat stronger than that of ferulic acid, the much greater phenotypic
variability in ferulic acid content allows for a much greater selection response in ferulic acid content. Additionally, since there is a large phenotypic variance in ferulic acid content present in typical Midwestern maize germplasm, it appears that breeders would be able to improve the ferulic acid content of maize using Midwestern maize germplasm rather than less developed germplasm sources. Therefore, not only do breeders have a germplasm source that appears to display a fairly rapid response to selection; the source has also already been developed to withstand the biotic and abiotic stresses encountered in the Midwestern U.S. and to be relatively productive in Midwestern cropping systems.

Furthermore, maize is a highly concentrated, inexpensive source of ferulic acid, relative to some fresh produce and other food products. Table 4.3 provides a listing of the ferulic acid content of various food sources examined in the literature. A more complete list can be found in Zhou and Moghadasian’s (2008) literature review, and this review is also supportive of using maize grain or even maize byproducts as a source of ferulic acid. Building upon the current literature base, this study indicates that typical, commercial maize is a much more concentrated source of ferulic acid than other food sources. However, it should also be noted that there are more concentrated sources of hydroxycinnamic acids than maize grain. This study only focuses on maize grain and suggests that it may be a candidate source of hydroxycinnamic acids for food additive production. More research is needed in this area to identify the most economical and practical source for food additive production.

If ferulic acid is to be extracted from natural plant sources for use as a food additive, then yellow dent maize grain is an inexpensive substrate. The parental inbreds used in this study
represent the genetic makeup of food grade corn used in the production of breakfast cereals as well as corn hybrids grown in the American Midwest. This grain is used in animal feed, ethanol production, corn syrup, and the production of various industrial products. Therefore, extracting food additives from yellow dent maize rather than sweet corn, popcorn, or other maize types most likely would enable the production of food additives from an inexpensive substrate which would not detract greatly from the U.S. food supply.

Furthermore, since maize is used in a number of industrial processes, it is possible that the hydroxycinnamic acids could be extracted as a byproduct of an industrial process. A similar approach was taken by Acosta-Estrada et al. (2014) in which soluble hydroxycinnamic acids were extracted from maize during the processing of tortilla chips and added to bread. This did not alter the taste, texture, or color of the fortified bread in comparison to bread in which the hydroxycinnamic acids were not added. Additionally, any extraction procedure making use of an alkaline reaction would break the ester bonds which bind the hydroxycinnamic acids to the cell wall arabinoxylans. This would release the hydroxycinnamic acids into their soluble, bioavailable state. Therefore, the extraction procedure can be incorporated into existing maize processing procedures (Thomas Patterson, Dow AgroSciences, and Roger Faughn, Engineering Consultant, Personal Communication, 2015). This also indicates that food additives of hydroxycinnamic acids can be extracted from maize and added to a variety of processed foods, not only processed maize food products.
Total Harvestable Hydroxycinnamic Acids and Yield Per Hectare

One of the main goals of plant breeding is to improve multiple traits simultaneously. Therefore, it was important to identify hybrids which exhibited favorable hydroxycinnamic acid contents and grain yields. However, it can also be argued that perhaps the total harvestable hydroxycinnamic acid content per hectare (or acre) is more important in a system in which hydroxycinnamic acids are to be extracted from maize and used as food additives. The reason for this is that as the amount of hydroxycinnamic acids per hectare increases, the number of people whose diets can be positively influenced by hydroxycinnamic acids increases.

On average, the hybrids in this study produced \(207.62 \times 10^6\) mg per hectare (\(84.02 \times 10^6\) mg per acre) of ferulic acid and \(1.92 \times 10^6\) mg per hectare (\(0.78 \times 10^6\) mg per acre) of p-coumaric acid. The top ten percent of the observations in this study contained more than \(241.40 \times 10^6\) mg / Ha (\(97.69 \times 10^6\) mg / ac) of ferulic acid. As a comparison, many marketable food supplements contain approximately 200 mg of ferulic acid per tablet, and it is suggested that two tablets be taken twice a day. At this rate of 400 mg of ferulic acid per person per day, a single hectare has the potential to provide enough ferulic acid for 1,653.4 people for an entire year. Since the expected number of hectares planted to corn in the United States this year is \(37.88 \times 10^6\) hectares (\(93.60 \times 10^6\) acres), the percentage of U.S. corn acres needed to meet half of the United States’ population’s ferulic acid intake (at 400 mg / day) is 0.26%. Furthermore, this number was calculated using older germplasm, and the yield of modern hybrids is much greater.
As can be seen in Figure 4.4, it appears that the total amount of ferulic acid per hectare tends to increase as the yield increases. Thus, it should be possible to improve both traits concurrently. Index selection might prove useful, especially if using an index that makes use of the profitability of these traits. However, without a currently well-established market for hydroxycinnamic acids, it may be best to use a composite trait, such as the total amount of harvestable ferulic acid per hectare in conjunction with a threshold yield value, until such a market is established. Additionally, the amount of ferulic acid per hectare in modern hybrids may be greater, as well. While a market for the production of food additives from maize most likely will not greatly detract from the current food supply, its small size may also not attract the attention of the seed industry. Such a breeding program may be better suited for university research programs.

Figure 4.4 also shows that it should be possible to select for hybrids which yield well and which produce high amounts of both ferulic acid and p-coumaric acid per hectare. These hybrids are represented by red dots in Figure 4.4. The five hybrids which displayed a favorable combination of characteristics are PHG39×PHG47, B73×Mo17, PHG39×PHZ51, PHG39×Mo17, and PHG39×LH123HT. In these five hybrids, the grain yield ranged from 11.19 Mg / Ha (178.52 bu / ac) to 11.96 Mg / Ha (190.72 bu / ac). The total amount of ferulic acid per hectare ranged from $243.58 \times 10^6$ mg / Ha ($98.57 \times 10^6$ mg / ac) to $275.30 \times 10^6$ mg / Ha ($111.41 \times 10^6$ mg / ac). Lastly, the total amount of p-coumaric acid per hectare ranged from $1.70 \times 10^6$ mg / Ha ($0.69 \times 10^6$ mg / ac) to $3.23 \times 10^6$ mg / Ha ($1.31 \times 10^6$ mg / ac). Therefore, it should be possible to breed for improved hydroxycinnamic acid yield and grain
yield concurrently, provided no strong and negative correlations exist between the hydroxycinnamic acid contents and important agronomic traits of interest.

Unsurprisingly, grain yield, ferulic acid per hectare, and p-coumaric acid per hectare appear to increase concurrently. Interestingly, there were some hybrids which produced high levels of ferulic acid per hectare and only moderate levels of p-coumaric acid per hectare, and vice versa. This may be due to varying rates of conversion of p-coumaric acid to ferulic acid. p-Coumaric acid is a precursor to ferulic acid. Specifically, p-coumaric acid is acted on by p-coumaric hydroxylase and then o-methyltransferase to produce ferulic acid. It is possible that hybrids that have high levels of ferulic acid per hectare but only moderate levels of p-coumaric acid per hectare may convert p-coumaric acid to ferulic acid relatively quickly. Conversely, hybrids that exhibit high levels of p-coumaric acid per hectare but only moderate levels of ferulic acid per hectare may not be able to convert p-coumaric acid into ferulic acid very quickly, possibly due to a reduced presence of the enzymes which carry out these processes. This warrants further investigation into the variability in the expression of p-coumaric hydroxylase and o-methyltransferase in Midwestern maize germplasm.

Correlations Between Key Traits of Interest

The Pearson correlation coefficients between yield, test weight, extractable ferulic acid content, and extractable p-coumaric acid content are shown in Table 4.4. In the inbreds, none of the correlations were significant. In the hybrids, there was a moderate and positive correlation between ferulic acid content and p-coumaric acid content and a very weak but negative correlation between p-coumaric acid content and test weight. Therefore, it appears that breeding
for ferulic acid and p-coumaric acid content in maize can transpire without negatively impacting the yield or test weight of commercial hybrids.

Under adverse biotic or abiotic stresses, these correlations could have been different. Ferulic acid and p-coumaric acid content are positively correlated with resistance to feeding by storage insects as well as resistance to ear pathogen infection (Arnason et al., 1992; Assabgui et al., 1993; Bily et al., 2003; Classen et al., 1990; Dowd, 1994; Dowd et al., 1997; García-Lara et al., 2004; Serratos et al., 1987). Plant breeders and early agriculturists selected for crops that would perform well under adverse conditions, including insect and pathogen pressure. It is therefore unsurprising that the average concentrations of these two phytochemicals are relatively high in maize. It is also not surprising that considerable variation exists in these phytochemicals and that there is not a significant correlation between these phytochemicals and agronomic traits of interest. With the exception of instances in which plants suffer from either biotic or abiotic stresses, hydroxycinnamic acids do not aid in preventing plant damage (Dixon and Paiva, 1995) and subsequent yield loss. Thus, while maize breeders have selected primarily for yield during the last several decades, they have not necessarily selected for or against hydroxycinnamic acid content.

If breeding maize for use in the manufacturing of food additives is to be successful, it is necessary that hydroxycinnamic acid content be genetically controlled, be variable between hybrids, be relatively high in comparison to other plant sources, and not be strongly and negatively correlated with important agronomic traits of interest, namely yield and test weight. This research shows that all of these qualifications have been met. Furthermore, the extraction
procedure from maize can be incorporated into pre-existing maize industrial processes, and other research indicates that food additives can successfully be added to processed food products. Therefore, while more research is needed, it appears that these food additives can be added to a variety of inexpensive food products. This would extend the chemopreventive health benefits of hydroxycinnamic acids to all individuals, regardless of socioeconomic status. In conclusion, this method of improving the chemopreventive phytochemical content of maize food products (and other food products) is much more efficient and successful than attempting to select maize hybrids which maintain favorable phytochemical properties throughout processing.
**Table 4.1. Significance of ANOVA model effects for ferulic acid and p-coumaric acid and their variance component estimates**

<table>
<thead>
<tr>
<th>Term</th>
<th>Mean Square</th>
<th>F Value</th>
<th>P-Value</th>
<th>Variance Estimate ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ferulic Acid</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>3,738,835</td>
<td>5.73</td>
<td>0.03</td>
<td>16,579.00</td>
</tr>
<tr>
<td>Rep(Year)</td>
<td>592,106</td>
<td>17.92</td>
<td>&lt;0.01</td>
<td>9,025.78</td>
</tr>
<tr>
<td>Block(Rep×Year)</td>
<td>29,435</td>
<td>1.37</td>
<td>0.02</td>
<td>2,250.59</td>
</tr>
<tr>
<td>Genotype</td>
<td>465,933</td>
<td>6.68</td>
<td>&lt;0.01</td>
<td>56,556.00</td>
</tr>
<tr>
<td>Genotype×Year</td>
<td>71,164</td>
<td>3.32</td>
<td>&lt;0.01</td>
<td>20,706.00</td>
</tr>
<tr>
<td>Error</td>
<td>21,425</td>
<td></td>
<td></td>
<td>21,2425.00</td>
</tr>
<tr>
<td><strong>p-Coumaric Acid</strong> †</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>1.01</td>
<td>3.84</td>
<td>0.06</td>
<td>0.0041</td>
</tr>
<tr>
<td>Rep(Year)</td>
<td>0.22</td>
<td>24.75</td>
<td>&lt;0.01</td>
<td>0.0034</td>
</tr>
<tr>
<td>Block(Rep×Year)</td>
<td>0.01</td>
<td>1.14</td>
<td>0.21</td>
<td>0.0002</td>
</tr>
<tr>
<td>Genotype</td>
<td>0.66</td>
<td>14.16</td>
<td>&lt;0.01</td>
<td>0.0897</td>
</tr>
<tr>
<td>Genotype×Year</td>
<td>0.05</td>
<td>6.48</td>
<td>&lt;0.01</td>
<td>0.0171</td>
</tr>
<tr>
<td>Error</td>
<td>0.01</td>
<td></td>
<td></td>
<td>0.0074</td>
</tr>
</tbody>
</table>

† The values in this table for p-coumaric acid are based off of the natural log transformed p-coumaric content.
‡ Variance estimates were computed directly from SAS 9.3. The units of the variance estimates for ferulic acid content are in (µg/g)^2. The units of the variance estimates for p-coumaric acid are in (ln(µg/g))^2.

**Table 4.2. Summary statistics of extractable ferulic acid and p-coumaric acid content.**

Summary statistics were calculated from the BLUPs of each of ferulic acid content and the natural log of p-coumaric acid content. The non-transformed data column for p-coumaric acid content is the exponential of the corresponding values in the third column.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Ferulic Acid (µg/g)</th>
<th>p-Coumaric Acid (ln(µg/g))</th>
<th>p-Coumaric Acid (non-transformed) µg/g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>2,759.90</td>
<td>5.66</td>
<td>314.70</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>2,114.97</td>
<td>5.36</td>
<td>212.72</td>
</tr>
<tr>
<td>Median</td>
<td>1,866.18</td>
<td>5.19</td>
<td>179.47</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>1,682.99</td>
<td>4.92</td>
<td>137.00</td>
</tr>
<tr>
<td>Minimum</td>
<td>1,234.63</td>
<td>4.38</td>
<td>75.83</td>
</tr>
<tr>
<td>Range</td>
<td>1,525.27</td>
<td>1.28</td>
<td>238.87</td>
</tr>
<tr>
<td>Overall Mean</td>
<td>1,902.44</td>
<td>5.14</td>
<td>170.72</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>224.00</td>
<td>0.30</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3. Ferulic acid content of other grains and food sources. Most of the ferulic acid in grains is available in the insoluble-bound form. Ferulic acid is primarily available in the soluble form in fruits and vegetables. As can be seen from this table, maize is the most concentrated source of extractable ferulic acid. This amount is also much higher than the concentration of soluble, bioavailable ferulic acid in many fruits and vegetables. For a more complete listing of the ferulic acid content in various food products, please refer to Zhou and Moghadasian (2008).

<table>
<thead>
<tr>
<th>Food Source</th>
<th>Ferulic Acid µg/g</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>1902.4</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>646.8</td>
<td>Adom and Liu, 2002</td>
</tr>
<tr>
<td>Oats</td>
<td>358.6</td>
<td>Adom and Liu, 2002</td>
</tr>
<tr>
<td>Rice</td>
<td>297.9</td>
<td>Adom and Liu, 2002</td>
</tr>
<tr>
<td>Grapefruit</td>
<td>116.0</td>
<td>Mattila et al., 2006</td>
</tr>
<tr>
<td>Orange</td>
<td>94.0</td>
<td>Mattila et al., 2006</td>
</tr>
<tr>
<td>Bananas</td>
<td>54.0</td>
<td>Mattila et al., 2006</td>
</tr>
<tr>
<td>Rhubarb</td>
<td>20.0</td>
<td>Mattila et al., 2006</td>
</tr>
<tr>
<td>Blueberry</td>
<td>12.9</td>
<td>Mattila et al., 2006</td>
</tr>
<tr>
<td>Avocado</td>
<td>11.0</td>
<td>Mattila and Hellström, 2007</td>
</tr>
<tr>
<td>Cherries</td>
<td>4.6</td>
<td>Mattila et al., 2006</td>
</tr>
</tbody>
</table>

Table 4.4. Correlations between key agronomic and phytochemical traits. Correlations were calculated for hybrids and inbreds separately.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Ferulic Acid</th>
<th>ln of p-Coumaric Acid</th>
<th>Grain Yield</th>
<th>Test Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferulic Acid</td>
<td></td>
<td>0.3748</td>
<td>0.1011</td>
<td>-0.0305</td>
</tr>
<tr>
<td>ln of p-Coumaric Acid</td>
<td>0.1369</td>
<td></td>
<td>0.0924</td>
<td>-0.1403</td>
</tr>
<tr>
<td>Grain Yield</td>
<td>-0.0927</td>
<td>0.2698</td>
<td></td>
<td>-0.4193*</td>
</tr>
<tr>
<td>Test Weight</td>
<td>0.0937</td>
<td>0.1530</td>
<td>-0.1508*</td>
<td></td>
</tr>
</tbody>
</table>

Correlations among hybrids are shown in the gray boxes, whereas correlations among inbreds are shown in the white boxes. Significant correlations are shown in red and bold font.

\*The correlations here differ slightly from those reported in Macke et al. (2016) due to the more stringent outlier removal conducted in this study.
Figure 4.1. Profile plot of the interaction between maize genotypes and their environments
Figure 4.2. Selection response for ferulic acid content. The expected selection response for improvement of ferulic acid content is shown at different selection intensities.

Figure 4.3. Selection response for p-coumaric acid content. The expected selection response for improvement of p-coumaric acid content is shown at different selection intensities.
Figure 4.4. Three-dimensional scatterplot of ferulic acid per hectare, p-coumaric acid per hectare, and grain yield. The five hybrids which exhibited the best combination of all three traits are highlighted in red.


CHAPTER 5: PREDICTING PHENOLIC ACID CONTENT IN MAIZE USING NIR SPECTROSCOPY AND INDIRECT SELECTION APPROACHES

ABSTRACT

Value-added traits are often difficult or time-consuming to measure. Both Near Infrared Spectroscopy (NIR) prediction models and prediction models which utilize indirect selection offer high-throughput alternatives to the laborious measures that are involved in the phenotyping of some traits. This study examines the feasibility of using either of these approaches to predict ferulic acid content and p-coumaric acid content in maize. A total of 702 ground maize kernel samples obtained from field replications of 78 entries were analyzed for their p-coumaric acid content and ferulic acid content. All samples were used to build NIR prediction models which were cross-validated using validation data sets. Both NIR models had poor RPD scores with the larger RPD score being 1.6 for ferulic acid content and the smaller RPD score being 1.13 for p-coumaric acid content. Agronomic, ear, cob, and kernel traits which were relatively easy to evaluate were used in an attempt to predict ferulic acid and p-coumaric acid content via indirect selection methods. Unfortunately, based upon low $R^2$ values, multiple regression for indirect selection also proved to be unreliable. Discriminant analysis was also attempted as an alternative to multiple regression, but this was likewise unsuccessful. Therefore, the most high-throughput and reliable method in this study for phenotyping the phenolic acid content in maize was high-throughput wet-lab chemistry.
INTRODUCTION

The quantification of value-added traits is often laborious. Examples of such traits include nutritional quality traits. For instance, while ferulic acid content and p-coumaric acid content can be quantified concurrently, past experience indicated that only 200 samples per week could be analyzed. In comparison, Near Infrared Reflectance Spectroscopy (NIR or NIRS) scans can be completed and electronically stored in approximately one minute per sample using computer programs like Bruker’s OPUS software.

NIR is a spectroscopic method that can be used in the quantification of various chemical compounds, including the various phytochemicals present in plant tissues. NIR makes use of the absorbance of the near infrared spectrum of light. Different types of chemical bonds tend to absorb light at different near infrared frequencies, thereby enabling the relative quantification of these types of bonds. From this information, multivariate statistics techniques (principal components, partial least squares regression models, etc.) can be used in building an NIR prediction model for a compound of interest.

Other researchers have had considerable success using NIR in predicting complex traits. NIR has successfully been used in the prediction of phenolic acid content in other crops (Ferrer-Gallego et al., 2011; Li and Qu, 2010). Ferrer-Gallego, et al. (2011) reported an $R^2$ of 0.947 and a Ratio of Performance to Deviation (RPD) of 6.3 when attempting to predict the phenolic acid content of grape skins using NIR. Li and Qu (2010) reported $R^2$ values greater than 0.85 for a number of different phenolic acids. While the results of these studies are promising, they do not necessarily mean that NIR can successfully be used in all instances to predict the phenolic acid
content in all plants and plant parts. For instance, Ferrer-Gallego, et al. (2011) used NIR to predict the phenolic acid content in grapes, not maize. Likewise, Li and Qu (2010) studied red sage, not maize. It is possible that the different phytochemical composition of maize may impact the ability of NIR to accurately predict the phenolic acid content in maize. To the author’s knowledge, no other study has examined the ability of NIR to predict the phenolic acid content in maize. Of the phenolic acids typically found in grapes and red sage, ferulic acid and p-coumaric acid are not among them. It is also possible that the reliability of a model can change based upon the genotypes under study and the environment in which they were grown (James Behrman, Personal Communication, 2015). This study evaluates the reliability of NIR prediction models for a large population of maize genotypes and for maize which was grown under non-irrigated conditions in Central Illinois.

Some multivariate statistics techniques, including discriminant analysis and multivariate regression, may drastically increase the speed of phenotyping complex traits. These statistical techniques, if successful, would enable the use of indirect selection. For these methods to be successful, a trait or multiple traits must be highly correlated with the trait of interest and both the correlated trait(s) and the trait of interest must be heritable. Furthermore, the traits which are used in building the prediction model should be collectively easier and/or less expensive to phenotype than the trait of interest. The overall objective of this study was to develop high-throughput techniques for the prediction of ferulic acid and p-coumaric acid in maize whole kernels. Several different prediction techniques were used, including NIR, multiple regression, and discriminant analysis. The reliability of each of these models was tested using cross-validation techniques.
MATERIALS AND METHODS

Plant Materials

The ground plant materials used in Chapter 4 were used here, as well. These plant materials were comprised of 702 samples obtained from 78 maize entries grown during 2009-2011 in a resolvable incomplete block design with three replications each year. Each NIR vial was filled halfway with a single ground maize sample.

Collection of Phenotypic Information

The ferulic acid and p-coumaric acid contents reported in Chapter 4 were used in this study. Briefly, the ferulic acid and p-coumaric acid contents were quantified using the protocol outlined in Butts-Wilmsmeyer and Bohn (2016). If there were technical replicates (i.e. subsamples), the average ferulic acid and p-coumaric acid contents of these subsamples were used as the final measured ferulic acid content and p-coumaric acid content.

NIR Spectra Collection

NIR spectra were collected using an Agilent NIR spectrophotometer and the Bruker Corporation’s OPUS software. Absorbance was measured between 3,000 and 12,500 wave numbers. A spectra for each sample was stored electronically in OPUS.

Creation of NIR Prediction Models

Spectra were normalized before being used to build a model. Partial least squares regression was then used to fit the model. The OPUS software internally calculated the best model based on optimization criterion provided by the user. Potential optimization criterion used
in this experiment included first derivative, standard normal variate (SNV), and multiplicative scatter correction (MSC) optimization techniques. Ratio of performance to deviation (RPD) and $R^2$ values were used in judging the reliability of the model, as were the number of partial least squares factors. Too many partial least squares factors in the model could lead to overfitting the model. Models with RPD values less than 3 were considered poor, models with RPD values between 3 and 5 were considered fair enough for categorical class assignment (i.e. high/low values for traits of interest), and models with RPD values greater than 5 were considered excellent and suitable for a quantitative prediction model (Williams, 2004). For each trait, ten percent of the values were randomly selected for building the calibration model. The remaining values were used as the validation set for cross-validating the model.

**Indirect Selection Techniques**

Plant height (PHT), ear height (EHT), days to silking (DTS), days to pollination (DTP), grain yield (GY), test weight (TW), kernel width (KW), kernel length (KL), kernel depth (KD), hundred kernel volume (HKV), hundred kernel weight (HKWt), number of rows (R), kernels per row (KPR), fill length (FL), ear width (EW), ear size (ES), cob length (CL), cob width (CB), cob circumference (CC), ear circumference (EC), and ear weight were all measured as part of Macke, et al. (2016). Table 1 contains a complete description of these traits. All these traits can be measured with simple tools and do not require the use of wet lab chemistry for their evaluation. In contrast to ferulic acid content and p-coumaric acid content, these traits are often routinely evaluated in plant breeding programs. Therefore, these agronomic, ear, cob, and kernel traits were included in a stepwise regression model so as to predict each of the value-added traits of
interest. For the remainder of this chapter, the agronomic, ear, cob, and kernel traits named above will be collectively referred to as prediction traits.

*Extraction of Principal Components*

Principal components were extracted from the correlation matrix of the prediction variables named above using PROC PRINCOMP of SAS (version 9.3). Principal components with eigenvalues greater than 1 were maintained for future stepwise linear regression and discriminant analysis models.

*Stepwise Linear Regression and Robust Regression*

All prediction traits were included in a stepwise linear regression model. This analysis was completed in PROC REG of SAS (version 9.3) using method=STEPWISE. SLENTRY and SLSTAY levels were set to 0.15. If many influential points were present, as indicated by Cook’s D, then robust regression was conducted instead. Robust regression was conducted in PROC ROBUSTREG of SAS with M estimation. The Huber weight function and weighted convergence criteria were specified. The same prediction variables that were identified by stepwise linear regression as being significant to the model were used in the robust regression model. Models with $R^2$ values less than 0.75 were considered poor indirect selection models.

As an alternative to using the prediction traits as part of the linear regression model, principal components were used. Again, PROC REG of SAS was used to conduct stepwise selection in a linear regression model. If many influential points were present, then robust regression was conducted instead with the same methods as previously discussed.
Discriminant Analysis

If stepwise linear regression or robust regression proved to be ineffective, then discriminant analysis was conducted. Two new classification variables, namely FERclass and pCOclass, were created. PROC UNIVARIATE was used in order to calculate the summary statistics for ferulic acid content and p-coumaric acid content. These results were used to establish a threshold at which there was a clear division between high values and low values for each trait. Table 2 describes these classification variables in more detail. Each observation was assigned a value of “Low” or “High” for each of the four new classification variables, depending on where the value fell relative to the established threshold.

Stepwise discriminant analysis was conducted in PROC STEPDISC with SLENTRY set to 0.2 and SLSTAY set to 0.1 for each value-added trait. All prediction traits were included in this procedure. Traits selected by the stepwise discriminant analysis were included in a discriminant analysis in which the homogeneity of the variance-covariance matrices of each group could be tested and cross-validation (jackknifing) could be used to validate the accuracy of the discriminant analysis model generated. Models with error rates greater than 0.15 were considered undesirable. This same procedure was conducted again using principal components rather than the prediction traits.
RESULTS AND DISCUSSION

Usefulness of NIR

The validation results of the optimal NIR models generated by the Bruker Corporation’s OPUS software can be seen in Figures 5.1-5.2. Both models possessed incredibly low RPD values, indicating that neither quantitative nor qualitative NIR models can be successfully employed in the prediction of ferulic acid or p-coumaric acid. Any model with an RPD value less than 2.3 is considered to be very poor (Williams, 2004), and the RPD values of the models in this study ranged from 1.13 for p-coumaric acid content to 1.6 for ferulic acid content. In additional support of this conclusion, all $R^2$ values were very low. $R^2$ values ranged from 0.22 for p-coumaric acid content to 0.61 for ferulic acid content.

Many different types of phytochemicals exist in biological materials that can make NIR prediction models difficult to build. NIR works particularly well in instances in which either an atypical bond (e.g. a sulfur containing bond) is present in the compound of interest or when the compound of interest is present at high concentrations relative to the other phytochemicals in the plant. One of the most likely reasons that the NIR prediction models did not work for either of the traits examined is that these phytochemicals do not contain atypical chemical bonds for biological systems. For instance, ferulic acid and p-coumaric acid are composed of a phenolic ring attached to an aliphatic chain which terminates in a carboxyl group. The only difference between ferulic acid and p-coumaric acid is the presence of a methoxyl group attached to the 3’ position of the phenolic ring in ferulic acid (Fig. 5.3, 5.4). Additionally, ferulic acid and p-coumaric acid were both present at relatively small levels in comparison to other phytochemicals in maize. At an average concentration of 1902.4 $\mu$g / g of ferulic acid and 170.72 $\mu$g / g of p-
coumaric acid, the typical percentage of these two phytochemicals in maize is 1.90% and 0.17%, respectively. Therefore, due to small concentrations and the presence of relatively common chemical bonds in these phytochemicals, NIR does not appear to be a useful alternative high-throughput phenotyping technique for ferulic acid or p-coumaric acid content in maize. However, it should also be noted that the findings of this study may have been different if more powerful NIR equipment had been used.

**Usefulness of Indirect Selection Techniques**

Table 1 contains a description of each of the prediction traits or principal components included in the indirect selection models. Five principal components were extracted from the correlation matrix of the prediction traits. Principal component 1 appeared to involve ear weight and ear size. Principal component 2 appeared to be a contrast between kernel width and the cob traits cob width, cob circumference, and the number of rows. The third principal component was a contrast between grain yield and the quality traits hundred kernel volume, hundred kernel weight, and kernel depth. The fourth principal component was a contrast between flowering traits and test weight. The fifth and final principal component was a contrast between kernel length and the four traits kernel depth, test weight, fill length, and cob length. Regardless of whether the prediction traits themselves were used in building the indirect selection model or their principal components were used, no model had a high $R^2$ value (Table 5.3). Therefore, indirect selection cannot be used at this time to predict ferulic acid content or p-coumaric acid content.
There are a number of possible reasons for this outcome. First, none of the prediction traits were highly correlated with either of the traits of interest (Table 5.4). Considering that the metabolic pathways that affect the expression of ferulic acid content and p-coumaric acid content are not directly tied to the expression of the prediction traits, this finding is not highly surprising. Another possible reason for these results is that while a diverse set of Midwestern maize germplasm was used, this set of germplasm is strictly yellow dent maize from the U.S. Midwest which was grown in Urbana, IL. Had a more diverse set of germplasm and phenotypes been measured (i.e. germplasm which included tropical lines, European lines, popcorn, etc.), then it might have been possible to identify correlations between traits of interest and prediction traits. However, due to the lack of variability in both environment and genotype, the phenotypes of all samples analyzed in this study were that of a typical yellow dent hybrid or inbred grown in a relatively ideal environment with only minor stress.

A third possible reason for this outcome is that traits which might be more tightly correlated with phenolic acid content were not measured. A requirement for indirect selection to be successful is that the trait of interest and the trait(s) which are used in the prediction model must be correlated. Ferulic acid content and p-coumaric acid content are directly tied to pathogen and insect resistance (Arnason et al., 1992; Assabgui et al., 1993; Bily et al., 2003; Classen et al., 1990; Dowd et al., 1997; García-Lara et al., 2004). These traits can all be fairly rapidly measured, but they were not measured as part of this study. Perhaps future studies will prove more successful with the inclusion of these traits in their analyses.
### Table 5.1. Description of traits used in stepwise linear regression analysis and discriminant analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Trait</th>
<th>Acronym</th>
<th>Basis</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agronomic</td>
<td>Plant height</td>
<td>PHT</td>
<td>whole plot</td>
<td>Distance from ground to node with flag leaf</td>
<td>cm</td>
</tr>
<tr>
<td></td>
<td>Ear height</td>
<td>EHT</td>
<td>whole plot</td>
<td>Distance from ground to node with ear</td>
<td>cm</td>
</tr>
<tr>
<td></td>
<td>Days to silking</td>
<td>DTS</td>
<td>whole plot</td>
<td>Days from planting to 50% silk emergence</td>
<td>days</td>
</tr>
<tr>
<td></td>
<td>Days to pollination</td>
<td>DTP</td>
<td>whole plot</td>
<td>Days from planting to 50% pollen shed</td>
<td>days</td>
</tr>
<tr>
<td></td>
<td>Grain yield</td>
<td>GY</td>
<td>whole plot</td>
<td>Weight of grain per plot at 15.5% moisture</td>
<td>Mg / ha</td>
</tr>
<tr>
<td>Dry Milling</td>
<td>Test weight</td>
<td>TW</td>
<td>whole plot</td>
<td>Weight of 1 bushel of grain adjusted to 15.5% moisture</td>
<td>kg / hL</td>
</tr>
<tr>
<td>Ear</td>
<td>Cob length</td>
<td>CL</td>
<td>5 ears</td>
<td>Length of whole ear including cob</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Cob width</td>
<td>CW</td>
<td>5 ears</td>
<td>Width of cob at median length</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Cob circumference</td>
<td>CC</td>
<td>5 ears</td>
<td>Circumference of shelled cob at median length</td>
<td>mm</td>
</tr>
<tr>
<td>Ear fill length</td>
<td>FL</td>
<td>5 ears</td>
<td></td>
<td>Length of ear with kernels</td>
<td>mm</td>
</tr>
<tr>
<td>Ear width</td>
<td>EW</td>
<td>5 ears</td>
<td></td>
<td>Width of ear at median length</td>
<td>mm</td>
</tr>
<tr>
<td>Ear circumference</td>
<td>EC</td>
<td>5 ears</td>
<td></td>
<td>Circumference of cob with kernels at median length</td>
<td>mm</td>
</tr>
<tr>
<td>Ear rows</td>
<td>R</td>
<td>5 ears</td>
<td></td>
<td>Number of kernel rows per ear</td>
<td>count</td>
</tr>
<tr>
<td>Ear weight</td>
<td>EWt</td>
<td>5 ears</td>
<td></td>
<td>Mass of all kernels from ear and cob</td>
<td>g</td>
</tr>
<tr>
<td>Ear size</td>
<td>ES</td>
<td>5 ears</td>
<td></td>
<td>FL × EW × π</td>
<td>mm²</td>
</tr>
<tr>
<td>Kernels/row</td>
<td>KPR</td>
<td>5 ears</td>
<td></td>
<td>Number of kernels per row</td>
<td>count</td>
</tr>
<tr>
<td>Hundred kernel volume</td>
<td>HKV</td>
<td>5 ears</td>
<td></td>
<td>100 kernels measured in graduated cylinder</td>
<td>mm³</td>
</tr>
<tr>
<td>Hundred kernel weight</td>
<td>HKWt</td>
<td>5 ears</td>
<td></td>
<td>Weight of 100 kernels</td>
<td>g</td>
</tr>
<tr>
<td>Kernel</td>
<td>Kernel width</td>
<td>KW</td>
<td>10 kernels 2x</td>
<td>Length of 10 kernels arranged side by side</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Kernel length</td>
<td>KL</td>
<td>10 kernels 2x</td>
<td>Length of 10 kernels arranged tip to crown</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Kernel depth</td>
<td>KD</td>
<td>10 kernels 2x</td>
<td>Length of 10 kernels stacked on top of each other</td>
<td>mm</td>
</tr>
<tr>
<td>Principal Components</td>
<td>Principal component 1</td>
<td>correlation matrix</td>
<td></td>
<td>Positive loadings on both EWt and ES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal component 2</td>
<td>correlation matrix</td>
<td></td>
<td>Contrast between KW and the group of CW, CC, and R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal component 3</td>
<td>correlation matrix</td>
<td></td>
<td>Contrast between GY and the group of HKV, HKWt, and KD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal component 4</td>
<td>correlation matrix</td>
<td></td>
<td>Contrast between TW and the group of DTP and DTS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principal component 5</td>
<td>correlation matrix</td>
<td></td>
<td>Contrast between KL and the group of KD, TW, FL, and CL</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2. Description of classification variables used in discriminant analysis

<table>
<thead>
<tr>
<th>Classification Variable</th>
<th>High/Low</th>
<th>Content Range (μg / g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FERclass</td>
<td>High</td>
<td>&gt; 1900</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>≤ 1900</td>
</tr>
<tr>
<td>pCOclass</td>
<td>High</td>
<td>&gt; 200</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>≤ 200</td>
</tr>
</tbody>
</table>

Table 5.3. $R^2$ values and error rates associated with indirect selection models.  Note that both the $R^2$ and the error rate are expressed as proportions rather than percentages.

<table>
<thead>
<tr>
<th>Model</th>
<th>Ferulic Acid Content</th>
<th>p-Coumaric Acid Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Regression $R^2$</td>
<td>0.0041</td>
<td>0.0000</td>
</tr>
<tr>
<td>Discriminant Analysis Error Rate</td>
<td>0.3115</td>
<td>0.3000</td>
</tr>
</tbody>
</table>
Table 5.4. Correlations between phenolic acid traits and prediction traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Ferulic Acid</th>
<th>p-Coumaric Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHT</td>
<td>0.27097</td>
<td>-0.01091</td>
</tr>
<tr>
<td>EHT</td>
<td>0.21774</td>
<td>-0.03982</td>
</tr>
<tr>
<td>DTS</td>
<td>0.25076</td>
<td>0.02079</td>
</tr>
<tr>
<td>DTP</td>
<td>0.24555</td>
<td>0.02513</td>
</tr>
<tr>
<td>GY</td>
<td>0.03304</td>
<td>0.01143</td>
</tr>
<tr>
<td>TW</td>
<td>-0.02078</td>
<td>0.0038</td>
</tr>
<tr>
<td>KW</td>
<td>0.16226</td>
<td>0.04793</td>
</tr>
<tr>
<td>KL</td>
<td>0.13821</td>
<td>0.00585</td>
</tr>
<tr>
<td>KD</td>
<td>0.15221</td>
<td>0.06017</td>
</tr>
<tr>
<td>HKV</td>
<td>0.37793</td>
<td>0.10652</td>
</tr>
<tr>
<td>HKWt</td>
<td>0.35602</td>
<td>0.10005</td>
</tr>
<tr>
<td>R</td>
<td>-0.1245</td>
<td>-0.04332</td>
</tr>
<tr>
<td>KPR</td>
<td>0.16735</td>
<td>0.03781</td>
</tr>
<tr>
<td>FL</td>
<td>0.23603</td>
<td>0.05075</td>
</tr>
<tr>
<td>EW</td>
<td>0.19873</td>
<td>-0.01797</td>
</tr>
<tr>
<td>ES</td>
<td>0.281</td>
<td>0.02726</td>
</tr>
<tr>
<td>CL</td>
<td>0.11995</td>
<td>0.01928</td>
</tr>
<tr>
<td>CW</td>
<td>0.22915</td>
<td>-0.04244</td>
</tr>
<tr>
<td>CC</td>
<td>0.22915</td>
<td>-0.04244</td>
</tr>
<tr>
<td>EC</td>
<td>0.19873</td>
<td>-0.01797</td>
</tr>
<tr>
<td>Ewt</td>
<td>0.26423</td>
<td>0.05004</td>
</tr>
</tbody>
</table>
Figure 5.1. Validation of NIR predictions of ferulic acid content. The predicted ferulic acid values are on the vertical axis, and the actual values (as measured by wet-lab chemistry) are on the horizontal axis.

Figure 5.2. Validation of NIR predictions of p-coumaric acid content. The predicted p-coumaric acid values are on the vertical axis, and the actual values (as measured by wet-lab chemistry) are on the horizontal axis.
Figure 5.3. Chemical structure of ferulic acid, 4-hydroxy-3-methoxycinnamic acid

Figure 5.4. Chemical structure of p-coumaric acid, 4-hydroxycinnamic acid
Figure 5.5. NIR spectra following vector normalization.
WORKS CITED

Arnason, J., J. Gale, B. Conih de Beyssac, A. Sen, S. Miller, B. Philogene, J. Lambert, R.
Fulcher, A. Serratos, and J. Mihm. 1992. Role of phenolics in resistance of maize grain to
stored grain insects, Prostephanus truncates (horn) and Sitophilus zeamais (motsch). J.

erulic acid content of maize with resistance to Fusarium graminearum. Phytopathology
83:949-953.

Pauls, and J. Arnason. 2003. Dehydrodimers of ferulic acid in maize grain pericarp and

infrared-near-infrared-and-raman-spectroscopy/opus-spectroscopy-software.html>

Butts-Wilmsmeyer, C.J. and M.O. Bohn. 2016. Development of a high-throughput protocol for
the extraction of insoluble-bound ferulic acid in maize. Crop Sci. In Review.

Correlation of phenolic acid content of maize to resistance to Sitophilus zeamais, the


APPENDIX A: GLOSSARY

ad libitum – Latin for “at one’s pleasure.” Used in the context of food consumption, it means to eat at one’s leisure.

Aging-related disease – A chronic disease that is more often seen in older people and believed to be the result of an accumulation of the effects of aging.

Alkali hydrolysis – A hydrolysis reaction that involves the cleavage of chemical bonds by hydroxide ions. Examples of these types of reactions are those involving NaOH as a reagent. In this thesis, the hydroxide ion from NaOH attacks the ester bond which binds insoluble-bound phenolics to arabinoxylans, resulting in the release of the phenolic acid from the arabinoxylan backbone.

Alkaline reaction – A reaction which involves hydroxide ions.

Antioxidant – A molecule that inhibits the oxidation of other molecules by reactive oxygen species (ROS), often by donating an electron to these ROS and thereby preventing damage to other biological molecules.

Arabinoxylan – A fibrous component of the hemicellulose of plant cells which is composed of arabinose and xylose.

Bioavailable – A compound which can be absorbed and be available for use in the body when consumed. Anonym: biologically unavailable.

Carboxyl – A functional group which can act as a weak acid. Characterized by a carbon atom which is both double bonded to an oxygen atom and single bonded to a hydroxyl group. These can be involved in the formation of an ester bond.

Chemopreventive – A chemical which helps prevent the occurrence of particular diseases when consumed in sufficient quantities.

Derivatize – As in to derivatize a sample in preparation for gas chromatography analysis. This is a chemical change in the sample that makes it more amenable to analysis. For instance, the addition of a derivatizing agent in gas chromatography makes the chemicals in that sample more volatile and much easier to analyze using a gas chromatograph.

Diallel Cross – A type of cross used in breeding programs which examines every possible combination of the parental genotypes.

Diferulates – Two ferulate monomers coupled together.

Esterification – The process by which two molecules are joined together to form an ester bond.

Ester-Linked – A molecule is joined to another molecule via an ester bond.
**Fatty Acid Methyl Ester (FAME) Analysis** – A type of analysis which quantifies the amount and type of fatty acids present in a sample.

**Fraction** – In regards to chemistry, a fraction is a certain portion of a sample which is separated from the rest of the sample and then collected for subsequent analysis.

**Fourier Transform Infrared Spectroscopy (FTIR)** – A type of spectral analysis which collects absorbance data within the infrared spectrum and then requires the use of the Fourier Transformation in order to convert the raw data into the actual spectrum.

**Gas Chromatography (GC)** – An analytical chemistry technique which measures the concentrations of different compounds in a sample.

**Gas Chromatography-Mass Spectroscopy (GC-MS)** – An analytical chemistry technique which combines the use of gas chromatography and mass spectroscopy in order to identify the types of compounds present in a sample and then quantify their concentrations in that sample.

**General Combining Ability** – In quantitative genetics, general combining ability is the average performance of a parent in hybrid combinations.


**Heritability** – The proportion of phenotypic variance in a population explained by the genotypic variance in that population. Broad-sense heritability refers to the proportion of phenotypic variance explained by the total genotypic variance, whereas narrow-sense heritability refers to the proportion of phenotypic variance explained by the additive genetic variance.

**Hydroxycinnamic Acids** – A class of aromatic acids with a phenol group attached to a carboxyl group via an aliphatic chain. These are a subset of phenolic acids. They are hydroxy derivatives of cinnamic acid, and are also beneficial phytochemicals which are chemopreventives of several aging-related diseases.

**Hydroxyl** – A functional group that is characterized by an oxygen atom which is single-bonded to a hydrogen atom.

**in vitro** – Latin for “in glass.” These biological studies are conducted outside their normal biological environment.

**in vivo** – Latin for “within the living.” These biological studies are conducted within their normal biological environment or in similar biological environments (e.g. animal testing studies).

**Insoluble-Bound Hydroxycinnamic Acids** – Hydroxycinnamic acids which are covalently bonded to the cell wall constituents, usually via ester bonds.

**Methoxyl** – A functional group which is characterized by an oxygen atom which is single-bonded to a methyl group.
**Midparent Heterosis** – The difference in the F₁ phenotype and the midparent value, divided by the midparent value, and then multiplied by 100%. A measure of hybrid vigor.

**Midparent Value** – The average of the phenotypes of the two parents used in the creation of an F₁ hybrid.

**Monounsaturated Fatty Acid** – A fatty acid with only one double bond in the fatty acid tail.

**Near Infrared Spectroscopy (NIR)** – A spectroscopic method which uses the absorption of light in the near-infrared region in order to quantify the concentration of various compounds in laboratory and biological samples.

**Nuclear Magnetic Resonance Spectroscopy (NMR)** – A spectroscopic method that exploits the magnetic properties of certain atomic nuclei. A non-destructive phenotyping method which can be used to quantify the concentrations of certain phytochemicals in plant materials, including harvested grain.

**Obesity-Related Disease** – A disease that is more likely to be seen in obese people. Furthermore, the incidence of these diseases is believed to be a consequence of obesity.

**Phenolic Acids** – Aromatic acids which are sometimes called phenolcarboxylic acids. These are phytochemical chemopreventives of aging-related diseases. A subclass of phenolics.

**Phenolics** – A class of chemical compounds in which a hydroxyl group is single-bonded to an aromatic hydrocarbon. These are phytochemical chemopreventives of aging-related diseases.

**Phenotype** – The physiological, biochemical, morphological, and other observable traits of an organism which are influenced by its genotype. *adj.* phenotypic.

**Phytochemical** – Chemical compounds that occur naturally in plants.

**Polyunsaturated Fatty Acid** – An unsaturated fatty acid with two or more double bonds in the fatty acid tail.

**Prebiotic** – Food ingredients or components which, when consumed, selectively stimulate the growth and/or activity of beneficial bacteria in the colon. They are associated with improved systemic health and a decrease in the incidence of some aging-related diseases.

**Reagent** – A compound or an element added to a system in order to cause a chemical reaction.

**Ratio of Performance to Deviation (RPD)** – A statistical measure which provides an indication of how predictive a model is.

**Socioeconomic Status** – A combined measure of an individual’s sociological and economic position in relation to others.
Soluble Hydroxycinnamic Acids – Hydroxycinnamic acids which are not bound to the cell wall constituents and which are typically located in the vacuoles of a plant.

Solution – A homogenous mixture composed of two or more substances. Solutions are composed of a solute and a solvent. The solute is a substance which is dissolved in another substance, the solvent.

Specific Combining Ability – In quantitative genetics, specific combining ability is the genetic deviation from the sum of the population mean and the sum of the general combining ability of two parents used in the creation of the F₁. It is an interaction between the parental genotypes.

Tocopherols – Chemical compounds which partially compose Vitamin E and which exhibit antioxidant activity.

Ultra Performance Liquid Chromatography (UPLC) – An analytical chemistry technique which uses the separation of compounds via use of different types of chemical columns in order to identify and quantify the chemical compounds.