DISSERTATION

# METABOLOMIC PROFILES OF *ORYZA SATIVA* AND INFLUENCE OF GENETIC DIVERSITY

Submitted by

Adam Lawrence Heuberger

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2011

Doctoral Committee:

Advisor: Mark Brick
Co-advisor: Jan Leach
Co-advisor: Elizabeth Ryan

Patrick Byrne
Henry Thompson

ABSTRACT

METABOLOMIC PROFILES OF *ORYZA SATIVA* AND INFLUENCE OF GENETIC DIVERSITY

Food crops with enhanced health characteristics are being developed in many breeding programs. Rice (*Oryza sativa* L.) is an ideal candidate to study traits related to health due to its importance as both a global staple food and a model system for cereal crops. Evaluating metabolite profiles can be a high-throughput method to identify variation in health properties of dietary components. Metabolomics is a useful tool to assess the influence of genetics on total metabolite variation in the cooked grain. Cooked rice metabolite profiles for 10 diverse varieties were determined using ultra performance liquid chromatography coupled to mass spectrometry (UPLC-MS) on aqueous-methanol extracts. A total of 3,097 molecular features were detected, and 25% of the features varied among the 10 varieties (ANOVA, $p < 0.001$). Both z-score and partial least squares-discriminant analysis (PLS-DA) showed variation consistent with subspecies-based varietal groupings, and indicated genetic control over the metabolite profiles. Variation in total phenolics and vitamin E was also consistent with varietal groupings. Genes in biochemical pathways for health-related metabolites were interrogated for allelic variation by single nucleotide polymorphisms (SNPs). SNP variation may serve as an important mechanism

by which genes influence metabolic variation. The influence of genetic diversity on the metabolite profile of the rice grain was also assessed for two interacting effects: genotype-environment interactions (GEI) and genotype-fermentation interactions (GFI). GEI was assessed by growing two diverse rice varieties in the field and the greenhouse. Gas-chromatography-MS (GC-MS) was used to detect primary metabolites from aqueous-methanol extracts of cooked rice. Genotype, environmental, and GEI effects were observed for many metabolites, including the amino acid phenylalanine, a precursor for many secondary metabolites related to human health. Genes associated with phenylalanine synthesis were screened in rice gene expression databases, and variation within and among the genes suggests they are a potential source of genetic variation for phenylalanine synthesis. Both the metabolite and gene expression patterns indicate a potential interaction between phenylalanine and serine synthesis. The GC-MS data implies the GEI effects on primary metabolism may correspond to variation in secondary metabolites that are predicted to affect human health. Additionally, human health attributes of the grain may be dependent on fermentation of rice metabolites by gut microorganisms. GFI effects were assessed by fermenting three highly similar rice varieties with *Saccharomyces boulardii*, a probiotic yeast. Metabolites were extracted and detected by GC-MS. A PLS-DA model showed evidence of fermentation (F) effects, but not GFI. However, when extracts were assessed for the ability to inhibit viability of lymphoma cells, both F and GFI effects were apparent. It is therefore likely that GFI effects may exist among diverse rice varieties, and that interactions affect the bioactivity of rice

metabolites. In summary, total metabolite variation is largely influenced by the rice genotype, including interactions with environment and fermentation. These data describe both heritable and non-heritable sources of variation. Thus, although genetic variation in rice is sufficient to establish metabolite profiles specific to human health characteristics, the heritability of a secondary metabolite-associated health trait is likely influenced by both environment and fermentation effects.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank everyone who has helped me along the way.

A special thanks to my advisors: Drs. Mark Brick, Jan Leach, and Elizabeth Ryan. Thank you for providing me the opportunity to learn so much science and grow so much as a person. I would also like to thank my committee members: Drs. Patrick Byrne and Henry Thompson. Thank you for the support, teachings, criticisms, and expertise that helped keep the project moving forward.

Thanks to the Proteomics and Metabolomics Facility at CSU for providing an important educational element to metabolomics.

Thanks to the Soil & Crop Science graduate students, whose camaraderie and scholarship was truly special.

And finally, thanks to my family, for their emotional support, to my animals, who proved an excellent audience for practice talks, and to my wife Courtney, who was there every step along the way, and it just wouldn't be the same without her.

TABLE OF CONTENTS

# CHAPTER ONE: INTRODUCTION

## Importance of diversity in rice for humanity

Rice (*Oryza sativa* L.) is an annual cereal crop domesticated over 10,000 years ago in Southeast Asia from the progenitor *Oryza rufipogon*. The three major rice varietal groups, also referred to as “subspecies,” form the major unit of genetic, physiological, and morphological diversity in rice. The distinction between *indica* and *japonica* subspecies occurred several thousand years ago through divergent evolution [1,2]. The *aus* subspecies is a derivative of *indica* and is grown during the pre-monsoon season in India and Bangladesh.

Within-subspecies diversity is attributed to breeding varieties adapted to discrete environments. Currently, rice is grown in both temperate and tropical climates across six continents. There are three major types of rice: irrigated (“paddy”), rainfed (“upland” or “lowland” type) and deepwater (“floating”) [3]. Rice growers choose varieties based on availability of natural rainfall, freshwater, the shape of the land, season of production (wet or dry), latitude, and presence of abiotic and biotic stresses. Today, there are greater than 100,000 rice accessions that highly differ in plant physiology and morphology (International Rice Research Institute, personal communication). Upon integrating molecular biology and plant breeding, thousands of genes, genetic loci, and genetic variants have been identified that control phenotypic diversity in rice.

The wide diversity in rice phenotypes may include variation in health properties of the grain. It is an excellent staple food as a reliable source of calories. Rice has well-

established methods for production under varying environmental conditions, and milled rice can be stored at room temperature for long periods of time. The most prominent nutrients in rice are starch, protein, lipids, free sugars, and several vitamins and minerals. In rice, differences in macronutrient content among varieties may relate to differences in health properties of the grain. For example, rice starch is either amylose or amylopectin, and the ratio varies based on genotype. “Waxy” rice varieties contain mostly amylopectin, which results in an opaque grain that is sticky and quickly digested. High-amylose varieties are less sticky and require more time to digest. The ratio of amylose:amylopectin is highly variable and controlled by a small number of genes [4,5], but is also affected by environmental growing conditions [6]. Rice varieties also differ in protein and lipid content of the grain [7,8]. In addition to basic nutrients, the grain contains a suite of “bioactive compounds,” which are defined as chemicals with no apparent nutritional value, but that affect human health [9]. The influence of plant genetic variation on diversity in bioactive compounds is not well understood. This study will evaluate the effects of genotype, environment, and fermentation on bioactive compounds in rice.

**Rice as a model system for cereals**

Rice is unique as both an important staple food and a model system for cereal genetics. Model plant species require relatively simple, sequenced, and annotated genomes. Rice is a diploid organism with 12 chromosomes and 430 Mb of DNA. Rice has a sequenced and well-annotated genome. The genome for wheat (*Triticum spp.*), another staple food, is a hexaploid and can have 16,000 Mb of DNA. Rice and wheat contain a similar number of genes (approximately 40-50,000) and their genomes are syntenic [10]. Approximately 50% of rice DNA is repetitive, predominantly due to short repeats and transposable elements, and there is evidence of a genomic duplication and

reduction event in its evolution [11]. Additionally, well-established methods are used to transform rice for functional genetic analyses, and rice is ideal for production in both growth chambers and greenhouses. Rice has a relatively quick growth cycle, does not require vernalization, and many molecular and bioinformatics resources are available for gene-based investigations. Thus, rice contains the genetic and bioinformatic tools necessary for determining the influence of genetics on the wide diversity of traits.

**Enhancing the health properties of the rice grain is an important goal for breeding programs**

Developing plant foods with enhanced health characteristics is a breeding target for many staple foods, including rice, wheat, corn (*Zea mays* L. ssp. *mays*), beans (*Phaseolus vulgaris* L.), and potatoes (*Solanum tuberosum* L.). The first health-focused breeding efforts emphasized macronutrient content, such as “high-oil” or “high-lysine” crops. Traditional breeding methods were successful in altering nutrient content, as seen in the Illinois long-term corn experiment that created varieties with higher and lower oil and protein content [12]. The amino acid lysine is the most limiting nutrient in rice, and a number of efforts sought to improve its quantity via molecular techniques [13,14]. Unlike macronutrient-based breeding programs, micronutrient efforts focus on preventing deficiencies. In rice, there are global efforts to improve iron, zinc, and vitamin A content in the cooked grain [15].

Breeding for enhanced health traits is an important goal for both breeders and nutritionists. Diet is important in the development of many chronic diseases, such as with obesity, heart disease, cancer, and diabetes [16]. Staple foods with disease-prevention properties are ideal because they are widely consumed. However, the extent of the preventative effects may differ based on the variety/cultivar of crop being consumed. For example, dry beans show an effect on breast cancer development in a rat model,

however the extent of prevention was cultivar-dependent [17]. For rice, the variation in amylose and amylopectin may differentially promote or protect against the development of diabetes, as the two types of starch vary in glycemic and insulin indices [18]. Furthermore, while variation in health-related properties has been documented, the extent of variation and heritability of the disease-prevention traits is largely unknown. The extent of metabolite variation has been investigated in uncooked rice [19], however temperature and pressure upon cooking may alter the final metabolite content in the diet. The present study will assess variation induced by natural genetic diversity in the grain of cooked rice, variation induced by environment, and variation in rice bran induced by fermentation with a probiotic.

**Genetic diversity in rice may relate to variation in health characteristics**

Rice grain is predicted to vary in many metabolites related to health. Many dietary bioactive compounds produced by the plants are secondary metabolites. Secondary metabolites, unlike primary metabolites, are chemical compounds produced by plants for functions not associated with primary metabolism. For example, phenylpropanoids can counteract ultraviolet radiation [20], and terpenoids can protect against predatory herbivores [21]. Both phenylpropanoids and terpenoids are also important for human health [22,23,24]. While rice contains diversity in macro- and micronutrients, variation in secondary metabolite content has been described for only a few compounds. Unlike brown rice, red and purple rice are high in anothocyanin content and the brown/red phenotype is controlled by allelic differences in a single gene [25]. Like anthocyanins, many studies emphasize the importance of a single plant metabolite (or a single class of metabolites) to health, however they fail to address the evidence for synergistic, rather than independent action [26,27]. Thus, there is a need to establish

profiles of grain metabolites for different rice varieties to breed for more than one metabolite at a time.

**Genetic-environment interactions (GEI) may influence cooked grain metabolite content and reduce the heritability of health traits**

Heritable phenotypes are at the heart of all plant breeding programs. Heritability (narrow-sense) is defined as the proportion of phenotypic variation (0-100%) that is both (i) influenced by genetic factors and (ii) transmitted from parent to offspring. An example non-genetic factor in rice is the growing environment, whereby fertilizer application can increase the protein content in the grain [28], and this response is not inherited to succeeding plant generations. The interaction between a plant genotype and its growing conditions ("genotype-environment interactions," GEI) is a common source of non-heritable variation. GEI are defined as the change in response of genotypes across diverse environments, and some genotypes do better than others in specific environments. Genetic and environmental effects have been widely observed in cereal grains. A recent study implicated genotype as having significant effects on grain metabolite diversity in uncooked rice [29]. Another study assessed uncooked corn metabolite profiles from approximately 100 varieties grown in three environments and identified metabolite variation [30], but did not assess GEI effects. The effect of GEI on rice grain metabolite quantities is unclear, and GEI may affect the heritability of health-related metabolites. Breeding efforts that seek to improve health traits, including bioactive compounds, may observe metabolic phenotypes with low heritability due to interactions between environmental and genotypic effects.

Evaluating the variation in primary metabolism is a critical step in determining the heritability of bioactive compounds. Regulatory networks for primary metabolites, such as negative or positive feedback loops, can influence the formation of secondary

metabolites [31]. Furthermore, variation among secondary metabolites is highly complex, and most of the potential molecular variants within a chemical class are not well characterized. Thus, the genetic control over secondary metabolism is difficult to assess. Primary metabolism may be better to predict variation in total quantities of secondary metabolites, as well as identifying key genetic regulatory processes. This study evaluated potential health properties of bioactive compounds in the cooked grain by assessing GEI effects on primary metabolites.

**Genetic-fermentation interactions (GFI) are likely to influence the health properties of rice**

Gut microbes were recently identified as central players in health. They are predominantly situated in the colon and metabolize dietary components that bypassed absorption in the small intestine. Colonic microbes treat these metabolites as an energy source, a process referred to as fermentation. Byproducts of microbial fermentation, short chain fatty acids (SCFAs), provide another opportunity to access nutrients from the diet. SCFAs may also induce cancer apoptosis and reduce colonic inflammation [32,33,34]. Gut fermentation is also associated with the presence of metabolites in plasma and urine, including varying quantities of amino acids, creatinine, citrate, plasma lipoproteins, and plasma glucose [35,36]. Therefore, fermentation of rice compounds may alter the metabolite profile absorbed into the circulatory system.

Rice contains components that may interact with gut microbes. Metabolites that stimulate colony multiplication in microorganisms are referred to as “prebiotics”. The endosperm also contains prebiotics, such as incompletely digested starches that are subsequently fermented by microbes in the large intestine and colon. Both forms of starch, amylose and amylopectin, act as prebiotics [37]. Unlike starch, dietary fibers (both soluble and insoluble) can partially evade human digestion and are fermented by

gut microbes [38,39]. Gut microbes break down insoluble components of rice bran, including cellulose, hemicelluloses, lignin, and arabinoxylan [39]. Bacteria that specifically ferment dietary insoluble fibers have been observed in the human colon [40,41]. Whole foods that are a complex mixture of fibers, such as brown rice, also show prebiotic effects [42,43]. Rice bran contains both insoluble and soluble fiber, and is thus likely to be fermented by gut microorganisms.

Non-carbohydrate components may also act as prebiotics. Ferulic acid is often found in plant cell walls and is covalently bound to lignin, and therefore may directly affect the growth of gut microbes upon fermentation of plant cell wall components. Phenolic metabolites can be metabolized by gut microbes and some phenolics display antimicrobial effects [44]. Interestingly, pathogenic bacteria are more sensitive to the antimicrobial effect of phenolics than beneficial microbes [45,46,47]. Therefore, dietary phenolics may influence the gut microbial community to contain more beneficial bacteria.

Because of the large amount of starch in the rice endosperm, and the large amount of fiber in the bran layer, it is likely that rice serves as an excellent dietary source of prebiotics. As with bioactive compounds, the large genetic diversity in rice is likely to translate to variation in the ability to act as a prebiotic. Rice may differentially stimulate microbial growth, and further alter the metabolite content exposed to the human body. This is another influence of genetic diversity on metabolites associated with human health. This is referred to as a genotype-fermentation interaction (GFI), whereby the chemical variation among rice varieties is differentially fermented by gut microbes to result in a metabolite profile with altered bioactivity. This report uses a probiotic and *in vitro* cancer cell culture system as a model to determine the effects of fermentation on rice metabolite variation, and thus can estimate the heritability of the health phenotype by measuring bioactivity of the extracts.

**Research Objectives**

It is hypothesized that genotype controls a substantial proportion of variation in the rice grain. A second hypothesis is that variation in the health properties of the rice grain can be affected by interactions between the rice genotype and environmental and fermentation effects. The influence of genetic diversity was assessed as a single effect (Study 1), as well as with environmental and fermentation interactions (Study 2 and 3). These three studies collectively demonstrate the influence of genetics on the metabolite profile of the rice grain, and further show potential associations to health properties of the cooked grain.

All three studies used metabolomics as a method to assess chemical diversity in rice. Metabolomics is a high-throughput chemical detection technique that couples chromatography, mass spectrometry (MS), and statistical analyses to determine qualitative and quantitative metabolite variation among biological tissues. The two forms of chromatography used in this study were liquid chromatography (LC) and gas chromatography (GC). LC-MS excels at characterizing secondary metabolites. LC columns are excellent at retaining more nonpolar compounds, including flavonoids, phenylpropanoids, terpenoids, anthocyanins, and lipids. GC is more sensitive at detecting small, polar metabolites involved in primary metabolism. This study used metabolomics to evaluate the effect of genetic diversity on rice metabolite profiles of the cooked grain (Study 1), GEI effects on cooked rice primary metabolites (Study 2) and GFI effects with rice bran (Study 3).

In Study 1, LC-MS was conducted on aqueous-methanol extracts of lyophilized cooked rice. Ten varieties were chosen that span all three varietal groups (subspecies), and also contain morphological and physiological diversity in the grain, center of origin, growth habits, and other aspects related to metabolism. For the ten varieties, metabolite profiles were evaluated for two types of bioactive compounds: phenolics and vitamin E.

The metabolite data was integrated with genetic information to determine the influence of genotype on the metabolite profile.

In Study 2, two varieties were grown in two environments to estimate the potential for GEI to alter the heritability of bioactive compounds. Two varieties were chosen based on large differences in metabolite profiles and based on data from Study 1. The varieties were grown in two highly different environments, the field and the greenhouse, and cooked rice was assessed for GEI effects on primary metabolites. Amino acids were evaluated in the context of a genetic network to determine genes that may interact with phenylalanine, a precursor to secondary metabolite synthesis. The data was then supported by assessing co-expression information from rice-specific RNA databases for the interacting genes.

In Study 3, three rice varieties were screened for GFI effects on the metabolite profiles and in an assay that evaluates bioactivity of the rice extracts on cancer cells. The metabolite profile was determined using a modified extraction technique to detect both primary and secondary metabolites by GC-MS. Only the bran layer was used to concentrate the bioactive compounds, which are mostly bran components, and to better represent the plant material exposed to gut microbes. Rice bran was fermented by *Saccharomyces boulardii*, a probiotic yeast with well-characterized health benefits. Both non-fermented and fermented rice bran extracts were assessed for an effect on lymphoma cell viability, and GFI effects were observed by evaluating lymphoma cell viability for each of the three varieties.

These studies explain the influence of genetic diversity on metabolites in the rice grain, with an emphasis on metabolites related to health. There are decades of evidence for beneficial effects of functional foods and bioactive compounds. Thus, the potential to prevent disease lies beyond the basic supply of macro- and micronutrients, but more specifically the type (and potentially the quantity) of phytochemicals or metabolites in the

diet. Incorporating bioactive-compound based health traits into rice breeding programs will require a better understanding of the genetic control and heritability underlying the production of these compounds in the grain.

## REFERENCES


1. Chang T-T, Bardenas EA, AC DR (1965) The morphology and varietal characteristics of the rice plant. IRRI Tech Bull 4: 1-40.
2. Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schaal BA (2006) Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. Proceedings of the National Academy of Sciences 103: 9578-9583.
3. Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35: 25-34.
4. Sano Y (1984) Differential regulation of waxy gene expression in rice endosperm. Theoretical and Applied Genetics 68: 467-473.
5. Wang Z-Y, Zheng F-Q, Shen G-Z, Gao J-P, Snustad DP, et al. (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. The Plant Journal 7: 613-622.
6. Chen M-H, Bergman C, Pinson S, Fjellstrom R (2008) Waxy gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. Journal of Cereal Science 47: 536-545.
7. Juliano BO, Antonio AA, Esmama BV (1973) Effects of protein content on the distribution and properties of rice protein. Journal of the Science of Food and Agriculture 24: 295-306.
8. Goffman F, Pinson S, Bergman C (2003) Genetic diversity for lipid content and fatty acid profile in rice bran. Journal of the American Oil Chemists' Society 80: 485-490.
9. Kris-Etherton PM, Hecker KD, Bonanome A, Coval SM, Binkoski AE, et al. (2002) Bioactive compounds in foods: their role in the prevention of cardiovascular disease and cancer. The American Journal of Medicine 113: 71-88.
10. Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, et al. (2003) Comparative DNA Sequence Analysis of Wheat and Rice Genomes. Genome Research 13: 1818-1827.
11. Kurata N, Nonomura K-I, Harushima Y (2002) Rice Genome Organization: the Centromere and Genome Interactions. Annals of Botany 90: 427-435.
12. Dudley J, Lambert R (2004) 100 generations of selection for oil and protein in corn. Plant Breeding Reviews 24: 79-110.
13. Schaeffer GW, Sharpe FT (1987) Increased Lysine and Seed Storage Protein in Rice Plants Recovered from Calli Selected with Inhibitory Levels of Lysine plus Threonine and S-(2-Aminoethyl)cysteine. Plant Physiology 84: 509-515.
14. Kumamaru T, Sato H, Satoh H (1997) High-lysine mutants of rice, Oryza sativa L. Plant Breeding 116: 245-249.
15. Virk P, G B (2009) Biofortified rice—towards combating human micronutrient deficiencies. http://kmfaoorg/.
16. Anon (2002) Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases. WHO technical

report series, Geneva, Switzerland 916.
17. Thompson MD, Brick MA, McGinley JN, Thompson HJ (2009) Chemical Composition and Mammary Cancer Inhibitory Activity of Dry Bean. Crop Sci 49: 179-186.
18. Kabir M, Rizkalla SW, Champ M, Luo J, Boillot J, et al. (1998) Dietary Amylose-Amylopectin Starch Content Affects Glucose and Lipid Metabolism in Adipocytes of Normal and Diabetic Rats. The Journal of Nutrition 128: 35-43.
19. Storck CR, Silva LPd, Alves Fagundes CA (2005) Categorizing rice cultivars based on differences in chemical composition. Journal of Food Composition and Analysis 18: 333-341.
20. Warpeha KM, Gibbons J, Carol A, Slusser J, Tree R, et al. (2008) Adequate phenylalanine synthesis mediated by G protein is critical for protection from UV radiation damage in young etiolated Arabidopsis thaliana seedlings. Plant, Cell & Environment 31: 1756-1770.
21. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3: 408-414.
22. Hollman PCH, Katan MB (1997) Absorption, metabolism and health effects of dietary flavonoids in man. Biomedecine & Pharmacotherapy 51: 305-310.
23. Grassmann J, Hippeli S, Elstner EF (2002) Plant's defence and its benefits for animals and medicine: role of phenolics and terpenoids in avoiding oxygen stress. Plant Physiol Biochem 40: 471-478.
24. Lin XB, Ma L, Racette SB, Spearie CLA, Ostlund RE (2009) Phytosterol glycosides reduce cholesterol absorption in humans. Am J Physiol Gastrointest Liver Physiol 296: G931-G935.
25. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught Red-Handed: Rc Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. The Plant Cell Online 18: 283-294.
26. Liu RH (2003) Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. American Journal of Clinical Nutrition 78: 517S-520S.
27. Lila MA (2007) From Beans to Berries and Beyond. Annals of the New York Academy of Sciences 1114: 372-380.
28. Vaughan DA, Womack M, Smith RT, Wiser WJ (1980) Protein quality of rice as affected by application of nitrogen fertilizer. Journal of Agricultural and Food Chemistry 28: 894-896.
29. Mochida K, Furuta T, Ebana K, Shinozaki K, Kikuchi J (2009) Correlation exploration of metabolic and genomic diversity in rice. BMC Genomics 10.
30. Skogerson K, Harrigan GG, Reynolds TL, Halls SC, Ruebelt M, et al. (2010) Impact of Genetics and Environment on the Metabolite Composition of Maize Grain. J Agric Food Chem 58: 3600-3610.
31. Freeman M (2000) Feedback control of intercellular signalling in development. Nature 408: 313-319.
32. Wang YF, Chen NS, Chung YP, Chang LH, Chiou YH, et al. (2006) Sodium butyrate induces apoptosis and cell cycle arrest in primary effusion lymphoma cells independently of oxidative stress and p21(CIP1/WAF1) induction. Molecular and Cellular Biochemistry 285: 51-59.
33. Kim J, Park H, Im JY, Choi WS, Kin HS (2007) Sodium butyrate regulates androgen receptor expression and cell cycle arrest in human prostate cancer cells. Anticancer Research 27: 3285-3292.

34. Inan MS, Rasoulpour RJ, Yin L, Hubbard AK, Rosenberg DW, et al. (2000) The luminal short-chain fatty acid butyrate modulates NF-kappa B activity in a human colonic epithelial cell line. Gastroenterology 118: 724-734.
35. Li M, Wang BH, Zhang MH, Rantalainen M, Wang SY, et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. Proceedings of the National Academy of Sciences of the United States of America 105: 2117-2122.
36. Martin FPJ, Wang Y, Sprenger N, Yap IKS, Lundstedt T, et al. (2008) Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model. Molecular Systems Biology 4.
37. Wang X, Conway PL, Brown IL, Evans AJ (1999) In vitro utilization of amylopectin and high-amylose maize (amylomaize) starch granules by human colonic bacteria. Applied and Environmental Microbiology 65: 4848-4854.
38. Topping DL (1991) Soluble fiber polysacchardies - effects on plasma-cholesterol and colonic fermentation. Nutrition Reviews 49: 195-203.
39. Flint HJ, Bayer EA (2008) Plant cell wall breakdown by anaerobic microorganisms from the mammalian digestive tract. In: Wiegel J, Maier RJ, Adams MWW, editors. Incredible Anaerobes: from Physiology to Genomics to Fuels. pp. 280-288.
40. Robert C, Bernalier-Donadille A (2003) The cellulolytic microflora of the human colon: evidence of microcrystalline cellulose-degrading bacteria in methane-excreting subjects. Fems Microbiology Ecology 46: 81-89.
41. Wedekind KJ, Mansfield HR, Montgomery L (1988) Enumeration and isolation of cellulolytic and hemicellulolytic bacteria from human feces. Applied and Environmental Microbiology 54: 1530-1535.
42. Sunvold GD, Hussein HS, Fahey GC, Merchen NR, Reinhart GA (1995) In vitro fermentation of cellulose, beet pulp, citrus pulp, and citrus pectin using fecal inoculum from cats, dogs, horses, humans, and pigs and ruminal fluid from cattle. Journal of Animal Science 73: 3639-3648.
43. Bird AR, Hayakawa T, Marsono Y, Gooden JM, Record IR, et al. (2000) Coarse brown rice increases fecal and large bowel short-chain fatty acids and starch but lowers calcium in the large bowel of pigs. J Nutr 130: 1780-1787.
44. Selma MV, Espin JC, Tomas-Barberan FA (2009) Interaction between Phenolics and Gut Microbiota: Role in Human Health. Journal of Agricultural and Food Chemistry 57: 6485-6501.
45. Lee HC, Jenner AM, Low CS, Lee YK (2006) Effect of tea phenolics and their aromatic fecal bacterial metabolites on intestinal microbiota. Research in Microbiology 157: 876-884.
46. Cueva C, Moreno-Arribas MV, Martin-Alvarez PJ, Bills G, Vicente MF, et al. (2010) Antimicrobial activity of phenolic acids against commensal, probiotic and pathogenic bacteria. Research in Microbiology 161: 372-382.
47. Hervert-Hernandez D, Pintado C, Rotger R, Goni I (2009) Stimulatory role of grape pomace polyphenols on Lactobacillus acidophilus growth. International Journal of Food Microbiology 136: 119-122.

# CHAPTER TWO: METABOLOMIC AND FUNCTIONAL GENOMIC ANALYSES REVEAL VARIETAL DIFFERENCES IN BIOACTIVE COMPOUNDS OF COOKED RICE

A modified version of this chapter is published as Heuberger AL, Lewis MR, Chen MH, Brick MA, Leach JE, Ryan EP. Metabolomic and Functional Genomic Analyses Reveal Varietal Differences in Bioactive Compounds of Cooked Rice. PLOS One 2010, 5 (9), e12195.

**SUMMARY**

Emerging evidence supports that cooked rice (*Oryza sativa* L.) contains metabolites with biomedical activities, yet little is known about the genetic diversity that is responsible for metabolite variation and differences in health traits. Metabolites from ten diverse varieties of cooked brown rice were detected using ultra performance liquid chromatography coupled to mass spectrometry. A total of 3,097 compounds were detected, of which 25% differed among the ten varieties. Multivariate analyses of the metabolite profiles showed that the chemical diversity among the varieties cluster according to their defined subspecies classifications: *indica, japonica,* and *aus*. Metabolite-specific genetic diversity in rice was investigated by analyzing a collection of single nucleotide polymorphisms (SNPs) in 91 genes from biochemical pathways of nutritional importance. Two types of bioactive compounds, phenolics and vitamin E, contained nonsynonymous SNPs and SNPs in the 5' and 3' untranslated regions for genes in their biosynthesis pathways. Total phenolics and tocopherol concentrations were determined to examine the effect of the genetic diversity among the ten varieties. Per gram of cooked rice, total phenolics ranged from 113.7 to 392.6 μg (gallic acid equivalents), and total tocopherols ranged between 7.2 and 20.9 μg. The variation in the cooked rice metabolome and quantities of bioactive components supports that the SNP-based genetic diversity influenced nutritional components in rice, and that this approach may guide rice improvement strategies for plant and human health.

## INTRODUCTION

Rice (*Oryza sativa* L.) is a valuable model system for cereal plant genetics due to its sequenced and annotated genome, capacity for transformation, and similarity to other major cereal crop species. Most importantly, rice is a vital source of calories as a food crop. Cereals are the primary source of energy for over 50% of the global population, of which rice is the third largest contributor [1]. The global dependence on rice has led to the development of thousands of varieties with large genetic and morphological diversity. Rice is structured into several well-defined gene pools via the subspecies classification of *indica, japonica,* and *aus*. This classification was recently confirmed with the genome resequencing of 20 representative varieties and subsequent documentation of single nucleotide polymorphisms (SNPs), referred to as the OryzaSNP set [2]. Across and within each classification, rice contains significant diversity in plant architecture and growing habits [3], and in grain phenotypes such as width, weight, cooking properties, aroma, and texture [4]. The extensive phenotypic and genotypic variation within the OryzaSNP set makes these varieties a powerful tool to study rice chemical diversity such that methods can be developed to enhance health promoting qualities of rice.

Metabolites present in the rice grain have demonstrated human disease protective activities following dietary intake, and also have beneficial effects on the immune system [5-7]. Specific rice components, such as phenolics (mono- and polyphenols), vitamin E (tocopherols and tocotrienols), phytosterols, and linolenic acid, have nutrient value to human health [8-11]. Phenolic bioactivity is largely due to the efficiency of donating hydrogen atoms to oxygen radicals [12], a process associated with anticancer activity [13]. Unlike phenolics, tocopherols are lipid-soluble antioxidants incorporated into lipoproteins, and are predicted to counteract the inflammatory effects of lipoprotein oxidation in blood [14]. While brown rice is an efficient source of both

phenolics and tocopherols, little is known regarding the genetic basis for the variation in type and quantity of these components in cooked rice across genetically diverse varieties.

The functional impact of SNP-derived genetic variation in pathways that regulate the production of dietary bioactive compounds in rice is also unclear. Metabolomics, the comprehensive analysis of low-molecular-weight compounds in biological samples, provides a high-throughput and sensitive approach to assess the outcome of different genotypes on metabolites in the cooked grain. New evidence supports the utility of this technique to capture the complexity of the rice metabolome and to evaluate changes in metabolic responses [15,16]. However, there has been minimal integration of the rice metabolomic signature with genomic data sets and the use of this information to assess components of dietary importance. A systems biology approach was applied herein to reveal the synthesis and metabolic regulation of nutritionally important phytochemicals, by profiling multiple rice varieties for pathway-specific SNPs and metabolomics.

## MATERIALS AND METHODS

### Rice materials

Rice seeds of ten OryzaSNP accessions were acquired from the International Rice Research Institute (IRRI, Los Baños, Philippines) and are listed in Table 2.1. Rice plants were grown at the Dale Bumpers National Rice Research Center in Stuttgart, Arkansas to produce seed used in this study. The grain was isolated from the husk using a manual stone dehusker, left unpolished, and then cooked by boiling in a 2:1 volume of water/rice ratio for 15 minutes or until soft. Cooked brown rice was lyophilized over a period of 48 hours immediately after cooking and stored at -80°C until further analysis.

**Rice processing and extractions**

Metabolites in cooked rice were extracted by first grinding rice to a powder with a mortar and pestle in liquid nitrogen. One mL of ice-cold methanol/water (4:1) was added to 100 mg of rice powder. Samples were incubated for one hour at -80°C to precipitate protein, centrifuged at 1500 × g for five minutes at 4 °C, and the supernatant was collected and stored at -20°C until further analysis.

**Ultra Performance Liquid Chromatography-Mass spectrometry**

Rice extract separation was performed using an Acquity UPLC® controlled with MassLynx software, version 4.1 (Waters, Milford, MA, USA). Samples were held at 8°C in a sample manager during the analysis to minimize evaporation. The complete sample set was randomized and profiled in two independent iterations. Sample injections of 2 µL were made to a 1.0 x 100 mm Waters Acquity UPLC® BEH C8 column with 1.7 µm particle size held at 40°C. Separation was performed by reverse phase chromatography at a flow rate of 0.14 mL/min. The eluent consisted of water and methanol (Fisher, Optima LC-MS grade) supplemented with formic acid (Fluka, LC-MS grade) in the following proportions: Solvent A = 95:5 water:methanol + 0.1% formic acid; Solvent B = 5:95 water:methanol + 0.1% formic acid. The separation method is described as follows (25 minutes total): 0.1 min hold at 30% B, 1.9 min linear gradient to 70% B, ten min linear gradient to 100% B, 6 min hold at 100% B, 0.1 min linear gradient to 30% B, and 6.9 min hold at 30% B for column equilibration prior to the next injection.

Eluate was directed to a Q-TOF Micro quadrupole orthogonal acceleration time-of-flight mass spectrometer (Waters/MicroMass, Millford, MA, USA) using positive mode electrospray ionization (ESI+). Mass data were collected between 50 and 1000 m/z at a rate of one scan per second. The voltage and temperature parameters were tuned for general profiling as follows: capillary = 3000 V; sample cone = 30 V; extraction cone =

2.0 V; desolvation temperature = 300°C; and source temperature = 130°C. Mass spectral scans were centered in real time producing centroid data. Leucine Enkephalin was infused via a separate orthogonal ESI spray and baffle system (LockMass) which allowed ions to be detected for a single-second scan every ten seconds in an independent data collection channel. The standard mass was averaged across ten scans providing a continuous reference for mass correction of analyte data.

**Allele frequencies**

Allele frequencies were calculated for each SNP site based on the 20 varieties of the OryzaSNP set (www.oryzasnp.org). Base calls for each SNP were determined using TIGR Pseudomolecule v5 in the OryzaSNP database. Frequencies were determined by evaluating the proportion of adenine, guanine, cytosine, and thymine nucleotides among the 20 varieties for each SNP site. Unresolved nucleotides were reported as "N." The number of nonsynonymous SNPs for vitamin E varied depending on the gene model, which are derived by determining all possible combinations of introns and exons. Genes in the vitamin E pathway contained either two, three, or four nonsynonymous SNPs based on different gene models, and all gene models were analyzed for allele frequency calculations.

**SNP Dendrogram**

An unweighted, unrooted neighbor-joining tree with 1000 bootstraps was constructed using DARwin (http://darwin.cirad.fr/darwin). Inputs for each variety consisted of a collection of base calls specific to either the phenolic or vitamin E pathway. For each pathway, SNP sites with greater than 50% unknown nucleotides were not included in the analysis, and varieties with greater than 50% missing information were also removed.

**Total Phenolics Assay**

Total phenolic concentrations in rice extracts were determined as previously described [40] with minor modifications. Briefly, 150 µL of Folin-Ciocalteu reagent/water (1:9) was added to 35 µL of rice metabolite extract and was incubated at room temperature for five minutes. Sodium bicarbonate (115 µL of a 7.5% solution) was then added and samples were incubated at 37°C for 30 minutes. Samples were allowed to cool to room temperature and absorbance was measured at 765 nm. Metabolite extractions were performed in triplicate. Total phenolics were calculated using a standard curve generated using a series of gallic acid concentrations and were expressed as micrograms of gallic acid equivalents (GAE) per gram of rice.

**Vitamin E quantification**

Tocopherol homologs, α-, γ-, and δ-tocopherols, were purchased from Cayman Chemicals (Ann Arbor, MI; ≥ 98% purity). Tocotrienol homologs, α-, γ-, and δ-tocotrienols, were purchased from Matreya Biochemicals (Pleasant Gap, PA; ≥ 97% purity). Methanol and acetonitrile were HPLC grade from Fisher Scientific (Fair Lawn, NJ).

Tocopherols (α-, γ-, and δ-tocopherols) and tocotrienols (α-, γ-, and δ-tocotrienols) were determined using HPLC (Waters, Milford, MA) based on the method described in [41] with modifications. The HPLC was equipped with a Waters 2695 Alliance Separation Module, a Waters 2996 Photodiode array detector (PDA), a Waters 474 Scanning Fluorescence detector, and Empower$^{TM}$ 2 software for data acquisition. The cooked and lyophilized rice powders were extracted with 100% methanol twice at the bran to solvent ratio of 1 to 33 (w/v). For each extraction, the mixture was flushed with nitrogen gas and shaken (300 rpm) for 2h at room temperature. After centrifugation

at 2000 × g for ten minutes at room temperature, the supernatants were pooled and filtered through a 0.45 µm polyvinylidene fluoride (PVDF) membrane (Waters, Milford, MA), injected through a Symmetryshield RP C-18 guard column (3.5 µm, 3.0 x 20 mm; Waters) and separated on a Symmetryshield RP C-18 analytical column (3.5 µm, 3.0 x 150 mm; Waters). The filtrate was eluted with a gradient mobile phase consisting of (A) 100% acetonitrile, (B) 100% methanol, and (C) 1% acetic acid in 50 % methanol at 0.5 mL / min at 25ºC. The gradient was used as follows: 0-1 min, 45% A, 35% B, and 20% C; 1-2 min, linear gradient to 45% A, 45% B, and 10% C; 2-16 min, linear gradient to 30% A, 65% B, and 5% C; 16-20 min, linear gradient to 25% A and 75% B; 20-22 min, linear gradient to 100% B; 22-25.4 min, isocratic at 100% B; 25.4-25.5 min, linear return to 45% A, 35% B, and 20% C; 25.5-35 min, isocratic at 45% A, 35% B, and 20% C to re-equilibrate. The tocopherol and tocotrienol homologs were detected by the fluorescence detector at the excitation and emission wavelengths of 298 and 328 nm, respectively. The peak identification of tocopherols and tocotrienols was performed by comparing their retention time with those of standards. The concentration of each tocopherol and tocotrienol homolog was calculated using the standard curve plotted as peak area against a series of concentrations of each tocopherol and tocotrienol homolog and indicated as µg/g rice. The coefficient of determinations ($R^2$) ranged from 0.9962 to 0.9999. The β- and γ-forms of tocopherols and tocotrienols are isomers and co-elutes on reversed-phase C18 columns. Rice bran contains only trace amounts of β-form, nevertheless, the concentrations of γ-forms of tocopherols and tocotrienols in bran reflect the sum of β- and γ-forms in this study.

**Statistical Analysis**

Chromatographic and spectral UPLC-MS peaks were detected, extracted, and aligned using MarkerLynx software (Waters, Millford, MA, USA). Chromatographic peaks

were detected between 0 and 14 min with a retention time error window of 0.1 min. Apex track peak detection parameters were used, automatically detecting peak width and baseline noise. No smoothing was applied. To reduce the detection and inclusion of noise as data, an intensity threshold value of 40 counts and a noise elimination value of 6 were used. Mass spectral peaks were detected between 50 and 1000 m/z with a mass error window of 0.07 m/z, and the de-isotoping function was enabled. A matrix of features as defined by retention time and mass was generated, and the relative intensity (proportional to quantity) of each feature (metabolite), as determined by area of the peak, was calculated across all samples. Potential effects of instrument variability were minimized by normalizing the total ion current (TIC) among all samples such that the summation of all feature intensities in each sample yielded a constant value. Furthermore, the relative intensity of each feature was averaged over the two replicate injections preformed for each sample to provide a reliable data matrix with minimal technical artifacts. Mean centering was applied, and the data matrix was analyzed in SIMCA-P+ v. 11.5 (Umetrics, Umeå, Sweden). Pareto scaling was applied to the data, and a score plot was generated to describe the data using partial least squares discriminant analysis (PLS-DA). The PLS-DA model was validated by testing new PLS-DA models built from 20 random permutations of the data. Significant UPLC-MS-detected metabolites were determined using a Kruskal-Wallis test on relative intensities of features with identical masses and retentions times with a threshold value of $P < 0.001$ and n=5 replicates per variety. Z-scores were calculated for each metabolite based on the mean and standard deviation of the reference variety Nipponbare. Statistical significance for total phenolics and vitamin E analyses was determined by ANOVA with a Tukey post-test and a threshold value of $P < 0.05$.

## RESULTS

### Metabolite variation in rice varieties and subspecies

A comparison of metabolite profiles was conducted to determine the extent of variation in cooked brown rice across ten varieties from the OryzaSNP set (Table 2.1). The subset of OryzaSNP varieties used in this study represent the extensive phenotypic and genetic diversity present in the three subspecies (*aus, japonica, indica*) of consumed rice varieties [2]. They also represent different levels of improvement through breeding [3]. Metabolites from cooked brown rice were extracted in 80:20 methanol:water and detected by ultra performance liquid chromatography coupled with mass spectrometry (UPLC-MS). A metabolomic profile for each rice variety was resolved as a sum of its features, and each feature (assumed here to be a unique metabolite) consists of a retention time, mass, and quantity. Across the ten varieties, 3,097 metabolites were detected, and these metabolites were distributed across a wide range of molecular masses Figure 2.1A). Approximately 25% (763 out of 3,097) of the metabolites differed in quantity among the ten varieties (Kruskal-Wallis test, $P < 0.001$) (Figure 2.1B). A z-score analysis applied to the set of 763 metabolites showed extensive metabolite variation relative to Nipponbare, a Japanese variety with a sequenced genome (Figure 2.1C). A sum of squares for the 763 z-scores showed that the metabolite profiles of all nine varieties were different from the profile of Nipponbare, and that profiles of *indica* subspecies varieties show larger differences from Nipponbare than did *japonica* subspecies profiles (Table 2.2).

Based on a partial least squares discriminant analysis (PLS-DA), metabolite profiles cluster according to subspecies (*indica, japonica, aus*) (Figure 2.2A). The first component of the PLS-DA model explained approximately 64% of the variation, and the second component explained an additional 35% of the variation. Varieties were then clustered into the *indica, japonica,* and *aus* subspecies, and 194 metabolites were

determined to be significantly different among the three subspecies (Kruskal-Wallis test, $P < 0.001$) (Figure 2.2B). Hierarchical clusters were determined using Euclidian distances, and the metabolite profiles of the *aus* varieties were nearer to the *japonica* than the *indica* varieties. The differences in the chemical profiles among the ten varieties suggest the potential for variation in metabolites important for human nutrition.

**SNP analysis reveals allelic differences in phytochemical pathways of nutritional importance in rice**

Relevant metabolic pathways, including those involved in the biosynthesis of phenolics, vitamin E, phytosterols, and linolenic acid, were chosen for functional genomic analysis of SNPs across the diverse rice varieties. The RiceCyc database (www.pathway.gramene.org/rice) was used to align the four classes of metabolites to biochemical pathways, and then to identify genes from the associated chemical reactions (Table 2.3). Pathways for phenolics combined both phenylpropanoid and flavonoid synthesis due to conservation of structure and function, and also included isoflavone-7-O-methytransferase 9 and leucodelphinidin biosynthesis genes that synthesize tricin, a phenolic unique to rice [17]. Vitamin E genes encode components of the tocopherol and tocotrienol synthesis pathways, which includes α-, β-, γ-, δ-tocopherol and tocotrienol-related enzymes, as well as tocopherol O-methyltransferase and homogentisic acid geranylgeranyl transferase genes [18]. Genes involved in phytosterol synthesis were derived from sterol synthesis pathways [19], and linoleic acid genes were derived from lipid desaturation pathways [20]. Genes were screened for SNPs using the rice OryzaSNP database (www.oryzasnp.org), which classified rice SNPs based on up to four gene models.

SNPs, base calls, and SNP classifications were associated to their respective class of metabolites by cross-referencing locus identifiers to the metabolite pathway

database. A number of SNPs were detected in pathways associated with synthesis of metabolites important to human health (Table 2.4). SNPs in gene pathways responsible for the synthesis of phenolics were evenly distributed among synonymous, nonsynonymous, and intron classes. Phenolics also contained a greater amount of nonsynonymous SNPs per gene than phytosterols, vitamin E, or linolenic acid, and had a higher probability of a change in enzymatic function or regulation. One large-effect SNP is predicted to alter the function of the ferulate 5-hydroxylase enzyme (gene: LOC_Os06g24180) and was classified as potentially altering regular intron splicing events. A larger percentage of SNPs in the sterol, vitamin E, and linoleic acid pathways were within introns compared to phenolics.

To identify the unique nonsynonymous SNPs in our rice collection, allele frequencies were calculated for genes involved in the phenolics and vitamin E biochemical pathways (Figure 2.3). Only two alleles existed for each of the 28 SNPs. Seven SNPs (25%) had one variety that contained its own unique allele. The remaining 21 SNPs (75%) had alleles that were shared among multiple varieties, and the average allele frequency per SNP was 0.52. The subset of 21 SNPs represent rice metabolic pathways that are common to a cluster of varieties rather than solitary occurrences.

**SNP diversity predicted subspecies variation in phenolics and vitamin E content**

To further characterize the genetic control of nutritionally important metabolites, a dissimilarity matrix was constructed using a concatenated sequence of SNPs specific to phenolics or vitamin E pathways. The phytosterol and linolenic acid pathways had low SNP abundance, and therefore low variation (data not shown). Clustering based on SNPs in both phenolic and vitamin E pathways grouped the rice varieties according to the *indica, japonica,* and *aus* subspecies classifications (Figure 2.4A, 1.5A). The total phenolic concentration differed among the ten varieties (Figure 2.4B). The overall mean

total phenolic concentration was 256 µg of gallic acid equivalents (GAE) $g^{-1}$ of cooked rice. The variety Dular had the highest total phenolics with a mean of 393 µg GAE $g^{-1}$ cooked rice. IR64 and Nipponbare had the least amounts with means of 114 and 136 µg GAE, respectively. The mean total phenolics was 179 µg GAE for the *indica*s, 288 µg GAE for the *japonicas*, and 302 µg GAE for the *aus* groups.

The SNP diversity in vitamin E-relevant genes was larger for *indicas* than *japonicas* (Figure 2.5A). For vitamin E, the low mean number of nonsynonymous SNPs per gene predicted high conservation in total rice vitamin E concentration. The ten varieties were analyzed for total tocopherols in the cooked grain, as well as the contribution by each of the main constituents: α-, γ-, and δ- tocopherol. N22 had the lowest levels of total tocopherols at 7.2 µg $g^{-1}$ of cooked rice, and M202 had the highest concentration at 20.9 µg $g^{-1}$ (Figure 5B). Because α-and γ-tocopherols vary in bioactivity, the contribution of α- and γ- to the total tocopherol pool was determined as a ratio of α:γ for each variety (Figure 2.5C). The levels of δ-tocopherol were consistently low and had a negligible contribution to total vitamin E. The ratio of α:γ significantly differed among the ten varieties. The *indica* varieties contained the highest levels of γ-tocopherols with a mean α:γ ratio of 0.75, whereas the *japonica* varieties contained higher levels of α-tocopherols with a mean ratio of 6.6. The variety Dular had the smallest α:γ ratio with a value of 0.27, and the Nipponbare variety had the largest α:γ ratio of 18.8. The tocopherol ratios of the two *aus* varieties (Dular and N22) were very different. None of the SNPs collected in Table 2.3 could directly explain the variation in tocopherol components. SNP diversity was smaller for predicting levels of vitamin E when compared to phenolics, however there was clear variation in the quantity of phenolics, and both the type and quantity of vitamin E metabolites among the ten rice varieties.

## DISCUSSION

The diversity in genetic and morphological rice traits from the OryzaSNP set was interrogated herein by applying metabolomic analysis to the cooked grain. Previous studies have established metabolite profiles for crop varieties [21,22], however metabolites were extracted from raw plant material. The screening of metabolites in cooked rice enhanced the dietary relevance of our findings, as the nutritional differences detected resembles actual metabolite intake following heat and moisture. An open-boiling technique was standardized for this study because of the global utilization of this cooking method.

Recent reviews emphasize the need for sustainable, breeding-based approaches to enhance plant food nutritional quality [23,24]. An integrated genomic and metabolomic method has been proposed as a useful measure to improve food crops [25]. A number of studies successfully correlated genomics with metabolomics, such as in the associations of quantitative trait loci with metabolite profiles in *Arabidopsis* [26] and of restriction fragment length polymorphism markers with nuclear magnetic resonance-generated metabolite profiles in uncooked rice [22]. An analysis of SNPs provides a new functional relevance for the differences detected in the rice metabolome. The integration of SNP-based bioinformatics with metabolomics as conducted herein may now be utilized to assist in selection of rice varieties with enhanced nutritional and health-promoting value.

The extensive metabolite variation in different varieties of cooked rice was approximately 25% of the total metabolites detected. The z-score analysis using Nipponbare as a reference was a compelling example of the metabolite diversity among the varieties (Figure 1.1C). Z-scores were calculated to determine metabolites that vary between one variety and a reference variety. An excessively high or low z-score (roughly higher or lower than five) usually indicated a metabolite present in one variety and

absent in another, and may provide direction in identifying unique metabolites. The sum of squares of the z-scores suggested that the *indica* varieties were more different from Nipponbare than the *japonica* or *aus*, and was expected given that Nipponbare is a *japonica* variety.

Another strong link between the rice genome diversity and cooked rice metabolome was the PLS-DA model that clustered the cooked rice metabolome for each variety according to subspecies (Figure 2.2A). Genomewide, *aus* is more homologous to the *indica* subspecies [2], however N22 (*aus*) grouped closely with the *japonicas* following metabolite analysis with both z-scores and the PLS-DA. The hierarchical clustering of the 763 metabolites that represent total metabolite variation also grouped the *aus* varieties closer to the *japonicas* than the *indicas*. This contrast between observed genomic homology and metabolomic profiles is likely due to introgressions of metabolite-related loci into the *aus* background. Such introgressions are frequent in rice [27], and have been utilized for genetic association strategies to identify loci important for synthesizing trait-specific metabolites in *Arabidopsis* [28] and tomato [29].

The genes in nutritionally important biochemical pathways contained SNP variation (Figure 2.4A, 1.5A) that associated with the UPLC-MS-derived metabolome for cooked rice (Figure 2.2A). SNPs with functionally-relevant classifications were found in genes in the phenolics, vitamin E, phytosterol, and linolenic acid pathways, with a larger mean number of SNPs per gene in the phenolics and vitamin E pathways. The total number of nonsynonymous SNPs may be larger than described in Table 2.4 because many genes and enzymes for key biochemical reactions remain unknown. Furthermore, our SNP analysis was limited to a subset of rice varieties that were diverse but represent a small proportion of the total rice genetic diversity. Additional types of genetic variation, such as insertions, deletions, and translocations, may explain the lack of evidence for individual SNP associations with metabolite variation.

The SNP homology in the phenolic and vitamin E pathways for the ten rice varieties coincided with the *indica, japonica,* and *aus* subspecies classifications. However, the phytosterol and linolenic acid pathways lacked sufficient information to function in a SNP homology-based model. The SNP dendrograms for phenolic and vitamin E related metabolites predicted that a given rice variety will be more similar to a variety of the same subspecies than of another subspecies. It can be postulated that distinct haplotypes for the synthesis and regulation of nutritionally important phytochemicals are present in select rice varieties. Thus, it is plausible that a SNP haplotype was responsible for a given variety's metabolite profile, and that haplotype breeding approaches could be used to optimize the metabolite profiles of rice for nutritionally important health traits.

The total phenolic concentration varied both among and within subspecies. In general, the *japonica* varieties contained a higher level of total phenolics (288 μg GAE $g^{-1}$) than the *indicas* (179 μg GAE $g^{-1}$). However, Nipponbare contained a lower abundance of phenolics (136 μg GAE $g^{-1}$) than its *japonica* counterparts, and Zhenshan appears an *indica*-outlier due to its higher concentration of total phenolics (256 μg GAE $g^{-1}$) than other *indicas*. This was consistent with the z-score analysis, in which Zhenshan contained the largest difference from Nipponbare (Table 2.2). N22 (*aus*) also grouped with the *indicas*, and both the z-score and PLS-DA models grouped N22 closer to Nipponbare (*japonica*) than Dular (*aus*). The solvents used herein are known to extract phenolic compounds from rice [30,31], and therefore a proportion of the variation observed in the z-score and PLS-DA models was likely due to differences in phenolics.

The total quantity of tocopherols per gram of rice showed slight variation among the ten varieties, and the observed range in quantities were similar to those found in various plants and plant tissues [32]. The α and γ forms of tocopherol have different bioactive functions and metabolism [33-35]. The tendency for the α:γ ratio to link a

variety within its subspecies is consistent with the observed trends in both the metabolome (Figure 2.2A) and total phenolics (Figure 2.4B) analyses, however specific varieties also deviate from the larger subspecies trends (Figure 2.5C). For the α:γ ratio, N22 (*aus*) clusters closer to all *japonicas* except for Nipponbare (*japonica*). Dular (*aus*) clusters with the *indicas*, which all contain a lower ratio of α:γ tocopherol than the *japonicas*. SNPs were not able to explain the *indica*/*japonica* division in α:γ ratios, as none were identified in γ-tocopherol-O-methyltransferase (γ-TMT), the enzyme that converts γ- to α-tocopherol by the addition of a methyl group. Enhanced γ-TMT expression has been shown to increase the α:γ ratios in various plants and tissues, but does not alter the overall quantity of tocopherols [36-38]. Thus, the variation among the ten rice varieties may be due to differential γ-TMT gene expression rather than a SNP-driven change in function. Furthermore, the α:γ tocopherol ratios were consistent with observed ratios of tocotrienols (data not shown), which further supports the importance of the γ-TMT in determining the overall composition of vitamin E. SNPs were not identified in the 5' untranslated region of the rice γ-TMT gene, and therefore it is likely that a diverse set of vitamin E gene regulators exists for tocopherol accumulation in rice.

The identification of the genetic basis for important agronomic traits, such as yield and abiotic/biotic stresses has led to considerable advances in accumulating desirable traits into rice breeding programs. The incorporation of nutritional traits, however, has been generally overlooked due to an emphasis on total plant yield [39]. Here, the SNP findings provide evidence for regular, systematic evolution at loci important to nutritional metabolite synthesis. A deeper understanding of the genetic basis for the type and quantity of metabolites in the rice grain may allow for breeding plants that contain an optimal metabolite profile for enhanced health attributes.

## FIGURES

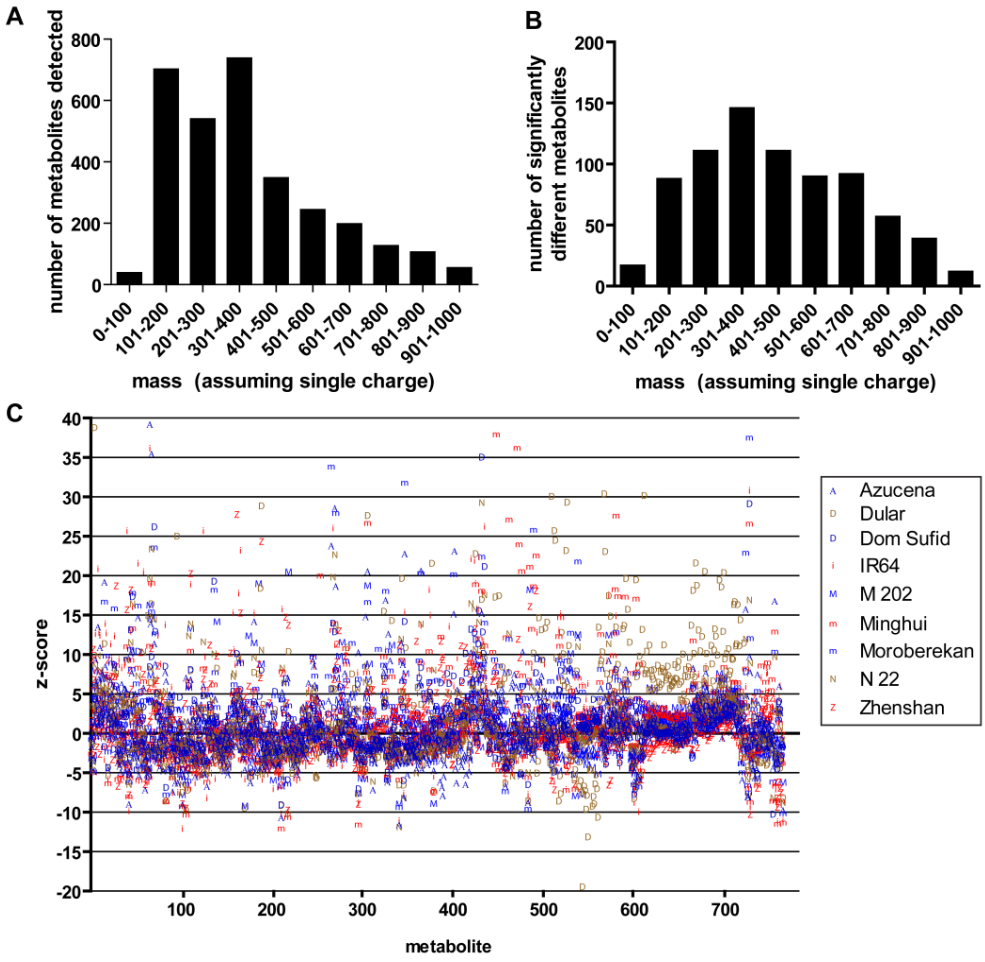


**Figure 1.1. Metabolite detection across ten rice varieties.** (A) Rice metabolites were detected by UPLC-MS and all 3,097 metabolites were sorted by size. (B) The 763 metabolites that differ among the ten varieties were dispersed across a similar mass distribution as the total metabolite profile. (C) Z-score analysis on the 763 metabolites was conducted using Nipponbare (*japonica*) as a reference. *Indica*, *japonica*, and *aus* varieties are shown in red, blue, and brown, respectively. A total of 32 data points with a z-score of greater than 40 were outside of the area shown.

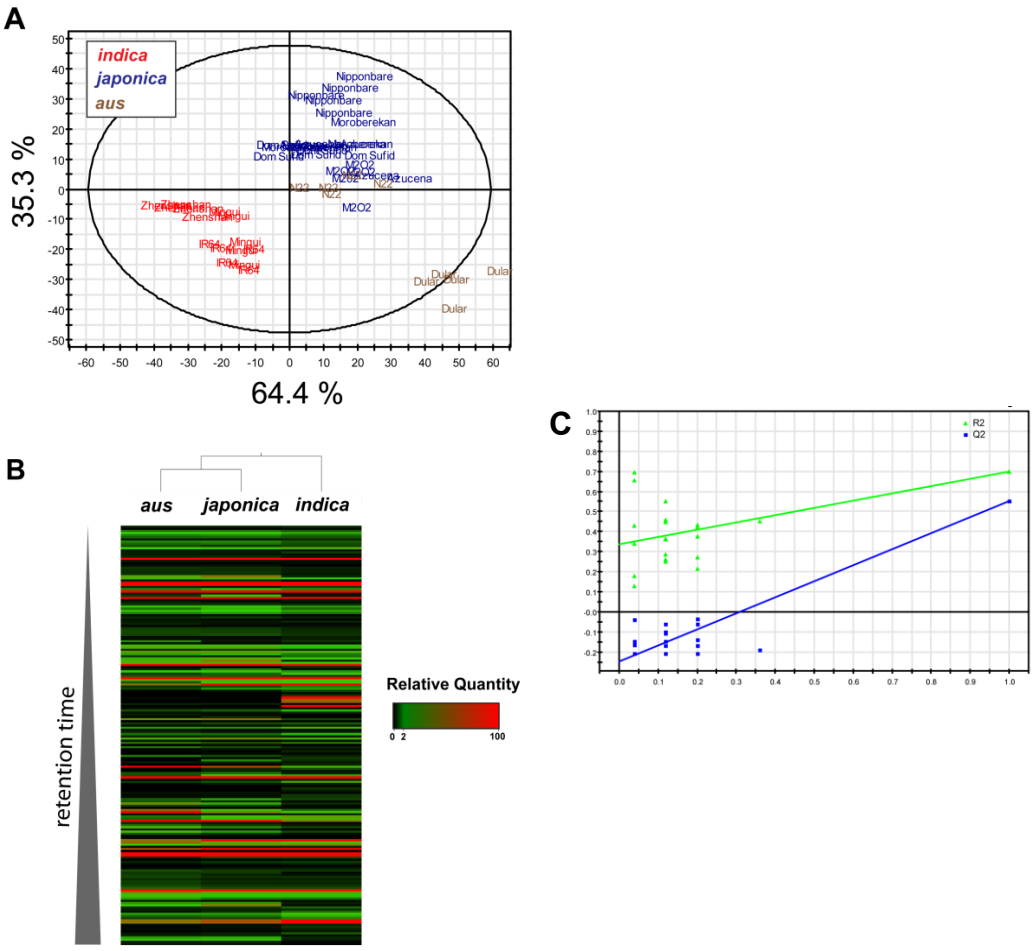


**Figure 2.2. Subspecies analysis of the cooked rice metabolome.** (A) PLS-discriminant analysis was conducted on ten rice varieties and was colored according to subspecies as *indica* (red), *japonica* (blue), and *aus* (brown). (B) The 194 metabolites that differ among the three subspecies were shown in a heat map whereby each cell represents a single metabolite. Metabolites were arranged according to retention time (0.5-12 minutes), and colors indicate relative quantities. Hierarchical clustering was performed using Euclidean distances. (C) Validation of the partial least squares discriminant analysis. The PLS-DA model for subspecies was validated using 20 permutations. Values for $R^2$ (0.7) and $Q^2$ (0.55) denote original and predictive data, respectively. A positive value of $Q^2$ when $R^2$ is zeros (x-axis=0) would suggest overfit in the model.

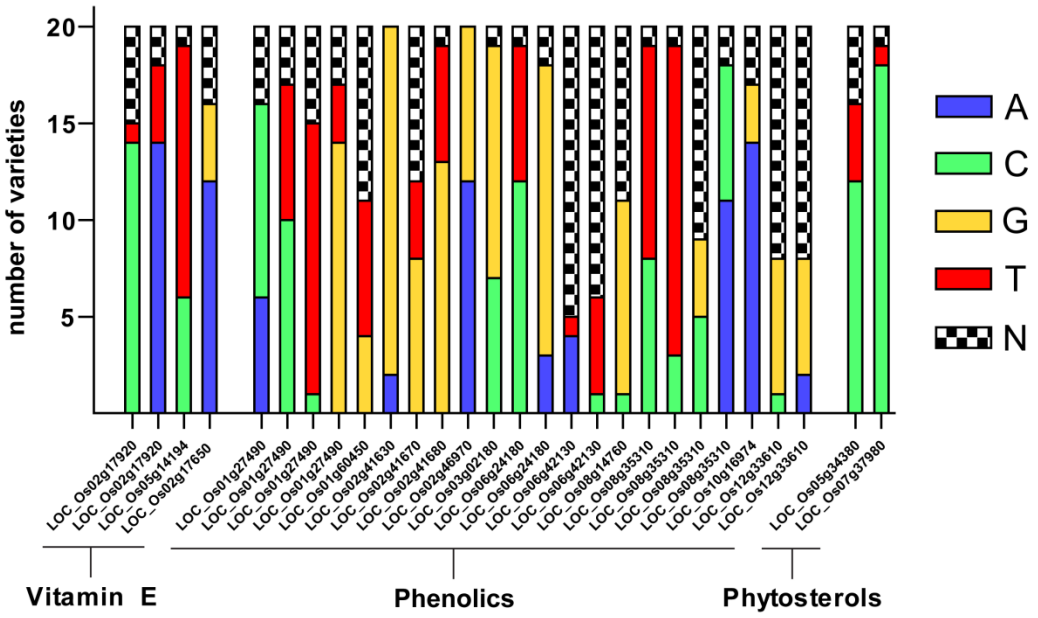


**Figure 2.3. Allele frequencies within the OryzaSNP set for nonsynonymous SNPs in genes for vitamin E, phenolic, and phytosterol pathways.** Allele frequencies are represented as the number of SNPs in common for each of the 20 varieties of the OryzaSNP set. X-axis labels correspond to the rice locus identifier for a given SNP.

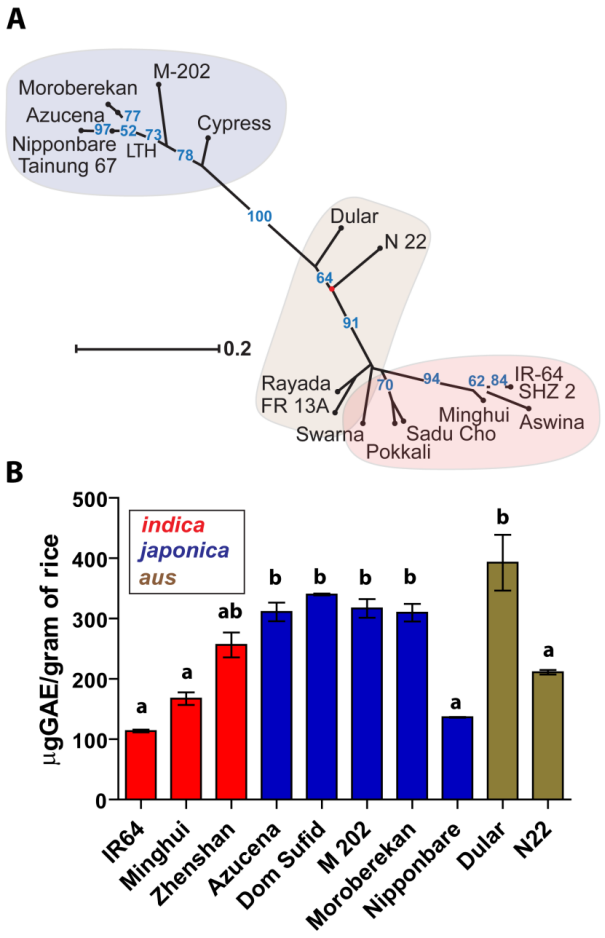


**Figure 2.4. Variation in total phenolics concentrations in cooked rice.** (A) An unrooted, neighbor-joining tree was developed based on total SNPs identified in the phenolic biochemical pathways. Clouds were colored according to subspecies: *indica* (red), *japonica* (blue), and *aus* (brown). (B) Total phenolics was measured in gallic acid equivalents (GAE) using Folin-Ciocalteau reagent. The letters a, b, and c denote significance (ANOVA, Tukey post-hoc, $P < 0.05$), and values are expressed as the mean ± the standard error of the mean.

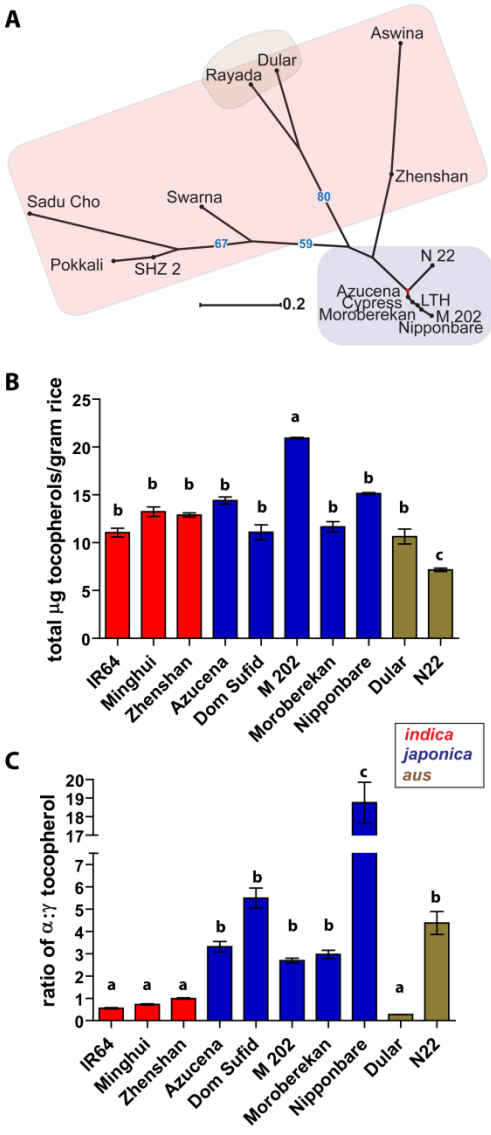


**Figure 2.5. Variation in vitamin E concentrations in cooked rice.** (A) An unrooted, neighbor-joining tree was developed based on total SNPs identified in the vitamin E synthesis pathway. Clouds were colored according to subspecies: *indica* (red), *japonica* (blue), and *aus* (brown). (B) The total quantities of tocopherols (α, γ, and δ) per gram of rice were determined. (C) Ratios of α:γ tocopherol were calculated for each variety. Values are expressed as the mean ± the standard error of the mean, and statistical groupings denoted by the letters a, b, and c (ANOVA, Tukey post-hoc, $P < 0.05$).

## TABLES

### Table 2.1 Rice materials

| Variety | Country of Origin | Subspecies | IRGC Accession ID | Breeding Classification | Traits of interest |
|---|---|---|---|---|---|
| Azucena | Philippines | *japonica* | 117264 | Landrace | Fragrant, tall stature, unique root structure |
| Dular | India | *aus* | 117266 | Landrace | Drought resistant, seed-shattering |
| Dom-Sufid | Iran | *japonica* | 117265 | Landrace | Similar to Basmati rice (aromatic) |
| IR64-21 | Philippines | *indica* | 117268 | Advanced | Widely grown, semidwarf, high yielding, abiotic and biotic stress tolerance |
| M 202 | United States | *japonica* | 117270 | Advanced | Erect leaf type, modern variety |
| Minghui 63 | China | *indica* | 117271 | Advanced | Parent used in hybrid breeding |
| Moroberekan | Guinea | *japonica* | 117272 | Landrace | Abiotic and biotic stress tolerance |
| N22 | India | *aus* | 117273 | Landrace | Red seed coat, stress tolerance |
| Nipponbare | Japan | *japonica* | 117274 | Advanced | First sequenced variety, short grain type |
| Zhenshan 97B | China | *indica* | 117280 | Advanced | Parent used in hybrid breeding |

**Table 2.2. Sum of squares of z-scores for 763 metabolites using Nipponbare (*japonica*) as a reference**

| Variety | Class | Sum of Squares |
|---|---|---|
| Zhenshan | *indica* | 49,099,871 |
| Minghui | *indica* | 6,683,571 |
| IR64 | *indica* | 4,709,880 |
| Dom Sufid | *japonica* | 568,329 |
| Azucena | *japonica* | 273,295 |
| M 202 | *japonica* | 79,634 |
| Moroberekan | *japonica* | 31,872 |
| Dular | *aus* | 71,600 |
| N22 | *aus* | 19,577 |

**Table 2.3. Genes associated with linolenic acid, phenolics, phytosterol, and vitamin E synthesis**

| LocusID | Enzymatic activity | SNPID | Class | Pathway |
|---|---|---|---|---|
| LOC_Os07g49310 | omega-3 fatty acid desaturase | TBGI337504 | Linoleic acid | lipid desaturation |
| LOC_Os07g49310 | omega-3 fatty acid desaturase | TBGI337502 | Linoleic acid | lipid desaturation |
| LOC_Os08g34220 | omega-6 fatty acid desaturase | TBGI358296 | Linoleic acid | lipid desaturation |
| LOC_Os08g34220 | omega-6 fatty acid desaturase | TBGI358297 | Linoleic acid | lipid desaturation |
| LOC_Os08g34220 | omega-6 fatty acid desaturase | TBGI358298 | Linoleic acid | lipid desaturation |
| LOC_Os08g34220 | omega-6 fatty acid desaturase | TBGI358299 | Linoleic acid | lipid desaturation |
| LOC_Os08g34220 | omega-6 fatty acid desaturase | TBGI358300 | Linoleic acid | lipid desaturation |
| LOC_Os08g34220 | omega-6 fatty acid desaturase | TBGI358301 | Linoleic acid | lipid desaturation |
| LOC_Os02g48560 | omega-6 fatty acid desaturase, endoplasmic reticulum isozyme 2, putative, expressed | no SNP | Linoleic acid | lipid desaturation |
| LOC_Os03g18070 | omega-3 fatty acid desaturase | no SNP | Linoleic acid | lipid desaturation |
| LOC_Os07g23410 | omega-6 fatty acid desaturase, endoplasmic reticulum isozyme 2, putative, expressed | no SNP | Linoleic acid | lipid desaturation |
| LOC_Os07g23430 | omega-6 fatty acid desaturase, endoplasmic reticulum isozyme 2, putative, expressed | no SNP | Linoleic acid | lipid desaturation |
| LOC_Os11g01340 | omega-3 fatty acid desaturase | no SNP | Linoleic acid | lipid desaturation |
| LOC_Os12g01370 | omega-3 fatty acid desaturase | no SNP | Linoleic acid | lipid desaturation |
| LOC_Os07g49310 | omega-3 fatty acid desaturase | TBGI337504 | Linoleic acid | lipid desaturation |
| LOC_Os07g49310 | omega-3 fatty acid desaturase | TBGI337502 | Linoleic acid | lipid desaturation |
| LOC_Os01g60450 | trans-cinnamate 4-monooxygenase | TBGI059140 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107353 | Phenolic | phenylpropanoid |
| LOC_Os02g41670 | phenylalanine ammonia-lyase | TBGI107368 | Phenolic | phenylpropanoid |
| LOC_Os02g41680 | phenylalanine ammonia-lyase | TBGI107403 | Phenolic | phenylpropanoid |
| LOC_Os02g46970 | 4-coumarate-CoA ligase | TBGI111976 | Phenolic | phenylpropanoid |
| LOC_Os03g02180 | ferulate 5-hydroxylase | TBGI124364 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289865 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289869 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347005 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479943 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479945 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107339 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107343 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107344 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107349 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107350 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107351 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107352 | Phenolic | phenylpropanoid |
| LOC_Os02g41670 | phenylalanine ammonia-lyase | TBGI107357 | Phenolic | phenylpropanoid |
| LOC_Os02g41680 | phenylalanine ammonia-lyase | TBGI107397 | Phenolic | phenylpropanoid |
| LOC_Os02g41680 | phenylalanine ammonia-lyase | TBGI107399 | Phenolic | phenylpropanoid |
| LOC_Os03g02180 | ferulate 5-hydroxylase | TBGI124360 | Phenolic | phenylpropanoid |
| LOC_Os02g41680 | phenylalanine ammonia-lyase | TBGI107408 | Phenolic | phenylpropanoid |
| LOC_Os01g60450 | | TBGI059143 | Phenolic | phenylpropanoid |
| LOC_Os01g60450 | | TBGI059145 | Phenolic | phenylpropanoid |
| LOC_Os02g46970 | 4-coumarate-CoA ligase | TBGI111975 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289866 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289873 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289874 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289875 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347006 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347008 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347011 | Phenolic | phenylpropanoid |
| LOC_Os01g60450 | phenylpropanoid biosynthesis | TBGI059141 | Phenolic | phenylpropanoid |
| LOC_Os01g60450 | phenylpropanoid biosynthesis | TBGI059147 | Phenolic | phenylpropanoid |
| LOC_Os01g60450 | phenylpropanoid biosynthesis | TBGI059151 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107354 | Phenolic | phenylpropanoid |
| LOC_Os02g41630 | phenylalanine ammonia-lyase | TBGI107355 | Phenolic | phenylpropanoid |
| LOC_Os02g41670 | phenylalanine ammonia-lyase | TBGI107371 | Phenolic | phenylpropanoid |
| LOC_Os02g41680 | phenylalanine ammonia-lyase | TBGI107406 | Phenolic | phenylpropanoid |
| LOC_Os05g35290 | phenylalanine ammonia-lyase | TBGI250746 | Phenolic | phenylpropanoid |
| LOC_Os05g35290 | phenylalanine ammonia-lyase | TBGI250747 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289867 | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289868 | Phenolic | phenylpropanoid |

| | | | | |
|---|---|---|---|---|
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289872 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347003 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347004 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347009 | Phenolic | phenylpropanoid |
| LOC_Os08g14760 | 4-coumarate-CoA ligase | TBGI347010 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479929 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479935 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479941 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479944 | Phenolic | phenylpropanoid |
| LOC_Os12g33610 | phenylalanine ammonia-lyase | TBGI479949 | Phenolic | phenylpropanoid |
| LOC_Os04g43760 | phenylalanine ammonia-lyase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os06g06980 | caffeoyl-CoA O-methyltransferase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os08g38900 | caffeoyl-CoA O-methyltransferase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os08g38910 | caffeoyl-CoA O-methyltransferase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os08g38920 | caffeoyl-CoA O-methyltransferase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os09g30360 | caffeoyl-CoA O-methyltransferase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os10g36848 | ferulate 5-hydroxylase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os11g48110 | phenylalanine ammonia-lyase | no SNP | Phenolic | phenylpropanoid |
| LOC_Os06g24180 | ferulate 5-hydroxylase | TBGI289864 | Phenolic | phenylpropanoid |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028393 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028394 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028395 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028396 | Phenolic | flavanoid biosynthesis |
| LOC_Os06g42130 | leucoanthocyanidin dioxygenase | TBGI304688 | Phenolic | flavanoid biosynthesis |
| LOC_Os06g42130 | leucoanthocyanidin dioxygenase | TBGI304691 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028392 | Phenolic | flavanoid biosynthesis |
| LOC_Os03g60509 | putative chalcone isomerase | TBGI173576 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028398 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g27490 | leucoanthocyanidin dioxygenase | TBGI028400 | Phenolic | flavanoid biosynthesis |
| LOC_Os01g44260 | dihydroflavonol-4-reductase | no SNP | Phenolic | flavanoid biosynthesis |
| LOC_Os04g56700 | naringenin,2-oxoglutarate 3-dioxygenase | no SNP | Phenolic | flavanoid biosynthesis |
| LOC_Os07g11440 | putative chalcone synthase | no SNP | Phenolic | flavanoid biosynthesis |
| LOC_Os10g17260 | flavonoid 3-monooxygenase | no SNP | Phenolic | flavanoid biosynthesis |
| LOC_Os11g32650 | putative chalcone synthase | no SNP | Phenolic | flavanoid biosynthesis |
| LOC_Os09g19734 | isochorismate synthase 1, chloroplast precursor, putative, expressed | no SNP | Phenolic | via chorismate, pyruvate pathway |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359460 | Phenolic | flavanoid biosynthesis |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359464 | Phenolic | flavanoid biosynthesis |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359466 | Phenolic | flavanoid biosynthesis |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359467 | Phenolic | flavanoid biosynthesis |
| LOC_Os10g16974 | flavonoid 3-hydroxylase | TBGI400908 | Phenolic | leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os10g16974 | flavonoid 3-hydroxylase | TBGI400906 | Phenolic | leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359462 | Phenolic | flavanoid biosynthesis |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359463 | Phenolic | flavanoid biosynthesis |
| LOC_Os10g16974 | flavonoid 3-hydroxylase | TBGI400913 | Phenolic | leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os10g16974 | flavonoid 3-hydroxylase | TBGI400915 | Phenolic | leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os10g16974 | flavonoid 3-hydroxylase | TBGI400916 | Phenolic | leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os10g16974 | flavonoid 3-hydroxylase | TBGI400920 | Phenolic | leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os08g35310 | isoflavone-7-O-methytransferase 9 | TBGI359465 | Phenolic | flavanoid biosynthesis |
| LOC_Os09g09230 | dihydroflavonol-4-reductase, putative | no SNP | Phenolic | leucodelphinidin |

| | | | | |
|---|---|---|---|---|
| LOC_Os10g17260 | flavonoid 3-hydroxylase | no SNP | Phenolic | biosynthesis (dihydrotricetin (demthylated tricin)) leucodelphinidin biosynthesis (dihydrotricetin (demthylated tricin)) |
| LOC_Os05g34380 | cytochrome P450 51, putative | TBGI249857 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329832 | Phytosterol | sterol synthesis |
| LOC_Os03g59040 | farnesyl-diphosphate farnesyltransferase | TBGI172275 | Phytosterol | sterol synthesis |
| LOC_Os03g59040 | farnesyl-diphosphate farnesyltransferase | TBGI172276 | Phytosterol | sterol synthesis |
| LOC_Os03g59040 | farnesyl-diphosphate farnesyltransferase | TBGI172274 | Phytosterol | sterol synthesis |
| LOC_Os07g10130 | farnesyl-diphosphate farnesyltransferase | TBGI317153 | Phytosterol | sterol synthesis |
| LOC_Os07g10600 | sterol 24-C-methyltransferase | TBGI317459 | Phytosterol | sterol synthesis |
| LOC_Os07g10600 | sterol 24-C-methyltransferase | TBGI317461 | Phytosterol | sterol synthesis |
| LOC_Os07g37970 | cytochrome P450 51, putative | TBGI329813 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329833 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329834 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329836 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329837 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329840 | Phytosterol | sterol synthesis |
| LOC_Os07g37980 | cytochrome P450 51, putative | TBGI329843 | Phytosterol | sterol synthesis |
| LOC_Os09g39220 | C-14 sterol reductase | TBGI391322 | Phytosterol | sterol synthesis |
| LOC_Os09g39220 | C-14 sterol reductase | TBGI391323 | Phytosterol | sterol synthesis |
| LOC_Os09g39220 | C-14 sterol reductase | TBGI391329 | Phytosterol | sterol synthesis |
| LOC_Os05g34380 | cytochrome P450 51, putative | TBGI249853 | Phytosterol | sterol synthesis |
| LOC_Os05g34380 | cytochrome P450 51, putative | TBGI249858 | Phytosterol | sterol synthesis |
| LOC_Os05g34380 | cytochrome P450 51, putative | TBGI249862 | Phytosterol | sterol synthesis |
| LOC_Os07g37970 | cytochrome P450 51, putative | TBGI329817 | Phytosterol | sterol synthesis |
| LOC_Os01g01369 | 3-beta-hydroxysteroid-delta-isomerase, putative, expressed | no SNP | Phytosterol | sterol synthesis |
| LOC_Os01g25189 | C-14 sterol reductase | no SNP | Phytosterol | sterol synthesis |
| LOC_Os02g04760 | cycloartenol synthase, putative | no SNP | Phytosterol | sterol synthesis |
| LOC_Os02g26650 | sterol delta7 reductase | no SNP | Phytosterol | sterol synthesis |
| LOC_Os03g04340 | S-adenosylmethionine-dependent methyltransferase | no SNP | Phytosterol | sterol synthesis |
| LOC_Os05g14800 | cycloartenol synthase, putative | no SNP | Phytosterol | sterol synthesis |
| LOC_Os07g28110 | cytochrome P450 51, putative | no SNP | Phytosterol | sterol synthesis |
| LOC_Os07g28160 | cytochrome P450 51, putative | no SNP | Phytosterol | sterol synthesis |
| LOC_Os11g18310 | cycloartenol synthase, putative | no SNP | Phytosterol | sterol synthesis |
| LOC_Os11g18340 | cycloartenol synthase, putative | no SNP | Phytosterol | sterol synthesis |
| LOC_Os11g19700 | cycloeucalenol cycloisomerase | no SNP | Phytosterol | sterol synthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090438 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090439 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090428 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090429 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090441 | Vitamin E | vitamin E biosynthesis |
| LOC_Os08g09250 | 4-hydroxyphenylpyruvate dioxygenase | TBGI343503 | Vitamin E | vitamin E biosynthesis |
| LOC_Os08g09250 | 4-hydroxyphenylpyruvate dioxygenase | TBGI343504 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g07160 | 4-hydroxyphenylpyruvate dioxygenase | TBGI081638 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090186 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090187 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090190 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090191 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090193 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090195 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090432 | Vitamin E | vitamin E biosynthesis |

| | | | | |
|---|---|---|---|---|
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090436 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17920 | 4-hydroxyphenylpyruvate dioxygenase | TBGI090437 | Vitamin E | vitamin E biosynthesis |
| LOC_Os05g14194 | 4-hydroxyphenylpyruvate dioxygenase | TBGI236102 | Vitamin E | vitamin E biosynthesis |
| LOC_Os05g14194 | 4-hydroxyphenylpyruvate dioxygenase | TBGI236103 | Vitamin E | vitamin E biosynthesis |
| LOC_Os05g14194 | 4-hydroxyphenylpyruvate dioxygenase | TBGI236108 | Vitamin E | vitamin E biosynthesis |
| LOC_Os05g14194 | 4-hydroxyphenylpyruvate dioxygenase | TBGI236109 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g17650 | tocopherol cyclase | TBGI090196 | Vitamin E | vitamin E biosynthesis |
| LOC_Os12g42090 | tocopherol cyclase | TBGI489786 | Vitamin E | vitamin E biosynthesis |
| LOC_Os02g47310 | tocopherol O-methyltransferase | no SNP | Vitamin E | vitamin E Biosynthesis |
| LOC_Os06g43880 | homogentisic acid geranylgeranyl transferase | no SNP | Vitamin E | vitamin E Biosynthesis |
| LOC_Os06g44840 | homogentisate phytyltransferase | no SNP | Vitamin E | vitamin E biosynthesis |
| LOC_Os07g49310 | omega-3 fatty acid desaturase | TBGI337504 | Linoleic acid | glycolipid desaturation |
| LOC_Os07g49310 | omega-3 fatty acid desaturase | TBGI337502 | Linoleic acid | glycolipid desaturation |

**Table 2.4. SNPs in genes that regulate phenolics, phytosterols, vitamin E, and linolenic acid.**

| Class | Gene | SNPs | SYN | NS | 5' | 3' | INT | SNPs/gene | NS SNPs/gene |
|---|---|---|---|---|---|---|---|---|---|
| Phenolics | 30 | 78 | 24 | 22 | 2 | 12 | 17 | 2.60 | 0.73 |
| Phytosterols | 15 | 22 | 4 | 2 | 0 | 2 | 14 | 1.27 | 0.09 |
| Vitamin E | 9 | 23 | 2 | 2-4 | 3 | 2-3 | 14 | 2.55 | 0.09-0.17 |
| Linolenate | 7 | 8 | 0 | 0 | 0 | 1 | 7 | 1.14 | 0.00 |

SYN: synonymous
NS: nonsynonymous
5', 3': untranslated regions
INT: intron

## REFERENCES


1. Food and Agriculture Organization of the United Nations (2010) FAO Statistical Databases: Analytical Graphs: Global Source of Calories.
2. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci U S A 106: 12273-12278.
3. McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, et al. (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. Plant Physiol 141: 26-31.
4. Fitzgerald MA, McCouch SR, Hall RD (2009) Not just a grain of rice: the quest for quality. Trends in Plant Science 14: 133-139.
5. Sierra S, Lara-Villoslada F, Olivares M, Jimenez J, Boza J, et al. (2005) Increased immune response in mice consuming rice bran oil. Eur J Nutr 44: 509-516.
6. Boateng J, Verghese M, Panala V, Walker LT, Shackelford L (2009) Protective Effects of Rice Bran on Chemically Induced Colon Tumorigenesis may be Due to Synergistic/Additive Properties of Bioactive Components. Int J Cancer Res 5: 153-166.
7. Xia XD, Ling WH, Ma J, Xia M, Hou MJ, et al. (2006) An anthocyanin-rich extract from black rice enhances atherosclerotic plaque stabilization in apolipoprotein e-deficient mice. J Nutr 136: 2220-2225.
8. Pryor WA (2000) Vitamin E and heart disease: Basic science to clinical intervention trials. Free Radic Biol Med 28: 141-164.
9. Grassmann J, Hippeli S, Elstner EF (2002) Plant's defence and its benefits for animals and medicine: role of phenolics and terpenoids in avoiding oxygen stress. Plant Physiol Biochem 40: 471-478.
10. Lin XB, Ma L, Racette SB, Spearie CLA, Ostlund RE (2009) Phytosterol glycosides reduce cholesterol absorption in humans. Am J Physiol Gastrointest Liver Physiol 296: G931-G935.
11. Billman GE, Kang JX, Leaf A (1999) Prevention of sudden cardiac death by dietary pure omega-3 polyunsaturated fatty acids in dogs. Circulation 99: 2452-2457.
12. Duthie G, Crozier A (2000) Plant-derived phenolic antioxidants. Curr Opin Lipidol 11: 43-47.
13. Valko M, Rhodes CJ, Moncol J, Izakovic M, Mazur M (2006) Free radicals, metals and antioxidants in oxidative stress-induced cancer. Chem Biol Interact 160: 1-40.
14. Singh U, Devaraj S, Jialal I (2005) Vitamin E, oxidative stress, and inflammation. Annu Rev Nutr 25: 151-174.
15. Tarpley L, Duran AL, Kebrom TH, Sumner LW (2005) Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period. BMC Plant Biol 5.
16. Ishikawa T, Takahara K, Hirabayashi T, Matsumura H, Fujisawa S, et al. (2010) Metabolome Analysis of Response to Oxidative Stress in Rice Suspension Cells Overexpressing Cell Death Suppressor Bax Inhibitor-1. Plant Cell Physiol 51: 9-20.
17. Zhou JM, Fukushi Y, Wollenweber E, Ibrahim RK (2008) Characterization of two O-methyltransferase-like genes in barley and maize. Pharm Biol 46: 26-34.
18. Chaudhary N, Khurana P (2009) Vitamin E biosynthesis genes in rice: Molecular characterization, expression profiling and comparative phylogenetic analysis. Plant Sci 177: 479-491.

19. Nes WD, Venkatramesh M (1999) Enzymology of phytosterol transformations. Crit Rev Biochem Mol Biol 34: 81-93.
20. Shanklin J, Cahoon EB (1998) Desaturation and related modifications of fatty acids. Annu Rev Plant Physiol Plant Mol Biol 49: 611-641.
21. Skogerson K, Harrigan GG, Reynolds TL, Halls SC, Ruebelt M, et al. (2010) Impact of Genetics and Environment on the Metabolite Composition of Maize Grain. J Agric Food Chem 58: 3600-3610.
22. Mochida K, Furuta T, Ebana K, Shinozaki K, Kikuchi J (2009) Correlation exploration of metabolic and genomic diversity in rice. BMC Genomics 10.
23. Sands DC, Morris CE, Dratz EA, Pilgeram AL (2009) Elevating optimal human nutrition to a central goal of plant breeding and production of plant-based foods. Plant Sci 177: 377-389.
24. Simon PW, Pollak LM, Clevidence BA, Holden JM, Haytowitz DB (2009) Plant Breeding for Human Nutritional Quality. Plant Breed Rev 31: 325-415.
25. Oksman-Caldentey KM, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. Curr Opin Biotechnol 16: 174-179.
26. Keurentjes JJB, Fu JY, de Vos CHR, Lommen A, Hall RD, et al. (2006) The genetics of plant metabolism. Nat Genet 38: 842-849.
27. Zhao K, Wright M, Kimball J, Eizenga G, McClung A, et al. (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS One 5: e10780.
28. Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, et al. (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. Plant J 53: 960-972.
29. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, et al. (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol 24: 447-454.
30. Goffman FD, Bergman CJ (2004) Rice kernel phenolic content and its relationship with antiradical efficiency. J Sci Food Agric 84: 1235-1240.
31. Hudson EA, Dinh PA, Kokubun T, Simmonds MSJ, Gescher A (2000) Characterization of potentially chemopreventive phenols in extracts of brown rice that inhibit the growth of human breast and colon cancer cells. Cancer Epidemiol Biomarkers Prev 9: 1163-1170.
32. DellaPenna D (2005) A decade of progress in understanding vitamin E synthesis in plants. J Plant Physiol 162: 729-737.
33. Jiang Q, Elson-Schwab I, Courtemanche C, Ames BN (2000) gamma-tocopherol and its major metabolite, in contrast to alpha-tocopherol, inhibit cyclooxygenase activity in macrophages and epithelial cells. Proc Natl Acad Sci U S A 97: 11494-11499.
34. Saldeen T, Li DY, Mehta JL (1999) Differential effects of alpha- and gamma-tocopherol on low-density lipoprotein oxidation, superoxide activity, platelet aggregation and arterial thrombogenesis. J Am Coll Cardiol 34: 1208-1215.
35. Leonard SW, Paterson E, Atkinson JK, Ramakrishnan R, Cross CE, et al. (2005) Studies in humans using deuterium-labeled alpha- and gamma-tocopherols demonstrate faster plasma gamma-tocopherol disappearance and greater gamma-metabolite production. Free Radic Biol Med 38: 857-866.
36. Demurin Y, Skoric D, Karlovic D (1996) Genetic variability of tocopherol composition in sunflower seeds as a basis of breeding for improved oil quality. Plant Breed 115: 33-36.
37. Shintani D, DellaPenna D (1998) Elevating the vitamin E content of plants through metabolic engineering. Science 282: 2098-2100.

38. Van Eenennaam AL, Li GF, Venkatramesh M, Levering C, Gong XS, et al. (2004) Elevation of seed alpha-tocopherol levels using plant-based transcription factors targeted to an endogenous locus. Metab Eng 6: 101-108.
39. Morris CE, Sands DC (2006) The breeder's dilemma - yield or nutrition? Nat Biotechnol 24: 1078-1080.
40. Singleton VL, Orthofer R, Lamuela-Raventos RM (1999) Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent. Oxidants and Antioxidants, Pt A. San Diego: Academic Press Inc. pp. 152-178.
41. Chen MH, Bergman C (2005) A rapid procedure for analysing rice bran tocopherol, tocotrienol and gamma-oryzanol contents. J Food Compost Anal 18: 319-331.

# CHAPTER THREE: GENETIC DIVERSITY AND GROWING ENVIRONMENT JOINTLY INFLUENCE METABOLITE VARIATION IN COOKED GRAIN FROM TWO DIVERGENT RICE VARIETIES

A modified version of this chapter in preparation as Heuberger AL *et al.*

**SUMMARY**

Genetic diversity is a major source of variation in the cooked rice metabolome, and evidence for genotype-environment interaction (GEI) effects on plant secondary metabolism is limited. A metabolomics dataset identified two divergent rice varieties (*Oryza sativa* L.) that differ in secondary metabolite profiles in the cooked grain: IR64 and Moroberekan. Variation in primary metabolites, which are precursors to secondary metabolites, can estimate effects associated with secondary metabolite variation. GEI effects were estimated by growing IR64 and Moroberekan in the field and greenhouse, and metabolite profiling was conducted on aqueous-methanol extracts of cooked rice using gas-chromatography coupled to mass spectrometry (GC-MS). Of the 39 metabolites detected, 17 exhibited GEI as a source of metabolite variation, including the primary metabolite phenylalanine, a precursor to secondary metabolite synthesis. Informatic analyses characterized genetic variation that may contribute to phenylalanine variation by evaluating arogenate/prephenate dehydratase (ADT/PDT), the rate-limiting enzyme in phenylalanine synthesis. Gene expression varied both within and among 11 ADT/PDT genes for five rice varieties in the Rice MPSS seed database. The ADT/PDT gene LOC_07g49390 had the most expression in the developing seed. The RiceNet gene interaction database implicated six additional amino acid synthesis genes as potentially epistatic to LOC_07g49390, including two genes involved in serine synthesis. Expression patterns for phenylalanine and serine genes were similar in the developing seed, and the two metabolites covaried in the cooked rice analysis. An amino acid gene network was constructed to display the relationship between phenylalanine and serine, and implicated glutamate as an important intermediary between phenylalanine and serine. These data support the role of genetic and metabolic networks in regulating metabolism. Gene-metabolite network analysis is a new and important method to identify genetic and GEI components that contribute to variation in secondary metabolism.

## INTRODUCTION

Both genetic and environmental effects have been implicated as sources of metabolic variation in food crops. In rice (*Oryza sativa* L.), genomic variation contributes to metabolomic differences in the grain [1,2,3]. Environmental effects also contribute to metabolite variation in rice and maize [4,5]. While genotypic and environmental variation has been independently associated with metabolite profiles of the grain, limited information exists for the importance of gene-environment interaction (GEI) effects and the genes centrally involved in the interactions.

Primary and secondary metabolites influence agronomic and post-harvest traits in food crops, including tolerance to environmental stress and the nutritional content in the grain. Primary metabolites function to maintain normal plant respiration and energy balance, and serve as precursors to synthesize secondary metabolites. Secondary metabolites are a diverse class of compounds that, although not acting towards energy homeostasis, act as important plant hormones, pigments, and volatiles [6,7]. Terpenes and flavonoids are two major classes of secondary metabolites that mediate plant-environment interactions, such as responding to abiotic or biotic stresses [8,9]. Dietary intake of terpenes and flavonoids is now being recognized as important to human health [10,11,12].

In plants, many secondary metabolites are dependent on phenylalanine as a primary metabolite precursor [13]. In plants, the rate-limiting enzyme for phenylalanine synthesis is either arogenate or prephenate dehydratase (ADT/PDT) [14,15,16,17]. Regulation of ADT/PDT genes is important for secondary metabolism. For example, *Arabidopsis pd1* (PDT) mutants showed increased sensitivity to ultraviolet (UV) radiation due to the absence of UV-absorbing phenylpropanoids [18]. In rice, an ADT/PDT mutant with insensitivity to phenylalanine feedback regulation showed elevated quantities of both phenylalanine and several phenylpropanoids [19]. In both cases, the modification of

a single phenylalanine biosynthesis enzyme also affected synthesis of downstream secondary metabolites. There is limited information about the effects of genes external to the phenylalanine pathway on secondary metabolism. Analyzing metabolomic data in the context of genetic networks is a new approach to identify additional genes that may regulate secondary metabolism.

Estimating GEI effects on the grain metabolome is limited for rice. There are greater than 100,000 rice accessions among three subspecies-like groupings (*indica, japonica,* and *aus*) [20], and rice is grown across environments with highly variable temperatures, latitudes, and irrigation schemes. Single nucleotide polymorphism (SNP) profiles were established for a core collection of 20 rice varieties, referred to as the OryzaSNP set [21],[22]. Two OryzaSNP varieties, IR64 and Moroberekan, vary in morphology, physiology, and grain metabolite traits [1,23]. IR64 is a high-yielding *indica* variety from the Philippines and is the most widely grown variety across Asia. Moroberekan is a large and low-yielding *japonica* variety from Guinea and was selected from a pure-line as a landrace. Genome-wide SNP variation confirmed the evolutionary relationship between IR64 and Moroberekan as divergent [22].

In the present study, IR64 and Moroberekan were assessed for secondary metabolite variation in cooked brown rice. GEI effects on primary metabolism and phenylalanine were determined using gas-chromatography coupled to mass spectrometry (GC-MS). GC-MS was conducted for cooked brown rice extracts from IR64 and Moroberekan grown field and greenhouse environments. Phenylalanine regulates a major link between primary and secondary metabolism, and so variation in phenylalanine synthesis genes was investigated by expression analysis. A genetic network was constructed with primary and intermediate metabolites to highlight the relationships among non-phenylalanine synthesis genes that also may contribute to phenylalanine-associated GEI effects.

## MATERIALS AND METHODS

### Plant growing conditions

Rice plants were grown in greenhouse and field conditions. Greenhouse-plants were grown at Colorado State University in 2009 and germination, potting, and growing conditions were as previously described [23]. Briefly, seeds were germinated in the presence of fungicide for 3 days, potted in peat:soil:sand (4:4:1), and watered daily. Plants were grown at 25.5 ºC and 55% humidity with supplemental lighting for 16/8 (light/dark) photoperiod, and were fertilized twice a week. Upon maturation, rice seed was harvested, stored for 1 month at room temperature, and then stored at 4 ºC until further use. IR64 and Moroberekan were additionally grown at the International Rice Research Institute (Los Baños, Philippines) in 2009 in paddy conditions, and plants were watered and fertilized as required. Seed was harvested, dried, and stored at 4 ºC until further use

### Secondary metabolite analysis

Secondary metabolites were extracted and detected in a previous study using ultra performance chromatography coupled to mass spectrometry (UPLC-MS) [1]. The collision energy at 7 eV induced incomplete fragmentation, and neutral losses were observed to be consistent with secondary metabolites. Parent ions and fragments were screened in metabolite databases for likely matches using an error window of 10 ppm.

### Metabolite extraction and detection

For GC-MS, rice seeds were dehulled using a TR200 electronic dehusker (Kett, Japan), left unpolished, and cooked for 55 min in a 2:1 water:rice (v/v) ratio using an open boiling technique. Rice was cooked until the time at which the grain starch granules were completely gelatinized. Cooked rice was immediately frozen at -80 ºC,

lyophilized over the course of 48 hr, and stored at -80 ºC until further use. Metabolites were extracted in triplicate as previously described [1], whereby rice was ground to a fine powder with a mortar and pestle in liquid nitrogen, and 1 mL of cold aqueous: methanol (1:4) solvent was added to 100 mg of powder, incubated for 1 hr at -80 ºC, centrifuged at 1500xg for 5 min at 4 ºC, and 500 $\mu$L of the supernatant was transferred to a new 1.5 mL microcentrifuge tube.

For GC-MS derivatization and detection, the cooked rice extract was dried using a vacuum centrifuge, resuspended in 50 $\mu$L of pyridine containing 20 mg/mL of methoxyamine hydrochloride, and incubated at 37 ºC for 2 hr. A second derivatization step was conducted by adding 50 $\mu$L of N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS) (Thermo Scientific). Samples were incubated at 37 ºC for 1 hr, centrifuged at 3000xg for 5 min, cooled to room temperature, and 80 $\mu$L of the supernatant was transferred to a 200 $\mu$L glass insert. Metabolites were detected a Trace GC Ultra coupled to a Thermo DSQ II (Thermo Scientific), which scanned 50-650 *m/z* at 5 scans/sec in electron impact mode, and a 30 m TG-5MS column (Thermo Scientific, 0.25 mm i.d., 0.25 $\mu$m film thickness). Both the inlet and transfer line were set at 280 ºC. The samples were injected in a 10:1 split ratio twice in discrete randomized blocks with a 1.2 mL/min flow rate, and the program consisted of 80 ºC for 30 sec, a ramp of 15 ºC per min to 330 ºC, and then held for 8 min. GC-MS mass spectra were deconvoluted in AMDIS (http://chemdata.nist.gov/mass-spc/amdis) and screened in the National Institute for Technology Standards metabolite database (www.nist.gov) for most probable matches.

**Bioinformatics**

The percent of maximum gene expression was determined individually for each gene among all libraries in the MPSS database, including libraries not shown in this

report. Only transcripts uniquely associated with a gene were used, except for LOC_Os09g39260 and LOC_Os01g34450, which only contained shared transcripts. SNPs and SNP classifications were established base on the rice OryzaSNP database (www.oryzasnp.org). The amino acid network was constructed in VANTED software [24] by linking common metabolites and enzymes among MetaCrop [25] amino acid synthesis pathways.

**Statistics**

Quantities of each molecular feature were determined using XCMS software (http://metlin.scripps.edu/xcms), which normalized each injection based on total molecular feature intensity values and integrated peak areas based on individual molecular feature intensities. For both UPLC-MS and GC-MS, principal component analysis was conducted in R statistical software, and unit variance scaling was applied to the dataset. Variation among molecular features was determined using Student's t-test with varying p-value thresholds on peak areas of each feature. Z-scores for primary metabolites were calculated based on peak areas of the most abundant feature for a given spectrum. Z scores were calculated independently for IR64 and Moroberekan and were based on the mean and standard deviation of field-grown metabolite quantities. For molecular features, genetic, environment, and GEI effects were determined by ANOVA ($p < 0.05$). For metabolite analysis, a false discovery rate correction was applied by confirming a minimum of three significant effects at a single retention time.

## RESULTS

### Secondary metabolites in cooked grain vary between IR64 and Moroberekan

A UPLC-MS metabolomics technique was used to detect secondary metabolite variation between IR64 and Moroberekan. This technique assumes that each metabolite is detected as a "molecular feature," and variation in molecular features infers metabolomic properties of the rice extract. UPLC-MS detected 1,256 molecular features, and 55 (4.4%) varied between IR64 and Moroberekan (Student's t-test, $p < 0.001$). The molecular features were sorted by size (mass/charge, assuming single charge) for each variety (Figure 3.1A). The two mass distribution profiles were highly similar. This indicates that metabolite variation between IR64 and Moroberekan was mostly due to variable quantities of the same metabolite, rather than metabolites unique to only one variety. Principal component analysis (PCA) was conducted on the molecular features to confirm variation in secondary metabolite profiles, which was largely explained by PC1 (37.5%) (Figure 3.1B).

To determine metabolites that contribute to variation between IR64 and Moroberekan, molecular features were sorted by volcano analysis (2-fold differences, Student's t-test $p < 0.01$) (Figure 3.1C). Moroberekan contained 41 molecular features with a 2-fold difference or greater than IR64, and IR64 contained 18 molecular features with a 2-fold difference or greater than Moroberekan (Student's t-test, $p < 0.01$). The 41 molecular features greater in Moroberekan corresponded to 11 metabolites. None of the 18 molecular features higher in IR64 associated with a parent metabolite, indicating these molecular features were minor fragments rather than representatives of the parent compound. The 11 metabolites that were identified in higher amounts in Moroberekan were classified based on mass spectral analysis, and relative quantities were determined using the two most prominent molecular features for each metabolite (Figure 3.1D). Metabolite classes included terpenoids (phytosterols, plant hormones, and

carotenoids), flavonoids, saponins, and glycerophospholipids. One compound contained a fragmentation pattern associated with a flavonoid-glycoside, and may explain the darker, yellow pigmentation of the aqueous-methanolic extracts from Moroberekan compared to IR64 (personal observation). These data imply that IR64 and Moroberekan vary for secondary metabolite profiles in the cooked grain.

**GEI associated with metabolite profiles and phenylalanine**

A GC-MS-based metabolomics approach was performed to determine primary metabolite variation affected by growing plants in field (FD) or greenhouse (GH) environments. Among the two varieties, 4,972 molecular features were observed. A PCA was conducted on the molecular features to determine genetic (G), environment (E), and GEI effects on the metabolome (Figure 3.2A). The PCA showed four distinct clusters: IR64-GH, IR64-FD, Moro-GH, and Moro-FD, and suggests the source of metabolite variation was a combination of genotypic and environmental effects. However, neither PC1 nor PC2 explained the variation as discretely due to variety or environment, and suggests a series of complex interactions between the two effects.

To determine GEI on primary metabolite variation, an ANOVA was conducted on each molecular features with G, E, and GEI as factor as potential effects. Of the 4,972 molecular features, 1,279 showed genotypic effects (25.7%), 1,255 (25.2%) showed environmental effects, and 633 (12.7%) showed evidence of GEI ($p < 0.05$). A volcano analysis was conducted on molecular features to determine if the two varieties and environments equally contribute to metabolite variation (Figure 3.2B). Molecular features that varied by genotype were largely skewed towards Moroberekan, which had 188 features compared to 23 in IR64 (2-fold difference in quantity, $p < 0.01$). Cooked rice from plants grown in the field contained more molecular features at greater quantities (148) than in the greenhouse (50) (Figure 3.2C).

For each retention time, mass spectra were screened in the National Institute for Standards and Technology database for to determine the metabolite associated with each molecular features. At a single retention time, the peak area of the most abundant molecular feature was used to assess G, E, and GEI effects for each metabolite (ANOVA, $p < 0.05$) (Table 3.1). The effect of growing IR64 and Moroberekan in different environments was additionally visualized for each metabolite by z-score analysis (Figure 3.3). Z represents the number of standard deviations attributed to growing a given variety in the field, compared to growing a variety in the greenhouse. Metabolites with z-scores greater or less than 1.96 suggest evidence of significant environmental effects in many of the sugars, amino acids, fatty acids, and organic and inorganic acids, and other metabolites detected. Z-scores that differ between IR64 and Moroberekan show GEI effects on metabolite variation. The z-score also suggests substantial variation within each metabolite class, including the amino acids, for which metabolites did not all covary with phenylalanine.

**SNP variation in ADT/PDT genes**

Phenylalanine varied between IR64 and Moroberekan in the field and greenhouse, and SNP diversity in ADT/PDT genes may explain the genetic variation associated with GEI effects. ADT/PDT genes were identified in rice based on previous reports [19], and also by screening for enzymes with PDT activity the rice biochemical database Ricecyc (www.gramene.org/pathway), which included 11 genes with Enzyme Commission (EC) codes 4.2.1.91 (ADT) or 4.2.1.51 (PDT activity) (Table 3.2). Variation in the ADT/PDT genes was evaluated by screening for SNPs among 20 varieties of rice using the OryzaSNP database (www.oryzasnp.org). The open reading frame of three ADT/PDT genes contained SNPs, however, all were synonymous polymorphisms. Two SNPs were detected in the 3' UTR, suggesting potential variation in gene regulation,

specifically at LOC_07g49390 and LOC_Os10g37980. IR64 and Moroberekan showed highly similar ADT/PDT haplotypes, however a potential SNP in the 3' UTR on chromosome 7 should be further investigated. Two other varieties, Cypress and Nipponbare, were displayed to show that SNPs were observed among other OryzaSNP varieties.

**Varietal differences in ADT/PDT gene expression**

Genetic diversity associated with phenylalanine biosynthesis was further interrogated by gene expression analysis among the 11 ADT/PDT genes. Gene expression profiles were acquired from the rice Massively Parallel Signature Sequencing (MPSS) database (http://mpss.udel.edu/rice), which consists of genome-wide gene expression profiles for 70 rice libraries [26]. Profiles for LOC_Os09g39260 and LOC_01g34450 were indistinguishable and therefore merged. A heat map was constructed to display 37 libraries that encompass profiles for the developing seed, as well as abiotic and biotic stresses (Figure 3.4). While the libraries do not include IR64 and Moroberekan, the MPSS data shows evidence of ADT/PDT expression variation in rice. For all libraries, LOC_Os04g55780, LOC_Os06g45930, LOC_07g32774, and LOC_Os08g33260 were not expressed and are unlikely to contribute to phenylalanine biosynthesis. LOC_Os03g17730 was the most commonly expressed gene and appears constitutive relative to the other ADT/PDT genes. For the developing seed, variation in ADT/PDT expression was observed in five of the ten genes, and each variety contained a unique gene expression profile. The highest levels of ADT/PDT expression occurred in variety Nipponbare, and this was the result of the expression of LOC_Os03g17730, LOC_Os07g49390, and LOC_Os09g39260/01g34450. The varieties Ilpumbyeo and LaGrue exhibited lower total ADT/PDT expression than Nipponbare and were derived from two genes: LOC_Os03g17730 and LOC_Os07g49390. Alternatively, Cypress

ADT/PDT expression was due to two different genes: LOC_Os09g3920 and LOC_Os03g17730. Furthermore, ADT/PDT expression was upregulated in plants resistant to the plant pathogens *Xanthomonas oryzae* pv. *oryzae* (Xoo) and *Magnaporthe grisea* (Mg) compared to mock inoculation controls and susceptible plants. This implicates ADT/PDT genes as likely important to secondary metabolism-associated plant defense responses. The ADT/PDT expression analysis shows variation within and among ADT/PDT genes for the five varieties, and indicates that such variation may be an important genetic component of phenylalanine-related GEI effects.

**Amino acid synthesis genes that interact with ADT/PDT gene LOC_Os07g49390**

Because LOC_Os07g49390 was the most abundant ADT/PDT gene in the five MPSS developing seed libraries, predicted gene interactions were determined using RiceNet, a database that establishes gene networks by characterizing gene interactions from 24 libraries (www.functionalnet.org/ricenet). The RiceNet algorithm identified 38 genes predicted to interact with LOC_Os07g49390, including six genes associated with amino acid and other metabolite syntheses (Table 3.3) The LOC_Os07g49390 ADT/PDT gene was predicted to interact with genes involved in tyrosine biosynthesis and the shikimate pathway, both of which are closely linked to phenylalanine synthesis. There was an additional predicted interaction with two genes involved in serine biosynthesis and an amino acid synthesis regulatory gene. The wide variation in MPSS-ADT/PDT gene expression profiles and predicted interactions with other amino acid biosynthesis enzymes suggest the ADT/PDT genes are centrally involved in amino acid metabolite variation.

The RiceNet gene interaction model was evaluated by establishing expression patterns of each gene as described in the rice MPSS database, and determining patterns similar to LOC_Os07g49390 gene expression within each library (Figure 3.5).

None of the predicted interaction genes had identical expression patterns to LOC_Os07g49390. However, LOC_Os03g06200, a serine metabolism gene, contained similar expression patterns as LOC_Os07g49390 in the developing seed, specifically between the varieties Cypress and Nipponbare. LOC_Os04g55720, a second serine metabolism gene, contained a profile more consistent with the ADT/PDT gene LOC_Os09g39230, whereby expression was significantly higher in Cypress compared to the other four rice varieties. The different genes by which the five MPSS varieties regulate phenylalanine metabolism may explain genotype and GEI effects involved in both primary and secondary metabolism in the cooked grain.

**Genetic network for genes associated with phenylalanine synthesis**

A primary metabolite network for amino acids was constructed to explain the predicted gene interactions associated with phenylalanine synthesis. Amino acid synthesis pathways from MetaCrop [25], a metabolic database for crops, were networked using VANTED software [24] (Figure 3.6). Nodes were constructed by linking common metabolites and enzymes among each of the amino acid biosynthesis pathways. The amino acid network confirms the link between phenylalanine and serine synthesis genes as potentially mediated through glutamate-associated transaminase activity. Phenylalanine and serine covaried and were both significant for GEI effects.

**DISCUSSION**

This is the first report of genetic, environmental, and GEI effects on the grain metabolite profile in rice. Varietal and environmental influences on metabolic variation are well documented in many crops. However, the GEI observed herein demonstrates an additional level of complexity in metabolome-genome analyses, as interactions may

confound non-interactive main effects if they are a major source of the total phenotypic variation. The GC-MS study design was limited to only two genotypes and two environments, and nevertheless GEI effects were central to many primary metabolites. Because GEI effects are expected to increase with the addition of more varieties and environments, a more narrow approach that effectively estimates GEI may be critical to apply metabolomics to plant studies. A metabolite network was constructed to visualize GEI at a systems-level. The network also allowed for concurrent analysis of multiple gene-metabolite relationships, and the characterization of previously undefined links among primary metabolites.

The UPLC-MS metabolomics approach assessed variation in secondary metabolites. IR64 and Moroberekan varied in secondary metabolite profiles, however the total variation was lower than previously observed in rice. Only 4.4% of the molecular features ($p < 0.001$) differed, compared to 25% when 10 rice varieties were assessed together [1]. The UPLC-MS analysis detected 11 metabolites that exhibited a greater than 2-fold difference between IR64 and Moroberekan (Figure 3.1D), and all 11 were in higher amounts in Moroberekan. The two varieties differed in only one flavonoid, a metabolite class with implications for both human health and plant immunity. Flavonoids are generally known for facilitating plant immunity in dicots, but are also important for monocots. Rice flavonoids are important for resistance to UV-irradiation and to fungal and bacterial pathogens [27,28,29]. In other cereals, flavonoids provide protection against pathogenic nematodes [30]. In rice, terpenoids and flavonoids act synergistically in plant defense responses [31]. Furthermore, the structure of flavonoids results in high antioxidant activity [32], and thus they are predicted to act as bioactive compounds in the human diet. The UPLC-MS detected flavonoid was characterized as a glycoside with two sugar molecules (ribose and glucose) based on the accurate masses of previously reported fragmentation patterns [33,34]. Flavonoid glycosides have altered antioxidant

activity and bioavailability [32,35]. Because IR64 and Moroberekan vary in secondary metabolites, identifying primary metabolic networks that influence these compounds is important when breeding for specific secondary metabolite profiles.

The GC-MS analysis identified variation in primary metabolites due to genetic, environmental, and GEI effects. As with the UPLC-MS volcano analysis (Figure 3.1C), the GC-MS volcano plot (Figure 3.2B, 3.2C) was skewed, and showed more features for 'Moroberekan' and 'field' effects. The z-score analysis (Figure 3.3) established the greenhouse as a fixed effect, and suggests that a greenhouse environment induced variation in many of the primary metabolites. The influence of growing environment was expected, as temperature is known to alter metabolite profiles of plants, including changes in lipids and phenolics due to cold stress [36,37,38]. Abscisic acid, a key plant abiotic stress hormone, alters terpenoid content in plants [39], and in barley, the content of flavonoid glycosides differed according to type of nutrient application [40]. Additionally, the genetic regulation of such environmental influences has been described in rice [41,42]. This suggests that genetic diversity in such regulators may account for GEI effects and in the distribution of the effect across the metabolome.

Gene expression models were recently reviewed as a valuable component of systems-level analyses [43]. Gene expression analysis was used to determine variation in ADT/PDT genes, as they are the genes most likely to interact with environmental factors in phenylalanine synthesis. The gene expression profiles for phenylalanine biosynthesis (Figures 3.4 and 3.5) showed many interesting trends. The upregulation of the 11 ADT/PDT genes varied among many of the libraries. This indicates that total cellular ADT/PDT function may be due to many genes, and is specific for a given situation, such as seed development, pathogen resistance, or abiotic stress. For example, resistance to *M. grisea* appears to be primarily dependent on LOC_Os04g33380 and LOC_Os03g17730, for which gene expression was much

greater in resistant plants compared to susceptible. For the developing seed, LOC_Os07g49390 was the most abundant among the five rice accessions in the MPSS database. Furthermore, LOC_Os07g49390 contains SNPs in the 3' UTR for the varieties Cypress and Nipponbare, the same gene for which a rice mutant exhibited enhanced phenylalanine synthesis [19]. Cypress had much greater expression of LOC_Os09g39230 compared to the other accessions. Thus, not only does total ADT/PDT expression vary, but varieties rely on different genes to achieve similar ADT/PDT function. This may explain the GEI observed in phenylalanine, as environmental influences may be dependent on the genetic background, and epistatically affect gene regions containing different ADT/PDT genes. In addition, the gene interaction model predicted interactions between LOC_Os07g49390 and six other genes. Four of the genes are involved in tyrosine and serine biosynthetic processes, and both metabolites contained similar GEI patterns to phenylalanine (Figure 3.6). Like the ADT/PDT gene LOC_Os07g4939, the serine biosynthesis gene LOC_Os03g06200 had lower expression in Cypress compared to Nipponbare. As with the ADT/PDT heat map (Figure 3.4), Cypress exhibited greater expression in an alternate biosynthesis gene, LOC_Os04g55720 (Figure 3.5). Together, these data support the importance of genetic background in primary and secondary metabolism.

The GEI observed in the GC-MS profiling suggests that the extent to which environment affects more complex metabolites (flavonoids) is largely influenced by genotype. A network approach was thus employed to determine relationships among metabolites, and a network of amino acids was constructed for metabolites detected by GC-MS. (Figure 3.6). Glutamate was the most central node in the network, which is expected given its role in transaminase reactions. In rice, glutamate is critical in amino acid metabolism, but is also associated regulating other primary and secondary metabolite classes [44]. While no genetic or GEI effects were observed for glutamate,

the significant environmental effect may alter regulation of other amino acids. For serine and glycine, which are substrates for each other's synthesis, Moroberekan showed opposing changes in quantity due to growing environment, and mildly so with IR64. Phenylalanine showed similar changes as serine (Figure 3.6), and the genes involved in serine metabolism were predicted to interact with the phenylalanine ADT/PDT synthesis gene (Table 3.3). The importance of phenylalanine in secondary metabolite synthesis suggests that serine may also be a critical component in breeding efforts to alter quantities of metabolites in the grain.

Developing crops with highly specific secondary metabolite profiles in the cooked grain is a breeding strategy of growing interest [45]. Metabolite and networks have been described as an innovative addition to explain variation in breeding [46]. A number of studies document the importance of genetic diversity in metabolite networks [47,48,49]. Integrating genetic and metabolomic analyses into networks can provide substantial predictive power of genotypic influences [50]. For plant breeding, it is important to characterize the genetic components that contribute to GEI to evaluate the heritability of a metabolite profile. The gene/metabolite network approach may be useful to explain genotypic, environmental, and GEI effects on primary and secondary metabolism.

**ACKNOWLEDGEMENTS**

Support for this student training project was provided by the USDA National Needs Graduate Fellowship Competitive Grant No. 2007-38420-17766 from the National Institute of Food and Agriculture, Colorado Agricultural Experiment Station, National Institute for Health, and the Bill and Melinda Gates Foundation.

## FIGURES

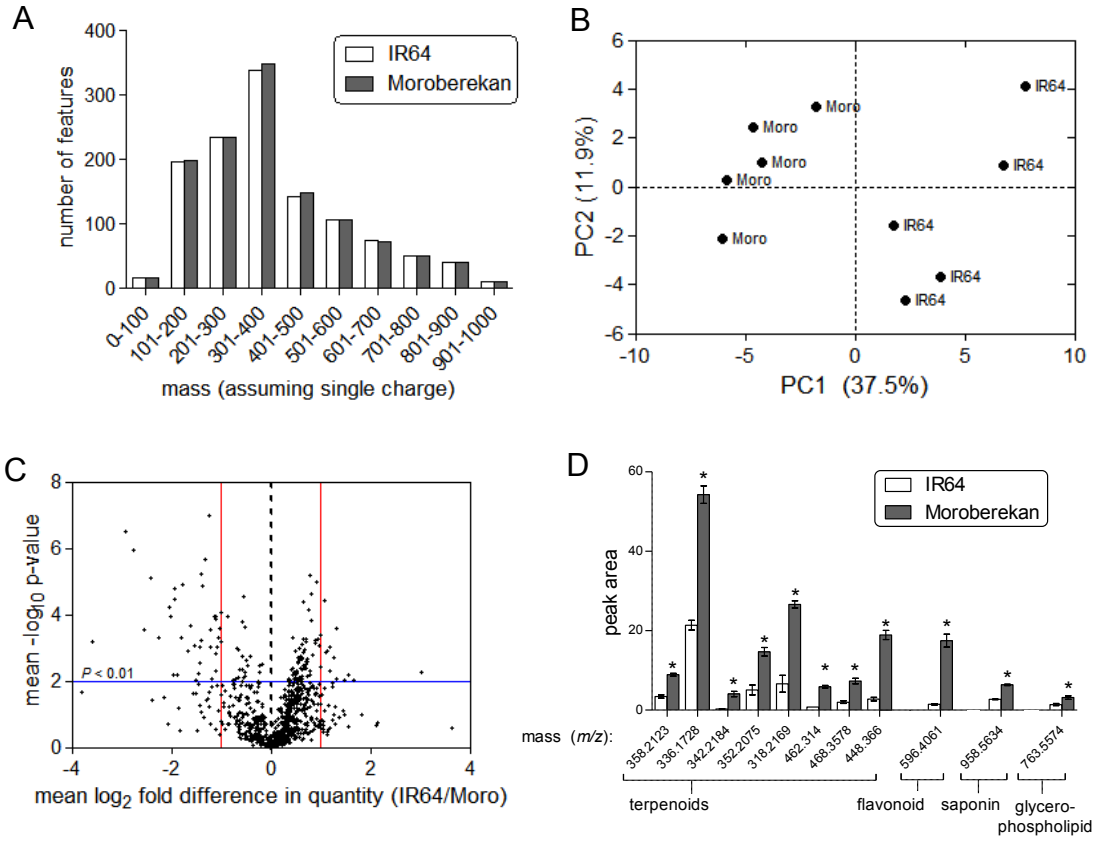


**Figure 3.1**. **Differences between IR64 and Moroberekan for metabolites detected by UPLC-MS.** (A) Mass distribution of molecular features from IR64 and Moroberekan. (B) Principal component analysis of approximately 1200 molecular features from cooked rice extracts show metabolome differences between IR64 and Moroberekan (n=5). (C) Volcano plot showing each molecular feature plotted by fold-difference in quantity between IR64 and Moroberekan (x-axis) and p-value calculated by student's t-test (y-axis). Vertical red bars indicate cutoff for 2-fold differences in molecular feature quantities, and the horizontal blue bar indicates a cutoff for $p < 0.01$. (D) Molecular features with greater than 2-fold differences between IR64 and Moroberekan (students t-test, *$p < 0.05$), whereby masses listed are of the predicted metabolite mass with no charge. Values are expressed as mean peak area ± SEM, and metabolite classes were deduced by matching masses and mass spectra to metabolite databases.

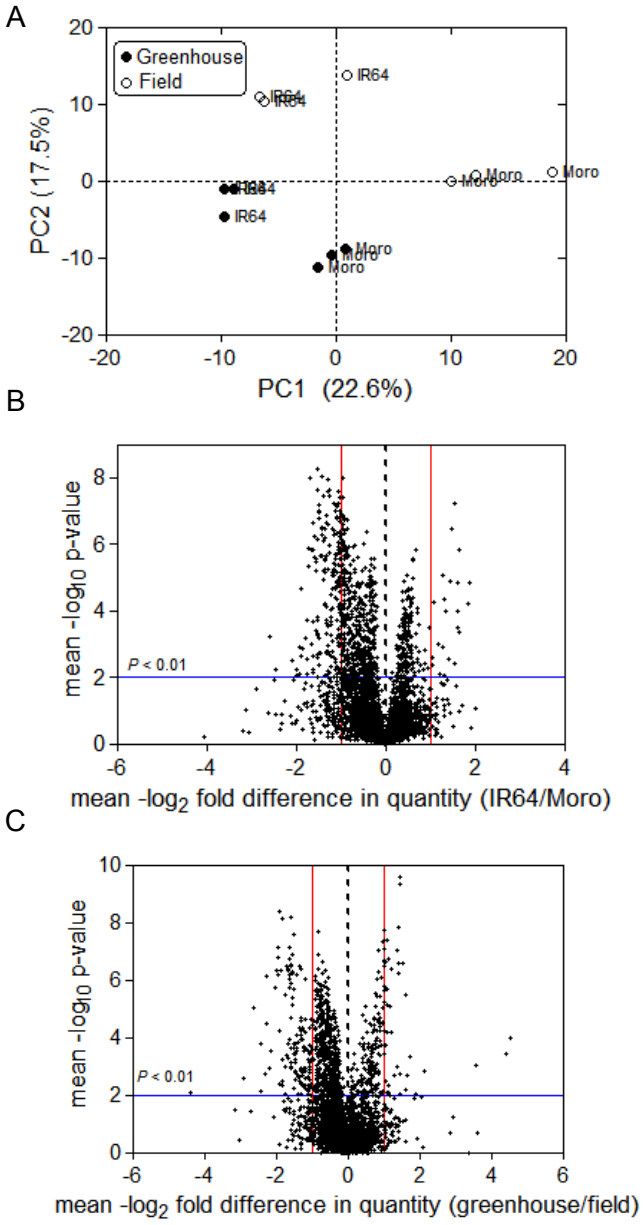


**Figure 3.2. Differences between IR64 and Moroberekan for metabolites detected by GC-MS from plants grown in the field and greenhouse.** (A) Principal component analysis of 4972 molecular features suggests grain metabolite variation due to both genotypic and environmental effects. Molecular features were assessed for differences among (C) genotype and (D) environment based on fold-changes in feature quantities and p-values (student's t-test).

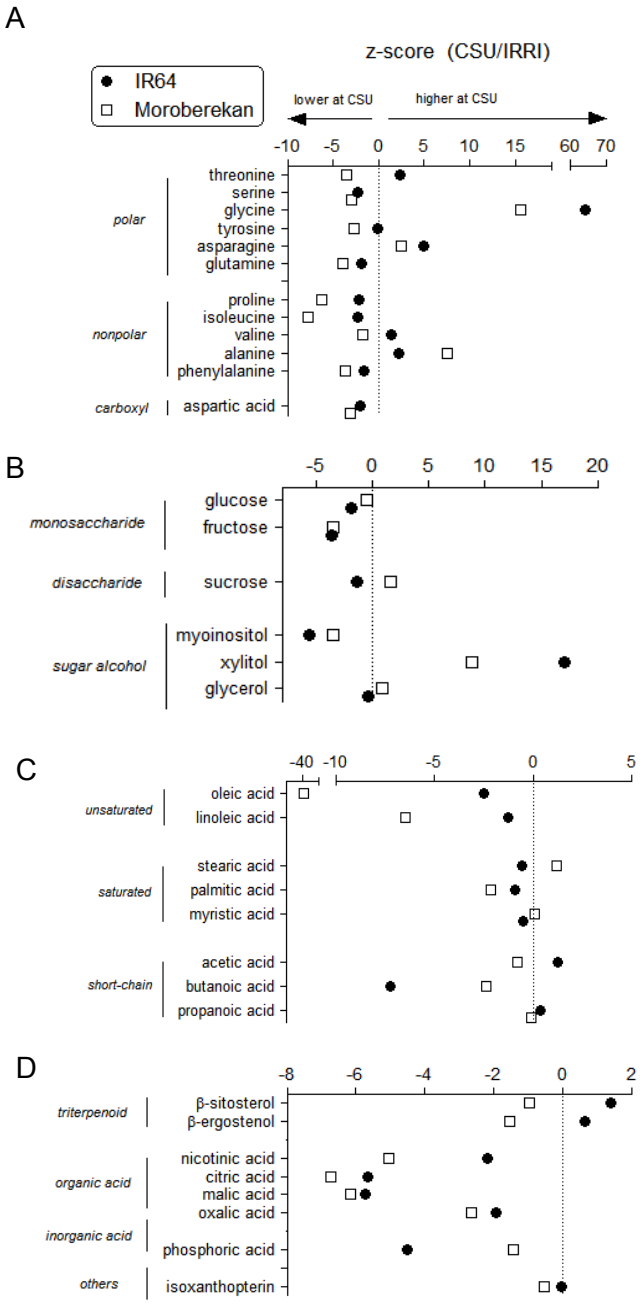


**Figure 3.3. Z-score analysis to assess metabolite variation associated with a greenhouse environment.** Z-scores were calculated for metabolites that were classified based on mass spectral database matching as an (A) amino acid (B) sugar (C) fatty acid or (D) triterpenoid, organic and inorganic acids, and other compounds. Z-scores represent the number of standard deviations each metabolite differed by growing IR64 or Moroberekan in the greenhouse (CSU), compared to field-grown rice (IRRI) as a control.

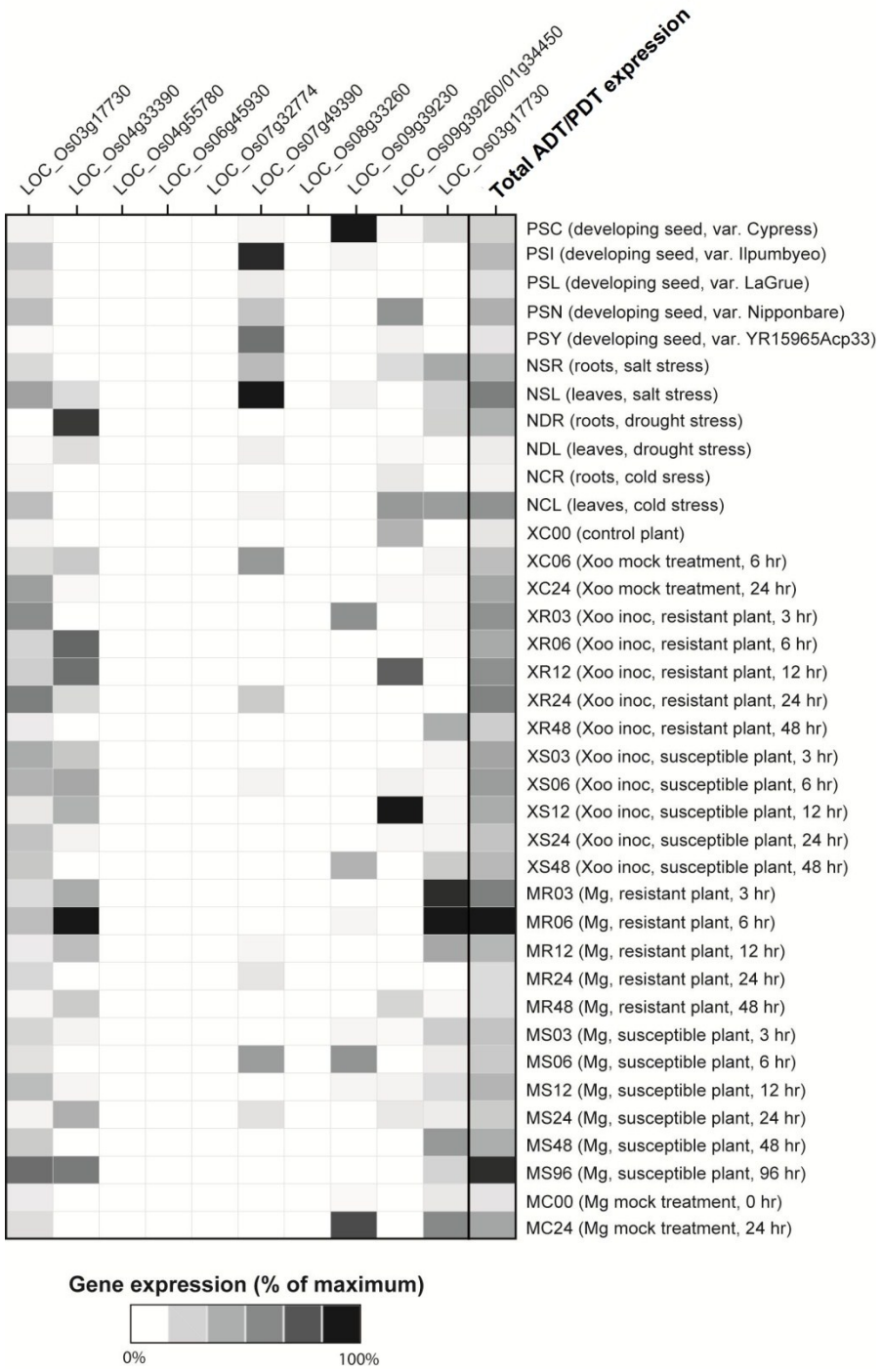


**Figure 3.4. Heat map of ADT/PDT gene expression.** The 10 ADT/PDT genes showed variation in expression among rice MPSS libraries for the developing seed, and abiotic and biotic stresses. Shade represents the percent of gene expression relative to the library with the maximum gene expression, assessed by transcript copy number. Rice pathogen libraries were for *Xanthomonas oryzae* pv. *oryzae* (Xoo) and *Magnaporthe grisea* (Mg).

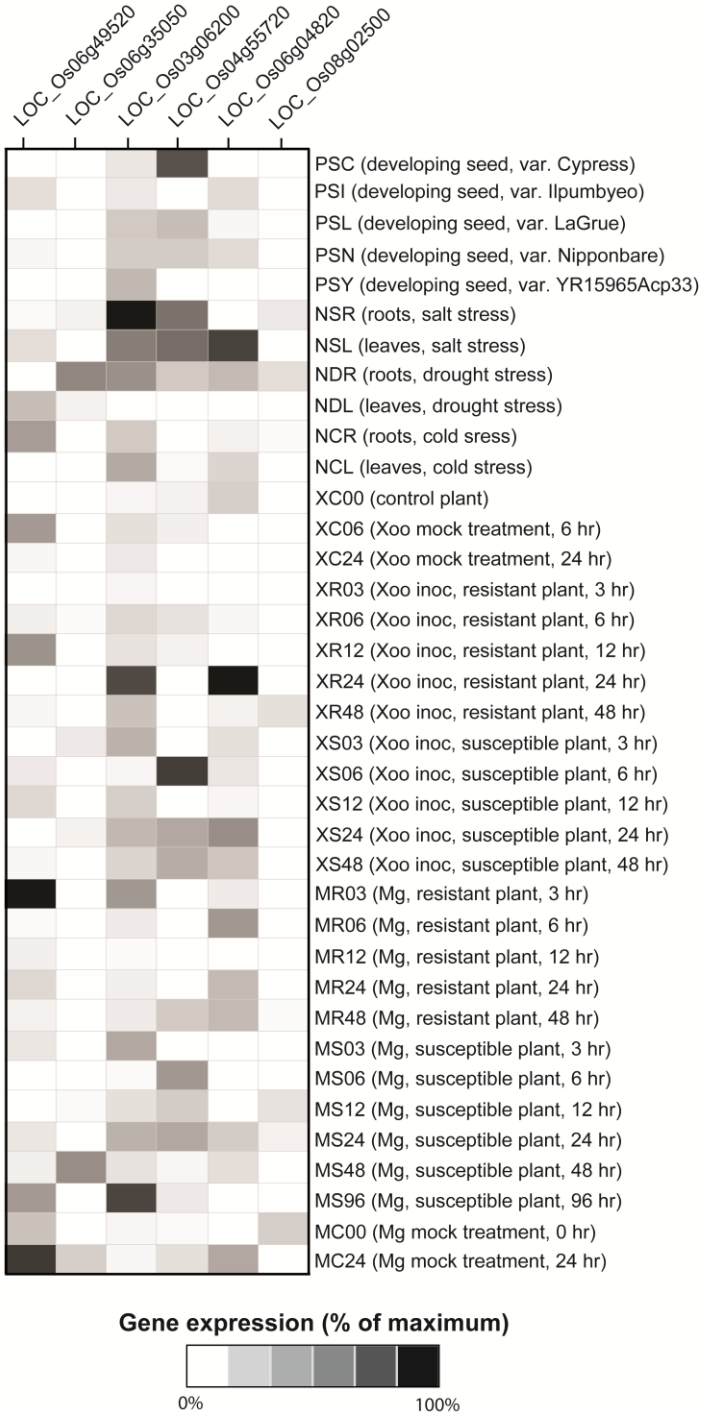


**Figure 3.5. Heat map of genes predicted to interact with Loc_Os07g49390.** Gene expression proflies for the six genes predicted to interact with Loc_Os07g49390, described in Table 3. Shade represents the percent of gene expression relative to the library with the maximum gene expression, assessed by transcript copy number. Rice pathogen libraries were for *Xanthomonas oryzae* pv. *oryzae* (Xoo) and *Magnaporthe grisea* (Mg).

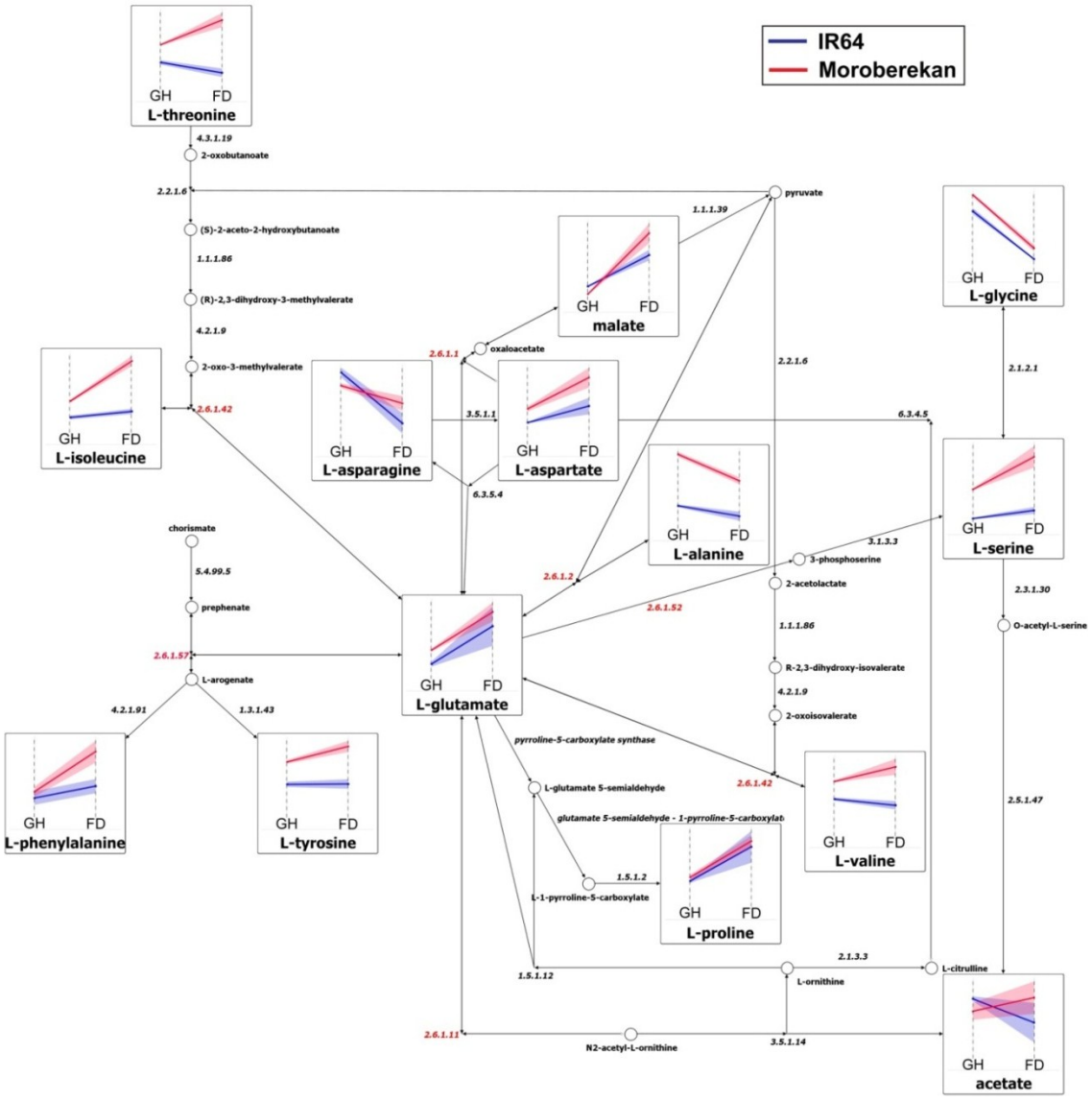


**Figure 3.6. Amino acid network for GC-MS detected metabolites from IR64 and Moroberekan grown in two environments.** Amino acid synthesis pathways from the MetaCrop database [25] were linked using shared amino acids, metabolic intermediates of amino acid biosynthesis, and related enzymes. Enzymes are represented as Enzyme Commission number [51], and red lettering indicates chemical reactions that use glutamate as a substrate in a transaminase reaction. GC-MS detected metabolites for IR64 (blue) and Moroberekan (red) are displayed for greenhouse-grown (GH) and field-grown rice (FD). Variation due to G, E, and GEI are described in Table 1.

**Table 3.1. Quantities (peak area units) for metabolites detected by GC-MS and effect of genotype (G), growing environment (E), and genotype-environment interactions (GEI).**

| | IR64 | | Moroberekan | | |
|---|---|---|---|---|---|
| **Amino Acids** | **Greenhouse** | **Field** | **Greenhouse** | **Field** | Effect p<0.05 |
| alanine | 110.76 ± 2.43 | 77.92 ± 8.42 | 273.98 ± 4.31 | 190.44 ± 6.41 | G, E, GEI |
| valine | 51.48 ± 0.98 | 44.97 ± 2.6 | 70.49 ± 0.6 | 85.88 ± 4.92 | G, GEI |
| isoleucine | 16.66 ± 0.66 | 20.35 ± 0.94 | 26.82 ± 0.54 | 51.36 ± 1.83 | G, E, GEI |
| glycine | 135.02 ± 3.47 | 43.83 ± 0.82 | 165.89 ± 1.81 | 64.52 ± 3.77 | G, E |
| serine | 28.76 ± 0.8 | 43.16 ± 3.62 | 80.54 ± 0.9 | 138.33 ± 11.04 | G, E, GEI |
| threonine | 18.06 ± 0.59 | 13.31 ± 1.14 | 26.26 ± 0.22 | 37.61 ± 1.88 | G, E, GEI |
| aspartic acid | 76.55 ± 0.9 | 113.57 ± 10.87 | 107.22 ± 3.22 | 177.26 ± 12.77 | G, E |
| proline | 65.56 ± 1.8 | 123.23 ± 15.5 | 71.89 ± 2.5 | 132.77 ± 5.67 | E |
| glutamate | 35.32 ± 1.99 | 84.44 ± 14.73 | 53.49 ± 0.87 | 102.97 ± 7.18 | E |
| asparagine | 46.16 ± 1.51 | 18.24 ± 3.26 | 38.93 ± 0.74 | 29.21 ± 2.31 | E, GEI |
| phenylalanine | 17.89 ± 2.21 | 24.66 ± 2.35 | 21.53 ± 1.65 | 44.26 ± 3.57 | G, E, GEI |
| tyrosine | 16.95 ± 0.58 | 17.07 ± 1.04 | 25.37 ± 0.38 | 31.17 ± 1.22 | G, E, GEI |
| **Sugars** | | | | | |
| glycerol | 1958 ± 693 | 2299 ± 605 | 1949 ± 677 | 1406 ± 374 | ns |
| xylitol | 55.53 ± 1.63 | 23.68 ± 1.08 | 53.23 ± 0.3 | 30 ± 1.52 | E, GEI |
| fructose | 31.26 ± 1.43 | 67.08 ± 5.68 | 120.67 ± 0.9 | 139.38 ± 3.15 | G, E, GEI |
| galactose | 75.39 ± 3.46 | 143.11 ± 7.44 | 187.11 ± 1.8 | 231.51 ± 8.05 | G, E |
| glucose | 37.82 ± 2.07 | 45.1 ± 2.33 | 40.98 ± 0.46 | 42.81 ± 2.31 | E |
| myoinositol | 26.35 ± 0.74 | 49.95 ± 2.42 | 67.84 ± 1.36 | 92.11 ± 3.98 | G, E |
| sucrose | 3075 ± 87 | 3367.3 ± 124 | 4197 ± 104.7 | 3777 ± 149 | G, GEI |
| **Fatty Acids** | | | | | |
| propanoic acid | 165.73 ± 41.85 | 153.21 ± 20.2 | 127.09 ± 28.65 | 129.69 ± 15.42 | ns |
| acetic acid | 27.5 ± 0.54 | 18.34 ± 4.38 | 22.66 ± 1.89 | 27.96 ± 3.61 | GEI |
| butanoic acid | 99.98 ± 4.55 | 284.18 ± 14.7 | 315.33 ± 6 | 417.63 ± 24.83 | G, E, GEI |
| myristic acid | 18.61 ± 2.13 | 20.56 ± 2.21 | 14.57 ± 1.41 | 14.53 ± 0.52 | G |
| palmitic acid | 167.8 ± 18.68 | 202.07 ± 20.75 | 170.34 ± 17.41 | 190.08 ± 5.22 | ns |
| linoleic acid | 105.85 ± 5.96 | 123.15 ± 7.62 | 86.27 ± 2.58 | 118.68 ± 2.89 | G, E |
| oleic acid | 18.89 ± 0.61 | 29.63 ± 2.49 | 21.88 ± 0.33 | 38.1 ± 0.24 | G, E |
| stearic acid | 185.88 ± 41.92 | 227.92 ± 42.3 | 174.84 ± 26.66 | 157.12 ± 8.95 | ns |
| **Others** | | | | | |
| isoxanthopterin | 21.82 ± 1.16 | 22.17 ± 4.86 | 6.34 ± 0.12 | 7.74 ± 1.47 | G |
| beta-sitosterol | 30.06 ± 3.82 | 26.51 ± 1.49 | 27.16 ± 3.46 | 31.21 ± 2.38 | ns |
| beta-ergostenol | 14.54 ± 1.58 | 13.71 ± 0.72 | 11.7 ± 0.98 | 14.42 ± 1.01 | ns |
| phosphoric acid | 64.14 ± 4.44 | 95.42 ± 4.01 | 74.9 ± 0.21 | 80.36 ± 2.19 | E, GEI |
| oxalic acid | 21.75 ± 1.04 | 28.24 ± 1.95 | 13.49 ± 1.53 | 19.96 ± 1.41 | G, E |
| nicotinic acid | 29.8 ± 1.21 | 38.39 ± 2.26 | 23.36 ± 0.85 | 40.08 ± 1.91 | E, GEI |
| malic acid | 43.01 ± 0.6 | 95.03 ± 5.26 | 30.14 ± 0.67 | 131.24 ± 9.46 | E, GEI |
| citric acid | 123.15 ± 2.43 | 232 ± 11.24 | 105.31 ± 0.87 | 144.61 ± 3.37 | G, E, GEI |
| pyridine | 98.05 ± 2.28 | 96 ± 3.34 | 105.66 ± 4.5 | 97.44 ± 4.3 | ns |

**Table 3.2. Genes associated with or ADT/PDT activity in rice.**

| Rice PDT/ADT | SNP ID | Ch. | Coordinate | Class | Cyp. | Nipp. | IR64 | Moro. |
|---|---|---|---|---|---|---|---|---|
| LOC_Os01g34450 | no snp | 1 | - | - | - | - | - | - |
| LOC_Os03g17730 | no snp | 3 | - | - | - | - | - | - |
| LOC_Os04g33390 | no snp | 4 | - | - | - | - | - | - |
| LOC_Os04g55780 | no snp | 4 | - | - | - | - | - | - |
| LOC_Os06g45930 | no snp | 6 | - | - | - | - | - | - |
| LOC_Os07g49390 | TBGI337566 | 7 | 29578537 | 3' UTR | C | T | N | N |
| LOC_Os07g49390 | TBGI337567 | 7 | 29578969 | 3' UTR | T | A | N | N |
| LOC_Os07g32774 | TBGI326181 | 7 | 19597711 | intron | G | A | G | G |
| LOC_Os07g32774 | TBGI326183 | 7 | 19598604 | syn SNP | A | G | G | G |
| LOC_Os07g32774 | TBGI326187 | 7 | 19599150 | intron | G | G | G | N |
| LOC_Os08g33260 | no snp | 8 | - | - | - | - | - | - |
| LOC_Os09g39230 | no snp | 9 | - | - | - | - | - | - |
| LOC_Os09g39260 | no snp | 9 | - | - | - | - | - | - |
| LOC_Os10g37980 | TBGI413593 | 10 | 20015366 | 3' UTR | A | A | G | A |
| LOC_Os10g37980 | TBGI413595 | 10 | 20016479 | syn SNP | G | G | N | G |
| LOC_Os10g37980 | TBGI413596 | 10 | 20016700 | syn SNP | G | G | A | G |

**Table 3.3. Predicted metabolite synthesis genes that interact with Loc_Os07g49390.**

| Interacting Gene | Pathway | Function | Location | Score | Evidence |
|---|---|---|---|---|---|
| LOC_Os06g49520 | tyrosine biosynthetic process | prephenate dehydrogenase | --------- | 1.33 | OS-GN |
| LOC_Os06g35050 | tyrosine biosynthetic process | prephenate dehydrogenase | --------- | 1.97 | OS-GN |
| LOC_Os03g06200 | L-serine biosynthetic process | phosphoserine transaminase | --------- | 2.61 | OS-GN |
| LOC_Os04g55720 | L-serine biosynthetic process | phosphoglycerate dehydrogenase | Mito, chloro | 2.42 | AT-DC |
| LOC_Os06g04280 | aromatic amino acid biosynthetic-shikimate pathway | ---------- | chloro | 1.96 | OS-GN |
| LOC_Os08g02500 | regulation of amino acid metabolic process, response to cytokinin stimulus, response to abscisic acid stimulus | amino acid binding | cytosol | 2.63 | AT-DC |

OS-GN: Gene neighborhoods of bacterial and archaeal orthologs of rice genes
AT-DC: Co-occurrence of domains among Arabidopsis protein
Mito: mitochrondria
Chloro: chloroplast

## REFERENCES


1. Heuberger AL, Lewis MR, Chen M-H, Brick MA, Leach JE, et al. (2010) Metabolomic and functional genomic analyses reveal varietal differences in bioactive compounds of cooked rice. PLoS ONE 5: e12915.
2. Mochida K, Furuta T, Ebana K, Shinozaki K, Kikuchi J (2009) Correlation exploration of metabolic and genomic diversity in rice. BMC Genomics 10.
3. Kusano M, Fukushima A, Kobayashi M, Hayashi N, Jonsson P, et al. (2007) Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. Journal of Chromatography B 855: 71-79.
4. Chen M-H, Bergman C, Pinson S, Fjellstrom R (2008) Waxy gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. Journal of Cereal Science 47: 536-545.
5. Shewry PR, Piironen V, Lampi A-M, Edelmann M, Kariluoto S, et al. (2010) The HEALTHGRAIN Wheat Diversity Screen: Effects of Genotype and Environment on Phytochemicals and Dietary Fiber Componentsâ€ Journal of Agricultural and Food Chemistry 58: 9291-9298.
6. Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry 68: 2831-2846.
7. Fraenkel GS (1959) The Raison d'Etre of Secondary Plant Substances. Science 129: 1466-1470.
8. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3: 408-414.
9. Treutter D (2006) Significance of flavonoids in plant resistance: a review. Environmental Chemistry Letters 4: 147-157.
10. Hollman PCH, Katan MB (1997) Absorption, metabolism and health effects of dietary flavonoids in man. Biomedecine & Pharmacotherapy 51: 305-310.
11. Grassmann J, Hippeli S, Elstner EF (2002) Plant's defence and its benefits for animals and medicine: role of phenolics and terpenoids in avoiding oxygen stress. Plant Physiol Biochem 40: 471-478.
12. Lin XB, Ma L, Racette SB, Spearie CLA, Ostlund RE (2009) Phytosterol glycosides reduce cholesterol absorption in humans. Am J Physiol Gastrointest Liver Physiol 296: G931-G935.
13. Shih C, Chu H, Tang L, Sakamoto W, Maekawa M, et al. (2008) Functional characterization of key structural genes in rice flavonoid biosynthesis. Planta 228: 1043-1054.
14. Jung E, Zamir LO, Jensen RA (1986) Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. Proceedings of the National Academy of Sciences 83: 7231-7235.
15. Maeda H, Yoo H, Dudareva N (2011) Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. Nat Chem Biol 7: 19-21.
16. Wakasa K, Ishihara A (2009) Metabolic engineering of the tryptophan and phenylalanine biosynthetic pathways in rice. Plant Biotechnology 26: 523-533.
17. Cho M-H, Corea ORA, Yang H, Bedgar DL, Laskar DD, et al. (2007) Phenylalanine Biosynthesis in Arabidopsis thaliana. Journal of Biological Chemistry 282: 30827-30835.
18. Warpeha KM, Gibbons J, Carol A, Slusser J, Tree R, et al. (2008) Adequate phenylalanine synthesis mediated by G protein is critical for protection from UV radiation damage in young etiolated Arabidopsis thaliana seedlings. Plant, Cell & Environment 31: 1756-1770.

19. Yamada T, Matsuda F, Kasai K, Fukuoka S, Kitamura K, et al. (2008) Mutation of a Rice Gene Encoding a Phenylalanine Biosynthetic Enzyme Results in Accumulation of Phenylalanine and Tryptophan. The Plant Cell Online 20: 1316-1329.
20. Zhao K, Wright M, Kimball J, Eizenga G, McClung A, et al. (2010) Genomic Diversity and Introgression in *O. sativa* Reveal the Impact of Domestication and Breeding on the Rice Genome. PLoS ONE 5: e10780.
21. McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, et al. (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. Plant Physiol 141: 26-31.
22. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci U S A 106: 12273-12278.
23. Jahn CE, McKay JK, Mauleon R, Stephens J, McNally KL, et al. (2011) Genetic Variation in Biomass Traits among 20 Diverse Rice Varieties. Plant Physiology 155: 157-168.
24. Junker BH, Klukas C, Schreiber F (2006) VANTED: A system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics 7.
25. Grafahrend-Belau E, Weise S, Koschutzki D, Scholz U, Junker BH, et al. (2008) MetaCrop: a detailed database of crop plant metabolism. Nucleic Acids Research 36: D954-D958.
26. Nobuta K, Venu RC, Lu C, Belo A, Vemaraju K, et al. (2007) An expression atlas of rice mRNAs and small RNAs. Nat Biotech 25: 473-477.
27. Kodama O, Miyakawa J, Akatsuka T, Kiyosawa S (1992) Sakuranetin, a flavanone phytoalexin from ultraviolet-irradiated rice leaves. Phytochemistry 31: 3807-3809.
28. Gandikota M, de Kochko A, Chen L, Ithal N, Fauquet C, et al. (2001) Development of transgenic rice plants expressing maize anthocyanin genes and increased blast resistance. Molecular Breeding 7: 73-83.
29. Padmavati M, Sakthivel N, Thara KV, Reddy AR (1997) Differential sensitivity of rice pathogens to growth inhibition by flavonoids. Phytochemistry 46: 499-502.
30. Soriano IR, Asenstorfer RE, Schmidt O, Riley IT (2004) Inducible Flavone in Oats (Avena sativa) Is a Novel Defense Against Plant-Parasitic Nematodes. Phytopathology 94: 1207-1214.
31. Edwards WR, Hall JA, Rowlan AR, Schneider-Barfield T, Sun TJ, et al. (2008) Light filtering by epidermal flavonoids during the resistant response of cotton to Xanthomonas protects leaf tissue from light-dependent phytoalexin toxicity. Phytochemistry 69: 2320-2328.
32. Rice-Evans CA, Miller NJ, Paganga G (1996) Structure-antioxidant activity relationships of flavonoids and phenolic acids. Free Radical Biology and Medicine 20: 933-956.
33. March RE, Lewars EG, Stadey CJ, Miao X-S, Zhao X, et al. (2006) A comparison of flavonoid glycosides by electrospray tandem mass spectrometry. International Journal of Mass Spectrometry 248: 61-85.
34. Vukics V, Guttman A (2008) Structural characterization of flavonoid glycosides by multi-stage mass spectrometry. Mass Spectrometry Reviews 29: 1-16.
35. Graefe EU, Wittig J, Mueller S, Riethling AK, Uehleke B, et al. (2001) Pharmacokinetics and bioavailability of quercetin glycosides in humans. The Journal of Clinical Pharmacology 41: 492-499.
36. Miquel M, James D, Dooner H, Browse J (1993) Arabidopsis requires polyunsaturated lipids for low-temperature survival. Proceedings of the National Academy of Sciences 90: 6208-6212.

37. Solecka D, Boudet A-M, Kacperska A (1999) Phenylpropanoid and anthocyanin changes in low-temperature treated winter oilseed rape leaves. Plant Physiology and Biochemistry 37: 491-496.
38. Nishida I, Murata N (1996) Chilling Sensitivity in Plants and Cyanobacteria:The Crucial Contribution of Membrane Lipids. Annual Review of Plant Physiology and Plant Molecular Biology 47: 541-568.
39. Mansouri H, Asrar Z, Szopa J (2009) Effects of ABA on primary terpenoids and Δ9-tetrahydrocannabinol *Cannabis sativa* L. at flowering stage. Plant Growth Regulation 58: 269-277.
40. Aaboer DBF, Bleeg IS, Christensen BT, Kondo T, Brandt K (2002) Flavone C-Glycoside, Phenolic Acid, and Nitrogen Contents in Leaves of Barley Subject to Organic Fertilization Treatments. Journal of Agricultural and Food Chemistry 51: 809-813.
41. Wang J, Ming F, Pittman J, Han Y, Hu J, et al. (2006) Characterization of a rice (Oryza sativa L.) gene encoding a temperature-dependent chloroplast [omega]-3 fatty acid desaturase. Biochemical and Biophysical Research Communications 340: 1209-1216.
42. Ithal N, Reddy AR (2004) Rice flavonoid pathway genes, OsDfr and OsAns, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, OsC1-MYB. Plant Science 166: 1505-1513.
43. Aoki K, Ogata Y, Shibata D (2007) Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. Plant and Cell Physiology 48: 381-390.
44. Kusano M, Tabuchi M, Fukushima A, Funayama K, Diaz C, et al. (2011) Metabolomics data reveal a crucial role of cytosolic glutamine synthetase 1;1 in coordinating metabolic balance in rice. The Plant Journal 66: 456-466.
45. Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends in Genetics 25: 39-48.
46. Keurentjes JJB (2009) Genetical metabolomics: closing in on phenotypes. Current Opinion in Plant Biology 12: 223-230.
47. Keurentjes JJB, Sulpice R, Gibon Y, Steinhauser MC, Fu JY, et al. (2008) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. Genome Biology 9.
48. Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, et al. (2007) Linking Metabolic QTLs with Network and <italic>cis</italic>-eQTLs Controlling Biosynthetic Pathways. PLoS Genet 3: e162.
49. Allen E, Moing A, Ebbels T, Maucourt M, Tomos AD, et al. (2010) Correlation Network Analysis reveals a sequential reorganization of metabolic and transcriptional states during germination and gene-metabolite relationships in developing seedlings of Arabidopsis. BMC Systems Biology 4: 62.
50. Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics 26: i255-i260.
51. Bairoch A (2000) The ENZYME database in 2000. Nucleic Acids Research 28: 304-305.

# CHAPTER FOUR: RICE BRAN FERMENTED WITH *SACCHAROMYCES BOULARDII* GENERATES NOVEL METABOLITE PROFILES WITH BIOACTIVITY

A modified version of this chapter is published as Ryan EP, Heuberger AL, Weir TL, Barnett B, Broeckling CD, Prenni JE. *Rice Bran Fermented with* Saccharomyces boulardii *Generates Novel Metabolite Profiles with Bioactivity*. Journal of Agricultural and Food Chemistry. 2011. 59 (5): 1862-1870.

**SUMMARY**

Emerging evidence supporting chronic disease fighting properties of rice bran has advanced the development of rice bran for human use as a functional food and dietary supplement. A global and targeted metabolomic investigation of stabilized rice bran fermented with *Saccharomyces boulardii* was performed in three rice varieties. Metabolites from *S. boulardii*-fermented rice bran were detected by gas chromatography-mass spectrometry (GC-MS) and assessed for bioactivity compared to non-fermented rice bran. Global metabolite profiling revealed significant differences in the metabolome that led to discovery of candidate compounds modulated by *S. boulardii* fermentation. Fermented rice bran extracts from three rice varieties reduced growth of human B lymphomas compared to each variety's non-fermented control and revealed that fermentation differentially altered bioactive compounds. These data support that integration of global and targeted metabolite analysis can be utilized for assessing health properties of rice bran phytochemicals that are both enhanced by yeast fermentation and that differ across rice varieties.

## INTRODUCTION

Rice bran is a unique whole food that naturally contains protein, vitamins, minerals, complex carbohydrates, phytonutrients, phospholipids, essential fatty acids, and more than 120 antioxidants (1). Dietary rice bran intake and rice bran components have demonstrated chronic disease fighting activity, particularly for protection against cardiovascular disease and certain cancers (2-8). We and others have shown that rice varieties are not equal in content and composition of bioactive rice bran components (9, 10). How these phytochemicals are altered by microbial fermentation and metabolism is an emerging area of research that merits scientific investigation when assessing bioactivity and health benefits. A few studies have evaluated rice bran as a dietary supplement or functional food ingredient (5, 11-13); however, little is known about how chemical content changes with and without fermentation.

The yeast, *Saccharomyces cerevisiae* var. *boulardii (S. boulardii)* has probiotic activity and is widely used as a dietary supplement for intestinal disease prevention and treatment (14-16). The spectrum of biomedical activities and food processing applications reported with *S. boulardii* has significantly grown over the last decade and includes, but is not limited to, protection against enteric pathogens, modification of lymphocyte proliferation, as well as differential release of plant secondary metabolites from foods such as wine, sourdough and cheese (17-19). *Sacchromyces boulardii* has been shown to be beneficial for modification of food components such as breakdown of dietary phytate and biofortification of folate to improve the nutritional value and health properties of food (20, 21). The health benefit of *S. boulardii* as both a probiotic and for fermented foods was recently reviewed and a meta-analysis of placebo-controlled treatment trials supports its safety and efficacy for protection against several types of diarrhea (16, 22, 23). Protection against specific enteric bacterial pathogens by *S. boulardii* may, in part, be due to anti-inflammatory actions and effects on immunity (15,

19, 23). *In vitro* studies using mammalian cell cultures have shown that *S. boulardii* modifies host cell signaling pathways associated with pro-inflammatory responses, and that the mechanism may be based on blocking activation of nuclear factor-kappa B (NFκB) and mitogen activated protein kinase (MAPK) (24, 25). Inhibition of these cell-signaling pathways is also an important mechanism for reducing cancer cell growth. Rice bran components have been reported to inhibit activation and promote apoptosis of malignant lymphocytes and to inhibit growth of intestinal cancers (3, 11, 26, 27). In this report, we examined the effects of *S. boulardii* fermented rice bran across rice varieties on viability of human B lymphoma *in vitro*. Rice bran chemical contents and the compounds altered by fermentation have not been previously assessed for effects on human B lymphomas, and were assessed using global and targeted metabolite profiling techniques.

A significant lack of knowledge exists regarding the ability of probiotics to alter the phytochemistry of rice bran for health benefit, and global metabolite profiling represents a novel approach to detect changes in rice bran phytochemical content due to fermentation without a bias towards certain chemical classes. A metabolite profiling approach based on gas chromatography−mass spectrometry (GC-MS) was recently used to investigate time-dependent metabolic changes during the germination of rice (*28*) and more targeted studies have sought to identify bioactive and volatile compounds from rice bran oil or in bran polished from red and black rice varieties (9, 29). Bran from three rice varieties was predicted to vary in bioactive chemical content after fermentation with *S. boulardii* and to differentially inhibit human B lymphoma viability.

## MATERIALS AND METHODS

### Reagents and Cell Culture

Caffeic acid, p-coumaric acid, ferulic acid, salicylic acid, β-sitosterol, and α-tocopherol standards were purchased from Sigma-Aldrich (St. Louis, MO). *Saccharomyces boulardii* was isolated from the commercial probiotic, Proboulardi® (Metagenics Inc., San Clemente, CA), and confirmed by morphological tests. Cultures were maintained on yeast nitrogen base (YNB) amended with 0.5% (w/v) ammonium sulfate and 2% (w/v) dextrose. Raji B lymphomas were purchased from American Type Culture Collection. Whole blood from healthy volunteers was collected into 8ml Cell Preparation Tubes (CPT) with sodium citrate as an anticoagulent (Becton Dickinson Vacutainer Systems, Franklin Lakes, NJ). CPT tubes were centrifuged at 1500 × g for 30 minutes for separation of normal human peripheral blood lymphocytes. Raji B lymphomas and freshly isolated normal blood lymphocytes were cultured in RPMI medium supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 10 mg/mL penicillin, 10,000 IU/mL streptomycin, 25mg/mL amphotericin, 1 mM sodium pyruvate, and 1x MEM nonessential amino acids.

### Rice Bran Collection and Storage

Three rice varieties selected for investigation were Neptune, Wells, and Red Wells (Table 4.1). Red Wells is derived from Wells, and the line resulted from a spontaneous mutation that confers a red-colored seed coat. Red Wells is assumed to be near-isogenic to Wells (30). Rice bran was provided as a generous gift from Dr. Anna McClung at the United States Department of Agriculture, Rice Research Center (Stuttgart, Arkansas). Bran was isolated by standard milling process, heat stabilized at 110°C for 3 min, and stored at -20°C.

### *S. boulardii* Fermentation and Metabolite Extraction

Rice bran, water and probiotic yeast fermentations were carried out using a modification of the methods described in (*31, 32*). Briefly, 1.6 g of rice bran was added to 11.4 mL of sterilized water in the presence and absence of *S. boulardii* concentration of $6x10^5$ cells $mL^{-1}$, and samples were incubated at 37°C for 24 hrs with gentle shaking (n=3). Metabolites were extracted using two separate solvents: either (A) isopropanol:acetonitrile:water (3:2:2) for metabolite profiling or (B) methanol:water (80:20) for measuring bioactivity of *in vitro* lymphoma cultures. Solvent A was used for metabolite profiling and was previously shown to extract both lipids and organic acids (*33, 34*). Solvent B was used to standardize cell culture treatments and conditions, as a similar single-phase aqueous-alcohol solvent was previously used to assess effects of rice bran compounds (*26, 27*). After 24 hours of fermentation in water, either isopropanol:acetonitrile or methanol was added to the culture for final 3:2:2 or 80:20 ratios, respectively. Samples were vortexed, incubated at room temperature for five minutes, and bran material and yeast cells were pelleted using centrifugation (1500 × g) for ten minutes followed by filtration. The supernatant was collected and stored at -80°C until further chemical and biological analyses.

### Gas chromatography-mass spectrometry

Rice bran metabolites were detected by transferring 500 µL of extract to a new tube and dried using a vacuum centrifuge. The extract was derivatized by first adding 50 µL of a solution containing 20 mg/mL of methoxyamine hydrochloride in pyridine and incubating at 37°C for two hours. Next, 50 µL of N-methyl-N-trimethylsilyltrifluoroacetamide with 1% trimethylchlorosilane (MSTFA + 1% TMCS)

(Thermo Scientific) was added and the reaction was incubated at 37°C for 60 min. Samples were centrifuged at 3000 × g for 5 min, and 80 µL of the supernatant was used for GC-MS analysis. Caffeic acid, coumaric acid, ferulic acid, salicylic acid, β-sitosterol, and α-tocopherol standards (Sigma-Aldrich, St. Louis, MO) were dissolved in an isopropanol/acetonitrile/water solution (3:2:2), evaporated, and derivatized under identical conditions.

The derivatized samples were equilibrated to room temperature, transferred to a 200 µL glass insert, and analyzed using a Trace GC Ultra coupled to a Thermo DSQ II scanning from m/z 50–650 at a rate of 5 scans/second in electron impact mode. Samples were injected at a 10:1 split ratio, and the inlet and transfer line were held at 280°C. Separation was achieved on a 30m TG-5MS column (Thermo Scientific, 0.25 mm ID, 0.25 µm film thickness) using a temperature program of 80°C for 0.5 min, then ramped at 15°C per minute to 330°C and held for 8 minutes, at a constant flow of 1.2 mL per minute. A single feature or known metabolite was defined by a given metabolite's retention time and mass, and the peak area was used to determine the relative quantity of each feature or known metabolite.

**Cell Viability**

Raji B lymphomas and normal peripheral blood lymphocytes were plated to a density of $2.5x10^5$ cells per mL. Rice bran extracts were dried in a vacuum centrifuge, resuspended in cell culture medium, and cells were incubated in the presence of rice bran extract for 24 hours. Cells were centrifuged at 1500 × g for 5 minutes, resuspended in a solution consisting of cell culture media and 1% resazurin sodium salt, and incubated at 37°C for one hour. Fluorescence was measured at 765nm and viability was expressed as percent fluorescence relative to the vehicle control (*35, 36*).

### *S. boulardii* growth on rice bran

*Saccharomyces boulardii* cultures were maintained in YNB (MP Biomedicals, Solon, OH) with 0.5% ammonium sulfate and 2% dextrose at 37°C. A liquid growth medium containing 5% rice bran and water was made with each rice variety. *S. boulardii* was added to the rice bran/water mixture at a final $OD_{600}$ of 0.02 (approximately $6x10^5$ cells $mL^{-1}$). Cultures were incubated at 37°C and sampled at 24, 48, and 72 hours. Yeast cells were enumerated by drop plating serial dilutions on YNB plates to determine total colony forming units (CFUs).

### Statistical Analysis

Chromatographic peaks between 2 and 25 min were detected by GC-MS and aligned using MarkerLynx software (Waters, Millford, MA, USA) with a retention time error window of 0.05 min. Masses used for analyses ranged between 50 and 650 m/z with a mass error tolerance of 0.4 m/z. Multivariate statistical analysis was performed using SIMCA P+ (v 12.0, Umetrics, Umeå, Sweden). Mean centering and pareto scaling was applied for all principal component, partial least squares, and orthoganol projection to latent structures (OPLS) analyses. Each feature was analyzed independently in a linear mixed-effects model to determine the significance and percent variance attributed to fermentation (fixed effect), or variety and variety-fermentation interactions (random effects). Significance was determined with a *P*-value threshold of 0.05, and percent variation was determined using the sum of squares partitions of each random effect relative to the total sum of squares of the model. Fold changes due to fermentation were calculated for each feature using the peak areas of fermented divided by the non-fermented.. Bioactive compounds were compared among varieties by one-way ANOVA

($P < 0.05$), and z-scores were calculated for metabolites from fermented varieties based on the mean and standard deviation of the non-fermented control. Effects on lymphoma viability and increased growth of *S. boulardii* on rice bran varieties were determined using a one-way ANOVA and Tukey's HSD. Significant differences between treatments (rice bran varieties) and controls (YNB) were confirmed by a Dunnet's 2-tailed comparison. These tests were performed using R software (v2.11.1), GraphPad Prism (v 5.0, GraphPad Software, Inc., La Jolla, CA) and XLStat-Pro (Addinsoft USA, New York, NY).

## RESULTS

### Rice bran metabolome differences among rice varieties before and after fermentation with *S. boulardii*

Bran from three rice varieties (Table 4.1) was extracted for metabolite profiling and analyzed by gas chromatography coupled to mass spectrometry (GC-MS). The rice varieties Neptune and Wells are U.S. semi-dwarf varieties, and Red Wells is isogenic to Wells apart from a single deleted base pair in the proanthocyanidin gene Rc. This mutation results in the production of red pigment in the bran layer of the seed (*30*). Principal component analysis was used to elucidate varietal differences in the metabolome of non-fermented rice bran (Figure 4.1A). Varietal differences were largely explained by the first component (56%), and the second component differentiated between biological replicates (14%). Rice bran was then incubated in the presence and absence of *S. boulardii* and metabolites were extracted and detected by GC-MS. The GC-MS chromatograms of non-fermented rice bran, *S. boulardii* fermented rice bran, and *S. boulardii* extracts showed unique differences among the treatments (Figure 4.1B). Partial least squares-discriminate analysis (PLS-DA) was used to detect differences in the metabolome among all three varieties and with or without fermentation

with *S. boulardii* (Figure 4.1C). The first two components of the PLS-DA model explained 66% and 19% of the variation, respectively. This model demonstrates the ability to apply metabolite-profiling techniques to differentiate chemical contents of fermented rice bran from non-fermented rice bran, irrespective of the rice variety tested.

**Rice bran metabolites modulated by *S. boulardii* fermentation**

*S. boulardii* fermentation induced changes in metabolite content for all three varieties, determined by quantitative analysis of peak areas of 10,260 GC-MS derived features. For Neptune, Wells, and Red Wells, 448, 127, and 311 features varied due to fermentation, respectively (Student's t-test, $P < 0.05$). A linear mixed model analysis was performed for each feature using fermentation (fixed), and variety and a fermentation-variety interaction (random) as effects. The mean percent variance explained by the model for all features was 14.8%. Genotype and genotype*fermentation effects explained a mean of 10% and 4.8% of the total variation, respectively.

A series of PLS-DA and OPLS models were applied to each rice variety to determine metabolites with the most significant changes due to fermentation (Figure 4.2). The effect of fermentation on each variety's metabolome was explained by the first component in each PLS-DA model. OPLS analyses were conducted to determine metabolites quantitatively altered by yeast probiotic fermentation. Metabolites of interest contained p(1) and p(corr) values greater than 0.02 and 0.8, respectively, and were analzyed for quantitative differences between samples by fold-change due to fermentation and by Student's t-test ($P < 0.05$) (Table 4.2). Mass spectra of significant peaks were screened in the National Institute of Technology Standards metabolite database for probable matches. Rice varieties differed in candidate metabolites altered by *S. boulardii* fermentation and by chemical classes. For the three varieties, there was wide variation in both the relative quantities of metabolites increased and the types of

predicted metabolites. *S. boulardii* fermentation of rice bran differs with regard to variety and was next evaluated for impact on anti-cancer properties of rice bran.

### *S. boulardii* fermented rice bran extracts differentially inhibit lymphoma viability

Polar rice bran extracts were previously shown to inhibit tumor promotion of lymphoblastoid B cells, and rice bran agglutinin inhibited growth of monoblastic leukemia U937 cells (*26*). Methanol-soluble metabolites from Neptune rice bran were screened for dose dependent effects on viability of normal human peripheral blood lymphocytes (PBL) and malignant human B-cell lymphoma (Figure 4.3A). Viability was measured by resazurin stain after 24 hours of incubation with fermented and nonfermented rice bran extracts. The rice bran extracts did not affect the viability of normal PBL (Figure 4.3A). A significant reduction in lymphoma viability was demonstrated at the 500 μg $ml^{-1}$ dose of Neptune rice bran extract, while the 125 and 250 μg $mL^{-1}$ were not significantly reduced from vehicle control (Figure 4.3A). The 500 μg $mL^{-1}$ dose of rice bran extract was next used to examine effects of both non-fermented and fermented rice bran extracts across varieties on normal PBL and lymphoma. None of the rice bran extracts altered the viability of normal PBL (Figure 4.3B). The *S. boulardii*-fermented rice bran significantly inhibited lymphoma viability compared to vehicle controls for all varieties tested (Figure 4.3C). The non-fermented Neptune rice bran extracts showed a 23% reduction in viability, and *S. boulardii*-fermented Neptune rice bran extracts reduced viability by 85% compared to control. At 500 μg $mL^{-1}$, unfermented extracts of Wells and Red Wells had no effect on lymphoma viability relative to the control, however fermented extracts inhibited viability by 75% and 51%, respectively (Figure 4.3C). The percent reduction in viability differed among varieties of the three fermented extracts (ANOVA, Tukey post-hoc, $P < 0.05$). The differential reduction in viability by fermented rice bran among rice varieties supports that variation in metabolite contents as detected in Figure 4.1 may be

important for bioactivity. The isopropanol:acetonitrile:water (3:2:2) solvent used for metabolite profiling of fermented bran extracts (Figure 4.2) was also examined for effects on lymphoma viability, however this solvent demonstrated suboptimal background activity as a vehicle control and was therefore not utilized to compare effects across rice varieties (data not shown).

### *S. boulardii* modulation of bioactive rice bran compounds

Given the varietal differences in anticancer activity of *S. boulardii* fermented rice bran extracts, a number of bioactive rice bran compounds were selected for relative quantification. Rice bran contains a number of metabolites with reported anticancer effects, notably phenolics and phytosterols (*37-39*). Salicylic, p-coumaric, ferulic, and caffeic acid, and also α-tocopherol and β-sitosterol were detected in non-fermented and fermented rice bran from each of the three varieties by comparing the initial GC-MS chromatograms to purchased standards (Figure 4.4). Non-fermented extracts from Wells contained a greater quantity of salicylic acid than both Red Wells and Neptune (Figure 4.4A). Red Wells contained higher amounts of ferulic acid than Neptune, and Neptune contained significantly less β-sitosterol than both Wells and Red Wells (ANOVA, Tukey post-hoc, $P < 0.05$). A z-score analysis was conducted to determine significant changes in metabolite quantity due to fermentation with *S. boulardii*, using non-fermented rice bran as a control. The data shown in Figure 4.4C supports that *S. boulardii* fermentation reduced the quantity of p-coumaric acid in Red Wells, and increased ferulic acid in Neptune (Figure 4.4C).

### Rice bran as sole carbon source for *S. boulardii*

The ability of *S. boulardii* to utilize and quantitatively alter chemical components of rice bran was confirmed by measuring its growth on rice bran as a sole carbon

source. Overnight cultures of *S. boulardii* inoculated into medium containing 5% rice bran from each of the varieties grew significantly better than cultures inoculated into YNB broth with dextrose as the primary carbon source (Figure 4.5). In addition, *S. boulardii* cultures maintained viability and cell numbers for 3 days on rice bran media, while the number of cells in the YNB cultures steadily declined. No significant differences in the growth of *S. boulardii* were detected among the three rice varieties (ANOVA, Tukey post-hoc, $P < 0.05$).

## DISCUSSION

This study demonstrates the utility of integrating global and targeted metabolite profiling for analysis of rice bran phytochemicals in the presence and absence of *S. boulardii* fermentation, and has advanced our knowledge about how probiotic fermentation of rice bran can enhance anticancer properties. Metabolomics is one strategy used to measure the wide array of phytochemicals that are typically evaluated in the "free" forms from food extracts; as these small molecules dissolve quickly and are immediately absorbed into the bloodstream. This high throughput, yet sensitive approach is also useful to assess the "bound" forms of rice molecules, which are attached to the plant cell walls and must be released by microbes during digestion before they can be absorbed. These findings set the stage for developing metabolomics as a tool for investigating rice bran phytochemical diversity and digestion by probiotics.

Metabolite profiles in this study showed variation among the three U.S. rice varieties Neptune, Wells, and Red Wells (Figure 4.1A, Table 4.1). The Neptune metabolite profile separated from both Wells and Red Wells. This cluster was expected given the near-isogenic state of the “Wells” varieties and provided strong rationale for investigating differential bioactive properties. Rice bran fermentation with the *S. boulardii* probiotic enhanced metabolite diversity (Figure 4.1C), and showed rice varietal

differences in bran extract-mediated reduction of lymphoma growth (Figure 3B). Candidate metabolites that were significantly increased post-fermentation also differed among the three varieties (Table 4.2). The ability of *S. boulardii* to utilize rice bran as a sole carbon source substrate for cellular metabolism and growth (Figure 4.5) suggests that rice bran contains prebiotic characteristics. Although no differences were detected among the Neptune, Wells and Red Wells rice varieties, these findings warrant further investigation of rice bran pre-biotic components and the synergistic effects of prebiotic/probiotic combinations on human health. To our knowledge, only two studies have examined distinct rice bran varieties for differential anti-carcinogenic activity [27;28]. Data from these studies demonstrate that rice varieties with pigmented seed coat also exhibit differential activity when compared to non-pigmented. The findings presented in this report suggest that some of the inconsistent results of past rice bran investigations on cancer cell growth may be due to the rice variety tested and not just those chemicals responsible for pigment.

Another plausible explanation for inconclusive data on rice bran is not only differences in metabolite content among varieties, but also the influence of probiotics altering the bioavailability of cancer-protective compounds in select tissues. The ability of phenolics, particularly ferulic, salicylic, caffeic, and p-coumaric acids and α-tocopherol (a lipid-soluble antioxidant) found in rice bran, to scavenge free radicals, alter enzymes, affect biochemical pathways, and interfere with gene expression has attracted the attention of researchers in search of cancer-fighting agents [21,29,30]. The efficacy of ferulic acid, which remains in the bloodstream longer than other known antioxidants and therefore may provide more protection, is dependent on its bioavailability and dosage [31]. However, plant phenols are often found in a biologically unavailable form due to an ester-bond to cell wall polysaccharides. Therefore, the optimal dose of rice bran required to achieve cancer-fighting levels of ferulic acid is unknown. Humans and rats have been

shown to release diferulic acid from bran fiber using gastrointestinal esterases found in the large and small intestines, thus enhancing the bioavailability of this compound [32]. The data shown in Figure 4.4C supports that the Neptune rice variety may exhibit higher probiotic-induced ferulic acid release and bioavailability than the other two varieties, and that consuming the whole food post-fermentation with *S. boulardii* may be a viable alternative for achieving enhanced levels of this compound without losing the benefits of the others. Yeast cells typically only maintain viability for several hours after they have reached stationary phase and depleted their carbon source. Our results show that rice bran medium allows *S. boulardii* cells to maintain viability over several days, suggesting that secondary fermentation by the yeast may be occurring and may further alter the phytochemical content of the rice bran (Figure 4.5). Thus, it will be necessary to optimize fermentation times to advance our understanding of the kinetics of rice bran phytochemical metabolism and release by *S. boulardii*.

Emerging evidence supports additive and/or synergistic effects of rice bran components for protection against certain cancers (5, 11, 40, 41), however few studies have examined differences in phytochemical contents in commercially available rice varieties. Our data support that many rice bran components were fermented by the yeast probiotic (Figure 4.2) and these components work together to enhance probiotic growth (Figure 4.5). One study examined a yeast fermentation of rice bran for changes in the stability, palatability, and nutritional status (carbohydrate, methionine, calcium, and ash content) of the bran, but did not address the alteration of potentially bioactive phytochemicals (42). Given the evidence for cancer fighting activities of rice bran phytochemicals, the data presented herein support that *S. boulardii*-fermented rice bran should be next tested for bioavailability of bioactive components and for reducing lymphoma viability *in vivo*. Chemopreventive single agent compounds found in rice bran include, but are not limited to tocopherols, polyphenols, inositol hexaphosphate (IP6),

non-starchy polysaccharides, γ-oryzanol and phytosterols (2, 4, 43-45). Our metabolite profile analysis of fermented rice bran revealed extensive rice bran chemical diversity, and can be used to further the identities of novel combinations of bioactive compounds that display phytochemical teamwork (46, 47). Whole rice bran consumption is undoubtedly recognized as important for providing more comprehensive protection against cancer cells when compared to supplementation with isolated ingredients, and the metabolite profiling techniques and chemical analyses presented herein support further interrogation of rice bran effects on intestinal microbe interactions as well as probiotic growth and metabolism. The metabolomics strategy applied herein has advanced our understanding of the health importance of rice bran phytochemical diversity in the presence and absence of fermentation and for disease fighting activity. Single agent nutritional “magic bullets” too often fail to achieve the health benefits indicated by cell-based assays, and available methodologies have limited investigations to these reductive approaches. By utilizing global metabolomic profiling, we can now more holistically approach these complex mixtures of small molecules and improve studies linking bioactive food components and human health.

**FIGURES**

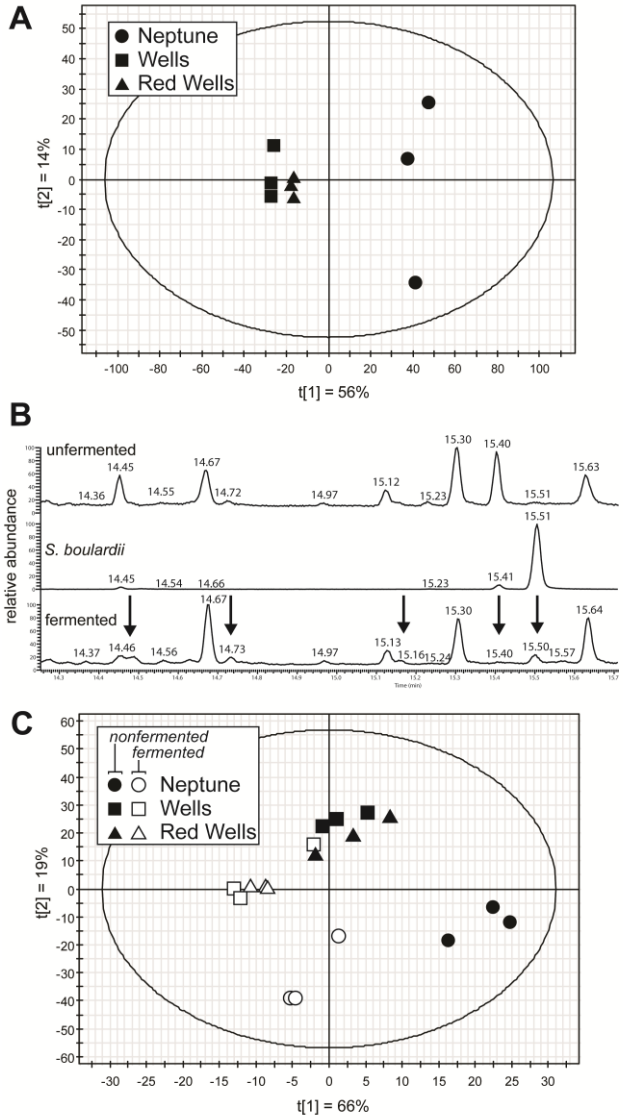


**Figure 4.1. Metabolite Profiling of Rice Bran from three varieties with and without fermentation by *S. boulardii*** (A) Principal Components Analysis (PCA) of bran extracts from three rice varieties (Neptune, Wells and Red Wells) show diversity in metabolite profiles. The first principal component separated the three varieties, and the second component was mostly composed of variation among replicates within a single variety. (B) A representative portion of a GC-MS chromatograph showed change in metabolites in fermented rice bran. Neptune rice variety alone (top), *S. boulardii* extract alone (middle) and bran from Neptune variety fermented with *S. boulardii* (bottom). Some peaks are present in only one sample, and others are present in both but vary in quantity, as indicated by arrows above differential peaks**.** (C) PLS-DA model of three varieties of non-fermented (black shading) or fermented with *S. boulardii* (white). The first component separated each variety from its fermented counterpart, and the second component separated Neptune from Wells and Red Wells.

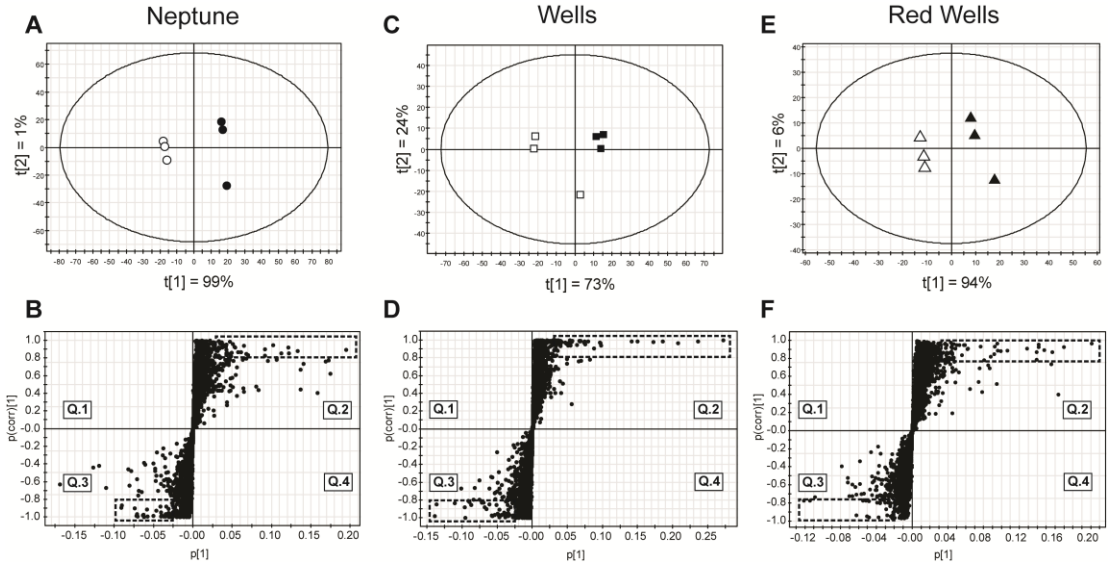


**Figure 4.2. PLS-DA and OPLS models to determine metabolite variation induced by fermentation with *S. boulardii.*** Each variety was independently analyzed for metabolite differences induced by *S. boulardii* fermentation. (A) Neptune PLS-DA showed the metabolome differs between unfermented (black) and fermented (white) samples. (B) The Neptune OPLS analysis showed metabolites that highly differ based on fermentation, indicated by the dashed box for Quadrant 2 (non-fermented) and Quandrant 3 (fermented). P(corr) values correspond to deviation across replicates, and p(1) values are proportional to the quantity of metabolite. Wells (C and D) and Red Wells (E and F) also showed altered metabolite content induced by *S. boulardii*.

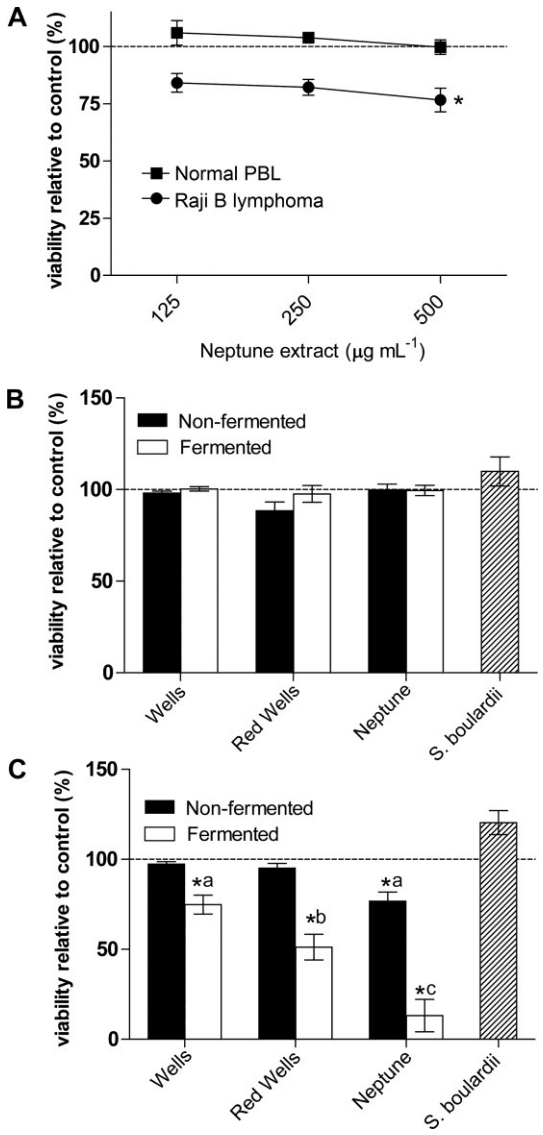


**Figure 4.3. *S. boulardii* fermented rice bran inhibits lymphoma viability.** (A) Different doses of non-fermented methanolic rice bran extracts (Neptune) were added to normal human peripheral blood lymphocytes (PBL) and Raji B lymphoma cultures for 24 hours. Values are expressed as the mean percent viable cells relative to the vehicle control ± S.E.M. Extracts of non-fermented Neptune reduced lymphoma viability at 500 µg $mL^{-1}$ (student's t-test, $P < 0.05$). (B) Fermented and non-fermented extracts of all three varieties at 500 µg $mL^{-1}$ had no effect on viability of normal PBL. (C) Fermented and non-fermented extracts of all three varieties at 500 µg $mL^{-1}$ differentially affected lymphoma viability, as measured by cell fluorescence after the addition of resaruzin (ANOVA, Tukey post-hoc, $P < 0.05$). Significance from vehicle control is represented by an asterisk, and statistical groupings are denoted by the letter a, b, and c.

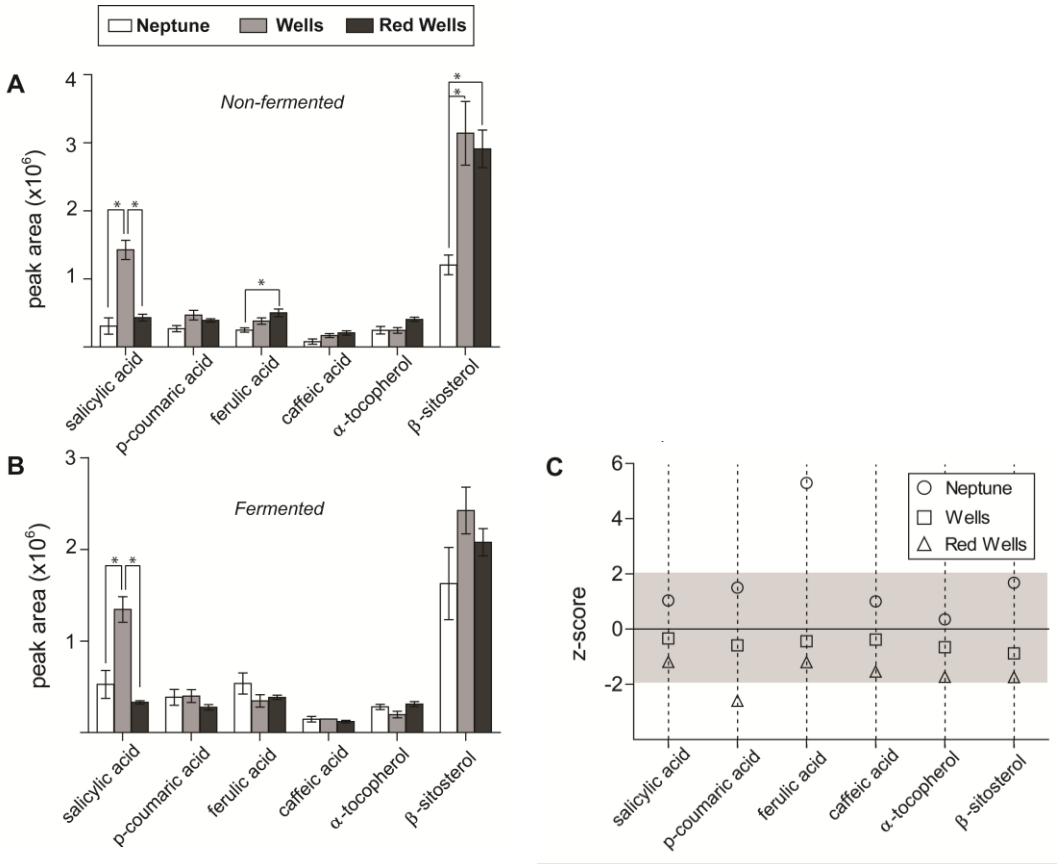


**Figure 4.4. Bioactive food components in rice bran with and without *S. boulardii* fermentation.** Relative quantification of metabolites without fermentation (A) and with fermentation (B) was based on the area of the GC-MS chromatograph. Metabolites showing a significant difference by relative quantity and between two varieties were indicated by an asterisk (ANOVA, Tukey post-hoc, $P < 0.05$). Values are expressed as mean peak area ± S.E.M. (C) Z-score for metabolites from fermented extracts using the non-fermented as a control. Significant changes in metabolite quantity are indicated by z-score values outside of the shaded region. Increased ferulic acid was detected in *S. boulardii* fermented Neptune rice bran, and decreased p-coumaric acid was detected from fermented Red Wells compared to non-fermented.

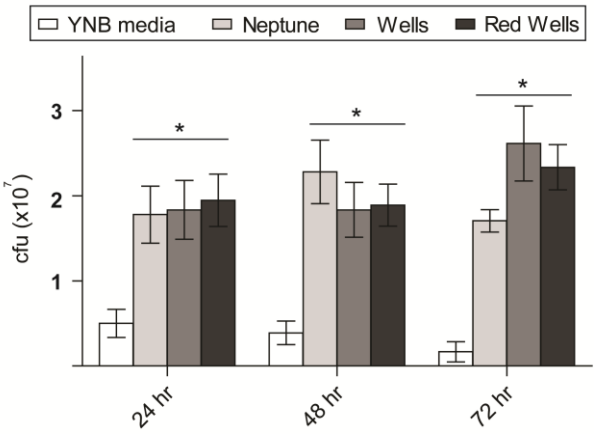


**Figure 4.5. Rice bran enhances growth of *S. boulardii*.** Rice bran from all three rice varieties significantly increased the growth of *S. boulardii* compared to YNB media alone after 24, 48 and 72 hours. An asterisk indicates difference in the quantity of yeast colonies compared to the YNB control (one way ANOVA, Tukey post-hoc, $P < 0.05$).

## TABLES

### Table 4.1. Rice Varieties Assessed for Bioactivity of the Bran

| Trait | Neptune | Wells | Red Wells |
|---|---|---|---|
| Plant ID | PI 655959 | PI 612439 | |
| yield | high | high | high |
| grain type | long | medium | medium |
| leaf type | erect | erect | erect |
| pericarp | brown | brown | red |
| sheath blight | moderately susceptible | moderately susceptible | moderately susceptible |
| rice blast | moderately resistant | moderately susceptible | moderately susceptible |

**Table 4.2. Varietal Differences in Candidate Compounds Altered by S. boulardii Fermentation**

| variety | compound | class | fold change after fermentation | p-value |
|---|---|---|---|---|
| Neptune | galactose | sugar | -10.32 | 0.03 |
| | palmitic acid | fatty acid | -1.2 | 0.04 |
| | alpha-linoleic acid | fatty acid | -1.23 | 0.04 |
| | uknown disaccharade | sugar | 26.44 | 0.001 |
| | xylitol | sugar-alcohol | 14.79 | < 0.001 |
| | glucitol | sugar-alcohol | ** | < 0.001 |
| | alanine | amino acid | 4.63 | 0.02 |
| | phosphoric acid | mineral | 2 | < 0.001 |
| | 1,2,3-propanetricarboxylic acid | organic acid | 4.34 | 0.02 |
| Wells | D-fructose | sugar | -6.38 | 0.005 |
| | ribitol | sugar-alcohol | * | < 0.001 |
| | linoleic acid methyl ester | fatty acid | 1.4 | 0.04 |
| Red Wells | palmitic acid | fatty acid | ** | 0.03 |
| | uknown disaccharade | sugar | ** | < 0.001 |

*metabolite only present in nonfermented extracts
**metabolite only present in fermented extracts

## REFERENCES


1. Kahlon, T. S., Rice Bran: Production, Composition, Functionality and Food Applications, Physiological Benefits. In *Fiber Ingredients: Food Applications and Health Benefits*, Samuel, S. S. C. a. P., Ed. Taylor and Francis Group, LLC: Boca Raton, 2009; pp 305-321.
2. Cai, H.; Al-Fayez, M.; Tunstall, R. G.; Platton, S.; Greaves, P.; Steward, W. P.; Gescher, A. J., The rice bran constituent tricin potently inhibits cyclooxygenase enzymes and interferes with intestinal carcinogenesis in ApcMin mice. *Mol Cancer Ther* 2005, 4, (9), 1287-92.
3. Kannan, A.; Hettiarachchy, N.; Johnson, M. G.; Nannapaneni, R., Human colon and liver cancer cell proliferation inhibition by peptide hydrolysates derived from heat-stabilized defatted rice bran. *J Agric Food Chem* 2008, 56, (24), 11643-7.
4. Kong, C. K.; Lam, W. S.; Chiu, L. C.; Ooi, V. E.; Sun, S. S.; Wong, Y. S., A rice bran polyphenol, cycloartenyl ferulate, elicits apoptosis in human colorectal adenocarcinoma SW480 and sensitizes metastatic SW620 cells to TRAIL-induced apoptosis. *Biochem Pharmacol* 2009, 77, (9), 1487-96.
5. Phutthaphadoong, S.; Yamada, Y.; Hirata, A.; Tomita, H.; Taguchi, A.; Hara, A.; Limtrakul, P. N.; Iwasaki, T.; Kobayashi, H.; Mori, H., Chemopreventive effects of fermented brown rice and rice bran against 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced lung tumorigenesis in female A/J mice. *Oncol Rep* 2009, 21, (2), 321-7.
6. Kestin, M.; Moss, R.; Clifton, P. M.; Nestel, P. J., Comparative effects of three cereal brans on plasma lipids, blood pressure, and glucose metabolism in mildly hypercholesterolemic men. *Am J Clin Nutr* 1990, 52, (4), 661-6.
7. Kuriyan, R.; Gopinath, N.; Vaz, M.; Kurpad, A. V., Use of rice bran oil in patients with hyperlipidaemia. *Natl Med J India* 2005, 18, (6), 292-6.
8. Wilson, T. A.; Nicolosi, R. J.; Woolfrey, B.; Kritchevsky, D., Rice bran oil and oryzanol reduce plasma lipid and lipoprotein cholesterol concentrations and aortic cholesterol ester accumulation to a greater extent than ferulic acid in hypercholesterolemic hamsters. *J Nutr Biochem* 2007, 18, (2), 105-12.
9. Sukhonthara, S.; Theerakulkait, C.; Miyazawa, M., Characterization of Volatile Aroma Compounds from Red and Black Rice Bran. *Journal of Oleo Science* 2009, 58, (3), 155-160.
10. Heuberger AL, L. M., Chen MH, Brick MA, Leach JE, and EP Ryan Metabolomic and Functional Genomic Analyses Reveal Varietal Differences in Bioactive Compounds of Cooked Rice. *PLOS One* 2010, 5, (9), e12195.
11. Katyama, M.; Yoshimi, N.; Yamada, Y.; Sakata, K.; Kuno, T.; Yoshida, K.; Qiao, Z.; Vihn, P. Q.; Iwasaki, T.; Kobayashi, H.; Mori, H., Preventive effect of fermented brown rice and rice bran against colon carcinogenesis in male F344 rats. *Oncol Rep* 2002, 9, (4), 817-22.
12. Koh, J. H.; Suh, H. J., Biological Activities of Thermo-tolerant Microbes from Fermented Rice Bran as an Alternative Microbial Feed Additive. *Applied Biochemistry and Biotechnology* 2009, 157, (3), 420-430.
13. Koh, J. H.; Yu, K. W.; Suh, H. J., Biological activities of Saccharomyces cerevisiae and fermented rice bran as feed additives. *Letters in Applied Microbiology* 2002, 35, (1), 47-51.
14. Zanello, G.; Meurens, F.; Berri, M.; Salmon, H., Saccharomyces boulardii effects on gastrointestinal diseases. *Curr Issues Mol Biol* 2009, 11, (1), 47-58.
15. Czerucka, D.; Piche, T.; Rampal, P., Review article: yeast as probiotics -- Saccharomyces boulardii. *Aliment Pharmacol Ther* 2007, 26, (6), 767-78.

16. Yan, F.; Polk, D. B., Probiotics as functional food in the treatment of diarrhea. *Curr Opin Clin Nutr Metab Care* 2006, 9, (6), 717-21.
17. Fleet, G. H., Yeasts in foods and beverages: impact on product quality and safety. *Curr Opin Biotech* 2007, 18, (2), 170-175.
18. Jespersen, L., Occurrence and taxonomic characteristics of strains of Saccharomyces cerevisiae predominant in African indigenous fermented foods and beverages. *Fems Yeast Res* 2003, 3, (2), 191-200.
19. Martins, F. S.; Nardi, R. M.; Arantes, R. M.; Rosa, C. A.; Neves, M. J.; Nicoli, J. R., Screening of yeasts as probiotic based on capacities to colonize the gastrointestinal tract and to protect against enteropathogen challenge in mice. *J Gen Appl Microbiol* 2005, 51, (2), 83-92.
20. Rekha, C. R.; Vijayalakshmi, G., Bioconversion of isoflavone glycosides to aglycones, mineral bioavailability and vitamin B complex in fermented soymilk by probiotic bacteria and yeast. *J Appl Microbiol*.
21. Sindhu, S. C.; Khetarpaul, N., Effect of probiotic fermentation on antinutrients and in vitro protein and starch digestibilities of indigenously developed RWGT food mixture. *Nutr Health* 2002, 16, (3), 173-81.
22. Skovgaard, N., New trends in emerging pathogens. *Int J Food Microbiol* 2007, 120, (3), 217-24.
23. de Vrese, M.; Marteau, P. R., Probiotics and prebiotics: effects on diarrhea. *J Nutr* 2007, 137, (3 Suppl 2), 803S-11S.
24. Dalmasso, G.; Cottrez, F.; Imbert, V.; Lagadec, P.; Peyron, J. F.; Rampal, P.; Czerucka, D.; Groux, H.; Foussat, A.; Brun, V., Saccharomyces boulardii inhibits inflammatory bowel disease by trapping T cells in mesenteric lymph nodes. *Gastroenterology* 2006, 131, (6), 1812-25.
25. Chen, X.; Kokkotou, E. G.; Mustafa, N.; Bhaskar, K. R.; Sougioultzis, S.; O'Brien, M.; Pothoulakis, C.; Kelly, C. P., Saccharomyces boulardii inhibits ERK1/2 mitogen-activated protein kinase activation both in vitro and in vivo and protects against Clostridium difficile toxin A-induced enteritis. *J Biol Chem* 2006, 281, (34), 24449-54.
26. Nam, S. H.; Choi, S. P.; Kang, M. Y.; Koh, H. J.; Kozukue, N.; Friedman, M., Bran extracts from pigmented rice seeds inhibit tumor promotion in lymphoblastoid B cells by. phorbol ester. *Food and Chemical Toxicology* 2005, 43, (5), 741-745.
27. NoriyuM Miyoshi, Y. K., * Yuko Katsuno,* Sumio Hayakawa,* Takashi Mita,*; Toshiro Ohta/ Kaziihiko Kaji, f. a. M. I., Apoptosis Induction Associated with Cell Cycle Dysregulation by Rice Bran Agglutinin. *J. Biochem* 2001, 130, 799-805.
28. Shu, X. L.; Frank, T.; Shu, Q. Y.; Engel, K. H., Metabolite profiling of germinating rice seeds. *J Agric Food Chem* 2008, 56, (24), 11612-20.
29. Goffman, F. D.; Pinson, S.; Bergman, C., Genetic diversity for lipid content and fatty acid profile in rice bran. *Journal of the American Oil Chemists Society* 2003, 80, (5), 485-490.
30. Jackson, S. A. B. W. Y. A. K.; Deren, C. W., A natural mutation in rc reverts white-rice-pericarp to red
and results in a new, dominant, wild-type allele: Rc-g. *Theor Appl Genet* 2008, 117, 575-580.
31. Sindhu, S. C.; Khetarpaul, N., Effect of probiotic fermentation on antinutrients and in vitro protein and starch digestibilities of indigenously developed RWGT food mixture. *Nutrition and Health* 2002, 16, (3), 173-181.
32. Sindhu, S. C. a. N. K., Fermentation with one step single and sequential cultures of yeast and lactobacilli: Effect on antinutrients and digestibilities (in vitro) of starch and protein in an indigenously developed food mixture. *Plant Foods for Human Nutrition* 2003, 58, 1-10.

33. Palazoglu, M.; Fiehn, O., Metabolite identification in blood plasma using GC/MS and the Agilent Fiehn GC/MS metabolomics RTL library. *Agilent Application Note* 2009, 5990-3638EN.
34. Skogerson, K.; Harrigan, G. G.; Reynolds, T. L.; Halls, S. C.; Ruebelt, M.; Iandolino, A.; Pandravada, A.; Glenn, K. C.; Fiehn, O., Impact of Genetics and Environment on the Metabolite Composition of Maize Grain. *J Agric Food Chem* 2010, 58, (6), 3600-3610.
35. Xiao, J.; Zhang, Y.; Wang, J.; Yu, W.; Wang, W.; Ma, X., Monitoring of Cell Viability and Proliferation in Hydrogel-Encapsulated System by Resazurin Assay. *Appl Biochem Biotechnol*.
36. Duarte, M.; Giordani, R. B.; De Carli, G. A.; Zuanazzi, J. A.; Macedo, A. J.; Tasca, T., A quantitative resazurin assay to determinate the viability of Trichomonas vaginalis and the cytotoxicity of organic solvents and surfactant agents. *Exp Parasitol* 2009, 123, (2), 195-8.
37. Taniguchi, H.; Hosoda, A.; Tsuno, T.; Maruta, Y.; Nomura, E., Preparation of ferulic acid and its application for the synthesis of cancer chemopreventive agents. *Anticancer Res* 1999, 19, (5A), 3757-61.
38. Bouic, P. J. D.; Etsebeth, S.; Liebenberg, R. W.; Albrecht, C. F.; Pegel, K.; Van Jaarsveld, P. P., Beta-sitosterol and beta-sitosterol glucoside stimulate human peripheral blood lymphocyte proliferation: Implications for their use as an immunomodulatory vitamin combination. *International Journal of Immunopharmacology* 1996, 18, (12), 693-700.
39. Vieno, P.; David, G. L.; Tatu, A. M.; Jari, T.; Anna-Maija, L., Plant sterols: biosynthesis, biological function and their importance to human nutrition. In 2000; Vol. 80, pp 939-966.
40. Mori, H.; Kawabata, K.; Yoshimi, N.; Tanaka, T.; Murakami, T.; Okada, T.; Murai, H., Chemopreventive effects of ferulic acid on oral and rice germ on large bowel carcinogenesis. *Anticancer Res* 1999, 19, (5A), 3775-8.
41. Hudson, E. A.; Dinh, P. A.; Kokubun, T.; Simmonds, M. S. J.; Gescher, A., Characterization of Potentially Chemopreventive Phenols in Extracts of Brown Rice That Inhibit the Growth of Human Breast and Colon Cancer Cells. In 2000; Vol. 9, pp 1163-1170.
42. Vivian Feddern, E. B. F., and Leonor Almeida de Souza SOARES, Effects of fermentation on the physicochemical and nutritional properties of rice bran. *Ciênc. Tecnol. Aliment., Campinas* 2007, 27, (4), 800-804.
43. Takeshita, M.; Nakamura, S.; Makita, F.; Ohwada, S.; Miyamoto, Y.; Morishita, Y., Antitumor effect of RBS (rice bran saccharide) on ENNG-induced carcinogenesis. *Biotherapy* 1992, 4, (2), 139-45.
44. Aoe, S.; Oda, T.; Tojima, T.; Tanaka, M.; Tatsumi, K.; Mizutani, T., Effects of rice bran hemicellulose on 1,2-dimethylhydrazine-induced intestinal carcinogenesis in Fischer 344 rats. *Nutr Cancer* 1993, 20, (1), 41-9.
45. Norazalina, S.; Norhaizan, M. E.; Hairuszah, I.; Norashareena, M. S., Anticarcinogenic efficacy of phytic acid extracted from rice bran on azoxymethane-induced colon carcinogenesis in rats. *Exp Toxicol Pathol* 2009.
46. Adom, K. K.; Sorrells, M. E.; Liu, R. H., Phytochemical Profiles and Antioxidant Activity of Wheat Varieties. *Journal of Agricultural and Food Chemistry* 2003, 51, (26), 7825-7834.
47. Lila, M. A., From beans to berries and beyond: teamwork between plant chemicals for protection of optimal human health. *Ann N Y Acad Sci* 2007, 1114, 372-80.

# CHAPTER FIVE: CONCLUSIONS AND FUTURE DIRECTIONS

## REVIEW OF WORK

### Influence of Genetic Diversity on Metabolite Variation in Rice

Incorporating metabolomics into the study of staple crops is an innovative method to assess metabolite-related phenotypes [1,2]. The influence of genetic diversity on metabolite profiles of staple foods has not been well described. My experiments confirm that genetic diversity influences metabolite variation in both cooked rice and stabilized bran. Previous studies show metabolite diversity in cereals; however none describe the potential mechanisms by which genetic diversity influences metabolic variation. My work suggests that the content of plant-derived metabolites in the diet may be influenced by single nucleotide polymorphisms (SNPs), the environment, and gut fermentation. Both liquid and gas chromatography/mass spectrometry (LC-MS, GC-MS) techniques may be useful to assess the effects of SNP, environment, and fermentation factors on the metabolite profiles of staple crops.

Study 1 determined that genetic diversity is associated with variation in cooked grain metabolite profiles and health-related metabolites. Ten rice varieties were profiled using LC-MS. Approximately 25% of the molecular features varied ($p < 0.001$), including metabolite variation among subspecies (*indica, japonica, aus*). The multivariate partial least squares (PLS) analysis confirmed that metabolite variation was associated with the *indica, japonica*, and *aus* subspecies. Principal component analysis (PCA) models were heavily skewed by the variety Dular, and varieties did not cluster via subspecies when

Dular was included in the PCA (data not shown). Both PCA and PLS analysis are valuable methods to characterize metabolite profiles, and both techniques can be integrated with genetic selection algorithms to aid in selecting varieties with specific metabolite profiles [3]. PLS was more efficient for distinguishing genetic variation by subspecies. The model showed that a variety could be classified as *indica* or *japonica* based on its LC-MS-detected metabolite profile. No within-subspecies metabolite variation was characterized due to the limited number of varieties included in the study. The study confirmed the sensitivity of the LC-MS technique, and within-subspecies variation can be determined by including more varieties in the experimental design.

The genetic analysis conducted in Study 1 suggests SNPs are associated with metabolite diversity in cooked rice. The SNP homology in the phenolics and vitamin E biochemical pathways was similar to genome-wide SNP variation described in [4]. Both the phenolics and vitamin E dendrograms were limited by the number of SNPs present in both biochemical pathways. Expanding the biochemical and genetic models for both classes of metabolites would potentially increase the number of SNPs available for dissimilarity matrix analyses, and therefore increase statistical power. However, even with a low-power dataset, the trends in SNP variation within each pathway suggest that there is genetic variation in enzymes related to metabolite biosynthesis. Furthermore, SNP patterns were consistent with evolutionary relationships among the varieties. This supports that SNPs profiling may be a potential target to predict metabolite profiles, and SNP variation may explain altered regulation and efficiency of metabolite biosynthesis enzymes.

Study 2 showed that the environment in which rice is grown can influence the metabolite profile. In this study, two rice varieties were grown in two environments: the field and the greenhouse. Metabolite profiling was conducted using GC-MS to determine variation in primary metabolites. Genotypic effects accounted for 25% of total variation

among detected molecular features, and environmental and genotype-environment interactions (GEI) accounted for 25% and 12.7% of the variation, respectively ($p < 0.01$). Bioinformatics analyses were conducted to identify allelic variation in phenylalanine synthesis genes, and also gene expression variation and gene-gene interactions were established.

The experimental design of the study was limited to two genotypes and environments. The limited genotypes and environments allowed for more detailed characterization of relationships among metabolites, as excessive GEI would make metabolite covariation difficult to detect. The GEI effects observed for 12% of molecular features suggest that their expression among genotypes would differ among genotypes across environments. For analysis of GEI on secondary metabolites, a phenylalanine network was constructed to determine metabolites that covary with phenylalanine, and genes that may be associated with the metabolite profiles. The GEI effects indicated that metabolite profiles vary across varieties and environments.

The third study confirmed genotype-fermentation interactions as a source of variation in the health properties of the grain. In this study, extracts form three rice varieties were fermented with *Saccharomyces boulardii,* a probiotic yeast. Metabolite profiles were established for the fermented and nonfermented extracts. Both nonfermented and fermented extracts were added to a lymphoma cell culture, and cell viability was assessed. There was mild inhibition of viability for the nonfermented extracts, but fermented extracts from all three varieties significantly reduced cell viability. From the metabolite profiles, it was determined that ferulic acid was significantly increased for fermented-Neptune, but not for the varieties Wells and Red Wells.

While Studies 1 and 2 assessed genetically diverse varieties, Study 3 described differences between two near isogenic varieties: Wells (brown rice) and Red Wells (red rice). The two varieties are thought to differ in a single anthocyanin synthesis gene [5]

due to a spontaneous mutation in Wells that restores the Rc gene, which is a loss of function allele due to an insertion in the allele that conditions brown rice vs. red rice seed coat [6]. Nonfermented rice bran extracts from Wells and Red Wells had no effect on lymphoma cell viability, however viability was reduced when treated with fermented rice bran extracts from both varieties. Genotype-fermentation interactions (GFI) were observed in that extracts from fermented-Red Wells resulted in a greater reduction of viability (50% reduction) than fermented-Wells (25%). A third variety, Neptune (of similar germplasm to both Wells and Red Wells), showed 90% reduction in lymphoma cell viability.

GFI effects may have important consequences in breeding for enhanced health traits. Gut microbiota have been recently recognized as important for human health. Metabolites unique to both the nonfermented and fermented extracts were observed, and thus fermentation may alter bioavailability and effect of some plant metabolites. The GFI effects in Study 3 suggest that a gut microorganism-associated health phenotype may be dependent on the plant variety consumed. Metabolomics may be a valuable technique to characterize rice metabolites that are known to be altered by fermentation in the gut.

There was a clear relationship between genotype and metabolite quantities in all three studies. Interaction elements (*e.g.* environment, fermentation) can significantly alter the heritability of some metabolites. Interaction effects were not present in all metabolites, and therefore significant nutritional gains can still be made in food crops, whereby breeders can incorporate metabolomic selection methods to develop varieties with similar profiles.

**Review of Metabolite Detection Techniques**

There are many methods to assess metabolite diversity in plants. Metabolite extraction solvents are inherently biased in polarity and pH, and detection techniques are limited by size of the compounds, column chemistry, and the physics and chemistry of compound separation. It is impossible to establish a complete metabolite profile using one technique. Therefore, the chosen extraction and detection methods should attempt to quantify a wide range of metabolites.

For all three studies, a single-phase solvent was used to extract metabolites. The single-phase extraction minimized the potential for degradation after metabolites were extracted from the biological matrix. For cooked rice, an aqueous:methanol (80:20) solvent was used to assess both polar and nonpolar compounds. For rice bran, a isopropanol:acetonitrile:water (3:2:2) technique was used to screen for more nonpolar compounds. Both solvents extracted a diverse set of metabolites to efficiently screen for genotypic effects. Single-phase extractions are relatively rapid procedures, and therefore reduce the potential for sample degradation. Both ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) and gas chromatography-MS (GC-MS) were used to assess the effects of genetic diversity. UPLC-MS is efficient in detecting secondary metabolites, and GC-MS excels at primary metabolite detection, and limiting metabolite detection to only one method further biases the dataset. It is therefore recommended to extract metabolites in a single-phase and to detect with both LC-MS and GC-MS.

## FUTURE RESEARCH DIRECTIONS

### Metabolomic Selection

Developing grain metabolite profiles for enhanced human health attributes requires identification and selection on a metabolome-wide scale. The current metabolite detection tools are high-throughput, and can be combined with multivariate statistics to initiate “metabolomic selection,” which incorporates metabolite profile as part of a selection index. Genomic selection has been proposed as one measure to expedite crop improvement in breeding programs [7]. Metabolomic selection requires first establishing metabolite profiles that have an ideal phenotype, such as health properties of the grain. Selection decisions can be made based on similarities to the ideal profile. Like genomic selection, it would involve both a training set and a validation set, and data from multiple years and environments. Establishing a clear set of methods for metabolomic selection for health-related metabolites would be a valuable addition for plant breeding. To conduct metabolomic selection, an ideal metabolite profile needs to be establish, such as with yield, disease resistance, or other agronomic traits. Varieties in a breeding population would be selected by establishing individual metabolite profiles, and then determining the similarity to the ideal profile. Varieties would then be crossed to obtain offspring with metabolite profiles most similar to the ideal phenotype.

A number of statistical methods should be explored to characterize the metabolite profile during metabolomic selection. The ANOVA model is effective in identifying metabolites that have high variability; however this approach may be skewed by false positives. A Bonferroni correction can be employed, which changes the probability of a false positive based on the number of tests performed. However, this approach may be too conservative and risk false negatives, and it is not recommended for metabolomics analyses. The PLS loadings plot successfully classified metabolites as

contributing to *indica, japonica,* or *aus*, and then scored samples based on how similar the profile was to the model classification. However, the PLS limitation is that it can be skewed towards highly abundant compounds, which may be abundant due to the chosen method of extraction, or properties of the mass spectrometer. A combination of univariate and multivariate techniques are recommended for metabolomics analysis.

Study 2 suggests that metabolite covariance may be important to consider in metabolomic selection. The network was extremely useful to interrogate the data within the study, but this approach may not be feasible on larger datasets. The correlation/covariance among metabolites can be used as method to assess variation. This can be done by using an Individual Differences Scaling method, which assesses metabolite profiles based on metabolite covariance matrices [8]. Establishing covariance for metabolomics datasets is ideal for assessing treatment effects, such as environment or fermentation. Another metabolite variation analysis approach is the Random Forest (RF) method, which is a biased classification algorithm that has shown promise for metabolomics [9]. The RF method uses a decision tree for each variable, for which each tree "votes" for a classification. RF analysis corrects for limitations of interrogating metabolomics data with PCA and PLS, notably an indifference to scaling, accurate modeling with nonparametric data, exclusion of variables with little contribution to phenotypic variation, correcting for datasets with low treatment groups and many variables, and establishing relationships among variables. Thus, RF models should be investigated for utility in metabolomic selection.

**Functional Genomics Associated with Metabolomics**

Metabolomic and genomic selection can be performed simultaneously with an understanding as to which genes affect specific metabolite quantities. These studies would highlight the field of 'genetical metabolomics,' or the integration of gene

expression and metabolite profiles [10]. A better understanding of gene-metabolite relationships can aid in selecting varieties with ideal metabolite profiles.

Mapping the genetic loci that contribute to a metabolite phenotype is an important first step. Metabolites have been mapped to loci in *Arabidopsis*, rice, *Populus*, and humans [11,12,13,14]. One genetic mapping approach is to identify quantitative trait loci in a recombinant inbred line population (RIL). Extracts from each RIL can be characterized for metabolite quantities, and each metabolite can be individually mapped. Gene regions can then be fine mapped and sequenced to identify genetic variants that correlate with metabolite variation. A genome-wide association analysis can yield similar results, and allows for more than two alleles/haplotypes at a given locus. Although both approaches can identify genetic markers that represent metabolite variation, they may not be able to describe metabolite variation due to environment or fermentation effects.

It is also important to establish the functional characterization by which genetic variation alters metabolite profiles. In Study 2, a set of genes was identified that regulate phenylalanine, phenylpropanoid, and flavonoid synthesis. Future investigations should validate the genetic control behind metabolite covariation using molecular techniques. Both IR64 and Moroberekan should be fully sequenced for the set of genes of interest from the dataset. Additional non-SNP based genetic variation should be investigated, such as inversions, insertions, deletions, or transpositions, and epigenetic patterns, such as methylation or chromosomal orientation. For example, a 26-bp deletion in an oxalate oxidase gene was important in disease resistance phenotypes in Moroberekan [15]. A series of silencing, complementation mutants, and domain swaps can further characterize functional differences among rice varieties. Rice phenylalanine lyase mutants should also be screened for variation in secondary metabolite variation. Furthermore, gene expression may be difficult to characterize in the developing seed. If leaf gene expression variation correlates to grain metabolite profiles, then leaf metabolite

and gene expression profiles may be able to predict metabolite variation in the seed. This would allow for metabolomic selection to occur prior to seed development, and can expedite breeding for enhanced health characteristics.

**Fermentation and Health**

Rice varieties may differ in the ability to stimulate growth of beneficial microbes, and therefore rice can be improved for prebiotics metabolite profiles. *In vitro* fecal fermentation models can be used as a high-throughput method to assess the prebiotic capacity of a food. Microbial growth can be assessed using fluorescence *in situ* hybridization coupled to flow cytometry (FISH-FCM) or with next-generation sequence platforms. Additionally, the production of total short chain fatty acids (SCFAs), byproducts of microbial fermentation with beneficial health effects, should be determined as related to rice as a substrate for microbial growth.

The capacity for a metabolite to encourage the growth of beneficial microbes is measured by the use of a “prebiotic index” (PI) [16,17]. The PI is a quantitative assessment of the ability for a metabolite to result in multiplication of beneficial microbes. Variations on PI have been used to compare different types of metabolites for *in vitro* analysis of fecal microbial population growth [18,19]. The types of microbes present can be measured by use of selective agars. However microbial 16S rRNA sequences are available for use of fluorescence *in situ* hybridization (FISH), which can be coupled with quantitative reverse transcription-polymerase chain reaction (qRTPCR) [20] or with flow cytometry (FCM) [21], and such techniques can rapidly assess a microbial population for the presence of probiotics. Both FISH-qRTPCR and FISH-FCM are commonly used to characterize microbial communities in both *in vitro* and *in vivo* systems.

In addition to influencing the gut microbial population, byproducts of fermentation are also of health importance. The fermentability of a prebiotic is a measure of the ability for a gut microbial community to breakdown the component to extract energy for microbial metabolism. SCFAs are common byproducts of fermentations and important to human health that can be detected using GC-MS. Therefore, in addition to affecting microbial growth, SCFAs are excellent targets to measure the effects of a candidate prebiotic.

**Long-term Goal**

Widely-consumed staple crops are the most effective means to promote health and disease prevention. The long-term goal of this research is to improve the phytochemical and nutrient profile of the cooked grain and bran for commonly consumed rice varieties. Although some of the varieties examined in this report are not widely consumed, they can be screened for health-related alleles and incorporated into traditional rice breeding programs. Identifying and characterizing the genetic factors that regulate plant secondary metabolites can have profound impacts on global health.

**REFERENCES**


1. Dixon RA, Gang DR, Charlton AJ, Fiehn O, Kuiper HA, et al. (2006) Applications of Metabolomics in Agriculture. Journal of Agricultural and Food Chemistry 54: 8984-8994.
2. Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends in Genetics 25: 39-48.
3. Ramadan Z, Jacobs D, Grigorov M, Kochhar S (2006) Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. Talanta 68: 1683-1691.
4. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci U S A 106: 12273-12278.
5. Brooks S, Yan W, Jackson A, Deren C (2008) A natural mutation in *rc* reverts white-rice-pericarp to red and results in a new, dominant, wild-type allele: *Rc-g*. TAG Theoretical and Applied Genetics 117: 575-580.

6. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught Red-Handed: Rc Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. The Plant Cell Online 18: 283-294.
7. Heffner EL, Sorrells ME, Jannink JL (2009) Genomic Selection for Crop Improvement. Crop Science 49: 1-12.
8. Jansen J, Szymańska E, Hoefsloot H, Jacobs D, Strassburg K, et al. (2011) Between Metabolite Relationships: an essential aspect of metabolic change. Metabolomics: 1-11.
9. Breiman L (2001) Random Forests. Machine Learning 45: 5-32.
10. Keurentjes JJB (2009) Genetical metabolomics: closing in on phenotypes. Current Opinion in Plant Biology 12: 223-230.
11. Keurentjes JJB, Fu J, de Vos CHR, Lommen A, Hall RD, et al. (2006) The genetics of plant metabolism. Nat Genet 38: 842-849.
12. Keurentjes JJB, Sulpice R, Gibon Y, Steinhauser MC, Fu JY, et al. (2008) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. Genome Biology 9.
13. Mochida K, Furuta T, Ebana K, Shinozaki K, Kikuchi J (2009) Correlation exploration of metabolic and genomic diversity in rice. BMC Genomics 10.
14. Morreel K, Goeminne G, Storme V, Sterck L, Ralph J, et al. (2006) Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. The Plant Journal 47: 224-237.
15. Carrillo MGC, Goodwin PH, Leach JE, Leung H, Cruz CMV (2009) Phylogenomic Relationships of Rice Oxalate Oxidases to the Cupin Superfamily and Their Association with Disease Resistance QTL. Rice 2: 67-79.
16. Palframan R, Gibson GR, Rastall RA (2003) Development of a quantitative tool for the comparison of the prebiotic effect of dietary oligosaccharides. Letters in Applied Microbiology 37: 281-284.
17. Roberfroid M (2007) Prebiotics: The concept revisited. Journal of Nutrition 137: 830S-837S.
18. Vardakou M, Palop CN, Christakopoulos P, Faulds CB, Gasson MA, et al. (2008) Evaluation of the prebiotic properties of wheat arabinoxylan fractions and induction of hydrolase activity in gut microflora. International Journal of Food Microbiology 123: 166-170.
19. Hughes SA, Shewry PR, Li L, Gibson GR, Sanz ML, et al. (2007) In vitro fermentation by human fecal microflora of wheat arabinoxylans. Journal of Agricultural and Food Chemistry 55: 4589-4595.
20. Franks AH, Harmsen HJM, Raangs GC, Jansen GJ, Schut F, et al. (1998) Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-Targeted oligonucleotide probes. Applied and Environmental Microbiology 64: 3336-3345.
21. Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, et al. (1990) Combination of 16S ribosomal-RNA-targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial-populations. Applied and Environmental Microbiology 56: 1919-1925.

# APPENDIX I: INTEGRATION OF METABOLOMIC AND MICROBIAL ANALYSES

## INTRODUCTION

In some studies, it may be important to integrate metabolomics with non-metabolomics data to characterize the functional consequences of treatment effects. Correlation tests may be an effective integration method, and subsequent hierarchical clustering can identify groups of metabolites that covary. Here, a correlation/clustering approach was used to combine a metabolomics dataset with microbial population fluxes from fecal samples. One human subject was on a dry bean (*Phaseolus vulgaris* L.) dietary intervention for four weeks, and fecal samples were collected at timepoints 0, 2, and 4 weeks. Fecal samples were evaluated for changes in metabolite and microbial profile at timepoints 2 and 4 compared to baseline. Here, these data were integrated to generate hypotheses for mechanisms by which the dietary intervention influenced the metabolite profile via shifts in microbial populations.

## METHODS

Fecal samples were ground to a fine powder in the presence of liquid nitrogen, and metabolites were extracted by adding 1 mL of methanol/water (80:20) to 100 mg of fecal sample, vortexing, and incubating at -80°C for one hour. Samples were centrifuged at 1500 x g for five minutes, and 400 μL of supernatant was transferred to a new tube. Extracts were centrifuged under reduced pressure to remove solvent, and derivatized by resuspending in 50 μL methoxyamine hydrochloride in pyridine (15 mg/mL, 2 hour incubation at 37°C), and then 50 μL of MSTFA + 1% TMCS (1 hour at 37°C). Samples

were centrifuged at 1500 x g for five minutes, and 80 µL of the supernatant was transferred to a glass vial. One µL was used for analysis, and GC-MS parameters and peak detection methods were as previously described [1].

Statistics were performed in R statistical software (v2.11.1), and relied on the packages pcaMethods, qpgraph, and gplots. Data for principal component analysis was mean-centered and Pareto-scaled data. Z-scores were calculated using timepoint 0 as control. Pearson's correlation dendrograms were constructed using hierarchical clustering.

**RESULTS**

Fecal metabolites profiles for three timepoints were assessed by principal component analysis (PCA), and first two components explained 49% and 32% of the variation, respectively (Figure 1A). The three timepoints clustered, indicating fecal metabolite profiles changed due to the dietary intervention. PC1 explained variation due to the 2-week timepoint, whereas PC2 should a steady change from 0 to 2 to 4 weeks. Z-scores were used to assess differences in metabolites using time 0 as a control. From timepoint 0, 2, and to 4 weeks, there was a steady decrease in seven metabolites, which were mostly amino acids and carbohydrates (Figure 1B). There was a steady increase across timepoints in six metabolites, which were mostly stanols and fatty acids (Figure 1C). Transient increases are not shown. The largest increase was in an unknown compound at 19.05 minutes with mass/charge ratios of 291 (100% intensity) and 306 (75% intensity). This compound was also increased in a canine study of a similar diet (data not shown), and should be further investigated.

Fecal microbial profiles were kindly provided by Dr. Tiffany Weir at Colorado State University, and each microbe was valued as percent of the total microbial population at each timepoint. PCA was applied to the dataset, and there were

differences in the microbial population at time 0, 2, and 4 weeks (Figure 2A). The first component (65% of the variation) showed a steady change from baseline to 2 to 4 weeks. The second component (35%) was mostly attributed to variation at 2 weeks. The PCA loadings show Methanobrevibacter and Ruminococcus as microbial genus with the largest changes over the course of 4 weeks (Figure 3A).

The metabolite and microbial profiles both show changes that may due to the dietary intervention. The data were integrated by establishing a Pearson's correlation matrix using mean metabolite (peak area values) and microbial values (percent of population) across the three timepoints, and hierarchical clustering was performed (Figure 3). The data identified three major clusters for metabolites: those that steadily decrease over 4 weeks, those that steadily increased over 4 weeks, and those that increased at 2 weeks, and then decreased at 4 weeks. The heat map was too large to display the correlation matrix for both metabolite and microbial profiles simultaneously, therefore a second Pearson's correlation matrix was generated using metabolite classes instead to individual compounds (Figure 4). The same three clusters of microbes were observed.

**DISCUSSION**

The metabolite and microbial data both showed steady increases, steady declines, and transient increases over the three timepoints. Thus, these data are a good candidate to evaluate the efficacy of a correlation/clustering-based integration. The major limitation to this study was the low sample size for determining correlations (n=3 timepoints). Although a metabolite and microbe may have a high correlation (say 0.9 or greater), statistical significance could not be established at $p < 0.05$ for most of the data. Thus, while the data is highly informative for establishing connections between metabolites and microbes, it is dependent on non-significant correlations. Furthermore,

this analysis assumes normal distributions, and Pearson's correlation analysis is highly dependent on normal data. This metabolomics procedure has been performed before, and the data was determined to be mostly normal. However, the distribution characteristics of microbial profiles are unclear. In this case, a Spearman's correlation matrix may be more informative.

In summary, a correlation/clustering analysis may be informative for combining large datasets. For future studies, including more individuals and timepoints can increase the power of the analysis. It can be a powerful tool to generate hypotheses and to characterize the functional outcome of a treatment on the metabolome. However, validation will be necessary to confirm the established correlations.

## REFERENCES

1. Ryan E, Heuberger A, Weir T, Barnett B, Broekling C, et al. (2010) Rice bran fermented with Saccharomyces boulardii generates a novel metabolite profile with bioactivity. J Ag Food Chem: 1862–1870.

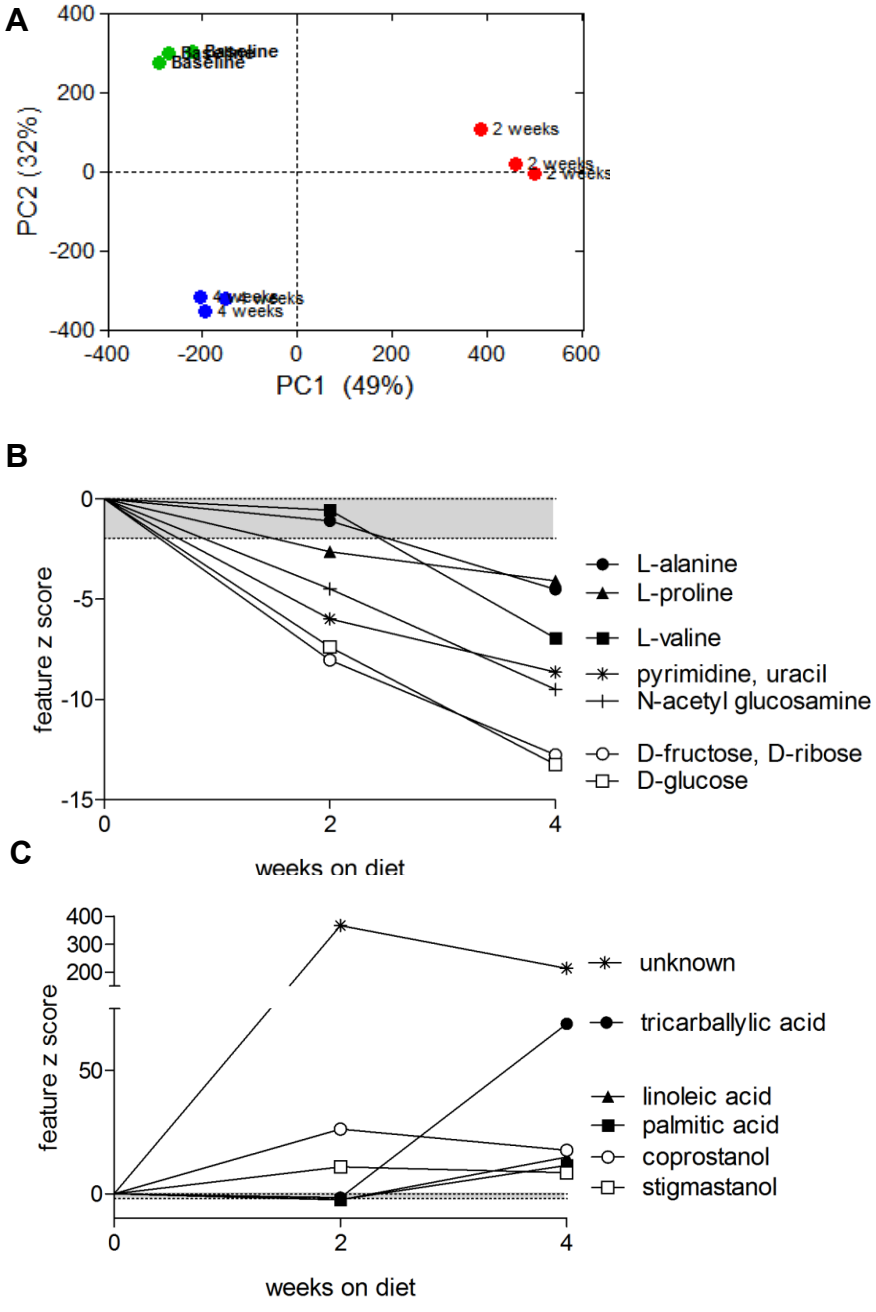


**Figure 1. Metabolite profiling of fecal samples from a person on a 4-week diet.** (A) Principal Component Analysis for metabolites from fecal samples at three timepoints: baseline (0), 2, and 4 weeks. (B) Z-scores for metabolites that steadily decrease at 2 and 4 weeks. (C) Metabolites that steadily increase at 2 and 4 weeks.

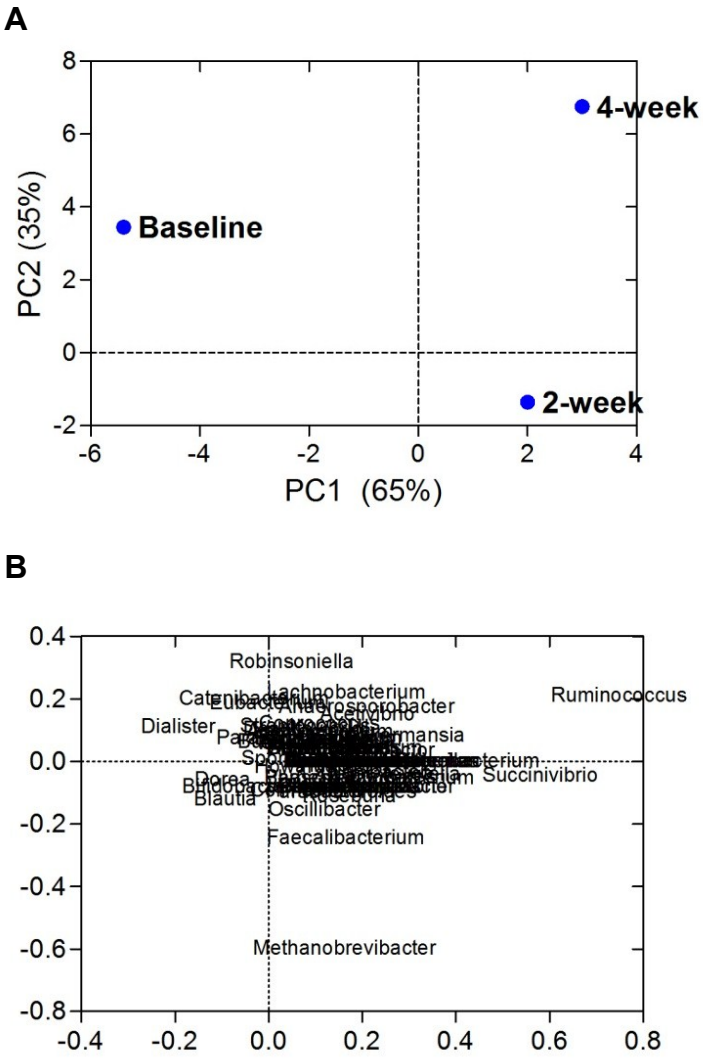


**Figure 2. Metabolite profiling of fecal samples from a person on a 4-week diet.** (A) Principal Component Analysis for metabolites from fecal samples at three timepoints: baseline (0), 2, and 4 weeks. (B) Z-scores for metabolites that steadily decrease at 2 and 4 weeks. (C) Metabolites that steadily increase at 2 and 4 weeks.

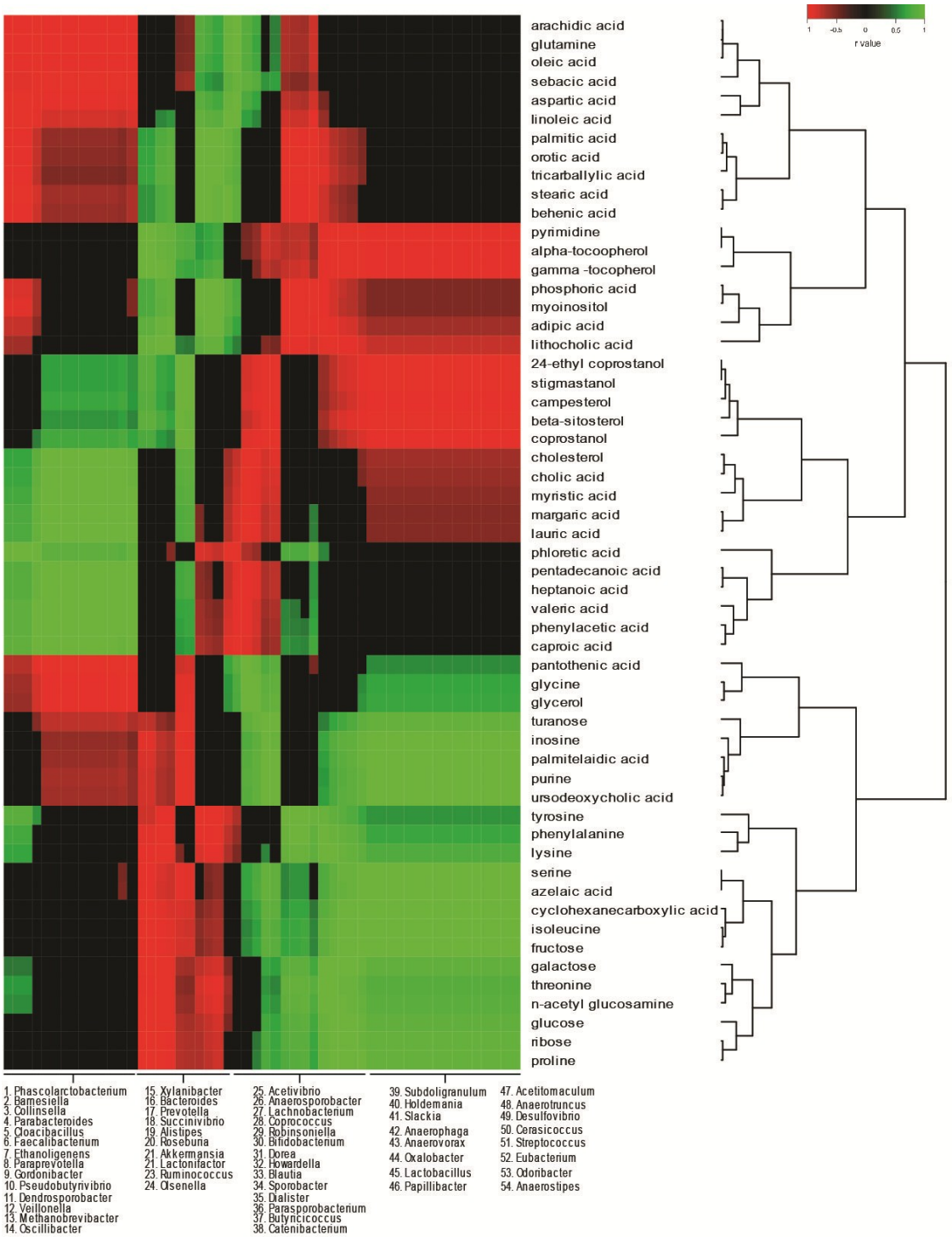


**Figure 3. Integration of metabolite and microbial profiles.** Fecal metabolites and microbial population parameters were correlated and expressed as a heat map. Color represents r values from -1 (red) to 1 (green). Dendrogram denotes hierarchical clustering.

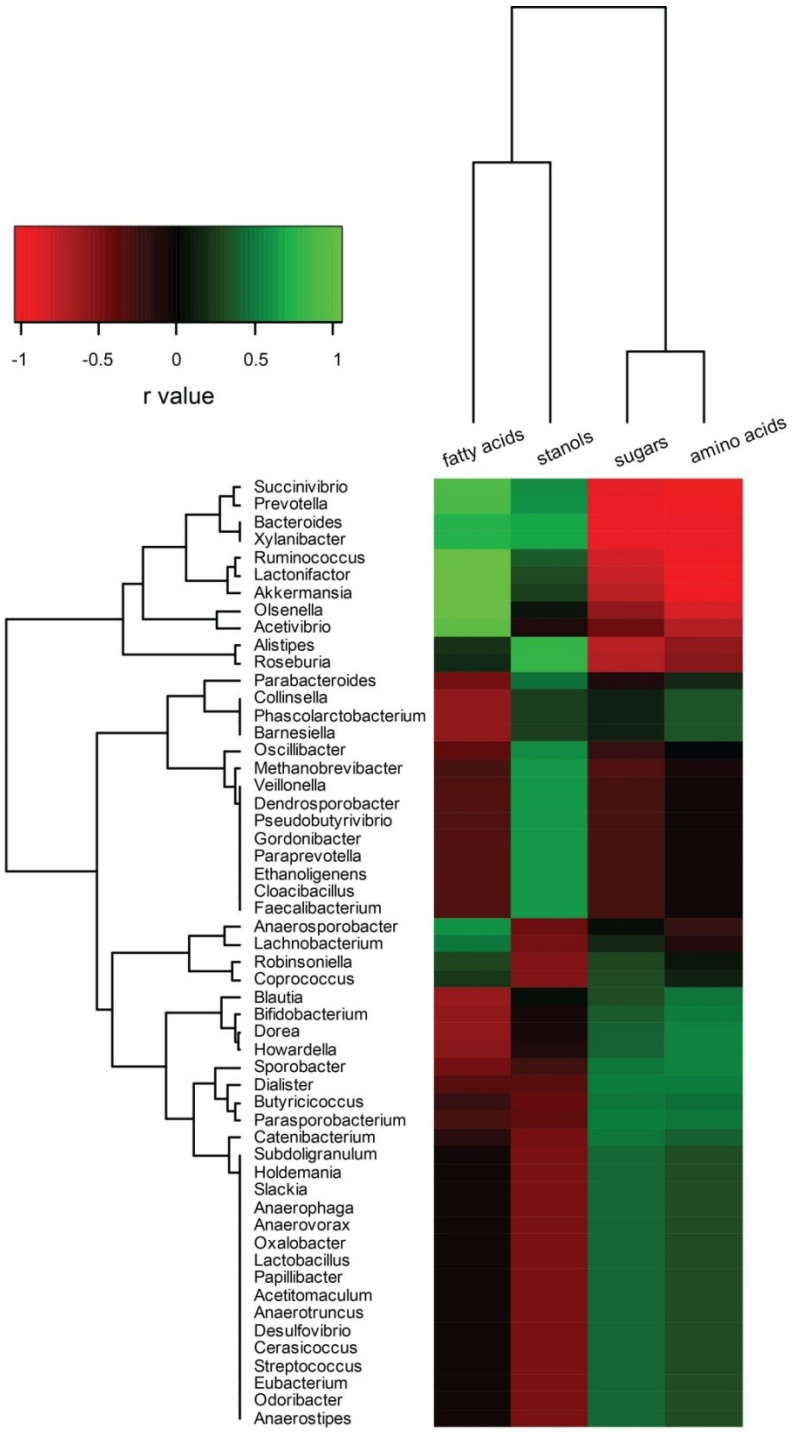


**Figure 4. Integration of metabolite classes and microbial profiles.** Fecal metabolite classes and microbial population parameters were correlated and expressed as a heat map. Color represents r values from -1 (red) to 1 (green). Dendrogram denotes hierarchical clustering.