Spring 2016

Investigation Into The Genetic Basis Of Leaf Shape Morphology In Grapes

Brigette Rachelle Williams

INVESTIGATION INTO THE GENETIC BASIS OF LEAF SHAPE MORPHOLOGY IN GRAPES

A Masters Thesis

Presented to

The Graduate College of

Missouri State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science, Plant Science

By

Brigette R. Williams

May 2016

INVESTIGATION INTO THE GENETIC BASIS OF LEAF SHAPE

MORPHOLOGY IN GRAPES

Agriculture

Missouri State University, May 2016

Master of Science

Brigette R. Williams

ABSTRACT

Leaves are a highly distinguishing characteristic in grape (*Vitis vinifera*) and display great diversity in comparison to other crops. However, little is known about the genetic basis of leaf shape in grape. Here, morphometrics and 40,724 single-nucleotide polymorphisms (SNPs) are correlated to describe the relationship between genotype and phenotype of leaf shape in grape. Parent vines (Norton and Cabernet Sauvignon) and their F1 progeny make up the mapping population located at the State Fruit Experiment Station, Genomics Research Vineyard on Missouri State University campus in Mountain Grove, Missouri. Important leaf shape characteristics (n=17) were identified based on the venation pattern, lobes, and sinuses. Morphometric analysis quantified overall leaf shape and SNP data were used to identify potential quantitative trait loci (QTL) responsible for leaf shape. A General Procrustes Analysis (GPA) produced trait measurements in the form of principal component (PC) scores. Generalized linear model (GLM) and composite interval mapping (CIM) analyses correlated trait measurements and SNPs data to identify three possible QTLs (located on chromosomes 1, 8, and 17) associated with leaf shape in this population. Post-hoc statistical analyses (Bonferroni correction and Benjamini-Hochberg adjustment) indicated SNPs on each of the three chromosomes with statistically significant association ($p<0.08$) to leaf shape in this population.

KEYWORDS: grapes, leaf shape, morphology, genetics, genome-wide association study (GWAS), single-nucleotide polymorphism (SNP), genotyping-by-sequencing (GBS)

This abstract is approved as to form and content

_____

Dr. Chin-Feng Hwang
Chairperson, Advisory Committee
Missouri State University

INVESTIGATION INTO THE GENETIC BASIS

OF LEAF SHAPE MORPHOLOGY IN GRAPES

By

Brigette R. Williams

A Masters Thesis
Submitted to the Graduate College
Of Missouri State University
In Partial Fulfillment of the Requirements
For the Degree of Master of Science, Plant Science

May 2016

Approved:

_____
Dr. Chin-Feng Hwang

_____
Dr. Anson Elliott

_____
Dr. Melissa Remley

_____
Julie Masterson, PhD: Dean, Graduate College

ACKNOWLEDGEMENTS

I would like to thank the following people for their support during the course of my graduate studies: my advisor, Dr. Chin-Feng Hwang, for allowing me the opportunity to work in his lab as his advisee; my dedicated committee members in the MSU Darr School of Agriculture, Dr. Anson Elliott and Dr. Melissa Remley for their essential guidance and support; Dr. Dan Chitwood at the Donald Danforth Plant Science Center in St. Louis, MO for his vital contribution and participation in this project; Hwang Lab manager, Li-Ling Chen, for her direct role in my training at the bench; Dr. Josh Smith in MSU Biomedical Sciences for his support and encouragement in both my current and continued academic pursuits; my fellow lab mates and other graduate students in various MSU programs for their support and camaraderie; and all faculty and staff in the MSU Darr School of Agriculture in Springfield, MO and the State Fruit Experiment Station in Mountain Grove, MO, for their collective efforts, support, and knowledge which make all graduate student work possible.

I dedicate this thesis to my parents, Eddie & Kelly Williams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Historically, plant breeders have been well served to rely upon personal, direct observations in selective breeding of agricultural crops. This method has allowed breeders to develop stronger, heartier, and more productive crops that better suit consumer demands. Cold tolerance, disease resistance, drought tolerance, yield increases, and reduced shattering of seeds are examples of traits that have been selected for successfully by observation. The methods aiding artificial selection in agricultural crops, as well as the traits for which we may select, seemingly advance and expand in tandem, driven by the manifestation of new demands, threats, and problems. With the advent of agricultural biotechnology, characteristics that may have once been considered immutable by farmers are now within reach of selection and controlled modification. As well, broadening ecological and environmental concerns continually arise from climate change research, and traits that were previously of secondary consideration (if at all) are now subject to further scrutiny in an attempt to infer an adaptive context from highly managed agricultural crops with extensive domestication histories.

This new era of precision breeding has agricultural scientists redefining what 'crop improvement' means by working to unlock the genetic code of plant diversity and variation. Among the many plant characteristics that display diversity, leaf shape is perhaps one of the most easily observed though the possible causal relationships remain unclearly defined. The study of leaf shape is deep and complex, with new vigor being breathed into it in pursuit of clarifying the inseparable relationship between plant form and function with a modern, ecological bent for agricultural and conservation purposes.

In this thesis, I provide appropriate historical and modern, theoretical and applied contexts to the importance of genetic characterization of leaf shape morphological diversity in the agricultural crop of wine grapes. I offer sensible reasoning for investigating leaf shape diversity in wine grapes, with an aim to contribute to the foundation of descriptive knowledge in morphology upon which further work may stand to better elucidate the practical implications of morphological study specifically in wine grapes.

Vitis

Grapevines are classified into the genus *Vitis*, which is part of the flowering plant family, Vitaceae (Terral et al., 2009; Myles et al., 2011). They are perennial lianas; woody, climbing vines that are naturally found in paratropical forests of summerwet biomes (Willis & McElwain, 2002). Wild grapevines are typically found growing in loose soil along riverbanks and sandy, loamy soil is the preferred type for cultivated vineyards (Terral et al., 2009). Grapes are a long-season crop and commonly will not produce fruit until 3-5 years of age, depending on the cultivar (Avery et al., 2003).

The evolution and history of grapevine domestication is long and rich. Grapevines are one of the oldest and most valuable horticultural crops in the world, with cultivation and domestication dated to between the seventh and fourth millennia BC in an area between the Black Sea and modern-day Iran (Terral et al., 2009; Myles et al., 2011). Winemaking is dated to antiquity with discoveries of vinification residues in clay jars and archaeological grape seeds, thought to be from cultivated grapevines, dated to the mid Bronze Age. Viticulture has been tracked from its beginnings in the Near East (the

geographical region of countries in western Asia), spreading gradually westward into the Middle East and then Central Europe. However, this model of viticultural and vinicultural expansion still lacks detail of temporal and spatial varietal diversification. Morphological data used in the identification of seed and wood archaeological remains has been collected in recent years, and is being used to clarify patterns of cultivation of ancient varieties (Terral et al., 2009).

Currently, identification of modern grapevine cultivars is easier, more reliable, and understood at a much finer scale by analyzing genetic diversity. The accurate identification of grapevines has been historically important in cultivar development due to the expansive variety of grapevines that share great similarities from having been crossbred throughout domestication. Before the advent of genetic technologies, identification by vine characteristics was extremely important in breeding and tracking lineage (Chitwood et al., 2015). Not only is this information useful in outlining the domestication of grapevines, but it also supplements present understanding of biological changes that have taken place during domestication by way of comparative studies between wild grapevines and modern cultivars (Terral et al., 2009).

Two cultivars are of specific interest in this study: *Vitis aestivalis*-derived 'Norton' and *Vitis vinifera* 'Cabernet Sauvignon'. The origination of Norton has been traced to the 1820's in Richmond, Virginia, and credited to a local medical doctor by the name of Daniel Norborne. Dr. Norborne was an amateur viticulturalist and plant breeder who is thought to have produced Norton via an unintentional breeding by open-pollination of a mother vine, Bland, with an as-of-yet unconfirmed pollen parent (Ambers & Ambers, 2004; Ambers, 2013). The pollen parent is speculated to be a cultivar of *Vitis*

aestivalis from observations on vine typology, leaf morphology, berry traits, and genotyping analysis. Based on this information, we know that Norton is a North-American native grape with a lineage stemming from *Vitis aestivalis*. Norton has been grown in Missouri since the 1830's when it was brought from Virginia by German settlers who appreciated its cold-hardiness and disease resistance (Ambers & Ambers, 2004; Ambers, 2013; Teeter, 2014; Missouri Wines, 2016). Missouri winemakers brought the Norton grape to worldwide recognition with its rich, full-bodied red wine. Norton's increasing popularity as a wine grape was thwarted in 1920 with the onset of Prohibition in the United States (Teeter, 2014). The Norton wine grape was largely forgotten by the end of Prohibition in 1933, failing to resume popularity until 1989 when native Missourian, Dennis Horton, established a small vineyard in Virginia and dedicated a small section of it to bringing back Norton (Roberts, 1999; Teeter, 2014).

Cabernet Sauvignon is a far more familiar grape in the global wine industry than Norton. It is widely regarded as one of the highest quality French wine grapes. The origins of Cabernet Sauvignon are believed to be the Bordeaux region of southwest France, approximately 400 years ago, with the first documented importation to North America in 1852 by French nurseryman, Antione Delmas, to the Santa Clara Valley. Vineyards in the Santa Clara Valley began planting Cabernet Sauvignon around 1858 (Sweet, 2008).

Genetic testing has proven Cabernet Sauvignon was produced by crossing Cabernet franc and Sauvignon blanc. This was likely a spontaneous crossing of the two as there was no documented grape breeding in the Bordeaux region during that time. Cabernet Sauvignon is a favored wine grape due to its excellent wine quality and it is

considered easy to grow due to hard berries and wood. The vine prefers a moderately warm climate and berries ripen late season and slowly. The berries are also insensitive to harvest time and can remain on the vine for a relatively long time (Sweet, 2008). The wine produced from Cabernet Sauvignon, like Norton, is a full-bodied red wine and has distinct dark fruit flavors and savory taste (Puckette, 2012).

Norton and Cabernet Sauvignon display differences in their berry and leaf characteristics. Cabernet Sauvignon berries are small, round, thick skinned, dark blue-black in color, and the clusters are small to medium, loose but well filled, and conical in shape. Cabernet Sauvignon leaves are medium in size, with deep lobing, medium serration, an overlapping petiolar sinus, a dark green, smooth adaxial side, and lighter abaxial surface with light, scattered hair (Wolpert, 2006). Norton berries are similar to Cabernet Sauvignon in that they are small, round, dark blue-black, with a tough skin. However, Norton berries form small to medium compact clusters that are typically not uniform, and taper into a single-shouldered cylindrical shape (Smiley, 2008). Norton leaves are very different from Cabernet Sauvignon leaves. They are medium to large (in some cases, nearly twice the size of mature Cabernet Sauvignon leaves), dark green leaves with shallow lobing and small teeth. Easily observed differences in phenotype, as well as, genetic differences based in monophyletic origins and distinct evolutionary lineages combined with artificial selection during domestication make these two cultivars ideal candidates for comparative studies.

Grapes have been demonstrated as an important plant economically, agriculturally, biologically, and ecologically. The fields of agriculture, evolution, botany, ecology, genetics, and anthropology and archaeology all have individual interests driving

a shared pursuit of advancing contemporary knowledge about grapes. Therefore, thoughtful questions, study design, and interdisciplinary collaborations centering on grapes can result in broadly impactful results.

Morphology

Plant morphology is a unifying discipline that focuses on convergences of presumably unrelated forms by seeking to understand that which underlies observed variation. The origin of morphology is pre-Darwinian, dating to more than 200 years ago when German philosopher Johann Wolfgang von Goethe published a highly influential work titled "Versuch die Metamorphose der Pflanzen zu erklären" (translated "An Attempt to Explain the Metamorphosis of Plants") (Kaplan, 2001; Jensen, 2014). His ideas were a departure from Carl Linnaeus' reigning botanical theory in Europe at the time (Jensen, 2014). Linnaeus classified plants in a taxonomic schema that identified relation to other species, genera, and kingdom according to external characteristics such as size, number, and location (Waggoner, 2006). Recognizing that developmental aspects were ignored entirely in Linnaeus' system, Goethe determined that a fundamental organizational theme linked morphological diversity among plants. He greatly expanded understanding with empirical observation of patterned changes throughout development as plants interact with their environment (Kaplan, 2001; Jensen, 2014).

Plant morphology, in its purest form, remains largely Germanic in practice. Origins stem from studying the natural history of plants and focusing on homologies. In the United States, the field has morphed in its own right, focusing heavily on plant systematics and relying on phylogenetic tools and methods to explain morphological

relationships. Although some tension remains between idealistic and phylogenetic approaches in morphology, there is no actual conflict between these two perspectives; Darwinian evolution (and subsequently, phylogenetics) supplies the necessary explanation for observed homologies based in idealistic morphology. This relationship provides the critical argument in support of a mutual importance to morphology in that the diversity of plant form observed today is a result of evolution based on ties between phylogenetic formulations and form/structure relationships (Kaplan, 2001).

The historical trends of plant morphology follow a trajectory familiar to most fields of science: description of phenomena, classification of phenomena, and the investigation of causal links. Most plant morphologists' work is interdisciplinary in nature and necessarily focused at the interface of morphology and tangential disciplines. Plant morphology emphasizes analogies constructed from developmental and evolutionary theories and discoveries but it also includes information from many other diverse fields of study to comprise sound, holistic explanations of form. Plant systematics, plant ecology, plant genetics, and plant physiology all contribute to plant morphology. Morphology does not duplicate other fields; rather, there is a mutual exchange of information between and among them (Kaplan, 2001).

Many leaf shapes have evolved numerous times as independent characters in the evolutionary history of angiosperms (Nicotra et al., 2011). From a modern perspective, plant morphology contributes directly and equally to plant genetics in the characterization of phenotype. Molecular geneticists and plant morphologists share increasingly overlapping interests regarding the causal aspects of plant morphology, in an attempt of

explaining what combination of environmental and genetic circumstances may create certain leaf shapes (Kaplan, 2001; Chitwood et al., 2015).

Ampelography

Leaf morphology can be further quantified by ampelography, which is a morphology based classification and identification of grapevines (*Vitis*) based on leaf traits and characteristics. Leaf morphology covers many aspects of leaf structure and make-up, much of which have been useful to ampelographers: venation patterning, leaf mechanics, hirsuteness, blade outgrowth, leaf positioning, lobing, sinuses, serration, color, size, and contour (Chitwood et. al., 2014). Even within a species, notably seen in grapes, the morphology of a leaf can change depending upon the developmental stage of the leaf. Such changes in allometry are referred to as heteroblasty (Costa et al., 2012).

Many examples of leaf form diversity can be observed in natural plant communities and ancestral form variation is echoed in populations of domestic agricultural crops like grapes. In particular, grape displays profound leaf shape diversity from heteroblastic juvenile forms to maturity, and across its many cultivars, more so than many other domesticated crops (Chitwood et al., 2014). At maturity, some cultivars have extreme lobing and serration and are so heavily dissected they are technically a compound leaf form. Others are more orbicular with little to no lobing or serration. This expansive variation in leaf shape warrants *Vitis* special attention from morphologists, who come to conclusions regarding structural relationships of leaves by performing comparative studies of plant form between species (Kaplan, 2001). Ampelography has tailored this study design concept to grapes for the specialized purpose of cultivar

identification. Grape is an ideal plant for morphological study because extensive leaf shape variation among its many cultivars provides the necessary variant forms for a comparative study. Additionally, ampelographic methods serve as an archetypal design for accurately capturing the nuances of continuously varying leaf shape.

In 1952, Frenchman Pierre Galet published "A Practical Ampelography: Grapevine Identification" detailing a standardized system for measuring the phenotypic shape characteristics of grape leaves. He measured length from the petiolar branching point to the tips each of the midvein, distal, proximal, and petiolar veins. He also measured the distance from petiolar branching point to the upper and lower lateral sinuses, and corresponding angulation between each vein (Galet, 1952; Chitwood et al., 2014). Using this system, Galet determined the phenotypical classification of roughly 9,600 vines (Galet, 1952). Galet's 1952 publication remains the authoritative reference on ampelography 62 years later; it is upon this that we have based the phenotyping design for this study.

Goethe's perceptivity and unique insights along with Galet's standardized system of trait measurements, have built their respective fields of morphology and ampelography into substantial areas of study. Though morphology and ampelography do not share the same motivations behind their respective inceptions, they are certainly complementary as investigatory subjects.

Functional Significance

Leaves continue to be of ecological and environmental importance, transcending boundaries between scientific disciplines. Leaf shape variation can be categorized into

three classes: across communities, within lineages, and within individuals. It is generally accepted that there is some functional or adaptive significance to leaf shape given its role as the major photosynthesizing organ (Nicotra et al., 2011).

There are numerous theories regarding functional significance of leaf shape: thermoregulation, hydraulic constraints, patterns of leaf expansion during development, mechanical constraints, adaptations to avoid herbivory, optimization of light interception, and representing an effect of selection on flower form (Nicotra et al., 2011). Leaf lobing, serration, and dissection affect light penetration through the canopy. Intensity of light filtering through the canopy affects berry composition and fruit cluster development. Increased light penetration decreases fungal infections common in grapevines such as powdery mildew and black rot (Spotts, 1977; Austin et al., 2011; Austin et al., 2012). Venation patterning, which greatly affects overall leaf shape (Galet, 1952; Chitwood, 2014), and distance of veins to laminar mesophyll, also influences photosynthetic performance and hydraulic efficiency (Brodribb et al., 2010; Sack et al., 2013).

Many of the aforementioned functions are determined as part of a leaf economics spectrum, which consists of key chemical, structural, and physiological properties that are scaled against a quick-to-slow 'return on investment' continuum in the form of modulation of leaf traits and trait relationships (Royer et al., 2005; Nicotra et al., 2011; Wright et al., 2004). Therefore, diverse leaf traits likely represent structure-function trade offs due to the conflicting requirements of physiology, anatomy, and morphology for its given environment. For instance, leaf positioning that maximizes light interception also negatively impacts heat dissipation. The Optimization Theory suggests that the number of

phenotypic solutions that allow for different, but equally successful, trait combinations increases as the number of trade-offs increase (Nicotra et al., 2011).

There are several leaf traits known to be homologous to all vascular plants, which share a monophyletic origin from a leafless common ancestor. In general, evolutionary changes in the angiosperms created great opportunity for diversity of form while maintaining functional possibilities of a leaf. Reticulate venation pattern is a shared evolutionary change observed in a vast majority of angiosperms; it releases restrictions on leaf shape by allowing for novel growth of leaf margins, which is where discrete tissue production is limited. Variation in angiosperm venation patterning revolutionized water distribution by creating equitable water access for leaf tissue furthest from the base (Nicotra et al., 2011). This allowed plants to grow and maintain leaf tissue in diverse shapes because the higher vein densities of reticulate venation patterns minimize risk of water loss due to transpiration. Accordingly, high vein densities and reticulate venation patterns are both associated with leaf shape diversity (Nicotra et al., 2011; Zwieniecki et al., 2002).

Leaf shape morphology is one aspect of plant environmental adaptations that has been documented in fossil record and continues to be of importance to paleoclimatologists and paleobotanists (Chitwood, et. al., 2014). Plant physiognomy uses leaf size and shape as proxies for temperature and moisture variables to infer and reconstruct paleoclimate (Royer et al., 2005; Seward, 1892; Parrish, 1998). Interest in ecological relationships that may be represented by leaf shape continues today as scientists attempt to predict adaptive capacities of natural plant communities and agricultural crops alike in response to anticipated climate change.

In historically warmer and wetter climates, fossil records indicate larger, entire leaves were the primary leaf type, while smaller, dissected leaves were predominant in drier and cooler climates (Wilf et al., 1998; Chitwood et al., 2012). Large leaves are thought unlikely to have evolved while CO_2 levels remained high because stomatal density would have been insufficient for large leaves to perform the evaporative cooling required under conditions of intensified solar radiation (Royer et al., 2005; Pires et al., 2012). Small leaves are associated with harsh conditions in general, suggesting a lower energy investment in leaf size when faced with environmental factors that challenge plant survival (Nicotra et al., 2011; Pires et al., 2012). Although while variability of leaf size has been shown to have predictability across environments, current knowledge regarding shape diversity in natural plant communities indicates that a singular leaf shape may not be most ecologically beneficial in a generalized context (Royer et al., 2005; Nicotra et al., 2011; Pires et al., 2012). Interestingly, venation patterning (which has been shown to strongly influence overall leaf shape in many species) has been observed to vary with climate throughout paleohistory (Sack et al., 2013).

Perhaps the most extensively studied physiognomic trait, leaf margin analysis (LMA) is based on a single character state – the presence or absence of teeth on the leaf margin. It is known to be the oldest and most reliable physiognomic technique (Royer et al., 2005). LMA based on modern forests have found that the percentage of woody perennial species with untoothed leaf margins correlates significantly to mean annual temperature (Bailey & Sinnott, 1915; Bailey & Sinnot, 1916; Wolfe, 1979; Wilf, 1997). LMA can also be used for placement of fossil leaf species when molecular data is

incomplete because leaf physiognomy has been demonstrated to reflect convergent response to climate in different lineages (Royer et al., 2005).

The evolution of plant leaf morphology in response to factors of climate change indicates functional importance of leaves (Chitwood et al., 2014). Rising CO_2 levels is one factor in climate change known to affect leaf morphology. The rise of CO_2 levels causes increases in grape leaf surface area and partial stomatal closure, both of which affect thermoregulation (Moutinho-Pereira et al., 2009). No current data exists regarding to what extent rising CO_2 levels impact leaf size in *Vitis*. In spite of some documented favorable morphological adaptations in response to elevated CO_2 levels, grape yield and quality are still expected to suffer due to climate change (Schultz et al., 1998; Kolbe et al., 2001; Chitwood et al., 2014). Studies have shown that as UV-B radiation increases, leaf expansion, overall biomass, and photosynthetic capacity decrease (Jansen, 1998; Schultz et al., 1998; Kolbe et al., 2001). Overall, leaf shape does not appear to definitively explain ecological differentiation between species, rather it represents varied structure-function trade-offs. A recurrent theme in structure-function trade-offs is the balancing of influential leaf-water relations (Nicotra et al., 2011; Royer et al., 2005). Undoubtedly, thorough understanding of plant leaf morphology and how it relates to climate change is a timely issue.

Morphometrics

Morphological understanding is being furthered through morphometrics, a quantitative analysis of form (Chitwood et al., 2014). There are several ways to quantify the relationship between size and shape. The morphological traits of specific interest in

Vitis are aspect ratio (AR), circularity, symmetrical shape variation, and venation patterning. In particular, for this study, I will focus on those characteristics and traits that are easiest to measure, as well as previously identified in historical literature to greatly influence the overall shape of the leaf; venation pattern, lobing, and sinuses. Venation pattern also captures AR data.

ImageJ, a public domain image-processing program, accurately and efficiently measures and compares morph traits (Abramoff, 2004; Chitwood et al., 2014). Leaves are scanned using a large-format scanner and uploaded into ImageJ. Landmarks (n=17) are sequentially plotted on each leaf image in ImageJ so that each landmark on each image represents the same landmark as on each subsequent image for every leaf sampled throughout the population. Structural points of interest for landmarking are as follows: petiolar junction, leaf tip opposite the petiolar junction or tip of the midvein, upper lateral sinuses, tips of upper lateral lobes, lower lateral sinuses, tips of lower lateral lobes, and the first major branching points for each of the midvein, proximal, distal, and petiolar veins.

Placement of these landmarks is critical because form differs between individual leaves. Each landmark in a set of 17 represents a fixed structural point of the leaf form that can be linearly compared across a population. Landmarking structural points in ImageJ converts phenotypic characteristics into a set of linear values, which can then be used for morphometric analysis.

AR is the ratio of the major axis to the minor axis of a fitted ellipse. AR tells us how balanced a leaf is in regard to length and width, and it is often measured in partnership with circularity. Circularity is a ratio of the surface area to perimeter of the

shape outline. Circularity reflects the presence of lobing and serration of the leaf outline (Chitwood et al., 2014).

Symmetrical shape variation describes the distinct sinuses, lobes, position of petiolar veins, and the angular distances between major veins. Symmetrical shape variation is not represented by AR or circularity but instead is captured by analysis of leaf outlines and the resultant principal components (PC). The PCs from individual leaf assays can then be analyzed together to describe symmetrical shape variation of a population. Venation patterning affects overall leaf shape to a great extent; positioning of lobes and sinuses appears to be determined by the branching angles of the veins. The use of principal component analysis (PCA) based on inner and outer venation pattern landmarks can describe leaf shape variation among a population (Chitwood et al., 2014).

Genetics

Phenotype can readily be described by observation but relying solely on observation results in limited understanding of complex traits (i.e., continuously varying / quantitative traits). A more thorough description of phenotype can be made using genetics (Nordborg et al., 2008). Development of genotyping resources in concert with phenotypic analysis of segregating populations helps to construct a clearer picture of the relationship between genotype and phenotype (Chitwood et al., 2014). Several methods currently exist for genotyping, but the ultimate goal is to identify chromosomal locations likely to contain genetic information for traits of interest.

Markers are perhaps the single most useful component in genetic characterization studies. They are required to pinpoint quantitative trait loci (QTLs) without either of

which we would not be able to fully determine the genetic basis of observable traits and characteristics. There are three main kinds of markers used for QTL identification and map construction: morphological markers, biochemical markers, and DNA/molecular markers. Morphological markers are the phenotypic traits themselves (like flower color), biochemical markers are isozymes (enzymes which differ in amino acid sequence but retain the same function), and DNA markers identify a specific location within a genome (Collard et al., 2005).

DNA markers are not usually part of the target gene itself, but rather indicate regions of DNA flanking a gene. They allow us to quantify recombination events that take place during meiosis, as well as providing information about genomic events like mutations, insertions and deletions (indels), and duplication events (Myles et al., 2009). Genomic and molecular genetic information is widely useful in the broad subfields of biology by giving insight into evolutionary events, disease resistance and susceptibility, developmental processes, cell signaling, heritability, and more. However, much of this information would be unknown without molecular markers.

Different methods of discovery are used to produce many kinds of markers; DNA-DNA hybridization, polymerase chain reaction (PCR), and genotyping-by-sequencing (GBS) based. For the purposes of this study, we will focus on the GBS method of discovery and the relevant molecular markers, single-nucleotide polymorphisms (SNPs). SNPs are the result of an individual base pair change in a DNA sequence. They can be created by mutations and the effect of a SNP can be none at all or greatly impactful depending on the location and nature of the mutation. SNPs are usually

found in the non-coding region of DNA between genes, making them useful as markers in the identification of genes for traits of interest (Lodish et al., 2008; Weaver, 2012).

The method by which SNPs are discovered is known as Genotyping-By-Sequencing, or GBS, which is based on high-throughput next-generation sequencing (NGS) of genomic subjects that are targeted by restriction enzymes (REs). GBS is currently very popular for many reasons; it's technically simple, highly multiplexed, inexpensive, fast, specific, highly reproducible, and analysis demands minimal computational power (Glaubitz et al., 2014).

The GBS process requires high-weight molecular DNA to be digested with an appropriate RE in consideration of desired coverage and genome size. Selection of appropriate REs may vary depending on species and study purposes. It is critical to select a RE that leaves a 2-3 basepair (bp) "sticky end" or overhang for ligation of adapter/barcode primers. Up to 96 unique DNA samples can be simultaneously processed, making this a highly efficient and cost-effective process. Genomic DNA samples are placed into individual wells that also contain adapters, and the selected RE is added. The DNA is digested, and then adapters are ligated to the sticky ends of the digested DNA fragments. There are generally two kinds of adapters used in GBS; the first adapter ends in a "barcode" sequence consisting of a 4-8 bp motif on the 3' end of the top strand, and 3 bp overhang that is complementary to the sticky end of the digested DNA on the 5' end of the adapter's bottom strand. The second adapter is called a "common" adapter and it only has a sticky end RE digested compatible bp sequence (Elshire et al., 2011).

Aliquots of the digested DNA fragments with ligated adapters are pooled and the samples are cleaned up using a size exclusion column to filter out unreacted adapters. Oligonucleotide primers that are complementary to the adapter sequences, which have been ligated to the DNA fragments, are added to the sample and PCR is performed to amplify the pool of fragmented DNA to create a "sequencing library" for analysis on an automated electrophoresis machine.

Fragment libraries are evaluated to identify those that have less than 0.5% adapter dimers, as well as fragments of 170-350 bp in length. These fragments are then single-end sequenced on an Illumina Genome Analyzer and the resulting raw sequence data is filtered to produce fastq files for good reads that align perfectly to the reference genome for a minimum of 64 bases. SNPs from the resulting library can be used to identify significant associations between SNPs and the trait of interest, indicating potential QTL(s) (Collard et al., 2005; Elshire et al., 2011; Weaver, 2012; Glaubitz et al., 2014).

QTLs are regions of chromosomes that have been identified to contain genetic information relevant to target genes and trait expression for a particular quantitative trait. QTLs can be identified for segregating traits of a parent population used for mapping or by association studies that are based on allele frequencies that correlate to a measured phenotype (Collard et al., 2005; Sleper et al., 2006; Myles et al., 2009). They are referred to as 'quantitative' because QTLs are regions of polygenic inheritance, i.e., they contain more than one gene affecting trait expression (Nordborg et al., 2008; Wei et al., 2014). Many phenotypic traits are due to polygenic inheritance. QTLs can be identified via two different forms of mapping: linkage mapping and association mapping. Linkage mapping, which relies on a breeding population, relies on the theoretical basis of recombination

frequency between genes that are arranged in a linear fashion at different loci on chromosomes (Sleper et al., 2006; Hyde, 2009). Linkage mapping can be slow, labor intensive, and expensive due to the need to construct a breeding population, as well as the laborious genotyping process, traditionally requiring microsatellite marker analysis (Myles et al., 2009). For the purposes of this study, we will focus on association mapping.

Association mapping is based on the rate at which linkage disequilibrium (LD) decays and the strength of correlation between genotype and phenotype is a function of the distance between two markers; closer in distance equates to stronger LD. LD is the non-random association of alleles. It varies among loci within populations and the rate at which it decays depends on the species of interest. In outcrossing populations, like grape, LD breaks down much faster than self-fertilizing populations. It is estimated that LD decays within approximately 100-200 bp in grapevines (Lijavetzky et al., 2012; Myles et al., 2009). Based on this, association mapping may not offer much, if any, advantage over family-based linkage mapping in the detection of QTL depending on how extensive LD decay really is in grapevines and other similarly high outcrossing species. The ultimate goal in association mapping is to identify genetic variants that are responsible for phenotypic variation. In this way, genotyped markers act as proxies for functional variants (Myles et al., 2009).

Association mapping has some strengths that family-based linkage mapping does not. As the name implies, this method is based on association of allele frequencies with the measured trait of interest. It searches for correlations of genotype and phenotype in large populations of unrelated individuals; meaning, it does not require a breeding

population, which can save time, money, and labor that would otherwise go toward the development of such a population. Because of this, it can be a good choice for organisms that cannot be crossed or cloned, as well as those that have long generation times. Association mapping is able to exploit all recombination events that have occurred within the evolutionary history of a sample population, resulting in higher resolution than linkage mapping. The number of QTL identified are not limited to parental segregation of a cross; rather, it relies on the number of real QTL underlying the trait and the degree to which the sampled population captures the total naturally available genetic diversity (Myles et al., 2009). In particular, this can be very useful in marker-assisted selection (MAS) plant breeding.

Because association mapping does not require a breeding population that has been created with the explicit intentions of producing progeny that will segregate for the trait of interest, it is absolutely necessary that the sampled population have a diverse enough germplasm to capture historical recombination events within the species for reliable correlative analysis (Myles et al., 2009). Diversity of germplasm is crucial. If care is not taken during sample collection, certain alleles may show up more frequently in certain populations sampled than is expected were mating to occur truly at random. It can create false-positive associations by loading up on specific alleles found in one subset of a population. Enough markers need to be genotyped across the genome so functional alleles are probabilistically in LD with at least one of the genotyped markers. The number of markers and density necessary are determined based on genome size and LD decay, both of which depend on the species of interest (Myles et al., 2009). Fortunately, the

continually decreasing costs of NGS make obtaining full sequence data of large population samples monetarily feasible.

Using a family-based design for an association study has advantages over either individually. As mentioned above, false-positive associations have been problematic in association studies and population structure can skew results. However, when a family-based design is used for an association study, linkage and association are always implied by significant findings because population substructure is a non-issue in that allele frequencies should be consistent (i.e., Mendelian transmission patterns are observed) throughout a breeding population (Laird & Lange, 2006). In this study, we have chosen to take advantage of the benefits offered by such a study design and use a family-based breeding population paired with correlation analysis of the phenotype and genotype for an association study to provide genetic characterization of leaf shape in this population.

Problem Statement

Grape leaves exhibit impressive diversity of shape to an unparalleled extent compared to other crops. However, in spite of the generally accepted importance of leaves, the genetics controlling leaf shape in grape are not clear. Lacking genomic and genetic context acts as a barrier to fully understanding the functional implications and consequences of leaf shape relevant to plant physiology, development, organismal interactions, and environmental adaptations. If the genetics controlling leaf shape in grapevines were better understood, it could allow breeders to consider the optimum leaf shape for each cultivar, dependent upon individualized needs and breeding goals, using marker-assisted selection (MAS) breeding techniques. In broader terms, the genetic

characterization of grape leaf shape could assist breeders in preserving terrior - a sustainable interaction between genotype, environment, and cultural demands (Chitwood et al., 2014).

Hypothesis

In order to find a statistical correlation between phenotype and genotype, we will measure phenotypic traits of a segregating population between two morphologically distinct grape varieties and perform a SNP association study using both Mixed Linear Model and Generalized Linear Model approaches to look for associations and corresponding strengths. We will measure venation patterning, and lobe and sinus positioning of both of the parent populations and the F1 generation.

I will use a library of SNP markers constructed by the Cornell University Genotyping-By-Sequencing Project as part of a collaborative effort in the VitisGen1 project. I will attempt to identify statistically significant associations between phenotype and genotype data to identify potential QTLs responsible for leaf shape in an F_1 Norton x Cabernet Sauvignon progeny population.

METHODS

Study Design

In this study, I attempt to describe the correlated relationship between grape leaf shape and its genetic basis. The grape varieties of Norton and Cabernet Sauvignon were selected due to their highly distinct leaf morphs; Norton leaves are large with subtle lobing and shallow sinuses, and Cabernet Sauvignon leaves are notably smaller and display prominent lobing and deep sinuses (Figure 1).

Leaves were scanned on a large format scanner and images were processed in ImageJ using the landmark and measurement tools. Phenotype was measured by landmarking (n=17) the venation pattern, lobes, and sinuses of each leaf. Genotyping consists of a SNP library constructed from the F_1 progeny population genomes; this was performed at the Cornell University Institute of Biotechnology as part of the collaborative VitisGen1 project. Genotype-phenotype correlational analysis was carried out by Missouri State University's molecular breeding lab under the direction of Dr. Chin-Feng Hwang. Morphometrics were completed in the software program R. Association analysis was performed in Trait Analysis by aSSociation Evolution & Linkage (TASSEL).

Collection & Scanning

Leaf samples (n=4 per plant in 2014, n=6 per plant in 2015) were collected from the mapping population maintained at the Missouri State University Mountain Grove Fruit Experiment Station Genomic Research Vineyard in late July and early August of 2014, and again in early August 2015. Mature leaves were collected from the Norton

(n=14) and Cabernet Sauvignon (n=18) parent populations. Three F_1 populations consisting of CSxN vines (n=18) planted in 2005, NxCS vines (n=71) planted in 2005, and NxCS vines (n=158) planted in 2011 were also collected. In 2014, leaves (n=2) were sampled from shoots (n=2) on each F_1 plant such that leaves (n=4) were collected from most plants. In 2015, leaves (n=2) were sampled from shoots (n=3) on each F_1 plant such that leaves (n=6) were collected from most plants. Only mature, fully expanded leaves for which venation pattern and leaf margin remained intact and easily identifiable were selected and stored in Ziploc bags in a travel cooler until scanning, which immediately followed collection each day.

Leaves were scanned the same day as collection on a large format, color scanner (Mustek A3 1200S) at a resolution of 300 dpi and arranged face down to scan the adaxial side with the petiolar junction oriented toward the bottom of the image. Large leaves were scanned individually. Scans containing multiple leaves were limited to leaves of the same plant and the leaves were arranged so as to prevent overlap. Images were saved as JPEGs with a file name indicating individual genotype and vineyard location.

Landmarking

Leaf images were landmarked in the image-processing program ImageJ using the 'landmarking' tool. Morphological points for grape leaves (n=17) have been previously identified by Chitwood and colleagues based on Galet's work (Chitwood et al., 2014). Each leaf point was landmarked in precisely this order (Figure 2):

1. Petiolar junction
2. Tip of leaf opposite petiolar junction (tip of midvein)

3. Left upper lateral sinus
4. Right upper lateral sinus
5. Left distal lobe tip
6. Right distal lobe tip
7. Left lower lateral sinus
8. Right lower lateral sinus
9. Left proximal lobe tip
10. Right proximal lobe tip
11. Left petiolar vein tip
12. Right petiolar vein tip
13. First major branching point of midvein
14. First major branching point of left distal vein
15. First major branching point of right distal vein
16. First major branching point of left proximal vein
17. First major branching point of right proximal vein

The landmark points were always plotted in the exact sequence as listed above for continuity of the process, ensuring all landmarks in a sequence represented the same morphological position on each leaf for linear comparison of measurements between points in the principal component analysis. After points were plotted on each leaf of an image, the landmarks for each image were then converted into (X, Y) values and saved in ImageJ using the "measure" function. Each leaf image has 17 landmarked points for conversion or, in the case of images containing multiple leaves, a multiple of 17

landmarked points. Each additional image's landmarks were added successively to the most current measured data file. The ImageJ measurements table was saved as a comma-delimited (CSV) file until all landmarking was complete and ready for statistical analysis.

Additionally, all landmark data was verified for correct order of landmark placement by running a graphical leaf check script in R using the ggplot2 package, and ggplot() and ggsave() functions. The output was one graph per set of 17 landmarked data points wherein the landmark points were sequentially connected by colored lines to produce a crude line drawing of each leaf. If any points were plotted out-of-order, the graph produced would appear as an abstract scribble of lines, indicating the order of landmarks should be corrected for that leaf image. Upon correction of any improperly ordered landmarks, the graphical leaf check script was ran a second time to verify the mistake had been corrected.

Genotyping

A SNP library was constructed for this population using methods developed by Cornell University Genotyping-By-Sequencing Project. DNA samples from grape leaves of the mapping population, barcode adapters, and adapter pairs were plated on 96-well plates and dried. Samples were then digested with a selected RE and double-stranded oligonucleotide adapters were ligated to target DNA fragment ends using T4 ligase. Heat shock was used to inactivate the ligase, after which a sample of each aliquot was pooled and applied to a size exclusion column to remove unreacted adapters. Primers with binding sites on ligated adapters were added and PCR amplification was then performed.

PCR products were cleaned up and fragment sizes of the library were checked on a BioRad Experion DNA Analyzer (Elshire et al., 2011).

Statistical Analyses

Landmark data from the raw images processed in ImageJ were used to perform a General Procrustes Analysis (GPA) for general shape analysis based on the venation pattern, sinuses, and lobes (Chitwood et al., 2014). The GPA produced the shape comparison results in the form of diagrams of the mean shape and of shape variation two standard deviations from the mean representative of each Principal Component (PC) identified. The Principal Component Analysis (PCA) results were produced in the form of Principal Component scores that were retrieved in R using the "shapes" package, and procGPA() and shapepca() functions (Dryden, 2013; R Core Team, 2013; Chitwood et al., 2014). Principal Component Scores, ranked by percent variation explained, were produced from the Principal Component Analysis (PCA). Data was then sorted and filtered based on various parameters to find the most meaningful data for analysis. I reduced the replicate PC scores per genotype (four replicates per in 2014, six replicates per in 2015) to one per genotype. The representative replicate scores for each genotype were sorted based on the Maximum PC1 value for each group of replicates per genotype, and the Maximum PC1 score was kept along with the corresponding PC2 and PC3 scores for each genotype. The left over replicate genotype PC scores were deleted from the list and the remaining scores compiled for the complete list of genotypes with their assigned VitisGen identifiers and PC scores 1-3 based on the Maximum PC1 score. This file was the final phenotyping data ready for association analysis in TASSEL.

Morphometric traits were then correlated with genotype data. Correlation analyses were conducted in Trait Analysis by aSSociation Evolution & Linkage (TASSEL) software. Genotype and phenotype data were uploaded into TASSEL. The genotype data was a variant call form (.vcf) file provided by Cornell University, and the phenotype file was a .txt file containing the VitisGen IDs and corresponding PC scores. The .vcf file was converted to a HapMap file (.hmp) to reduce file size, making it easier to load and analyze. Next, marker density was reduced to retain only sites that were scored in at least 10% of individuals. Minor SNP states were removed by converting them to missing data "N" to force only two types of segregation at a locus; this is suggested to reduce sequencing error. Parental genotypes were also removed. Then the phenotype and genotype files were combined into one file using the "intersection join" function in TASSEL. The combined phenotype and genotype file was used to generate a GLM analysis. Results of the various GLM analyses were produced in the form of QQ Plots and Manhattan plots.

From the "Association" folder in TASSEL, the "GLM_stats" file was exported as .txt files for post hoc analysis of p-values in R. Type I error rate was estimated in R under the following analyses: no correction, Bonferroni, and Benjamini-Hochberg (BH). Using the multtest() package in R, p-values were adjusted stringently using a Bonferroni correction to control for the family-wise error rate and minimization of false-positive associations, as well as a BH correction to minimize false-positives while also minimizing false-negatives. A well-known criticism of association studies is based in statistical analysis of large data sets that are common to genomics work. It is widely accepted nowadays that 'no correction' of p-values in association studies is unacceptable

due to the alarmingly high rate at which false-positive associations are identified. This is because the more attributes you attempt to compare and incorporate into tests, the greater the likelihood of observing a correlated difference of at least one attribute. Therefore, in this case of working with > 43,000 SNPs, with which I am attempting to correlate a continuously varying trait, the chances of identifying spurious associations is very high and cannot be ignored. The problem of multiple testing can be written as:

$$Pr(\text{at least one significant result}) = 1 - Pr(\text{no significant results})^{k}$$

Where 'k' represents the number of attributes being tested. For this study, I set the Type I error rate at the generally accepted $\alpha=0.05$, to produce the following equation:

$$Pr(\text{at least one significant result}) = 1 - (1 - 0.05)^{43,000}$$

When calculated, for all intents and purposes, this shows us the *probability of "at least one significant result"* to be equal to 1, i.e., a 100% probability of false-positive associations due to the size of our data set and the number of independent tests that must be run, if no correction is applied.

I then estimated the Type I error rate when using a Bonferroni correction. The Bonferroni correction is a stringent analysis that reduces false-positive associations to an extreme degree by dividing the standard Type I error rate ($\alpha=0.05$) by the number of attributes.

$$Pr(\text{at least one significant result}) = 0.05/43,000$$

The Bonferroni correction brings the estimated Type I error rate appropriately down to $\alpha\approx0.049442$, or an approximate 5% false-positive association rate. However, the Bonferroni correction brings the Type II error (β) rate (the rate of false-negatives) up a

great deal to $\beta \approx 0.964419$. Due to such a high Type II error rate estimation, I proceeded with a BH procedure.

The BH procedure is often preferred to the Bonferroni correction in the case of large data set analysis as is so common in genomics. A BH analysis corrects p-values to minimize false-positive associations, while also minimizing false-negatives. It provides a correction of p-values without being too stringent, resulting in an appropriately small Type I error rate without producing a prohibitively large Type II error rate. The BH procedure sorts and orders raw p-values, "*k*", from smallest to largest, with the smallest being assigned the rank of "1", the next smallest "2", and so on until the largest p-value is assigned the largest rank of "N" representing the total number of p-values in a given data set. Each ranked p-value is multiplied by "N" and then divided by its assigned rank to produce the corrected BH p-value, as follows:

*raw p-value * N / rank = BH corrected p-value*

RESULTS

Graphical Leaf Check & Principal Component Analysis

Plotted landmarks were verified for correct order of placement by running a graphical leaf check script in R using the ggplot2 package, and ggplot() and ggsave() functions. The output was one graph per set of 17 landmarked data points wherein the landmark points were sequentially connected by colored lines to produce a crude line drawing of each leaf (Figure 3). In the 2014 population, one leaf (NxCS89) was incorrectly landmarked and identified for correction. In the 2015 population, 13 leaves (N13, CSxN 27 and 18, and NxCS 02, 37, 38, 39, 55, 123, 124, 143, 151, and 154) were incorrectly landmarked and identified for correction. Upon correction of the improperly landmarked leaves, the graphical leaf check script was run a second time to verify the mistakes had been corrected (Figure 3).

Landmark data from the raw images processed in ImageJ were used to perform a General Procrustes Analysis (GPA) for general shape analysis. The results of the GPA are shown in diagrams of the mean shape and corresponding shape variation two standard deviations from the mean representative of each Principle Component (PC) identified (Figure 4). In the 2014 data, together, the first three PCs explain 57.3% of the variation of shape in our population; PC1 represents 30.4%, PC2 accounts for 15.3%, and PC3 represents 11.6% (Figure 4). In the 2015 data, the first three PC's explain 62.4% of the total variation in our population; PC1 represents 32.5%, PC2 represents 18.8%, and PC3 represents 11.1% (Figure 5).

Filtered Genotypes & Association Analyses

Genotype calls were received in the form of a .vcf file containing a SNP library for our NxCS F_1 population, organized by their assigned VitisGen identifiers that incorporated the field plant number for ease of cross-referencing, from Cornell University as part of the collaborative Cornell-based nationwide VitisGen1 project. The original .vcf file contained 190 taxa, or grapevine genotypes, and 43,971 SNPs. After removal of parental genotypes, 182 genotypes remained. Marker density was reduced by removal of likely duplicates, bringing the total number of SNPs included in the analysis to 40,724. Genotypes were further reduced to be included in analysis due to plants that died, were determined missing, and those for which we lacked genotyping data (n=22, total); these genotypes were identified for each year and all were removed from both years' analysis for continuity (Table 1). Phenotyping data, which corresponded with the removed genotypes, were removed from the PC file as well prior to analysis in TASSEL. The resulting number of genotypes and SNPs included in the association analysis for both years were 160 and 40,724, respectively.

GLM analyses were performed independently on both years' data. In 2014, when phenotyping data was sorted for Maximum PC1 score, analysis indicated ambiguous results for PCs 1 and 3 (Figure 6, Figure 7-top, and Figure 7-bottom). However, significant results were observed for PC2 on Linkage Group (LG) 17 (LOD=6.75) (Figure 7-center). As well, LGs 1 and 8 showed uniform clustering of SNPs with relatively less significant LOD scores of 4.5 and 4.75, respectively (Figure 7-center). In addition to analyzing these results in TASSEL, a correlation analysis was performed in R/qtl in the form of composite interval mapping (CIM) (Figure 8). The results produced

in R/qtl confirmed the GLM analysis results from TASSEL. This analysis allowed for identification of species/parental source associations. Significance was indicated on peak 36, which is representative of LG17 on this output (Figure 8). Therefore, LG 17, in Cabernet Sauvignon, was indicated above the significance threshold (LOD=3.1) at LOD=~4.5. Though neither broke the significance threshold (LOD=~3.1), LG 1 (from Cabernet Sauvignon, peak 20) and LG8 (from Norton, peak 8) were the next highest observed; both peaks reach approximately LOD=~2.8 (Figure 8). When data was sorted for Low PC1 score, only PC2 was marginally significant (Figure 9 and Figure 10-center); PC1 and PC3 displayed no significance (Figure 10-top and 10-bottom).

2014 phenotyping data were also sorted for Maximum PC2 score and the Average PC score for all PCs. When sorted for Maximum PC2 score, the results for PCs 1 and 3 were relatively ambiguous (Figure 11 and Figure 12-top and 12-bottom). Yet again, the results for PC2 were interesting; LG8 was indicated with significance this time at LOD=6.25, and LG17 peaked at LOD=4.5 with very nice grouping (Figure 12-center). When data were sorted for Low PC2 score, as observed in the Maximum PC2 based results, LG8 was indicated significantly (LOD=6.3) (Figure 13 and Figure 14-center). No significant association was observed with PC1 or PC3 (Figure 14-top and 14-bottom). When data were sorted for the average score of all three PCs, PC2 showed significance on LG8 (LOD=5.75) and LG17 (LOD=5.9); PC1 and PC3 remained insignificant (Figure 15, Figure 16-top, 16-center, and 16-bottom).

For 2015, the same sorting of PC scores and analyses were repeated. When 2015 data were sorted for Maximum PC1 score, the GLM analysis results were again ambiguous for PCs 1 and 3 (Figure 17, Figure 18-top and 18-bottom). The results for

PC2 were significant at LOD=5 on LG1 (Figure 17 and Figure 18-center). The results based on sorting for Low PC1 showed some significance on PC1 at LOD=6.0 (Figure 19 and Figure 20-top). PC2 and PC3 were ambiguous this time (Figure 20-center and 20-bottom). When the data were sorted for Maximum PC2 score, PCs 1 and 3 were ambiguous but significance was indicated for PC2 on LG1 at LOD=~5.9 (Figure 21 and Figure 22). When sorted for Low PC2 score, the results for PCs 2 and 3 were not significant but PC1 was significant for LG1 at LOD=5.75 (Figure 23 and Figure 24). No significant associations were identified in 2015 when data were sorted for Average All PCs (Figure 25 and Figure 26).

Post-hoc Statistical Analysis of P-Values

Type I and Type II error rates were estimated in R based on the following different analyses and corrections: no correction, Bonferroni correction, and a BH procedure. Without correction, we estimated a Type I error rate of 0.05, which is a good target, however, the corresponding Type II error rate is estimated at 0.087209. This is deceptively small. When put in context of the size of the data set, with that large of a Type II error, around 4,000 statistically significant SNPs are potentially missing. The Bonferroni correction brought the estimated Type I error rate down appropriately to $\alpha \approx 0.049442$, or an approximate 5% false-positive association rate. This is indisputably a far more acceptable Type I error rate for the purposes of this study. It should be noted, however, that it brought the Type II error (β) rate up a great deal to $\beta \approx 0.964419$. In consideration of the too high Type II error rate estimation, the Bonferroni correction is

probably too conservative. Better error estimations are observed with the BH analysis, which estimated the Type I error rate, $\alpha \approx 0.003395$, and Type II error rate, $\beta \approx 0.403256$.

Based on the resulting estimated Type I and Type II error rates discussed above, performing a Bonferroni correction and a Benjamini-Hochberg (BH) correction allowed identification of the most probable real SNP-trait associations for our population. The Bonferroni correction identified 66 total statistically significant SNPs in 2014, and a total of zero statistically significant SNPs in 2015 (Tables 2, 3, 4, 5 & 6). In contrast, the BH correction identified 92 total SNPs in 2014 and (if we adhere to a desired Type I error rate of $\alpha = 0.05$) zero total SNPs in 2015 as significant (Tables 7, 8, 9, & 10). In comparing the two post-hoc analyses of p-values, the BH analysis captured an additional 78 statistically significant SNPs in total that were not included in the more conservative Bonferroni analysis.

While the BH procedure gives an estimated Type II error rate of ~40% (that is still relatively high), the Type II error rate cannot be controlled, rather it can only be optimally minimized. Therefore, efforts are focused on controlling the Type I error probability at the desired $\alpha = 0.05$ level, while minimizing the volume of false-negatives (relative to the target probability for a Type I error).

Comparison of statistically significant p-values as identified by each of the post-hoc analyses for and between both years was also completed. Statistically significant p-values were indicated with each of the post-hoc analyses for the 2014 Maximum PC1 sort, SNPs relevant to PC2 were indicated in all three analyses with the BH analysis indicating nine more SNPs as significant than the Bonferroni, for a total of 13 statistically significant SNPs on LG17 and one from LG13 (Figure 7; Table 2). This supports the

results shown in the Manhattan plot. No SNPs had statistically significant p-values after Bonferroni correction or the BH procedure to support the association indicated from the Low PC1 sort data on the Manhattan plot (Figure 10; Table 3). A total of nine SNPs are statistically significant in association with PC2, based on the Maximum PC2 sort, after both the BH analysis that identified five more SNPs than the Bonferroni correction; seven SNPs are in LG8, one in LG16 and one in LG12 (Figure 12; Table 4). Analysis of the p-values from the Low PC2 sort indicated association again with PC2; the Bonferroni correction identified four statistically significant SNPs and the BH procedure identified 27 statistically significant SNPs (20 on LG8, one on LG17, one on LG11, and one on LG16) (Figure 14; Table 5). Association with PC2 was supported again in the averaged sort of all PC's with a total of 42 SNPs, largely in LGs 17 and 8 (only three SNPs appeared from different individual linkage groups – LG5, LG13, and LG16), showing statistical significance (Figure 16; Table 6). For this analysis, the Bonferroni correction was far more conservative and only identified one significant SNP, but the BH correction identified a total of 42 SNPs. In each iteration, and for both years, the statistically significant SNPs from the BH procedure are inclusive of the SNPs identified via the Bonferroni correction.

The 2015 data were less clear and indicated no statistical significance in the post-hoc analyses. In the 2015 results from the Maximum PC1 sort, the BH analysis only identified four SNPs that may be of some significance in association with PC2 on LG1, though all violate the desired 0.95 confidence level, while the Bonferroni correction resulted in much higher p-values still (Table 7). This contradicts the results observed based on raw p-value in the Manhattan plot (Figure 18) where we see SNPs on LG1

showing significant association with PC2 as indicated by an approximate LOD=5.9. Similarly high p-values were observed for the Low PC1 sorted data; the 10 lowest p-values from the BH analysis all violated our desired 0.95 confidence level (Figure 20; Table 8), and for the first time, (though statistically weak) an association with PC1 indicated on LG5 was identified. This is also what was observed in the Manhattan and QQ plots based on raw p-values for this analysis. Again, no statistical significance was identified in the post-hoc analyses for any p-values from either the Maximum PC2 or Low PC2 sorted data (Tables 9 & 10), yet both indicated some significance in the Manhattan and QQ plots based on raw p-values (Figures 21, 22, 23, & 24). For the Maximum PC2 sorted data, the QQ plot showed significant association with PC2, and this was supported with significance on LG1 in association with PC2 on the Manhattan plot (Figures 21 & 22). The Low PC2 data showed significant association with PC1 on the QQ plot, and this was supported in the PC1 Manhattan plot on LG1 (Figures 23 & 24). Finally, no significant SNPs were identified in the post-hoc analysis for the averaged sort of all PCs in 2015. Based on the raw p-values, the Manhattan and QQ plots indicated potentially significant SNPs associated with PC2 on LG1 but this was not supported in the more stringent post-hoc analyses with all adjusted p-values at 0.376346 and greater (Figures 25 & 26).

DISCUSSION

Plant morphology began with Goethe's perceptive observations on the expansion and contraction of leaf tissue during plant development (Boyle, 2015). Yet, more than 200 years later, the hypothesized operative purpose(s) of many leaf forms and structures remain unconfirmed. Morphology's interdisciplinary nature positions it advantageously to gain knowledge and insight from advancements in related fields. Technological progress in areas like genomics provides a framework for large-scale genetic characterization studies. As well, recent increased interest in phenotyping has drawn attention to the need for more precise methodology that can be applied to sampling large populations en masse. Together these have created a modern wheelhouse of opportunity for discovery in morphology. In turn, such discoveries are likely to mutually contribute information and data to the fields upon which morphology relies, and particularly so when interdisciplinary collaborators thoughtfully select an organism to maximize study impact. The choice to use an F_1 breeding population cross-bred from two Vitis species demonstrates the type of multifaceted consideration in organismal selection that should be the cornerstone of morphological studies in order to maximize the potential impact of results. This project rests firmly in the realm of descriptive science, making it both basic and necessary, at once. While immediate intentions have been, quite simply, to use genetics to better describe leaf shape without experimental manipulation, the outcome is a foundation from which to continue work that may potentially have effects in agriculture, plant physiology, ecology, evolution, viticulture, and so on. This study also highlights the complications surrounding reproducibility of results in association studies

and the importance of post-hoc statistical analysis to minimize the effects of confounding variables. At this time, focus will shift toward future work in the genetic characterization of leaf shape morphology that may expand on posited functional implications and the evolution of trait adaptation in Vitis.

Principal Component Analysis & Principal Component Scores

In reviewing the top three principal components identified in both 2014 and 2015, the mean shape of our population is broadly consistent. Anecdotally, it is observed that the mean shape from the PCA is representative of the typical F_1 progeny sampled in the population and is consistent between years. It's apparent that the SD of the shape captured by PC1, for both years, is stable with only a slightly smaller distance between the lower lateral sinus and tip of the proximal vein in 2015 compared to 2014 (Figures 4 & 5). PC1 also captures approximately 2% greater overall shape variation in the population in 2015 (32.5%) than in 2014 (30.4%) (Figures 4 & 5). In comparing the most extreme shape variations (±2 SD) identified by PC1, it seems that +2 SD in both years represents greater expansion in upper lateral sinus (ULS) and tips of the mid- and distal veins paired with compaction in the petiolar sinus, while -2 SD captures a restriction in the ULS region paired with expansion in the proximal veins and petiolar sinus (Figures 4 & 5). In 2014, +2 SD of PC1 represents such compaction of the petiole junction that total overlapping of the tips of the proximal branching veins is seen (Figure 4). Also, the interior-most pattern created by the major branching points and sinuses provide noteworthy variation. Referring to the shape represented by -2 SD PC1, it is noted that

the interior landmarking pattern makes a pentagon shape in contrast to the nearly isosceles trapezoid shape produced by the +2 SD PC1 shape (Figure 4).

Looking at PC2 for 2014 and 2015, again it is observed that the identified shape variation is very similar between years (Figures 4 & 5). It appears that PC2 likely represents shape variation (15.3% in 2014 and 18.8% in 2015) along species lines when comparing to the adult/mature leaf morphologies of both Norton and Cabernet Sauvignon; the shape -2 SD from the mean looks like Cabernet Sauvignon and the shape +2 SD from the mean looks most like Norton (Figures 4 & 5). Again, minor variations exist between years for PC2, mostly identifiable at -2 SD away from the mean shape, possibly indicating more shape variation stemming from Cabernet Sauvignon for 2015 than seen in 2014 (Figures 4 & 5).

Interestingly, PC3 observed in 2015 is reversed to that which was captured in 2014. The variation relative to PC3 is most strongly influenced by the midvein, specifically beginning at the first major branching point and running the remaining length of the leaf to the tip of the midvein (Figures 4 & 5). It is between these two points where a distinct curve is seen in the midvein that highlights an expansion of the leaf side opposite the direction of the vein curve, and compaction of the leaf on the same side as the direction of the curve (Figures 4 & 5). In spite of this distinct (and consistent) shape variation, PC3 is only estimated to account for 11.6% and 11.1% for 2014 and 2015, respectively (Figures 4 & 5). As well, no significant association with PC3 was identified in any of the analyses for either year.

TASSEL Results & P-Value Analysis

Overall, the TASSEL results for year 2014 present an overarching theme of statistically significant SNPs from LGs 17 and 8 in association with PC2. These results were supported in the analysis ran in R/qtl which identified statistical significance in LG17, in addition to statistical support for LGs 1 and 8 that was just short of the significance threshold (Figure 8). As mentioned above, PC2 appears to identify shape variation along species lines. This corresponds with the results in our R/qtl analysis, which differentiates LG effects between species. Recall, from Figure 8, that LG 8 was attributed to Norton, and LGs 1 and 17 were attributed to Cabernet Sauvignon. Referring back to the shape variation of PC2 as shown in Figure 4, strong support is observed for an association between genotype (LGs 1, 8, and 17) and phenotype (PC2) within the NxCS F_1 progeny (Figures 4, 6, 7, 8, 11, 12, 13, 14, 15, & 16). As well, the statistical significance of identified SNPs from both LGs 8 and 17 withstand the rigorous post-hoc p-value analyses performed (Tables 2, 4, 5, & 6).

As acknowledged above, the 2015 data lacks statistically significant support for an association between genotyping and phenotyping data. However, the potential biological significance of the data should not be dismissed. In the case of the Maximum PC1 sorted data for 2015, four SNPs were observed with the smallest adjusted p-values to be approximately 0.07 (Table 7). While an outright claim cannot be made for strong statistical support, it should be acknowledged that these four SNPs are all identified on LG1 in association with PC2 (Figure 18-center; Table 7). In context, these data are in-line with what was observed in 2014 in that PC2 appears to consistently represent shape variation along species lines for both years (Figure 4) and, although LG1 was not

indicated with statistical significance in either year (as defined by adjusted p-values), it is noted that it is indicated with nearly the same LOD score as LG8 in the R/qtl analysis from 2014 (Figures 7, 8, & 18-center).

Based on the combined association analyses and adjusted p-values from 2014 data, and in consideration of the possible biological significance of our 2015 results, it appears this study identified LGs 8 and 17 to contain SNPs associated with leaf shape morphology in this breeding population, and, perhaps, some SNPs on LG1 play a minor role. The extent to which any of the identified SNPs influence or control leaf shape morphology is, as-of-yet, unclear.

Limitations of Association Analysis

The potential shortcomings of association studies are widely known though rarely tested and, therefore, frequently unconfirmed (Casadevall, 2010). Reproducibility (prediction of results) and replicatibility (identical results) are, perhaps, equally elusive (Casadevall, 2010). One possible source of unreliability in association studies, as in many kinds of studies, may stem from a variation in conditions or departure from an original protocol. For instance, in this study, in 2014 four leaves per vine were collected and in 2015 that number increased to six leaves per vine. It is generally accepted that a larger sample size should enhance the robustness of measurements taken. However, by sampling an additional two leaves from an additional shoot per vine in 2015, poorer quality leaves from each vine may have sometimes been collected in order to obtain six leaves total from three shoots per vine. (Poorer quality may be considered those leaves with shape features that may have been influenced by disease or herbivory.) Shape

features that are impacted by insects and disease may provide enough influence on PC measurements and resulting scores to affect association with genotyping data. The original protocol outlining sample collection in 2014 was not followed exactly in the subsequent year, making it difficult to linearly compare and contrast results between years. Failure to replicate 2014's results may indicate a small genetic affect influencing leaf shape, and it provides further support for leaf shape and development being highly environmentally labile.

Another consideration is that continuously varying, i.e., complex, traits do not manifest in simplistic predictor-response variable relationships (Casadevall, 2010). It is generally accepted that numerous aspects of plant form and function respond to environmental factors but that the involved confounding variables are frequently imprecisely defined (Ioannidis et al., 2001; Myles et al., 2009; Casadevall, 2010; Vilhajalmsson & Nordborg, 2013). For the time being, the best counterbalance to those potentially unknown confounding variables lies in statistical analysis. This recognition should greatly emphasize the importance of the thoughtful selection of statistical tests used in association studies and the necessary restraint and skepticism required on the part of the researchers for interpretation of initially significant results.

Future Directions

The opportunity for future work in grape leaf morphology is diverse. The underlying genetic architecture of any complex trait is difficult to predict and, frequently, studies focused on one population do not yield all possible loci of varying degrees of effect (McCarthy et al., 2008). Therefore, inclusion of populations with different

ancestry, i.e., a more diverse germplasm, may lead to discovery of additional loci. As well, results from association studies are biased toward additive effects and, therefore, they often overlook such effects as gene x gene (GxG) and gene x environment (GxE) (McCarthy et al., 2008). Evaluation of the same genotypes in different environments would allow for observation of possible GxE effects, i.e., the plasticity potential of current genotypes. Identifying the population's capacity for plasticity may offer further explanation for the variability of identified association between years. In order to provide appropriate context for plasticity, location(s) of new environment should be considered carefully in order to control for degree of environmental exposure to which plants are subjected, and records should be kept for all potentially conceivable confounding environmental variables. Another prospective approach for understanding grape-environmental interactions would be a leaf margin analysis (LMA). This could be conducted using grape to look for correlations between leaf physiognomic variables, like leaf dissection, and environmental factors, like temperature (Royer et al., 2005).

REFERENCES

Abramoff, M.D., Magalhaes, P.J., and Ram, S.J. (2004). Image processing with ImageJ. Bio. Intl. **11**: 36-42

Ambers, C.P. (2013). A historical hypothesis on the origin of the Norton grape. Journal of Wine Research. **24**(2): 85-95.

Ambers, R.K.R., and Ambers, C.P. (2004). Dr. Daniel Norborne Norton and the origin of the norton grape. American Wine Society Journal. **Fall 2004**: 77-86.

Austin, C.N., Grove, G.G., Meyers, J.M., and Wilcox, W.F. (2011). Powdery mildew severity as a function of canopy density: associated impacts on sunlight penetration and spray coverage. American Journal of Enology and Viticulture. **62**: 23-31.

Austin, C.N., Wilcox, W.F. (2012). Effects of sunlight exposure on grapevine powdery mildew development. Phytopathology. **102**: 857-866.

Avery, J.D., Byers, P.L., Howard, S.F., Kaps, M.L., Kovacs, L.G., Moore Jr., J.F., Odneal, M.B., Qiu, W., Saenz, J.L., Teghtmeyer, S.R., Townsend, H.G., and Waldstein, D.E. (2003). In P. Byers et al., (Ed.), Growing grapes in Missouri. State Fruit Experiment Station, Missouri State University, Department of Fruit Science.

Bailey, I.W., and Sinnott, E.W. (1915). Botanical index of Cretaceous and Tertiary climates. Science. **41**: 831-834. doi: 10.1126/science.41.1066.831

Bailey, I.W., and Sinnott, E.W. (1916). The climatic distribution of certain types of angiosperm leaves. American Journal of Botany. **3**: 24-39. doi: 10.2307/2435109

Boyle, N. (2015). Johanne Wolfgang von Goethe. Retrieved January 1, 2016 from Encyclopedia Brittanica online: http://www.britannica.com/biography/Johann-Wolfgang-von-Goethe

Brodribb, T.J., Field, T.S. (2010). Leaf hydraulic evolution led a surge in leaf photosynthetic capacity during early angiosperm diversification. Ecology Letters. 13: 175-183. doi: 10.1111/j.1461-0248.2009.01410.x

Casadevall, A. (2010). Reproducible science. Infection And Immunity. **78**(12):4972-4975. doi: 10.1128/IAI.00908-10

Chitwood, D.H., Headland, L.R., Filiault, D.L., Kumar, R., Jimenez-Gomez, J.M., Schrager, A.V., Park, D.S., Peng, J., Sinha, N.R., and Maloof, J.N. (2012).

Native environment modulates leaf size and response to simulated foliar shade across wild tomato species. PLoS ONE **7**: e29570

Chitwood, D.H., Ranjan, A., Martinez, C., Headland, L., Thiem, T., Kumar, R., Covington, M.F., Hatcher, T., Naylor, D.T., Zimmerman, S., Downs, N., Raymundo, N., Buckler, E.S., Maloof, J.N., Aradhya, M., Prins, B., Li, L., Myles, S., and Sinha, N. (2014). A modern ampelography: A genetic basis for leaf shape and venation patterning in grape. Plant Physiology. **164**: 259-272.

Chitwood, D.H., Klein, L.L., O'Hanlon, R., Chacko, S., Greg, M., Kitchen, C., Miller, A., and Londo, J. (2015). Latent developmental and evolutionary shapes embedded with the grapevine leaf. New Phytologist. doi:10.1111/nph.13754

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., and Pang, E.C.K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica **142**: 169-196.

Costa, M.M.R., Yang, S., Critchley, J., Feng, X., Wilson, Y., Langlade, N., Copsey, L., and Hudson, A. (2012). The genetic basis for natural variation in heteroblasty in *Antirrhinum*. New Phytologist. **196**: 1251-1259.

Dryden, I.L. (2013). Shapes: Statistical shape analysis. R package version 1.1-8. http://CRAN.R-project.org/package=shapes Retrieved October 14, 2014

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE **6**: 1-10.

Galet, P. (1952) A Practical Ampelography: Grapevine Identification. Translated by L.T. Morton (Ithaca, NY: Cornell University Press).

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE **9**(2): e90346. doi: 10.1371/journal.pone.0090346

Hyde, D. (2009). Linkage and Mapping in Eukaryotes. In P.E. Reidy (Ed.), *Introduction to Genetic Principles* (pp 122-165). New York, NY: McGraw-Hill.

Ioannidis, J.P.A., Ntzani, E.E., Trikalinos, T.A., and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. Nature Genetics **29**: 306-309.

Jansen, M.A.C., Gaba, V., Greenberg, B. (1998). Higher plants and UV-B radiation: Balancing damage, repair, and acclimition. Trends in Plant Science. **4**: 131-135.

Jensen, A.K. (2014). 19th century European philosophers: Johann Wolfgang von Goethe (1749-1832). Internet Encyclopedia of Philosophy. City University of New York / Lehman College. http://www.iep.utm.edu/goethe/ Retrieved: September 19, 2014

Kaplan, D.R. (2001). The science of plant morphology: Definition, history, and role in modern biology. American Journal of Botany. **88**(10): 1711-1741.

Kolbe, C.A., Kaser, M.A., Kopecky, J., Zotz, G., Riederer, M., and Pfundel, E.E. (2001). Effects of natural intensities of visible and ultraviolet radiation on epidermal ultraviolet screening and photosynthesis in grape leaves. Plant Physiology. **127**: 863-875.

Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. Nature Reviews Genetics. **7**: 385-394.

Lijavetzky, D., Carbonell-Bejerano, P., Grimplet, J., Bravo, G., Flores, P., Fenoll, J., Hellin, P. (2012). Berry flesh and skin ripening features in Vitis vinifera as assessed by transcriptional profiling. PLoS One **7**(6):e39547. doi: 10.1371/journal/pone.0039547

Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., and Matsudaira, P. (2008). Molecular Cell Biology. 6th ed. (New York: W. H. Freeman and Company).

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A., Hirschhorn, J.N. (2008) Genome-wide association studies for complex traits: Consenus, uncertainty and challenges. Nature Reviews Genetics. **9**: 356-369. doi: 10.1038/nrg2344

Author Not Given. (n.d.). "Norton." MissouriWines.org, Norton | MO Wine. Retrieved on January 1, 2016 from http://missouriwine.org/wines/varietals/Norton

Moutinho-Pereira, J., Goncalves, B., Bacelar, E., Cunha, J.B., Coutinho, J., and Correia, C.M. (2009). Effects of elevated CO2 on grapevine (*Vitis vinifera L.*): physiological and yield attributes. Vitis. **48**: 159-165.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., and Buckler, E.S. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. The Plant Cell. **21**: 2194-2202.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J-M., Ware, D., Bustamante, C.D., and Buckler, E.S. (2011). Genetic structure and domestication history of the grape. Proceedings of the National Academy of Sciences. **108**(9): 3530-3535

Nicotra, A.B., Leigh, A., Boyce, C.K., Jones, C.S., Niklas, K.J., Royer, D.L., and Tsukaya, H. (2011). The evolution and functional significance of leaf shape in the angiosperms. Functional Plant Biology. **38**: 535-552.

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. Nature **456**: 720-723.

Parrish, J.T. (1998). Interpreting pre-Quaternary climate from the geologic record. (New York, NY: Columbia University Press).

Pires, N.D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. The Royal Society of Biology. **367**: 508-518.

Puckette, M. (2012). Guide to Cabernet Sauvignon red wine. WineFolly.com. Retrieved on January 1, 2016 from http://winefolly.com/review/guide-to-cabernet-sauvignon-red-wine/

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/ Retrieved October 14, 2014

Roberts, P.L. (1999) Norton, America's true grape: Whence, and whither? MissouriWineCountry.com, Missouri Wine Articles. Retrieved on December 30, 2015 from www.missouriwinecountry.com/articles/wines/norton-true.php

Royer, D.L., Wilf, P., Janesko, D.A., Kowalski, E.A., and Dilcher, D.L. (2005). Correlations of climate and plant ecology to leaf size and shape: potential proxies for the fossil record. American Journal of Botany. **92**(7): 1141-1151.

Sack, L., Scoffoni, C. (2013). Leaf venation: Structure, function, development, evolution, ecology and applications in the past, present and future. New Phytologist. **198**: 983-1000.

Schultz, H.R., Lohnertz, O., Bettner, W., Balo, B., Linsenmeier, A., Jahnisch, A., Muller, M., Gaubatz, B., and Varadi, G. (1998). Is grape composition affected by current levels of UV-B radiation? Vitis. **37**: 191-192.

Seward, A.C. (1892). Fossil plants as tests of climate. (London, UK: C.J. Clay).

Sleper, D.A. and Poehlman, J.M. (2006) Breeding Field Crops. 5th ed. (Ames, IA: Blackwell Professional).

Smiley, L. (2008). Cynthiana/Norton. Iowa State University, Horticulture.

Spotts, R. (1977). Effect of leaf wetness duration and temperature on the inefectivity of Guignardia bidwellii on grape leaves. Phytopathology. **67**: 1378-1381.

Sweet, N. (2008). Cabernet Sauvignon at FPS. Foundation Plant Services Grape Program Newsletter. **October 2008**: 16-32.

Teeter, A. (2014). The fascinating history behind America's oldest native grape: Norton. Vinepair.com. Retrieved on January 1, 2016 from http://vinepair.com/wine-blog/fascinating-history-behind-americas-oldest-native-grape-norton/

Terral, J-F., Tabard, E., Bouby, L., Ivorra, S., Pastor, T., Figueiral, I., Picq, S., Chevance, J-B., Jung, C., Fabre, L., Tardy, C., Compan, M., Bacilieri, R., Lacombe, T., and This, P. (2009). Evolution and history of grapevine (*Vitis vinifera*) under domestication: New morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. Annals of Botany. **105**: 443-455.

Vilhjalmsson, B.J., and Nordorg, M. (2013). The nature of confounding in genome-wide association studies. Nature Reviews Genetics. **14**: 1-2.

Waggoner, B. (2006). Carl Linnaeus (1707-1778). University of California Museum of Paleontology at Berkley. http://www.ucmp.berkeley.edu/history/linnaeus.html. Retrieved September 3, 2014.

Weaver, R.F. (2012). Molecular Biology. 5th ed. (New York: McGraw-Hill).

Wei, Z., Zhang, G., Du, Q., Zhang, J., and Zhang, D. (2014). Association mapping for morphological and physiological traits in *Populus simonii*. BMC Genetics. **15**: 1-8.

Wilf, P. (1997). When are leaves good thermometers? A new case for leaf margin analysis. Paleobiology. **23**: 373-390.

Wilf, P., Wing, S.L., Greenwood, D.R., Greenwood, C.L. (1998). Using fossil leaves as paleoprecipitation indicators: an Eocene example. Geology. **26**: 203-206.

Willis, K.J., and McElwain, J.C. (2002). The evolution of plants. (Oxford, UK: Oxford University Press.)

Wolfe, J.A. (1979). Temperature parameters of humid to mesic forests of Eastern Asia and relation to forests of other regions of the Northern Hemisphere and Australasia. U.S. Geological Survey Professional Paper. **1106**: 1-37.

Wolpert, J. (2006) Cabernet Sauvignon. Wine Grape Varieties in California. Retrieved on December 27, 2015 from http://iv.ucdavis.edu/files/24325.pdf

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C., Midgley, J.J., Navas, M-L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J., and Villar, R. (2004). The worldwide leaf economics spectrum. Nature. **428**: 821-827. doi: 10.1038/nature02403

Zwieniecki MA, Melcher PJ, Boyce CK, Sack L, Holbrook NM (2002) Hydraulic architecture of leaf venation in Laurus nobilis L. Plant, Cell & Environment **25**: 1445–1450.

Table 1. Plants (n=22) excluded from analysis based on missing genotype and/or phenotype data.

Year	Variety	Plant ID Number
2005	Cabernet Sauvignon x Norton	015
2005	Cabernet Sauvignon x Norton	016
2005	Cabernet Sauvignon x Norton	017*
2005	Norton x Cabernet Sauvignon	008*
2005	Norton x Cabernet Sauvignon	014*
2005	Norton x Cabernet Sauvignon	021*
2005	Norton x Cabernet Sauvignon	029*
2005	Norton x Cabernet Sauvignon	034*
2005	Norton x Cabernet Sauvignon	045
2005	Norton x Cabernet Sauvignon	061
2005	Norton x Cabernet Sauvignon	069
2005	Norton x Cabernet Sauvignon	070
2005	Norton x Cabernet Sauvignon	071
2011	Norton x Cabernet Sauvignon	121
2011	Norton x Cabernet Sauvignon	137*
2011	Norton x Cabernet Sauvignon	142*
2011	Norton x Cabernet Sauvignon	143
2011	Norton x Cabernet Sauvignon	155
2011	Norton x Cabernet Sauvignon	164
2011	Norton x Cabernet Sauvignon	171*
2011	Norton x Cabernet Sauvignon	176*
2011	Norton x Cabernet Sauvignon	178

*Indicates exclusion based on plant data missing in 2015.

Table 2. 2014 Maximum PC1 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	Index	Marker	Raw p-value	Bonferroni	BH
1	75450	S17_7176322	2.23E-07	2.72E-02	1.14E-02
2	75451	S17_7176341	2.23E-07	2.72E-02	1.14E-02
3	74890	S17_3696433	2.80E-07	3.42E-02	1.14E-02
4	75012	S17_4735635	5.76E-07	7.04E-02	1.53E-02
5	66963	S13_6777085	6.26E-07	7.65E-02	1.53E-02
6	74944	S17_4085532	9.00E-07	1.10E-01	1.65E-02
7	74838	S17_3380545	9.46E-07	1.16E-01	1.65E-02
8	74911	S17_3893921	1.40E-06	1.71E-01	2.14E-02
9	75395	S17_6936627	2.05E-06	2.50E-01	2.78E-02
10	75061	S17_4983586	2.94E-06	3.59E-01	3.34E-02
11	74754	S17_2900500	3.01E-06	3.68E-01	3.34E-02
12	74765	S17_2953957	3.93E-06	4.80E-01	4.00E-02
13	74949	S17_4116736	4.65E-06	5.68E-01	4.37E-02
14	74851	S17_3456248	6.23E-06	7.61E-01	5.44E-02

Table 3. 2014 Low PC1 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	Index	Marker	Raw p-value	Bonferroni	BH
1	56979	S8_10702966	3.73E-06	4.56E-01	4.56E-01
2	14119	S7_8428333	7.87E-05	1.00E+00	6.69E-01
3	15114	S7_23884675	5.63E-05	1.00E+00	6.69E-01
4	15115	S7_23884687	5.63E-05	1.00E+00	6.69E-01
5	36265	S18_6323650	1.04E-04	1.00E+00	6.69E-01
6	36266	S18_6323651	1.04E-04	1.00E+00	6.69E-01
7	36267	S18_6323654	1.04E-04	1.00E+00	6.69E-01
8	36274	S18_6325557	9.89E-05	1.00E+00	6.69E-01
9	40126	S19_10608663	6.76E-05	1.00E+00	6.69E-01
10	42989	S1_16747362	3.28E-05	1.00E+00	6.69E-01

Table 4. 2014 Maximum PC2 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	Index	Marker	Raw p-value	Bonferroni	BH
1	57358	S8_13698124	6.00E-07	7.33E-02	2.10E-02
2	57359	S8_13698153	6.00E-07	7.33E-02	2.10E-02
3	57360	S8_13698178	6.00E-07	7.33E-02	2.10E-02
4	73510	S16_15835946	6.88E-07	8.41E-02	2.10E-02
5	57773	S8_16809245	2.32E-06	2.83E-01	4.95E-02
6	57789	S8_16833307	2.43E-06	2.97E-01	4.95E-02
7	58430	S8_21562808	4.11E-06	5.02E-01	5.84E-02
8	58431	S8_21562812	4.11E-06	5.02E-01	5.84E-02
9	105234	S12_4068492	4.30E-06	5.25E-01	5.84E-02

Table 5. 2014 Low PC2 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	Index	Marker	Raw p-value	Bonferroni	BH
1	57773	S8_16809245	6.32E-07	7.72E-02	2.39E-02
2	57358	S8_13698124	7.84E-07	9.58E-02	2.39E-02
3	57359	S8_13698153	7.84E-07	9.58E-02	2.39E-02
4	57360	S8_13698178	7.84E-07	9.58E-02	2.39E-02
5	73510	S16_15835946	1.87E-06	2.28E-01	3.36E-02
6	58430	S8_21562808	2.17E-06	2.65E-01	3.36E-02
7	58431	S8_21562812	2.17E-06	2.65E-01	3.36E-02
8	58361	S8_21139733	2.20E-06	2.69E-01	3.36E-02
9	57771	S8_16760411	2.69E-06	3.29E-01	3.53E-02
10	21701	S11_1797543	3.04E-06	3.71E-01	3.53E-02
11	57789	S8_16833307	3.18E-06	3.89E-01	3.53E-02
12	58294	S8_20475153	3.48E-06	4.25E-01	3.54E-02
13	56979	S8_10702966	4.49E-06	5.49E-01	4.22E-02
14	57996	S8_18640543	5.67E-06	6.93E-01	4.34E-02
15	56854	S8_9041374	8.28E-06	1.00E+00	4.34E-02
16	56855	S8_9041385	8.28E-06	1.00E+00	4.34E-02
17	56856	S8_9041386	8.28E-06	1.00E+00	4.34E-02
18	56857	S8_9041387	8.28E-06	1.00E+00	4.34E-02
19	56858	S8_9041392	8.28E-06	1.00E+00	4.34E-02
20	56859	S8_9041398	8.28E-06	1.00E+00	4.34E-02
21	56860	S8_9041402	8.28E-06	1.00E+00	4.34E-02
22	56861	S8_9041403	8.28E-06	1.00E+00	4.34E-02
23	56862	S8_9041404	8.28E-06	1.00E+00	4.34E-02
24	58001	S8_18737788	8.52E-06	1.00E+00	4.34E-02
25	58096	S8_19446588	9.43E-06	1.00E+00	4.61E-02
26	74944	S17_4085532	9.85E-06	1.00E+00	4.63E-02
27	57767	S8_16747787	1.17E-05	1.00E+00	5.29E-02

Table 6. 2014 average of all PCs statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	index	Marker	Raw p-value	Bonferroni	BH
1	73510	S16_15835946	5.11E-08	6.24E-03	6.24E-03
2	75450	S17_7176322	1.29E-06	1.58E-01	1.82E-02
3	75451	S17_7176341	1.29E-06	1.58E-01	1.82E-02
4	74838	S17_3380545	1.39E-06	1.70E-01	1.82E-02
5	57789	S8_16833307	1.59E-06	1.94E-01	1.82E-02
6	74944	S17_4085532	1.92E-06	2.35E-01	1.82E-02
7	56854	S8_9041374	2.24E-06	2.74E-01	1.82E-02
8	56855	S8_9041385	2.24E-06	2.74E-01	1.82E-02
9	56856	S8_9041386	2.24E-06	2.74E-01	1.82E-02
10	56857	S8_9041387	2.24E-06	2.74E-01	1.82E-02
11	56858	S8_9041392	2.24E-06	2.74E-01	1.82E-02
12	56859	S8_9041398	2.24E-06	2.74E-01	1.82E-02
13	56860	S8_9041402	2.24E-06	2.74E-01	1.82E-02
14	56861	S8_9041403	2.24E-06	2.74E-01	1.82E-02
15	56862	S8_9041404	2.24E-06	2.74E-01	1.82E-02
16	75061	S17_4983586	2.42E-06	2.96E-01	1.85E-02
17	74911	S17_3893921	4.42E-06	5.40E-01	3.18E-02
18	57358	S8_13698124	5.92E-06	7.23E-01	3.40E-02
19	57359	S8_13698153	5.92E-06	7.23E-01	3.40E-02
20	57360	S8_13698178	5.92E-06	7.23E-01	3.40E-02
21	75170	S17_5812218	6.08E-06	7.43E-01	3.40E-02
22	75395	S17_6936627	7.07E-06	8.64E-01	3.40E-02
23	74949	S17_4116736	7.12E-06	8.70E-01	3.40E-02
24	74753	S17_2895107	7.33E-06	8.96E-01	3.40E-02
25	74765	S17_2953957	7.47E-06	9.13E-01	3.40E-02
26	75358	S17_6576048	7.54E-06	9.21E-01	3.40E-02
27	75131	S17_5530331	7.78E-06	9.50E-01	3.40E-02

Table 6 continued. 2014 average of all PCs statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	index	Marker	Raw p-value	Bonferroni	BH
28	75026	S17_4849625	7.79E-06	9.52E-01	3.40E-02
29	75012	S17_4735635	9.06E-06	1.00E+00	3.82E-02
30	57996	S8_18640543	1.16E-05	1.00E+00	4.57E-02
31	74860	S17_3530583	1.14E-05	1.00E+00	4.57E-02
32	75091	S17_5222677	1.33E-05	1.00E+00	5.08E-02
33	57771	S8_16760411	1.47E-05	1.00E+00	5.35E-02
34	75110	S17_5274871	1.49E-05	1.00E+00	5.35E-02
35	66963	S13_6777085	1.70E-05	1.00E+00	5.77E-02
36	74890	S17_3696433	1.67E-05	1.00E+00	5.77E-02
37	58242	S8_20189704	1.76E-05	1.00E+00	5.81E-02
38	75058	S17_4957347	1.82E-05	1.00E+00	5.85E-02
39	74531	S17_1336009	1.95E-05	1.00E+00	5.93E-02
40	74712	S17_2581596	1.99E-05	1.00E+00	5.93E-02
41	74754	S17_2900500	1.97E-05	1.00E+00	5.93E-02
42	51029	S5_20807066	2.05E-05	1.00E+00	5.96E-02

Table 7. 2015 maximum PC1 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	index	Marker	Raw p-value	Bonferroni	BH
1	46873	S1_21085379	1.37E-06	1.81E-01	7.55E-02
2	46874	S1_21085394	1.37E-06	1.81E-01	7.55E-02
3	46401	S1_16155521	2.23E-06	2.94E-01	7.55E-02
4	46296	S1_13065418	2.29E-06	3.02E-01	7.55E-02
5	46733	S1_20396204	6.95E-06	9.17E-01	1.76E-01
6	46613	S1_18628414	8.01E-06	1.00E+00	1.76E-01
7	5224	S3_3019494	3.96E-05	1.00E+00	2.49E-01
8	10223	S5_9396256	3.21E-05	1.00E+00	2.49E-01
9	10224	S5_9396276	3.21E-05	1.00E+00	2.49E-01
10	23283	S11_1797543	4.53E-05	1.00E+00	2.49E-01

Table 8. 2015 low PC1 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	index	Marker	Raw p-value	Bonferroni	BH
1	9653	S5_9854486	7.55E-07	9.22E-02	9.22E-02
2	9485	S5_7982850	2.87E-06	3.51E-01	1.23E-01
3	9625	S5_9694291	3.62E-06	4.42E-01	1.23E-01
4	9755	S5_11887637	7.25E-06	8.86E-01	1.23E-01
5	73510	S16_15835946	7.51E-06	9.18E-01	1.23E-01
6	9484	S5_7982800	1.02E-05	1.00E+00	1.23E-01
7	9659	S5_10108831	1.11E-05	1.00E+00	1.23E-01
8	9712	S5_10723002	1.03E-05	1.00E+00	1.23E-01
9	9734	S5_11333806	1.01E-05	1.00E+00	1.23E-01
10	25495	S13_1280795	8.46E-06	1.00E+00	1.23E-01

Table 9. 2015 maximum PC2 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	index	Marker	Raw p-value	Bonferroni	BH
1	43109	S1_18079176	1.29E-06	1.58E-01	1.58E-01
2	140	S0_6635996	2.71E-05	1.00E+00	3.31E-01
3	141	S0_6636001	2.71E-05	1.00E+00	3.31E-01
4	142	S0_6636004	2.71E-05	1.00E+00	3.31E-01
5	35857	S18_2398499	2.53E-05	1.00E+00	3.31E-01
6	43278	S1_20396204	1.38E-05	1.00E+00	3.31E-01
7	43411	S1_21085379	1.81E-05	1.00E+00	3.31E-01
8	43412	S1_21085394	1.81E-05	1.00E+00	3.31E-01
9	47529	S4_4336219	2.49E-05	1.00E+00	3.31E-01
10	55930	S7_24677404	2.37E-05	1.00E+00	3.31E-01

Table 10. 2015 low PC2 statistically significant SNPs relative to raw p-value, Bonferroni correction, and Benjamini-Hochberg procedure.

P-value rank	index	Marker	Raw p-value	Bonferroni	BH
1	2691	S1_21088363	1.78E-06	2.17E-01	1.26E-01
2	84656	S2_1919469	2.94E-06	3.59E-01	1.26E-01
3	118352	S18_10366622	3.10E-06	3.79E-01	1.26E-01
4	1469	S1_6729328	3.05E-05	1.00E+00	2.94E-01
5	2130	S1_12893206	1.46E-05	1.00E+00	2.94E-01
6	2495	S1_19564318	8.29E-05	1.00E+00	2.94E-01
7	2816	S1_22788309	3.54E-05	1.00E+00	2.94E-01
8	2949	S1_24054279	2.94E-05	1.00E+00	2.94E-01
9	4127	S2_16736744	4.54E-05	1.00E+00	2.94E-01
10	23319	S12_830617	8.03E-05	1.00E+00	2.94E-01

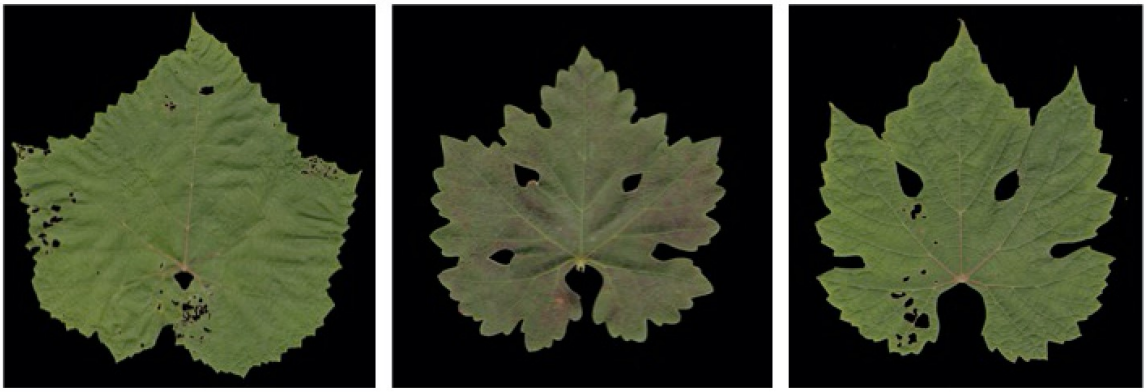

Figure 1. Images of leaves from breeding population located in Genomics Research Vineyard at State Fruit Experiment Station, Missouri State University, Mountain Grove, Missouri Campus. Left: Norton parent; Center: Cabernet Sauvignon parent; Right: Norton x Cabernet Sauvignon F_1 progeny.

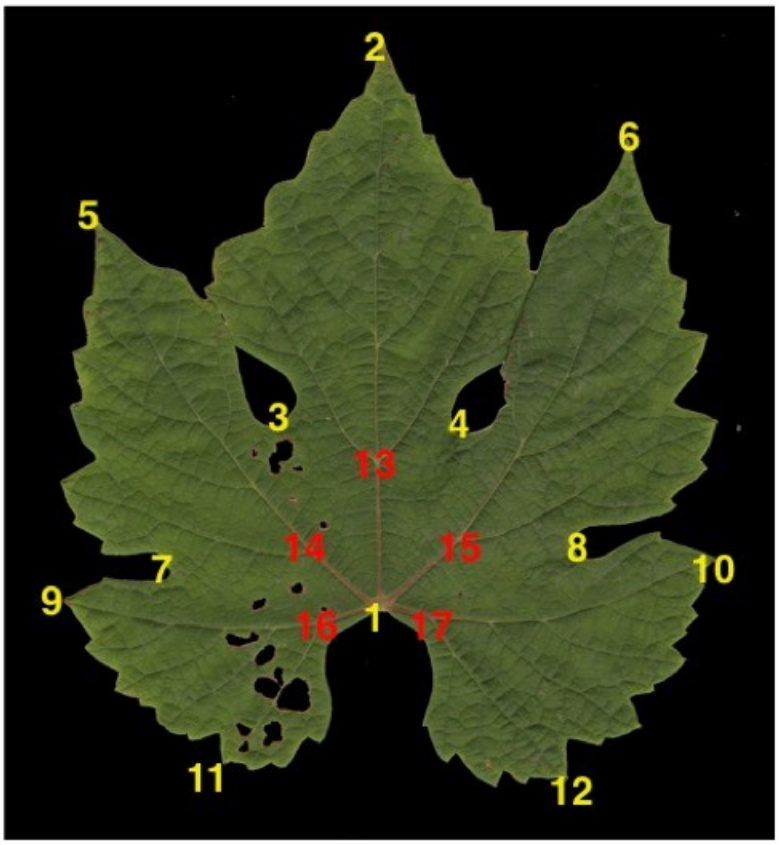

Figure 2. Landmark placement on leaf surface, adaxial side. Exterior Venation Pattern (1-12): 1-Petiole Junction, 2-tip of midvein, 3-left upper lateral sinus, 4-right upper lateral sinus, 5-tip of left distal vein, 6-tip of right distal vein, 7-left lower lateral sinus, 8-right lower lateral sinus, 9-tip of left proximal vein, 10-tip of right proximal vein, 11-tip of left petiolar vein, 12-tip of right petiolar vein. Interior Venation Pattern (13-17): 13-first major branching point of midvein, 14-first major branching point of left distal vein, 15-first major branching point of right distal vein, 16-left petiole vein branching point from proximal, 17-right petiole vein branching point from proximal.

Figure 3. Image examples of graphical leaf check output from 'leaf check' script in R. Top: Incorrectly landmarked Norton parent leaf 13. Bottom: Corrected landmark placement of Norton parent leaf 13.

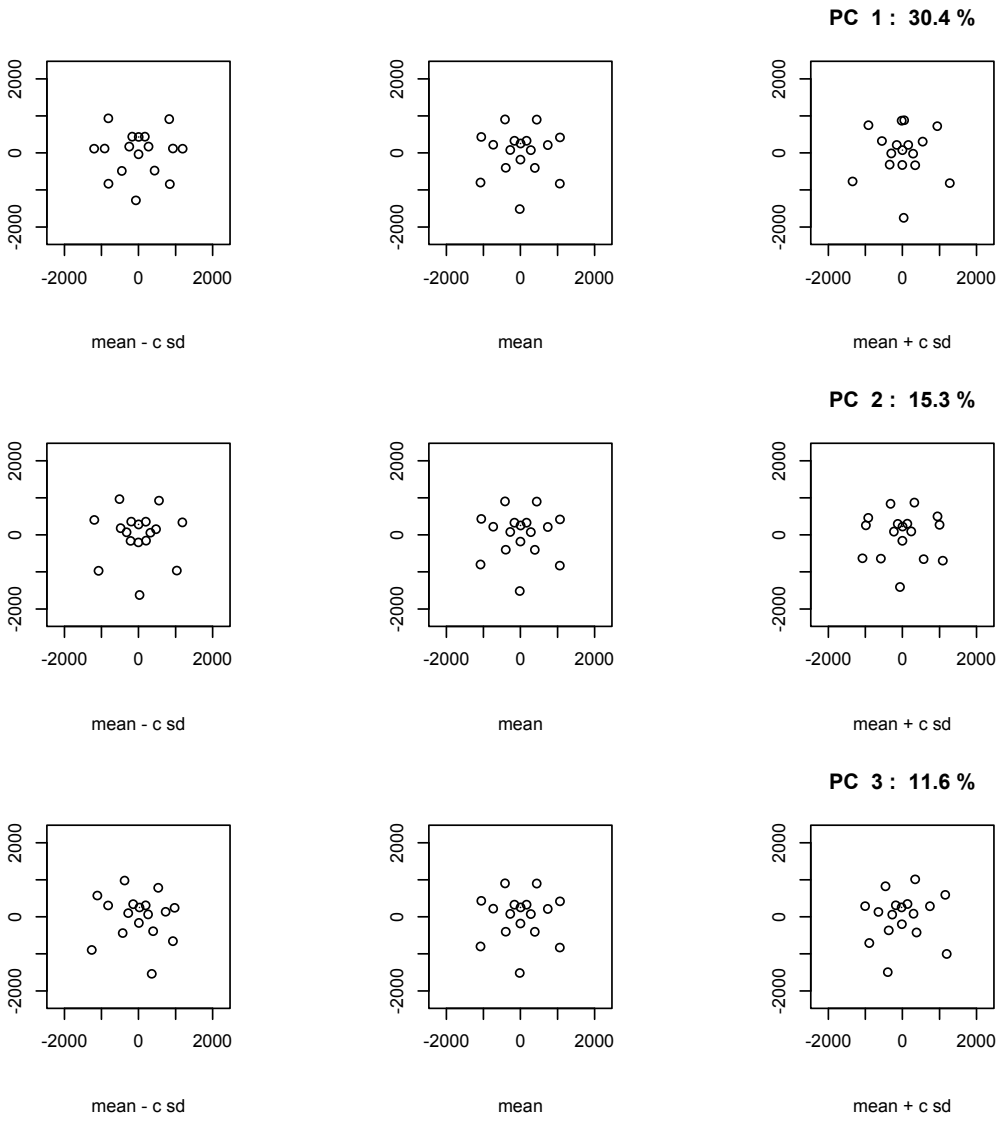

Figure 4. 2014 top three Principal Components produced from Generalized Procrustes Analysis. PC1, PC2, and PC3 are estimated to account for approximately 60% of the shape variation measured in this population. PCs are displayed individually by rows with three plotted images each to demonstrate the range of shape variation explained per PC. The mean shape of the population is shown at center in each row for reference. For each PC, -2 SD from the mean is shown left of center, and +2 SD from the mean is shown right of center.

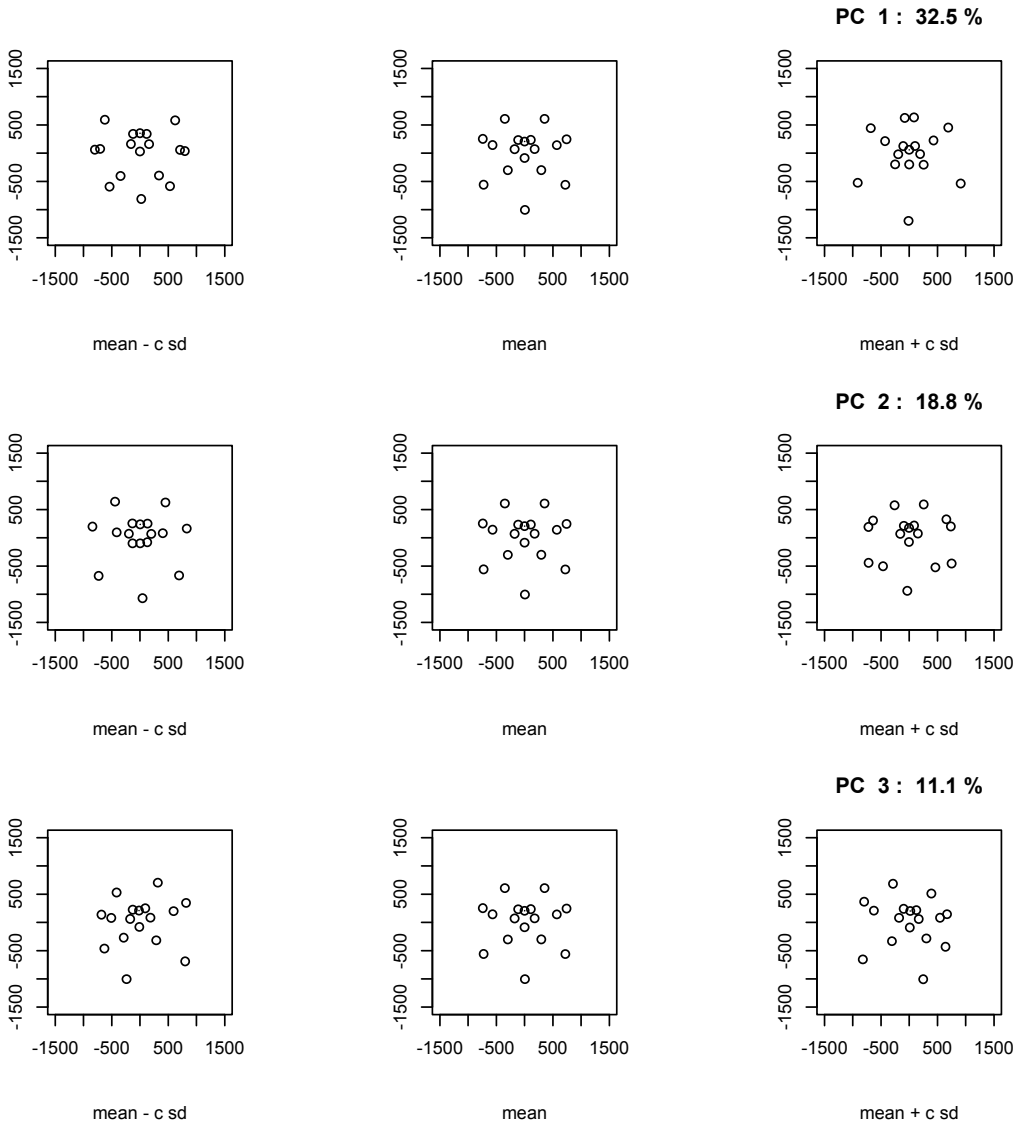

Figure 5. 2015 top three Principal Components produced from Generalized Procrustes Analysis. PC1, PC2, and PC3 are estimated to account for approximately 60% of the shape variation measured in this population. PCs are displayed individually by rows with three plotted images each to demonstrate the range of shape variation explained per PC. The mean shape of the population is shown at center in each row for reference. For each PC, -2 SD from the mean is shown left of center, and +2 SD from the mean is shown right of center.

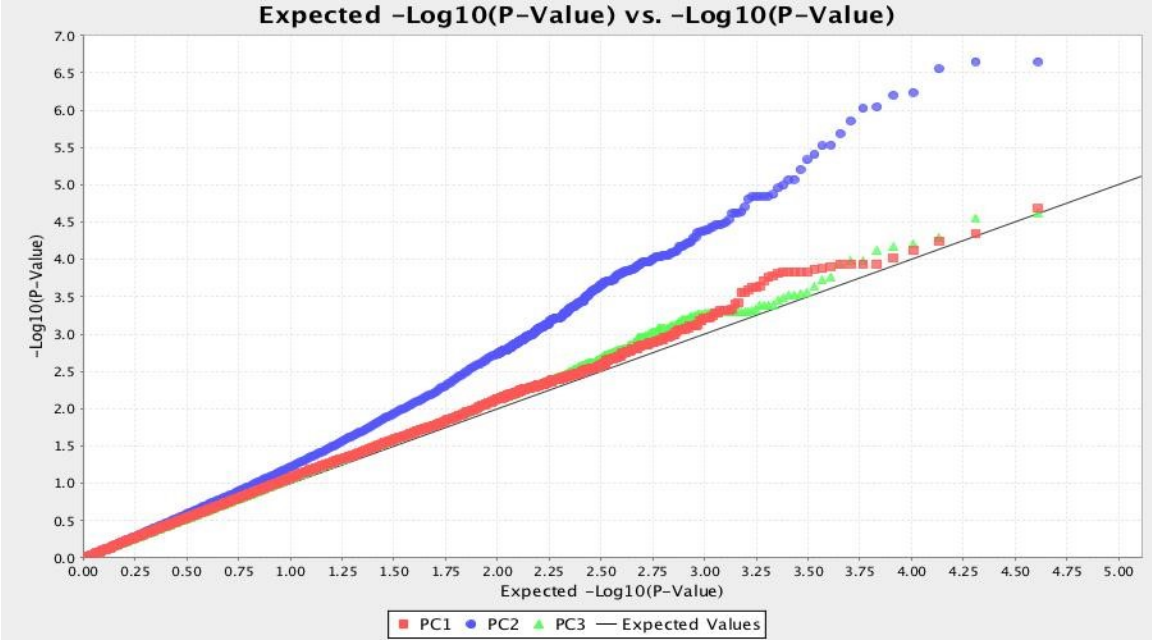

Figure 6. QQ Plot for 2014 Maximum PC1 GLM analysis. Significant association indicated with PC2 when PC scores are sorted according to Maximum PC1.

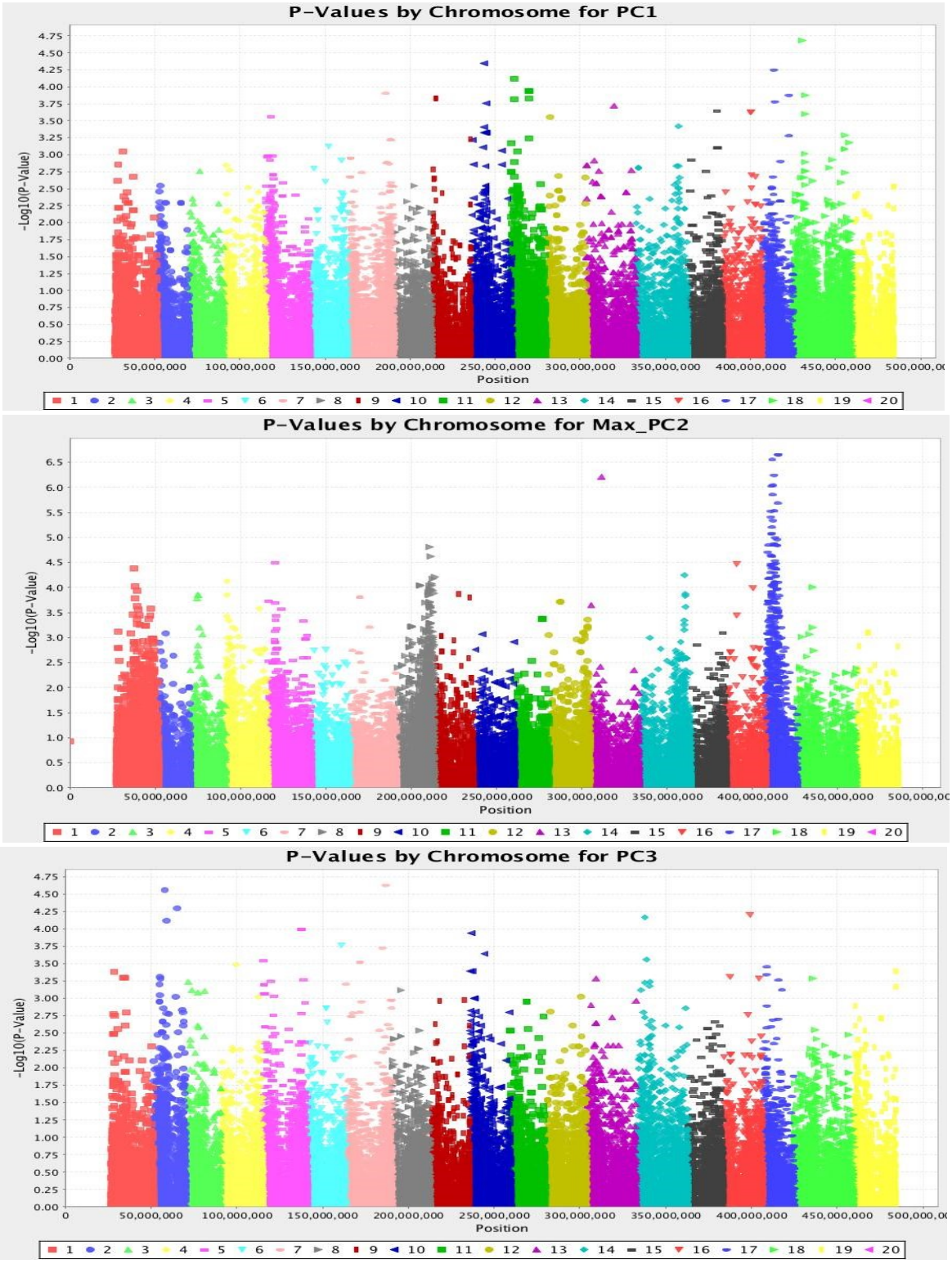

Figure 7. Manhattan Plots for 2014 Maximum PC1 GLM analysis. Top: PC1; Center: PC2; Bottom: PC3. Chromosomes 8 & 17 show association with PC2; chromosome 1 shows some clustering in association with PC2.However, significant results were observed for PC2 on Linkage Group (LG) 17 (LOD=6.75).

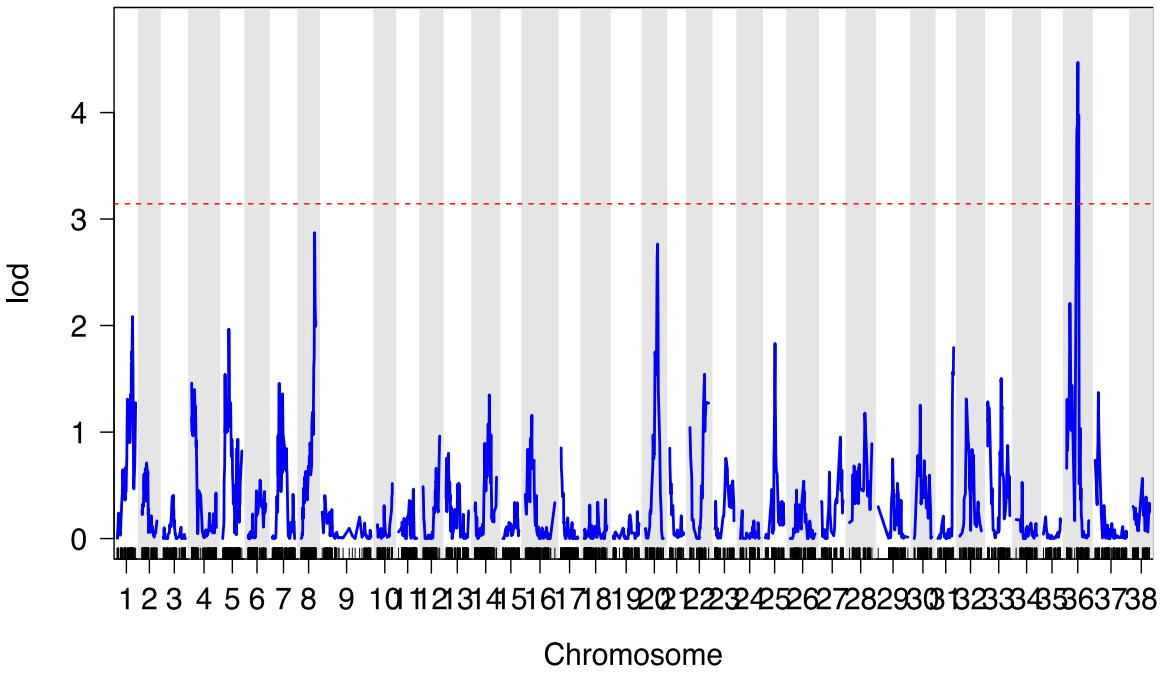

Figure 8. R/qtl consensus map of 2014 year data. Significance shown by LOD score; threshold for significance represented by red dotted line on y-axis. Chromosomes are shown on x-axis; 1-19 represents 19 chromosomes of Norton and 20-38 represents 19 chromosomes of Cabernet Sauvignon. Chromosome 17 from Cabernet Sauvignon shows statistically significant association (LOD=4.5). The peak representing Chromosome 8 from Norton reaches just below the significance threshold (LOD=2.9) and peak for Chromosome 1 on Cabernet Sauvignon is third highest (LOD=2.8).

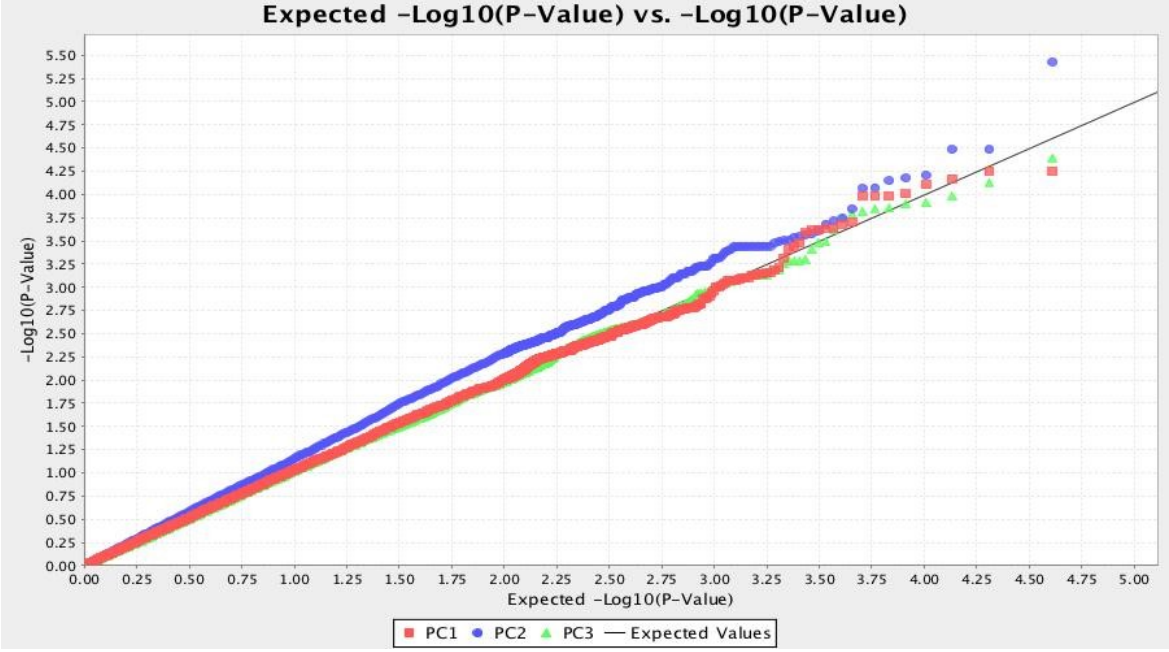

Figure 9. QQ Plot for 2014 Low PC1 GLM analysis. Significant association with PC2 indicated PC scores are sorted (before analysis) according to Low PC1.

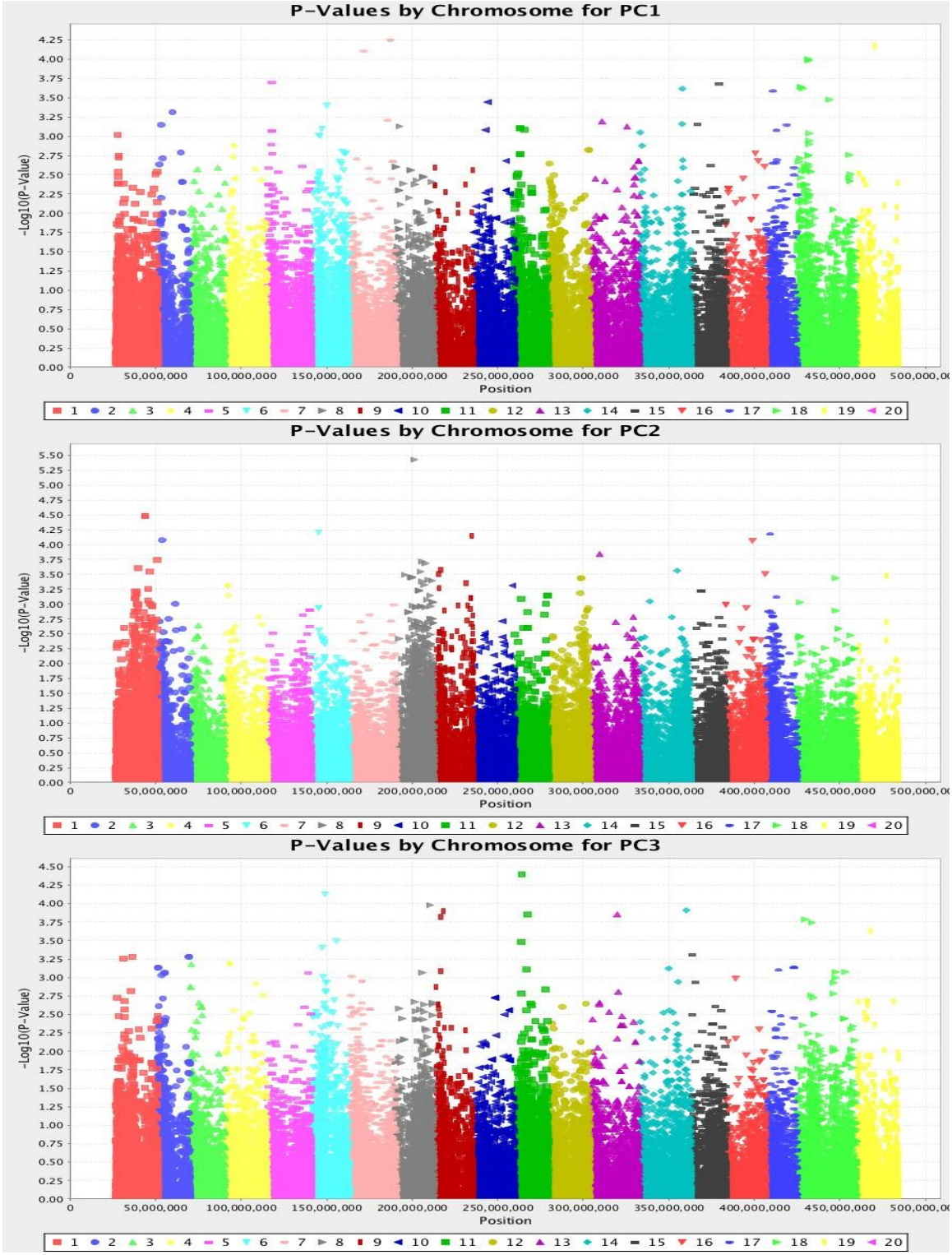

Figure 10. Manhattan Plots for 2014 LowPC1 GLM analysis. Top: PC1; Center: PC2; Bottom: PC3. No significant association indicated.

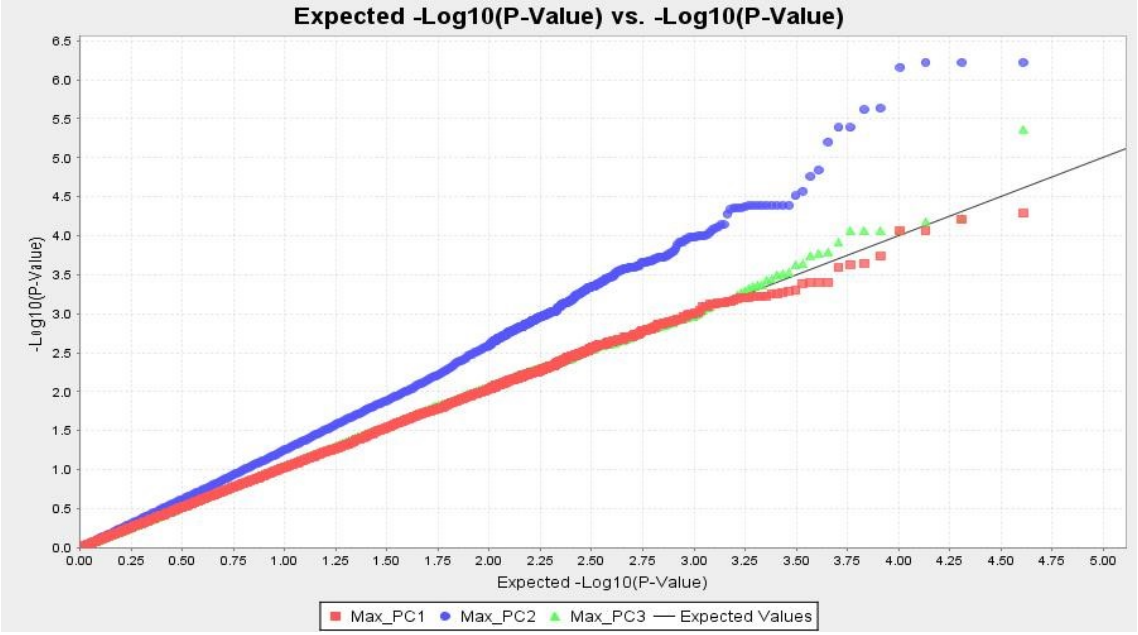

Figure 11. QQ Plot for 2014 Maximum PC2 GLM analysis. Significant association with PC2 indicated when PC scores are sorted (before analysis) according to Maximum PC2.

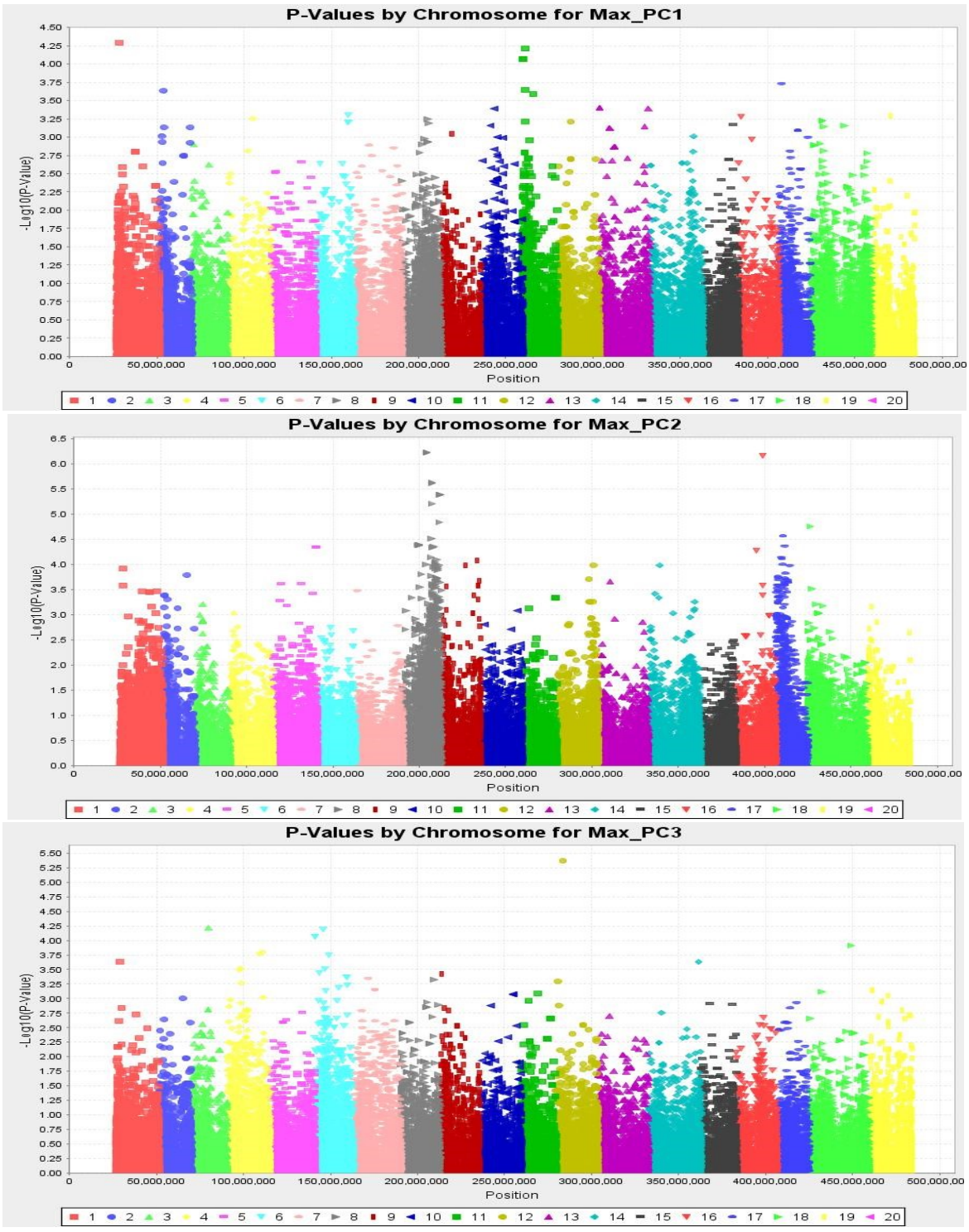

Figure 12. Manhattan Plots for 2014 Maximum PC2 GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Chromosomes 8 (LOD=6.25) & 17 (LOD=4.6) appear to show association with PC2; one SNP on Chromosome 16 shows association with PC2 (LOD=6.1).

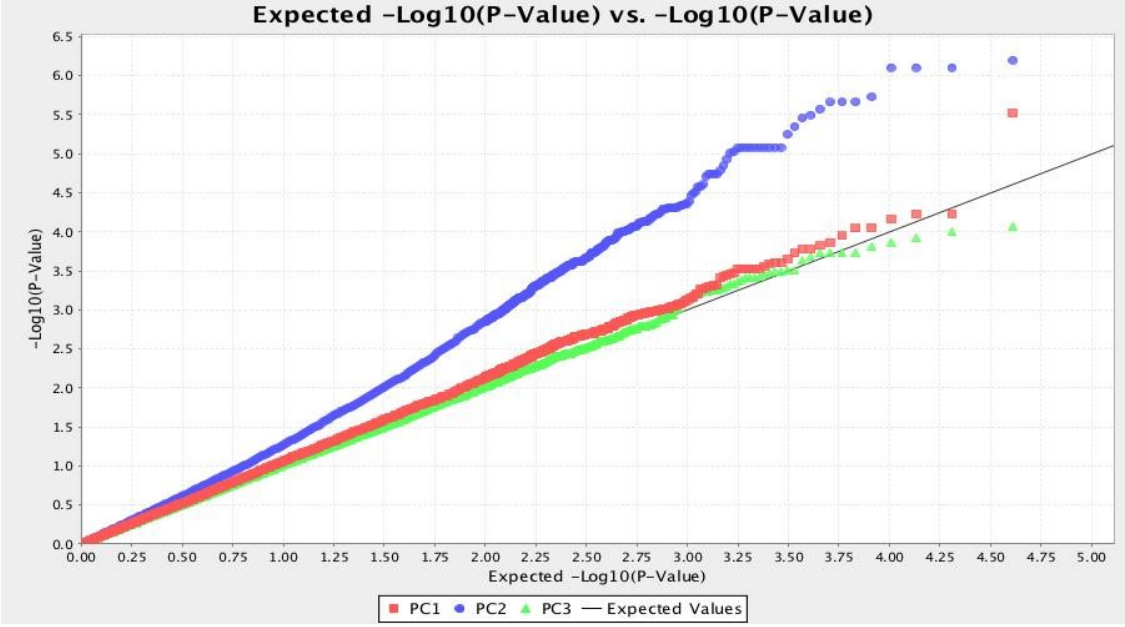

Figure 13. QQ Plot for 2014 Low PC2 GLM analysis. Significant association with PC2 when PC scores are sorted (before analysis) according to Low PC2.

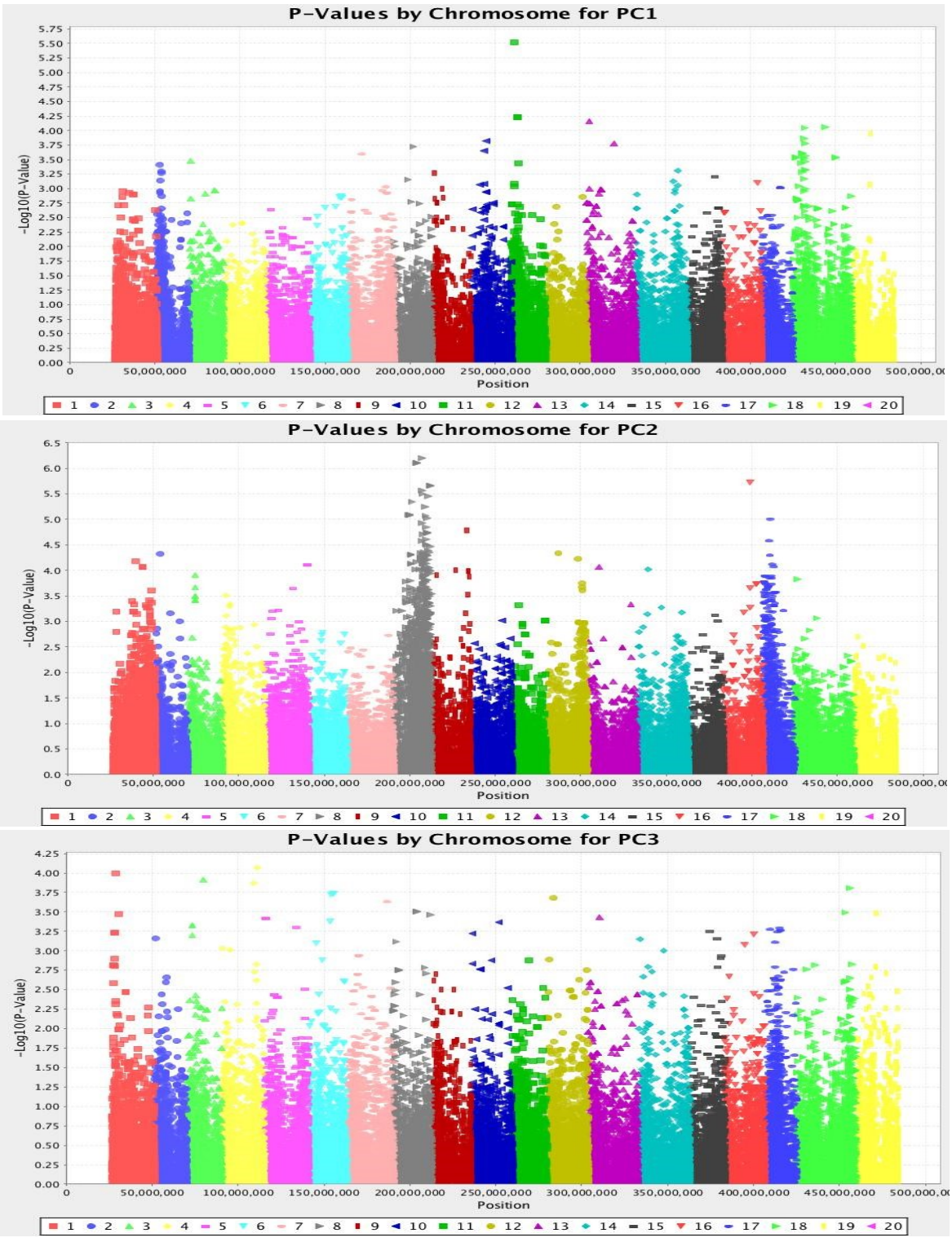

Figure 14. Manhattan Plots for 2014 Low PC2 GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Chromosomes 8 (LOD=6.25) & 17 (LOD=5.0) show association with PC2; one SNP on Chromosome 16 appears to associate with PC2 (LOD=5.75).

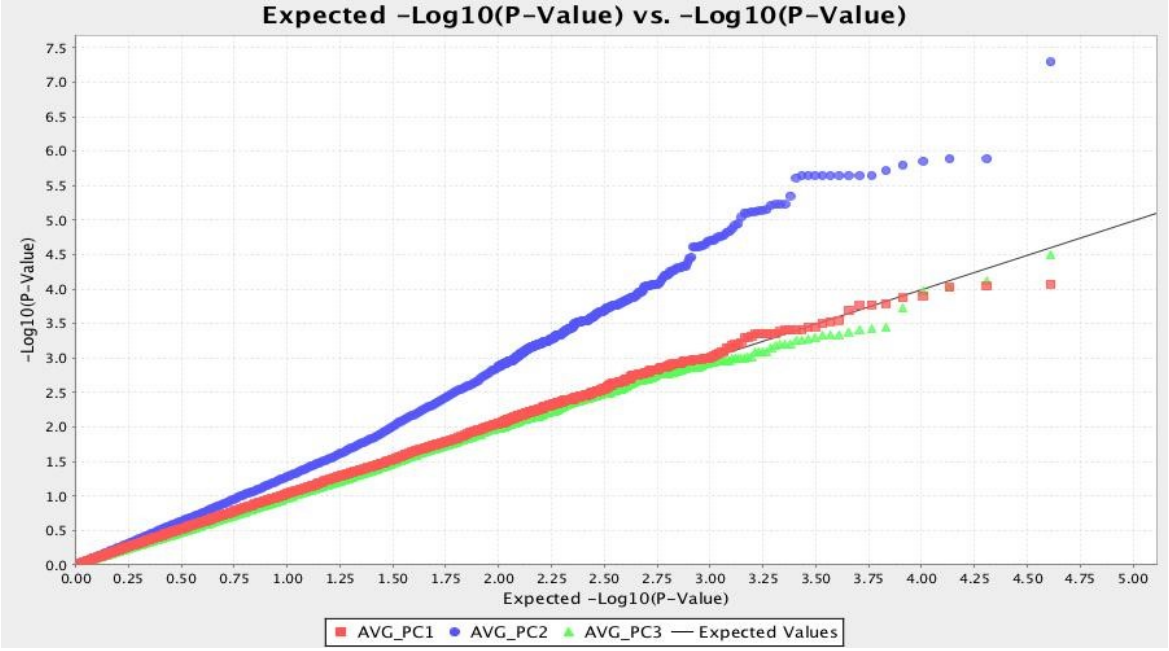

Figure 15. QQ Plot for 2014 All PCs Averaged GLM analysis. Significant association indicated with PC2 when PC scores are sorted (before analysis) according to All PCs Averaged.

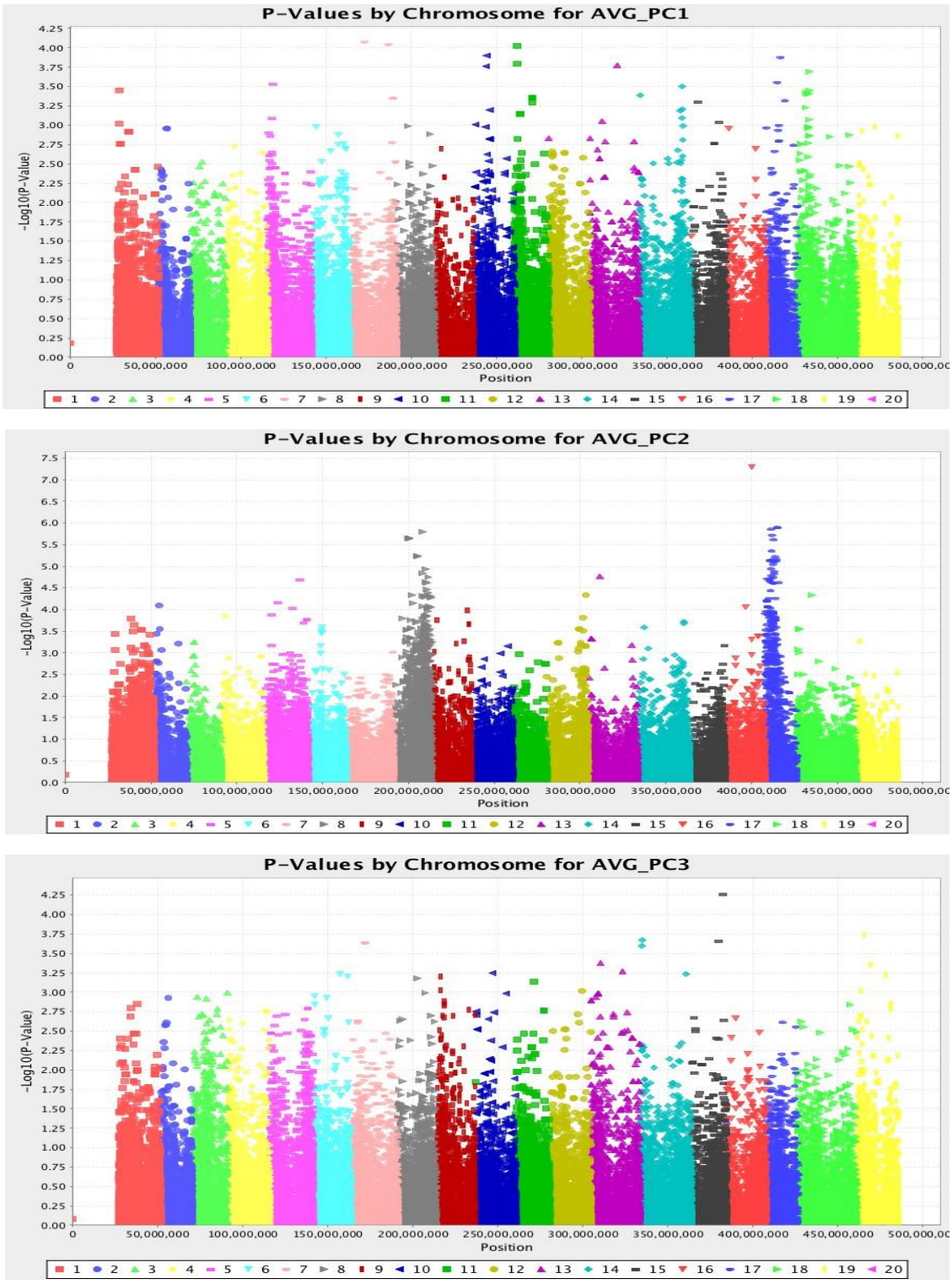

Figure 16. Manhattan Plots for 2014 All PCs Averaged GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Chromosomes 8 (LOD=6.0) & 17 (LOD=6.0) show association with PC2; one SNP on Chromosome 16 shows association with PC2 (LOD=7.25).

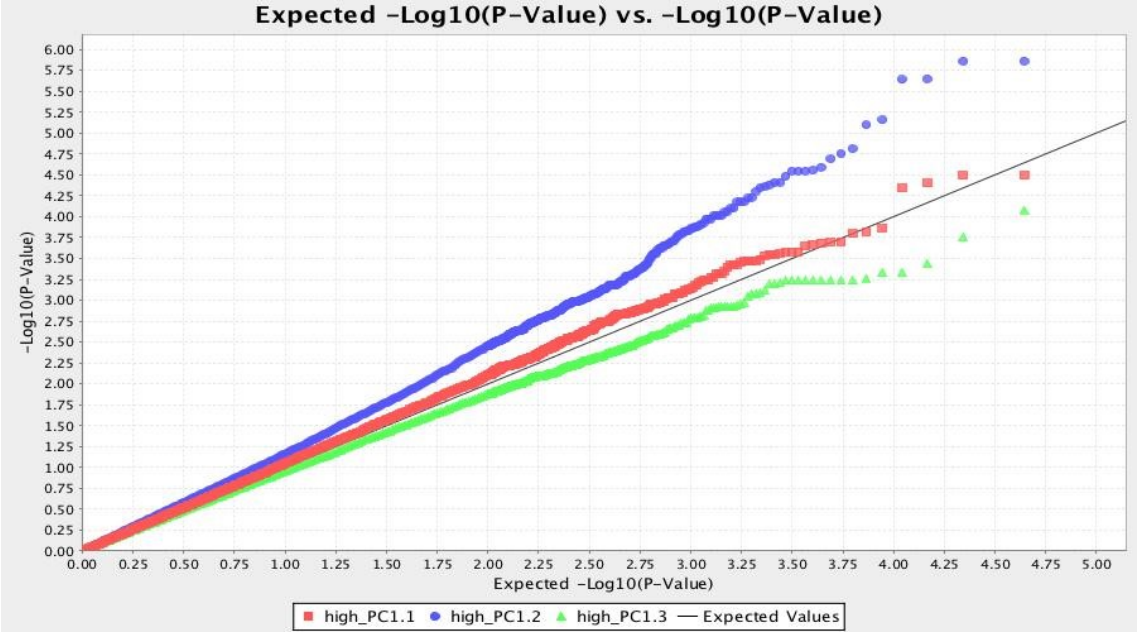

Figure 17. QQ Plot for 2015 Maximum PC1 GLM analysis. Significant association indicated by raw p-values with PC2.

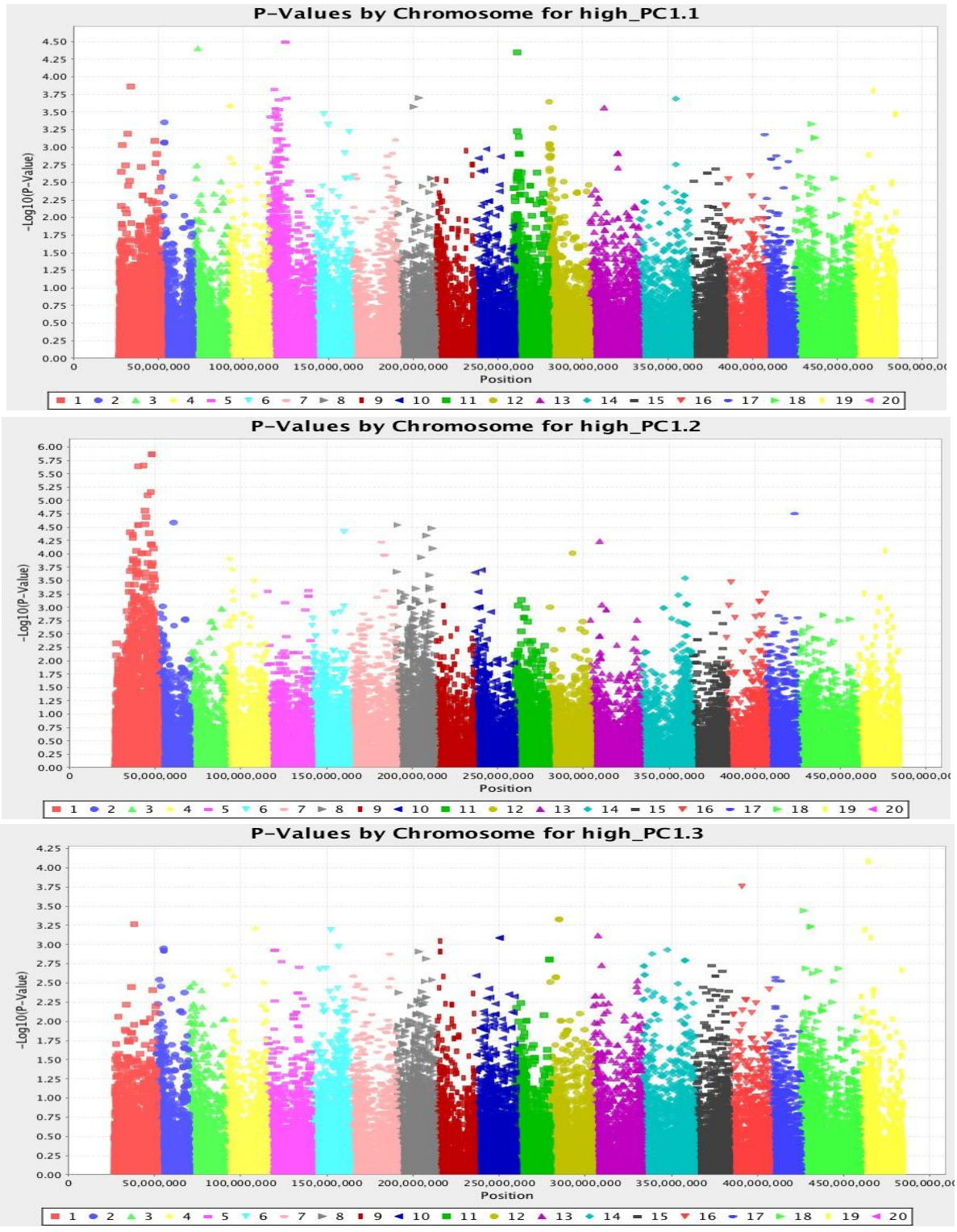

Figure 18. Manhattan Plots for 2015 Maximum PC1 GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Raw p-values display clustering on chromosome 5 in association with PC1, and on chromosome 1 in association with PC2.

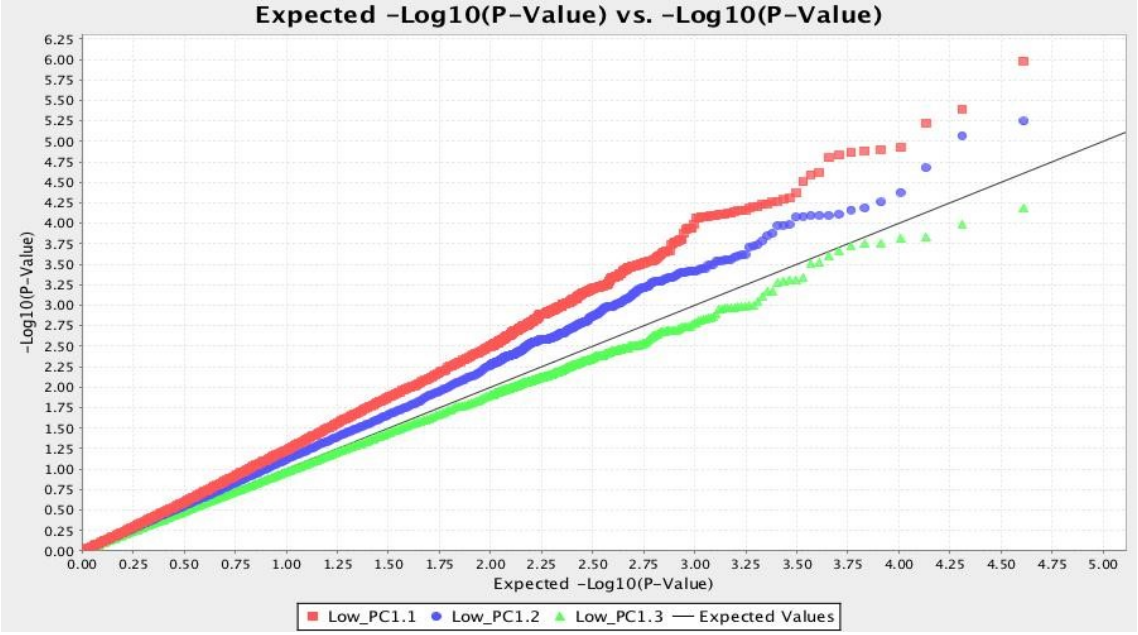

Figure 19. QQ Plot for 2015 Low PC1 GLM analysis. Significant association indicated by raw p-values with PC1 and PC2.

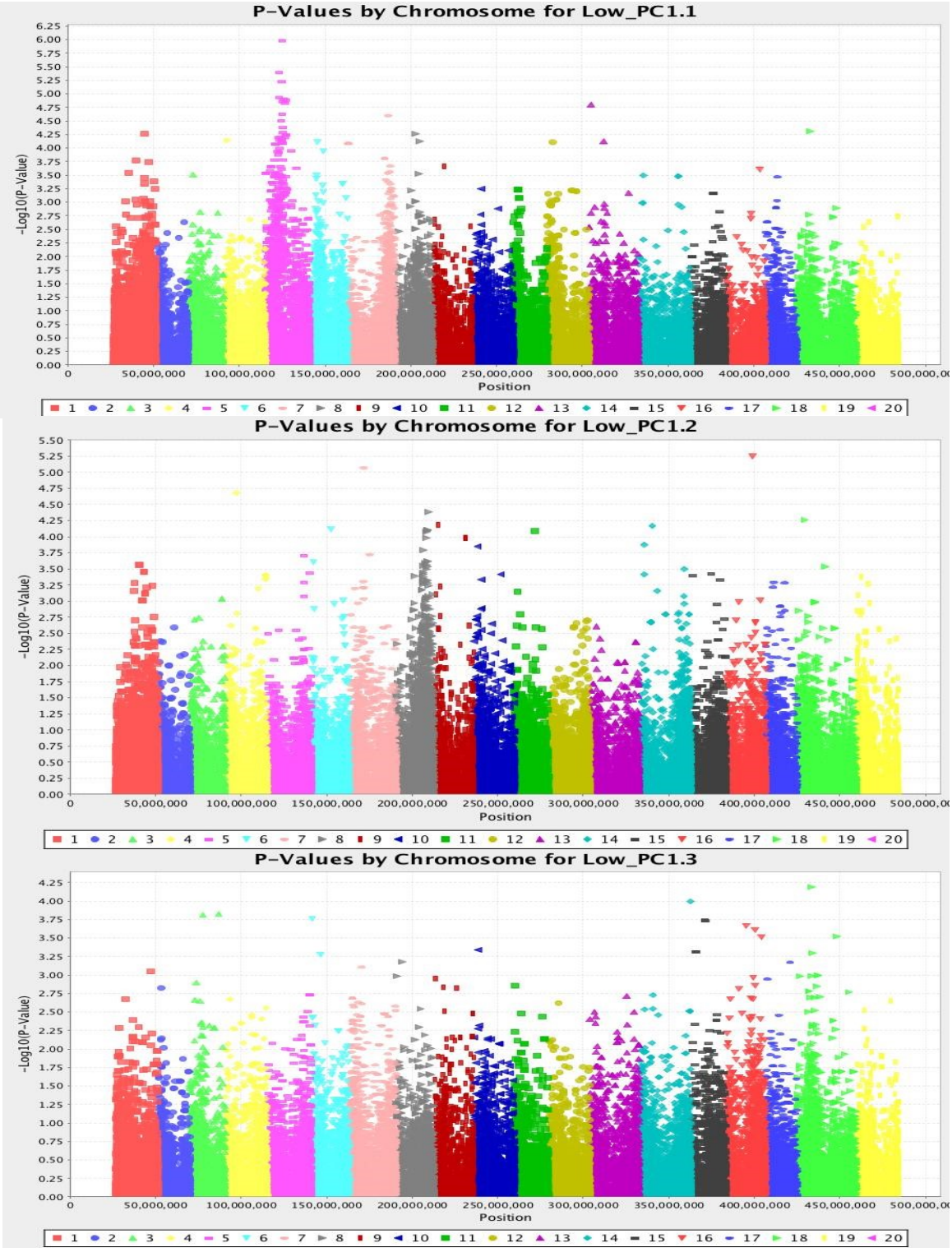

Figure 20. Manhattan Plots for 2015 LowPC1 GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Raw p-values indicate association with PC1 on Chromosome 5. SNPs cluster according to raw p-values on chromosome 8 in association with PC2.

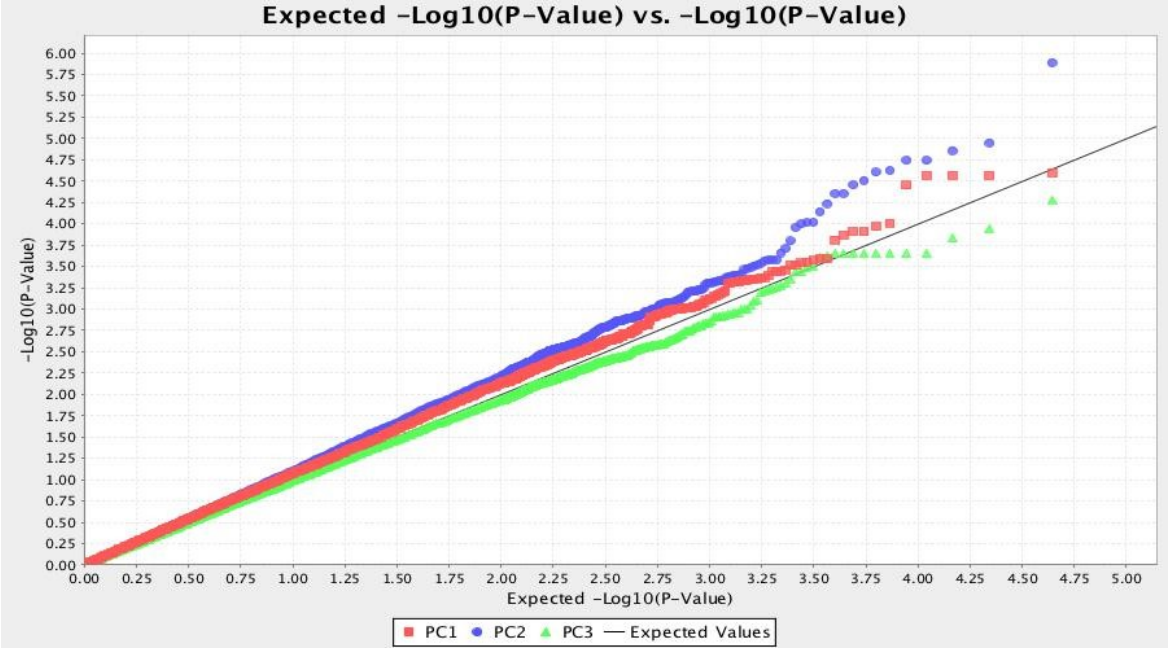

Figure 21. QQ Plot for 2015 Maximum PC2 GLM analysis. Significant association indicated by raw p-values with PC2.

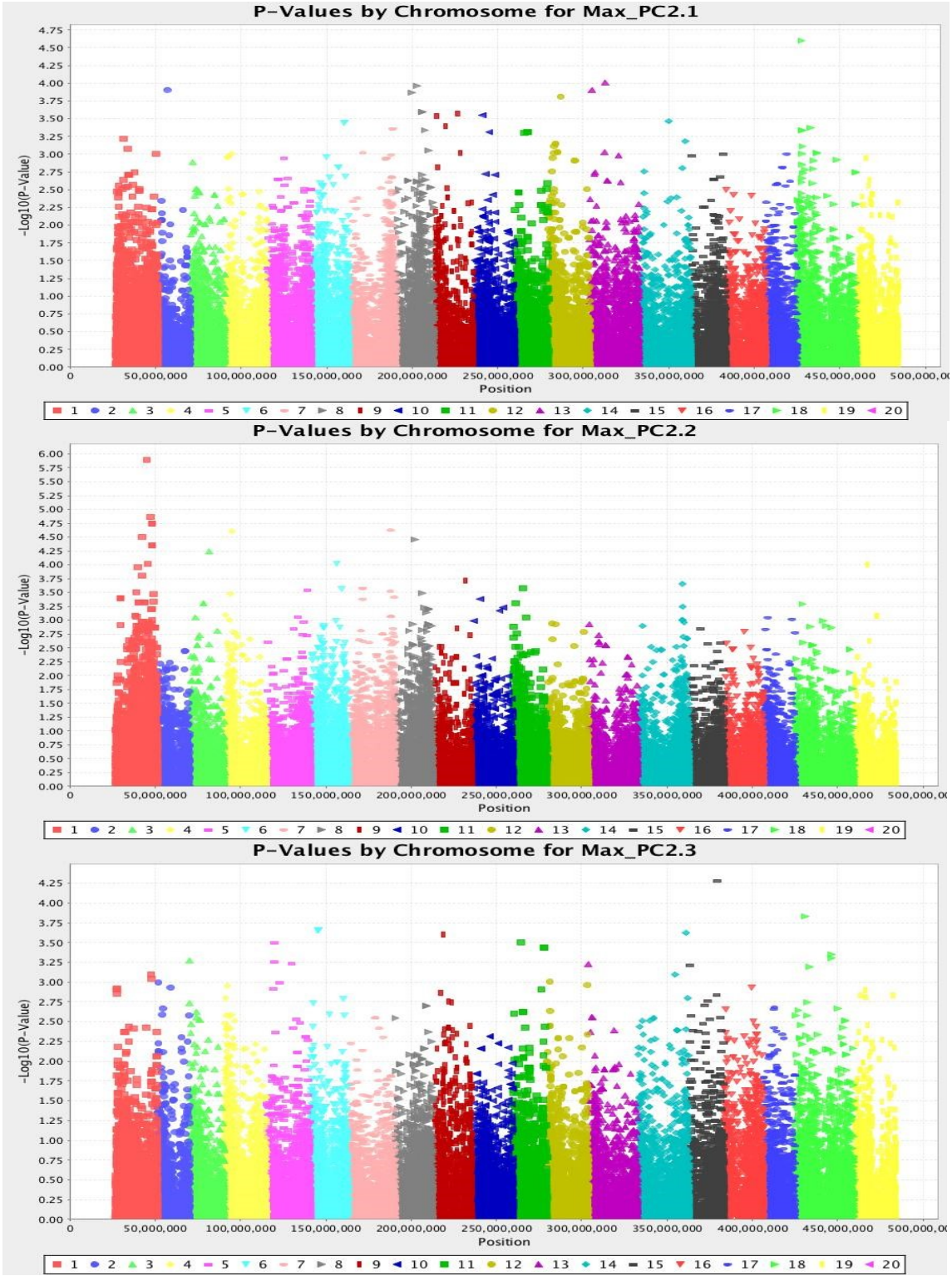

Figure 22. Manhattan Plots for 2015 Maximum PC2 GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Raw p-values display clustering on chromosome 11 in association with PC1, and on chromosomes 8 and 17 in association with PC2.

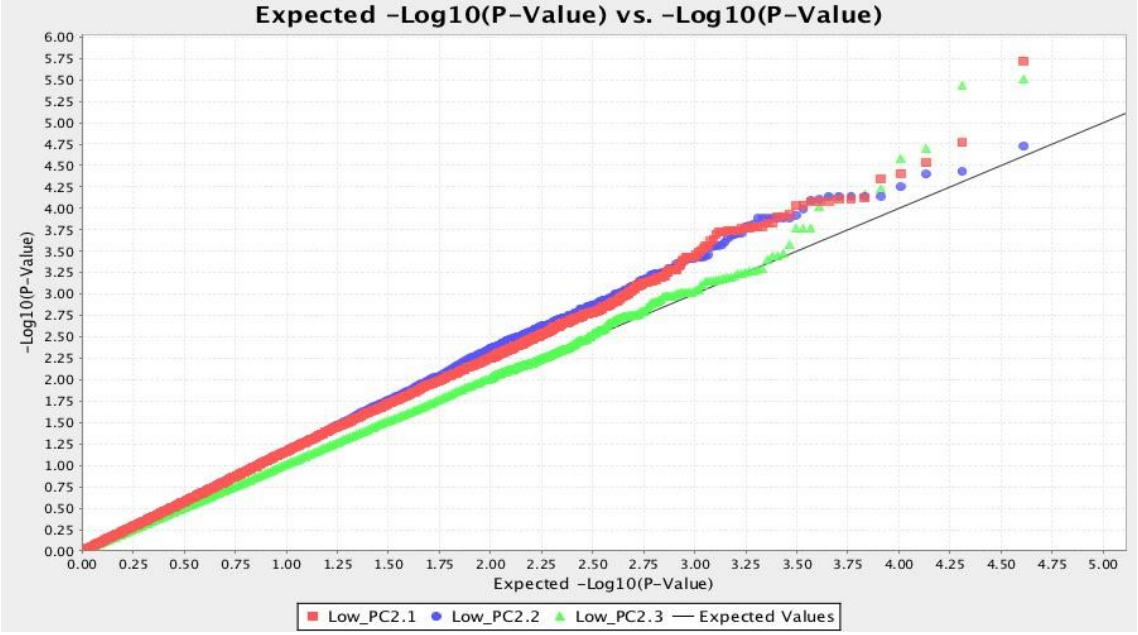

Figure 23. QQ Plot for 2015 Low PC2 GLM analysis. Significant association indicated by raw p-values with PC1 and PC3.

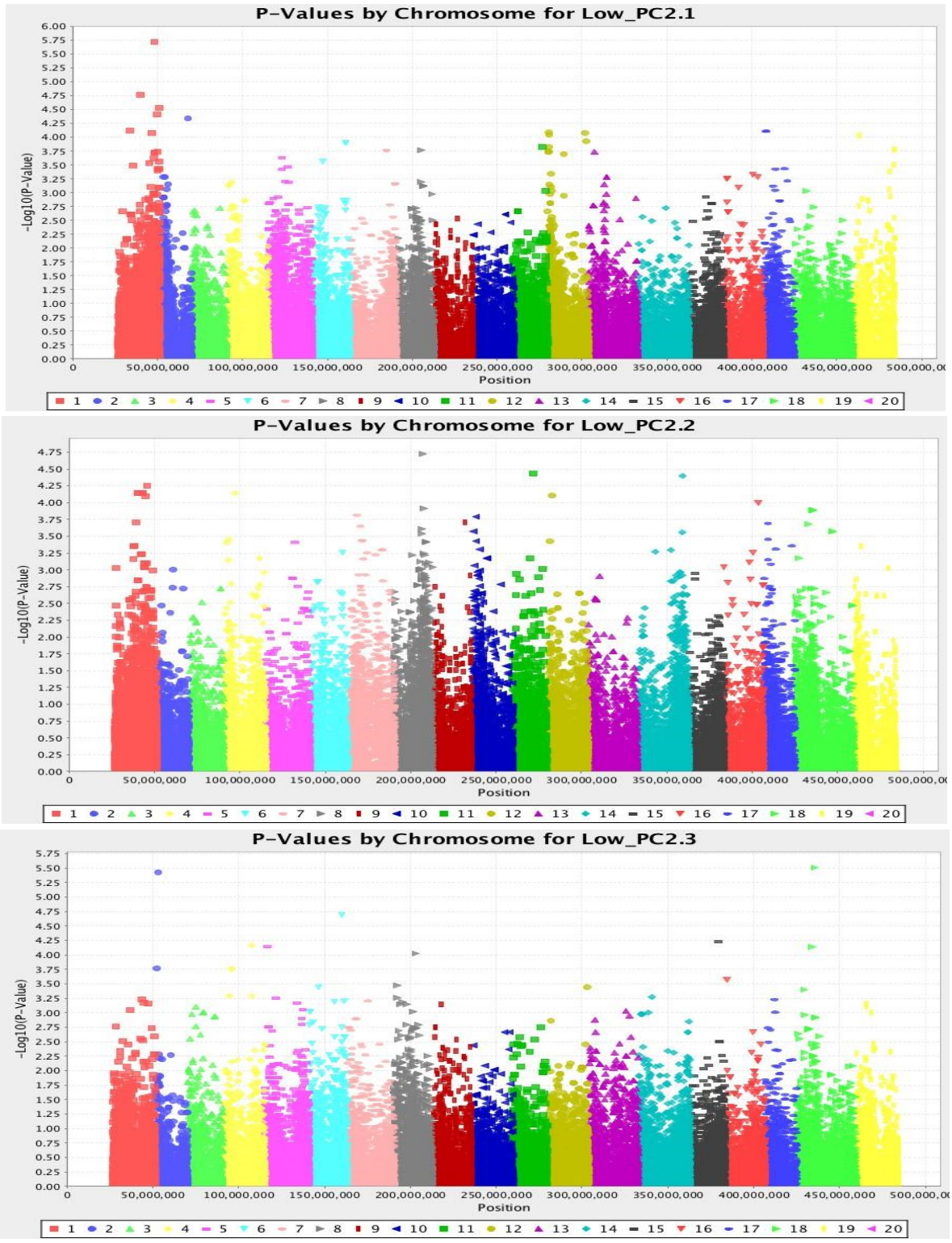

Figure 24. Manhattan Plots for 2015 Low PC2 GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Raw p-values display clustering and significance on chromosome 1 in association with PC1, and on chromosomes 1 and 8 in association with PC2.

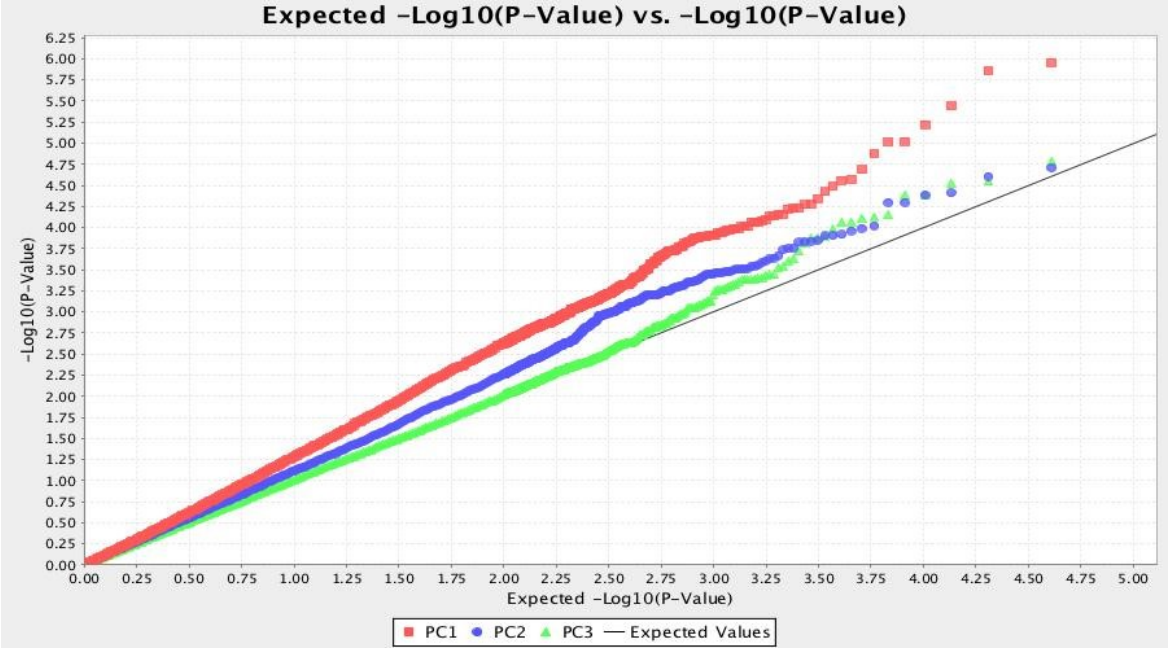

Figure 25. QQ Plot for 2015 All PCs Averaged GLM analysis. Moderate association detected with PC1 and PC2.

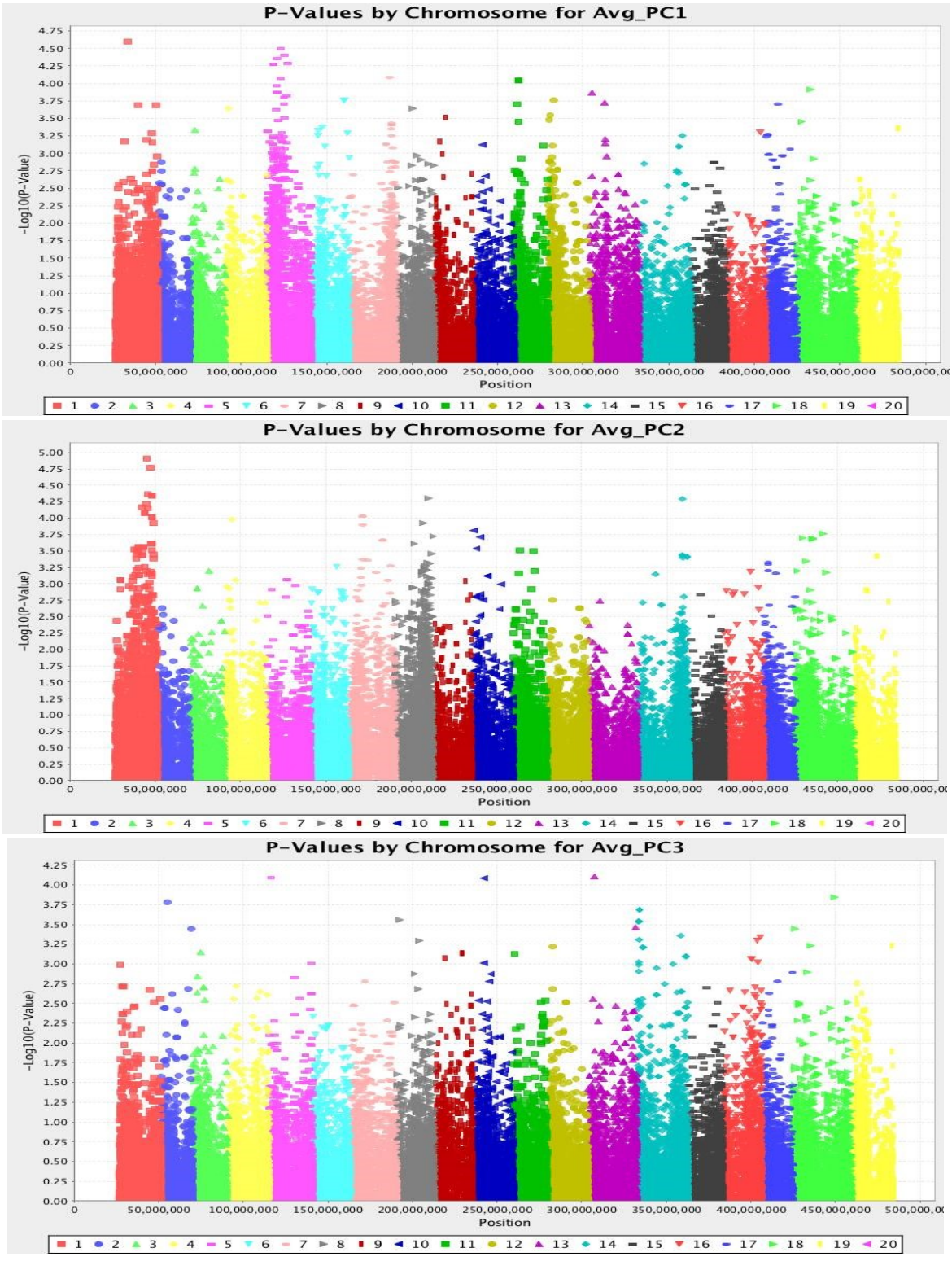

Figure 26. Manhattan Plots for 2015 All PCs Averaged GLM analysis. PC1 (top), PC2 (center), PC3 (bottom). Raw p-values display clustering on chromosomes 1 and 5 in association with PC1, and on chromosomes 1 and 8 in association with PC2.