ABSTRACT

BUSSLER, WESTON WOODROW. Uncovering Novel Loci in Crop Genomes to Maximize their Health Benefits. (Under the direction of Dr. Slavko Komarnytsky and Dr. Mary Ann Lila)

A healthy diet is one of the controllable factors that influences the development or severity of multiple chronic diseases, including metabolic, inflammatory, and neoplastic disorders. Diets rich in plant based foods have been connected to reducing the risk of deadly chronic diseases. Researching individual metabolites in plants has revealed many compounds that elicit bioactive properties to human beings. Crop breeding and production has been successful at producing high yielding plants able to withstand harsh conditions and last long enough to reach consumers by understanding the genetic architecture within plants that leads to their optimization. However, continued focus on breeding for these traits has inadvertently selected against health properties of plants inherent to wild species, and led to cultivars that are sub-optimal to face the current public health challenges. This dissertation applies the genome association techniques used in plant breeding to locate plant gene regions that significantly control bioactive properties of plants.

Chapter one reviews existing research surrounding genetic association studies that focus on health relevant traits in plants by way of quantitative trait loci (QTL) studies and genome-wide association studies (GWAS). These studies focus on bioaccumulation of a medicinal, nutritional, or nutraceutical compounds, but can also focus on the removal of toxic or anti-nutritional metabolites. Functional traits that lead to health relevance such as glycemic index and antioxidant properties are also outlined.

In the first study—chapter two—the ability of broccoli (Brassica oleracea variety italica) extracts, from a nearly isogenic bi-parental population, to reduce growth of HT-
29 human colorectal cells was associated with molecular markers corresponding to the broccoli genetic linkage map. Three primary QTLs responsible for reducing cancer cell growth were found, and gene searches within the *Brassica oleracea* genome sequence identified nine putative candidate genes. Genes favoring an increased ratio of glucoraphanin over gluconapin or progoitrin were discovered to be a predictor of toxicity to cancer cells.

The second study—chapter three—investigates the interrelationships between broccoli glucosinolates and cancer cell growth inhibition by comparing pure glucosinolates in various mixtures to determine synergistic and/or antagonistic relationships. Results indicated an antagonistic relationship between compounds that is amplified by open competition for hydrolysis by the enzyme myrosinase. Predicted ideal ratios for broccoli glucosinolate profiles were identified and can be used as targets for future plants and products aiming to maximize chemopreventative properties.

In the third study—chapter four—a medicinal property of a plant was mapped onto the genome of a different crop. Oats (*Avena sativa*) from a diverse population were screened in two *in vitro* models for skin health. This population was investigated for the ability to attenuate inflammation by reducing the production of nitric oxide radicals from LPS stimulated macrophages and for its ability to alter skin cell migration in a simulated wound. With a diverse population, more genetic variation was captured than if a biparental source would have been used. A GWAS was then conducted to find loci that contributed to these two phenotypes. A total of 22 loci were uncovered, significantly associated with wound healing.
There is great variation between plant species, subspecies, and cultivars. Understanding how the underlying genetics in plants leads to health properties is essential for tailoring the next generation of crops for public health benefits. This work establishes the efficacy of using biological models to screen large mapping populations of crops and associating them back to plant genomes to find relevant loci able to guide development of new, healthier, plants lines.
Uncovering Novel Loci in Crop Genomes to Maximize their Health Benefits

by
Weston Woodrow Bussler

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Nutrition

Raleigh, North Carolina

2018

APPROVED BY:

Dr. Slavko Komarnytsky
Committee Co-Chair

Dr. Mary Ann Lila
Committee Co-Chair

Dr. Gabriel Keith Harris

Dr. Corey E. Scott
External Member

Dr. Massimo Iorizzo
DEDICATION

To all the unwavering support from family, friends, and colleagues.
BIOGRAPHY

Weston Woodrow Bussler was born March 11th, 1991 during a snowstorm in Minneapolis, Minnesota to Brett and Deanne Bussler. During which, as he is often reminded by his mother, his father was in the corner of the hospital room studying for his Ph.D. It seemed as if Weston was destined to obtain a Ph.D. of his own ever since. Weston attended Webster Groves High School in St. Louis, Missouri where he was a standout wrestler. It became second nature to think about how food effects the human body because of the necessity to be conscious of all the things he was consuming during wrestling season in order to maintain weight and remain at peak performance. This drove him toward pursuing Nutrition at the University of Minnesota. Before leaving high school, Weston met his future wife Anna.

At Minnesota, Weston was fostered into a house with six engineers where they spent most of their time doing nothing but staying at home and studying… He learned how to balance life and excel in his studies by learning from those around him. Weston joined the collegiate handball team and went to four collegiate national tournaments, where his doubles team placed second in 2012. During his summers, he would return to St. Louis to work at the Monsanto Company conducting sensory and quality studies on high omega-3 soybean oil. The glamorous gig saw him working in a lab coat in a makeshift food lab running eight deep fryers constantly every day. From this he learned how he could make a place for himself combining Agriculture, Food Science, and Nutrition Science. Upon returning to Minneapolis for his final years of study he rotated as an undergraduate research assistant for the nutrition department working on a clinical trial and in the agronomy department studying plant bioactive compounds from wild perennial plants.
In May of 2013 Weston interviewed over video during a snowstorm for a graduate student position at NC State University. He was accepted into a Nutrition Ph.D. program where he would be a graduate research assistant working in the Plant Pathways Elucidation Project (P2EP). This position required he move almost immediately to Kannapolis, North Carolina. His first summer was a whirlwind of excitement and frustration. He was tasked with leading five undergraduate students in a number of different research projects. The learning curve was steep but he was able to thrive in the chaotic environment. After this summer he moved to Raleigh, North Carolina, to attend the NC State main campus for a condensed class schedule.

In Raleigh, Weston was able to study both Nutrition and Food Science; he overloaded to finish all his courses for the Major and Minor in three semesters in person. He continued to research and to write during this time, completing two publications. He was also able to act as the Handball coach for NC State and aid as a volunteer instructor for the handball class. Anna was able to start a graduate program at the University of North Carolina at Charlotte just in time for his move back to the Kannapolis area to continue his research.

Now able to research full time under the advisement of Dr. Komarnytsky, Weston’s projects began to take off. Paired with an army of now eight summer interns he and his advisor began utilizing the multidisciplinary program to associate human health relevance to plant genomes, specifically the bioactive properties of oats and broccoli. In total he directly mentored 30 interns; he focused on developing his leadership skills as he began to really see research as a team effort. He encouraged his interns to shoot for the moon on their projects, a strategy that lead to great progress in a short amount of time. As
projects become more successful, he began communicating his results to the public. The P2EP program started showcasing research in an end of the summer research symposium.

Weston’s team won first prize three of four years. In the spring of 2016 he submitted his research to the Experimental Biology conference in San Diego. Here he unveiled his work publically for the first time. He won first prize in the emerging leaders competition as part of the Translational Nutrition research section. In addition, he was chosen to give a seminar in 2016 and again in 2017 when he returned with another ambitious project. In October of 2017 Weston participated in the NC State 3-minute Thesis competition, where he once again worked at communicating his research to the public.

Now as he is finishing up his studies, Weston is excited to begin the next chapter where he puts all that he learned into practice. He hopes to apply the nutrition research and many emerging discoveries/technologies to help make foods that promote human health.
ACKNOWLEDGMENTS

At times, working toward a Ph.D. felt like a lonely endeavor and an impossible undertaking, but those around me proved to be my saving grace. Most especially my amazing wife Anna, you are the absolute perfect partner. Your love and support propelled me through this and I am eternally grateful to have you in my life.

Thank you to all my family—Dad, Mom, Wyatt, Anna, Kirk, Jyll, and Nora—for setting such a great example for me, being willing to lend an ear to my concerns, or reviewing and editing my work. You were all essential to keeping my spirits up and directly helping me maintain my sanity through this extensive undertaking.

To my advisors, Slavko and Mary Ann, thank you for taking the chance on me in this program. It was a once in a lifetime opportunity to learn about my passion for research while getting to give back to others throughout the process. The freedom you gave me to undertake any project I could come up and support to help me pursue it is something I know I will miss going forward. Thank you for everything you taught me.

To everyone in K-lab—Debbie, Mickey, Kim, John, Sierra, Jade, PJ, and Thirumurugan—it was fantastic to work alongside you for the past four and a half years. Thank you for putting up with my insanity and helping me learn how to be a researcher all along the way. I loved our endless musings about how the world works and thoughtful banter.

Thank you to my NC State plant researchers—Allan Brown and Massimo Iorrizo—who helped me get up to speed in an area totally foreign to me. You helped me to break ground into a new field.
A thank you to all the professors and support staff in the NC State FBNS department—especially Dr. Harris, Dr. Allen, and Ms. Cooper—you all welcomed me into the program and helped me to excel as a nutritionist.

To everyone involved in the Plant Pathways Elucidation Project, there have been so many great memories from this program. It was a pleasure to be part of such an amazing undertaking, as a graduate student. I learned lifelong skills that I am excited to take with me to my next stage. To my fellow P2EP graduate mentors—Richard, Scott, Samantha, Kelsey, Aaron, Neha, Bryan, Shawn, Adrian, Rachael, Koyt, and Jaime. It was amazing to share the graduate school experience with all of you. Thank you for pushing me to be better every day.

Thank you to the P2EP support team—Ebony, Tara, Susan, Claire, Aubrey and Megan—we would have not been able to handle any of this without you managing the intern program along the way. To the P2EP industry partners—Corey Scott, Eric Jackson, and Nick Gillet—thank you for allowing us to be part of a bigger picture for our research. It was fun to get all of your perspectives throughout the process. Thank you for all the help and advice you gave along the way.

To the UNCC bioinformatics team—Rob, Steven, Jeremy, and Dr. Brower—I loved how you all let me be a part of the team from day one, provided a much needed education in all things technology, helped out on all sorts of research projects, and contributed essential edits to my writing and presentations.

Finally, thank you to all of my amazing interns. It was quite an experience to be able to work with and mentor all of you. Some of you so graciously stayed on for all the fun and excitement of working with me beyond just the one summer—Jessica Alley,
Holli Chandler, Jocsa Cortez, Jessica Wiltshire, Katelyn DeZego, Nikhita Vemulapalli, Charles Wagner—and others just stayed for the summers—Amber, Tamara, Jordan, Burley, Mason, Brianna, Chika, Jonothan, Claire, Ashley, Enrique, Jessica, Mikayla, Connor, Jeremiah, Kristen, Natalie. Thank you for making my life easier, challenging me to be better, and helping me to grow throughout. You made my experience so unique, I will always be able to look back fondly on my summers mentoring.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... xii
LIST OF FIGURES ......................................................................................................... xiii
ABBREVIATIONS ........................................................................................................... xiv

CHAPTER 1. ASSOCIATING HEALTH PROMOTING TRAITS TO CONSUMPTION CROPS AND MEDICINAL PLANT GENOMES .......... 1
1.1. Diet-related health problems ................................................................. 2
1.2. Plant diversity, quality and current status ........................................... 3
1.3. Developing health-promoting plants ...................................................... 4
1.4. Proteins and amino acids ................................................................. 6
1.5. Minerals ............................................................................................ 7
1.6. Toxic metals .................................................................................... 9
1.7. Carotenoids .................................................................................. 10
1.8. Tocopherols .................................................................................. 12
1.9. Folic Acid ..................................................................................... 13
1.10. Ascorbic acid ............................................................................... 13
1.11. Essential fatty acids ..................................................................... 14
1.12. Phytosterols ................................................................................ 16
1.13. Dietary fibers ............................................................................... 16
1.14. Glycemic Index ............................................................................ 19
1.15. Polyphenols ................................................................................ 20
1.16. Antioxidant activity ..................................................................... 22
1.17. Glucosinolates .......................................................................... 23
1.18. Goitergenic glucosinolates .............................................................. 24
1.19. Pharmaceutical compounds ............................................................. 25
1.20. Phytate ........................................................................................ 27
1.21. Oxalic acid ................................................................................. 28
1.22. Cyanide ........................................................................................ 29
1.23. Moving forward ........................................................................... 29
REFERENCES ............................................................................................................. 31

CHAPTER 2. HEALTH-MODIFYING LOCI AFFECT HUMAN COLORECTAL CANCER CELL GROWTH IDENTIFIED ON GENETIC MAP OF BROCCOLI (BRASSICA OLERACEA L. VAR. ITALICA) ....... 42
2.1. Abstract ......................................................................................... 43
2.2. Introduction .................................................................................. 44
2.3. Methods ........................................................................................ 48
    2.3.1. Development of breeding population ........................................ 48
    2.3.2. Development of genetic resource .......................................... 48
    2.3.3. Plant Material Preparation .................................................... 48
    2.3.4. Extract prep for cell culture study .......................................... 49
    2.3.5. Digestion of extract for anticancer assay ................................ 49
    2.3.6. Cell culture parameters ......................................................... 49
    2.3.7. SRB Viability Assessment ....................................................... 50
LIST OF TABLES

Table 1.1. Leading causes of death in the United States..................................................2
Table 1.2. Consumption statistics for food groups ..............................................................3
Table 2.1. Multivariate regression individual metabolite contributions .............................63
Table 2.2. HT-29 cell growth inhibition QTLs and putative candidate genes...............65
Table 3.1. Predicted IC50’s and calculated combination indexes of mixtures of four major aliphatic glucosinolates for inhibition of HT-29 human colorectal cancer cell.................................................................94
Table 4.1. Variation in phenotypic data for population gathered from the public T3/oat database for all agronomic and metabolite traits..............................................124
Table 4.2. Correlation matrix between phenotypes .........................................................127
Table 4.3. Significant loci for inflammatory and migration effects..................................129
# LIST OF FIGURES

| Figure 1.1. | Diet related risk factors for cardiovascular disease | 22 |
| Figure 2.1. | Effect of broccoli extracts on cell growth inhibition | 58 |
| Figure 2.2. | Whole genome multiple QTL mapping | 60 |
| Figure 2.3. | Broccoli genetic linkage map | 61 |
| Figure 2.4. | Forward stepwise and reverse multivariate metabolite regression vs cancer cell growth inhibition | 62 |
| Figure 2.5. | Biosynthesis pathway for aliphatic glucosinolates and phenolic compounds in broccoli | 66 |
| Figure 3.1. | Summary graphic for broccoli glucosinolate mixture analysis study | 82 |
| Figure 3.2. | Biosynthesis pathway for three classes of glucosinolates | 86 |
| Figure 3.3. | Equations for calculating combination index and three-point mixture analysis design | 90 |
| Figure 3.4. | Chemopreventative effect of aliphatic glucosinolates from broccoli | 92 |
| Figure 3.5. | Ternary plots depicting antiproliferative activity for three glucosinolate mixtures | 98 |
| Figure 3.6. | Antiproliferative property and glucosinolate profile of near optimal and average broccoli lines | 99 |
| Figure 4.1. | Oat sample location map | 117 |
| Figure 4.2. | RAW cell and HDFa cell viability and population inflammatory and migration bioactivity spread | 123 |
| Figure 4.3. | Nitric oxide radical production and migration activity induced by significant oat cultivars | 125 |
| Figure 4.4. | Time lapse of fluorescent dyed human dermal fibroblast migration into exclusion zone | 130 |
| Figure 4.5. | Inflammatory and migration QTLs on oat genetic linkage map | 131 |
ABBREVIATIONS

BNC = Broccoli parent line 2
CCD = Healthy human colon cells
CFR = Code of federal regulations
CLC = Lacombe, Alberta, Canada
DMSO = Dimethylsulfoxide
Glycemic Index = GI
GN = Gluconapin
GR = Glucoraphanin
GWAS = Genome Wide Association Study
HDFa = Human dermal fibroblast
HT-29 = Human colon cancer cell line
IC50 = Inhibitory concentration to achieve 50% reduction
IDAC = Aberdeen, Idaho, USA
INF = Inflammation
LOD = Logarithm of the odds
LPS = Lipopolysaccharide
MIG = Migration
MYR = Myrosinase
PG = Progoitrin
QTL = Quantitative trait loci
RAW = RAW 264.7 Maurine macrophage cells
RIL = Recombinant inbred line
SEP = Separate myrosinase hydrolysis
SNG = Sinigrin

SNP = Single nucleotide polymorphism

SRB = Sulforhodamine B

VAT = group myrosinase hydrolysis

VI-158 = Broccoli parent line 1
CHAPTER 1. ASSOCIATING HEALTH PROMOTING TRAITS TO CONSUMPTION CROPS AND MEDICINAL PLANT GENOMES
1.1. Diet-related health problems.

Although human beings have surmounted many challenges throughout history, they still struggle to manage worldwide malnutrition while stunting the development and progression of non-communicable diseases. For areas with prevalent nutritional deficiencies, the solution is being able to provide individuals with a complete diet with adequate levels of essential macro and micronutrients. A healthy diet is also one controllable factor influencing the development of multiple metabolic, inflammatory, and neoplastic disorders. Rising rates of obesity in the young and old adds another major risk factor for diseases connected to diet. Six of the ten leading causes of death in the United States list “diet” as a major risk factor for the development and progression of these diseases (Table 1.1). These diseases are on the rise in other developed nations, and even faster in the developing world as previous challenges with infectious disease and famine are becoming less prevalent.

<table>
<thead>
<tr>
<th>Causes of Death</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Disease*</td>
<td>633,842</td>
</tr>
<tr>
<td>Cancer*</td>
<td>595,930</td>
</tr>
<tr>
<td>Chronic Lower Respiratory Diseases</td>
<td>155,041</td>
</tr>
<tr>
<td>Accidents</td>
<td>146,571</td>
</tr>
<tr>
<td>Stroke*</td>
<td>140,323</td>
</tr>
<tr>
<td>Alzheimer’s Disease*</td>
<td>110,561</td>
</tr>
<tr>
<td>Diabetes*</td>
<td>79,535</td>
</tr>
<tr>
<td>Influenza &amp; Pneumonia</td>
<td>57,062</td>
</tr>
<tr>
<td>Kidney Diseases*</td>
<td>49,959</td>
</tr>
<tr>
<td>Intentional Self-Harm</td>
<td>44,193</td>
</tr>
</tbody>
</table>

* Denotes diet as a major contributing risk factor. Modified from National Center for Health Statistics.

The 2015-2020 dietary guidelines for Americans laid out updated goals and approaches for combatting and preventing diseases and increasing lifespan by recommending the consumption of nutrient-dense foods, specifically an increased amount of whole fruits and vegetables. The western diet is lower in consumption of these whole...
fruits and vegetables (Table 1.2), but they still represent a large quantity of the American diet. This begs the question: How can we increase the health benefits of the plants we already consume on a regular basis?

<table>
<thead>
<tr>
<th>Food Source</th>
<th>Avg Consumption from 2007-2008 (lbs/year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetables</td>
<td>161.96</td>
</tr>
<tr>
<td>Fruits</td>
<td>119.37</td>
</tr>
<tr>
<td>Grains</td>
<td>135.30</td>
</tr>
<tr>
<td>Nuts</td>
<td>8.54</td>
</tr>
<tr>
<td>Fats &amp; Oils</td>
<td>65.03</td>
</tr>
<tr>
<td>Sugars</td>
<td>79.69</td>
</tr>
<tr>
<td>Eggs</td>
<td>18.97</td>
</tr>
<tr>
<td>Meat</td>
<td>145.73</td>
</tr>
<tr>
<td>Dairy</td>
<td>211.35</td>
</tr>
</tbody>
</table>

Modified from US Food Commodity Availability by Food Source, 1994-2008

1.2. Plant diversity, quality and current status

Plants hold an incredible genomic and metabolic diversity, stemming from the existence of different species, subspecies, and cultivars. The variation of metabolites in each plant is controlled by a combination of the genetics of the plant, the environment in which the plant is grown, and the processing that occurs following harvest. This diversity has been utilized to identify whole plants, plant extracts, and single ingredients that can treat specific ailments, but this diversity becomes problematic when connecting dietary patterns to health outcomes as it is nearly impossible to recommend the specific amount of a plant that should be consumed to achieve a desired effect.

Additionally, the agricultural community has historically focused breeding objectives for commercial crops primarily on increasing yield, pest resistance, environmental stress resistance, and herbicide tolerance rather than increasing the nutritional properties of crops, a likely result of preferentially selected traits being negatively correlated with beneficial health effects that ultimately get left behind. These levels are not trivial either; the documented evidence has shown a significant depletion of nutrient density in
current crop varieties, when compared to their historical predecessors. The cause of this depletion has been attributed to the continued selection for yield-promoting traits with less regard for nutritionally relevant traits in commercial varieties.

To reverse this, more breeding programs have moved toward locating and understanding the genetic architecture of nutritional relevance, so health-promoting vitamins, minerals, and beneficial phytochemicals can be maximized in future cultivars, providing people with plants that confer greater health. Precise breeding measures need to be utilized so the gains in value-added agronomic traits (yield, biotic, and abiotic stress resistance) are, alongside nutritional improvements, retained.

1.3. Developing health-promoting plants

Plant health improvements can be made through bio-fortification and functional food development. *Biofortification* occurs when the nutrient content of a crop has been increased, and *functional foods* are foods that deliver a health benefit beyond nutrition content. Targeting the nutritional improvement of food crops to address nutritional insecurity and confer positive health effects on other diseases will add value to foods for consumers and producers. Goals for these processes can be achieved by:

1) Changing plant metabolite profiles to enhance beneficial traits—like essential nutrients, beneficial phytochemicals, or active medical compounds.

2) Limiting the accumulation of negative traits—like allergens, toxic levels of compounds, or specific anti-nutritional components.

The influx of new technologies in genome sequencing, marker-assisted breeding, and predictive breeding have rapidly increased the accessibility and capability of creating new functional plant lines. Selection and production of more desirable crop lines relies
on the exploitation of natural genetic variation in the world’s gene banks to produce breeding populations that can be used for genome association studies. Genes affecting quantitative traits can be detected indirectly through associations between trait scores and molecular markers and then matched with the positions on chromosomes or marker linkage groups to identify quantitative trait loci’s (QTL). These QTL can be used to find underlying genes in completed genome sequences or directly incorporated into future lines through marker-assisted breeding. To successfully implement nutritionally enhanced breeding programs, a few aspects need to be met:

1) The plant has significant variation in a heritable health-relevant phenotype.
2) The phenotype is reliably measurable across many different samples.
3) Adequate genetic resources exist in the species of interest.

Health-relevant phenotypes can be the accumulation of a specific beneficial metabolite, limiting the accumulation of a problematic component or using an indirect trait associated with a health risk. Beneficial traits in plants, relevant for humans start with the 41 essential nutrients to survive, but also include conditionally essential nutrients and those associated with reducing chronic disease risk. Indirect phenotypes can be a taste of a food, a functional property, like starch pasting, or a collective chemical effect, such as the total antioxidant activity of a plant extract. In this review, the goal is to summarize the many different ways that quantitative trait loci and genome association studies have been used to link phenotypes related to human health to areas of plant genomes.
1.4. Proteins and amino acids

Protein malnutrition is especially prevalent in the developing world, where plants are the primary source of protein in the diet. Plants are normally low in one or more essential amino acids and thus must be paired with other protein sources in order to assure adequate status is reached. Maize, for example, is normally deficient in the amino acid lysine, and as a staple crop with great functionality there was interest in using breeding techniques to create lines with a more complete protein profile. A mutant high-lysine corn variety was crossed with another high-lysine mutant and a wild type to develop a total of 325 recombinant inbred lines (RILs), which were all assessed for lysine concentration in seeds. Both populations identified the same marker loci on chromosome 8 significantly controlled the lysine content. This region was found to explain 9.1-12.45% in the dual mutant cross and 6.46-6.71% in the mutant crossed with wild type. Another marker on chromosome 7 controlled approximately 40% of the phenotypic variation only in the dual mutant cross population. An underlying gene for lysine content was further investigated from these regions to further outline the genetic control of accumulation of this essential amino acid.

Rice protein has a more complete amino acid profile compared with corn, but it has a low total concentration (8.5%) in relation to other staple grains, like wheat (12.3%) or barley (12.8%)—making it an adequate crop for interventions in areas where people are at high risk of protein malnutrition. However, the amount of total protein in rice is much lower than what would be ideal for these regions. To initiate the development of high-protein rice lines, researchers used an RIL population to investigate total crude protein and accumulation of four different protein factions (albumin, globulin, prolamín,
and glutelin). Within the analysis of the protein fractions two QTLs explaining 9.1% and 19.4% of albumin content, four QTLs explained between 10.6% and 18.1% of globulin content, three QTLs explained between 11.5% and 23.3% of the prolamin content, and four QTLs explained 9.7-14.8% of glutelin content, respectively. Three QTLs explained 11.2-14.4% of the total crude protein content. Three QTL showed up as significant in more than one trait—one for crude protein and glutelin on chromosome 12, another for globulin and glutelin, and the last for glutelin and prolam on chromosome 10. This work presents an early step in the process for molecular development of a rice with higher nutrition quality. Another protein-related phenotype worthwhile to explore for increasing protein quality of plants would be to map the digestibility differences for cultivars.

1.5. Minerals

Inorganic elements present in all forms of life, minerals serve many essential functions for humans, with at least 28 having been deemed essential. There are two groups of essential minerals for human beings: major minerals, which are required in large amounts, and trace minerals, which are necessary but of which decreased quantities are required. The major minerals are calcium (Ca), magnesium (Mg), potassium (K), sodium, (Na), chloride (Cl), phosphorus (P), and sulfur (S); the trace minerals are iodine (I), zinc (Zn), selenium (Se), iron (Fe), manganese (Mn), copper (Cu), cobalt (Co), molybdenum (Mo), fluoride (F), chromium (Cr), and boron (B). Functions minerals play in the body include: Ca as a component of bone and teeth, Zn as a component of many enzymes involved in a wide array of functions, Fe as a component of hemoglobin allowing for transport of oxygen throughout the blood, amongst many more.
Inadequate mineral status is related to many health problems, including but not limited to anemia, osteoporosis, goitergenesis, and development issues, in addition to having a connection to the development of chronic diseases. Minerals are derived from dietary sources, which accumulate minerals from the soil, but species and cultivars show differing ability for uptake and accumulation in the edible part of the plant. There are many areas in the world where mineral status is inadequate. Genomic association studies focusing on plant mineral content have been conducted on, soybeans, tomatoes, cabbage, and potato tubers.

Targeted programs have used a combination of QTL studies followed with marker-assisted selection to increase one specific mineral to fortify grains for growth in areas at risk for mineral deficiency. An estimated 17.3% of the world's population is at risk for Zn deficiency, with South, Southeast Asia, and Sub-Saharan Africa showing levels at or above 25% of the population at risk. People in these regions obtain the majority of their diet from cereal grains that are poor sources of Zn. Increasing Zn concentration in rice cultivars has been a goal of breeding programs to help alleviate deficiency related growth impairments for those consuming diets with inadequate Zn content. In the effort to produce high Zn rice, molecular and genetic associations to Zn accumulation in the edible portion of the plant have been carried out. Mineral status is tricky to determine through breeding alone because it is so dependent on the soil concentrations. Zn mobilization in the plant is also difficult to control as it may be possible to increase overall plant content, but to be relevant for nutrition this mineral must of course accumulate in the edible portion of the plant to have a true benefit. QTL studies have been run in multiple RIL populations. A total of 26 QTL from these...
studies on Zn bioaccumulation in rice grain have been found. Of these QTL, 3 regions consistently associated with Zn concentration on chromosome 7, 11, and 12—explaining between 5.3-35.0% of the Zn content variation. Fine mapping of these regions still must be carried out to find the best estimates for applying a marker-assisted breeding program to produce an enhanced level of Zn.

1.6. Toxic metals

Breeding interventions applied in staple crops to address mineral deficiencies can also be made to decrease the uptake of minerals in areas where this is of concern. Certain minerals or too much of a necessary one can become toxic to human beings, so it is important to be careful when increasing a plant's concentration. Minerals that become toxic when consumed excessively include Cu, Mn, Mo, Se, and Zn. Non-essential minerals that can cause negative effects in human beings are arsenic (As), cadmium (Cd), and lead (Pb). When researchers are breeding for high mineral-accumulating plants in regions with higher than normal levels of these minerals, they need to be especially aware of the potential to cause harm. Studies for selective uptake of heavy metals have been carried out in cereal grains to identify the genetic architecture to the bioaccumulation of toxic components in edible portions. In rye, 25 QTL corresponded to mineral accumulation. Of them, 1 QTL was found for As (explaining 10.0% of the phenotypic variation), 1 for Cd (explaining 12.1% of the phenotypic variation), 1 for Cu (explaining 12.2% of the phenotypic variation), 2 for Fe (explaining 8.9% and 13.8% of the phenotypic variation), 2 for Mn (explaining 10.8% and 12.0% of the phenotypic variation), 1 for Mo (explaining 18.7% of the phenotypic variation), 1 for Pb (explaining 10.8% of the phenotypic variation), and 3 for Zn (explaining 10.8-12.3% of the
phenotypic variation each). Five of these QTL regions saw mineral QTL co-localize with both toxic and essential minerals 47.

1.7. Carotenoids

Vitamin A (retinol & retinoic acid) is necessary for many biological actions in human beings, including the production of rhodopsin in eyes, maintenance of epithelial and immune cells, and other roles in development. Vitamin A can be derived from provitamin sources found in plants, with the most notable form being β-carotene 48. Deficiency of vitamin A can lead to a weakened immune system, anemia deterioration of low-light vision, and an irreversible form of blindness called xerophthalmia 49. Worldwide, it is estimated that nearly 200 million school-aged children and 20 million pregnant women have serum retinol levels below 0.70 µmol/L, marking deficiency 50. Biofortification of plants with higher levels of carotenoids would not likely lead to toxicity (in the way that direct retinol consumption could), making increasing plant concentrations of carotenoid content of major interest for breeding programs worldwide, especially in impoverished nations where vitamin A–poor cereal grains make up the primary bulk of the diet.

One study with an RIL population of 233 maize lines selected for differences in provitamin-A content associated the concentration of carotenoids to a genetic linkage map. The study outlined a total of 31 loci significant for some type of carotenoid bioaccumulation, but the majority was observed in 2 regions that explained 6.6-27.2% of the phenotypic variation. One of these most significant regions was on chromosome six, corresponding to functional marker psy1—a known gene in the carotenoid biosynthesis pathway 51. Several breeding programs have used results from QTL and gene association
studies to increase the carotenoid content of maize to fight the common worldwide nutritional deficiency for vitamin A. Resulting high-carotenoid maize products have been assessed for their ability to promote eye development in at-risk populations from Zambia in a randomized placebo-controlled clinical trial, which found orange maize cultivars significantly increased biological reserves of retinol comparably to supplementation with vitamin A oil.

Other types of carotenoids have been identified as bioactive molecules associated with reduced risk of chronic diseases. Lycopene is a naturally present carotenoid, lycopene-containing foods have been associated with reduced risk of prostate cancer and cardiovascular disease. Lycopene’s beneficial health effects have been proposed to be derived from radical scavenging ability, and the main source of lycopene for humans comes via tomato consumption. In a QTL study focusing on lycopene accumulation, two significant regions showed repeated significant control of lycopene accumulation—one on chromosome 7 and another on chromosome 12 controlling 9.0-23% and 6.0-15% of the phenotypic variation, respectively.

Carotenoids are directly tied to the color development of a crop. Researchers have used color as an indirect measure of carotenoid content or changing the profile from one carotenoid to another. Sorghum is the staple cereal grain prevalent in Africa. Based on the accumulation of carotenoids and other pigments, its edible portion ranges from white, yellow, red, and purple. In a 352 RIL sorghum population developed from white and yellow parents, QTLs for different carotenoids (lutein, zeaxanthin, and β-carotene) and endosperm color were identified. A total of five QTLs significantly associated to β-carotene—with only one in both growing environments. This QTL explained 11.6-
15.15% of the carotenoid variation and contained a newly identified phytoene synthase gene. Genes for color co-localized with QTLs for β-carotene, zeaxanthan, and lutein; however, not all traits were shared so color should not alone be used as an indicator when breeding plants to be better at preventing night blindness 58.

1.8. Tocopherols

Tocopherols are fat-soluble essential vitamins for humans that function as cellular antioxidants, allow for platelet adherence, and play key roles in the body’s immune system 59. Supplementation with vitamin E above the recommended level has also been investigated for potential protective effects against cardiovascular disease development, but this has not been found to have any substantial effect 60. In plant oils, tocopherols are important for preventing auto-oxidation of fatty acids 61. Vitamin E deficiency is rare in human beings but is possible under specific diseases that impair fat metabolism 59. In studies looking to breed higher tocopherol-containing plant seeds, the goal may be more related to increasing the shelf stability of the resulting food products. One such study looking at sunflower oil used a 123 RIL population from parents with significantly different tocopherol concentrations. The analysis outlined 5 QTL that pertained to total tocopherol concentration of seed oil, controlling between 4.0% and 15.0% of the phenotypic variation for tocopherol content. A gene analysis for known tocopherol biosynthesis found a few genes corresponding with this region able to be used as functional markers for future fine mapping and potentially for marker-assisted breeding going forward 62.
1.9. **Folic Acid**

Folate, an important water-soluble nutrient, is especially important for women of childbearing years because of the connection between folate deficiency and neural tube defects in infants. Many foods have folate added during industrial processing to ensure that adequate content is achieved, but biofortification presents another approach to widespread public health efforts to maintain adequate folate consumption. Adequate folate status has also been associated with the reduction of mortality from chronic diseases. Studies have been undertaken to map folate accumulation onto genetic linkage maps of dry beans and in rice. In dry beans, four loci associated with folate accumulation and these loci explained 8-19% of the total phenotypic variation in a single population. Another study mapped folate accumulation in rice. This resulted in three significant QTLs close together that explained 7.8%, 11.1-15.8%, and 25.3% of phenotypic variation from two separate populations. A unique aspect of the rice study was that it was running simultaneously with an experiment that increased the expression of known folate biosynthesis genes where researchers did not observe a significant increase in rice folate concentration following increased expression of the biosynthesis genes. This outlines the unique nature of QTL studies, where genes might be contributing to a phenotype in one way that is beyond the genes located in biosynthesis pathways.

1.10. **Ascorbic acid**

Ascorbic acid (vitamin C) plays essential roles in biosynthesis of collagen, carnitine, and neurotransmitters. It functions as a metal chelator and making non-heme iron more bioavailable and in fat and cholesterol metabolism. Deficiency of vitamin C
causes scurvy, a condition particularly common in days of long sea voyages, ameliorated by consuming fresh fruits and vegetables or a concentrated citrus fruit tonic \(^6^9\). In the present day, vitamin C deficiency has become rare; however, increased consumption of vitamin C has been linked to beneficial effects on chronic diseases—though information is conflicting \(^7^0\). Much like tocopherols, the antioxidant activity of vitamin C has been proposed as the source of its health association, and plant breeders have investigated increasing its concentration in plants through genomic-association studies. In tomatoes, vitamin C QTL were mapped in 3 biparental populations uncovering 15 unique loci significant for ascorbic acid. Of the 15 vitamin C QTL, 6 showed up in more than one of the populations analyzed, increasing the confidence in their ability to control vitamin C bioaccumulation. One of these loci, found on chromosome 9, corresponded with the known ascorbic acid biosynthesis gene monodehydroascorbate reductase \(^7^1\).

### 1.11. Essential fatty acids

Human beings do not contain delta 15 desaturase necessary to synthesize a double bond at the omega-6 and omega-3 position of fatty acids, making these polyunsaturated fatty acids a necessary component of the human diet. Increased consumption of omega-3’s has been found to have many positive metabolic effects, including: decreased inflammation, improved immune functions, bone mass maintenance, improved insulin sensitivity, proper brain development, optimized visual signaling, and amelioration of various risk factors for heart disease \(^7^2\). Essential omega-3 fatty acids (DHA and EPA) can be obtained directly from fatty fish or elongated from a precursor obtained from plant sources \(^7^3\). However, the plant sources with significant proportions of omega-3 over omega-6 fatty acids are likely to be poor sources of fat \(^7^2\). Furthering this issue,
polyunsaturated fatty acids are known to cause autoxidation of lipids, causing selective breeding programs to focus on eliminating linolenic fatty acids because of the negative effects on shelf life, oxidation, and sensory aspects of oils. Recent efforts have been made to combat this by producing more crop lines with dramatically changed fatty acid profiles of high fat seeds through transgenic approaches, and the same programs that identified ways to remove linolenic fatty acids can be repurposed to increase content of omega-3 fatty acids.

One example of genomic associations to the fatty acid profile makeup was run in soybeans. This study found six QTLs associated with linolenic acid content, four QTLs associated to linoleic acid, four associated to oleic acid, and four for palmitic acid. Each QTL controlled from 2.53-37.30% of the phenotypic variation. Three significant loci for linolenic acid content were replicated over four environments, increasing the confidence in their ability to be used for marker-assisted selection for new soybean varieties. Olive trees, another important oilseed crop, are also low in the omega-3 fatty acids in comparison with their concentration of omega-6 fatty acids. One study identified QTLs for different fatty acids present in olives. Seven unique QTLs were found to control accumulation of the different fatty acids present in olives. The QTL for oleic and linoleic fatty acid was shared—explaining 41.1% and 69.7% of the phenotypic variation respectively—and showcased an inverse effect on concentration for each. Four QTL explained between 15.0-28.0% of linoleic fatty acid, and one QTL was identified to control palmitoleic fatty acid, explaining 22.5% of the variation.
1.12. Phytosterols

Naturally accumulating plant phytosterols are non-essential molecules that, when consumed, have shown the ability to decrease the absorption of dietary cholesterol\(^\text{78}\). Phytosterols can be found in many foods, but are numerous in oilseeds specifically. While increased phytosterol consumption suggests positive effects on some cardiovascular disease risk factors, the consumption of naturally accumulating dietary sources of phytosterols has not been able to show consistent positive effects on CVD mortality\(^\text{79}\). One study of rapeseed identified QTL for total phytosterol content, finding three loci that controlled the bulk of the total variation in phytosterol levels. Further investigation of these regions in the closely related \textit{Brassica rapa} completed genome was able to identify multiple candidate genes relevant to phytosterol biosynthesis\(^\text{80}\). Further studies can focus on the regions or directly target the putative candidate genes that were found.

1.13. Dietary fibers

Dietary fibers are compounds found in the consumed portion of a plant or food resistant to digestion and absorption in the small intestine. Different forms of dietary fiber include: arabinoxylan, inulin, lignin, pectin, bran, cellulose, \(\beta\)-glucan, fructan, and resistant starches\(^\text{81}\). Meta-analysis of the consumption of dietary fiber has shown consumption is inversely associated with incidence of cardiovascular disease\(^\text{82}\), colon cancer\(^\text{83}\), and diabetes\(^\text{84}\). Different types of fiber can elicit differing effects on health, but they all seem to work in a similar way; thus, a health claim for total consumption of dietary fiber and reduction in cancer and another for reduction in heart disease has been issued by the FDA. Mechanistic data for the relationship between fiber and diseases are
still unclear but hypotheses range from slowing transit in GI, functioning as a prebiotic, being a source for bioactives released upon microbial digestion, to helping with the removal of metabolic byproducts \(^{81}\).

Beta-glucan (1-3, 1-4)-\(\beta\)-D-glucan (\(\beta\)-glucan) is a hemicellulose native to oats and barley. When ingested, \(\beta\)-glucan acts as a soluble fiber and has been linked to many health benefits. Dietary \(\beta\)-glucan is associated with lowering cholesterol, blood pressure, and blood glucose, modifying the immune response, and reducing the risk for colorectal cancer \(^{85}\). In the United States a health claim is issued for foods with at least 0.75g of \(\beta\)-glucan of reducing the risk of heart disease (21 CFR 101.81). This has made increasing \(\beta\)-glucan concentration a goal for plant breeding. A diversity panel of 431 oat landraces and breeding lines was used to run a genome-wide association study to find molecular markers associated with increased \(\beta\)-glucan content. The range of beta glucan in the collection was 1.44-6.20% of seed weight. The association study identified 3 marker loci that associated to \(\beta\)-glucan content. The markers were used to explore potential candidate genes by way of sequence homology to the rice genome. Homologous regions were used to find \(\beta\)-glucan biosynthesis pathway genes. This study located multiple \(CslF\) genes that are known to play a role controlling \(\beta\)-glucan content accumulation as genes of interest \(^{86}\). Recently, another \(\beta\)-glucan genome association study was carried out in barley. In the diverse population of 3,069 varieties, two loci were strongly associated to beta glucan content. Investigation into the sequence data corresponding to this region identified glucan endo-1,3-\(\beta\)-glucosidase gene to as a candidate for controlling this effect \(^{87}\). Breeding based on this allele could lead to a functional barley line in the future.
The main non-cellulose polysaccharide in wheat is arabinoxylans, a dietary fiber that has been associated with lowering cholesterol, enhancing mineral absorption, promoting fecal bulking, and aiding immunomodulatory activity. Using a RIL population of 240 lines, 15 QTLs were mapped that explained 3.5-14.6% phenotypic variation for this specific fiber type. Wheat is a particularly interesting plant because of its hexaploid genomic, which contains 3 different ancestral genomes, and showcases many repetitive genes. This complication makes identifying underlying genes responsible for traits difficult to study and breed for in wheat. Progress in advancing sequencing of this species will help to mitigate the challenges brought on by working with a polyploidy genome and open the door to precisely increasing dietary fiber in wheat based on these identified regions.

Two other common dietary fibers that have been associated to plant genomes are the fructans and the resistant starches. Onions are a good source of the fructans; their primary health associations are acting as a prebiotic, fueling the growth of beneficial microbes in the human gastrointestinal tract, lowering blood lipid levels, and managing insulin levels. In a biparental onion population, fructan content was associated to genetic markers mapped to the chromosome 8 region. By means of these regions, fructan concentration increase can be targeted to reap the health benefits from the dietary fiber for the general population. But for people with inflammatory bowel disease who follow a low FODMAP diet, decreasing fructan concentration would be desirable. They would want to consume onions that do not have fructans which would exacerbate their symptoms. A resistant starch QTL study was undertaken to find genetic control of accumulation in rice using a RIL population. The study found two QTL on chromosome...
7, explaining 7.6-17.3% of the phenotypic variation for resistant starch accumulation in the grain \(^{92}\). Breeding for resistant starch content may become an area of emphasis now that the FDA has approved a specific health claim for its association for reduced risk of type 2 diabetes \(^{93}\).

**1.14. Glycemic Index**

A great example of an indirect approach to associate health-related phenotypes to plant genomes can be seen in studies that look at the glycemic index (GI) of cereal grains. High GI diets have shown correlation to increased risk of type 2 diabetes and heart disease \(^{94}\). Increasing levels of amylose in grains have been correlated to lower blood glucose levels, \(^{95}\) and there is a large amount of variation in amylose levels across different cultivars \(^{96}\). Because of this, molecular mapping efforts to produce grains with a lower GI have started by finding QTLs associated with increased amylose content. One major QTL on chromosome 6 was found to explain 67.3% of the phenotypic variation for amylose content. The major gene identified in this region attributed with the activity was identified as the \(W_x\) gene. By incorporating this trait into new lines, researchers have successfully increased amylose content rice lines exhibiting a lower GI \(^{97}\). In a study looking at barley, the population’s \(\beta\)-glucan concentration was mapped. High \(\beta\)-glucan barley was also found to have a lower GI response, but this trait was not mapped \(^{98}\). One study used a more direct method of associating the health relevance of GI by screening 235 rice varieties for percent available carbohydrates following *in vitro* simulated human digestion. This method resulted in a strong correlation between amylose content and GI responses. The study associated the activity from each cultivar to the 4 different \(W_x\) genotypes identifying increased GI in cultivars with a genes that decreased overall
amylose content \(^96\). Conducting a genome-wide association study to discover more markers that may be contributing to glycemic response or modifying the phenotype to more closely model true human GI response to screen all lines in the population could expand upon this work further.

1.15. Polyphenols

Plant polyphenols have many roles in plants but have also been found to elicit effects on human beings. These can be organoleptic effects like odor, flavor, bitterness, and astringency; or they can have beneficial health associations such as protection from the development of cancer, diabetes, heart disease, osteoporosis, and neurodegenerative diseases \(^99\). The compounds are classified into four groups on the basis of their chemical backbones: lignans, stilbenes, flavonoids, and phenolic acids. While the direct mechanisms of this benefit are still unclear, one proposed method has to do with the ability of these compounds to accept free radicals and limit impending oxidative damage. These compounds spare the need to deplete endogenous antioxidants and/or helping to absorb pro-oxidant components \(^100\). **Figure 1** outlines the effect that polyphenols can have to limit risk factors for chronic disease, with a focus on cardio-protective effects.

Efforts to increase the total phenolic content in crops have recently moved to utilizing a genome association approach to use total phenolic accumulation as a phenotype to associate to genetic linkage maps. Studies have examined total phenolic compound QTLs in barley, strawberry, peanut, tomato, and broccoli \(^87,102–105\). More targeted programs have focused on increasing one specific polyphenol with a well-understood health connection. Isoflavones are a type of flavonoid found in soybeans that have been associated with a variety of beneficial health effects, on diabetes, heart disease,
osteoporosis, and various cancers (breast and prostate). In 1999, based on evidence for positive effects on heart disease, the United States issued a health claim for foods with adequate soy protein \(^{106}\). Individual isoflavones—daidzein, genistein, and glycitein—and total isoflavone content in soybeans were associated to a genetic linkage map using a RIL population built to have different proportions of total oil and total protein content. This QTL study also looked at whether any of these associated regions also were affecting protein or oil concentration, so that bioactive flavones could be increased without creating a negative drag on other important traits.

The results from this study were that 21 total QTLs affected isoflavone concentration before factoring in protein and oil. Specifically five for DZ (23.1% explained), 3 for GC (15.49% explained), six for GT (31.43% explained), and seven for total isoflavone content (41.49% explained). When oil and protein were analyzed in tandem with the isoflavone levels, two of these QTLs were independent from any effect.

**Figure 1.1** Diet related risk factors for cardiovascular disease.
on protein and only one QTL did not impact oil content\textsuperscript{107}. The results from this study outline the potential consequences of marker-assisted selection on other traits due to localized associations with other important traits. Other examples of specific flavonoids that have been mapped are cyanidin in raspberries\textsuperscript{108} and total anthocyanins in grape\textsuperscript{109}.

1.16. Antioxidant activity

Because there are many metabolites from the plant matrix able to contribute antioxidant activity, the collective action of these molecules are thought to synergistically provide the health benefits associated with consuming plants high in antioxidant vitamins, flavonoids, and other phytochemicals\textsuperscript{110}. The collective antioxidant capacity of a fruit extract can be measured using different chemical assays—ABTS, DPPH, FRAP, and ORAC\textsuperscript{111}. Diverse populations of crops have been assessed for total antioxidant activity by chemical assays and correlated to metabolite concentrations\textsuperscript{102,112–114} finding associations between phenolic compounds and specific metabolites and antioxidant capacity. This work can be then paired with QTL studies for these molecules. However, other studies have looked at total antioxidant capacity as the phenotype to map. One example analyzed a 120 RIL of peanuts for flavonoids, total phenolic compounds, and antioxidant capacity (measured by DPPH). All three of these phenotypes were then used to run a QTL study. The results found 10 QTLs for the three phenotypes. One QTL between the three traits overlapped, being found to significantly explain both total flavonoid and total phenolic compound concentration—explaining 25.2\% and 12.5\% of the phenotypic variation respectively. Four QTLs were identified for antioxidant capacity, explaining 11.5-33.0\% of the phenotypic variation. No QTL for antioxidant capacity corresponded to total flavonoid or phenolic compound QTLs suggesting that
either specific individual polyphenols or different compounds are controlling antioxidant activity in plants\textsuperscript{104}. In a genome-wide association study conducted on 416 black and red rice accessions, 41 marker loci significantly associated to rice color, phenolic content, flavonoid content, and total antioxidant capacity were identified. Among these loci, six were found conserved across the three genotypes as well as for color—a result more in line with the hypothesis that increased flavonoids lead to stronger antioxidant foods\textsuperscript{115}. The conflicting results from these studies suggest the complication of using correlative information about foods and more complicated phenotypes. Other studies have looked at antioxidant mapping in in lettuce\textsuperscript{116} and in broccoli\textsuperscript{117}.

\textbf{1.17. Glucosinolates}

The metabolic derivatives of glucosinolates—known as isothiocyanates—have been associated with beneficial effects on slowing the development and progression of heart disease and, more notably, cancer. Epidemiological evidence has outlined a consistent correlation to consumption of high glucosinolate containing foods and reduced risk for lung,\textsuperscript{118} colorectal,\textsuperscript{119} breast,\textsuperscript{120} and prostate cancer\textsuperscript{121}. Glucosinolates become isothiocyanates when they interact with thioglucosidases in the plant following physical damage to the cell wall or from microbial digestion in lower gastrointestinal tract of human beings. Specific glucosinolates are more potent inducers of chemopreventative activity, with the derivative of glucoraphanin—sulphoraphane showing an especially significant effect\textsuperscript{122}. The glucosinolate pathway has been well characterized, and there have been successful efforts to increase specific and total glucosinolate content in broccoli by crossbreeding broccoli with \textit{Brassica villosa}\textsuperscript{123}. Recently, a genetic linkage map of broccoli was completed and eight different glucosinolates were associated to the
A total of 14 QTL were identified that contributed between 7.5% and 56% of the phenotypic variation of a single compound. A region on chromosome nine was significant for progoitrin, a compound two biosynthesis steps following glucoraphin—explained 40% of the phenotypic variation—and this region was in a similar location to the region that gets displaced in the high glucoraphanin Brassica villosa crossbreed broccoli cultivar.  

### 1.18. Goitergenic glucosinolates

Rapeseed (Brassica napus) is an economically important oilseed crop. Once the oil is extracted, the resulting byproduct was considered to be a protein source in animal feed. However, supplementation with this product led to a dose-dependent enlargement of the kidney and liver, along with induced problems with thyroid function. These problems were lessened when the meal was replaced with a low glucosinolate line. A glucosinolate-derived byproduct called oxaolidine-2-thione was identified as responsible for the negative effects. Oxaolidine-2-thione is produced following the hydrolysis of glucosinolates—such as progoitrin—contain a hydroxyl group at the C-2 position and the isothiocyanate produced is unstable and cyclizes. Tailoring glucosinolate production to lean to those associated with beneficial health effects over the compounds with negative associations at levels seen in plants has been a primary breeding goal for Brassica species, and it is one where marker-assisted breeding has been employed. An integrated mapping of two double haploid lines of Brassica crosses was used to associate genetic loci to eight glucosinolate compounds. The resulting analysis found 16 significant loci that controlled variation of aliphatic glucosinolate concentration from 1.9-43% More recent studies using a RIL line have repeated this assessment and narrowed down
these findings to three regions that control 5-22% of the phenotypic variation. A transgenic approach was used to silence the GSL-ALK controlling the conversion of glucoraphanin into gluconapin in *Brassica napus*. This process was greatly aided by the completion of a genetic mapping that outlined loci where these genes were located in different *Brassica* species because of the redundancy of this trait in plant genomes. One resulting experimental line showed a reduction of goitergenic progoitrin from 80µM g⁻¹ to 32µM g⁻¹ and an increase of glucoraphanin from 0µM g⁻¹ to 42.6µM g⁻¹. Further understanding of the regions and specific genes in each controlling glucosinolate accumulation will be enable researchers to advance the development of future elite lines of the many agronomically important *Brassica* crops.

1.19. Pharmaceutical compounds

Beyond enhancing crops for nutritional and functional food purposes, breeding methods can be applied to plants that have pharmaceutical importance. Chemicals derived from plants accounted for 51% of the newly approved small molecule drugs between 1981 and 2014, and many active components are present in low levels. While breeding methods have been undertaken to increase the content of bioactive traits, only a few examples exist where gene-association techniques have been applied. To source active ingredients, the pharmaceutical industry has instead focused on optimizing synthetic production, metabolic engineering, and tissue culture techniques. Nevertheless, increasing the bioaccumulation of active ingredient in new cultivars presents a potentially cheaper and more sustainable option. One example where research in this area has been done is in connecting morphine content in opium poppy to underlying genes. This was especially important for Turkish poppy producers because the locally adapted strains
that grow in the area were nearly two-fold lower in morphine content than other elite cultivars. A genome-wide association study was employed to capture the whole genome diversity of the species and was able to identify four markers that significantly contributed to morphine accumulation—accounting for between 10% and 32% of the phenotypic variation. The resulting regions are now being used to produce reliable breeding lines that will thrive in the region.\textsuperscript{131}

The anticancer molecules, and sources for even more potent synthetic analogs, of vinblastine and vincristine are uniquely derived from the Madagascar periwinkle (\textit{Catharanthus roseus}). These compounds are incredibly important for treating a variety of cancers, but the bioaccumulation in the plant for either of them is exceedingly low. Adding to the complication, the biosynthesis of the medicinal alkaloids are long multistep processes with many challenges to increasing accumulation in the plant or to produce through other biotechnology based methods.\textsuperscript{132} A biparental mapping population of periwinkle was constructed and analyzed for various alkaloid compound content. A combined content of vinblastine and vincristine and two precursor molecules to them—vindoline and catharanthine—were associated to the genetic linkage map through composite interval mapping. The results found 10 significant QTLs, one for combined vinblastine and vincristine contributing 30.9% of the phenotypic variation; this overlapped with one of the two QTLs for vindoline accumulation that explained 7.6% variation (the other explained 12.6%); and two catharanthine QTLs in both the leaf and root content explaining 12.1%, 14.8%, 27.3%, and 28.3% variation, respectively. This plant also contains important compounds for managing cardiovascular disease: serpentine
and ajmalicine. One QTL for serpentine explained 31.9% of the total variation and two QTLs for ajmalicine explained 9.9% and 23.8%, respectively.\textsuperscript{133}

Throughout the world, artemisinin combination treatments are used to treat malaria. These compounds are derived from the \textit{Artemisia annua} L. plant. A bi-parental mapping population of this medicinal crop was used to make a genetic linkage map and associate artmenisin content to the genome. This resulted in 4 QTLs on a single chromosome that accounted for 20-38\% of the phenotypic variation\textsuperscript{134}. Breeding efforts have been able to raise the artemisinin content from the naturally present 0-0.39\% dry weight to a height of 2.4\% of dry weight focusing on breeding these regions; however, there has been great difficulty in stabilizing these lines and also in making them be high yielding. Continued work in this area is necessary to lower the cost of artemisinin and increase access of the drug in areas affected by the disease\textsuperscript{135}.

\textbf{1.20. Phytate}

Phytate (phytic acid) is a chemical with a strong ability to chelate cations such as Ca, Mg, Zn, Cu, Fe, and K. This action creates an insoluble salt known as phytate, decreasing the availability of essential nutrients from plant sources. Phytate can also decrease the digestibility of protein, carbohydrates and fats from plant sources\textsuperscript{136,137}. Phytate is found primarily in cereal grains and legumes. It is not all negative; there has been an association between phytate and some positive health outcomes because of the ability of it to bind metal ions that induce free radicals and preventing excessive mineral accumulation in the form of kidney stones. Phytate can be degraded by the enzyme phytase, and it is often added to plant feed in industrial animal production. Breeders and public health officials familiar with the areas should carefully weigh these options when
employing a low-phytate crop variety\textsuperscript{138}. In a study with soybeans, the concentration of phytate was associated to a genetic linkage map. This study identified 2 QTLs controlling 15.4\% and 24.3\% of the variation for phytate concentration. QTLs for phosphate, iron, manganese, and zinc were also identified, but only the QTL for phosphorous localized in the same region as that for phytate. This suggests that these areas can be targeted to decrease phytate in a breeding program and only see reduced mineral accumulation for phosphate\textsuperscript{139}. One thing to think about would be if mineral accumulation from consuming these soybeans were used as the phenotype. Blood levels in mice mapped to the genome could identify areas to target that encompass both the inhibiting action of phytic acid and the minerals available to be digested.

1.21. Oxalic acid

Oxalic acid is another chemical found in plants, particularly leafy greens, which can react with minerals during digestion, reducing mineral absorption and contributing to the formation of kidney stones\textsuperscript{140}. Oxalic acid content in spinach is high, but there is a good degree of variation among different varieties. To help produce new spinach cultivars with low oxalic acid content, researchers preformed a genome-wide association study on 310 spinach genotypes to map loci significantly controlling oxalic acid concentration. Six total markers were significantly associated to oxalic acid content. Of these, two markers strongly contribute to oxalic acid content in all five of the models used to analyze the population. These findings suggest that oxalic acid is a trait in spinach where multiple genes and gene regions are involved in its bioaccumulation in the plant. The regions identified here can be further investigated for candidate genes as genome
resources are made available as well as be used to guide precision and molecular
breeding going forward to create spinach lines with low oxalic acid. 141

1.22. Cyanide

There is an understood link between the bitterness of apricot seeds and the
concentration of cyanogenic glycosides, compounds that are converted into toxic
hydrogen cyanide when the plant is damaged. Apricot seeds are not normally consumed
for this reason. The seeds are also not utilized as a source for edible byproducts either, in
spite of being a great source of oil, protein, and fiber, due to the presence of the toxic
compounds. A QTL study investigating the accumulation of amygdalin, a specific
cyanogenic glycoside in apricot seeds, was looking to find ways to breed varieties
without these negative traits. NMR quantification of amygdalin in a RIL population of
118 apricot lines found associations to seven loci explaining 12-16.9% of the phenotypic
variation. None of these QTL lined up with functional markers for amygdalin
biosynthesis, therefore other factors may be involved in cyanogenic compound
accumulation in apricot seeds. 142 Almonds and cassava are other commonly consumed
crops that can produce hydrogen cyanide and have negative health effects for human
beings.

1.23. Moving forward

The identification of a QTL for a specific trait is only an initial step toward
producing a new variety; advanced breeding techniques need to be used to apply the
information found in an association study. Development of new varieties with
information from QTL studies can be done by using marker-assisted selection. 26 This
process becomes difficult if the genetic control for the phenotype is the sum of many
different loci that contribute small amounts of phenotypic control respectively. Genomic selection pools large amounts of marker data for a particular line based on identified values relevant to each genotype in a QTL to create a genome estimated breeding value and choose parents to create the most optimal crosses for new lines \(^{143}\). Additionally, methods of gene editing are advancing rapidly so a better understanding of the underlying genetic control for candidate genes for health relevance will become necessary to get the most from the new technology \(^{144}\).

One area for improvement in this area is introducing more biochemically relevant phenotyping for human health. Future studies can start using \textit{in vitro}, \textit{ex vivo}, and \textit{in vivo} models for human beings to screen these plant populations to capture more of the complex path to that genomic component of plants ultimately leading to the health benefits of foods. Although, the huge sample load of treatments necessary for screening plant mapping populations will make many screening programs difficult to achieve.

Combating diseases with a healthy diet can be further aided by the agricultural sector producing plants that have been tailored to favor increased health-promoting aspects of the plants. As presented in this review, many studies have been undertaken that were able to associate a phenotype relevant for human health to different genomic loci. Generating QTL data is important to further the understanding of how the genes controlling the biosynthetic pathways of specific plant bioactive compounds, and the regions can be used to guide specific breeding studies.
REFERENCES


(47) Yun, L.; Larson, S. R.; Jensen, K. B.; Staub, J. E.; Grossl, P. R. Quantitative trait loci (QTL) and candidate genes associated with trace element concentrations in perennial grasses grown on phytotoxic soil contaminated with heavy metals. *Plant Soil* 2015, 396 (1-2), 277–296.


(53) Gannon, B.; Kaliwile, C.; Arscott, S. A.; Schmaelzle, S.; Chileshe, J.; Kalungwana, N.; Mosonda, M.; Pixley, K.; Masi, C.; Tanumihardjo, S. A. Biofortified orange maize is as efficacious as a vitamin A supplement in Zambian children even in the presence of high liver reserves of vitamin A: a community-


Hoadley, J. E.; Balentine, D. *Petition for a Health Claim for High-Amylose Maize Starch (Containing Type-2 Resistant Starch) and Reduced Risk Type 2 Diabetes Mellitus (Docket Number FDA-2015-Q-2352)*; Office of Nutrition and Food Labeling, Center for Food Safety and Applied Nutrition.


Weston Bussler; Joseph Hildebrand; Catherine Mixon; Ty Wagoner; Slavko Komarnytsky; GabrielKeith Harris. Flavonoid Supplementation and Cardiovascular Disease. In *Dietary Supplements in Health Promotion*; CRC Press, 2015; pp 281–318.


CHAPTER 2. HEALTH-MODIFYING LOCI AFFECT HUMAN COLORECTAL CANCER CELL GROWTH IDENTIFIED ON GENETIC MAP OF BROCCOLI (BRASSICA OLERACEA L. VAR. ITALICA)
2.1. Abstract

Plant genomes are an invaluable resource for expanding understanding of the relationship bioactive food components and human diseases. Diets that emphasize cruciferous vegetables have shown a decreased incidence in development of colorectal cancer—the third deadliest cancer for both men and women. A quantitative trait loci study was performed on a nearly isogenic broccoli population to associate reducing colorectal cancer cell growth to the broccoli genome. To quantify the anticancer phenotype, extracts were characterized from a collection of 125 near isogenic broccoli samples grown at two locations in two different years for their ability to selectively inhibit HT-29 colon cancer cell growth. The population exhibits broad variation in growth inhibition and mapping to the broccoli genetic linkage map results in six health-modifying QTLs. Three QTLs located on chromosomes two, four, and nine combine to account for 53.8% of the variation in cell growth inhibiting activity. Metabolite correlations to HT-29 cell growth inhibition showed sinigrin and quercetin both corresponded to reduced potency and glucoraphanin strongly correlated to increased cell growth inhibiting activity. Within the QTL region associated to cancer cell growth inhibition, a total of 13 putative candidate genes from the recently annotated Brassica oleracea genome involved in biosynthesis and retention of metabolites relevant to cancer cell growth inhibition were identified. These genetic loci and underlying genes provide a valuable resource for guiding breeders to produce broccoli with strong and consistent chemo-preventative properties.
2.2. Introduction

For most foods connected to reducing chronic disease risk, the variability of metabolite concentration of what is actually consumed by a person is large, making it impossible to know how much must be consumed regularly to achieve health benefits. Plants provide the majority of macronutrients, vitamins, minerals, and diverse bioactive phytochemicals humans consume, and it is understood that people who consume diets rich in whole plant sources are associated with having a reduced risk of developing certain chronic diseases such as cancer. As research to better understand this connection has expanded, results have not always showcased this generally understood protective effect. Epidemiological research needs to be paired with controlled models, in vitro and in vivo, to confirm any mechanistic activity for preventing a disease and would greatly benefit from more direct collaboration between plant science and nutritional sciences. Complicated metabolite profiles in plants present an extreme challenge for determining how consumption of plant-based food leads to disease preventing effects.

Plant metabolites are incredibly plentiful, with approximately 30,000 types of metabolites in every plant. For many chemical structures, the biological activity is completely unknown. In addition, plant metabolites undergo many changes during postharvest processing which may affect bioavailability in people. Human and microbial metabolism can further alter the structure and bioactivity of food derived compounds. In addition, the metabolite concentration in crops varies greatly between different plant species and different cultivars. An example of the vast genetic differences between the plant cultivars can be seen in rice (Oryza sativa) where this one crop contains over 40,000 different cultivars worldwide.
Advanced genomics and metabolomics technologies have provided plant breeders with better tools for identifying plants with measurable disease preventing effects \(^{12}\). Crop breeding programs have traditionally focused on controlling quantitative traits such as yield, shelf life, pest resistance, disease resistance, and optimizing sensory aspects for consumers \(^{13}\). Utilizing genetic linkage maps along with a genotyped plant population, it is possible to associate different traits to genetic markers based off of ratios of recombination between genes and how those compare to fluctuation in a quantitative trait value. The identified regions are known as quantitative trait loci’s (QTLs) and within them candidate genes can be found, enabling marker assisted breeder selection and incorporation of desired regions into plant offspring \(^{14}\). QTL studies have been used in the past to determine genome locations of dynamic traits such as crop yield \(^{15}\). The continued focus on maximizing these economically important traits ignores how the concentrations of important health promoting bioactive compounds in plants have been affected \(^{13}\). Utilizing high throughput biological disease models with plant populations suitable for genetic mapping, breeders can use these same techniques to identify health-modifying QTLs. However, as beneficial this sounds, breeding plants with added health benefits will only flourish if economically viable varieties can be developed which incorporate traits such as yield together with QTLs \(^{16}\).

Cancers of the colon and rectum are the third most diagnosed (132,700 per year) and third most deadly (49,000 per year) of all cancers in the United States according to the American Cancer Society \(^{17}\). \textit{Brassica oleracea} subspecies \textit{italica}, otherwise known as \textit{broccoli}, has a well-deserved reputation for being a healthy food. Increased consumption of broccoli has been associated with reduced risk of colorectal cancer in
humans 18,19. Broccoli contains vitamins, previtamins, minerals, phenolic compounds, flavonoids, and organosulfur compounds that could all be contributing to its cancer preventing characteristics 20-21.22.

Broccoli contains organosulfur compounds called glucosinolates, which are precursor molecules to isothiocyanates. These breakdown products have shown extensive chemo preventative activity and are specifically known for their effect on colorectal cancer 23-26. Glucosinolates are not active while they are bound to their glycoside, but they become active when plant cells are damaged allowing the enzyme myrosinase (MYR) to come into contact with the glucosinolates and cleave the glycoside. Once the glycoside is cleaved, an unstable aglycone is formed which breaks down into isothiocyanates but only at neutral pH 27,28. These isothiocyanates are the main bioactive components with respect to anti-cancer activity. In broccoli, the most notable anti-cancer isothiocyanate is sulphoraphane (derived from glucoraphanin). Glucosinolate break-down products, most notably sulphoraphane, inhibit cancer cell growth by inducing apoptosis and cell cycle arrest, acting as direct antioxidants, and modulating phase-I and phase-II detoxification enzymes 24,26,29-34.

Broccoli also contains pre-vitamin A, alpha and beta carotene which have shown a correlation to reduced cancer rates; although experimental results indicate more of a supplementary/indirect effect (Tocopherols & Tocotrienols) 35. Although broccoli is rich in minerals, it is unclear whether mineral concentrations deter the development of colon cancer Se 36. Two flavonoids in broccoli that have shown connections to cancer prevention and therapy are kaempferol 37 and quercetin 38,39. Both have shown activity for reducing growth of colon cancer in animal models and in vitro studies 40. There is a
potential for compounds to work together to produce biological effects beyond those of individual nutrients, but little evidence exists to suggest that this is the case for metabolites found in broccoli, with the exception of glucosinolate breakdown products.

Producing broccoli with enhanced colorectal cancer preventing properties has been undertaken in the past, using multiple techniques. In one such effort glucosinolates in the edible portion of the plant were fortified by supplementing soil with increased levels of sulphur. Another group managed to cross breed broccoli cultivars containing high amounts of glucosinolates with cultivars containing other desirable market qualities but low glucosinolates. A third approach involved crossing a wild broccoli ancestor *Brassica villosa* to an economically superior breeding line to vastly increase glucoraphanin and total glucosinolates concentration. All of these methods focused on increasing the overall health of broccoli by changing the glucosinolates profile or increasing a specific bioactive metabolite. None make a point to conduct actual biological tests to prove the increase in health potential of the new broccoli lines. In almost all efforts to produce nutritionally enhanced crop varieties it is extremely rare to see any attempt to validate the health benefits.

Changes in health potential of a food are derived from genetic differences in plants, environment inputs, and postharvest processing. Controlling for these factors will reduce the complexity of a foods overall health value. With advances in marker assisted breeding, genetic transformation, and CRISPR Cas9 genome editing, gene targets for increasing health potential of crops must be identified. In this study genetic loci on the broccoli genome were identified that associated to the ability to slow the growth of cancer cells. The objective was to find candidate genes and metabolites that are
responsible for these effects and gain insights as to how significant metabolites may interact to further reduce the growth of human colon cancer cells. An additional benefit of these findings is to ultimately provide broccoli breeders with the genetic resource to create cultivars with superior cancer preventing properties.

2.3. Methods

2.3.1. Development of breeding population

The genetic mapping population was developed from two broccoli parents, a calabrese-type double haploid (VI-158) crossed with a brocolette neri-type (BNC), with vastly different phytochemical profiles. The 150 line isogenic broccoli population was created by bud pollinating a F1 cross to produce the F2 generation followed by a second bud pollination of the F2 plants to create the F3 population. Details of the population development are described in detail in Brown et al. 2007.

2.3.2. Development of genetic resource

The genetic linkage map has been described previously by Brown et. al 2014, with minor changes. Markers for the genetic linkage map were constructed with single nucleotide polymorphism markers (SNPs), from the sequenced Brassica napus and Brassica rapa genomes. The isogenic population was analyzed for presence of either parent SNPs, then markers were placed into linkage groups representing different chromosomes and ordered based off of linkage disequilibrium with JoinMap software.

2.3.3. Plant Material Preparation

Broccoli florets were grown in the North Carolina State Piedmont Research Station, Salisbury, NC in 2009 and 2010. Plants were harvested individually based on maturity and shipped on ice to a facility where they were flash frozen with liquid nitrogen
and stored in -80°C to prevent enzymatic degradation. Details of the harvesting procedure are further described in Brown et. al 2014. Liquid nitrogen frozen samples were freeze-dried and ground into a fine powder. Methanol extracts (50%) filtered through 2µM pore filters from this population were prepared for cell treatments. Further description of the extraction procedure described in Brown et. al 2014 51.

2.3.4. Extract prep for cell culture study

For cell culture preparation, methanol extracts were transferred to new pre-weighed 1.5 mL centrifuge tubes and the solvent was removed by 4 hours of vacuum concentration using a SPD1010 SpeedVac™, Thermo Scientific Waltham, MA, under low heat setting. Dried samples were weighed and diluted to 30 mg/mL with dimethyl sulfoxide (DMSO) and stored in 80 °C until ready for treatment.

2.3.5. Digestion of extract for anticancer assay

To model plant enzymatic activity during mastication and bacterial induced myrosinase, an external myrosinase (EC Number 3.2.1.147), purchased from Sigma-Aldrich, St. Louis, MO, was used to hydrolyze sulphur containing compounds into their breakdown products. Broccoli samples were enzymatically digested with 0.025U/mL myrosinase for 2 hours in neutral pH media52 prior to any cell treatment.

2.3.6. Cell culture parameters

Human colorectal adenocarcinoma cells (HT 29) obtained from American Type Culture Collection (ATCC), Manasses, VA, were cultured and maintained in DMEM supplemented with 10% FBS and 1% antibiotic incubated in a 5% CO₂ environment at 37°C. Non-cancerous human colon epithelial like cells (CCD 841 CoN) purchased from ATCC, Manasses, VA, were used as a non-cancerous control. CCD cells were also
cultured and maintained in DMEM supplemented with 10% FBS and 1% antibiotic incubated in a 5% CO₂ environment at 37°C.

2.3.7. **SRB Viability Assessment**

Each broccoli extract was tested for anticancer effect by measuring growth inhibition effects on HT-29 cells using a Sulforhodamine B (SRB) test. The HT-29 cells were seeded at a concentration of 1x10⁴ cells per well in a 96 well plates with 100μL of complete media. Cells were incubated for four hours to allow cell treatments to be prepped and digested. The extracts were administered at a concentration of 100μg/mL in a 100μL vehicle of 0.7% DMSO, 1.17% double distilled water plus enzyme and complete media. A DMSO only vehicle control and a paclitaxel positive control, Sigma-Aldrich, St. Louis, MO, were used to validate each assay and compare to broccoli samples. Cells were allowed to grow uninhibited for 72 hours in a 5% CO₂ environment at 37°C.

Following the incubation period, the media was removed and living cells were fixed to the plate with 10% trichloroacetic acid (TCA) Sigma-Aldrich, St. Louis, MO, for 1 hour at 4°C. Plates were then washed four times with H₂O to remove dead cells and dried overnight. The colorimetric assay was performed by adding 100μL of 0.056% SRB dye Sigma-Aldrich, St. Louis, MO, in 1% acetic acid treatment for 30 minutes to stain the remaining cells. Dye solution was then removed and plates were washed with 1% acetic acid four times to remove any residual dye. The remaining SRB dye bound to cells could then be extracted with addition of 200μL of 10mM Trizma (Sigma-Aldrich, St. Louis, MO), solution and rotation for 30 minutes. Extracted dye was moved to a new plate and measured at 510nm with a BioTek Synergy H1 spectrophotometer. Cell viability was
determined as a percent reduction of absorbance values compared to the vehicle control and listed as percent cell growth reduced by each treatment. All samples were tested in triplicate and the average of the three was recorded. Two broccoli extracts were obtained from each plant line grown in 2009 and 2010. Assessment of average cell viability of both broccoli parent line extracts with and without the addition of myrosinase was calculated, along with a dose response to calculate the inhibitory concentration to kill 50% of cells (IC$_{50}$) and determine effective treatment range for each parent separately.

2.3.8. Apoptosis, Necrosis Assessment

Non-cancerous CCD cells were cultured in 6 well plates at a concentration of 2.5x10$^5$ cells per well and administered a 200µg/mL digested broccoli sample or a negative control (DMSO only) for 72 hours. Cell apoptosis following addition of broccoli extracts was measured with the YO-PRO®-1 Iodide fluorescent dye analysis kit (Thermo Fisher Scientific) as instructed on a BD Accuri flow cytometer\textsuperscript{55}.

2.3.9. QTL study

MapQTL\textsuperscript{®}6 software by Ooijen et. al 2002\textsuperscript{56}, was used to associate single nucleotide polymorphism (SNP) markers from the broccoli genetic linkage map with percent growth reduction in both HT-29 cells and control. Populations grown in 2009 and 2010 were analyzed separately with a third analysis carried out using the average of two years. Each analysis was conducted using 10,000 iterations over a 0.5cM step distance. Kruskall Walls single factor analysis was run to showcase significantly different Brassica SNPs and the growth reduction of HT-29 cells. Interval mapping was used to find presumptive locations and identify potential cofactor markers. The non-restricted MqM Mapping tool was then used to find QTLs. Regions were refined by manually adding
cofactor markers. A location was considered a QTL if the observed values were greater than the Logarithm of the odds (LOD) score was over 2.5 (determined by a permutations test, equivalent to a p-value < 0.05). The span of a QTL was determined by the first marker adjacent to the QTL below the 4.0 LOD threshold when at least a 2X LOD drop-off for adjacent markers was observed. If the regions from two or more analyses overlap, they are considered shared and the highest of the three is taken as the LOD value. For each QTL, the chromosome position, additive gene effect, dominance gene effect, and percent variation in trait explained were recorded. Visual depiction of broccoli map and QTL locations were made with R.

2.3.10. Metabolite analysis and correlations

Values for 27 metabolites for every broccoli extract have been previously quantified by Brown et. al (2014 ref, 2015 ref, & in preparation). Metabolites identified include: boron, calcium, copper, iron, potassium, magnesium, manganese, sodium, phosphorus, sulphur, zinc, glucoiberin, progoitrin, glucoraphanin, sinigrin, gluconapin, total aliphatic glucosinolates, glucobrassicin, neoglucobrassicin, total indol glucosinolates, gluconasturtiin, total glucosinolates, gamma tocopherol, alphatocopherol, total tocopherol, lutein epoxide, lutein, noxanth, beta carotene, total carotenoids, phenolic acids, quercetin, kaempferol, and total phenolics. These metabolite values were correlated to the cancer cell growth inhibition. For glucosinolates, the percent makeup of each class was correlated using GraphPad PRISM. Significant associations (p-value < 0.05) between metabolites and activity were compared to regions identified in the QTL gene pathway analysis. A forward and reverse stepwise multivariable regression model JMP pro 11 (SAS Institute, Cary, NC) was run to determine the significant groupings of
metabolites that explain HT-29 cell growth inhibition. The resulting variables were used to make a partial least squares estimate of the model. The analysis used a cutoff of entry to the model p-value of < 0.09 and a p-value to stay in the model < 0.07. Metabolite pathways that together contribute an association to cancer cell growth inhibition would be included in the gene search if they were not already identified as significant in the individual correlation analysis.

2.3.11. Gene and pathway analysis

To identify the Brassica genes and their relevant metabolic pathways, a candidate gene search list was generated. The genome and predicted genes from the Brassica oleracea reference sequence genome assembly were provided by Parken et. al, and are available at NCBI genes (NCBI Bioproject accession: PRJNA293438). Metabolic pathways of interest were retrieved from the KEGG gene database. All genes near QTL markers of interest within ±500,000 nucleotides were retrieved. These genes and their corresponding predicted proteins were compared to KEGG orthologs in a number of different pathways of interest.

2.4. Results

2.4.1. HT-29 cell apoptosis

The IC50 of broccoli extracts from both broccoli parents (BNC and VI-158) inhibiting growth of HT-29 cells over 72 hours was seen at 205 µg/mL with a 95% confidence interval of 157-267µg/mL (Figure 2.1B). A treatment level of 200µg/mL was thereby chosen to test the broader broccoli population for all further assays. The addition of myrosinase to broccoli extract treatments was essential to see growth inhibition of HT-29 colon cancer cells (Figure 2.1A), due to the flash freezing with liquid nitrogen during
the extraction process. Without myrosinase, HT-29 cell growth was unimpeded. VI-158 and BNC were both able to reduce cell growth but BNC appears more potent than VI-158 (Figure 2.1A).

In the flow cytometer analysis, the number of apoptotic cells in both healthy and cancerous cells after being treated with extracts from either of the broccoli parent lines were compared (Figure 2.1D). Non-cancerous CCD cells treated with 200µg/mL of digested broccoli extract showed less than 1.6% of apoptosis for both parents; a similar result was seen in the vehicle control (DMSO, enzyme, and media). When the cancerous HT-29 cells were treated with either parent extract, an increase in cells undergoing apoptosis was observed. The HT-29 cells exhibited a 4x higher level of observed apoptosis compared to the CCD cells for all treatments and an apoptosis increase compared to the control. This shows that treatments are selectively inhibiting the growth of cancer cells by apoptosis and not damaging the growth of a healthy cell line. Overall, the BNC line extract has greater ability to induce HT-29 cell apoptosis than The VI-158 broccoli extract. This was consistent with the SRB results where BNC was the more potent growth HT-29 cell growth inhibiting broccoli extract (Figure 2.1A).

Extracts from the 150 F2:F3 isogenic population samples derived from the two parents inhibited HT-29 growth from 15-60% over 72 hours (Figure 2C). A large portion of the population extracts were active, reducing cell growth 35-55%. Only a very small proportion of the population provided less than 25% cell growth reduction. The most active samples displayed cell growth reduction over 60% compared to control. The large normally distributed spread of growth inhibiting potential for plant extracts provides a
viable dataset to conduct a QTL study using the genetic linkage map previously
developed by Brown et. al 2014 50
Figure 2.1: Effect of broccoli extracts on cell growth inhibition. (A) HT-29 cell growth inhibition following 72-hour treatment with broccoli extracts from parent lines (BNC & VI-158 with and without MYR), the nearly isogenic broccoli population of broccoli extracts with MYR, 10µM positive control of anticancer drug Paclitaxel and a Vehicle Control (0.67% DMSO and 1% ddH₂O). Dunnett’s test for significant differences from vehicle control marked by * <0.05 and **<0.01. (B) Dose response of broccoli extracts for reducing the growth of HT-29 cells over 72 hours. (C) Frequency histogram of the broccoli population extracts inhibition cell growth % reduced over 72 hours compared to control. (D) Apoptosis: Percentage of cells in apoptosis (R1) or still viable (R2) after 200µg/mL of digested broccoli parent extract treatment over 72 hours. Top row represents non-cancerous CCD 841 CoN colon cells (CCD) and bottom row represents cancerous HT-29 colon cells. The y-axis (FL4A) is fluorescence at 640/675nm and the x-axis (FL3A) is fluorescence at 488/670nm.

2.4.2. QTL study results

Six health-modifying loci were found on five linkage groups representing broccoli chromosomes that eclipsed the 4.0 LOD significance threshold mark (Figure 2.2 and Table 2.1). The regions contain 15 significant markers out of ~540 total, with an average
size of 3.2 cM. The overall QTL region corresponds to an average of 1,952,828 base pairs in the *Brassica oleracea* genome (Table 1). QTL-2, QTL-5, and QTL-6 recurred in multiple analyses and represented the bulk of the HT-29 cell growth reduction attributed to genotypes in this model. QTL-2 on chromosome two explained 18.2% of the variation in HT-29 cell growth inhibition, while QTL-5 on chromosome four explained 17.4%. These QTLs produced differing results in the two parent lines. The BNC parent homozygous for QTL-2 provided 11.87% more cell inhibition than the homozygous VI-158 parent, while BNC homozygous for QTL-5 provided 14.4% more inhibition than homozygous VI-158. QTL-6 on chromosome nine explained 18.2% of the variation in cell inhibition but only genotypes from BNC increased HT-29 cell growth inhibition (16.34% increase). QTL-6 also showcased the highest overall percent cell inhibition average increase compared to all other QTLs. For the weaker QTLs (QTL-1, QTL-3 and QTL-4; which accounted for substantially less variation in cell inhibition, effects of homozygous VI-158 line were stronger than in the BNC line in QTL-1 and QTL-3 but the opposite in QTL-4 (Table 2.2). The three other QTL explained 5.7% (QTL-1), 7.4% (QTL-3), and 5.2% (QTL-4) of the variation. None were found in more than one analysis, providing less confidence of their overall contribution to reducing the growth of colon cancer cells. Together these six regions provide a feasible amount of genomic information to be searched for relevant cancer cell inhibiting gene candidate searches. QTLs are all shown on the broccoli genetic linkage map (Figure 2.3).
Figure 2.2: Whole genome multiple QTL mapping results for 3 different analyses red dots are broccoli grown in year 2009, blue dashes are broccoli grown in 2010, and green dash-dots are average of both years. Each graph represents one chromosome linkage group with positions listed as centimorgans (cM).
Figure 2.3: Brassica genetic linkage map. Significant QTL are shown as red bars. Distances are in centimorgans (CM) and the start and end distances are included with the QTL. Markers contained in the regions are shown to the right of each chromosome.
2.4.3. Multivariate metabolite regression

Using multivariate regression modeling, four metabolites were identified (Sinigrin, gluconapin, glucoraphanin, and quercetin) that account for a substantial portion ($r^2 = 0.58$) of the cell growth inhibition activity (Figure 2.4). Glucoraphanin is the most significant metabolite (FDR log worth = 18.909) contributing increased cell growth inhibition. Sinigrin has a much less significant, but also positive, association with cell growth inhibition FDR log worth = 2.244). Gluconapin and quercetin concentrations appear to be negatively associated with growth inhibition activity.

Figure 2.4: Forward stepwise and reverse multivariate metabolite regression vs cancer cell growth inhibition (95% confidence intervals red dotted line) between metabolites significantly associating to HT-29 cell growth inhibition over 72 hours. Gluconapin, glucoraphanin, sinigrin and quercetin are metabolites included in the model (adjusted $r^2$=0.56, FDR p-value <0.05)
Table 2.1: Multivariate regression individual metabolite contributions

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Effect Estimate</th>
<th>Log Worth</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucoraphanin</td>
<td>19.81</td>
<td>18.91</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Quercetin</td>
<td>-7.67</td>
<td>2.73</td>
<td>0.0019</td>
</tr>
<tr>
<td>Sinigrin</td>
<td>5.11</td>
<td>2.24</td>
<td>0.0057</td>
</tr>
<tr>
<td>Gluconapin</td>
<td>-4.76</td>
<td>1.96</td>
<td>0.0109</td>
</tr>
</tbody>
</table>

2.4.4. Candidate gene hunt

The biosynthesis pathways of significantly associated metabolites, identified in the correlation model, were compiled utilizing curated databases (KEGG, Plant Cyc, etc.) and prior research (Figure 2.5). Using biosynthesis genes from the annotated *Brassica oleracea* genome as search targets, all corresponding genes contained in each QTL region were identified. Annotated sequencing data from the *Brassica oleracea* genome assembly that covered an average of 1.95 million base pairs was used to retrieve base pair codes found within each region. The sequence data was searched for 45 genes from aliphatic glucosinolate and phenolic compound biosynthesis pathways. It was found that 13 putative candidate genes were within regions that could explain the variation in broccoli’s ability to inhibit the growth of HT-29 cells. Relevant genes identified are outlined in green (Figure 2.5) and listed as putative genes (Table 2.2). Nine of the genes were homologous to glucosinolate biosynthesis genes and two were homologous to genes in the production of phenolic compounds. In the glucosinolate pathway genes were found to be homologous to the core biosynthesis genes for aliphatic glucosinolates—CYP83A1, GSTF11, and GGP1. Three other genes from the glucosinolate biosynthesis pathway identified in the gene hunt responsible for glucosinolates side chain modifications including: FMO, APOP2, APOP3, and GSOH. These genes control the accumulation of glucoraphanin, gluconapin, progoitrin, glucoiberin, and sinigrin. From the phenolic pathway, one upstream pathway gene, CHS, was identified. Another phenolic pathway
gene, DFR, was the lone gene found in QTL-4. DFR is a branch-point gene in the conversion of dihydroquercetin to leucocyanidin. No candidate genes of interest were found in the assembly from the *Brassica oleracea* sequences from QTL-1.
Table 2.2: HT-29 cell growth inhibition QTLs and putative candidate genes.

<table>
<thead>
<tr>
<th>QTL</th>
<th>Chr</th>
<th>LOD</th>
<th>( \mu P_1 )^a</th>
<th>( \mu H )</th>
<th>( \mu P_2 )^b</th>
<th>PE</th>
<th>Markers Contained</th>
<th>BP(^c) Begin</th>
<th>BP(^d) End</th>
<th>Putative Candidate Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C01</td>
<td>5.1</td>
<td>40.9</td>
<td>35.2</td>
<td>33.2</td>
<td>5.7</td>
<td>Bn-C1-p05199797, Bn-C1-p06110842, Bn-C2-p48985239, Bn-C2-p49174598, Bn-C2-p49178699, Bn-C2-p49684079</td>
<td>4,699,797</td>
<td>6,610,842</td>
<td>Flavin-containing monooxygenase-like protein</td>
</tr>
<tr>
<td>2</td>
<td>CO2</td>
<td>10.7</td>
<td>27.0</td>
<td>36.6</td>
<td>38.8</td>
<td>18.2</td>
<td>48,485,239</td>
<td>50,184,079</td>
<td>Flavin monooxygenase-like protein, GMP synthase [glutamine-hydrolyzing]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CO3</td>
<td>4.2</td>
<td>37.7</td>
<td>31</td>
<td>36.4</td>
<td>5.2</td>
<td>Bn-C3-p60769039, Bn-C3-p61187831</td>
<td>60,269,039</td>
<td>61,687,831</td>
<td>Flavin monooxygenase-like protein, GMP synthase [glutamine-hydrolyzing]</td>
</tr>
<tr>
<td>4</td>
<td>CO4</td>
<td>6.7</td>
<td>30.2</td>
<td>42.6</td>
<td>43.9</td>
<td>7.7</td>
<td>Bn-C4-p03415373, Bn-C4-p03809977, Bn-C4-p32269978, Bn-C4-p32309236, Bn-C4-p35007222</td>
<td>2,915,373</td>
<td>4,309,977</td>
<td>Dihydroflavonol 4-reductase family</td>
</tr>
<tr>
<td>5</td>
<td>CO4</td>
<td>8.0</td>
<td>26.4</td>
<td>39.5</td>
<td>40.8</td>
<td>7.7</td>
<td>Bn-C4-p32309236, Bn-C4-p35007222</td>
<td>31,769,978</td>
<td>35,507,222</td>
<td>Cytochrome P450 2C family 83 2C subfamily A</td>
</tr>
<tr>
<td>6</td>
<td>CO9</td>
<td>11.7</td>
<td>29.7</td>
<td>40.2</td>
<td>46.1</td>
<td>18.2</td>
<td>Bn-C9-p01016890, Bn_A09_00100554</td>
<td>516,890</td>
<td>2,073,330</td>
<td>Chalcone synthase, Glutathione S-transferase, 2-oxoglutarate-dependent dioxygenase</td>
</tr>
</tbody>
</table>

^a Population growth location  
^b Chromosome linkage group number  
^c Position on genetic linkage map (cM)  
^d Logarithm of the odds  
^e Mean for homozygous QTL marker allele from parent 1 (VI-158)  
^f Mean for heterozygous marker allele  
^g Mean for homozygous QTL marker allele from parent 2 (BNC)  
^h Percent of phenotypic variation explained by QTL  
^i Beginning base pair region used for candidate gene search in *Brassica oleracea* genome  
^j End base pair region used for candidate gene search in *Brassica oleracea* genome
Figure 2.5: Biosynthesis pathway for aliphatic glucosinolates (A) and phenolic compounds (B) in broccoli. Green boxes highlight genes identified in QTLs.
2.5. Discussion

This research confirms previous work showing that broccoli extracts are able to reduce the growth of human HT-29 colon cancer cells by inducing apoptosis. This understanding was furthered by showing broccoli extract treatments did not induce apoptosis of non-cancerous CCD cells. A large difference of activity between the two broccoli parents and an overall broad range of cancer cell inhibiting activity (between 15-65% cell growth inhibition) was found. This provided a quantitative phenotype ideal for QTL mapping as the range of activity in the population.

This work was able to identify 6 QTLs that eclipsed the 4.0 LOD genome-wide thresholds where loci were associated to reducing the growth of HT-29 cells. The vast majority of variation in HT-29 growth inhibition was explained by QTL-2, QTL-5, and QTL-6. All three of these loci showcased a large increase in growth inhibiting activity when the genotypes for the broccoli that produced the extract derived from Broccoli Parent 2 (BNC). Breeding programs can use these identified regions to guide production of stronger chemo-preventative broccoli. The BNC parent line was found to hold most all of the preferential genotypes related to inhibiting growth of HT-29 cells and could be used as a parent for future breeding lines. The population produced broccoli that are homozygous for each QTL favoring higher growth inhibiting activity and can be used for inclusion into a new broccoli population. With the genome-anchored markers in *Brassica oleracea* identified for the QTLs, marker assisted breeding and genomic predictive breeding strategies can be applied to attempt to make new strong chemopreventative broccoli cultivars with different populations. Further confirmation of these QTLs can be made with bi-parental populations derived from different parents,
using fine-mapping to target a smaller region of the genome that encompasses a specific QTL, or a by using a diversity panel of the same crop to conduct a genome-wide association study.

It is no surprise that glucoraphanin has the greatest effect since it is the precursor to the isothiocyanate sulforaphane, a compound from broccoli shown to have ant carcinogenic activity \(^{23-26}\). Although sinigrin also produces a small positive increase in cell growth inhibition, it has shown little ability to slow the growth of colon cancer cells in previous work \(^60\). The significance of sinigrin could be due to many of the enzymes involved in its biosynthesis coinciding with the biosynthesis of glucoraphanin. The gluconapin metabolite concentration negatively impacted HT-29 cell growth inhibition and any work to develop broccoli lines with improved cancer reducing properties should strive to minimize or eliminate this metabolite. The only non-glucosinolate compound found to be significant in the model is the flavonoid quercetin. Similar to gluconapin, quercetin has a slight drag on the growth inhibition potential of the broccoli population extracts. This finding is consistent with large-scale studies correlating individual flavonoids with colon cancer risk that have shown no effect by quercetin \(^61\)

Health-modifying QTL regions found in this study were similar to those found in other broccoli mapping studies that focused on aliphatic glucosinolates or phenolic compounds. Mithen et. al 2003 found that broccoli produced with alleles on chromosome 9 from the Brassica villosa genome in a similar loci to this study’s QTL-6, had an increase in both glucoraphanin concentration and ability to induce phase II detox enzymes \(^44\). Sotelo et al. 2014 focused on identifying loci that contribute to non enzymatic thermal degradation of glucosinolates also looked at QTLs for glucoraphanin.
They found that the major contributor to variation in glucoraphanin content was coming from a locus at the beginning of chromosome nine, in the same area as this study’s QTL-6. However, this QTL did not provide a concise area or candidate gene to allow for a more robust comparison. A third mapping study identified QTLs for all of the glucosinolates found within broccoli but no regions were found to co-localize with this study’s QTLs. This study conducted a candidate gene analysis for relevant glucosinolate biosynthesis genes, but did not find a QTL that contained the most important gene for retention of glucoraphanin (GSL-OH). The most closely related study to the work done here was by Brown et. al 2015, where the same population was used to map glucosinolate accumulation onto the same genetic linkage map used here. In their work, each type of glucosinolate along with total glucosinolates (indole, aliphatic, and aromatic) was mapped to the broccoli genome, resulting in 14 significant loci. Of these QTLs, three share either the same genetic loci or overlapping physical locations in the Brassica oleracea assembly to QTL-2, QTL-4, and QTL-6 identified in this analysis. However, these glucosinolate related QTLs were from different types of glucosinolates. This study’s QTL-2 matched up to a significant region responsible for accumulation of glucoraphanin and glucoiberin. In this same region Brown et. al identified two candidate genes for MAM and MYB28, but neither were found in the current analysis because the marker region stopped short (by 714,526 base pairs) from where the sequence associated to their QTL ended. This shift in genomic location did allow for the identification of a unique candidate gene for flavin monooxygenase. This study’s QTL-4 matched up with a gluconapin derived QTL, both sharing marker Bn-C4-p03809977, in which a lower average accumulation of gluconapin by broccoli offspring that contained BNC genotypes.
was seen. This loci was unable to be associated to a gene relevant to glucosinolates in previous work, but the new region covered a gene for CYP450 83 subfamily A. This study’s QTL-6 was also identified in the Brown study, in which it showed a significant effect in the bioaccumulation of progoitrin when VI-158 genotypes were passed to the offspring. Gene investigation in this region was able to relate the significant branch-point gene for aliphatic glucosinolates necessary for the conversion of the health relevant glucoraphanin to the detrimental gluconapin as well as progoitrin. Both genes catalyzed from this region are 2-oxoglutarate dependent dioxygenases. In this analysis, when broccoli offspring had BNC genotypes at these loci they had the largest associated increase in HT-29 cell growth inhibition.

The importance of down-regulating GSL-OH to increase the concentration of glucoraphanin was also seen by Robin et. al 2016, where they analyzed inbred lines of cabbage. They found that glucoraphanin and sinigrin could only be detected when there was decreased expression of GSL-OH genes. Controlling for the bioaccumulation of a profile of aliphatic glucosinolates should be a primary objective of breeding programs attempting to produce health-promoting broccoli going forward.

QTL-4 matched with a gluconapin derived QTL, both sharing marker Bn-C4-p03809977, in which a lower average accumulation of gluconapin by broccoli offspring that contained BNC genotypes was seen. This QTL was unable to be associated to a gene relevant to glucosinolates in the previous work, but the new region covered a gene for CYP450 83 subfamily A. In addition, this study’s QTL-6 was also identified in the Brown study, showing a significant effect in the bioaccumulation of progoitrin when parent 1 (VI-158) genotypes were passed to the offspring. Gene investigation in this
region was able to relate the significant branch-point gene for aliphatic glucosinolates necessary for the conversion of the health relevant glucoraphanin to the detrimental gluconapin as well as progoitrin. Both genes catalyzed from this region are 2-oxoglutarate dependent dioxygenases. When broccoli offspring had BNC genotypes at these loci they had the largest associated increase in HT-29 cell growth inhibition \(^{51}\). The observation of the importance of down-regulating GSL-OH to increase the concentration of glucoraphanin was also seen by Robin et. al 2016, where they could only detect glucoraphanin and sinigrin when there was decreased expression of GSL-OH genes \(^{64}\). Controlling for the bioaccumulation of an aliphatic glucosinolate profile should be included alongside yield and disease traits for any breeding program producing broccoli going forward.

An increased concentration of quercetin was related to lowering a broccoli extract’s ability to inhibit growth of HT-29 cells. Two of this study’s QTLs, four and six, contain genes known to be part of the biosynthetic pathway for quercetin. These genes are both crucial to the bioaccumulation of quercetin. Dihydroflavonol reductase (DFR), a branch point enzyme in the quercetin biosynthesis pathway was found in QTL-4 and in QTL-6 gene homologous to chalcone synthase (CHS), an upstream enzyme for quercetin accumulation, was found. DFR is the key step for the conversion of quercetin to leucocyanidin, and when genes from BNC in this region were present the broccoli was a more effective cancer cell growth inhibiting extract, therefore active DFR should be incorporated into future broccoli lines. CHS increases the overall accumulation of phenolic compounds, including quercetin, but without an active DFR it would be leading
to increased accumulation of quercetin and be a potential negative drag on chemopreventative potential.

Gardner et. al 2016 mapped the phenolic compound accumulation and radical scavenging activity in broccoli and found 5 QTLs that overlapped in physical base pairs to this study’s QTL-1, QTL-2, QTL-3, QTL-4, and QTL-6. This study’s QTL-4 was found significant in their analysis but their candidate gene search did not identify DFR. They found CHS in QTL-6 in the same location identified in this analysis, and they found a significant QTL in the same region as this study’s QTL-1. However, they were unable to find any annotated candidate gene from this area in the current *Brassica oleracea* assembly. MYB genes associated with accumulation of phenolic compounds were found in the Gardner et. al 2016’s gene search and co-localized to this study’s QTL-2 (MYB 50), QTL-3 (MYB 48/59), and QTL-6 (MYB 50). Although phenolic findings are not as clear-cut as previous glucosinolate research, breeding programs may also find that changing the expression of DFR and CHS genes in new broccoli will lead to new varieties with optimized health benefits. Phenolic compounds, including quercetin, have been found to have many connections to health promotion so any work in this area should look to retain total phenolic production, but favor more cyanidin over quercetin.

2.6. Conclusion

This work successfully identified human colon cancer relevant QTLs onto a plant’s genome for is understood to be the very first time. This opens up a new method to investigate and build upon the complicated relationships between medicinally and nutritionally relevant plants and how they impact human health. This method has the added benefit of providing a path toward the inclusion and promotion of the nutritional
aspects of crops in commercial lines to add value for both consumers and producers of these crops. The genetic loci can be immediately used to guide breeding efforts for strong chemopreventative broccoli. Genes within the genetic loci can be further investigated to functionally characterize the relationship between plant genes, metabolites produced, and the effects on development of colon cancer in humans. Foods claiming to help prevent colorectal cancer should be further analyzed with the aid of their genomes for their ability to reduce disease risk. In broccoli the genome-aided approach identified different aliphatic glucosinolates contributing to the bioactivity, glucoraphanin and sinigrin helping to increase cancer cell growth inhibition and more gluconapin contributed to decreased cell inhibition potency of broccoli extracts. Future work should expand on the role these different glucosinolates play in changing bioactivity of a broccoli extract. Quercetin concentration showed a slight drag on cancer cell growth inhibition. On a broader scope this study outlines the importance of the effects of plant variety on the interactions of diet and disease.

2.7. Acknowledgements

This work was supported in part by Plant Pathways Elucidation Project (W.B., M.A.L., S.K.), PHHI seed grant 2015–2017 (S.K.), and NCSU faculty start-up funds (S.K.).
REFERENCES


CHAPTER 3. COMPARATIVE INTERACTIONS BETWEEN ALIPHATIC GLUCOSINOLATES INHIBIT CHEMOPREVENTATIVE ACTIVITY.
3.1. Abstract

Broccoli (Brassica oleracea L. Italica Group) is an economically important cruciferous crop that is known for its beneficial effects on human health. Upon damage or consumption, broccoli glucosinolates undergo enzymatic hydrolysis by myrosinase to yield bioactive isothiocyanates. The inherent glucosinolates in the plant tissue and the conditions of enzymatic hydrolysis are critical to the chemopreventive qualities of broccoli, but the individual contributions and potential interactions between glucosinolates are poorly understood. In this study, varying mixtures of four aliphatic glucosinolates from broccoli were investigated putative interactions they have with one another on reducing colorectal cancer cell proliferation following individual compound enzymatic hydrolysis (SEP) and combined compound enzymatic hydrolysis (VAT) for different proportions of aliphatic glucosinolates. Combination indexes for mixtures showed weak to moderate antagonistic interactions in modulating antiproliferative activity among glucoraphanin, gluconapin, progoitrin, and sinigrin that were more prominent when the mixture was VAT hydrolyzed. To determine optimal ratios of aliphatic glucosinolates, a three-point mixture analysis was used. Maximal chemopreventive potential of an “anticancer” broccoli cultivar was predicted for glucosinolate profiles of 81%-84% glucoraphanin, 9%-19% gluconapin, 0-7% others to (adjusted $R^2 > 0.80$). A broccoli cultivar with a glucosinolate profile resembling this ratio was found to have a 2-fold increase in ability to limit growth of HT-29 cells. Future breeding programs for broccoli should use this makeup of aliphatic glucosinolates as a target for future broccoli-based functional foods and dietary supplements centered on chemoprevention.
3.2. Introduction:

Colorectal cancer is the third leading cause of all cancer-related deaths in both men and women in the United States\(^1\). Broccoli (Brassica oleracea var. italica) consumption has long shown an inverse correlation with colorectal cancer risk in epidemiological studies\(^2\). More recent meta-analyses continue to support this finding\(^3,4\). As such, multiple broccoli breeding programs target increased cancer chemopreventive qualities through manipulation of broccoli secondary metabolites, especially glucosinolates and their hydrolysis product profiles\(^5,6,7\).

Glucosinolates are a major class of sulfur-containing secondary metabolites commonly found in cruciferous vegetables,\(^8\) of which 132 unique, naturally occurring forms have been identified and classified as indole, aromatic or aliphatic depending on the amino acid precursor they were derived from\(^9\). Glucosinolates synthesized from valine, alanine, leucine, isoleucine, and methionine are considered aliphatic\(^10\). Glucosinolates themselves do not show any relevant chemo-preventive activity,\(^11\) but exposure to plant or microbial myrosinase (thioglucoside glucohydrolase, EC 3.2.3.1) catalyzes a breakdown into isothiocyanates;\(^12,13\) metabolites most often

![Figure 3.1: Summary graphic for broccoli glucosinolate mixture analysis study](image)
implicated with the chemo-preventative and other health benefits in humans\textsuperscript{14,15}.

Multiple methods of action have been suggested for isothiocyanate bioactivity including: decreasing activity of phase-I detoxification enzymes,\textsuperscript{16} increasing activity of phase-II detoxification enzymes,\textsuperscript{16} modulating the NRF2/ARE pathway,\textsuperscript{17} inducing cell cycle arrest and apoptosis of cancerous cells,\textsuperscript{18} and inhibiting histone deacetylases\textsuperscript{19}.

Studies identifying total glucosinolates from different broccoli cultivars find that the genetics, environmental conditions, and postharvest processing can drastically change the amount and chemical profile of individual glucosinolates\textsuperscript{20–23}. The most common broccoli glucosinolates are: progoitrin (PG), gluconapin (GN), glucoraphanin (GR), sinigrin (SGN), glucoiberin (GIB), glucobrassicin (GB), neoglucobrassicin (NGB), and gluconasturitiin (GNI)\textsuperscript{24}. Their schematic biosynthetic pathway is shown (Figure 3.2). Initial efforts to increase glucosinolates in broccoli have been made by increasing sulfur content in soil,\textsuperscript{25} but the diverse glucosinolate profiles in broccoli raises the question about individual contributions, potential synergisms and antagonistic effects each metabolite has on biological activity, which may make direct breeding for total glucosinolate content impractical. With the wide variability in concentration and the current lack of understanding about interactions at the level of myrosinase degradation and downstream molecular targets, it is difficult to estimate overall potential of broccoli cultivars to reduce colorectal cancer risks.

Glucoraphanin, the precursor to \textit{sulforapane}, is the best characterized glucosinolate in broccoli for chemopreventative properties\textsuperscript{26}. Previous targeted approaches have increased GR by crossbreeding broccoli with a wild relative \textit{Brassica villosa},\textsuperscript{5,6} but bringing in a new species can lead to undesirable functionality of the plant,
and combination effects on increasing other glucosinolates in addition to GR have not been investigated. Future efforts can now use a newly constructed, high-density genetic linkage map for targeted breeding programs for broccoli with optimized glucosinolate profiles. Previously, biological activity within multi-compound interventions was investigated using the method by Chao et al. that establishes presence of additive, antagonistic, or synergistic interactions within the test substance (Figure 3.3A), see Material and Methods section for details). An alternative method used in food science is a multiple point mixture analysis, commonly applied to maximize a specific factor, such as consumer desirability or a functional property, by finding the optimal proportion to combine variables.

This study found that individual chemopreventive contributions of aliphatic glucosinolates are weakened when introduced as a mixture by measuring their effect on HT-29 human colon cancer cell proliferation. Two methods were used to hydrolyze aliphatic glucosinolate mixtures prior to challenging HT-29 cells, the first treatment used pre-hydrolyzed glucosinolates added in proper ratios separately (SEP), and the second treatment used glucosinolates combined together before undergoing enzymatic hydrolysis (VAT). The VAT treatment allows for compounds to be together as the enzyme is catalyzing their breakdown mimicking how the compounds would be in a plant matrix introduced to either plant or microbial myrosinase. Additionally, a three-point mixture analysis (Figure 3.3B) was used to produce a model to predict cell growth inhibition of glucosinolate mixtures of different proportions at various concentrations and determine the optimal ratios of aliphatic glucosinolates to maximize their chemopreventive potential.
Figure 3.2: Biosynthesis pathway for three classes of glucosinolates. Main genes and intermediate metabolites involved in side chain elongation, core structure synthesis, and side chain modification of major (A) aliphatic (B) indole and (C) aromatic glucosinolates.
3.3. Materials and Methods

3.3.1. Chemicals and myrosinase treatment

HPLC grade aliphatic glucosinolate standards of glucoiberin, sinigrin, glucoraphanin, gluconapin, and progoitrin were purchased from Chromodex (Irvine, CA). Myrosinase EC: 3.2.1.147 from Sinapis alba (white mustard) seed was purchased from Sigma-Aldrich (St. Louis, MO). All chemicals and solvents (anhydrous and ACS grade) were purchased from Sigma-Aldrich unless specified otherwise. Individual glucosinolates were enzymatically digested as either separate treatments (SEP) or combined pooled treatments (VAT) in the dose range of 0.5-50 µM with 0.025 U/ml myrosinase for 2 hours in neutral pH media prior to exposure to cells. Methanolic extracts from F$_{2:3}$ broccoli mapping population were kindly provided by Dr. Alan Brown as a natural source of variable glucosinolate concentrations derived from a plant matrix. Methanol extracts (200 µg/mL) underwent the same enzymatic treatment as VAT samples.

3.3.2. Cell culture

Human colorectal adenocarcinoma cells (HT-29), obtained from American Type Culture Collection (Manasses, VA), were cultured and maintained in DMEM supplemented with 10% FBS and 1% antibiotic in a 5% CO$_2$ environment at 37°C. Cells were routinely passaged every three to four days at <90% confluence and only passages 2-16 were used for measuring glucosinolate-induced cytotoxicity and cell proliferation studies.

3.3.3. Viability assessment

Each treatment was tested for anticancer activity by measuring growth inhibition effects on HT-29 cells using a Sulforhodamine B (SRB) test. HT-29 cells were
seeded at a concentration of 1x10^4 cells per well in 96 well plates with 100 µL of complete media. After a four-hour incubation myrosinase digested treatments were administered in a 100 µL vehicle (0.7% DMSO in complete medium). Paclitaxel (Taxol), an anti-cancer cyclodecane isolated from the bark of the pacific yew tree Taxus brevifolia, was used as a reference positive control at 0.5 µM. Following a 72 hour incubation with treatments, the media was removed and living cells were fixed to plates with 10% trichloroacetic acid for 1 hour at 4°C. Plates were then washed four times with water to remove dead cells and dried overnight. The colorimetric assay was performed by adding 100 µL of 0.056% SRB dye Sigma-Aldrich, St. Louis, MO, in 1% acetic acid treatment for 30 minutes to stain the fixed cells. Dye solution was then removed and plates were washed with 1% acetic acid four times to remove any residual dye. The remaining SRB dye bound to cells was extracted with addition of 200 µl of 10 mM Trizma and incubation with shaking for 30 minutes. Extracted dye was measured at 510 nm with a BioTek Synergy H1 spectrophotometer (Winooski, VT). Cell proliferation was determined as a percent reduction of absorbance values compared to the vehicle control and listed as percent cell growth reduced by each treatment. All samples were tested in triplicate.

3.3.4. Determination of IC_{50} and combination index

IC_{50} values for each treatment were calculated using GraphPad Prism 6.0 (San Diego, CA). Combination Index (CI) was calculated to determine the nature of interactions between compound mixtures at various concentrations using the formula in Figure 3.3A. Interactions were scored as very strong synergism (<0.1), strong synergism (0.1-0.3), synergism (0.3-0.7), moderate synergism (0.7-0.85), weak synergism (0.85-
0.9), nearly additive (0.9-1.1), weak antagonism (1.1-1.2), moderate antagonism (1.2-1.45), antagonism (1.45-3.3), strong antagonism (3.3-10), and very strong antagonism (>10) \(^{27,36}\).

### 3.3.5. Mixture analysis

Dose effects from exposure to aliphatic glucosinolates on HT-29 cell growth inhibition were calculated using a three-point mixture analysis (Figure 3.3B). Each sample mixture was hydrolyzed by myrosinase as individual compounds (SEP) or as combined pool samples (VAT) in specific ratios at dose ranges 0.5 \(\mu\)M, 5 \(\mu\)M and 50 \(\mu\)M. Statistical mixture analysis was performed with JMP Pro 12.0 (SAS Institute, Cary, NC) using a standard least squares model to create predictions for cell growth inhibition derived from each compound ratio.

Next, three-point mixture ternary plots were generated to visualize the combined effects of aliphatic glucosinolates and determine the optimal proportion of individual compounds to achieve maximum inhibition of HT-29 cancer cell proliferation. Interactions between each variable were investigated for significance using an alpha of 0.01 and corrected for false discovery rate (FDR) and parameter estimate.
Figure 3.3: Methods for calculating combination index and three-point mixture analysis. (A) Combination index was determined from IC$_{50}$ where Dn is concentration of the compound n that produces 50% effect, and scored for synergistic, additive, or antagonistic interaction. (B) Three-point mixture analysis was performed at doses of 5-50 µM to generate ternary plots of alipathic glucosinolate interactions and determine the optimal proportion of individual compounds to achieve maximum inhibition of HT-29 cancer cell proliferation.
3.4. Results

3.4.1. Comparative potency of aliphatic glucosinolates

When separately assayed for cytotoxic activity, four major aliphatic glucosinolates from broccoli (sinigrin, glucoraphanin, gluconapin, and progoitrin) in their native form showed no effect on HT-29 human colon cancer cell proliferation in concentrations up to 50 𝜇M. Myrosinase-catalyzed hydrolysis of target aliphatic glucosinolates was then performed before addition to cells at the dose range of 0.5 𝜇M, 5 𝜇M, and 50 𝜇M for 72 hours. Breakdown products from glucosinolates showed a clear antiproliferative effect on HT-29 cells in the following order of potency: glucoraphanin \( (IC_{20} = 5.61 \, \text{µM}, \, IC_{50} = 11.80 \, \text{µM}, \, IC_{80} = 24.8 \, \text{µM}) > \) sinigrin \( (IC_{20} = 16.58 \, \text{µM}, \, IC_{50} = 29.51 \, \text{µM}, \, IC_{80} = 52.51 \, \text{µM}) > \) gluconapin \( (IC_{20} = 17.92 \, \text{µM}, \, IC_{50} = 44.76 \, \text{µM}, \, IC_{80} = 111.80 \, \text{µM}) > \) progoitrin \( (IC_{20} = 229.8 \, \text{µM}, \, IC_{50} = 2100.6 \, \text{µM}, \, IC_{80} = \text{N/A}) \) (Figure 3.4).
Figure 3.4: Chemopreventive effects of aliphatic glucosinolates from broccoli. HT-29 human colon cancer cells were treated with 0.5-50 μM myrosinase-hydrolyzed glucosinolates to establish the IC\textsubscript{50} dose required for median effect.
3.4.2. Combination index of aliphatic glucosinolates

The combination index of aliphatic glucosinolates was calculated at IC50 concentration for all compounds in SEP and VAT myrosinase treatments with the exception of some combinations with high amounts of progoitrin, which did not reach adequate cell proliferation at the concentrations analyzed to calculate an IC50 value. Most of combinations clearly showed moderate to antagonistic interactions with glucoraphanin, thus confirming this molecule to be the precursor of primary hydrolysis products responsible for broccoli chemopreventive activity (Table 3.1). Direct comparisons between SEP and VAT myrosinase treatments suggested that there were no potentiating actions between hydrolyzed glucosinolate compounds at the concentrations tested. Most compound combinations showcased antagonistic interactions between aliphatic glucosinolates in either method of myrosinase hydrolysis, but VAT sample mixtures exhibited a trend of higher IC50 value and larger antagonistic index compared to SEP treated counterparts. Two combinations—1/3 GN, 1/3 PG, 1/3 SNG and 1/6 GR, 2/3 GN, 1/6 PG—were nearly additive CI’s when hydrolyzed separately. One treatment 1/6 GN, 2/3 PG, 1/3 SNG hydrolyzed separately displayed a 0.55 CI indicating synergism, but this same treatment was unable to achieve an IC50 after VAT hydrolysis. Multiple present glucosinolates impaired growth inhibiting effects on cancer cells greater than the antagonism seen when glucosinolates are exposed to equal myrosinase alone.
Table 3.1: IC50’s and calculated combination indexes of mixtures of four major aliphatic glucosinolates for inhibition of HT-29 human colorectal cancer cells.

<table>
<thead>
<tr>
<th>Proportion of Each Compound</th>
<th>IC50 (mM)</th>
<th>Combination Index (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SEPT</td>
<td>VATT</td>
</tr>
<tr>
<td></td>
<td>SEPT</td>
<td>VATT</td>
</tr>
<tr>
<td>1 - - - -</td>
<td>11.80</td>
<td>-</td>
</tr>
<tr>
<td>- 1 - -</td>
<td>44.76</td>
<td>-</td>
</tr>
<tr>
<td>- - 1 -</td>
<td>2100.62</td>
<td>-</td>
</tr>
<tr>
<td>- - - 1</td>
<td>29.51</td>
<td>-</td>
</tr>
<tr>
<td>0.33 0.33 0.33 0.33</td>
<td>64.19</td>
<td>52.25</td>
</tr>
<tr>
<td>0.33 - 0.33 0.33</td>
<td>28.76</td>
<td>50.16</td>
</tr>
<tr>
<td>0.33 0.33 0.00 0.33</td>
<td>27.69</td>
<td>46.18</td>
</tr>
<tr>
<td>0.00 0.33 0.33 0.33</td>
<td>54.23</td>
<td>62.06</td>
</tr>
<tr>
<td>0.67 0.17 0.17 -</td>
<td>24.66</td>
<td>45.24</td>
</tr>
<tr>
<td>0.67 0.17 - 0.17</td>
<td>29.84</td>
<td>43.07</td>
</tr>
<tr>
<td>0.67 - 0.17 0.17</td>
<td>25.94</td>
<td>35.03</td>
</tr>
<tr>
<td>0.17 0.67 0.17 -</td>
<td>48.01</td>
<td>56.32</td>
</tr>
<tr>
<td>0.17 0.67 0.00 0.17</td>
<td>47.74</td>
<td>50.23</td>
</tr>
<tr>
<td>0.17 0.17 0.67 -</td>
<td>205.83</td>
<td>62.74</td>
</tr>
<tr>
<td>0.17 - 0.67 0.17</td>
<td>58.31</td>
<td>59.65</td>
</tr>
<tr>
<td>0.17 0.17 - 0.67</td>
<td>48.18</td>
<td>50.00</td>
</tr>
<tr>
<td>0.17 - 0.17 0.67</td>
<td>47.09</td>
<td>48.24</td>
</tr>
<tr>
<td>0.00 0.67 0.17 0.17</td>
<td>66.16</td>
<td>62.67</td>
</tr>
<tr>
<td>0.00 0.17 0.67 0.17</td>
<td>57.15</td>
<td>*</td>
</tr>
<tr>
<td>0.00 0.17 0.17 0.67</td>
<td>47.82</td>
<td>51.77</td>
</tr>
</tbody>
</table>

(-) Not present in mixture, (*) Unable to calculate

3.4.3. Mixture analysis of chemopreventive activity

To determine optimal ratios of aliphatic glucosinolates to maximize chemopreventive potential in broccoli, a three-point mixture analysis for both SEP and VAT digested glucosinolate treatments in the dose range of 0.5-50 µM was conducted. The dots on the triangle represented the experimental points gathered to create a standard least squares model to generate the prediction equation to determine values for the rest of the possible mixtures (Figure 3.3B). No substantial inhibition of HT-29 cell growth was seen at the lowest treatment level (0.5 µM) for any of the mixtures; however, growth inhibition became prevalent when concentrations increased to 25 and 50 µM. In all sample mixtures, the SEP glucosinolate treatments exhibited larger areas of increased cell growth inhibition than the respective VAT glucosinolate treatments: signifying that a
portion of the antagonism on chemopreventative properties seen between aliphatic glucosinolates comes from competition for myrosinase digestion (Figure 3.5).

Interactions between variables for the GR_GN_PG mixture model using both SEP \( (R^2 \text{ adj} = 0.924) \) and VAT \( (R^2 \text{ adj} = 0.934) \) myrosinase treatments revealed that the most significant positive effects for variables contributing to growth inhibition of HT-29 cells in both modes were GR and GN proportion \( (\text{FDR log worth} > 2, p = 0.01) \), and GR and GN proportion interacting with total concentration \( (p<0.0001) \). In the SEP model, additional significance was seen in the positive interaction between the GR*GN*PG compound mixture \( (p<0.0001) \). Negative interactions were evident between GR*GN proportion \( (p<0.0001) \) and GR*PG proportion \( (p = 0.03) \). The VAT model showed the same 2 compound antagonisms \( (p = 0.12 \text{ and } p = 0.15, \text{ respectively}) \) (Figure 3.5A). The difference in significance of the 3 compound SEP model suggested that the increased dispersion in the glucosinolate distribution of VAT samples contributes a negative effect on bioactivity that has overlapped with the dual compound antagonism.

In the GR_GN_SNG mixture, a more bioactive SNG replaced PG. This model showed highest \( R^2 \text{ adj} \) values in both SEP \( (R^2 \text{ adj} = 0.948) \) and VAT \( (R^2 \text{ adj} = 0.967) \) modes. There were many significant positive associations for interaction variables in both of the treatments. In the SEP mode, all interactions besides GR*GN proportion*concentration were significantly associated with bioactivity \( (p<0.01) \). In the VAT mode, GR*GN*SNG proportion was a non-significant factor compared to being highly significant to the SEP model \( (p<0.0001) \). This echoed the result seen in GR_GN_PG and was visualized by increased area of higher growth inhibition of cells in the response surface graphs (Figure 3.5B). These results further confirm a relationship
between chemopreventative properties of aliphatic glucosinolates and their competition for myrosinase. This is especially relevant, since the compounds were tested in nearly equal concentrations. The dual compound interaction of GR proportion * SNG proportion and GN proportion* SNG proportion showcased a negative interaction in both models, suggesting that SNG had a weak negative effect on antiproliferative activity of GR and GN – a pattern also seen in the CI values for these compounds (Table 3.1).

In the GN_ SNG_ PG mixture model in both SEP (R^2 adj = 0.894) and VAT (R^2 adj = 0.855), interactions showed significant positive associations for SNG and GN individual proportion (p<0.0001), and for SNG and GN proportion*concentration (p<0.0001). A negative association was once again seen in GN*SNG proportion for SEP (p<0.0001), as well as for VAT (p = 0.01). New negative associations were found between GN*PG proportion in both SEP (p<0.0001) and VAT (p<0.02) modes, and a negative association between SNG*PG proportion was significant in the SEP mode (p = 0.02). It is again observed that 3 compound proportion interactions significantly contributing to growth inhibition in the SEP model, but not the VAT model, was observed once again (Figure 3.5C).

The final combination of GR_ SNG_ PG showed the lowest R^2 adj of all mixtures SEP (R^2 adj = 0.870) and VAT (R^2 adj = 0.806). This mode only found significant associations (p<0.01) with inhibition of HT-29 cell proliferation in single compounds GR and SNG proportions, and GR/PG proportions*concentration. A nearly significant negative association between GR*SNG proportion (p = 0.13) was present in the SEP model only (Figure 3.5D).
Taken together, the surface response models predicted that 80% threshold inhibition of proliferation of the HT-29 human colon cancer cells was achieved at 50 µM dose by the following ratio: 81%-84% GR, 9%-19% GN, 0-7% other glucosinolates (Figure 3.5A). Similar levels can be achieved by increasing SNG, but broccoli samples do not exhibit high levels of this glucosinolate normally. To test the predicted optimal glucosinolate ratio with real broccoli samples, methanol extracts from a nearly isogenic broccoli lines from a F2:3 mapping population32 were screened for glucosinolate proportions. This identified broccoli line VB067 that contained the closest aliphatic glucosinolate profile proportions (77% glucoraphanin, 14% gluconapin, and 9% others) to the ideal ratio (Figure 3.6A). When tested for its chemopreventative activity in HT-29 cells, this line showed 44% increase in mean cancer antiproliferative properties over the initial broccoli breeding line VI-158 or an average isogenic sister line VB019 (Figure 3.6B).
Figure 3.5: Mixture analysis revealed optimal proportions of aliphatic glucosinolates in broccoli to maximize antiproliferative properties. Ternary plots demonstrate three-compound interactions of (A) GR_GN_PG mixture, (B) GN_PG_SNG mixture, (C) GR_PG_SNG mixture, and (D) GR_GN_SNG mixture in the dose range of 5-50 µM with color scale indicative of the proportion of cell growth inhibition relative to vehicle control. Each sample mixture was hydrolyzed by myrosinase as individual compounds (SEP) or combined pooled samples (VAT) before combined exposure to HT-29 cells.
Figure 3.6: Antiproliferative properties of near-optimal VB067 isogenic broccoli line. (A) Glucosinolate profiles and (B) inhibition of HT-29 cell proliferation by myrosinase-treated crude methanolic extracts of VB067 (near-optimal) and VB019 (average) breeding lines from F_{2:3} broccoli mapping population^{32} as compared to broccoli variety (VI-158). Paclitaxel (Taxol) was used as a positive reference drug (10 µM). Results are expressed as means ± SEM (n = 3). * p < 0.05 and ** p < 0.01 when compared to vehicle control by one-way ANOVA followed by Dunnett’s post hoc test.
3.5. Discussion

The number of compounds present in plants are estimated to exceed 200,000, providing an large number of possible interactions. The recommended way for people to obtain health-beneficial phytochemicals is from food sources due to the potential synergies produced by the sum of metabolites. A few classes of compounds found within the food matrix have been studied for their connection to chemoprevention. Glucosinolates are one phytochemical class prominently associated with colorectal cancer chemopreventative properties, specifically from glucoraphanin, and they are major constituents of brassica vegetables such as broccoli. However, there are 7 other glucosinolates common to broccoli which total to concentrations between 8.37 and 19.51 µM/100g of fresh weight in commercial varieties. On their own, most of these glucosinolates have been shown to have weak chemopreventative properties. Their combined contributions to broccoli chemopreventive properties are poorly understood due to unknown interactions among glucosinolates and potential competition for myrosinase catalyzed formation of bioactive isothiocyanates that modulate downstream cell and molecular targets.

In this study, a mixture analysis of combined chemopreventive activity from major aliphatic glucosinolates from broccoli (sinigrin, glucoraphanin, gluconapin, and progoitrin) was preformed to determine synergistic and antagonistic relationships that contribute to overall bioactivity of the crop. To access combined action of phytochemicals, the classical methods such as isobologram and combination index are typically used, however they are often inadequate to capture complex phytochemical interactions that would be present in the food matrix. Instead, a mixture analysis is a
statistical-model based method, routinely used to perform sensory optimization to understand desirability, product functionality, and customer acceptance, \(^{44}\) that can be applied to characterize complex interactions and identify optimal proportions of 3 unique compounds. To my knowledge, this work is the first attempt to use a mixture analysis to determine interactions between crop-derived phytochemicals for anti-proliferative activity in cell culture. The superimposed ternary contour profile plots showed the optimal ratios of aliphatic glucosinolates in broccoli crop for enhancing its chemopreventive bioactivity (Figure 3.5). The data supported glucoraphanin as the primary source of anti-proliferative activity in a mixture of aliphatic glucosinolates native to broccoli, and a lesser degree of activity derived from GN and SNG. PG alone showed insignificant ability to inhibit cell growth. Analysis of combination index between different glucosinolates showed clear moderate to antagonistic interactions between GR and other glucosinolates used in this study with a finite concentration of myrosinase, however it could not always account for PG contribution due to low bioactivity of this compound (Table 1). This obstacle was circumvented by application of the mixture analysis.

Surface response mixture models predicted the optimal ratio of aliphatic glucosinolates to be 81%-84% GR, 9%-19% GN, and 0-7% other glucosinolates as this combination achieved 80% threshold inhibition of proliferation of the HT-29 human colon cancer cells at 50 µM dose (Figure 3.5). Direct comparison between chemopreventive bioactivity of individual (SEP) and pooled (VAT) glucosinolate samples revealed negative associations within GR*GN, GR*SNG, GN*SNG and GN*PG proportions, suggesting that the presence of other glucosinolates in the mixture had a
negative effect on chemopreventive activity in both SEP and VAT modes. At the level of
direct competition for myrosinase hydrolysis, three compound mixtures were consistently
significant in only the SEP models and more antagonistic in VAT combinations based on
CI values, suggesting the competition for digestion via myrosinase muted individual
contributions to bioactivity. In a previous study, the induction of phase 2 detoxification
enzymes at low levels of isothiocyanate administration was assumed to drive the
antagonistic relationship between GR and other glucosinolates; 45 while not confirming
this action, the findings support that there is some level of non-myrosinase competition-
derived antagonism that could be explained by detoxifying capabilities of other
glucosinolates.

The findings of the optimal glucosinolate profile were directly confirmed by
screening 150 isogenic lines from F_{2:3} broccoli mapping population 32 and identifying line
VB067 that contained the near-optimal aliphatic glucosinolate profile (77% GR, 14%
GN, and 9% others) (Figure 3.6A). When tested in HT-29 cells, VB067 showed a two-
fold increase in its antiproliferative properties as compared to commercial broccoli
cultivar VI-158 (Figure 3.6B). A typical aliphatic glucosinolate profile of commercial
broccoli contains 47% GR, 4% GN, 42% PG and 1% SNG 23. Ignoring bioavailability
and assuming 100% myrosinase-based hydrolysis of glucosinolates, the model predicted
that this makeup will exhibit between 9.9% and 27.7% cell growth inhibition in one raw
broccoli serving (36 g). Under similar assumptions, consuming a serving of VB067
broccoli line would result in over 90% cancer cell growth inhibition. Plasma
concentrations of sulphoraphane, a major enzymatic breakdown metabolite of
glucoraphanin, reach 15-20 µM level in humans 46,47 putting these findings in the range of
physiologically relevant bioactivity. However, one of the limitations of this study is not including other classes of glucosinolates into the analysis. Hydrolysis products of GB (indole-3-carbinol) and NGB also possess a degree of anticarcinogenic activity; but, these compounds have also been implicated with contributing to off flavors for Brassica vegetables. Their possible interactions with hydrolysis and bioactivity of aliphatic glucosinolates in broccoli warrant further investigation. Expanding the study to look at other factors for chemoprevention would provide a better mechanistic understanding of how these compounds are synergistic or antagonistic to one another.

3.6. Conclusion

In conclusion, the data described here provides strong support for the molecular breeding of broccoli containing increased concentration and an increased proportion of glucoraphanin and limiting the development of other glucosinolates that have an antagonistic relationship. Direct SEP and VAT hydrolysis models suggested antagonistic interaction between increased ratios of aliphatic glucosinolates at a standard level of myrosinase determine changes in chemopreventative activity. A bioactivity prediction model developed based on the three-point mixture analysis method allowed for detection and characterization of complex interactions between phytochemicals, even when their bioactivity is not sufficient to generate IC\textsubscript{50} values for the desired health promoting effects. Finding optimal levels of individual phytochemicals or classes of phytochemicals in target crops will be critical to produce acceptable new enhanced cultivars targeting promotion of human health.
3.7. Acknowledgements

This work was supported in part by Plant Pathways Elucidation Project (W.B., M.A.L., S.K.), PHHI seed grant 2015–2017 (S.K.), and NCSU faculty start-up funds (S.K.).
REFERENCES


CHAPTER 4. GENOME-WIDE ANALYSIS FOR TRAITS IN OAT (AVENA SATIVA) THAT PROMOTE WOUND HEALING
4.1. Abstract

Oats (*Avena sativa*) are seen as a key component of skin care both for their topical application and nutritional benefits following consumption. Bioactive components of oat are of major interest for the development of new herbal-based remedies for skin health and treatments for acute and chronic wounds. Diversity in cultivars has been found to modify the ability for plants to elicit physiological responses in human cell models, understanding the genotypic contribution of oats to skin health promotion can help guide future production efforts for medicinal and cosmetic purposes. 109 diverse oat varieties sourced from 5 different continents were grown in two locations, milled into flour, extracted, and screened with two *in vitro* models relevant to skin healing. Oat samples show different abilities to either promote or reduce production of inflammation and migration. RAW 264.7 macrophages inflamed with 1µg/mL of lipopolysaccharide treated with oat extracts were found to change nitric oxide radical production -36.0–40.0% and to change human dermal fibroblast cells (HDFa) movement into an exclusion zone following a simulated wound -34.0-65.0%. Ethanol insoluble fiber content correlated with reducing inflammatory response and increasing cell migration. A genome-wide association study (GWAS) was performed to identify quantitative trait loci’s (QTLs) on the oat genetic linkage map that significantly associate to either marker of wound healing. A total of identified a total of 22 QTLs were found in the analysis, 15 for inflammation response and 7 pertaining to migration. Two oat cultivars, UFRGS881971 and Boudrias, were identified as strong reducers of inflammation response and consistently aided faster cell migration respectively. Results indicate a differing ability for oat extracts to promote
wound healing, and identified targets for breeding programs interested in producing oats for skin health.

4.2. Introduction

Skin is the largest organ in the body; its main function is to serve as the protective barrier between the body and the outside environment. When the integrity of the skin is disrupted, it must go through a repair process so the tissue section can quickly regain its function. Most wounds heal within two weeks, but do not always regain the original structure and function of undamaged tissue. Skin repairing is divided into 4 main phases—hemostasis, inflammation, tissue formation, and remodeling—that overlap and work together to restore skin. During the inflammation phase, cellular and vascular responses clear a wound of damaged tissue, remove any foreign material, and work together to prevent bacterial colonization or infection from taking hold. These cells phagocytize bacteria, debride damaged tissue by releasing proteases, respond to chemotactic factors generated in the wound, and produce growth factors—PDGF, TGF-beta, IL’s, and Tumor Necrosis Factor. Once the area is clear of pathogens, it is essential that the wound progress out of the inflammatory phase, the actions of the inflammatory process negatively impact wound healing, which results in a chronic wound.

In the proliferative phase, new keratinocytes, melanocytes, fibroblasts, myofibroblasts and endothelial cells need to be produced and to migrate around the injury site. Fibroblasts need to proliferate and migrate to the wound area for the first 2-3 days, and once enough have accumulated they begin the production of collagen, fibronectin and glycosaminoglycan. Continuing the process, myofibroblasts grip wound edges and
pull them closer to one another, unnecessary cells begin to undergo apoptosis, and epithelialization occurs by epithelial cells proliferating and moving over the scaffold matrix from either end of the wound until the sides meet. The wound healing process culminates with a scar containing little fibrosis, a small amount of wound contraction, and renewed tissue function; however, further tissue remodeling can then continue for the next 2 years following initial healing.

Wounds can be chronic or acute; chronic wounds impact an estimated 6.5 million patients in the United States per year costing 15.3 billion dollars in annual expenditure on wound care products. Chronic wounds, referred to as ulcers, occur when the repair system is disrupted, resulting in injuries not healing after 3 months. Chronic wounds are especially damaging in diabetic populations where they often occur on feet. Type two diabetes has been increasing rapidly, the number of people with the disease quadrupled between 1980 and 2014, reaching a total of 422 million people. Foot ulcers can cause frequent hospitalizations due to their difficulty to treat, increased risk for infection and extended time required for healing. Foot ulcers precede the overwhelming majority of amputations attributed to diabetes complications. Diabetic patients suffering from foot ulcers see a sharp increase in their mortality rate, especially after amputation. The increasing incidence of diabetes adds to the need for effective wound healing treatments to increase quality of life for people suffering from these complications. In addition to the wound healing, there is interest in limiting excessive scarring following an injury which can lead to problems such as adverse aesthetics, loss of function, restriction of tissue movement, and adverse psychological effects. Accomplishing scar-free tissue regeneration has been a target of medical sciences, but has not been achievable to this
point for adult humans\textsuperscript{15}. New treatments for skin ailments have been utilized for increasing healing speed and limiting development of scar tissue\textsuperscript{16}.

While much progress has been made in both techniques and advances in understandings of growth factors, traditional plant therapies for wound healing remain under-examined\textsuperscript{17}. Plants are great sources of metabolites and bioactive properties that may be relevant to skin health and wound healing; in fact, many continue to be applied to induce skin healing\textsuperscript{18}. There remains a great diversity of plants that may have properties not fully understood\textsuperscript{2}, and an additional amount of variation is introduced by the variety of chemical profiles seen in plants grown across the globe\textsuperscript{19}. These plant treatments can target many aspects of wound healing and can promote overall skin health through stimulating or inhibiting specific growth factors and cytokines, aiding specific cell recruitment, preventing bacterial colonization, and modulating an excessive inflammatory response\textsuperscript{20}.

Over 22 million metric tons of oats (\textit{Avena sativa}) are produced worldwide each year to be used for food and animal feed, making them one of the more commonly available plants throughout the world\textsuperscript{21}. Being rich in vitamins, minerals, fibers, proteins, peptides, beta glucans, phenolic acids, and avenanthramides makes oats an excellent nutritional source for humans, as well as a great source for bioactive compounds\textsuperscript{22–24}. Additionally, oats have long been of interest for a connection to skin health\textsuperscript{25}. \textit{Colloidal oats} are finely milled oats suspended and evenly dispersed in water or a gel that have historically been used as a treatment for eczema, poison ivy, poison oak, and acne vulgaris\textsuperscript{26,27}. Oat extracts have been evaluated for their effect on increasing wound healing speed in animals, finding that ethanol extracted treatments significantly
aid the speed of wound closure. Oat bran extracts have been found to be protective components to reactive compound induced injury in fibroblasts as well as a protective from 320-370 ultraviolet A radiation. These effects were proposed to be derived from avenanthramides, phenolic compounds unique to certain cereal grains with known antioxidant properties.

Another component of oat of interest for skin health promotion are dietary fibers, β-glucans in particular. β-glucans are a unique compound prevalent in the oat plant, well known for providing health benefits in humans. Beta glucans are polysaccharides of D-glucose monomers linked by β-glycosidic bonds—β-(1 → 3) and β-(1 → 4) linkages. It is classified as a soluble dietary fiber, and it is one of the few food compounds that have received a full health claim status by the United States Food and Drug Administration in 1996 stating the efficacy for beta-glucans from oats reducing risk for coronary heart disease (CFR. 61: 296–313). For skin health, beta-glucans have been thought of as a source of oat moisturizing action; but they have also been used as a topical addition to aid wound healing. In mice, β-(1 → 3) glucan was has shown benefits on wound healing in diabetic mice, and the incorporation of beta-glucans into the water helped aid the healing process in carp following injury. Models for wound healing with human cells found oats to exhibit positive effects for increased wound healing speed. Mechanistic details for any oat-derived effects on wound healing are lacking and further complicated by the variability between different oat lines. One example being different antioxidant and anti-inflammatory effects different oat lines show in in vitro models. Differences have been seen in the TNF-alpha induced NF kappa-b expression in human kidney cells following 2mg/mL extract treatment and in the total antioxidant capacity of
the different extracts. Bioaccumulation of metabolites in plants are heavily influenced by the environment, processing conditions, and genetics of each plant.

Genetic association studies are used to identify areas of a plant genome significantly associated with the control of a particular phenotype. This method has long been used to associate the bioaccumulation of a compound known to have health benefits—such as b-glucan in oat—helping breeders select varieties to produce new health relevant varieties.

More recent examples have moved beyond this to find quantitative trait loci (QTL) directly connected to reducing disease risk in humans; one example being lowering glycemic index of rice. In this study a phenotypes for oat samples from worldwide diversity panel consisting of 109 accessions are being collected by two high-throughput screens modelling different stages of the wound healing process. A model of biological inflammation measuring changes in nitric oxide radical formation inhibition and the migration of human fibroblast cells into an exclusion zone of simulated injury. After screening, both wound-healing phenotypes, a GWAS was run using the newly constructed oat genetic linkage map, identifying loci that can be important for steps to understanding the medicinal potential of oats for human skin health and breeding for oat lines with greater health significance.
4.3. Methods and Materials

4.3.1. Reagents

All organic solvents were HPLC grade and obtained from VWR International (Suwanee, GA, USA).

4.3.2. Plant Material

Oat material was obtained from a world diversity panel, consisting of 109 accessions, assembled as part of the Collaborative Oat Research Enterprise (CORE) initiative. The world panel represents the genomic diversity among oat cultivars and includes lines from Asia, Europe, Australia, North America and South America. The world panel was grown and harvested in two locations, Lacombe, Alberta, Canada (CLC) and Aberdeen, Idaho, USA (IDAC). Accessions were planted in April or May of 2010 in 4.6 m² plots in CLC, while 1.2 to 2.2 m four- or five-row plots were used in IDAC. A database containing phenotypic data for all accessions is stored in the public T3/oat database (http://triticeatoolbox.org/oat) that outlines the content of beta glucan, alpha tocopherols, gamma tocopherol avenanthramide a, d, and e, plant height, heading date, grain yield, ethanol soluble fiber, and ethanol insoluble fiber for all oat lines. Oat seeds were de-hulled and uniformly milled into whole grain flour by General Mills in Minneapolis, MN.
4.3.3. Extraction

All oat samples were initially defatted by taking 50mg of oat flour and adding 1mL hexane for 1-hour vortexing at room temperature. Sample was centrifuged at 5000g for five minutes and the solvent was removed and discarded and the residual material was dried overnight. Defatted sample was then extracted by adding 1mL of 70% aqueous ethanol and vortexing for two hours. Sample was then centrifuged at 5000g for five minutes and the supernatant removed. The solvent was transferred to pre-weighed 1.5mL centrifuge tubes and removed by four hours of vacuum concentration using a SPD1010 SpeedVac™, Thermo Scientific Waltham, MA, under low heat setting. Dry sample was adjusted to 30mg/mL with dimethylsulfoxide (DMSO) for the inflammatory assay and 10mg/mL with 80% ethanol to be used for the migration cell assay.

4.3.4. Cell lines

Primary human dermal fibroblasts isolated from adult skin (HDFa, Invitrogen C-013-5C) were cultured in Medium 106 (Invitrogen M-106-500) with Low Serum Growth
Supplement (LSGS, Invitrogen S-003-10) supplemented with Antibiotics Penicillin/Streptomycin Solution 100 IU/100 µg/mL (Fisher MT-30-002-CI). RAW 264.7 Murine Macrophages from American Type Culture Collection (ATCC), Manasses, VA, were maintained using Dulbecco’s Modified Eagle Medium (DMEM) with with Low Serum Growth Supplement (LSGS, Invitrogen S-003-10) supplemented with Antibiotics Penicillin/Streptomycin Solution 100 IU/100 µg/mL (Fisher MT-30-002-CI). The cells were maintained at 37 °C in a humidified atmosphere containing 5% CO₂.

4.3.5. Cell viability and dose range determination

The cytotoxic activity against fibroblast cells and macrophages was evaluated using an MTT ([3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide] colorimetric assay essentially as described and quantified spectrophotometrically at 550 nm using a microplate reader SynergyH1 (BioTek). Percent of inhibition of each cell type was calculated against the vehicle in decreasing concentration (two-fold serial dilutions) of the test samples. The concentrations of test reagents that showed no changes in cell viability compared with that vehicle (ethanol) were selected for further studies.

4.3.6. In vitro inflammatory stimulation

Raw 264.7 macrophage cells were seeded at approximately 1.0x10⁴ per well in a sterile 96 well plate and allowed to fix to plates overnight. Cells were exposed to 100 µL of sample extract mixed with DMEM (100 µg/mL final concentration) and then 100 µL of DMEM containing lipopolysaccharide (LPS, from Escherichia coli 026:B6) and incubated for 24 h. The antioxidant dexamethasone (DEX) was used as a positive control at 10 µM in place of oat samples and a vehicle control consisting of DMSO (0.33%)
alone were run alongside each sample. All sample treatments were conducted in triplicate.

4.3.7. **In vitro nitric oxide reduction quantification**

The ability of test samples to inhibit nitric oxide radical formation was determined in RAW 264.7 cells using a colorimetric assay according to a protocol previously reported (Grace et al., 2016). Initially 100 µL of cell culture medium was removed and combined with to 100 µL of Griess reagent (Promega, Madison, WI)—first 50µL 1% sulfanilamide and second 50µL 0.1% naphthylethylenediamine in 5% phosphoric acid incubated protected from light for 10 minutes at room temperature following the addition of each chemical. The absorbance at 520 nm was read (Synergy H1, Biotech, Winooski, VT) and a calibration curve built with serial dilutions of sodium nitrite (R² = 0.99) was used to calculate µM of nitric oxide concentration. Results were expressed as % NO change compared to the vehicle control.

4.3.8. **In vitro wound healing assay**

Ability of samples to influence cell migration was evaluated on a 2-D assay of adherent HDFa cell lines using Oris™ 96-well tissue culture treated plates (AMS Biotechnology, Cambridge, MA)⁴⁵. For this, 100 µL of human HDFa fibroblasts previously treated with fluorescent NucBlue® Live Cell Stain were seeded in 96-well plates (max. 2×10⁵ cells/well) containing centrally positioned stoppers. The plates were incubated for 24 h at 37 °C and 5% CO₂. Thereafter, the stoppers were removed and 1 µL of test samples (to a final concentration of 50 µg/mL) were added to each well. The progress of cell migration was monitored after 0, 4, 8, and 22 hours after sample addition by measuring the fluorescence at 360 nm (excitation) and 460 nm (emission) on a
microplate reader (Synergy H1, Biotech, Winooski, VT). Positive (cells treated with 10 µL of FBS) and negative controls (no cells and only cells and growth media) were included in every experiment.

4.3.9. Fluorescence image analysis

Bright field and fluorescent images were observed using EVOS® FL Auto Cell Imaging System (Life Technologies). Images were captured at the center of each treatment. Three representative images of the areas from each well under each condition were photographed at 0 and 24 h to estimate the wound closure. The images were analyzed using EVOS software and % wound closure was calculated relative to vehicle control. Area of the space where the insert template was printed was measured at initial and 24-hour time periods. The percent difference of wound closure was found between the two time points. The inverse determines the percentage of wound space enclosed by cell movement.

4.3.10. Statistical analysis

Statistics were performed using the software GraphPad Prism v6 (GraphPad Software Inc., La Jolla, CA). All data were analyzed by one-way ANOVA with drying technique as a factor. Post hoc analyses were conducted using the Dunnett’s multiple comparison tests at 5% level of significance. All samples were obtained in triplicate and analyzed in three repetitions (N=9), unless specified. All results are expressed as means ±SEM.

4.3.11. Genome wide association study

Linkage between all marker pairs within each linkage group was calculated as the squared difference between observed and expected haplotype frequencies divided by the
product of the major and minor allele frequencies of both markers ($r^2$). Each measure of
LD (y-axis) was plotted by linkage distance (cM, x-axis) and average linkage decay
across all linkage groups for each dataset was determined by fitting a spline to all the data
and setting a linkage cutoff of $r^2 = 0.2$. Population structure was determined using a
principal component analysis (PCA) and no kinship analysis was used since this
population was of diverse lineage. The top 10 principal components were determined
for the diversity panel with JMP Genomics (SAS, Cary, NC). Analysis of scree plots of
eigenvalues was used to determine how many principal components were used to account
for population structure in each GWAS analysis. Multivariate analysis of the
relationship matrix principal components and the marker data principal components
determined the two measures of population structure were correlated. The genome-wide
association study was performed using the JMP Genomics (SAS, Cary NC) PCA for
population stratification procedure which is based on the Eigenstrat method. This
method uses the top principal components of the PCA analysis to control for population
structure.

The number of principal components to use for each panel was determined using
eigenvalue scree plots. Independent association studies were performed using the
predetermined number of principal components for the diversity panel from individual
locations the plants were grown and the average of the two locations. $P$-values were
converted to $-\log_{10}$ values where values $\geq 2.5$ were considered significantly associated
markers. QTLs were designated as having one or more linked markers that were
significant in any of the independent association studies: either location, or the average of
both locations. QTLs in close proximity (≤ 4 cM) to each other were considered to be within linkage disequilibrium and combined into a single QTL.

4.4. Results

4.4.1. Phenotyping

In total, flour samples from 109 unique oat varieties were extracted by 80% aqueous ethanol and applied as treatments in an in vitro model for two key aspects of skin healing, inflammation response modulation and migration of fibroblasts into a simulated wound. These oat lines were primarily derived from breeding programs in the United States and Europe, but also came from Asia, South America, and Australia (Figure 4.1). Cytotoxicity determination (Figure 2A) outlined no toxic effects of oat extracts up to 200µg/mL in RAW macrophages, but saw growth inhibiting effects for HDFa cells start at 200µg/mL. For subsequent treatments to screen the population a single dose of 100µg/mL was applied to the RAW cells and 50µg/mL was used to treat the HDFa cells.

Results from both screens found the collective populations of oats were unable to induce a consistent response in the cells. For inflammation (production of nitrite over a vehicle control) the average for the oat population was -0.24% ± 3.47% and for migration (fluorescence intensity increase of labeled cells into a simulated wound area) was 1.688% ± 3.324%. However, the different cultivars had a wide range in activity -36.0–40.0% for inflammation and -34.0–65.0% for migration. For each location a similar spread in responses were seen (Figure 4.2B), but low correlation was found between the locations for inflammation r=0.25 (p=0.03) and r=0.37 (p<0.01) for migration, suggesting a large environmental component driving the responses.
Figure 4.2: Viability and population inflammatory and migration bioactivity spread. (A) Cell viability for RAW macrophages (circles) and HDFa cells (squares) following exposure to oat ethanol extracts for 24 hours. Significant differences are calculated vs. vehicle control using Dunnet's test for multiple comparisons $p < 0.05$. (B) Population scores for nitric oxide radical percent change from control following 24-hour treatment for CLC (upright triangles) and IDAC (downward triangles). Population scores for percent migration change of HDFa cells over 24 hours at CLC (circles) and IDAC (squares).
Table 4.1: Variation in phenotypic data for population gathered from the public T3/oat database for all agronomic and metabolite traits. ([http://triticeaetoolbox.org/oat](http://triticeaetoolbox.org/oat))

<table>
<thead>
<tr>
<th>Mean &amp; 95% Confidence Interval</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heading Date (julian date)</td>
<td>190.5 ± 0.90</td>
</tr>
<tr>
<td>Plant Height (cm)</td>
<td>104.7 ± 3.00</td>
</tr>
<tr>
<td>Grain Yield (g m-2)</td>
<td>520.8 ± 17.69</td>
</tr>
<tr>
<td>Avenanthramide 2c (ug g-1)</td>
<td>9.412 ± 1.063</td>
</tr>
<tr>
<td>Avenanthramide 2f (ug g-1)</td>
<td>6.626 ± 0.766</td>
</tr>
<tr>
<td>Avenanthramide 2p (ug g-1)</td>
<td>8.677 ± 1.132</td>
</tr>
<tr>
<td>Avenanthramide 5p (ug g-1)</td>
<td>1.570 ± 0.191</td>
</tr>
<tr>
<td>Total Avenanthramide (ug g-1)</td>
<td>26.28 ± 3.010</td>
</tr>
<tr>
<td>Tocopherol Alpha (ug g-1)</td>
<td>8.098 ± 0.304</td>
</tr>
<tr>
<td>Tocotrienol Alpha (ug g-1)</td>
<td>23.04 ± 0.809</td>
</tr>
<tr>
<td>Tocopherol Beta (ug g-1)</td>
<td>1.087 ± 0.072</td>
</tr>
<tr>
<td>Tocotrienol Beta (ug g-1)</td>
<td>2.470 ± 0.165</td>
</tr>
<tr>
<td>Tocotrienol Delta (ug g-1)</td>
<td>0.163 ± 0.020</td>
</tr>
<tr>
<td>Tocotrienol Gamma (ug g-1)</td>
<td>0.217 ± 0.022</td>
</tr>
<tr>
<td>Total Tocopherol (ug g-1)</td>
<td>35.07 ± 1.070</td>
</tr>
<tr>
<td>Beta Glucan (% DW)</td>
<td>4.528 ± 0.099</td>
</tr>
<tr>
<td>Ethanol-insoluble Dietary Fiber (% DW)</td>
<td>4.777 ± 0.130</td>
</tr>
<tr>
<td>Ethanol-soluble Dietary Fiber (% DW)</td>
<td>5.898 ± 0.242</td>
</tr>
<tr>
<td>Dietary Fiber Total (% DW)</td>
<td>11.23 ± 0.320</td>
</tr>
<tr>
<td>Inflammation (% Change)</td>
<td>-0.24 ± 3.47</td>
</tr>
<tr>
<td>Migration (% Change)</td>
<td>1.688 ± 3.324</td>
</tr>
</tbody>
</table>

Additional agronomic traits and metabolite information for the population in the study was summarized from the public T3/oat database (Table 4.1). Significant correlations (p<0.05) between the different traits were identified with a correlation matrix (Table 4.2). Of the known agronomic traits, ethanol insoluble fiber showcased a significant association with inflammation (r = -0.47) and migration activity (r = 0.39), suggesting a benefit to increasing this trait in plants for both reducing excess nitrite production and inducing the speed of migration for dermal fibroblasts. A small association was found between plant height and inflammation (r=0.23), perhaps indicating a slight decrease in ability for extracts from taller oat samples to mitigate anti-
inflammatory activity. No significant association was seen for any of the metabolites in the oat population to either of the markers for wound healing.

Nitrite production for two oat varieties showing consistent responses for inflammation reduction or increased inflammation production are outlined (Figure 4.3A). The oat line Prescott, derived from the Ottawa Research and Development Centre for Agriculture and Agri-food (Ottawa, Canada), is known for having a high yield, good milling qualities, an average metabolite concentration, and having an excellent resistance to current strains of crown rust. The oat line UFRGS881971 is from U Federal do Rio Grande do Sul (UFRGS), Brazil and has average to low growth yield and metabolite concentrations. For both of these lines, the total fiber statistics were unavailable.

**Figure 4.3:** Nitric oxide radical production and migration activity induced by significant oat cultivars. (A) Nitric oxide radical (nitrite) produced in cell medium by RAW 264.7 macrophages stimulated with 1 µg/mL lipopolysaccharide over 24 hours and treated with oat extracts or controls. Plain circle is the vehicle control of DMSO, , squares are a positive control of 10µM anti-inflammatory dexamethasone (DEX), triangles are for oat variety Prescott, and grey outlined circles are oat variety UFRGS881971. (B) Percent of Human dermal fibroblast cell migration into an exclusion zone from 0-48 hours. Plain circle is the vehicle control of 80% aqueous ethanol, squares are a positive control of 10% fetal bovine serum (FBS), triangles are for oat variety Boudrias, and grey outlined circles are oat variety Shadow. * Denotes significant differences from vehicle treatment Dunnett’s multiple comparison test.

Visual migration changes were captured using fluorescent microscopy over the course of 48 hours with time-points at 0, 12, 24, 36, and 48 hours. The four samples outlined in (Figure 4.4) are a 80% aqueous ethanol vehicle, a 10% fetal bovine serum positive control and two oat varieties—a high activity oat variety Boudrias (50µg/mL)
and (D) low activity oat variety (50μg/mL) Shadow. Visually, the vehicle control
treatment shows that very little cell movement takes place but when the cells are treated
with the positive control they begin to move in mass into the empty region by 24 hours
continuing through 48 hours until they populate all areas of the zone. The Boudrias
sample resembles the positive control much more than the Shadow treatment. Fluorescent
analysis was paired with this analysis (Figure 3.3B) finding that both the FBS and
Boudrias increased cell migration significantly, 45% ± 28.7% and 31% ± 6.4%
respectively. Boudrias oat is a popular hulless oat variety with good yielding properties
and average metabolite contents; it also displays a good amount of resistance to crown
rust and yellow dwarf virus 49.
### Table 4.2: Correlation matrix between phenotypes

<table>
<thead>
<tr>
<th></th>
<th>HD</th>
<th>PH</th>
<th>GY</th>
<th>AV2C</th>
<th>AV2F</th>
<th>AV2P</th>
<th>AV5P</th>
<th>aTo</th>
<th>aTe</th>
<th>bTo</th>
<th>bTe</th>
<th>BG</th>
<th>dTo</th>
<th>dTe</th>
<th>gTe</th>
<th>IDF</th>
<th>sDF</th>
<th>TAV</th>
<th>TDF</th>
<th>TT</th>
<th>MIG</th>
<th>INF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heading Date (julian date)</strong></td>
<td><strong>0.38</strong></td>
<td></td>
<td></td>
<td>-0.09</td>
<td>0.07</td>
<td>-0.11</td>
<td>-0.01</td>
<td>-0.08</td>
<td>0.03</td>
<td>0.02</td>
<td>0.10</td>
<td>0.09</td>
<td><strong>0.41</strong></td>
<td>-0.08</td>
<td>-0.07</td>
<td>0.09</td>
<td>-0.24</td>
<td>-0.27</td>
<td>-0.01</td>
<td>-0.38</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Plant Height (cm)</strong></td>
<td>0.14</td>
<td>-0.19</td>
<td>-0.25</td>
<td>-0.22</td>
<td>-0.20</td>
<td>-0.04</td>
<td>-0.08</td>
<td>-0.07</td>
<td><strong>0.31</strong></td>
<td>0.10</td>
<td>0.03</td>
<td>0.08</td>
<td>-0.38</td>
<td>-0.34</td>
<td>-0.22</td>
<td>-0.44</td>
<td>0.01</td>
<td>-0.14</td>
<td><strong>0.23</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grain Yield (g m-2)</strong></td>
<td>-0.15</td>
<td>-0.08</td>
<td>-0.09</td>
<td>0.02</td>
<td>0.19</td>
<td>-0.12</td>
<td>-0.20</td>
<td>-0.10</td>
<td>-0.02</td>
<td>0.18</td>
<td>0.13</td>
<td>-0.03</td>
<td>-0.35</td>
<td>-0.36</td>
<td>-0.10</td>
<td>-0.27</td>
<td>-0.07</td>
<td>-0.10</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Avenanthramide 2c (ug g-1)</strong></td>
<td><strong>0.81</strong></td>
<td>0.93</td>
<td><strong>0.85</strong></td>
<td><strong>0.45</strong></td>
<td>0.02</td>
<td>-0.01</td>
<td>0.08</td>
<td>-0.02</td>
<td>-0.07</td>
<td>-0.07</td>
<td>0.10</td>
<td>0.16</td>
<td>0.18</td>
<td><strong>0.96</strong></td>
<td>0.13</td>
<td>-0.10</td>
<td>-0.03</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Avenanthramide 2f (ug g-1)</strong></td>
<td>0.86</td>
<td><strong>0.89</strong></td>
<td><strong>0.32</strong></td>
<td>0.00</td>
<td>-0.04</td>
<td>0.11</td>
<td>0.07</td>
<td>-0.07</td>
<td>-0.11</td>
<td>0.06</td>
<td>0.12</td>
<td>0.05</td>
<td><strong>0.92</strong></td>
<td>0.05</td>
<td>-0.08</td>
<td>-0.09</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Avenanthramide 2p (ug g-1)</strong></td>
<td>0.88</td>
<td><strong>0.40</strong></td>
<td></td>
<td>0.09</td>
<td>0.01</td>
<td>0.18</td>
<td>-0.02</td>
<td>-0.08</td>
<td>-0.06</td>
<td>0.07</td>
<td>0.17</td>
<td>0.14</td>
<td><strong>0.98</strong></td>
<td>0.12</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Avenanthramide 5p (ug g-1)</strong></td>
<td>-0.33</td>
<td></td>
<td></td>
<td>0.01</td>
<td>-0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>-0.06</td>
<td>-0.12</td>
<td>0.05</td>
<td>0.07</td>
<td>0.22</td>
<td><strong>0.92</strong></td>
<td>0.15</td>
<td>-0.09</td>
<td>-0.12</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocopherol Alpha (ug g-1)</strong></td>
<td>0.20</td>
<td>0.32</td>
<td></td>
<td>0.10</td>
<td>-0.04</td>
<td>-0.04</td>
<td>0.06</td>
<td>-0.25</td>
<td>-0.07</td>
<td>-0.30</td>
<td>-0.41</td>
<td>-0.17</td>
<td><strong>0.48</strong></td>
<td>-0.11</td>
<td>-0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocotrienol Alpha (ug g-1)</strong></td>
<td>0.31</td>
<td>0.75</td>
<td></td>
<td>0.09</td>
<td>-0.02</td>
<td>-0.13</td>
<td><strong>0.28</strong></td>
<td>0.15</td>
<td>0.07</td>
<td>0.04</td>
<td>0.07</td>
<td><strong>0.95</strong></td>
<td>0.01</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocopherol Beta (ug g-1)</strong></td>
<td>0.40</td>
<td>-0.22</td>
<td></td>
<td>-0.08</td>
<td>-0.02</td>
<td><strong>0.24</strong></td>
<td>0.08</td>
<td>-0.15</td>
<td>-0.01</td>
<td>-0.11</td>
<td><strong>0.45</strong></td>
<td>-0.03</td>
<td>-0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocotrienol Beta (ug g-1)</strong></td>
<td>-0.02</td>
<td>-0.01</td>
<td></td>
<td>-0.07</td>
<td><strong>0.22</strong></td>
<td>0.14</td>
<td>0.22</td>
<td>0.13</td>
<td>0.22</td>
<td><strong>0.78</strong></td>
<td>0.09</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Beta Glucan Dwb (%)</strong></td>
<td>-0.05</td>
<td>-0.13</td>
<td></td>
<td>-0.10</td>
<td><strong>0.51</strong></td>
<td><strong>0.89</strong></td>
<td>0.01</td>
<td><strong>0.91</strong></td>
<td>0.03</td>
<td>-0.03</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocopherol Delta (ug g-1)</strong></td>
<td>0.10</td>
<td>-0.07</td>
<td></td>
<td>-0.13</td>
<td>0.16</td>
<td>0.07</td>
<td>-0.07</td>
<td>0.03</td>
<td>-0.04</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocotrienol Delta (ug g-1)</strong></td>
<td>0.03</td>
<td>-0.30</td>
<td></td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.10</td>
<td>-0.08</td>
<td>-0.02</td>
<td>-0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocopherol Gamma (ug g-1)</strong></td>
<td>-0.28</td>
<td>0.05</td>
<td></td>
<td>0.08</td>
<td>0.03</td>
<td><strong>0.36</strong></td>
<td>0.03</td>
<td>-0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dietary Fiber EtOH-insoluble (%)</strong></td>
<td>0.34</td>
<td>0.16</td>
<td></td>
<td><strong>0.58</strong></td>
<td>0.13</td>
<td><strong>0.39</strong></td>
<td><strong>-0.47</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dietary Fiber EtOH-soluble (%)</strong></td>
<td>0.14</td>
<td><strong>0.90</strong></td>
<td></td>
<td>0.02</td>
<td>-0.19</td>
<td>-0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Avenanthramide Total (ug g-1)</strong></td>
<td>0.11</td>
<td>-0.06</td>
<td></td>
<td>-0.05</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dietary Fiber Total (%)</strong></td>
<td>0.05</td>
<td>-0.16</td>
<td></td>
<td>-0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tocol Total (ug g-1)</strong></td>
<td>-0.01</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inflammation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* is significant (p<0.05)  ** is significant (p<0.01)
4.4.2. Genome wide association study

The genome wide association study found a total of 22 significant loci for wound healing traits. Of these traits, 15 were for inflammation and 7 were for migration (Table 4.3). All traits $R^2$ for estimated trends were between 0.10-0.16. Trends were found to exhibit both beneficial and negative effects on either phenotype. Multiple markers were found to be significant for 4 INF QTL (QTL-1, 10, 16, and 20) but no QTL were shared in both locations for a single trait. The highest negative log10 for estimated trend of any locus was 3.45 for QTL-21 on merge group 23 shown to contribute approximately 16% of phenotypic control for INF grown in the CLC region. The marker at this location was avgbs_2344, the presence of this genotype had an estimated trend of increasing inflammation 10%. The most significant locus for a MIG controlling region was QTL-5, located on merge group 6 at 63.2cM, found with the CLC grown population with a log10 for estimated trend of 2.86 controlling 13% of the phenotypic variation in positive fashion (estimated trend = 7%). The marker associated to QTL-5 is avgbs_61371. QTL-8 and QTL-9 for IDAC INF and CLC MIG co-localized on merge group 9 with it having a positive association for reducing inflammation response but decreasing migration speed increases. QTLs from both growth locations are showcased in (Figure 4.5).
Table 4.3: Significant loci (Negative Log10 for effect over 2.5) for inflammatory and migration effects

<table>
<thead>
<tr>
<th>LG</th>
<th>Trait</th>
<th>Region</th>
<th>Position (cM)</th>
<th>ET^A</th>
<th>SE^B</th>
<th>Neg Log10</th>
<th>R^2</th>
<th>Significant Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>QTL-1</td>
<td>Mrg01</td>
<td>INF</td>
<td>CLC</td>
<td>105.9–109</td>
<td>-13%</td>
<td>4%</td>
<td>2.52</td>
<td>12%</td>
</tr>
<tr>
<td>QTL-2</td>
<td>Mrg01</td>
<td>MIG</td>
<td>CLC</td>
<td>57.2</td>
<td>13%</td>
<td>4%</td>
<td>2.53</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-3</td>
<td>Mrg05</td>
<td>INF</td>
<td>IDAC</td>
<td>130.9</td>
<td>10%</td>
<td>3%</td>
<td>2.89</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-4</td>
<td>Mrg06</td>
<td>INF</td>
<td>IDAC</td>
<td>28.5</td>
<td>7%</td>
<td>2%</td>
<td>2.86</td>
<td>13%</td>
</tr>
<tr>
<td>QTL-5</td>
<td>Mrg06</td>
<td>MIG</td>
<td>CLC</td>
<td>63.2</td>
<td>7%</td>
<td>2%</td>
<td>2.86</td>
<td>13%</td>
</tr>
<tr>
<td>QTL-6</td>
<td>Mrg08</td>
<td>MIG</td>
<td>CLC</td>
<td>105.5</td>
<td>-6%</td>
<td>2%</td>
<td>2.58</td>
<td>10%</td>
</tr>
<tr>
<td>QTL-7</td>
<td>Mrg08</td>
<td>INF</td>
<td>CLC</td>
<td>151.9</td>
<td>9%</td>
<td>3%</td>
<td>2.83</td>
<td>13%</td>
</tr>
<tr>
<td>QTL-8</td>
<td>Mrg09</td>
<td>INF</td>
<td>IDAC</td>
<td>50.9</td>
<td>-13%</td>
<td>4%</td>
<td>2.67</td>
<td>10%</td>
</tr>
<tr>
<td>QTL-9</td>
<td>Mrg09</td>
<td>MIG</td>
<td>CLC</td>
<td>51.2</td>
<td>-8%</td>
<td>3%</td>
<td>2.67</td>
<td>12%</td>
</tr>
<tr>
<td>QTL-10</td>
<td>Mrg09</td>
<td>INF</td>
<td>IDAC</td>
<td>71.3–73.5</td>
<td>14%</td>
<td>4%</td>
<td>3.01</td>
<td>12%</td>
</tr>
<tr>
<td>QTL-11</td>
<td>Mrg11</td>
<td>INF</td>
<td>IDAC</td>
<td>100.9</td>
<td>-13%</td>
<td>4%</td>
<td>2.82</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-12</td>
<td>Mrg12</td>
<td>MIG</td>
<td>IDAC</td>
<td>67.8</td>
<td>-6%</td>
<td>2%</td>
<td>2.57</td>
<td>10%</td>
</tr>
<tr>
<td>QTL-13</td>
<td>Mrg13</td>
<td>INF</td>
<td>IDAC</td>
<td>58.6</td>
<td>-8%</td>
<td>2%</td>
<td>2.84</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-14</td>
<td>Mrg15</td>
<td>INF</td>
<td>CLC</td>
<td>79.7</td>
<td>9%</td>
<td>3%</td>
<td>2.5</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-15</td>
<td>Mrg17</td>
<td>INF</td>
<td>CLC</td>
<td>13.7</td>
<td>-16%</td>
<td>5%</td>
<td>2.65</td>
<td>12%</td>
</tr>
<tr>
<td>QTL-16</td>
<td>Mrg17</td>
<td>INF</td>
<td>IDAC</td>
<td>83.9–87.8</td>
<td>-10%</td>
<td>3%</td>
<td>3.43</td>
<td>14%</td>
</tr>
<tr>
<td>QTL-17</td>
<td>Mrg17</td>
<td>MIG</td>
<td>CLC</td>
<td>104.8</td>
<td>8%</td>
<td>3%</td>
<td>2.54</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-18</td>
<td>Mrg18</td>
<td>MIG</td>
<td>CLC</td>
<td>104.8</td>
<td>8%</td>
<td>3%</td>
<td>2.54</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-19</td>
<td>Mrg19</td>
<td>INF</td>
<td>IDAC</td>
<td>45.4</td>
<td>13%</td>
<td>3%</td>
<td>3.45</td>
<td>14%</td>
</tr>
<tr>
<td>QTL-20</td>
<td>Mrg19</td>
<td>INF</td>
<td>IDAC</td>
<td>49.8–50</td>
<td>11%</td>
<td>3%</td>
<td>2.75</td>
<td>11%</td>
</tr>
<tr>
<td>QTL-21</td>
<td>Mrg23</td>
<td>INF</td>
<td>CLC</td>
<td>23.2</td>
<td>10%</td>
<td>3%</td>
<td>3.45</td>
<td>16%</td>
</tr>
<tr>
<td>QTL-22</td>
<td>Mrg24</td>
<td>INF</td>
<td>CLC</td>
<td>11.3</td>
<td>7%</td>
<td>2%</td>
<td>2.51</td>
<td>11%</td>
</tr>
</tbody>
</table>
Figure 4.4: Time lapse of fluorescent dyed human dermal fibroblast migration into exclusion zone over 48 hours following treatment. (A) vehicle control (0.5%) 80% aqueous ethanol (B) fetal bovine serum (10%) positive control (C) high activity oat variety boudrias (50µg/mL) and (D) low activity oat variety (50µg/mL) shadow
Figure 4.5: Oat QTL Results. Genetic linkage map for oats with marker QTL for migration (CLC = Circle, IDAC = diamonds) and inflammation (CLC = plus signs, IDAC = x’s). Position is in cM and Mrg groups are chromosome representations. *** denotes neg log > 2.5 (blue) **** denotes neg log > 3.0 (red)
4.5. Discussion

This is the first attempt to use a genome wide association study to connect biological inflammatory response effects and migration speed induction changes to a plant genetic linkage map. To date, the largest screening of the effect that oat variety extracts have on inflammatory responses has been limited to seven oat genotypes; the current study screened 109 unique oat varieties. This is also the first time an oat population was measured for ability to induce fibroblast migration changes. By using ORIS technology, migration was assessed as fluorescent intensity of labeled cells moving into a defined exclusion zone, allowing the assessment to be done in a reproducible manor for this large sample-set. Completion of the two screens showed that oat lines generally did not change nitric oxide radical production in an inflamed macrophage cell or increase fibroblast migration into a simulated injury area. However, some lines stood out as significant inducers of either trait from both growth locations. Two specific oat lines were identified from this screen to have substantial beneficial effects for the two phenotypes relevant to wound healing. South American oat variety UFRGS881971 lowered nitrite production from inflamed macrophage cells and the Canadian oats line Boudrias significantly increased migration speed of human dermal fibroblasts. These lines were two of the only lines that showed similar high beneficial effects when grown in both CLC and IDAC locations.

Metabolite correlations from the oat core database found only plant height and ethanol insoluble fiber to have significant correlations to migration and inflammation changes. An additional correlation was seen between plant height and increasing production of nitric oxide radical formation from RAW cells. Increased phenotyping to
include more specific metabolite and macronutrient data would be a good way to refine the gene pathways and metabolites relevant to stronger wound healing potential for oats going forward. Mechanistic data for inflammation response changes and wound healing relevant growth factor expression changes following oat treatments are currently underway.

The fact that few oat lines exhibited consistent responses in both locations suggests that there was a large amount of environmental control to INF and MIG changes. The completion of this GWAS found 22 total QTL, 15 that pertained to inflammation response modulation, and 7 for migration speed. None of the QTL found were conserved between the two growing locations; this is not surprising based on the low correlation observed between the two traits. The most significant marker had an $r^2$ of 0.16, corresponding to a relatively low percent explained.

Oat is also a relatively complicated plant to analyze for bioactivity because many oat phenolic compounds are bound within the plant, making them inaccessible to simple solvent extraction. Oat also possesses a large genome with 3 different ancestral genomes increasing the difficulty of any kind of genetic study. Of the defined merge groups representing chromosomes, there are 3 closely related genomes (A, C and D) that closely resemble one another and can have the overlapping functions. Pinpointing underlying candidate genes from QTL data will be difficult in oats until there is a functional genome sequence; currently there is a major sequencing project being undertaken. As sequence homology between the chromosomes is better outlined, there may be QTL identified in this study that are in regions on two chromosomes that mirror one another.
With the absence of conserved loci, none of these markers can be deemed relevant to build a breeding population around. To get a better understanding of the regions that are controlling the differences between oat varieties future research can utilize either the UFGRS881971 or Boudrias varieties to create a bi-parental populations able to better pinpoint the genetic control for skin healing relevance. Another option would be to increase the size of the population used to conduct the GWAS to increase statistical power for identifying QTL from plants grown in different regions.

Going forward, health relevant gene association studies should focus on using models with a low amount of variability between replicates and on traits that have been well established as being connected to the plants of interest. However, if the phenotypes are simple to phenotype then this method would be a valuable screen to identify interesting cultivar differences and potentially identify map based regions to direct breeding programs.

4.6. Conclusion

Given the widespread use of oats in cosmetic skin health treatments, it was worthwhile to screen potentiating factors for skin healing. Overall, oat cultivars seem to showcase different bioactive properties. A phenotypic screen paired with metabolic data correlations and a GWAS proved to be a good approach for analyzing oats further; however, the studies used to screen this population for inflammatory and migration effects were not reproducible across regions. The identification of oat cultivars with significant bioactivity can be used to analyze wound healing potential going forward.
4.7. Acknowledgments

Thank you to the Plant Pathways Elucidation Program and the General Mills and the Oat Core Initiative Program for providing the oat lines used for this study.
REFERENCES


CHAPTER 5. CONCLUSION
5.1. Conclusion

Plants we eat serve as a metabolite delivery medium for our bodies. Understanding the mechanisms that make plant compounds essential for growth, daily function, and prevention of diseases has helped our species thrive. Still, there is much area for improvement. Deadly diet-xrelated chronic diseases afflict a large number of people, and many impoverished areas still see diseases arise from inadequate nutrition. My previous work—appendix 1—helped me learn about bioactive plant components. This piece summarizes the connection between flavonoids (a common plant phytochemical) and cardiovascular disease. Two consistent themes emerged: (1) plants have lots of different compounds that may reduce disease risk, but it is difficult to find consistencies because of the variability from plant to plant; and (2) single compounds do not tell the whole story—the entire plant metabolite profile is important.

Three factors control the metabolites and impending health benefits of the plants we ingest: genetics, the external environment, and postharvest processing. By optimizing the plant genetics, we can work toward giving consumers the greatest potential to consume healthier foods. To better understand this field, I helped create a guide for novices to understand “omics” oriented fields—appendix 2. This piece reviews genomics, bioinformatics, and proteomics through a case study of the of blueberry genomic sequencing project. Understanding these disciplines was important for me, a nutrition scientist, to help usher in the next generation of health promoting plants.

The agricultural sector can help in the fight against diseases by producing plants tailored to health-promotion. Many studies have been able to associate a phenotype relevant for human health to different genomic loci, as presented in chapter one.
Generating QTL data is important for understanding how plants synthesize bioactive compounds. These regions can be used to guide targeted breeding approaches. Most studies in this area outline a concentration of a single compound as the phenotype, that elsewhere has been linked to health promotion. This ignores the complex nature of plants and humans, where many compounds can interact with one another to become more or less active, become bio-converted further in human metabolism, or confer activity in a way not presently understood. This work outlined how we can use in vitro models for human disease risk to screen plant mapping populations and associate them to the respective plant genomes. Identifying QTLs with direct biological relevance encapsulates more of the complex path between plant genes and human health benefits.

With the first study—chapter 2—this idea was put to the test. Whole plant extracts from a mapping population of broccoli were screened for the ability to slow the growth of cancerous human colon cells. This was chosen because chemopreventative properties of broccoli have been shown in many epidemiology studies. By comparing this trait to the underlying genotypes of each plant with a QTL study, regions of the broccoli genome that significantly associate to chemopreventative properties were uncovered. Different aliphatic glucosinolates were found to contribute to the bioactivity seen both by associating cell growth inhibiting properties to various metabolites and from investigating the underlying genes in significant QTLs. Glucoraphanin and sinigrin were found to increase bioactivity, and more gluconapin contributed to decreased cell inhibition potency of broccoli extracts. In chapter 3, interactions between aliphatic glucosinolates, at a standard level of myrosinase, were determined with a three-point mixture analysis. Antagonistic relationships between different glucosinolates were identified. These
interactions were somewhat mitigated when compounds were run through enzymatic
digestions alone. Individual compounds are relevant for drug and supplement
development, but the compounds we eat in foods are consumed together all at once.
Therefore, this work reinforced the importance to design and source optimal plant lines
for health properties.

By using a GWAS—chapter 4—two traits relevant to skin health were
identified on the oat genetic linkage map. *In vitro* inflammatory response effects and
migration speed induction changes induced by oat extracts were used to screen a diverse
mapping population. A total of 22 QTLs were found to significantly associate to wound
healing phenotypes. However, none of these regions co-localized with one another. This
decreased the confidence that any of these QTLs could be used for marker assisted
breeding. A better path would be to use the QTL as markers for predictive breeding to
select the best parent matches. Plant lines that consistently had a strong effect on wound
healing are currently being investigated for a mechanistic understanding of oat’s
association to skin health. Going forward, gene association studies should focus on well
understood health aspects of crops and use phenotypic screens where variability is
primarily derived from the plant genetics. This is a unique approach to gain in-depth
understanding of a medicinal or nutraceutical property of a plant.

The identification of a QTL for a specific trait is only an initial step toward
producing a new, healthier plant variety. Advanced breeding techniques can be used to
apply information found in an association study. As methods of gene editing continue to
advance, identifying direct targets in plants that play a role in their health properties will
be essential to get the most out of emerging technologies and to create plants with added functional value.
APPENDICES
APPENDIX A: FLAVONOID SUPPLEMENTATION AND CARDIOVASCULAR DISEASE, IN: DIETARY SUPPLEMENTS IN HEALTH PROMOTION

Published and reprinted with permission from:
13 Flavonoid Supplementation and Cardiovascular Disease

Weston Bussler, Joseph Hildebrand, Catherine Mixon, Ty Wagoner, Slavko Komarnytsky, and Gabriel Keith Harris

CONTENTS

13.1 Flavonoid Supplementation and Cardiovascular Disease ........................................... 282
13.2 Impact of Heart and Vascular Diseases ........................................................................ 283

13.3 Flavonoid Overview ...................................................................................................... 285
13.3.1 Flavonols .............................................................................................................. 286
13.3.2 Flavanols .............................................................................................................. 287
13.3.3 Anthocyanidins .................................................................................................... 287
13.3.4 Flavonones ........................................................................................................... 287
13.3.5 Isoflavones ........................................................................................................... 288
13.3.6 Flavones .............................................................................................................. 288
13.3.7 Absorption and Bioavailability ........................................................................... 288

13.4 Functions of Flavonoids in Metabolism .................................................................... 290
13.4.1 Flavonoids as Antioxidants ................................................................................ 290
13.4.2 Nitric Oxide Production ..................................................................................... 290
13.4.3 Inflammation Reduction .................................................................................... 291
13.4.4 Phase II Enzyme Detoxification ....................................................................... 292
13.4.5 Microflora Interactions ...................................................................................... 292
13.4.6 Potential Complications from Flavonoid Supplementation ................................ 292

13.5 Cardiovascular Disease .............................................................................................. 294
13.5.1 Risk Factors for CVD ......................................................................................... 294

13.6 Flavonoid Supplementation Effect on Chronic CVD ................................................ 294
13.6.1 Blood Lipid Profile .............................................................................................. 294
13.6.2 Hypercholesterolemia ....................................................................................... 296
13.6.3 Hypertension and Blood Circulation ................................................................ 297
13.6.4 Atherosclerosis ................................................................................................. 298

13.7 Flavonoid Supplementation and Acute CVD/CVD Mortality ................................... 299
13.7.1 Congestive Heart Failure .................................................................................... 299
13.7.2 Ischemic Heart Disease and Myocardial Infarction ........................................... 300
13.1 FLAVONOID SUPPLEMENTATION AND CARDIOVASCULAR DISEASE

The connection between flavonoids and cardiovascular health was observed as early as the 1930s, when the Hungarian scientist Albert Szent-Györgyi observed the effects of the flavone hesperidin on changes on the permeability of the vascular epithelium (Pearson 1957). Over the past 80 years, there has been a continued rise in cardiovascular diseases (CVD), reaching an annual $108.9 billion in healthcare costs (Heidenreich et al. 2011). CVD is a blanket term that refers to a number of diseases and conditions that affect the circulatory system. Within this classification are two of the leading causes of death in humans—heart attack and stroke. Researchers are studying treatments that may reduce the impact of risk factors, which are leading to these death-causing events. Growing interest has been placed on the relationship of plant-derived secondary metabolites and health, particularly the group of plant polyphenols known as flavonoids. Research on flavonoids has increased significantly over the past two decades (Figure 13.1).

This chapter will explore the prevalence and risk of CVD, the different types of flavonoids, the metabolic functions of flavonoids, the outcomes of flavonoid supplementation on CVD, the potential complications of high-dose supplementation, and the mechanisms of flavonoid–CVD interaction. In addition, this chapter will clarify

![Figure 13.1: PubMed Search of "Flavonoids and Cardiovascular Disease" 1960–2012.](image-url)
13.2 IMPACT OF HEART AND VASCULAR DISEASES

As diagnosis of various vascular and heart diseases rises, so do the overall healthcare costs. In the United States, CVD costs roughly $272.5 billion annually (Heidenreich et al. 2011). This amount includes the direct costs of healthcare services, medications, which does not include the indirect costs associated with losses in productivity. To put this into perspective, approximately $1 of every $6 spent on healthcare in the United States is spent on CVD treatments and services. The direct healthcare costs of four types of CVD are shown in Table 13.1.

According to the CDC, in 2009–2010, 35.7% of American adults and 16.9% of adolescents and children were considered overweight or obese. Also, 67 million, or roughly 1 in 3 Americans, have been diagnosed with hypertension. Another 25.8 million people, about 30%, have been diagnosed with Type II Diabetes. Obesity, hypertension, and Type II Diabetes are major risk factors for CVD. Today, 49% of the population has one or more of these risk factors, increasing their risk of being diagnosed with CVD.

The numbers shown in Table 13.1 reflect the massive number of people who suffer from CVD and the differences in the types of CVD—outlined in Table 13.2—that

<table>
<thead>
<tr>
<th>Cardiovascular Disease</th>
<th>Cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary heart disease</td>
<td>$108.9</td>
</tr>
<tr>
<td>Hypertensive disease</td>
<td>$93.5</td>
</tr>
<tr>
<td>Stroke</td>
<td>$53.9</td>
</tr>
<tr>
<td>Heart failure</td>
<td>$34.4</td>
</tr>
</tbody>
</table>

**TABLE 13.2**

Total Deaths in 2008 Due to Cardiovascular Diseases Globally (AHA 2011)

<table>
<thead>
<tr>
<th>Region</th>
<th>Ischemic</th>
<th>Cerebrovascular</th>
<th>Rheumatic</th>
<th>Overall CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Americas (North/South)</td>
<td>881,000</td>
<td>437,000</td>
<td>10,000</td>
<td>1,944,000</td>
</tr>
<tr>
<td>Europe</td>
<td>2,195,000</td>
<td>1,278,000</td>
<td>25,000</td>
<td>4,584,000</td>
</tr>
<tr>
<td>East Mediterranean</td>
<td>287,000</td>
<td>292,000</td>
<td>22,000</td>
<td>1,195,000</td>
</tr>
<tr>
<td>Africa</td>
<td>374,000</td>
<td>449,000</td>
<td>11,000</td>
<td>1,254,000</td>
</tr>
<tr>
<td>South-East Asia</td>
<td>1,834,000</td>
<td>1,192,000</td>
<td>56,000</td>
<td>3,616,000</td>
</tr>
<tr>
<td>Western-Pacific</td>
<td>1,383,000</td>
<td>2,504,000</td>
<td>96,000</td>
<td>4,735,000</td>
</tr>
<tr>
<td>Total deaths</td>
<td>7,254,000</td>
<td>6,152,000</td>
<td>220,000</td>
<td>17,327,000</td>
</tr>
</tbody>
</table>
# TABLE 13.3
Cardiovascular Events and Conditions that Contribute to Mortality

<table>
<thead>
<tr>
<th>Condition</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atherosclerosis</td>
<td>Plaque builds up inside arteries. Plaque can be made up of fat, cholesterol, calcium, and/or any other compounds found in blood. Atherosclerosis is a precursor to heart attack and stroke.</td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td>Hypercholesterolemia is a disorder characterized by high levels of blood cholesterol. Cholesterol is manufactured primarily in the liver and then carried to the cells throughout the body by LDL. As cholesterol and other fats do not dissolve in water, they cannot travel through the body unaided. Lipoproteins are particles formed in the liver to transport cholesterol and other fats through the bloodstream. For CVD, high cholesterol constitutes high levels of LDL cholesterol or low HDL cholesterol. Having either of these, increases the CVD mortality risk by allowing more LDL to be oxidized and congest the arteries.</td>
</tr>
<tr>
<td>Congenital heart disease</td>
<td>Birth defect that can cause a narrowing of the aorta or holes in the wall that separates the left and right ventricle (Kirkler 1992).</td>
</tr>
<tr>
<td>Rheumatic heart disease</td>
<td>Chronic condition that is characterized by the development of fibrosis in the heart valves following Rheumatic fever. Bacterial infection (Rheumatic fever) that causes fibrosis of heart valves, living in areas without access to antibiotics. A total of 15.6 million people are currently affected worldwide. About 1% of all school children in Africa, Asia, the Eastern Mediterranean region, and Latin America show signs of the disease.</td>
</tr>
<tr>
<td>Peripheral arterial disease</td>
<td>Atherosclerosis that usually affects the arteries and legs.</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>Circulatory problems that inhibit proper circulation primarily in the legs. People that have this disease will have pain in their calves, especially when they are walking.</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>An imbalance of prooxidants and antioxidants. The excessive formation of reactive species leads to atherosclerosis.</td>
</tr>
<tr>
<td>Chronic inflammation</td>
<td>Typically associated with prolonged OS. A low level chronic immune response leads to vascular damage and atherosclerosis.</td>
</tr>
<tr>
<td></td>
<td>A systolic blood pressure at or above 140 mmHg and/or a diastolic blood pressure at or above 90 mmHg.</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Systolic BP = the maximum pressure in the arteries when the heart contracts. Diastolic BP = the minimum pressure in the arteries between the hearts contractions.</td>
</tr>
<tr>
<td>Acute CVD (Death Causing Events)</td>
<td></td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>Problems circulating blood to the heart muscle. A lack of enough oxygenated blood (ischemia) causing angina (chest pain) and dyspnea (shortness of breath) and myocardial infarction, commonly known as a heart attack.</td>
</tr>
<tr>
<td>Cerebrovascular disease (Stroke)</td>
<td>Occurs when there are problems circulating blood through the blood vessels of the brain. A blockage with effects lasting less than 24 h is referred to as a transient ischemic attack. A complete blockage with long-term effects is referred to as a cerebrovascular thrombosis (clot) or a stroke. Sometimes, a blood vessel in the brain can burst, resulting in a long-term brain damage.</td>
</tr>
</tbody>
</table>

(Continued)
### Table 13.3 (Continued)

**Cardiovascular Events and Conditions that Contribute to Mortality**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ischemic stroke</td>
<td>Complete blockage of blood flow to the brain</td>
</tr>
<tr>
<td>2. Hemorrhagic stroke</td>
<td>Obstruction within a blood vessel supplying blood to the brain</td>
</tr>
<tr>
<td>3. Heart failure</td>
<td>The pumping action of the heart can no longer provide enough blood to the rest of the body to serve required functions. Symptoms include, shortness of breath from congested lungs, swelling of legs from water retention due to little blood reaching the kidneys, increased need for urination, stomach bloating/loss of appetite, dizziness/fatigue/weakness from less blood reaching the brain, and finally a rapid or irregular heartbeat.</td>
</tr>
</tbody>
</table>

are afflicted people from different areas. Ischemic heart disease deaths predominate Europe, both America and the Eastern Mediterranean; Cerebrovascular deaths are much more prominent in comparison to Africa, Southeast Asia, and the Western-Pacific. Lifestyle differences have been attributed to these different rates, with diet representing a major portion. Research over the past two decades has shown that flavonoid supplementation may have an effect on decreasing the risk factors of CVD and many other chronic diseases. Local consumption of flavonoids also might explain why there are different rates of CVD in different regions of the world.

### 13.3 Flavonoid Overview

Plants synthesize a number of polyphenolic compounds with many biological purposes such as biochemical defense and pigmentation. These compounds are classified by their chemical structure into ten different classes, one of which is the flavonoids. Recent evaluations state that more than 5500 different naturally occurring flavonoids have been discovered (Yamane and Kato 2012). They are characterized by an aromatic ring structure with one or more hydroxyl groups. The characteristic ring structures of flavonoid compounds and their subsequent functional groups contain a large amount of electron density, thereby creating partial negative charges. These partial negative charges become targets for chemical reactions such as functional group modification, metal chelation, redox reactions, or protein/enzyme interactions (Pourcel et al. 2007). They also provide plants with antimicrobial properties, protection from ultraviolet rays, and limit the ability of herbivore digestive enzymes. These features are able to help the plants survive and thrive (Pourcel et al. 2007). Flavonoids are categorized into six subclasses based on oxidation state or the connection between aromatic rings. The subclasses are listed in Table 13.4.

The majority of flavonoids are derived from a flavone backbone, namely, 2-phenyl-1,4-benzopyrene, shown in Figure 13.2, which gives them their characteristic structure. Isoflavones are derived from a slightly different 3-phenyl-1,4-benzopyrene backbone. Flavonoids generally exist in two forms: Flavonoid glycosides, which are
### TABLE 13.4
Classification of Flavonoids with Common Examples and Associated Foods

<table>
<thead>
<tr>
<th>Flavonoid Class</th>
<th>Examples</th>
<th>Associated Foods*&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flavonols</td>
<td>Catechin, epicatechin, gallicatechin</td>
<td>Tea, cocoa, wine</td>
</tr>
<tr>
<td>Flavonones</td>
<td>hesperetin, naringenin</td>
<td>Citrus, grapefruit, lemon</td>
</tr>
<tr>
<td>Isoflavones</td>
<td>Genistein, glycitein</td>
<td>Soybean, chickpea</td>
</tr>
<tr>
<td>Flavonols</td>
<td>Quercetin, myricetin, kaempferol</td>
<td>Onion, wine, most fruits</td>
</tr>
<tr>
<td>Flavones</td>
<td>Luteolin, apigenin</td>
<td>Thyme, oregano, celery, and parsley</td>
</tr>
<tr>
<td>Anthocyanidins</td>
<td>Cyanidin, malvidin, delphinidin</td>
<td>Red, blue, and purple berries</td>
</tr>
</tbody>
</table>

*<sup>a</sup> Arts and Hollman (2005).
*<sup>b</sup> de Pascual-Teresa, Moreno, and Garcia-Viguera (2010).

---

**FIGURE 13.2** Basic structural backbone of flavonoids.

attached to sugars, and aglycones, which are not. Glycosides are the more dominant form of most flavonoids with the exception of flavonols (Williamson 2003). Attachment to sugars plays an important role in bioavailability by altering absorption routes. The general structure of the flavonoid classifications is shown in Figure 13.3. Most features of the 2-phenyl-1,4-benzopyrene structure are shown below, while isoflavones are derived from 3-phenyl-1,4-benzopyrene.

### 13.3.1 FLAVONOLS

Flavonols are the most commonly consumed flavonoids because they are found in a variety of common fruits and vegetables. The most abundant flavonols in food are quercetin, myricetin, and kaempferol. These compounds are commonly associated with onions, apples, and wine, although they are present in most fruits and vegetables. Quercetin is the main dietary source of flavonols, with an estimated daily intake of 25–30 mg making up roughly 65% of the global daily flavonoid consumption (Ross and Kasum 2002; Perez-Vizcaino and Duarte 2010). Due to its abundance in food supply, quercetin is the most studied flavonoid to date (Perez-Vizcaino and Duarte 2010).
13.3.2 FLAVANOLS

Flavanols, often referred to as flavan-3-ols, have also been studied extensively because they are commonly found in two highly consumed foods: tea and chocolate. Other sources of flavan-3-ols are grape products and some legumes. The common flavanols catechin, epicatechin, and gallocatechin exist as monomers, mainly in the aglycone form (de Pascual-Teresa, Moreno, and García-Viguera 2010). Flavanols also commonly exist in a polymerized form as proanthocyanidins.

13.3.3 ANTHOCYANIDINS

Anthocyanidins are the aglycone derivative of anthocyanins, which are attached to sugars for solubility and stability. These compounds are extremely sensitive to pH, light, and oxidative conditions. They exhibit a well-characterized color change from red to blue with increasing pH. Anthocyanins are water-soluble pigments. Six types of anthocyanidins are commonly found in red and purple berries: pelargonidin, cyanidin, delphinidin, petunidin, peonidin, and malvidin. Other foods that contain anthocyanidins are red wines, cereals, and some leafy/root vegetables. Cyanidin is the most common anthocyanidin in foods; black grapes contain up to 600 mg/100 g and berries can contain up to 500 mg/100 g (Erdman et al. 2007).

13.3.4 FLAVONONES

Flavonones are present in highest concentrations in citrus fruits. These compounds are generally found in a glycosylated form. Flavonones such as neofisetin—which are commonly found in citrus fruits such as grapefruit—feature an attached
neohesperidose that imparts a bitter flavor (Manach et al. 2004). Research is limited on flavanones because natural sources are limited in human diets.

13.3.5 Isoflavones

While all of the flavonoid classes have been tested individually in vivo, isoflavones are the most widely tested in humans due to their abundance in food supply. There are conflicting scientific reports on the potential health benefits, but they are marketed online as cures to heart disease and inflammation (Espín, García-Conesa, and Tomás-Barberán 2007). They are unique flavonoids in that they belong to the phytoestrogen class. The two most common isoflavones are genistein and glycitein, which are commonly found in soybeans, chickpeas, and other legumes. Although they do not participate in hormonal signaling, they feature similar conformations of hydroxyl groups to estradiol giving them pseudo-hormonal properties (Manach et al. 2004).

13.3.6 Flavones

Flavones are the least common flavonoid found in fruits and vegetables (Manach et al. 2004). Apigenin and luteolin are the most common flavone glycosides, and are mainly found in celery and parsley (Erdman et al. 2007).

13.3.7 Absorption and Bioavailability

Like all nutrients, flavonoids pass through the digestive system and interact with many of the nutrient uptake mechanisms. Note that the ingestion of flavonoid-rich foods does not directly lead to increased levels in the body. Factors such as absorption efficiency, metabolism, protein interactions, and excretion of these compounds impact the bioavailability of flavonoids.

When consumed by humans, limited levels of flavonoid glycosides are absorbed through the SGLT-1 transporters in enterocytes. Other methods of flavonoid absorption come when the glycoside is hydrolyzed to the aglycone form after reactions in either the oral cavity or via microbial metabolism in the large intestine (Walle 2004). The glycoside forms have a greater ability for absorption (Hollman and Katan 1997).

More recent research has examined the synergistic effects of diet on absorption of flavonoids. Low-fat diets are recommended as a way to curb cardiovascular risk factors, but dietary fat may improve the absorption of potentially beneficial flavonoids. Plasma concentration of flavonoids taken with a high fat diet improved by 45% when compared to a fat-free meal (Guo et al. 2013). Studies with green tea catechins have shown that flavonoids can complex with lipoproteins and alter the lipid usability of fatty acids within lipoproteins, increasing the excretion of lipids (Koo and Noh 2007). Flavonoid derivatives have a high affinity for proteins, and flavonoids commonly interact with human serum albumin (HSA) in the blood. This affinity is amplified by increased dietary oleic acid, the type of fatty acid found in olive and peanut oils (Bolli et al. 2010).
The number of clinical trials investigating flavonoid bioavailability and absorption has gradually increased over the past two decades. These studies commonly measure flavonoid levels in blood plasma or urine to determine the extent at which these compounds are absorbed by the body. The methods used to determine flavonoid bioavailability are the post-prandial test, oral–intravenous balance, oral–fecal balance, and observation of effects of chronic consumption. Post-prandial tests look at the area under the curve in the blood after a single dose and are considered the most suitable test to measure flavonoid bioavailability (Erdman et al. 2007).

Flavonoids are a diverse group of molecules with many possible modifications, so it is not surprising that some are more efficiently absorbed than others. The conjugated form of flavonoids has an impact on the location of absorption, and the level of glucosylation has a direct impact on absorption (Arts and Hollman 2005). Early research suggested that the glycone conjugates of quercetin, which are naturally found in onions, have higher absorption than aglycone or rutinoside conjugates that may be present in supplements (Hollman et al. 1995). Manach also found onion sources of quercetin to be more bioavailable than quercetin commonly found in apples or green tea (Manach et al. 2005).

As flavonoids are metabolized, they undergo a variety of conformational modifications during the passage from the small intestine to the blood stream (Crozier, Jaganath, and Clifford 2009). These may include deglycosylation, glucuronidation, sulfation, and methylation. The resulting metabolic byproducts are different from the glycone and aglycone conjugates that are almost exclusively found in plants (Williams, Spencer, and Rice-Evans 2004). Typical flavonoid concentrations in plasma are less than 1 μmol/L (Manach et al. 2004). Isoflavones are known to be particularly well absorbed in humans, as are quercetin glucosides. The least well-absorbed flavonoids are proanthocyanidins and flavan-3-ols, but this could be a result of the instability of these compounds in the body rather than poor absorption. For example, a 50 mg dose of epicatechin translates in to a maximum plasma concentration between 0.4 and 2.5 μM (Erdman et al. 2007).

Modification of flavonoids may make them more available to absorption by humans. One such example is a solid dispersion of quercetin with polyvinylpyrrolidone, Kollidon 23%. This modified version increased solubility 436 times over quercetin (Costa et al. 2011). Moving forward, to properly assess flavonoids and their ability to alter the prognosis of CVD, the safety and effectiveness of the increased bioavailable forms must be proven. Flavonoid supplementation has the potential to provide a number of health benefits, but the possible negatives must be understood.

Without a specific uptake mechanism in humans, these alternate routes of absorption present many possible routes of flavonoid entry into the bloodstream. Further research in absorption and bioavailability of flavonoids is still needed. An important concept that has been lacking in the literature is the relationship between absorption efficiency and age. Not only are there inherent regulators of absorption, as mentioned above, age-related complications add another dimension to the puzzle. The body’s efficiency in all processes declines with age, which will have a direct impact on the metabolism and processing of nutrients, and may later affect flavonoid dose recommendations. Once absorbed, flavonoids impact a variety of metabolic pathways.
13.4 FUNCTIONS OF FLAVONOIDS IN METABOLISM

Flavonoids have the ability to induce many different biochemical responses in humans. The main functions that are applicable to CVD include antioxidant potential by scavenging and preventing the formation of free radicals, reduction of chronic inflammation, phase II enzyme detoxification, and interactions with microflora metabolism. This section also identifies potential negative metabolic effects (García-Lafuente et al. 2009; Lu, Xiao, and Zhang 2013). Each of these activities is unique, but they are also closely related to one another.

13.4.1 FLAVONOIDS AS ANTIOXIDANTS

Free radicals are extremely reactive molecules because they contain unbound electrons. Free radicals are constantly created in biological systems either intentionally or unintentionally, and they are involved in a number of required physiological functions, such as NADPH oxidase activity (Actis-Goretta et al. 2003). Unintended radical compounds can interfere with cellular functions, leading to oxidative stress (OS) (Aruoma, Kaur, and Halliwell 1991). OS results from an imbalance in the ratio of prooxidant and antioxidant compounds in humans (Jones 2006). Examples of damaging effects of OS in organisms include peroxidation of membrane lipids, oxidative damage to nucleic acids, and the oxidation of sulfhydryl groups in proteins. Chronic OS has been identified as an inducing risk factor for a number of fatal diseases in humans, including CVD. The primary free radicals of concern in humans are reactive oxygen and nitrogen species (ROS and RNS) (Wang et al. 2006).

13.4.2 NITRIC OXIDE PRODUCTION

Nitric oxide (NO) is a compound produced from the amino acid arginine and nitric oxide synthase (NOS). Multiple genes are used to make the NOS family of enzymes; each is regulated differently and produces NO for a different biological purpose. The functions and biological locations of the different NOS types are described in Table 13.5. Production of NO is often used as a marker of CVD treatment efficacy because epithelial NO is a key regulator of vascular health. The scientific literature offers confusing explanations of the positive and negative effects of NO. Ultimately,

<table>
<thead>
<tr>
<th>Nitric Oxide Synthase Type</th>
<th>Biological Location</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuronal nitric oxide synthase (nNOS)</td>
<td>Nervous tissue and skeletal muscle</td>
<td>Cell signaling and communication</td>
</tr>
<tr>
<td>Endothelial nitric oxide synthase (eNOS)</td>
<td>Endothelium</td>
<td>Vasodilation</td>
</tr>
<tr>
<td>Inducible nitric oxide synthase (iNOS)</td>
<td>Immune and cardiovascular systems</td>
<td>Immune response to pathogens and foreign materials</td>
</tr>
</tbody>
</table>
increasing concentrations of NO can either protect from or lead to CVD depending on where it is produced in the body.

NO alone can be cytotoxic by reacting with a superoxide anion, forming a damaging RNS called peroxynitrite. The development of RNS in humans focuses mainly on iNOS-derived NO rather than eNOS, in which production of NO can be viewed as a positive for preventing CVD mortality by increasing blood flow. The triggers for iNOS activity and NO release are lipopolysaccharides (LPS), bacteria, viruses, and proinflammatory cytokines. Excess NO production leads to more RNS and ROS, causing more OS. Antioxidants are molecules that have the ability to absorb the free radicals of oxidative compounds, negating the negative health impacts (García-Lafuente et al. 2009). Many flavonoids show great antioxidant potential in vitro by scavenging and negating radical compounds (Wang et al. 2006); however, the actual impact of flavonoids as antioxidants may not be as great because the concentration of flavonoids circulating in the blood is much lower than that of other free-radical scavenging molecules in humans (Williams, Spencer, and Rice-Evans 2004). This suggests that the antioxidant potential of flavonoids is not their only CVD related bioactivity.

In addition to decreasing the damaging effects of iNOS derived NO, flavonoids have the ability to increase the eNOS in vascular endothelial cells to promote the formation of NO. An increase in NO production promotes vasodilation that limits hypertension (Duffy and Vita 2003). By activating eNOS, flavonoids can reduce blood pressure. This is a primary reason why the effects of flavonoids are being investigated in patients with blood circulation issues, specifically those with hypertension and improper blood flow.

13.4.3 Inflammation Reduction

NO biosynthesis—along with prostaglandins—is involved in a variety of inflammatory reactions. The proper regulation of this pathway is crucial to management of cardiovascular health (González-Gallego, Sánchez-Campos, and Tuñón 2007). Inflammation is a biological process where the immune system responds to triggers of infection or injury by quickly creating a specifically targeted destructive environment to resolve the infection, repair any damage to the cell, and restore the original equilibrium. The ideal inflammation process is short, self-limiting, and specific to the intended spot. When inflammation functions properly there are no dangers; when left unchecked, overactive immune responses can lead to the development of debilitating chronic diseases (García-Lafuente et al. 2009). Cyclooxygenase-2 (COX-2) is responsible for the conversion of arachidonic acid to prostaglandins. COX-2 is needed for normal body function even though excess levels signal chronic inflammation (González-Gallego, Sánchez-Campos, and Tuñón 2007).

Flavonoids inhibit enzyme activity by acting on the protein kinases involved in signal transduction. This limits the ability of transcription factors, nuclear factor-κB (NF-κB), and activator protein-1 (AP-1), along with increasing antioxidant defenses through activation of nuclear factor erythroid-derived-2 (NF-E2). NFκB is a protein complex responsible for controlling DNA transcription in animal cells; it is involved
in cellular responses to stress, cytokines, free radicals, oxidized low-density lipo-protein (LDL), ultraviolet radiation, and bacterial/viral antigens. AP-1 is a group of leucine zipper transcription factors that bind to DNA and promote cellular differentiation, proliferation, and apoptosis (Fujio et al. 2004). NF-E2 has shown to be a primary cellular defense mechanism against cytotoxic effects of OS (Linker et al. 2011; Gilmore 2006; González-Gallego, Sánchez-Campos, and Tuñón 2007). These anti-inflammatory mechanisms work together to reduce reactive species that may lead to the oxidation of LDL and eventually atherosclerotic plaques.

13.4.4 Phase II Enzyme detoxification

Phase II enzymes play a role in the elimination and excretion of toxic substances from the body, including ROS and other electrophilic compounds. Increasing the activity of these enzymes leads to an increased resistance to oxidative damage that may lead to atherosclerosis (Li, Cao, and Zhu 2006). Green tea catechins have been shown to increase activity of these elements in HepG2 cells (Chen et al. 2000). Similar results were seen in rats when fed green tea leaves, which contain high levels of catechins, by increasing the activity of the glutathione-S-transferase (GST) family of Phase II enzymes in the liver (Lin et al. 1998).

13.4.5 Microflora interactions

More recent research proposes that the health related benefits of flavonoids might be due to their effects on intestinal microbiota as opposed to a direct benefit on human epithelial cells. Flavonoids have a flexible structure that allows for interactions with a wide variety of enzymes, suggesting that they may have an innate ability to regulate microbial metabolic enzymes. Recent studies have identified more than 900 metabolic targets of flavonoids that have homologues in the genomes of human intestinal bacteria. These pathways benefit the intestinal bacteria by inducing changes in metabolism that indirectly benefit humans (Lu, Xiao, and Zhang 2013). Future research should expand on the recent developments of the microbiome flavonoid interactions in humans.

13.4.6 Potential Complications from flavonoid Supplementation

As consumers look for a “magic bullet” to alleviate health problems, flavonoid supplements have gained attention. Supplements have limited regulation by the FDA, so claims and dosage information are not strictly monitored. There has been an increase in evidence showing specific drug interactions between flavonoids and conventional medications that can present detrimental health effects. One proposed mechanism for this interaction is the presence of flavonoids altering the bioavailability of drugs in the body.

Despite evidence suggesting health benefits from flavonoid consumption, dosage plays an enormous role; there are potential risks associated with high dose supplementation of flavonoids. The risks of consuming high levels of these compounds are not well understood due to the vast number of compounds in the flavonoid family,
the lack of adequate dietary intake information evaluating exact amounts that are consumed by people, and limited studies investigating the risk factors of flavonoid compounds. Many foods that have been long consumed by humans are rich in these compounds, giving them the Generally Recognized as Safe (GRAS) moniker. However, some hazards have been identified including anti-nutritional effects, thyroid toxicity, drug interactions, genotoxicity/carcinogenicity, and developmental effects. Human clinical studies supplementing quercetin show adverse effects after a single dose over 4 g; however, a treatment under 500 mg twice daily for a month did not show adverse effects (Russo et al. 2012).

The anti-nutritional effect of flavonoids can come from a number of mechanisms. There is a potential for inhibition of proteolysis within the gut, reduced glucose uptake, impaired food utilization, and impaired mineral absorption. Flavones can inhibit the activity of catechol-O-methyl-transferase, resulting in an increase of norepinephrine concentration and activation of fat oxidation (Dulloo et al. 1999). Epigallocatechin (EGC) and epigallocatechin gallate (EGCG) from green tea have the potential to inhibit glucose uptake by competing for sodium-glucose-cotransporter-1 (Kobayashi et al. 2000). While some researchers cite this action as a negative impact of consuming flavonoids, others conclude that this may be beneficial because it slows glucose uptake, protecting from diabetes mellitus and metabolic syndrome (Johnston et al. 2005). Dietary non-heme iron absorption was reduced by 50%–70% when consumed in tandem with 20–50 mg of polyphenols. This inhibition was increased to 60%–90% when the polyphenol consumption was increased to 100–400 mg (Hurrell, Reddy, and Cook 1999).

High levels of quercetin supplementation have been of some concern because of the potential toxicity of their oxidized product quercetin-quinone (QQ). QQ is a semiquinone radical that has the ability to arylate protein thiols, creating electrophilic compounds that can damage cells. Despite potential concerns, this theory has not yet been proved in any human-based models (Russo et al. 2012).

Plants produce some flavonoids as toxins to deter other animals from eating the plants, so it is reasonable to assume potential toxicity in humans as well. Although some of the beneficial health effects of flavonoids are related to their antioxidative properties, flavonoids at high concentrations can act as prooxidants in the presence of redox-active heavy metals (Galati and O’Brien 2004). This oxidative capacity can lead to the formation of ROS that can damage biological compounds. In the presence of Cu²⁺ and oxygen in vitro, flavonoids can produce radicals that damage DNA (Sakihama et al. 2002).

Potentially harmful effects on the liver have been seen in vivo mice models with EGCG and polymeric tannins (Li et al. 2002; Gali-Muhtasib, Yamout, and Sidani 2000). Both of these compounds can be found in herbal supplements, so there are potential safety implications with very large doses of these compounds. However, it is important to note that toxic doses of flavonoids are much greater than the levels found in dietary sources, so toxicity would more likely be associated with large dose supplements (Galati and O’Brien 2004). Moreover, the prooxidant effect of flavonoids is catalyzed by free copper, ascorbate, and peroxidase activity in vitro. When scaling up to a human some discrepancies occur, most notably the in vivo copper is not usually in the free form that is most damaging, and peroxidase is
compartmentalized making these in vitro effects not directly translatable to humans (Erdman et al. 2007).

One example that has been identified is the flavonoid naringenin—present in grapefruits—inhibiting cytochrome P450 3A4 activity (Veronese et al. 2003). These drug interactions need to be considered by persons attempting to consume excessive flavonoid concentrations in their diets (Morris and Zhang 2006). With these possible negative effects identified, the latent therapeutic potential of flavonoids on CVD and its risk factors can be explored in experiment models. The full range of flavonoid effects on metabolism are crucial to developing proper recommendations for them as supplements to treat CVD.

13.5 CARDIOVASCULAR DISEASE

CVD is not just a single disease; it is a complex collection of diseases and risk factors that take different forms in different individuals. These risk factors are better understood if classified into two categories: diseases that can directly lead to death (such as ischemia, heart failure, and stroke), and chronic diseases that contribute to these events (such as atherosclerosis, hypertension, and peripheral heart disease). Explanations of these events are presented in Tables 13.3. Other risk factors to consider for prevention of CVD are insulin resistance, obesity, smoking, lack of exercise, age, gender, family history, alcohol consumption, and race.

Studies use in vitro and animal models to understand some of the factors that can lead to chronic CVD. The complex mechanisms for altering risk factors such as atherosclerosis and inflammation are used to understand flavonoid CVD activity more than the acute events. Acute events such as MI and stroke are more easily recorded, so large-scale epidemiological and meta-analyses are used to correlate flavonoids with CVD risk reduction. Table 13.3 refers to both the chronic and acute forms of heart disease.

13.5.1 RISK FACTORS FOR CVD

The development of CVD cannot be isolated to one specific risk. Rather, a cascade of possible factors contributes to the development of CVD. Figure 13.4 attempts to outline some of the major modifiable risk factors that lead to CVD in individuals. The relationship is very complex and specific answers are hard to come by; however, any action that might reduce a risk factor should be investigated. The remainder of this chapter will review studies that observe the relationship between supplementing flavonoids and the impact they have on CVD and risk factors.

13.6 FLAVONOID SUPPLEMENTATION EFFECT ON CHRONIC CVD

13.6.1 BLOOD LIPID PROFILE

Blood lipid modifications are recommended by both the American Heart Association and National Cholesterol Education to prevent CVD (Kim et al. 2011). Specifically, they recognize the association of high levels of total cholesterol, LDL cholesterol, and blood triglycerides, and as well as the low levels of HDL cholesterol and
their relationship to cardiovascular morbidity and mortality (Kim et al. 2011). Hypercholesterolemia is a condition characterized by high cholesterol concentrations throughout the blood of an individual. High total cholesterol, high LDL cholesterol, high blood triglycerides, and low HDL cholesterol are risk factors predicting development of other more damaging CVD. Drug- and diet-formulated treatment plans are recommended to improve blood lipid profiles.

The oxidation of lipids in LDLs that become trapped in the extracellular matrix of the subendothelial space causes events that induce inflammation, which can eventually damage blood lipid profiles, leading to atherosclerosis and other forms of CVD (Berliner et al. 1995). Transcription factors are activated by the oxidized lipids, leading to the production of proteins and an inflammatory response that ultimately causes a fatty streak in arteries. This progression inhibits the mechanical abilities of the artery wall, while also allowing predisposition to plaque ruptures at sites of monocyctic infiltration. This eventually induces thrombosis, or other obstructions, making blood flow less efficient (Berliner et al. 1995).

Long-term supplementation of flavonoids may be effective in improving blood lipid profiles by decreasing serum concentrations of inflammation markers. Anthocyanins have been shown to reduce chronic inflammatory markers and cytokines in hypercholesterolemic adults (Zhu et al. 2012). Twice daily doses of 320 mg anthocyanins for 24 weeks resulted in a 21.6% decrease in serum levels of high sensitivity C-reactive protein (hsCRP), a 12.3% reduction in soluble vascular cell adhesion molecule-1 (sVCAM-1), and a 12.8% decrease in Interleukin-1 beta (IL-1beta).
Serum LDL cholesterol was also reduced by 10.4%, and HDL was increased by 14% (Zhu et al. 2012).

In vitro models have demonstrated that flavonoids have the ability to sequester chronic inflammation. When tested on primary human adipocytes, flavonoids were shown to reduce inflammation (Chuang et al. 2010). The anti-inflammatory action of flavonoids has been attributed to the inhibition of inflammatory enzymes. Flavonoids can reduce the expression of iNOS, cyclooxygenase (COX-2), lipoxygenase, prostanoids, leukotrienes, chemokines, adhesion molecules, and cytokines (García-Lafuente et al. 2009). Prostaglandins and NO biosynthesis are involved in the pathway for inflammatory response to stimuli. The isoforms of iNOS and COX-2 are responsible for producing the majority of these pathway mediators. Flavonoids have the capability to inhibit both of these enzymes, as well as other enzymes in the inflammatory process such as adhesion molecules and C-reactive protein. Overexpression of these molecules in the inflammation pathway is attributed to the development of chronic diseases (Tuñón et al. 2009).

Cocoa flavanols (catechin and epicatechin) have shown the ability to inhibit proinflammatory cytokine and leukotriene production, while also inducing vessel relaxation. These improvements on the blood lipid profile may reduce damage from oxidation and inflammation, which are leading risk factors for CVD (Mathur et al. 2002).

### 13.6.2 Hypercholesterolemia

Between 1988 and 2008, high total cholesterol rates have decreased for men and women, aged 45–75 years or older. The CDC reports that roughly one in four Americans in this demographic (about 32 million) take statins, cholesterol lowering drugs (CDC 2013). The high cholesterol rates for men and women aged 20–44 have increased over the same time period. This data trend in older adults has shown that interventions with statins have decreased cholesterol rates (CDC 2013). The data also show that cholesterol levels are on the rise for younger adults, increasing their need for early cholesterol-lowering interventions. One of the major problems with starting a statin regimen in younger individuals is adherence. Data have shown that only 57% of patients continue their treatment (Grundy 2013).

Reasons for stopping statin treatment include observed side effects (10%–20% of patients experience muscle problems), fear of some side effects reported by patients but not confirmed by FDA (some patients claim to experience memory loss, a fuzzy feeling, or lack of focusing ability), medication costs, lack of insurance coverage, not understanding the treatment benefits, and an overall lack of commitment to the treatment (Grundy 2013). Supplementation of flavonoids may provide a safer and more manageable alternative to the current prescribed statin treatment. Adequate flavonoid supplementation alone, or in addition to statins, has shown an active ability to improve blood lipid profiles.

Flavonoids are thought to act on uptake mechanisms for cholesterol by the small intestine effectively reducing circulating cholesterol (Ikeda et al. 1992). Investigating the changes brought on by consuming a metabolite on blood lipid profile is an easy way to determine if it may present a health benefit. One such study examined the protective effect of quercetin in hypercholesterolemic rats. The results showed that
Flavonoid Supplementation and Cardiovascular Disease

Quercetin supplementation in combination with a high cholesterol diet was able to decrease liver triglycerides by 24%, reduce total cholesterol by 22%, and reduce serum cholesterol by 20% (Mariee, Abd-Allah, and El-Beshbishy 2012).

These protective effects seem to translate to humans as well. One study assessed the effect of dietary flavonoid supplementation on CVD in post-menopausal women taking statins to treat Type II Diabetes. The dosage was set at 27 g/day of chocolate with 850 mg flavan-3-ols and 100 mg isoflavones, or a placebo, for one year. The results showed that the treatment group had a significant reduction in total cholesterol, LDL/HDL ratio, and total LDL cholesterol. They concluded that high flavonoid diets can work well with drug treatments to reduce lipid profile risk factors (Curtis et al. 2012).

A cohort study based in Finland examined the food intake of 2748 men and 2385 women, 30–69 years of age and free of any known heart disease. After a 26-year period, five major flavonoid intakes—quercetin, kaempferol, myricetin, luteolin, and apigenin—were estimated based on dietary data. Median flavonoid intake was found to be 3–4 mg/day with approximately 95% of the total intake coming from quercetin. The incidence of coronary mortality was found to be higher in populations with low dietary flavonoid intake. The protective effect of flavonoids was associated with a diet high in apple and onion intake. The proposed mechanism for the effect was the prevention of oxidation of LDL particles, however, other mechanisms could have also contributed. The beneficial effect of high fruit and vegetable diets correlated to their flavonoid content (Knekt et al. 2002).

Other blood lipid profile studies found that cocoa powder flavonoids reduced oxidative susceptibility of LDL cholesterol while keeping prostaglandin concentrations stable (Wan et al. 2001). Ingestion of green tea catechins may alter blood lipid profile by moderately reducing the intestinal absorption and delivery of cholesterol (Koo and Noh 2007).

13.6.3 HYPERTENSION AND BLOOD CIRCULATION

Proper blood circulation is necessary to life, and any impediments to blood circulation can be very damaging. One important CVD risk factor is atherosclerosis, which is the medical condition in which fat, cholesterol, and other substances build up on the walls of arteries and form hard structures called plaques. Plaques are created when these materials are oxidized, causing them to adhere to the walls of arteries. This can eventually lead to partial or full occlusion as well as creation of thrombogenic material potentially leading to stroke.

Atherosclerosis is a chronic condition; it progresses slowly and is asymptomatic until abnormal narrowing of a blood vessel occurs. This narrowing can become so severe that blood supply to tissues becomes insufficient or plaques break off and become thrombogenic material, leading to ischemia. Flavonoids interact with this narrowing mechanism by increasing the vasodilation of blood vessels. Specifically, a 2013 study showed high berry anthocyanin consumption was able to increase flow mediated dilation at two intervals, 1 and 6 h (Rodriguez-Mateos et al. 2013). In a meta-analysis published by the European Society of Cardiology, chronic and acute intakes of flavan-3-ol rich foods, such as chocolate or cocoa, reduced diastolic blood
pressure and improved flow-mediated dilation. Additionally, the improvements were found at all dose levels (Wright et al. 2012).

When treating hypertensive patients with quercetin supplements, an observed reduction in markers for OS—ferric reducing ability of plasma (FRAP) and poly ADP ribose polymerase (PARP)—was observed. However, this effect was only seen in patients with Stage I Hypertension; prehypertension patients were not significantly affected (Edwards et al. 2007).

The Kuopio Ischemic Heart Disease Risk Factor Study (KIHDS) consisted of 1380 middle-aged eastern Finnish men in which the mean common carotid artery intima-media thickness (CCA-IMT) was studied. The study identified a trend for an inverse association between intake of flavonoids and mean CCA-IMT (Mursu et al. 2008). A different double-blind, randomized study found the flavonoid intervention did not significantly change CCA-IMT, augmentation index, or blood pressure, but pulse pressure variability improved. In a subgroup with pulse wave velocity data, a 10% CV risk reduction over control patients was observed in high risk patients already using cholesterol lowering drugs (Curtis et al. 2013). These findings suggest that a high flavonoid diet would prevent further circulatory damage.

### 13.6.4 Atherosclerosis

Atherosclerosis is often the underlying pathological condition of CVD. The condition involves the initiation and perpetuation of atherosclerotic lesions, which erode or rupture, resulting in acute CVD events, such as poor blood flow, myocardial infarction, or cerebrovascular attack. Atherosclerosis is a risk factor for many CVD, and is caused by oxidation of cholesterol, lipids, and other products resulting in narrower passageways for blood circulation. It can be thought of as a combination of poor blood lipid status, constricted blood vessels, and blood circulation issues. Atherosclerosis can lead to any of the acute CVD death events, so prevention and treatment can reduce its impact on CVD. Flavonoids are often studied specifically for their therapeutic potential on atherosclerosis.

OS results from an imbalance between excessive formation of reactive oxygen or nitrogen species and limited antioxidant defenses. Endothelium and NO are key regulators of vascular health. NO bioavailability is modulated by ROS that degrade NO, uncouple NO synthase, and inhibit synthesis. These processes lead to the development of endothelial dysfunction, which has been identified as an initial step to atherosclerosis. Flavonoids have shown to improve vascular function through increased endothelial nitric oxide (eNOS) production in animals and humans, but many of these studies have lacked placebo controls and large sample sizes (Rimbach et al. 2009). Red wine has been evaluated as a way of reducing LDL oxidation and lowering atherosclerosis risk. The results indicate that cells treated with red wine showed a reduced amount of copper-induced LDL (Kerry and Abbey 1997).

Licorice root has been used in traditional Chinese medicine to treat a number of ailments; when examined for flavonoid content, licorice root extract yielded five flavonoids: liquiritin, liquiritigenin, isoliquiritigenin, and 7,4'-dihydroxyflavone (Jayaprakasam et al. 2009). Supplementation of licorice root extract to hypercholesterolemic patients for a period of 1 month, followed by an additional 1 month of
placebo consumption was examined. Licorice consumption resulted in a moderate reduction in the patients’ plasma susceptibility to lipid peroxidation (by 19%), and a marked reduction in the susceptibility of plasma LDL oxidation. LDL oxidation slowed by 55%, in comparison to the lag time of LDL isolated from plasma derived before licorice extract consumption (Aviram 2004).

A study by Kapetivadze examined the effect of flavonoid supplementation in combination with the anti-hypertension drug indapamide in diabetic patients. After 10 weeks, results showed an observed reduction of total cholesterol by 7%, LDL cholesterol was reduced by 10%, total glucose was reduced by 9.4%, and HDL cholesterol concentration increased by 8%. These results suggest a benefit for flavonoid supplementation in combination with drug treatment for atherosclerosis (Kapetivadze et al. 2010).

Various medications are recommended to assist in treating this condition including ACE inhibitors, antiplatelet drugs, and many others. Recent studies have found that flavonoids function in many of the same ways as ACE inhibitors and antiplatelet drugs. Individuals with high intakes of flavonoids show effective dilation of blood vessels. Flavonoids also have been shown to inhibit the formation of clots in the blood (Wright et al. 2012).

13.7 FLAVONOID SUPPLEMENTATION AND ACUTE CVD/CVD MORTALITY

13.7.1 Congestive Heart Failure

Congestive heart failure is a diagnosis of various issues that pertain to the efficiency at which the heart pumps blood through the body and the declining strength of the heart muscle. Congestive heart failure can be caused by many other conditions, including coronary artery disease (CAD), peripheral arterial disease (PAD), atherosclerosis, heart attack, cardiomyopathy, and conditions that overwork the heart, such as hypertension (AHA 2011).

There are different types of heart failure, affecting different areas in the heart and inhibiting the amount of blood filling the heart. This limits the amount of oxygen-rich blood pumped throughout the body. Systolic dysfunction occurs when the heart muscle does not contract with enough force; therefore, decreased amounts of oxygenated blood will be pumped throughout the body (AHA 2011). Diastolic dysfunction mainly affects the ventricles of the heart. The heart muscles contract normally; however, the ventricles cannot properly relax or become stiff, inhibiting the heart's normal filling capacity, meaning less blood will exit the heart; this inhibits major organs and tissues from receiving the required amounts of blood in order to normally function (AHA 2011). Angioplasty and stents are medical procedures that attempt to unblock heart arteries. In cases where the two above methods do not succeed, heart bypass surgery is necessary to unblock coronary arteries (AHA 2011). These are costly interventions, and they may not be able to be administered in time to prolong life. Flavonoid consumption has been investigated as a way to avoid the development of congestive heart failure (Wright et al. 2012).

According to various studies, many experts have recommended increased intakes of flavonoid-rich diets for heart failure preventative purposes (Peterson et al. 2012).
In the Dutch Zutphen Elderly Study cohort, flavonoids in the diet were analyzed for 805 men, aged 65–84. Their diet consisted of: tea (61%), onions (13%), and apples (10%). Their total flavonoid intake was 25.9 mg per day. At both 5 and 10-year follow-ups, flavones and flavonols were associated with reduced coronary heart disease (CHD) mortality. Also, higher intake of flavan-3-ols, specifically, catechins and epicatechins, showed a reduced risk of CHD after a 10-year follow-up. Similarly, in Finnish populations, quercetin and kaempferol were associated with lower CHD mortality; quercetin showed a significant decrease in mortality, while kaempferol showed a small effect on mortality (Knekt et al. 2002).

13.7.2 Ischemic Heart Disease and Myocardial Infarction

Ischemic heart disease is characterized by a blockage in the coronary arteries that reduces the flow of blood as it returns to the heart. Myocardial infarction (MI) is the most common diagnosis when referring to these types of blockages. MI is the irreversible death of heart muscle tissue due to the loss of oxygen via insufficient blood flow to a section of the heart. Many studies have analyzed the effect that various flavonoids have on the prevention of MI in select populations. Despite some disagreement, the results are promising, suggesting that flavonoid consumption may have a positive effect on the prevention of MI.

A cohort study that was conducted over an 18-year time span yielded promising information correlating anthocyanin intake with a lower risk of developing MI. When 93,600 women were examined over an 18-year period, 405 cases of MI were reported. When analyzing the diets of patients the biggest inverse association correlated with women who consumed 58–643 mg of flavonoids per day (2–35 mg were anthocyanins). The 15 mg increase in anthocyanins decreased risk of MI by 17%. Interestingly, the authors also concluded that combined intakes of blueberries and strawberries decreased risk of MI compared to participants who consumed less than three servings per week of those fruits (Cassidy et al. 2013).

Many cohort studies estimate dietary consumption of flavonoids and relate it to the observed incidence of CVD. Table 13.6 summarizes a portion of these results. The Rotterdam Study released by the American Journal of Clinical Nutrition compared MI incidence in black tea-drinkers and nondrinkers. A longitudinal study of 4807 individuals with no previous MI history reported that the relative risk of MI was lower for subjects consuming >375 mL tea per day compared with non-tea drinkers (RR = 0.57; 0.33–0.98). Drinking tea even more strongly lowered the risk of fatal MI (RR = 0.35; 0.13–0.98) when the highest and lowest tertiles of tea consumption were compared (Geleijnse et al. 2002).

The Finnish Mobile Clinic Health Examination Survey also found an inverse correlation between quercetin intake and ischemic heart disease. Comparing the highest and lowest levels of quercetin intake as determined from surveys shows RR = 0.79; 0.63–0.99, with statistically significant results ($p < 0.02$) (Knekt et al. 2002).

Two of the largest health studies in the United States did not find a significant inverse correlation between flavonoids and ischemia. The Women’s Health Study observed 38,445 women for 6.9 years for CVD risks and events. The highest level of flavonoid consumption had an age-adjusted RR = 0.88 (0.63–1.24), but the results
### Table 13.6 (Continued)
Table Summarizing Several Flavonoid Studies on Acute CVD and Their Outcomes

<table>
<thead>
<tr>
<th>Location</th>
<th>Population Mean Follow-Up</th>
<th>Outcome</th>
<th>Diet</th>
<th>Results (Relative Risk, 95% Confidence Interval)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurses health study (USA)</td>
<td>66,360 W</td>
<td>12 years MI and CHD</td>
<td>Flavonol, flavone</td>
<td>Women in the highest quintile of flavonol and flavone intake had RR for nonfatal MI (RR: 1.05, 0.85, 1.29,  p = 0.55) and CHD death (RR: 0.81, 0.57–1.16,  p = 0.29) compared to lowest quintile. Both results were non-significant.</td>
<td>Lin et al. (2007)</td>
</tr>
<tr>
<td>Iowa women's study (USA)</td>
<td>34,489 W</td>
<td>16 years CHD, CVD and</td>
<td>Anthocyanidins, flavones</td>
<td>Adjusted RR between highest and lowest consumption quintiles for anthocyanidins and CHD (RR: 0.88 (0.78, 0.99)) and CVD (0.83 (0.63, 0.99)); between flavonoids and CHD (RR: 0.78 (0.63, 0.96)); and between flavonoids and total mortality (RR: 0.88 (0.82, 0.96)). These results were all significant (p &lt; 0.05).</td>
<td>Mink et al. (2007)</td>
</tr>
</tbody>
</table>

### Table 13.6
Table Summarizing Several Flavonoid Studies on Acute CVD and Their Outcomes

<table>
<thead>
<tr>
<th>Location</th>
<th>Population Mean Follow-Up</th>
<th>Outcome</th>
<th>Diet</th>
<th>Results (Relative Risk, 95% Confidence Interval)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zutphen elderly study (Netherlands)</td>
<td>894 M</td>
<td>10 years CHD mortality and MI outcome</td>
<td>Flavonol</td>
<td>RR of CHD mortality for the highest tertile of daily flavonol consumption (more than 29 mg) compared to the lowest (less than 19 mg) was 0.67 (0.27–0.82) p = 0.006. RR for MI incidence for highest tertile compared to lowest was 0.62 (0.24–1.05) p = 0.078.</td>
<td>Heros et al. (1997)</td>
</tr>
<tr>
<td>Caerphilly Study (UK)</td>
<td>1490 M</td>
<td>14 years IHD mortality</td>
<td>Black tea</td>
<td>Flavonol intake, highest compared to lowest quartile, was weakly positively correlated to IHD mortality RR = 1.6 (0.8–2.9) p = 0.19.</td>
<td>Heros et al. (1997)</td>
</tr>
<tr>
<td>Rotterdam study (Netherlands)</td>
<td>4807 M and W</td>
<td>5.6 years MI incidence and MI mortality</td>
<td>Flavonoids, tea</td>
<td>The RR of MI incidence for highest tertile (&gt;375 mg/day) compared to non-drinkers was 0.57 (0.33–0.94). The intake of dietary flavonoids was significantly inversely associated with fatal MI (RR: 0.33 (0.13–0.89)) in upper compared with lower tertiles of intake.</td>
<td>Gislier et al. (2002)</td>
</tr>
<tr>
<td>Finish Mobile clinic health examination survey</td>
<td>10,034 M and W</td>
<td>28 years IHD mortality</td>
<td>Quercitin</td>
<td>The highest quartile of quercitin intake was correlated with lower mortality from IHD compared to lowest quartile, RR: 0.79 (0.63–0.99) p = 0.02.</td>
<td>Kaurk et al. (2002)</td>
</tr>
<tr>
<td>Women's health study (USA)</td>
<td>38,645 W</td>
<td>6.9 years CVD risk and events</td>
<td>Flavonoids, tea</td>
<td>For both CVD and important vascular events, age and treatment adjusted RR was 0.88 (0.63, 1.24) p = 0.15. A small proportion of women consuming &gt;4 cups tea a day had a non-significant (p = 0.07) reduction in event risk. Flavonoid intake alone was non-significantly inversely correlated with CVD risks and events.</td>
<td>Sesso et al. (2003)</td>
</tr>
</tbody>
</table>

(Continued)
weren't significant \( p < 0.07 \). Consumption of more than four cups tea per day also non-significantly correlated with the risk of cardiovascular events (Sesso et al. 2003). The Nurses Health Study also found non-significant relationships between total flavonol and flavone intake and non-fatal MI (RR = 1.05; 0.85–1.29; \( p < 0.55 \) and fatal CHD event (RR = 0.81; 0.57–1.16; \( p < 0.29 \)) (J. Lin et al. 2007). The Iowa Women's Study analyzed intake of anthocyanins, flavanones, and flavones separately, finding significant differences. Comparing the highest and lowest quintiles of consumption, anthocyanidins were associated with reduced risk of CHD (RR = 0.88; 0.78–0.99) and CVD (RR = 0.91; 0.83–0.99), while flavanones were associated with reduced risk of CHD (RR = 0.78; 0.82–0.96) and flavones with total mortality (RR = 0.88; 0.82–0.96)(Mink et al. 2007).

Several animal studies have also shown individual flavonoids to have a protective effect against ischemia. One study used rats to look at the modulation of vascular ion channels stemming from NO release after flavonoid consumption as a way to reduce ischemic disease. Pre-treatment with flavonoids with these substitutions decreased the time required to reach half-maximal contracture and the maximum ischemic contracture, while the non-substituted flavonoids (4'-hydroxyflavanone, 6-hydroxyflavone, and 7-hydroxyflavanone) did not improve post-ischemia functional parameters (Testai et al. 2013). Also, these flavonoids improved post-ischemia recovery.

The preventative effects of epicatechin on lysosomal alterations—a marker that accompanies ischemic myocellular damage—showed that when the lysosomal membrane is compromised there could be an elevation of undesirable enzymes in intra- and extracellular space. The ability to maintain normal levels of lysosomal enzymes in body fluids/tissues would also be compromised in this state. When there is localization of acid hydrolases in the cardiac myocytes of the lysosomes, enzyme release into the cytosol will follow, leading to myocardial injury or cellular death. Alterations in the activity of lysosomal enzymes have been connected to patients that experience MI. Epicatechin reduced myocardial damage in MI-induced rats (Stanely and Prince 2013). This suggests that certain flavonoids provide a specific benefit to treating and preventing ischemic heart disease (Testai et al. 2013).

13.7.3 Stroke

Stroke is the second leading cause of death both in developed and developing countries (WHO 2013), and there are a growing number of studies examining the relationship between flavonoids and stroke incidence. Some earlier studies examined total stroke occurrence, rather than separating into categories based on the nature of the cardiovascular event. These two main categories are ischemic and hemorrhagic stroke.

Ischemic stroke occurs when blood flow is restricted to the brain. This most often results due to narrowing of the blood vessels in the brain caused by atherosclerosis. Hypertension is the greatest controllable risk factor for ischemic stroke (Roger et al. 2012). In terms of both mortality and frequency of occurrence, ischemic stroke is more dominant than hemorrhagic stroke in the United States.

Hemorrhagic strokes occur after vascular rupture in the brain (intracerebral), or just outside the brain (subarachnoid), forces blood into the surrounding area causing compression on the brain tissue. The most important diet-related risk factor for
hemorrhagic stroke is hypertension, which strains the blood vessels, although the causes of hemorrhagic stroke are less known than ischemic stroke. Due to the different mechanisms of action, ischemic and hemorrhagic stroke should be analyzed separately in future cohort studies. Due to the difficulty of directly measuring the effects of flavonoids on stroke, especially in small, clinical trials, many studies assess a number of risk factors for stroke instead. These risk factors may include atherosclerosis, hypertension, HDL, and LDL cholesterol levels, and flow-mediated dilation. Table 13.7 describes the large, population-based studies that have been conducted to date examining the effects of flavonoid intake on stroke risk.

The potential health effects of dietary flavonoids have been studied extensively, but the relationship between flavonoids and stroke have been inconsistent. The Zutphen study (Keli et al. 1996) was an early prospective cohort study that examined total dietary intake of flavonoids based on participant evaluations. The relative risk (RR) of stroke incidence for flavonoid consumption of 28.6 mg/day or greater was 0.27 (0.11–0.70) compared to consumption of less than 18.3 mg/day. They also examined tea intake, showing a RR of 0.31 (0.12–0.84) for daily consumption of more than 4.7 cups of tea compared to less than 2.6 cups. While this study brought early attention to flavonoids as having possible benefits on stroke incidence, the small sample (n = 552) only examined men.

More recent studies have focused on tea as a primary source of flavonoids. Excluding water, tea is the most commonly consumed beverage in the world (Cheng 2006). Black tea is the most commonly produced variety, with an estimated 76%–78% of total production. In the USA, 80% of tea consumers are black tea (Basu and Lucas 2007). Tea varieties have been studied independently, including green tea (Tanabe et al. 2008) or black tea (Gans et al. 2010). More recent studies have looked at total tea intake (Lopez-Garcia et al. 2009; Larsson, Virtamo, and Wolk 2013; Leurs et al. 2010). The Tokamachi-Nakasato study focused specifically on several levels of green tea intake in Japan (n = 6,358) (Tanabe et al. 2008). Both middle (several times per week: RR = 0.43, 0.25–0.74; p < 0.002) and high (several cups daily: RR = 0.41, 0.24–0.70, p < 0.001) tea intakes were associated with a considerably lower risk of total stroke incidence. Even after separation of cerebral infarction and cerebral hemorrhage, the results were consistently statistically significant. Similar results were found when looking primarily at black tea consumption (Gans et al. 2010). After adjusting for age and sex, high tea consumption (4.1–6 cups a day) was associated with a reduced risk of total stroke (RR = 0.69, 0.50–0.95, p < 0.07). However, the trend was non-significant when adjusted for additional lifestyle variables such as physical activity, waist circumference and alcohol consumption (RR = 0.92, 0.66–1.28; p < 0.63). Despite the large sample size (n = 37,514), only 70 incidences of stroke-related mortality were observed. Additional studies are needed to evaluate this relationship.

A more recent Finnish study with a great number of stroke occurrences also found an inverse correlation between tea intake and total stroke (Larsson, Virtamo, and Wolk 2013). Consumption of more than four cups of tea per day showed a statistically significant impact (RR = 0.77, 0.61–0.98) on all strokes, split up by the type of event.

European studies have also embraced the potential protective power of the flavonoids in chocolate. Early research on the subject suggested that chocolate
### Table 13.7 (Continued)

#### Table Summarizing Several Flavonoid Studies on Stroke Risk and Their Outcomes

<table>
<thead>
<tr>
<th>Location</th>
<th>Population</th>
<th>Mean Follow-Up</th>
<th>Diet</th>
<th>Results</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>10,904 M</td>
<td>8 years</td>
<td>Cocoa</td>
<td>The adjusted RR for comparing top vs. bottom quartiles was 0.32 (0.30–0.38) for stroke.</td>
<td>Buijsse et al. (2010)</td>
</tr>
<tr>
<td>Tokushima-Nakano (Japan)</td>
<td>2087 M</td>
<td>5 years</td>
<td>Tea</td>
<td>A lower risk was observed for total stroke incidence in both the middle (multivariable HR, 0.43; 95% CI, 0.25–0.74; p = 0.02) and the high (multivariable HR, 0.41; 95% CI, 0.24–0.70; p = 0.001) categories of green tea consumption.</td>
<td>Takahge et al. (2008)</td>
</tr>
<tr>
<td>Nurses Health study (US)</td>
<td>83,076 W</td>
<td>24 years</td>
<td>Tea</td>
<td>RR = 0.79 (0.49–1.29) for stroke.</td>
<td>Lopez-Garcia et al. (2010)</td>
</tr>
<tr>
<td>Netherlands cohort study</td>
<td>120,852 M</td>
<td>10 years</td>
<td>Tea</td>
<td>For tea, moderate (1–2 cups/d) and high tea consumption (≥3 cups/d) were inversely related to risk mortality in men (HR: 0.75; 95% CI 0.61–0.93 and HR: 0.71; 95% CI 0.57–0.88, respectively (p = 0.007)) compared with 0–1 cup of tea consumption.</td>
<td>Leurs et al. (2010)</td>
</tr>
<tr>
<td>Epic-NL cohort (Dutch)</td>
<td>37,514 M</td>
<td>13 years</td>
<td>Tea</td>
<td>Tea consumption tended to be associated (p = 0.07) with a reduced risk of stroke with the lowest RR for 6.1 to 6.9 cups per day (0.69; 95% CI, 0.50–0.95). This relation attenuated to nonsignificant after multivariate adjustment (HR, 0.92; 95% CI, 0.66–1.28; P trend = 0.63).</td>
<td>de Knippen et al. (2010)</td>
</tr>
</tbody>
</table>
consumption may reduce blood pressure by increasing NO production. A review of the dietary intake of chocolate in Germany found that increased chocolate intake was associated with a lower total risk of stroke (RR = 0.52, 0.30–0.89). The authors of the study proposed that the beneficial effects of chocolate intake were due to reductions in blood pressure, a risk factor for stroke (Buijsse et al. 2010).

More recent studies have built on this framework by examining the effects of chocolate on both ischemic and hemorrhagic stroke, separately. The Swedish Mammography study evaluated women (n = 33,372) over 10.4 years, and found that a 50 g/week increase in chocolate consumption was associated with a decreased risk for total stroke (RR = 0.86, 0.77–0.96), cerebral infarction (RR = 0.88, 0.77–0.99), and hemorrhagic stroke (RR = 0.73, 0.54–0.99). However, the results were only significant for women in the highest quartile of chocolate consumption, suggesting higher doses are necessary for a protective effect (Larsson, Virtamo, and Wolk 2011). The Swedish cohort of Men (n = 37,103) found similar results after 10.2 years, but even higher chocolate consumption was necessary for a protective effect. Men in the highest quartile of consumption (62.9 g/week) had a relative risk for total stroke of 0.83 (0.70–0.99) compared to the lowest quartile (Larsson, Virtamo, and Wolk 2012).

These results of these three studies are promising, however, the caloric density of chocolate is a concern. Daily doses of 100 g dark chocolate have been shown to decrease systolic blood pressure, but the increased daily caloric intake (nearly 480 kcal) may counteract the cardiovascular benefits (Taubert D et al. 2003). An extract of flavonoids from cocoa, if effective, may provide the protective benefits in a less calorie-dense form.

Not all prospective cohort studies have found flavonoids to be protective against stroke. The Nurses Health Study in the United States (n = 69,622) found no evidence for an inverse association between flavonoid intake (flavonols and flavanones) and total, ischemic, or hemorrhagic stroke (Cassidy et al. 2012). However, there were only 253 cases of hemorrhagic stroke, so additional studies are needed. The same study also compared stroke risk with green and black tea consumption. High consumption levels (more than 4 cups/day) tended to be associated with a reduced risk of total stroke (RR = 0.79, 0.49–1.29), but the results were not significant (p < 0.19) (Lopez-Garcia et al. 2009).

There is a growing amount of evidence suggesting flavonoids may provide protective benefits from stroke, but additional studies are warranted before clear recommendations can be made. Individual flavonoids need to be evaluated in randomized studies to determine if the protective effects are due to the flavonoids, their metabolites, or synergistic effects between flavonoids and their food source.

13.8 PROPOSED MECHANISMS OF FLAVONOID-INDUCED CVD PREVENTION AND TREATMENT

The complexity of biological systems makes it difficult to understand the mechanisms flavonoids use to prevent and treat CVD. The flavonoid mechanisms in metabolism outlined in Section 3 present many possibilities for CVD interactions. The
many forms of CVD outlined in Table 13.3 provide different symptoms and risk factors. Over 5000 flavonoids have been characterized that provide different cellular impacts. The following focuses on the mechanisms of flavonoid CVD prevention and treatment.

Ingestion of green tea catechins may prevent CVD by reducing the intestinal absorption of cholesterol and other dietary lipids. It appears that EGCG interferes with the emulsification of fats by limiting digestion through inhibition of phospholipase A2 and micellar solubilization of lipids; thus, limiting fat absorption efficiency (Koo and Noh 2007).

In vitro studies of red wine flavonoids found an association with CVD prevention, which was attributed to the prevention of LDL oxidation through flavonoids antioxidant potential (Kerry and Abbey 1997). Flavonoids are efficient free-radical scavengers in vitro, but serum concentrations are typically much lower than other dietary antioxidants (Chun, Kim, and Lee 2003).

In animal models, flavonoids have the potential to reduce inflammation and protect against hypertension and atherosclerosis. Chronic quercetin supplementation in mice reduces LPS-induced inflammation (Patil et al. 2003). Green tea consumption protects against hypertension by reportedly decreasing OS and inflammation (Bhardwaj and Khanna 2013). Nuclear factor (erythroid-derived 2)-like 2 (Nrf2) is a transcriptional regulator of phase II antioxidant enzymes, and activation of Nrf2 has been suggested to be an important step in reducing OS associated with CVD (Reuland et al. 2013).

In vivo work has found that the epicatechins present in green tea inhibit the formation of oxidized serum cholesterol, which prevents progression toward atherosclerosis (Basu and Lucas 2007). Flavonoids can improve cerebral blood flow and prevent platelet aggregation (Perez-Vizcaino and Duarte 2010). Flavonoids also promote activation of eNOS, which increases the bioavailability of NO in the system by reducing formation of RNS and leads to dilatory effects (Sudano et al. 2012). Hypertension is considered the most important risk factor for cerebral infarction (Tanabe et al. 2008; Ikeda et al. 2007). Green tea also led to a reduction in atherosclerosis in rabbits by decreasing vascular endothelial growth factors expressed in atherosclerotic plaque (Kavantzas et al. 2006).

Flavan-3-ols, flavonols and proanthocyanidins have shown an effect on ACE activity by inhibiting the conversion of angiotensin, thus preventing blood pressure increases (Actis-Gorella et al. 2003; Balasuriya and Rupasinghe 2012). This suggests that the in vivo reduction in blood pressure is related to the ACE inhibition of flavonols. Flavonoids may also inhibit PDI reductase, an enzyme that participates in thrombus formation, which can lead to stroke. The flavonol rutin was shown to inhibit PDI reductase in vitro (Jasuja et al. 2012).

Modulation of vascular ion channels stemming from NO release is a major way flavonoids reduce ischemia mortality, likely due to flavonoids’ antioxidant capabilities (Testai et al. 2013). Previous studies conducted by these researchers have found that the flavones, apigenin, kaempferol, naringenin, and the isoflavone, daidzein, have demonstrated vasodilatory properties. There has also been recent evidence that flavonoids, vitexin, proanthocyanadin, orientin, and luteolin, exhibit cardioprotective effects in experimental models of myocardial I/R injury (Testai et al. 2013).
Compared to the hearts from the vehicle-treated rats, the rats pre-treated with the flavonoids that had the hydroxy or methoxy substitution exhibited higher rate pressure product and coronary flow values throughout the entire reperfusion time.

In addition to the prevention of stroke, several studies have proposed flavonoid supplementation as a means of reducing the severity of acute stroke damage (Chen et al. 2011; Perez-Vizcaino and Duarte 2010). Epicatechin doses administered to mice showed a neuroprotective effect during the 6-h window from stroke onset to irreversible neuron death (Shah et al. 2010). Reducing stroke severity may increase the likelihood of surviving the stroke. High anthocyanin consumption was found to impact neutrophil NADPH oxidase activity ultimately increasing blood flow (Rodriguez-Mateos et al. 2013). Inhibiting NADPH oxidase limited the ROS in the brain, reducing the damage after a stroke (Actis-Goretti et al. 2003).

Researchers have found that only the 5-hydroxy-substituted derivatives (5-hydroxy flavone, apigenin, chrysin, naringenin) improved post-ischemia functional recovery and a decrease in tissue injuries (Testai et al. 2013); 5-methoxy-flavone showed similar effects as well. Therefore, the capabilities of flavonoids were attributed to the 5-hydroxy-substituted derivatives (Testai et al. 2013). Further research is needed both in CVD development and flavonoid activity to pinpoint the exact way that flavonoid supplementation works to prevent CVD.

13.9 CONCLUSION

In this chapter, evidence was collected in order to better understand the mechanisms by which flavonoids prevent CVD and other risk factors. The powerful connection between flavonoid consumption and heart health has been identified through various studies. It should also be noted that there is a potential concern with over supplementation of flavonoid extracts, specifically those that have high absorption capabilities. The doses found in dietary sources of flavonoids appear to be safe.

In many previous studies, antioxidant and anti-inflammatory activity are separated; however, they work together. For example, antioxidants quench reactive reactions but regulation of iNOS prevents the formation of those radicals in the first place. The main functions of flavonoids in human metabolism include: antioxidant potential and formation prevention, reduction of chronic inflammation, phase II enzyme detoxification, microflora metabolism/interactions, and potential negative metabolic effects. The interactions between intestinal microbiota and flavonoids have not been characterized, which is a new frontier of possible research.

Many epidemiological studies use specific events such as heart attack and stroke to evaluate the protective benefits of flavonoids. The complexity of biological systems makes it difficult to understand exactly how flavonoids impact CVD. There are many possible effects of flavonoid supplementation: flavonoids lower LDL/blood lipid profiles by reducing absorption and decreasing delivery; they reduce chronic inflammation by preventing radical formation; they reduce COX enzyme activity and decrease iNOS production and flavonoids work as antioxidants to prevent LDL oxidation/foam cell development. The vasodilatory impact on hypertension, increasing eNOS (dilatory effects), preventing platelet aggregation, and performing phase II enzymatic detoxification; however, these correlations ignore the possible protection
that flavonoids may provide earlier in life. Age plays an important role in metabolism and absorption, and additional research on younger subjects may help evaluate the role of flavonoids before the onset and progression of CVD.

Moving forward, there will be foods that are developed specifically with higher or more bioavailable flavonoid contents. For example, genetically engineered tomatoes have been produced that have increased levels of rutin and kaempferol (Le Gall et al. 2003). Innovations in increasing bioavailability and creating isolated supplements need to be tested thoroughly before being distributed. Randomized trials are needed to establish the bioavailability of individual flavonoids and their metabolites, and to determine if the beneficial effects are actually due to the individual compounds or to synergistic effects between flavonoids and their food source. The importance of flavonoid dosage is absent from much of the literature. The preventative effects appear to be dose dependent, and high dose supplementation may lead to adverse health effects.

When investigating CVD and the impact of flavonoids, it is beneficial to understand that CVD is a collection of diseases and risk factors that affect individuals differently. The two categories of risk factors are better understood as diseases that can directly lead to death (such as ischemia, heart failure and stroke) and chronic diseases (such as atherosclerosis, hypertension, and peripheral heart disease). Therefore, the various flavonoids mentioned have different mechanisms and effects on these conditions. Separating the different types of CVD and specific causes can allow for cross-referencing flavonoid types, and the mechanisms of prevention they may exhibit.

For every $6 spent on health care, $1 goes toward treatment for CVD, and this trend is projected to increase over the next several decades. This is an exponential amount of resources that includes health care services, medications, and losses in productivity. Looking at the different types of CVD diagnoses on a global scale, ischemic heart disease predominates Europe, North and South America, and the Eastern Mediterranean. Cerebrovascular CVD deaths are more prominent in Africa, Southeast Asia, and the Western-Pacific. These regional differences are linked with lifestyle differences, mainly diet. Looking at these statistics and the different dietary habits of different populations, the data exhibits the effect that food has on heart health.

It can be concluded that flavonoid consumption has strong potential to improve heart and vascular health by inhibiting various risk factors that lead to CVD diagnoses in individuals all over the world.

REFERENCES


Flavonoid Supplementation and Cardiovascular Disease


Flavonoid Supplementation and Cardiovascular Disease


APPENDIX B: STUDENTS’ PERSPECTIVE ON GENOMICS: FROM SAMPLE TO SEQUENCE USING THE CASE STUDY OF BLUEBERRY


Copyright © 2013 Austin Mudd et. al.
Students’ perspective on genomics: from sample to sequence using the case study of blueberry

Austin B. Mudd1, Elizabeth J. White1, Michael P. Bolloskis1, Nicholas P Kapur1, Koyt W. Everhart1, Ying-Chen Lin1, Weston W. Bussler1, Robert W. Reid2 and Ryan H. Brown3 **

1 Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, USA
2 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Kannapolis, NC, USA
3 General Mills, Inc., Kannapolis, NC, USA

*Correspondence: Ryan H. Brown, General Mills, Inc., 150 North Research Campus Drive, Kannapolis, NC 28081, USA
a-e-mail: Ryan.Brown@genmills.com

INTRODUCTION

Increased public awareness about the health benefits of blueberries has elevated its worldwide demand in recent years; thus, researchers and breeders seek ways to make blueberry cultivation more efficient and fruitful. An increased understanding of the blueberry genome and pathways, for example, can facilitate selection for climatic adaptation, enabling blueberry growth in new regions as well as greater cultivation in existing growth regions. Through pathway elucidation, blueberry fruit quality can also be improved, providing consumers and breeders with the traits they desire. For consumers, these characteristics include increased antioxidants and dietary fiber, improved taste, cheaper costs, and added vitamin content. For breeders, disease resistance, pest resistance, increased yields, and increased zone hardness are qualities of interest. For a beginner in this discipline, a clear understanding of mapping, sequencing, genome assembly, and proteomics will provide the necessary framework to explore and contribute intellectually to the growing field of plant genomics. This paper presents an overview of these topics with an emphasis on how they pertain to the study of blueberries.

OVERVIEW OF BLUEBERRIES

Blueberries provide a perfect case study to explore current research in the fields of genomics, bioinformatics, and proteomics, as it is not one of the highly characterized plant species such as rice (Oryza sativa) or Arabidopsis thaliana. Humans have gathered and consumed blueberries for thousands of years. Blueberries are native to North America and were a staple in the diet of Native Americans and early settlers. Additionally, blueberries were used for medicinal purposes, relieving fevers, headaches, and persistent coughs (Trehane, 2004). Currently, North America is the largest commercial blueberry producer in the world, producing 600 million pounds of blueberries in 2012. Production in North America has been on the rise, with the total supply increasing by an average of 20% every two years since 2008 (Brazelton, 2013).

HEALTH BENEFITS

Consumption of blueberries has increased dramatically in the past 15 years due to consumers’ growing knowledge of its health benefits. Blueberries, along with several other berries, contain various types of anthocyanins, which are compounds with anticarcinogenic and therapeutic properties. These compounds have the ability to negate the effects of free radicals in the body, thereby protecting DNA integrity, improving brain function, preventing cancer cell formation, and reducing cardiovascular disease (Zafra-Stone et al., 2007). In fact, a study conducted in rats correlated the consumption of blueberries with protection against neurodegeneration and cognitive impairment. During an 8-week period, rats were injected with kainic acid (KA) and then fed either a diet containing 2% blueberry extract or a control diet. After analyzing a variety of performance variables, researchers concluded that KA-treated rats showed clear signs of impaired learning performance, but the blueberry diet reduced the impairment (Duffy et al., 2008).

GENETICS AND GROWTH

Blueberries have a varied genetic background. The basic chromosome number (x) is 12 (Rowland and Levy, 1994), and seven different ploidy levels have been found in the wild: 2x, 3x, 4x,
5x, 6x, 8x, and 12x (Brueederle and Vorsa, 1994). The blueberry genome is estimated to be 300–608 Mb/1C DNA, which is four to five times larger than the 125-Mb Arabidopsis thaliana genome and several times larger than other fruiting plant genomes such as the strawberry (240 Mb; Shulaev et al., 2011), grape, and raspberry genomes (Die and Rowland, 2013).

In terms of growth, blueberries require significant organic matter and thrive in acidic, sandy soil, which was originally considered marginal for crop production (Trehans, 2004; Rowland et al., 2012a). There are numerous varieties of blueberry in existence around the world, but four of the major varieties are Vaccinium corymbosum (Northern highbush), V. darrowii (Southern highbush), V. angustifolium (lowbush), and V. abei (rabbit-eye; Trehans, 2004). Different varieties, however, have particular requirements, such as chilling hours. The number of chilling hours is a measure of accumulated hours of temperatures below 7°C in the dormant season (Cesaraccio et al., 2004).

Genetic studies on blueberries have the potential to significantly improve fruit quality and the breeding process. Blueberry breeding will be simplified, for instance, through the use of genetic markers to identify desired seedlings. Rather than waiting for seedlings to mature and then examining their phenotypes, breeders can proactively select and cultivate individuals with desired genotypes, speeding up the screening process. Utilization of genetic markers will make traditional breeding methods more efficient and less random (Hancock et al., 2008).

SEQUENCING AND ASSEMBLY

Sequencing genomes has the potential to solve large-scale and small-scale biological problems by characterizing key genes involved in various biological processes. These include genes whose mutations lead to disease susceptibility (NOD2 and Crohn’s disease in humans; Ogura et al., 2001), genes involved in flowering and fruiting (FT and CO in Arabidopsis thaliana; Turch et al., 2008), and genes controlling anti-inflammatory responses (IL6 in humans; Xing et al., 1998). The recent development of easier and cheaper sequencing technologies has led to a rapid increase in the number of sequenced genomes. With regard to plants, Arabidopsis thaliana was the first sequenced plant genome in 2000 and has become the model organism for plant genomics due to its small genome size, short life cycle, and low chromosome number. These features are extremely desirable in the sequencing process. When sequencing a new plant species, various elements must be considered including the overall genome size, the presence of duplications and repetitive DNA, and ploidy factors. Crop species, in particular, are inherently difficult to work with due to large, repeated, and duplicated genomes from years of human induced inbreeding (Barthelson et al., 2011).

EVOLUTION OF SEQUENCING

There has been a proliferation of sequencing processes and technologies over the past decade. Sanger sequencing, the primary sequencing technology for the past 30 years, produces reads of up to 1,000 bases and remains the gold standard for accuracy. Unfortunately, the cost and required time for Sanger sequencing make it prohibitive, particularly given the recent rise of next-generation sequencing. The new technologies include Roche 454 pyrosequencing (2005), Solexa/Illumina (2006), SOLID (2007), and Helicos single-molecule sequencing (2008). These techniques center on placing millions of DNA fragments on a surface and then sequencing the fragments simultaneously. The fragments, however, are smaller in length, ranging from 25 to 400 base pairs depending on the technique (Pop, 2009). Although these methods are quickly advancing, third-generation sequencing technologies, which have longer-read lengths, shorter run times, and single-molecule resolution, have arisen over the past few years. These third generation technologies include Pacific Biosciences PacBio, Life Technologies Starlight, Oxford Nanopore, and Ion Torrent. Life Technologies Starlight, for example, has a read length of 1,500 base pairs and a run time of 20 minutes (Munroe and Harris, 2010; Igar et al., 2012). More detailed information about the mechanisms, strengths, and limitations of next-generation and third generation technologies can be found in Egan et al., 2012, and a review of plant-focused sequencing is available in Hamilton and Buell, 2012.

SHOTGUN SEQUENCING PROCESS

Genome sequencing projects in recent years have centered on whole-genome shotgun sequencing. During this process, a genome is broken into small fragments, and a subset that fits a prescribed size range is selected for sequencing from both ends, creating paired reads. Most sequencers utilize paired reads a fixed distance apart to compensate for small read fragment lengths. These sequences are then assembled with the end goal of reconstructing whole chromosomes (Pop et al., 2004b; Pop, 2009). In general, the genome assembly’s quality improves with more reads and depth, which is also known as coverage, or the ratio of total bases sequenced versus genome size (Pop, 2009; Kane et al., 2011). However, an increase in read depth will not solve the current difficulties with sequencing repetitive DNA. Repetitive DNA is a particular problem for computational approaches, as the repeats produce bias, ambiguities, and errors in the assembly process (Treangen and Salzberg, 2011). Various approaches have been suggested and utilized in recent years to avoid this problem, such as parallel sequencing with a similar species (Macas et al., 2007).

ASSEMBLY PROCESS

Using millions of sequenced short reads, genome assembly combines the reads into contiguous segments (contigs), which are ideally reconstructed into whole chromosomes (Boetzer et al., 2011). Initially, this process compares shotgun sequence reads and overlaps the reads using an indexing technique to identify the sequences that commonly overlap. After indexing, another algorithm aligns similar reads, laying out all of the alignments. To get a final DNA sequence, the layout is condensed down to a consensus (Pop et al., 2004b). Additional algorithms calculate the distance between mate pairs and construct supercontigs, also known as scaffolds, which are multiple contigs joined together (Boetzer et al., 2011). For example, if one end of a mate pair is located in one contig and the opposite end is located in another contig, the distance between the mate pairs ascertainsthe distance between these two contigs in the genome. Multiple mate pair links are desired in order to corroborate the correlation between two
contigs. Since the gap between these contigs lacks sequenced data, "N"s fill the empty space (Pop, 2009).

**DE NOVO AND COMPARATIVE ASSEMBLY**

Assembly of contigs and scaffolds utilizes two approaches: de novo assembly and comparative assembly. In de novo assembly, all reads are assembled based on algorithms, and no outside template is used. In comparative assembly, the sequence output can be aligned to a reference genome, a BAC (Bacterial Artificial Chromosome) library, or linkage groups. A reference genome utilizes a well-annotated genome that is similar to the species of interest. *Arabidopsis thaliana*, for example, is the primary reference genome for most plant studies. BAC libraries can also be used in comparative assembly. In this, a query genomic sequence of several thousand to over one hundred thousand bases is ligated onto bacterial vector DNA. Bacteria take up the DNA and then replicate on an agar plate. Following this replication, the bacterial DNA with the original query sequence is cut with restriction enzymes, and overlaying of different BACs can determine the restriction enzyme locations for a particular sequence (Shirasu et al., 1992; Shizuya and Kouros-Mehr, 2001). BAC libraries are useful because they only include the actual genome without additional cloning artifacts or alterations from the cloning process (Oseegawa et al., 2001). Finally, mapping of linkage groups can assist in assembly. For instance, mapping of restriction enzyme sites can result in an ordered directory of all restriction fragment length polymorphisms for the genome. The location of these restriction enzyme sites can be matched to the assembled contigs, helping to merge the contigs into scaffolds and eventually chromosomes (Pop, 2009).

**ASSEMBLY PROGRAMS**

A variety of programs exist for genome assembly including Arachne (Batzoglou et al., 2002), Bambus (Pop et al., 2004a), Euler (Pevzner et al., 2001), MIRA (Chevreux et al., 1999), Newbler (Margulies et al., 2005), and Velvet (Zerbino and Birney, 2008). These programs can be specialized for contig construction, scaffold building, or both, though most use a greedy approach, which starts with the reliable baseline data and then slowly merges additional data as long as this supplemental information agrees with the current sequence construction. Each assembler has unique strengths and weaknesses, as seen by comparing the MIRA and Newbler assemblers. The Newbler assembler, which is distributed by 454 Life Sciences, fails to account for repeats in an organism's genome, therefore misassembling the genome but creating fewer and larger scaffolds than the MIRA assembler (Lai et al., 2012). The MIRA assembler flags these repeats but results in a larger number of extremely small scaffolds. In addition, different assemblers work with different repeat types. For example, the Newbler assembler works best with 645 sequences. When deciding on assembling software, scientists must balance the strengths and limitations of each genome assembly program.

**GENOME MAPPING**

Genome mapping is an essential step for assembling highly repetitive genomes. There are two types of mapping: genetic mapping and physical mapping. Genetic mapping approximates the distance between genetic markers by comparing recombination frequencies. Genetic mapping requires cultivation of populations and acquisition of marker data that have been specifically designed for mapping. Physical mapping, on the other hand, anchors a linkage map to physical locations using sequencing, BAC libraries, or restriction enzyme sites (Collard et al., 2005). The end goal of genetic mapping is to create large linkage groups that sufficiently cover the organism's chromosomes and can be used to assemble the sequence into chromosomes.

**MAPPING POPULATION**

To make a genetic map, the distance between loci must be calculated by creating and genotyping the recombination frequencies between loci within a particular population. Accurately estimating recombination frequencies in a population requires a thorough understanding of the species, particularly the reproductive methods. For example, some plants can reproduce by self-pollination, while others require cross-pollination due to self-incompatibility. Using this knowledge, mapping populations create a population of individuals with stable segregation ratios. The genetic markers are then identified on each individual, and the recombination ratios for each marker are plotted in a matrix. The ratios of recombination determine the centiMorgan linkage distances between markers, such that 1 cm equals 1% recombination (Semagn et al., 2006).

**GENETIC MARKERS**

Genetic markers represent molecular differences between or within a species and are used to track particular sequences due to their proximity to genes of interest. Genetic markers distinguish polymorphisms between an individual offspring, assess genetic relationships, and assist in linkage map construction (Collard et al., 2005). As for marker types, RFLPs (restriction fragment length polymorphisms) are differences between homologous DNA sequences that are digested by restriction enzymes. Restriction enzymes cut specific locations on DNA, resulting in different length fragments. RAPDs (random amplification of polymorphic DNA) are DNA segments that are randomly amplified by PCR and total 8–12 nucleotides (Levi and Rowland, 1997). Another type of marker, EST-PCR (expressed sequence tag-polymermic chain reaction), is more commonly used today. ESTs are short DNA fragments of 200–500 nucleotides and are generated from sequencing one or both ends of an expressed gene, called complementary DNA (cDNA). Since EST markers are transcribed from mRNA, they only contain expressed genes and do not include introns (Dhananjay et al., 2004). In addition to ESTs, SSRs (simple sequence repeats) have regularly been used in plant genomics. SSRs, also called microsatellites, are short repeated DNA sequences of 2–6 base pairs (Boches et al., 2006).

**BLUEBERRY MAPPING**

The first blueberry genetic markers published were RFLPs (Haghighi and Hancock, 1992). This study utilized RFLP segregation in blueberry mitochondrial DNA to distinguish diverse highbush cultivars. Since the 1990s, other genetic markers have been analyzed in blueberry including RAPD (Rowland and Levi, 1994), SSR (Levi and Rowland, 1997), and EST-PCR markers (Dhananjay et al., 2004). Focusing on EST-PCR
markers, the first blueberry EST-PCR study examined cold acclimation genes using fruit, flower bud, leaf, and stem tissue (Dhanara et al., 2004). More recently 110 EST-PCR markers have been mapped into 16 genomic linkage groups (Rowland et al., 2012b). EST markers have also been used in evaluating blueberry species including lowbush and rabbiteye (Rowland et al., 2010). Furthermore, EST library development has assisted with finding blueberry SSR markers (Roche et al., 2005; Bassil, 2012). Continued marker studies will help with identification of blueberry cultivars and management of germplasm in gene banks.

**ANNOTATION**

**ANNOTATION PREPARATION**

Following assembly of contigs and scaffolds, various test statistics are used to determine if an assembly is ready for annotation. The most widely used statistic is N50, a scoring metric that describes the length of assembled scaffolds (International Human Genome Sequencing Consortium, 2001). The sequences are sorted according to size and then summed from the largest sequences in decreasing order until half of the total size of the sequences has been tallied. The N50 statistic is the size of the smallest contig or scaffold within this set of the largest sequences. Other statistics, such as percent gaps and percent coverage, ensure that there is not significant missing data (Yandell and Ence, 2012). Though N50 is frequently used, it only describes part of the assembly and has been disputed as an ideal metric for describing assemblies (Baker, 2012). Other metrics based on the N50 have also been proposed (Earl et al., 2011; Mäkinen et al., 2012).

**REPEAT MASKING**

After a sequence has been declared ready for annotation, repeats must be identified and masked. Repeats are short sequences that occur multiple times throughout a genome. Plant genomes contain a high percentage of repetitive DNA. This repetitive DNA can prevent the computer from gathering evidence and correctly assigning sequence locations, which is especially problematic for sections of coding DNA. Many computer programs can identify and mask repeats, such as RepeatMasker (Templ et al., 2012) and RepeatScout (Price et al., 2007), allowing the program to ignore the repeat (Bao and Eddy, 2002). Once the repeats are masked, the process of gene annotation commences.

**ANNOTATION PROCESS**

There are two types of gene annotation: ab initio and evidence-driven. Ab initio gene prediction uses computer-driven mathematical models to identify putative genes and determine their intron-exon structures. These predictions can be advantageous, as they do not require external evidence, saving time and money. Unfortunately, this approach caps the prediction’s accuracy at 70%. Evidence-driven gene annotation, on the other hand, uses data obtained from further analysis. This additional data, which can include gene expression using the transcriptome or ESTs, protein isolation, or experimental evidence based on cloning and characterization, results in a more accurate approach to gene prediction. This accuracy, however, comes at a price, being more costly and time intensive (Yandell and Ence, 2012). For a less expensive and time-consuming method, automated tools such as BLAST (Basic Local Alignment Search Tool) can query an input of known genes against the unknown genome and locate similarities between the two samples (Altschul et al., 1990). These results must be individually examined and interpreted in light of additional ab initio and evidence-based predictions to obtain the final genome annotation.

**PATHWAY EULSIDATION**

Pathway elucidation, which involves building gene expression pathways and discovering the identities of biochemical compounds, can provide insights into the biosynthesis of underinvestigated natural products. Specifically, many studies analyze transcriptomics and gene expression to characterize biosynthesis pathways. Following the introduction of stress to an organism, random mutations arise, potentially altering the natural processes and genetic functions. These conditional stress factors often test a plant’s ability to tolerate adverse conditions. By examining the alterations in the organisms, scientists can take a “top-down” or “bottom-up” genetics approach by studying phenotype to DNA or DNA to phenotype respectively (Flehn, 2001). This strategy and others, such as target analysis, profiling, finger-printing, and metabolomics of ESTs and RNA-seq data, enable deeper insight into an organism’s biomolecular pathways (Hirisi and Satô, 2004).

**BLUEBERRY PATHWAYS**

For blueberries, the anthocyanin and flavonoid biosynthesis pathways are a primary focus of research due to their health and anticarcinogenic properties. A recent study produced 1,000 transcripts and 800 transcription factors relating to antioxidant biosynthesis and identified 90 expressed genes involved in anthocyanin metabolism regulation (Li et al., 2012). A similar study examined the flavonoid biosynthesis pathway (Zikin et al., 2012). These studies are a first step toward characterizing important metabolic pathways, but they are centered on Arabidopsis thaliana, which is a distant relative to the blueberry, for their annotations. This distance results in a lack of conservation that could prevent full elucidation of the metabolic pathways. Future investigations should place focus on a closer relative to the blueberry, such as the grape (Vitis vinifera), and strive for full elucidation of the metabolic network interactions.

**BLUEBERRY GENOME**

Dr. Allan Brown of North Carolina State University is currently working on a draft genome of diploid V. corymbosum. Dr. Brown’s de novo approach utilizes both Roche 454 and Illumina GAIIx libraries. The working assembly consists of approximately 500 million base pairs, which encode 25,000 genes. Though the genome is not yet published, these recent developments will be valuable for understanding blueberry cultivation, pathways, and nutritional value (Die and Rowland, 2013).

**CONCLUSION**

As the disciplines of genomics, bioinformatics, and proteomics evolve, clear introductions can quickly become convoluted. For novices interested in these fields, a baseline understanding of key
concepts is essential. This background will enable these individuals to explore the currently relevant subjects, such as the sequencing of organisms or the elucidation of pathways. In fact, much is still unknown about many currently published genomes. Even the most complete and understood genome, the human genome, is only 10% characterized (Maher, 2012). Additionally, there are many organisms that have not yet been investigated. Examination of new organisms will yield breakthroughs and game-changing discoveries. Of the many organisms with a paucity of research, the blueberry is especially interesting, as scientists and consumers are conscious of its health benefits. Further exploration into the blueberry genome will allow researchers to relate genotype to phenotype, thus providing scientists and farmers with the necessary knowledge to produce blueberries that are more nutritious and desirable.

ACKNOWLEDGMENTS
This work was carried out as part of the Plant Pathways Elucidation Project using resources from a collaboration of the Cabarrus Economic Development Corporation, Catawba College, David H. Murdock Research Institute, Duke Nutrition Research Laboratory, Duke Energy Foundation, General Mills, North Carolina Research Campus, North Carolina State University Plants for Human Health Institute, Turner Construction, University of North Carolina at Charlotte Bioinformatics Services Division, and UNC General Administration.

AUTHOR CONTRIBUTIONS
Austin B. Mudd, Elizabeth J. White, Michael P. Bollolos, Nicholas P. Kapur, Koyt W. Everhart, and Ying-Chen Lin contributed to the research, writing, and revision of this paper. Ryan H. Brown oversaw the revision process as the paper’s corresponding author. Weston W. Buehler and Robert W. Reid provided critical reviews in the revision process and helped rewrite subsections of the paper. Austin B. Mudd spurred the initial idea behind the paper and led to the overall process.

REFERENCES