AN ABSTRACT OF THE THESIS OF

Ryan C. Graebner for the degree of Master of Science in Crop Science presented on November 21, 2014

Title: Germplasm Design, Gene Discovery, and Variety Release: Breeding Barley for the Pacific Northwest

Abstract Approved:

<div style="border-top:1px solid #000;"></div>

Patrick M. Hayes   Alfonso Cuesta-Marcos

This thesis consists of three manuscripts addressing separate components of barley breeding, plus an introduction and a conclusion summarizing the status of barley in the Pacific Northwest, plus the implications of the research presented in this thesis. In the first manuscript, a package written for the statistical software "R" that is designed to identify subsets of plant accessions that would be favorable for rare-trait discovery and genome-wide association studies is presented. In the second manuscript, genome-wide association scanning is used to identify thirteen single-nucleotide polymorphism markers that are significantly associated with tocochromanol concentration in a set of elite, spring barley accessions. In the third manuscript, the barley variety "Alba" was described in a germplasm release.

Germplasm Design, Gene Discovery, and Variety Release: Breeding Barley for the
Pacific Northwest

by
Ryan C. Graebner

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented November 21, 2014
Commencement June 2015

Master of Science thesis of <u>Ryan C. Graebner</u> presented on <u>November 21, 2014</u>.

APPROVED:

_____

Co-Major Professor, representing Crop Science

_____

Co-Major Professor, representing Crop Science

_____

Head of Department of Crop and Soil Science

_____

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

_____

Ryan C. Graebner, Author

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Pat Hayes, Alfonso Cuesta-Marcos, and the rest of the Oregon State barley project for guiding and contributing to the research presented in this thesis.

I would also like to thank Christina, my parents Paul and Linda, and my sister Lauren for providing me with support, encouragement, and inspiration.

CONTRIBUTION OF AUTHORS

Chapter 2. Christina Hagerty was involved in the writing of this chapter.

Chapter 3. Dr. Tom Blake, Dr. Shiaoman Chao, Dr. David Hole, Dr. Rich Horsley, Dr. Don Obert, Dr. Kevin P. Smith, and Dr. Stephen Ullrich contributed germplasm that was screened in this project. Dr. Tom Blake grew the set of germplasm used for genome-wide association scanning of genes controlling tocochromanol biosynthesis. Victoria Blake played a key role in curating data for this experiment. Dr. Mitchell Wise characterized the concentration of tocochromanol forms in the set of germplasm used for genome-wide association scanning of genes controlling tocochromanol biosynthesis. Matthew Geniza and Dr. Pankaj Jaiswal assisted with the identification of candidate genes that could be responsible for the quantitative trait loci for tocochromanol concentration identified in this chapter. Dr. Mitchell Wise and Matthew Geniza contributed to the writing of this manuscript.

Chapter 4. Dr. Stephen S. Jones, Brook O. Brouwer, and Scott Fisk assisted in the characterization of morphological, agronomic, and quality traits the barley variety Alba. Brook O. Brouwer assisted in the analysis of disease data for this manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF APENDIX A FIGURES

LIST OF APENDIX TABLES

Introduction

Barley was one of the earliest crops to be domesticated, between 8,200 and 11,700 years ago (Fuller 2006). Since its domestication, barley has been adopted by societies around the world, with each society selecting for barley with traits favorable for their local environments. Compared to other crops, barley tends to perform well in stressful environments (Muñoz-Amatriaín et al. 2014). In particular barley has demonstrated an incredible ability to survive at high altitudes, with germplasm collection records showing that the average pigmented barley variety in the USDA Core Collection for barley was collected from a site at 2,555m.

The four primary end-uses of barley are malt, food, forage, and animal feed. Malt barley has the strictest requirements, with uniform plumpness, low wort beta-glucan and protein, and alpha-amylase and enzyme contents that fall within a certain range (Mather et al. 1997). Food barley, when intended for whole-grain markets, is often characterized by non-adhering, "naked" hulls and colored grain. High grain beta-glucan concentrations can increase the health benefits of barley (Behall 1997), although this trait tends to receive less attention from growers and distributors. Barley varieties intended for forage are typically either "hooded" or awn-less, to make spikes more palatable to livestock. Feed barley has no strict requirements. As a consequence, feed barley originates from two primary sources: feed varieties that are selected for yield only, and malt, food, and forage barley that is either in surplus, or doesn't meet the criteria for its category.

Currently, barley production in the Willamette Valley, in Oregon, is relatively low. A major reason for this is the relatively low price of feed barley, compared to wheat. However, a recent surge in craft breweries in the Pacific Northwest region of the United States of America has led to an increased interest in barley that is both grown and malted locally. Also, recent improvements in naked food barley bred specifically for the Pacific Northwest may make whole-grain food barley production more profitable than it has been in the past.

The climate seen to the west of the Cascade Mountains in the Pacific Northwest can only be found in several places on earth. Due to a combination of its position relative

to earth's primary circulation cells, and its proximity to an ocean current that carries cold water from the arctic, this region experiences a long, wet winter, interrupted by approximately two months of drought in the summer. This climate generally allows plants that can be fall- or winter-planted to grow to maturity with no irrigation. As a bonus, for most grains and seed crops, maturity coincides with the summer drought, reducing the risk of crop-damaging rains before harvest.

While this climate has certain advantages, crops must also possess adaptations that are either not beneficial or irrelevant in many other parts of the world. For barley, this includes a winter or facultative growth habit (to increase yield and eliminate the need for irrigation), tolerance to wet soils, and resistance to barley stripe rust (incited by *Puccinia striiformis* f. sp. *hordei*) and scald (incited by *Rhynchosporium commune*). While scald and stripe rust have historically been the most prevalent diseases in the Willamette valley, in the 2014 growing season, heavy infestations of leaf rust (incited by *Puccinia hordei*), stem rust (incited by *Puccinia graminis* f. sp. *tritici*), and powdery mildew (incited by *Blumeria graminis* f. sp. *hordei*) were observed. While other characteristics, particularly flavor (real or perceived), are important for the public acceptance of locally grown barley, the ability to perform well agronomically is a prerequisite for any barley variety grown in the Willamette valley.

When using traditional breeding methods, the traits necessary to grow and market a crop in a particular climate are identified in individual genotypes from large germplasm collections. A more modern approach is to use a combination of genotypic and phenotypic information to focus on specific genome regions, by identifying significant marker-trait associations. While this approach has the ability to identify the location of genes or regulatory elements in a genome with relative ease, difficult choices must be made that often balance the ability to identify a specific causative sequence, the level of precision, the complexity of the analysis, and the work that needs to be conducted to characterize genotypes.

Population structure is perhaps the largest hurdle to overcome when attempting to identify significant marker-trait associations in a collection of accessions. If not considered, it can lead to a high number of false positives. For example, when testing

markers for association with row type in barley, many markers are correlated with row type, simply because 2-row and 6-row germplasm tends to have different genetic backgrounds. However, only several loci can actually cause barley to have 2-rowed or 6-rowed spikes (Muñoz-Amatriaín et al. 2014). Initially, to avoid false-positives, bi-parental mapping populations were used, which have no structure, thereby eliminating the need to account for it. However, this requires mapping populations to be developed for each study, and the approach suffers from relatively low resolution. More recently, Genome-Wide Association Studies (GWAS), which uses a mixed-model approach to account for the genetic structure present in existing lines, has gained popularity (Cantor et al. 2010; Evangelou and Ioannidis 2013; Muñoz-Amatriaín et al. 2014).

Low levels of recombination can enable genes to be detected by markers that are several cM away. However, this comes at the expense of the ability to find the precise location of a given gene. For example, the frequency of recombination events in a bi-parental mapping population derived using doubled haploid technology is only $n/(100cM)$, where n is the number of genotypes in the population. This means that for a population of 200 genotypes, it would typically be theoretically impossible to pinpoint the location of a gene to less than a 0.5cM region, due to an absence of recombination events.

The number of markers to use in GWAS raises an unlikely trade-off. Generally, using more genetic markers increases the probability that least one marker falls close to each target locus. However, when the number of markers used in an analysis begins to be larger than the number of independent tests that can be conducted in a genome, standard adjustments for multiple comparisons tend to over-adjust, because they assume that each marker is an independent test. Notably, while the probability of a true positive occurring is depressed at both of these extremes, the probability of a false-positive occurring is always equivalent to alpha (assuming population structure was correctly accounted for). A promising solution to this is a Bayesian approach, which instead adjusts for the number of independent tests that could be made. However, it is methodologically challenging do this, and most software currently available for GWAS is not equipped to conduct a Bayesian analysis (Evangelou and Ioannidis 2013; Cantor et al. 2010).

Statistically, increasing the population size may be the most surefire way to improve power in GWAS. However, this can be limited by the resources available to a study, especially for traits that are difficult to phenotype.

This thesis consists of three manuscripts, in chapters 2, 3, and 4: one concerning how to select subsets of genotypes that are efficient at detecting Quantitative Trait Loci (QTL) and rare traits, one investigating the genetic controls of tocochromanol (vitamin-E) synthesis in barley, and finally one describing Alba, a barley variety that was recently bred and released for the Pacific Northwest.

Briefly, the second chapter builds on the method used to select subsets of genotypes presented by Muñoz-Amatriaín et al. (2014). This resulted in two computationally efficient methods that will be available from the Comprehensive R Archive Network (CRAN). One function, which selects subsets based on the Polymorphism Information Content (PIC) criteria, returns identical results to the one presented by (Muñoz-Amatriaín et al. 2014), but uses an algorithm that is much more computationally efficient. This function was shown to be beneficial in rare-trait discovery, but not in GWAS. A second function is based on a criterion that we call the "Mean of Transformed Kinships" (MTK), which has not been previously described. This function was shown to be beneficial when used to select subsets for GWAS, but not for rare-trait discovery. This chapter will be submitted to Germplasm Resources and Crop Evolution.

In the third chapter, we describe GWAS for tocochromanol synthesis. Grain from 1534 barley genotypes, representing germplasm from eight breeding programs was characterized for each of the eight tocochromanol forms (which include vitamin E and seven similar antioxidants). We identified thirteen QTL. The positions of these QTL were then compared to sequences in the barley genome that were homologous to genes known to be associated with tocochromanol biosynthesis to identify five candidate genes. This chapter will be submitted to Plos One.

In the fourth chapter, we describe the six-rowed, winter feed barley "Alba" in a germplasm release. Alba was shown to have complete or nearly-complete resistance against both barley stripe rust and scald. Overall, Alba demonstrated few statistically

significant differences when compared to "Strider," which was a feed variety in these trials. However, it did have a non-significant yield advantage over Strider in high-rainfall environments, making it the "better bet" of these two varieties for yield in high-rainfall environments, when only considering these data. Alba may have applications for craft maltsters in these environments. This chapter was published in the Journal of Plant Registrations.

A comparison of Polymorphism Information Content and Mean of Transformed Kinships as criteria for selecting informative subsets from the USDA Barley Core Collection

Ryan C. Graebner[1], Patrick M. Hayes[1], Christina H. Hagerty[2], Alfonso Cuesta-Marcos[1]

[1]Department of Crop and Soil Science, Oregon State University, Crop Science Building, Corvallis, Oregon, United States of America
[2]Department of Botany and Plant Pathology, Oregon State University, Cordley Hall, Corvallis, Oregon, United States of America

Abstract

Recent advances in genetic technologies have given researchers the ability to characterize genetic marker data for large germplasm collections. While some studies are able to capitalize on entire germplasm collections, others, especially those that focus on traits that are difficult to phenotype, instead focus on a subset of the collection. Typically, subsets are selected using phenotypic or geographic data. One major hurdle in identifying favorable subsets is selecting a criterion that can be used to quantify the value of a subset. This study compares two such criteria, Polymorphism Information Content, and a new criterion based on kinship matrices, which will be called the Mean of Transformed Kinships. These criteria were explored in terms of their ability to select subsets that are favorable for Genome Wide Association Studies, and in their ability to select subsets that contain a high number of rare phenotypes. Using phenotypic and genotypic data that has been amassed from the USDA Barley Core Collection, evidence was found to support the hypotheses that subsets based on the Mean of Transformed Kinships were well-suited to select subsets intended for Genome-Wide Association Mapping, but the same was not found for Polymorphism Information Content. Inversely, evidence was found to support the hypothesis that subsets based on Polymorphism Information Content were well-suited to select subsets intended for rare-phenotype discovery, but the same was not found for subsets selected using the Mean of Transformed Kinships criterion. Tools to select

subsets using these two criteria have been released in the R package "GeneticSubsetter."

Introduction

Global efforts to preserve the genetic diversity of agriculturally important crops have resulted in a range of valuable germplasm collections. Projects screening germplasm collections for novel phenotypes and genes often do not have the resources to sample every accession in a given collection, so subsets of the total collections are made. Until recently, these subsets were generally made on the basis of phenotype and geographic origin of accessions, with the goal of maximizing genetic diversity (Holbrook et al. 1993; Mahajan et al. 1996; Upadhyaya et al. 2001; Upadhyaya et al. 2009; Zewdie et al. 2004). However, with the advent of high-throughput genotyping, complete sets of genotypic data are increasingly common for large germplasm collections (Muñoz-Amatriaín et al. 2014). This enables researchers to directly observe genetic diversity, as opposed to estimating it with phenotypic or geographic information.

Two principal components to any subsetting technique are the criterion used to quantify the value of a specific subset, and the method used to find the optimum subset, as judged by that criterion. For smaller collections, the method used to identify a favorable subset could be to simply test all possible subsets. However, this quickly becomes unfeasible as population sizes grow. For instance, in a circumstance where 100 accessions need to be chosen from a collection of 1000 accessions, there could be $6.385*10^{139}$ possible subsets. Given the large number of subset combinations, alternative methods are needed to reach a good, or ideally the best, subset for a given criterion. Without proper subsetting techniques, important phenotypes could be omitted, making them unavailable to breeders.

To quantify a population's diversity, Polymorphism Information Content (PIC) values were calculated with the following equation:

$$PIC = 1 - \frac{1}{m} \sum_{l=1}^{m} \sum_{i=1}^{n} f_{li}^2$$

where f_{li} is the frequency of the lth locus for m loci, and the ith allele for n alleles. This

equation was modified from an equation described by Smith et al. (1997). This is similar to the equation used to calculate heterozygosity, except that heterozygosity is defined as the probability of a single individual being heterozygous at any one allele under Hardy-Weinberg equilibrium, which is irrelevant for inbred lines (due to the near-complete fixation of alleles at each locus). Generally speaking, for bi-allelic markers, mean PIC values for a population can range from 0.0, where all markers are monomorphic, to 0.5, where the frequency of both alleles is 0.5 for every marker. While PICs are most frequently used to quantify the diversity of an existing set of genotypes, they have also been used to identify informative subsets in the program PowerMarker (Liu and Muse 2005), and in a study characterizing the USDA Barley Core Collection (Muñoz-Amatriaín et al. 2014). Because a complete description of the methods used by PowerMarker to identify subsets is apparently no longer available, it will not be evaluated in this study.

One shortcoming of the PIC criterion is that it does a poor job at removing similar or even identical genotypes from a population. This allows for redundancy in the selected subset, which should be detrimental to maximizing the discovery of rare traits, and conducting GWAS. To address this, an alternative approach has been developed based specifically on kinship matrices, where kinship values are risen to the power of 10 in order to increase the weight of pairs of similar genotypes. Subsets are compared by simply comparing the mean of these modified kinship values, or the Mean of Transformed Kinships (MTK).

The USDA Barley Core Collection provides an excellent opportunity to test these subsetting criteria. This collection contains 2,417 landraces, breeding lines, and cultivars that have been collected from around the world (Muñoz-Amatriaín et al. 2014). This collection was selected from the larger National Small Grains Collection (NSGC) for barley, by randomly selecting accessions based on the logarithm of the total number of entries from each country of origin, ensuring that a minimum of one accession from each country be included in the core collection (Muñoz-Amatriaín et al. 2014).

Our objectives in this study were to assess the utility of these subsetting criteria, both in terms of their ability to select subsets that are favorable for Genome-Wide

Association Studies (GWAS), and in their ability to select subsets that contain a high number of rare phenotypes. The functions used to identify favorable subsets in this study are available in the R package "GeneticSubsetter."

Methods

Description of Functions

To calculate the MTK for a set of genotypes, a kinship matrix was made using the "A.mat" function in the R package rrBLUP (Endelman et al. 2011), using the default options. Due to the way the A.mat function calculates kinship matrices, negative kinship values are created, and the cell describing an accession's kinship with itself has a degree of variability. To remove negative values, the kinship matrix was scaled to values ranging from zero to two (where the relative distance between kinship values were constant, and zero and two were the lowest and highest kinship values for the particular set of genotypes, respectively). Also, the diagonal values in the kinship matrix (the values describing a genotype's kinship with itself) were removed. Each value in the kinship matrix was raised to the power of 10. Finally, the mean of the values in the resulting transformed kinship matrix was calculated, to find MTK, which quantifies the extent to which a subset contains closely related accessions. To make this criterion computationally feasible for subsetting, transformed kinship values were calculated once using the whole population, and then subsets of the matrix are used to calculated MTK for subsets of genotypes.

The core functions, "SubsetterPIC" and "SubsetterMTK," in the R package GeneticSubsetter, remove one genotype at a time, on the bases of which genotype's removal will result in the highest PIC, or the lowest MTK, respectively. These functions return a list of ranked genotypes, from which subsets of any size can be obtained by taking the top lines. The SubsetterPIC function returns a list identical to the list returned by the Excel macro discussed in Muñoz-Amatriaín et al. (2014). However, the SubsetterPIC function uses a more efficient algorithm to identify this ranking, giving it a considerable advantage in computing time over the Excel macro.

Currently, these functions are only designed for homozygous, bi-allelic markers.

However, the concepts used to calculate PIC and MTK in these functions could be applied for heterozygous and poly-allelic markers.

Analysis of effect on GWAS

Subsets made using these criteria were assessed by their ability to identify simulated Qualitative Trait Loci (QTL) imbedded into heading date data from the USDA Barley Core Collection using GWAS. To create the simulated QTL, twenty Single Nucleotide Polymorphisms (SNPs) from the barley iSelect Illumina SNP platform (Muñoz-Amatriaín et al. 2014) were chosen at random, and the heading date data for genotypes with the "A" form of each allele was increased by five days (CITE). Genotypes were ranked using the SubsetterPIC and SubsetterMTK, and 200 times randomly, to make a total of 202 Set of Nested Subsets (SNSs). Each of these SNSs consisted of a series of subsets, one for each multiple of 50 genotypes between 150 and 1800 genotypes (a total of 35 subsets for each SNS), where each accession in a given subset was also present in the subsets that were larger than it in the given SNS. GWAS was conducted for each subset in each of the 202 SNSs. GWAS was conducted using the "GWAS" function in the R package rrBLUP, using the default parameters (Endelman et al. 2011).

Within each subset size, SNSs were assigned a rank based on how many simulated QTL were detected, relative to subsets of that size within other SNSs. The mean of a SNS's ranks across all 35 tested subset sizes was used to quantify a particular SNS's performance against other SNSs. Simple methods for combining p-values would not be appropriate here, as two subsets of a similar size from a single SNS are not independent from each other. While many random SNSs can be obtained from this collection, the SubsetterPIC and SubsetterMTK functions are determinate in nature, and were only able to return one SNS each. To test if a particular subsetting function returned a SNS that was better than a random SNS (with $p<0.05$), the non-random SNS was compared to the 200 random SNSs. A non-random SNS performing either better or worse than 97.5% of the random subsets would correspond to $p<0.05$, in which case it would be decided that there was a significant difference between the SNS made using the particular

criterion and the random SNSs, within the context of this collection.

Analysis of effect on Rare Phenotype Discovery

Eleven extreme phenotypes were identified, where extreme phenotypes were defined as either the highest or lowest ~2% of accessions for each given trait (Table 1). For example, the trait "plant height" had two sets of accessions that held an extreme phenotype: the 25 accessions that were shorter than 66 cm, and the 23 accessions that were taller than 117.5 cm. These extreme phenotypes were used to test whether these subsetting criteria were beneficial for the discovery of rare phenotypes. To circumvent the limitations of only having access to one large collection with extensive phenotypic and genotypic data available, we used 1,099 genotypes with thorough phenotypic information available to make 1,000 random "mini-sets" of 100 lines. While these mini-sets have similar population structures, pairs of mini-sets share an average of only 9.1% of their genotypes, making their results essentially independent from each other. Each mini-set was further subsetted three times to a subset size of 10 genotypes, once using the SubsetterMTK function, once using the SubsetterPIC function, and once randomly. Each 10-genotype subset was quantified by how many of the original 10 rare alleles were present in the final subset. Paired t-tests were used to determine if either the SubsetterPIC or the SubsetterMTK functions were able to identify subsets with more rare alleles than randomly selected subsets.

Phenotypic and Genotypic Information

To test the PIC and MTK criteria, we used phenotypic and genotypic data collected from the USDA Barley Core Collection. The collection was previously genotyped, using a barley iSelect Illumina SNP platform, which included 7,842 SNPs (Muñoz-Amatriaín et al. 2014). Data concerning heading date were collected in Corvallis, Oregon in 2012 (Muñoz-Amatriaín et al. 2014). All other phenotypic data were collected from the USDA-ARS Germplasm Resources Information Network website. A total of 1,852 lines had complete heading date and genotypic information available, and a total of 1,099 lines had complete information available for genetic markers and each of eleven

rare phenotypes assessed in this study.

<u>Results</u>

GWAS

 A Set of Nested Subsets (SNS) made using SubsetterMTK performed better than 199 out of the 200 random SNSs (Figure 2). This corresponds to a p-value of approximately 0.01, providing strong evidence that subsets made using the MTK criterion are more favorable for GWAS within the context of the USDA Barley Core Collection. A SNS made using SubsetterPIC performed better than 131 out of the 200 random SNSs, corresponding to a p-value of approximately 0.69. While this presents no evidence that subsets made using the PIC criterion are more favorable for GWAS for this specific germplasm collection, given the extremely low power of this test, this criterion may still have a benefit to subsetting for GWAS that was undetectable in this study.

Rare Phenotypes

 We found significant evidence that subsets identified using the PIC criterion were more likely to contain rare phenotypes than random subsets in the USDA Barley Core Collection (p<0.0001). However, we found no evidence that the same was true for subsets identified using the MTK criterion (p=0.83). On average, random subsets of ten genotypes from the USDA Barley Core Collection contained 1.974 rare or extreme phenotypes. In contrast, subsets of 10 genotypes selected from 100 random genotypes using the SubsetterPIC function contained an average of 2.456 rare or extreme phenotypes, representing a 24% increase over random subsetting.

Structure

 Principal Component Analysis (PCA) plots showing the full collection, a completely random subset, a subset of 200 genotypes made using the PIC criterion, and a subset of 200 genotypes made using the MTK criterion (Figure 1). These figures demonstrate how the PIC and MTK criteria differ in terms of the resulting population structure. While it appears that subsets made using SubsetterPIC maintain the general

structure of the full collection, the number of individuals in each group appears to differ from the random subset. This is likely because the PIC criterion will weight groups by their contribution to the subset's diversity, while the random subset weights groups by purely by how well they are represented in the full collection. Using SubsetterMTK instead appears to result in a population with very little structure. Interestingly, SubsetterMTK appears to prioritize genotypes that fall in the middle of the PCA plot, presumably because these genotypes are in fact the least related to the rest of the collection.

Discussion

 Within the context of the USDA Barley Core Collection, these results demonstrate that PIC is an acceptable subsetting criterion for rare phenotype discovery, and that MTK is an acceptable subsetting criterion for GWAS. Due to the limited number of core collections that have been extensively phenotyped and genotyped, it is currently difficult to assess these benefits on other sets of accessions.

 The PIC and MTK criteria are promising in their ability to help avoid a loss in power when making a subset for a specific purpose- either rare-phenotype discovery or GWAS. For dual-purpose subsets, it may be beneficial to use a combination of these two criteria (i.e. remove 100 lines based on MTK, then another 100 lines based on PIC). This approach may be able to maintain more than half of the benefit of only using one criteria, because these functions should first remove the lines that contribute very little to the collections diversity, or that are essentially redundant, depending on the criteria used. Alternatively, a hybrid criterion could be used, which considers how each accession's removal would affect both the PIC and the MTK values for the subset.

 We hope that the functions presented in the R package GeneticSubsetter can help to leverage "big data" in a way that substantially increases the efficiency of GWAS and rare-phenotype discovery: two tasks which are routinely conducted by plant breeding programs. While the R package GeneticSubsetter is currently only equipped to address homozygous accessions, we look forward to the possibility of building on these functions to expand this package's utility to species that tend to be heterozygous, including humans

and other animals.

References

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250-255.

Holbrook CC, Anderson WF, Pittman RN (1993) Selection of a core collection from the U.S. germplasm collection of peanut. Crop Sci 33:859-861.

Liu K, Muse SV. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.

Mahajan RK, Bisht IS, Agrawal RC, Rana RS (1996) Studies on South Asian okra collection: Methodology for establishing a representative core set using characterization data. Genet Resour Crop Evol 43:249-255.

Muñoz-Amatraín M, Cuesta-Marcos A, Endelman JB, Comadran J, Bonman JM (2014) The USDA barley core collection: genetic diversity, population structure, and potential for genome-wide association studies. PloS One 9:e94688.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ et al. (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays L.*): comparisons with data from RFLPS and pedigree. Theor Appl Genet 95:163-173.

Upadhyaya HD, Bramel PJ, Singh S (2001) Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci 41:206-210.

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG, Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci 49:1769:1780.

Zewdie Y, Tong N, Bosland P (2004) Establishing a core collection of Capsicum using a cluster analysis with enlightened selection of accessions. Genet Resour Crop Evol 51:147-151.

Figures
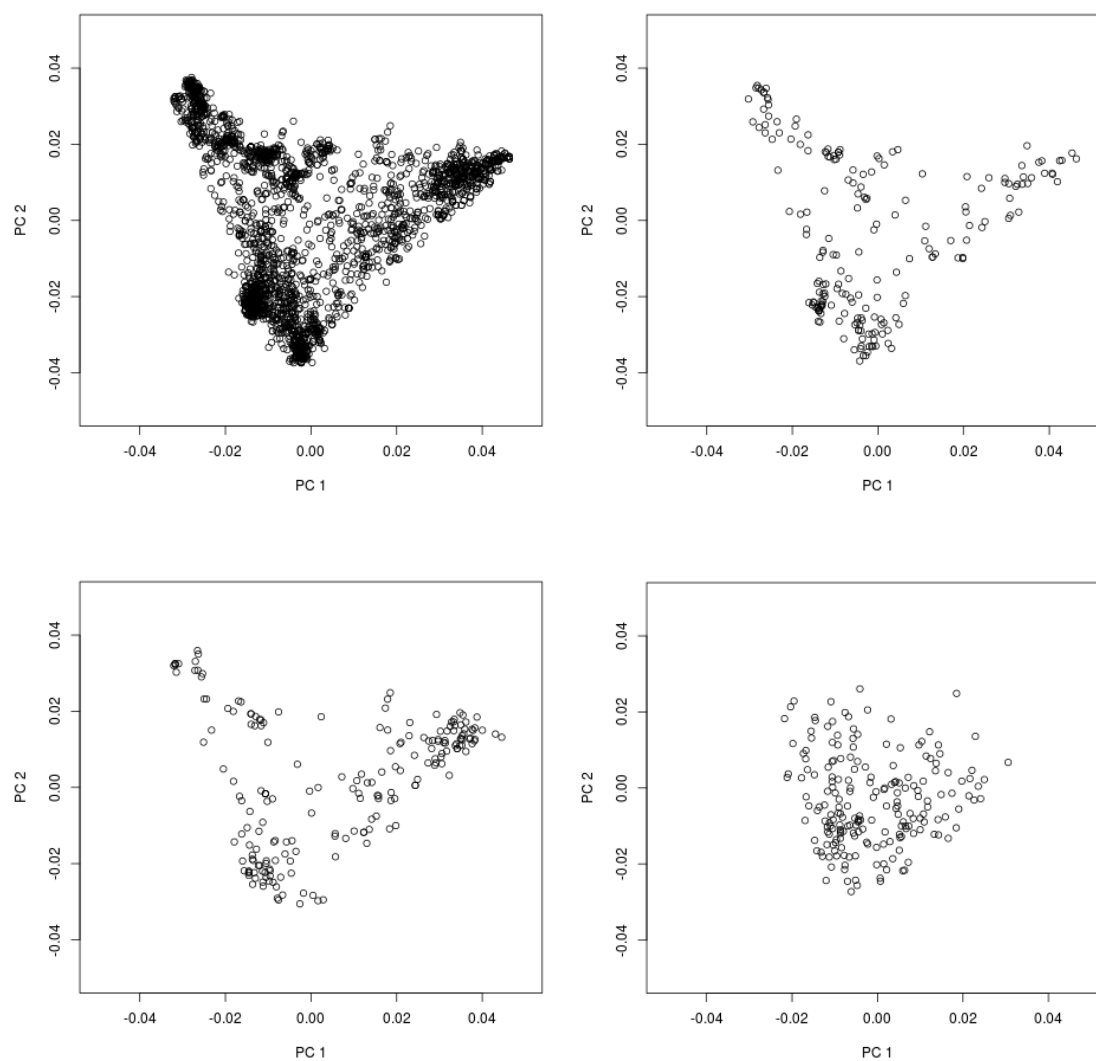
Figure 1. PCA plots of the USDA Barley Core Collection (top left), a random subset (top right) a 200-genotype subset made using the SubsetterPIC function (bottom left), and a 200-genotype subset made using the SubsetterMTK function (bottom right).
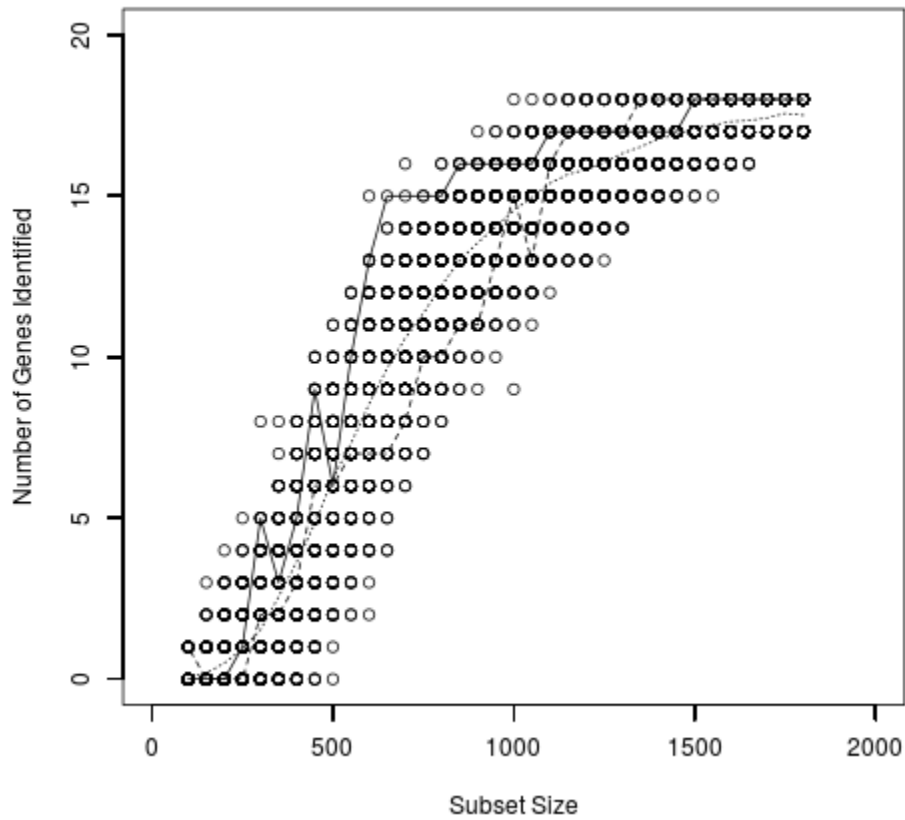
Figure 2. Comparison of subsets identified by SubsetterMTK (solid line), SubsetterPIC (dashed line), and 200 subsets that were randomly selected (circles, with dotted line showing mean). The x-axis shows the size of each subset, and the y-axis show the number of artificial genes detected by that subset.

Tables

Trait	Definition of Phenotype	Rare Phenotype Frequency
Spot Blotch Resistance	<4 on a 1-9 Scale	32
Russian Wheat Aphid Resistance	<7 on a 1-9 Scale	14
Early Heading	<31 Days After First Heading Date	18
Low Plant Height	<66 cm	25
High Plant Height	>117.5 cm	23
Low Beta-Glucan	<3.34%	20
High Beta-Glucan	>7.04%	20
Low Protein	<9.075%	20
High Protein	>18.15%	20
Low Kernel Weight	<31.75 mg	23
High Kernel Weight	>60.25 mg	19

Table 1. Rare phenotypes used in this study, the criteria to define the rare phenotypes, and the number of lines that fit these criteria in the set of 1,099 genotypes used to compare the abilities of subsetting criteria in rare-phenotype discovery.

Quantitative trait loci associated with the tocochromanol (vitamin E) pathway in barley

Ryan C. Graebner*, Mitchell Wise*, Alfonso Cuesta-Marcos, Matthew Geniza, Tom Blake, Victoria C. Blake, Shiaomen Chao, David J. Hole, Rich Horsley, Pankaj Jaiswal, Don Obert, Kevin P. Smith, Stephen Ullrich, Patrick M. Hayes

* These authors contributed equally to this work.

Abstract

In this study, the Genome-Wide Association Studies approach was used to detect Quantitative Trait Loci associated with tocochromanol concentrations in a panel of 1,466 barley accessions. All major tocochromanol types- α-, β-, δ-, γ-tocopherol and tocotrienol- were assayed. We found 13 single nucleotide polymorphisms associated with one or more of these tocochromanol concentrations in barley, seven of which were within 2.5 cM of sequences homologous to cloned genes associated with tocochromanol production in barley and/or other plants. The discovery of these loci could aid future breeding efforts to develop barley varieties with higher tocochromanol concentrations. At current recommended daily consumption amounts, barley would not be an effective sole source of vitamin E. However, it could be an important contributor in the context of whole grains in a balanced diet.

Keywords: Barley, GWAS, QTL, quantitative trait loci, tocochromanol, tocopherol, tocotrienol, vitamin E.

Introduction

The tocochromanols - including α-tocopherol (αT), α-tocotrienol (αT3), β-tocopherol (βT), β-tocotrienol (βT3), δ-tocopherol (δT), δ-tocotrienol (δT3), γ-tocopherol (γT) and γ-tocotrienol (γT3) forms - are credited with protecting polyunsaturated fatty acids from lipid peroxidation (Bruno et al. 2005). Tocopherol and tocotrienol fractions

are differentiated by the level of saturation on the polyprenyl. Of these eight tocochromanol forms, αT and γT receive the most attention: αT because it is the most concentrated in human plasma, and γT because it is the most abundant in many typical human diets (Traber and Arai 1999). While all tocochromanol forms have similar anti-oxidant properties and are in some cases referred to, cumulatively, as "vitamin-E", αT is the only tocochromanol form that meets the Recommended Daily Allowance (RDA) for vitamin-E, so the term "vitamin-E" commonly refers specifically to αT.

Despite the well-established nutritional requirement of tocochromanols for reproductive health and normal neurological development in mammals (Frank 2004), the precise physiological function of these compounds remains elusive. The scientific literature is replete with laboratory studies on the nutritional benefits of tocochromanols, particularly with respect to cardiovascular disease (Gey 1991). Oddly, depending on the specific health risk, human epidemiological studies have been equivocal (Cordero et al. 2010), with some reporting that the overall impact of αT is positive (Sanyal et al. 2010; Gey 1991), negative (Klein et al. 2011), or relatively neutral (Roberts et al. 2010). In one exceptionally large trial, in which 39,876 apparently healthy women were administered either vitamin E or a placebo over an average of 10.1 years, very little evidence was found that vitamin E reduced the risk of either cardiovascular diseases or cancer (Lee et al. 2005). However, most of the current literature is based on experiments where supplements, in the form of natural or synthetic αT, were used to test the effects of vitamin E on human health. High doses of αT are known to inhibit absorption of other tocochromanols in humans (Handelman et al. 1985; Traber et al. 1992), and these effects may be long lasting (Huang and Appel 2003). More research is needed to fully understand the effects of consuming tocochromanols in a natural form (i.e. in whole grains).

In addition to their possible implications for human health, tocochromanols play an important role in plant stress tolerance. One key function is that tocochromanols help to protect lipid membranes in the photosynthetic machinery from a range of oxidative stresses, primarily by deactivating $^1O_2$ and $OH^\cdot$ reactive oxygen species (Munné-Bosch 2005). When used to scavenge lipid peroxyl radicals in plants, tocochromanols must be

restored by ascorbate (vitamin C) to re-gain functionality, and in scavenging $^1O_2$, the anti-oxidant is irreversibly damaged (Munné-Bosch 2005). The functions of other tocochromanols in plant physiology remain to be elucidated.

To date, there have been two major studies of the genetic controls of tocochromanol synthesis in barley. In one study, the cDNA sequence encoding homogentisate geranylgeranyl transferase (*HGGT*), an enzyme necessary for tocotrienol synthesis, was isolated in barley (Cahoon et al. 2003). In the same study, the barley *HGGT* sequence was used for *Agrobacterium*-mediated transformation of maize, resulting in a six-fold increase of tocotrienols in the seed. However, the gene encoding this cDNA was not assigned a linkage or physical map position. In a more recent study (Oliver et al. 2014), a bi-parental mapping population was used to identify three Quantitative Trait Loci (QTL) associated with the concentrations of one or more tocochromanol forms in barley, one on chromosome 6H, and two on chromosome 7H. The QTL on chromosome 6H was attributed to *VTE4*, and one of the QTL on chromosome 7H was attributed to either *HGGT* or *VTE2*, based on orthology between rice and barley.

The availability of a comprehensive linkage map and a genome sequence in barley makes it possible to assign a map position to *HGGT* and to sequences for other genes in the tocochromanol pathway, using a Genome-Wide Association Studies (GWAS) approach. GWAS is now widely-used in a range of crop plants and is a powerful tool for rapidly detecting QTL and possibly candidate genes (Kang et al. 2010; Cantor et al. 2010). In barley, GWAS has been used to identify QTL related to flowering time (Berger et al. 2012; Muñoz-Amatriaín et al. 2014a), disease resistance (Berger et al. 2012), and food quality (Berger et al. 2012; Mohammadi et al. 2014).

Our objectives were to a) quantify the concentration of each tocochromanol form in cultivated barley using accessions from eight US spring barley breeding programs, b) identify QTL in the barley genome associated with the concentration of each tocochromanol form and fraction, and c) use identified QTL in conjunction with the barley genome sequence to identify candidate genes.

Methods

 This research was based on 1,534 spring-habit barley accessions from the Barley Coordinated Agriculture Project (Barley CAP), a predecessor to the Triticeae Coordinated Agricultural Project (TCAP; http://www.triticeaecap.org/, verified 26 October 2014). The "Barley CAP I" and "Barley CAP II" germplasm sets consisted of elite breeding lines and varieties from ten breeding programs participating in the Barley CAP: Montana State University (MT), North Dakota State University two-row and six-row (N2 and N6), the USDA-ARS program based at Aberdeen, Idaho (AB), the University of Minnesota (MN), Utah State University (UT), Washington State University (WA), and Busch Agricultural Resources Inc. (BA).

 The 1,534 spring barley accessions were grown, one time per accession, over a two year period (2006 and 2007) at Bozeman, Montana, USA. The crop was irrigated in 2006, but not in 2007. Plots consisted of four rows, and were 1.3m long.

 Tocochromanols were analyzed and quantitated using a modified saponification method (Fratianni et al. 2002). Grain, approximately 1 g, was ground in a Retsch ZM-1 mill (Haan, Germany) and an aliquot (approximately 0.5 g) was weighed and the weight recorded. The freshly-ground sample was then extracted by addition of 0.5 ml 10M KOH, 0.5 ml 95% ethanol, 0.5 ml 0.15M NaCl and 1.25 ml of a 0.5M solution of pyrogallol (in 95% ethanol) and shaken in a water bath at 70 ℃ for 30 min., vortexing every 10 min. The tubes were cooled on ice and an additional 3.75ml of 0.15M NaCl was added. This suspension was extracted twice with hexane/ethyl acetate (9:1 v/v) by vortexing and centrifuging at 1000g for 5 min and transferring the supernatant to a glass test tube. The combined organic phase was reduced to dryness in a Thermo-Savant SPD1010 speed-vac system (Asheville, NC) at 45 ℃. The dried extract was re-suspended in 1.0 ml hexane and centrifuged to remove particulates prior to analysis by High Performance Liquid Chromatography (HPLC). For HPLC analysis, each sample was analyzed with a Shimadzu LC-5a HPLC (Kyoto, Japan) using a 4.6×250 mm, 5 μm Adsorbosil silica column (Grace Co., Deerfield IL.) with an isocratic mobile phase at a flow rate of 2.0 ml/min. Samples from the barley CAP I germplasm were separated using a mobile phase of 0.5% isopropanol in hexane. Unfortunately this solvent did not effectively separate the

γT and the βT3 content, thus another solvent system was developed consisting of 2% ethylacetate and 2% dioxane in hexane, which did separate these two congeners, and was used for the Barley CAP II germplasm. Fluorescence detection was employed using a Shimadzu RF-10A spectrofluorometer with excitation at 295 nm and detection at 330 nm. Peaks were integrated and compared to tocochromanol standards. Tocotrienols were quantitated using the standard curve developed for the corresponding tocopherol (Thompson and Hatina 1979). Tocochromanol data for germplasm arrays are available at The Triticeae Toolbox (T3) (http://triticeaetoolbox.org/, verified 13 October 2014).

Barley accessions in the "Barley CAP I" and "Barley CAP II" germplasm arrays were genotyped for 3,072 single nucleotide polymorphism (SNP) markers with two GoldenGate Olionucleotide Pool Assays (OPAs), as described by Close et al. (2009) and Szücs et al. (2009). The genotyping was conducted at the USDA-ARS Small Grains Genotyping Center in Fargo, North Dakota. After excluding markers with missing data and markers that were cosegregating in this set of germplasm, 2,204 of the 3,072 SNP markers from the two OPAs were used in this analysis. Of the 1,534 accessions genotyped, 68 were excluded from the analysis because of missing genotypic data. Therefore, the GWAS is based on 1,466 barley accessions. SNP data was retrieved from The Triticeae Toolbox (T3) (http://triticeaetoolbox.org/, verified 13 October 2014) (Blake et al. 2012).

Linkage map positions from the barley consensus map (Muñoz-Amatriaín et al. 2011) were used to identify the position of SNP markers in this analysis. One SNP marker that was significant in this analysis, 11_20311, had not been assigned a position in this consensus map. Therefore, its position in the barley genome sequence (The International Barley Genome Sequencing Consortium 2012), relative to SNPs with known linkage map positions, was used to approximate its cM position. Linkage Disequilibrium (LD) between these markers was calculated using the "Measure.R2S" function in the R package "LDcorSV." The breeding program of each accession's origin was used to partially account for population structure for LD calculations in this panel.

An R script based on the "GWAS" function in the package rrBLUP version 4.1 (Endelman 2011), with minor modifications, was employed using R version 3.0.1, to

conduct GWAS. Markers with a minor-allele frequency below 5% or with more than 10% missing data, and genotypes with more than 10% missing data were excluded. The Efficient Mixed-Model Association eXpedited (EMMAX) method, using a kinship matrix and five principal components that were included as fixed effects, was used to efficiently account for genetic structure in this set of accessions (Kang et al. 2010). P-values were adjusted to account for multiple comparisons using the False Discovery Rate (FDR), developed by Benjamini and Hochberg (1995). In instances where multiple closely linked markers were significant, and one of the markers was more significant than every other marker in that region for every significant trait, only the most significant marker was reported. Marker effects were based on Best Linear Unbiased Estimates (BLUEs).

Data from 2006 and 2007 were combined into a single analysis, using a fixed effect to account for differences across years, as described by Evangelou and Ioannidis (2013). This method of combining years was also used to combine barley food-quality data from an overlapping set of trials by Mohammadi et al. (2014).

Positional information, Gene Orthology (GO) annotations (The Gene Ontology Consortium 2000), and InterPro assignments (Hunter et al. 2011) were obtained for barley genes (ISBC_1.0.030312v22) through the Gramene version of the BioMart (Jaiswal 2011). This information was scanned for genes that could be involved in the tocochromanol biosynthesis pathway, using a set of keywords to identify promising candidates. This list was manually curated to remove genes that were identified by the automatic search, but after further review, were not determined to be associated with the tocochromanol biosynthesis pathway. A manual search was also conducted in which sequences from other species that are associated with the tocochromanol biosynthesis pathway were compiled from NCBI (http://www.ncbi.nlm.nih.gov/, verified 29 October 2014). For each of these sequences, the Basic Local Alignment Search Tool (BLAST) at the IPK Barley BLAST Server (Deng et al. 2007; http://webblast.ipk-gatersleben.de/barley/, verified 29 October 2014) was used to identify regions of the barley genome homologous to these sequences of interest from other species. Annotations for barley SNPs were obtained from HarvEST (http://harvest.ucr.edu/, verified 15 December 2014).

To determine the linkage group and cM positions of candidate genes, OPA SNP markers were aligned with the barley genome sequence using the BLAST-Like Alignment Tool (BLAT) (Kent 2002). The base-pair position of SNP markers in the barley genome was determined from the BLAT output by percent identity and level of significance. The positions of candidate genes relative to their flanking markers in the genome sequence were then used to calculate approximate cM positions.

Results

Phenotypic data

There were detectable concentrations of all tocochromanols in all germplasm in both years (Table 1; Figures 1, 2, and 3). Including both years, αT concentrations ranged from 6.8 mg/kg to 23.9 mg/kg and total tocochromanol (TTC) concentrations ranged from 30.87 mg/kg to 94.06 mg/kg. Considering all forms, the average concentrations of αT3 were the highest, and the average concentrations of δT were the lowest. Means and standard errors for all tocochromanol forms are presented in Table 1. An analysis of variance showed that both year and breeding program had significant effects on αT and TTC concentrations (Table 2). Row-type had a significant effect on both αT and TTC (Table 2). αT, αT3 and δT concentrations were higher in 2006 (irrigated) than in 2007 (dryland), whereas the reverse was true for βT, δT3 and γT3 and TTC (p<0.0001 for all comparisons). As noted in the Materials and Methods, βT3 and γT were not distinguished in the analysis of 2006 samples. Therefore, it is not possible to assess the effect of year/management practice on these forms. The breeding program with germplasm having the highest average tocochromanol concentration varied by year and tocochromanol form. For example, in 2006, germplasm from MT had the highest average concentration of αT and TTC whereas in 2007, germplasm from UT had the highest average concentration of αT and germplasm from USDA-ARS-Idaho had the highest average concentration of TTC. In both years, germplasm from the MN had the lowest average concentration of both αT and TTC.

GWAS and marker-trait associations

Q-Q plots indicate that the model with five principal components adequately accounted for population structure, thereby controlling false positives (Figures 4, 5). Principal component analysis identified row-type and breeding program as major drivers of structure in this set of germplasm (Figure 6). Thirteen SNP markers were significantly associated with one or more of the tocochromanol forms and/or fractions (Table 3; Figure 7). Two significant SNPs were identified on chromosome 1H - one at cM 109.8 (associated with total tocotrienol (TT3) and TTC), and the second at cM 127.6 (associated with βT). Two SNPs were identified on chromosome 6H – one at 58.3 (associated with δT), and one at cM 70.6 (associated with γT3). The remaining eight SNPs were on 7H and formed three groups: one at cM 1.1 (associated with γT3); two at cM interval 95.0 – 95.7 (associated with βT and δT); and five at cM interval 136.0 – 145.2 (associated with αT3, βT3, δT, δT3, γT, γT3, TT3, and TTC). There were no significant associations of SNPs with αT or total tocopherol (TTP).

None of the significant markers were within an Expressed Sequence Tag (EST) that had a description in HarvEST that was obviously associated with tocochromanol biosynthesis.

Candidate genes

Of the thirteen significant markers, seven were within 2.5 cM of at least one sequence homologous with genes known to be associated with the tocochromanol biosynthesis pathway in barley and/or other plants (Table 4). On 1H, candidate gene MLOC_16149, which may encode *VTE2* (as described by Collakova and DellaPenna (2001)), homogentisate geranylgeranyltransferase (as described by Cahoon et al. (2003)), or one of multiple enzymes upstream of geranylgeranyl diphosphate (a precursor to all tocochromanols; Cahoon et al. 2003), including farnesyl diphosphate synthase (as described by Matsushita et al. (1996)) and homogentisate farnesyltransferase (as described by Sadre et al. (2006)), was identified at cM 107.7- 2.1 cM from marker 11_20021. No candidate genes were identified within 2.5 cM of marker 11_10586.

On 6H, candidate genes were identified within 2.5 cM of the marker at cM 58.3:

MLOC_72891, MLOC_44750, and MLOC_66290. Each of these candidate genes is likely associated with a gene upstream of geranylgeranyl diphosphate. On 6H at cM 71.5, the candidate gene MLOC_13082, which may encode the enzyme *VTE4* (as described by Shintani and DellaPenna; 1998) was 0.9 cM from marker 12_30637.

On 7H, no candidate genes were identified within 2.5 cM of QTLs at cM 1.08 or interval cM 95.0 – 95.7. In the region from cM interval 136.0 – 150.4 two candidate genes were identified: MLOC_12567 encoding *HGGT*, and MLOC_37476 encoding either *VTE2* or one of multiple enzymes upstream of geranylgeranyl diphosphate.

Allele effects and distributions

As shown in Table 5, the best linear unbiased estimators (BLUEs) for allele effects reveal substantial phenotypic variation associated with allele substitutions at the significant SNPs. Both alleles at each significant SNP were present in most breeding programs (Table 6). The accessions from the USDA-ARS- ID program and UT had the highest levels of allelic diversity, never having less than 9% and 6% of the minor allele, respectively.

Discussion

Phenotypic variation for tocochromanols

Differences were observed in tocochromanol concentrations over the two years of this study, although this was confounded by the different germplasm arrays grown each year (Table 1; Figures 1, 2, and 3). While the 2006 growing season in Bozeman, Montana was relatively typical, the 2007 growing season was characterized by extremes, with 18.5 cm of snowfall recorded on May 29th, followed by a July that was possibly the hottest on record, and had little precipitation (National Weather Service; http://nws.noaa.gov/climate/local_data.php?wfo=tfx, verified 3 November 2014). This was further confounded by the fact that barley was irrigated in the 2006 growing season, but not in the 2007 growing season, possibly leading to differential stress. Oliver et al. (2014) speculate that moisture availability could have an effect on tocochromanol concentration, but were not able to separate the effect of moisture availability from

location, since one location was irrigated and the other not. We observed that some tocochromanol forms (βT, δT3 and γT3 and TTC concentration) were higher in the second year, while others were lower (αT, αT3 and δT). Oliver et al. (2014) further identified temperature as an important environmental effect, with a one degree overall drop in temperature significantly increasing TTC concentrations across years. However, in their study, the location with the lowest average temperatures did have lower TTC. Therefore, we can only concur with Oliver et al. (2014) that environment has an important effect. Reserve seed of the barley GWAS panel used is available and could be a useful resource for experiments specifically designed to address the role of environmental factors on tocochromanol concentrations in barley.

Focusing on germplasm, the sample of 1,466 elite accessions that we analyzed followed the same trends reported in other barley studies (Oliver et al. 2014, and reviewed therein) in terms of the relative concentrations of specific tocochromanol forms, with αT3 generally being highest, and δT generally being the lowest. In general, the concentrations of tocochromanol forms in this study were comparable to previous studies (Peterson and Qureshi 1993; Oliver et al. 2014), suggesting that current barley cultivars grown in the western United States typically have average αT and TTC values of 10.83 mg/kg to 19.12 mg/kg, and 53.28 mg/kg to 75.92 mg/kg, respectively. The highest αT concentration observed was 1.72 times higher than the average αT concentration, and the highest TTC concentration observed was 1.49 times higher than the average TTC concentration. This finding supports one of the advantages of GWAS over biparental populations: by removing the limitation that all germplasm studied must be derived from the same two parents, more diverse and potentially more relevant germplasm can be sampled.

Exploring this panel more deeply, there were differences in tocochromanol concentrations between programs. While the relative ranking of these programs tended to vary between traits and years, the germplasm from the breeding program at MN, which is exclusively 6-row, consistently had the lowest concentrations for both αT and TTC concentration. However, the differences between 2-row and 6-row barley were generally smaller than those due to programs or years (Table 2). Differences between breeding

programs may be a result of selection for or against tocochromanol concentrations in source environments. However, another explanation could be that germplasm that is not adapted to Montana in terms of phenology and resistance to biotic and abiotic factors may experience additional stress, thus altering tocochromanol concentrations. At this time, the effects of these factors on tocochromanol concentrations are not fully understood.

Given the observed values for tocochromanol forms, a key question is whether barley can be a viable source of these compounds for human nutrition. The answer to this question is complicated by the fact that a RDA has only been established for αT, which is 15 mg/day for adults (National Institute of Health Office of Dietary Supplements, http://ods.od.nih.gov/factsheets/VitaminE-HealthProfessional/#h2; verified 29 October 2014). Using the accession with the highest αT concentration, 06MT-55 with 23.88 mg/kg αT, a healthy adult would need to consume approximately 628 g of barley (dry weight) per day to meet their RDA for αT. Therefore, it is not realistic to imagine barley as a sole source of αT in human diets. Nevertheless, other forms of the tocochromanols are reputed to provide nutritional benefits. γT, for example, is a superior detoxification agent of reactive nitrogen species than αT (Cooney et al. 1993), an important consideration in chronic inflammation and for smokers or individuals subject to air pollution. Furthermore, in cellular assays γT was shown to provide neuroprotective effects at concentrations 4 to 10 fold lower than typically found in human plasma (Khanna et al. 2003). Tocotrienols may provide several nutritional benefits not demonstrated by tocopherols (Theriault 1999; Sen 2007). For example, Qureshi found that non-polar extracts from barley reduced cholesterolgenesis in chicks and identified the active ingredient as αT3 (Qureshi et al. 1986). Later it was shown that tocotrienols from palm oil could effectively reduce serum cholesterol levels in humans and that γT3 is likely the most efficacious form (Qureshi et al. 1991). Tocopherols do not exhibit this property. In general, identification of beneficial nutraceutical effects of tocochromanols, beyond simple antioxidant activity, will likely increase in the future. Also noteworthy is that the food matrix provided by barley is likely more healthful then the high fat vegetable oils from soybean and corn that is the principal source of dietary vitamin E

intake in the US diet.

As shown in Figures 1, 2, and 3, the quantities of tocochromanols in barley present a fairly broad range, especially αT3 and γT3. Much of this variation is likely under genetic control (Peterson and Qureshi 1993). Thus, the ability to manipulate the amount and distribution of tocochromanols in barley through selective breeding holds much promise for the future of food barley. Whole grain barley also brings other valuable components to human diets, including β-glucan (Yang et al. 2003; Chutimanitsakun et al. 2013). Therefore, the focus could more realistically be on the total nutritional composition of barley, and not exclusively on its tocochromanol content. Other plant products are superior sources of αT. Sunflower seeds for example, are one of the best sources of vitamin E, and contain approximately 351.7 mg/kg αT (USDA-ARS National Nutrient Database; http://ndb.nal.usda.gov/ndb/foods/show/3658; verified 29 October 2014).

A final question to address in this context is the role of tocochromanols in barley growth, development, and reproductive fitness. Studies with *Arabidopsis* mutants deficient in tocopherol biosynthesis clearly illustrate a role for these metabolites in cold tolerance (Maeda 2006). Likewise, these mutants were employed to demonstrate a critical role for tocopherols in germination and seed storage (Sattler et al. 2004). There is also evidence of enhanced germination and root growth in barley cultivars correlating with higher γT concentration (Desel andKrupinska 2005). Given the reported role of these compounds in reducing oxidative stresses in the photosynthetic apparatus, one might expect higher concentrations in more stress-tolerant germplasm. In emmer wheat, this is supported by a study which showed that seeds collected from a location with higher abiotic stresses had higher tocochromanol concentrations than those collected from locations with lower abiotic stresses (Watts et al. 2014). With this dataset, it is not currently possible to correlate tocochromanol concentrations with stress resistance, either at the phenotype or at the QTL level. However, if this GWAS panel were also characterized for stress resistance, the resulting dataset could be used to identify an association between the two phenotypes.

QTL and candidate genes

Thirteen significant SNP: tocochromanol trait associations were detected on three chromosomes (1H, 6H, and 7H). The significant SNPs within each linkage group can be further subdivided into groups separated by substantial linkage distances. Chromosomes 1H and 6H have two groups each, separated by 17.8 cM and 12.3 cM, respectively. At least three groups are on chromosome 7H: groups 1 and 2 are 93.9 cM apart, and groups 2 and 3 are 40.2 cM apart. The third group on chromosome 7H is substantially larger than the others, spanning 9.8 cM. In our analysis, the pattern of association of SNPs with tocochromanol forms and fractions, and the presence of distinct candidate genes in some of the QTL regions, suggests that we identified five regions of the barley genome that contain genes and/or regulatory elements with important functions in the tocochromanol pathway, or that control other traits that indirectly affect tocochromanol concentration. An example of the latter would be a gene important for ascorbate biosynthesis, in which case increased ascorbate concentrations would in turn help to maintain αT concentrations in plant tissues (Munné-Bosch 2005). An analysis of LD suggests that all significant SNPs, except those found in groups 2 and 3 on chromosomes 7H, are not in linkage disequilibrium with each other (Supplementary File 1). In group 2 on chromosome 7H, the two significant markers are in LD, but the two markers were significant for different tocochromanol forms. In group 3 on chromosome 7H, in some cases, adjacent significant markers were in LD. However, the middle significant marker was not in LD with either the first or last significant marker, which suggests that multiple QTL lie in this region of the genome.

The most significant associations we detected were on 7H, at cM 136.0, where SNP 11_21209 was significantly associated with $\beta T3$, δT, and γT (with $-\log_{10}p$ values of 4.4, 13.5, and 10.1, respectively). The barley genome annotations for this region include a sequence that could encode either *VTE*2 or an enzyme upstream of geranylgeranyl diphosphate, and a sequence encoding *HGGT*. At SNP 11_0861 (2.8 cM from 11_21209) we detected significant associations with $\alpha T3$, δT, TT3, and TTC concentration. There were also associations with this SNP approaching the FDR threshold, for αT, $\gamma T3$, and TTP concentration. These regions are likely associated with a gene encoding an enzyme

upstream of all tocochromanol forms, because in both cases, various tocopherols and tocotrienols are affected, and the direction of the effect is uniform across all tocochromanol forms and fractions. SNP 11_10885, at 145.2 cM on chromosome 7H, was associated with βT3, δT3, and γT3. Additionally, associations with this SNP approached the FDR threshold for αT3, TT3, and TTC concentration, but no tocopherol forms approached the FDR threshold. This is consistent with the expected effect of *HGGT*, which is upstream of all tocotrienol forms, but no tocopherol forms. The sequence for *HGGT* identified by Cahoon et al. (2003) matches a sequence in the barley genome sequence (at cM 145.1 with the current SNP map), making it likely that *HGGT* is the gene driving the effects observed for SNP 11_10885. Oliver et al. (2014) found significant QTL only for βT3, γT3, and δT3 in this region using a biparental mapping population consisting of 142 Recombinant Inbred Lines (RILs). An alignment of the barley consensus map and the map used by Oliver et al. (2104) shows that this region by Oliver et al. encompasses some, but not all, of the significant markers in this analysis (Supplementary File 2). With 1,466 elite accessions and GWAS, we were able to find significant or near-significant associations with all tocochromanol forms in this region. This is consistent with the expected combined effect of the enzymes *VTE2* and *HGGT*, two enzymes that Oliver et al. (2014) proposed may be in this region.

Proximal to the QTL cluster coincident with *VTE2* and *HGGT*, we found two significant SNPs, one associated with βT and one associated with δT, at cM 95.0 and cM 95.7, respectively. This region also had associations that approached the FDR threshold for αT, TTP, and TTC. There are no annotated genes in this region that were obviously associated with tocochromanol biosynthesis. Oliver et al. (2014) did not report any QTLs or candidate genes in this region. Also on 7H, we detected a significant marker on the short arm (at cM 1.1) that was associated with γT3 but no other tocochromanol forms or fractions. There are no annotated candidate genes in this region. Oliver et al. (2014) identified a QTL close to this marker, but an alignment of the two maps shows that this marker and the QTL reported by Oliver et al. do not overlap (Supplementary file AL).

On chromosome 6H, we found a significant association of one SNP (12_30637) at cM 70.6 with γT3. *VTE4* is annotated in this region, at cM 71.5. Oliver et al. (2014)

reported QTLs for γT and δT in this same region and also identified *VTE4* as a candidate gene for the observed QTL effects. An alignment of the two maps shows that the QTL identified by Oliver et al. (2014) overlaps marker 12_30637, but not marker 12_30802 (Supplementary File AL). *VTE4*, which catalyzes the conversion of the γ and δ tocochromanol forms to α and β tocochromanol forms, respectively, could explain the accumulation of δT, δT3, γT, and γT3, if the enzyme was rate-limiting in this panel. Separately, on chromosome 6H, marker 12_30802 (at cM 58.3) was significantly associated with δT, and had an association that did not reach the FDR threshold for γT. There are three sequences that show similarities to genes that encode enzymes upstream of geranylgeranyl diphosphate. However, given that this QTL appears to affect only δT and γT concentrations, it seems more likely that it is driven by a gene or a regulatory sequence that could affect tocochromanol forms differently, similar to *VTE4*.

On chromosome 1H we detected two significant SNP markers, 11_20021 and 11_10586, that were 17.8 cM apart. Marker 11_20021 (at cM 110.0) was significantly associated with TT3 concentration and TTC concentration. This marker also had associations that approached the FDR threshold with αT, αT3, δT3, γT3, and TTP concentration. A sequence resembling genes encoding *VTE2*, as well as enzymes upstream of geranylgeranyl diphosphate, is annotated 2.1 cM from this marker, at cM 107.7. Marker 11_10586 was significantly associated with βT, and approached the FDR threshold for βT3, but was not associated with any other tocochromanol forms or fractions, and was not in close proximity to any sequences obviously associated with tocochromanol biosynthesis. Oliver et al. (2014) reported no QTLs on chromosome 1H.

GWAS can provide fundamental insights into the genetic bases of economically important traits, as evidenced by recent reports in a range of crop plants, including barley (Mohammadi et al. 2014; Muñoz-Amatriaín et al. 2014a). By providing estimates of the number and genomic context of sequences affecting target traits in relevant germplasm, GWAS can also provide targets for Marker Assisted Selection (MAS) that will increase the efficiency of development for superior varieties.

In this study, GWAS allowed us to assign linkage map positions to *HGGT* and *VT4*, corroborating the prior report of Oliver et al. (2014) regarding map positions of

these genes. In doing so, we confirm what many others have reported, namely that GWAS can identify the same QTLs, and candidate genes, as biparental QTL mapping populations. An advantage of GWAS in this context is that a panel can be assembled immediately, whereas with a biparental population, for self-pollinated crops, it can take several years after an initial cross to achieve the amount of seed and the desired level of homogeneity before phenotypic evaluations can begin (Muñoz-Amatriaín et al. 2014b). In addition to QTL likely associated with *HGGT* and *VTE4*, we also identified significant associations between SNPs and tocochromanol forms and fractions for which we were not able to identify a candidate gene, or were able to identify a candidate gene, but were unable to confidently predict its specific role. Given our data, prior data, and the availability of genetic stocks, barley is now a suitable candidate for genetic dissection of the tocochromanol biosynthetic pathway and deeper exploration of the effects of environment on tocochromanol concentration.

In terms of practical applications, two accessions from AB (2AB04-01084-6 and 2AB04-01084-15) have the favorable alleles at each of the 13 SNPs significantly associated with one or more tocochromanol forms and/or fractions. These two accessions may be a valuable resource for developing varieties with enhanced tocochromanols and agronomic performance. The TTC concentrations of these lines (81.79 mg/kg and 80.73 mg/kg, respectively) is higher than the average accession in 2007, the year that these lines were grown, but lower than the highest accession grown in that year (6B05-0788, from BA), which had a TTC of 90.02 mg/kg. The αT concentrations for these accessions, 17.51 mg/kg and 14.29 mg/kg, were also higher than the average accession in 2007, but lower than the highest accession grown in that year (04WA.122.24 from BA), which had an αT concentration of 15.33 mg/kg. In this set of germplasm, no accession had all negative alleles at the 13 significant SNPs.

Conclusions

This study demonstrates the ability of association mapping to detect genetic determinants of complex traits in a panel of elite germplasm. This approach for QTL and candidate gene identification can complement the use of bi-parental mapping populations

specifically tailored to each trait. The identification of these 13 significant marker-trait associations in the barley genome could be an important resource in breeding efforts to either increase the nutritional properties of barley, or to develop barley varieties with greater tolerance to abiotic stresses that could be alleviated by this group of antioxidants. The draft of the barley genome published by the International Barley Genome Sequencing Consortium (2012) allowed for the mapping of barley *HGGT*, and enabled the identification of sequences homologous to known tocochromanol-associated genes.

Acknowledgements

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25-29.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practival and powerful approach to multiple testing. J R Stat Soc 57:289-300.Berger GL, Liu S, Hall MD, Brooks WS, Chao S et al. (2012) Marker-trait associations in Virginia Tech winter barley identified using genome-wide mapping. Theor Appl Genet 126:693-710.

Blake VC, Kling JG, Hayes PM, Jannink JL, Jillella SR, et al. (2012) The hordeum toolbox: The barley coordinated agricultural project genotype and phenotype resource. Plant Genome 5:81-91.

Bruno RS, Leonard SW, Atkinson J, Montine TJ, Ramakrishnan R, Bray TM and Traber MG (2005) Faster plasmid vitamin E disappearance in smokers is normalized by vitamin C supplementation. Free Radic Biol Med 40:555-556.

Cahoon EB, Hall SE, Ripp KG, Ganzke TS, Hitz WD and Coughlan SJ (2003) Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased

antioxidant content. Nat Biotechnol 21:1082:1087.

Cantor RM, Lange K and Sinsheimer JS (2010) Prioritizing GWAS Results: A review of statistical methods and recommendations for their application. Am J Hum Genet 86:6-22.

Chutimanitsakun Y, Cuesta-Marcos A, Chao S, Corey A, Filichkin T, et al. (2013) Application of marker-assisted selection and genome-wide asociation scanning to the development of winter food barley germplasm resources. Plant Breeding 132:563-570.

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, et al. (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582.

Collakova E, DellaPenna D (2001) Isolation and functional analysis of homogentisate phytyltransferase from *Synechocytis* sp. PCC 6803 and Arabidopsis. Plant Physiol *127:1113-1124* .

Cooney RV, Franke AA, Harwood PJ, Hatch-Pigott V, Custer LM, Mordan LJ (1993). Gamma-tocopherol detoxification of nitrogen dioxide: superiority to alpha-tocopherol. Proc Natl Acad Sci 90:1771-1775

Cordero Z, Drogan D, Weikert C, Boeing H (2010) Vitamin E and risk of cardiovascular diseases: A review of epidemiological and clinical trial studies. Crit Rev Food Sci 50:420-440.

Deng W, Nickle DC, Learn GH, Maust B, Mullins JI (2007) ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics 23:2334-2336.

Desel C, Krupinska K (2005). The impact of tocochromanols on early seedling development and NO release. J Plant Physiol 162:771-776. Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250-255.

Evangelou E, Ioannidis JPA (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14:379-389.

Frank J (2004) Beyond vitamin E supplementation: An alternative strategy to improve

vitamin E status. J of Plant Physiology 162: 834-843.

Fratianni A, Caboni MF, Irano M, Panfili G (2002) A critical comparison between traditional methods supercritical carbon dioxide extraction for the determination of tocochromanols in cereals. Eur Food Res Technol 215: 353-358.

Gey FK, Puska P, Jordan P, Ulrich MK (1991) Inverse correlation between plasm and vitamin E and mortality from ischemic heart disease in cross-cultural epidemiology. American J Clinical Nutrition: 326S-334S.

Handelman GJ, Machlin LJ, Fitch AK, Weiter JJ, Dratz EA (1985) Oral α-tocopherol supplements decrease plasma γ-tocopherol levels in humans. J Nutr 115:807-813.

Huang HY, Appel LJ (2003) Supplementation of diets with α-tocopherol reduces serum concentrations γ- and δ-tocopherol in humans. J Nutr 133:3137-3140.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, et al. (2011). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res gkr948.

Jaiswal P (2011) Gramene database: a hub for comparative plant genomics. Methods Mol Biol 678:247–275.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348-354.

Kent WJ (2002) BLAT—The BLAST-Like Alignment Tool. Genome Res 12:656–664.

Khanna S, Roy S, Ryu H, Bahadduri P, Swaan PW, Ratan RR, Sen CK (2003) Molecular basis of vitamin E action: tocotrienol modulates 12-lipoxygenase, a key mediator of glutamate-induced neurodegeneration. J Biol Chem 278:43508-43515.

Klein EA, Thompson IM, Tangen CM, Crowley JJ, Lucia MS et al. (2011) Vitamin E and the risk of prostate cancer: the selenium and vitamin E cancer prevention trial (SELECT). JAMA 306:1549-1556.

Lee IM, Cook NR, Gaziano JM, Gordon D, Ridker PM et al. (2005). Vitamin E in the primary prevention of cardiovascular disease and cancer: the women's health study: a randomized controlled trial. JAMA 294:56-65.

Maeda H, Song W, Sage TL, DellaPenna D (2006) Tocopherols play a crucial role in low-

temperature adaptation and phloem loading in arabidopsis. Plant Cell 18:2710-2732.

Matsushita Y, Kang WK, Charlwood BV (1996) Cloning and analysis of a cDNA encoding farnesyl diphosphate synthase from Artemisia annua. Gene 172:207-209.

Mohammadi M, Endelman JB, Nair S, Chao S, Jones SS et al. (2014) Association mapping of grain hardness, polyphenol oxidase, total phenolics, amylose content, and β-glucan in US barley breeding germplasm. Mol Breeding 34:1229-1243.

Munné-Bosch S (2005) The role of α-tocopherol in plant stress tolerance. Plant Physiol 162:743-748.

Muñoz-Amatriaín M, Cuesta-Marcos A, Endelman JB, Comadran J, Bonman JM (2014a) The USDA barley core collection: genetic diversity, population structure, and potential for genome-wide association studies. PloS One 9:e94688.

Muñoz-Amatriaín M, Cuesta-Marcos A, Hayes PM, Meuhlbauer GJ (2014b) Barley genetic variation: implications for crop improvement. Brief Funct Genomics 13:341-350.

Muñoz-Amatriaín M, Moscou MJ, Bhat PR, Svensson JT, Bartoš J et al. (2011) An improved consensus linkage map of barley based on flow-sorted chromosomes and single-nucleotide polymorphism markers. Plant Genome 4:238-249.

Oliver RE, Islamovic E, Obert DE, Wise ML, Herrin LL et al. (2014) Comparative systems biology reveals allelic variation modulating tocochromanol profiles in barley (*Hordeum vulgare* L.). Plos One 9:e96276.

Peterson DM, Quresh AA (1993) Genotype and environment effects on tocols of barley and oats. Cereal Chem 70:157-161

Qureshi AA, Burger WC, Peterson DM, Elson CE (1986) The structure of an inhibitor of cholesterol biosynthesis isolated from barley. J Biol Chem 261:10544-10550.

Qureshi AA, Qureshi N, Wright JJK, Shen Z, Kramer G et al. (1991) Lowering of serum cholesterol in hypercholesterolemic humans by tocotrienols (palmvitee). Am J Clin Nutr 53:1021S-1026S.

Roberts JM, Myatt L, Spong CY, Thom EA, Hauth JC et al. (2010) Vitamins C and E to prevent complications of pregnancy-associated hypertension. N Engl J Med 362:1282-1291.

Sadre R, Gruber J, Frentzen M (2006) Characterization of homogentisate prenyltransferases involved in plastoquinine-9 and tocochromanol biosynthesis. Febs Lett 580:5357-5362.

Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM et al. (2010) Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med 362:1675-1685.

Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, DellaPenna D (2004) Vitamin E Is Essential for Seed Longevity and for Preventing Lipid Peroxidation during Germination. Plant Cell 16:1419-1432

Sen CK, Khanna S, Roy R (2007) Tocotrienols in health and disease: The other half of the natural vitamin E family. Mol Aspects Med 28:692-728.

Shintani D, DellaPenna D (1998) Elevating the vitamin E content of plants through metabolic engineering. Science 282:2098-2100.

Szűcs P, Blake VC, Bhat PR, Chao S, Close TJ, et al. (2009) An integrated resource for barley linkage map and malting quality QTL alignment. Plant Genome 2:134-140.

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-716.

Theriault A, Chao JT, Wang Q, Gapor A, Adeli K (1999) Tocotrienol: a review of its therapeutic potential. Clin Biochem 32:309-319.

Thompson JN, Hatina G (1979) Determination of tocopherols and tocotrienols in foods and tissues by high performance liquid chromatography. J Liquid Chrom 2: 327-344.

Traber MG, Arai H (1999) Molecular mechanisms of vitamin E transport. Annu Rev Nutr 19:343-55.

Traber MG, Burton GW, Huges L, Ingold KU, Hidaka H et al. (1992) Discrimination between forms of vitamin E by humans with and without genetic abnormalities of lipoprotein metabolism. J Lipid Res 33:1171-1182.

Watts EJ, Shen Y, Lansky EP, Nevo E, Bobe G, Traber MG (2014) High environmental stress yields greater tocotrienol content while changing vitamin E profiles of wild emmer wheat seeds. J Med Food.

Yang JL, Kim YH, Lee MS, Moon YK (2003) Barley beta-glucan lowers serum

cholesterol based on the up-regulation of cholesterol 7alpha-hydroxylase activity and mRNA abundance in cholesterol-fed rats. J Nutr Sci Vitaminol 49:381-387.

Tables

Program	Year	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TTC
AB	2006	15.84±0.19	34.63±0.55	0.72±0.01	-	0.32±0.01	0.64±0.02	-	3.30±0.11	61.68±0.83
AB	2007	12.86±0.18	32.68±0.58	1.01±0.02	8.23±0.23	0.35±0.01	1.37±0.05	3.92±0.16	7.07±0.17	67.48±0.87
BA	2006	16.26±0.25	33.89±0.47	0.70±0.02	-	0.43±0.01	0.67±0.02	-	4.42±0.17	63.96±0.95
BA	2007	12.18±0.16	31.64±0.36	0.97±0.02	8.51±0.22	0.30±0.01	1.40±0.04	3.46±0.13	7.00±0.21	65.45±0.63
MN	2006	13.30±0.18	30.95±0.49	0.65±0.02	-	0.36±0.01	0.67±0.01	-	2.64±0.08	53.28±0.77
MN	2007	10.83±0.10	31.94±0.32	0.84±0.02	6.66±0.13	0.33±0.01	1.04±0.03	2.46±0.07	4.48±0.09	58.59±0.50
MT	2006	19.12±0.18	40.46±0.52	0.72±0.02	-	0.49±0.02	0.60±0.01	-	4.87±0.13	75.92±0.79
MT	2007	11.61±0.20	29.04±0.33	0.82±0.01	8.24±0.29	0.26±0.01	1.43±0.07	3.18±0.09	9.14±0.18	63.72±0.71
N2	2006	16.99±0.19	35.78±0.56	0.63±0.02	-	0.36±0.01	0.52±0.01	-	3.61±0.11	64.23±0.73
N2	2007	11.09±0.12	30.87±0.37	1.13±0.02	7.68±0.24	0.28±0.01	1.16±0.03	2.97±0.09	7.09±0.14	62.28±0.74
N6	2006	15.71±0.14	33.82±0.53	0.57±0.01	-	0.27±0.01	0.67±0.01	-	3.13±0.08	58.87±0.76
N6	2007	11.48±0.17	33.64±0.44	1.09±0.02	7.34±0.18	0.27±0.01	1.06±0.03	2.24±0.06	4.94±0.10	62.06±0.75
UT	2006	15.84±0.42	28.95±0.98	0.63±0.03	-	0.47±0.03	0.69±0.05	-	3.91±0.29	56.90±1.56
UT	2007	14.08±0.22	33.28±0.52	0.88±0.02	6.97±0.21	0.34±0.01	1.09±0.05	3.68±0.12	7.01±0.19	67.34±0.93
WA	2006	13.95±0.17	30.21±0.4	0.58±0.010	-	0.49±0.02	0.62±0.02	-	4.36±0.15	59.11±0.87
WA	2007	12.17±0.14	30.82±0.31	0.87±0.02	8.83±0.22	0.38±0.01	1.28±0.04	4.74±0.15	7.12±0.16	66.21±0.73
Mean	2006	15.89±0.10	33.98±0.22	0.65±0.01	-	0.39±0.01	0.63±0.01	-	3.76±0.05	62.13±0.39
Mean	2007	12.04±0.07	31.73±0.15	0.95±0.01	7.81±0.08	0.31±0.00	1.23±0.02	3.34±0.05	6.74±0.08	64.15±0.28

Table 1. Means and standard errors for concentrations (mg/kg) of all tocochromanol forms and Total Tocochromanol (TTC) for accessions from each of the eight breeding programs, separated by year.

ANOVA Table for αT

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	1438.4	159.80	42.4	<0.001
Row Type	1	200.9	200.90	53.3	<0.001
Year	1	5533.5	5533.50	1466.6	<0.001
Residuals	1454	5485.8	3.80		

ANOVA Table for αT3

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	2527.0	280.82	11.5	<0.001
Row Type	1	265.0	264.75	10.8	0.001
Year	1	1772.0	1771.94	72.6	<0.001
Residuals	1454	35486.0	24.41		

ANOVA Table for βT

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	5.0	0.55	17.8	<0.001
Row Type	1	0.9	0.91	29.3	<0.001
Year	1	33.3	33.32	1076.3	<0.001
Residuals	1454	45.0	0.03		

ANOVA Table for βT3

Source	DF	SS	MS	F-value	Pr(>F)
Program	7	398.2	56.89	12.2	<0.001
Row Type	1	4.9	4.93	1.1	0.304
Residuals	755	3521.8	4.67		

ANOVA Table for δT

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	3.0	0.33	20.1	<0.001
Row Type	1	1.2	1.20	73.4	<0.001
Year	1	2.3	2.34	142.6	<0.001
Residuals	1451	23.8	0.02		

ANOVA Table for δT3

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	10.4	1.16	9.4	<0.001
Row Type	1	0.4	0.40	3.3	0.070
Year	1	129.7	129.74	1058.0	<0.001
Residuals	1452	178.1	0.12		

ANOVA Table for γT

Source	DF	SS	MS	F-value	Pr(>F)
Program	7	437.7	62.52	64.9	<0.001
Row Type	1	224.3	224.28	232.7	<0.001
Residuals	755	727.8	0.96		

ANOVA Table for γT3

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	1668.8	185.42	98.2	<0.001
Row Type	1	373.7	373.68	197.9	<0.001
Year	1	3135.2	3135.21	1660.4	<0.001
Residuals	1454	2745.5	1.89		

ANOVA Table for TTC

Source	DF	SS	MS	F-value	Pr(>F)
Program	9	21072.0	2341.30	35.1	<0.001
Row Type	1	1957.0	1956.55	29.3	<0.001
Year	1	1388.0	1388.50	20.8	<0.001
Residuals	1454	97001.0	66.71		

Table 2. Analyses of variances for α-tocopherol and TTC concentration in barley.

Marker	Chromosome	Position	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_20021	1H	109.8	2.47	5.31	0.33	0.19	0.07	1.60	0.67	2.88	6.17	2.38	5.16
11_10586	1H	127.6	0.13	0.26	7.17	3.03	0.63	0.16	0.63	0.06	0.21	0.06	0.13
12_30802	6H	58.3	1.26	0.42	1.82	0.23	3.77	0.23	2.63	0.45	0.57	1.60	0.23
12_30637	6H	70.6	0.08	0.09	0.04	0.40	0.70	0.99	0.72	5.09	0.75	0.10	0.97
12_30296	7H	1.1	0.20	0.91	0.20	0.52	0.64	1.26	0.14	4.83	1.89	0.17	1.27
11_21201	7H	95.0	1.51	0.34	6.26	0.22	3.56	0.25	0.37	0.02	0.21	2.13	0.97
11_20311	7H	95.7	3.60	1.99	2.73	0.01	4.16	0.41	1.53	0.38	1.58	4.10	3.50
11_21209	7H	136.0	0.35	1.33	1.42	4.39	13.53	1.21	10.05	0.06	1.20	0.73	2.48
11_10861	7H	138.8	3.13	6.37	1.95	1.34	5.06	0.90	0.35	3.29	7.04	3.69	6.52
11_10797	7H	141.4	0.22	2.43	0.14	1.11	7.50	0.17	1.97	0.95	2.54	0.38	3.08
12_10973	7H	142.4	0.66	2.07	0.46	6.90	4.91	3.57	2.24	2.25	2.79	0.42	2.91
11_10885	7H	145.2	0.25	2.32	0.30	9.84	0.08	7.06	0.07	4.84	3.87	0.24	3.28
12_31511	Unknown	Unknown	0.77	0.02	0.10	0.14	4.00	0.32	1.01	0.34	0.01	0.86	0.18

Table 3. Significance (-log$_{10}$(p)) of markers for concentration of all tocochromanol forms, fractions, and Total Tocochromanol (TTC). Highlighted values show significant marker-trait associations.

SNP Marker or Annotated Genome Sequence	Chromosome	Position (cM)*	Position (bp)**	Morex Contig Number	Sequence Annotation
MLOC_16149	1H	107.7	430,628,647-430,636,996	157254	Resembles genes encoding *HGGT* and *VTE2*, and enzymes upstream of geranylgeranyl diphosphate
11_20021	1H	109.8	Unknown	48282	SNP Marker
11_10586	1H	127.6	447,413,118	171284	SNP Marker
MLOC_72891	6H	56.3	84,648,047-84,652,085	62562	Resembles genes encoding enzymes upstream of geranylgeranyl diphosphate
MLOC_44750	6H	58.2	145,761,890-145,768,330	275292	Resembles genes encoding enzymes upstream of geranylgeranyl diphosphate
12_30802	6H	58.3	164,529,448	1592014	SNP Marker
MLOC_66290	6H	58.9	265,678,548-265,680,735	51352	Resembles genes encoding enzymes upstream of geranylgeranyl diphosphate
12_30637	6H	70.6	432,652,993	1559740	SNP Marker
MLOC_13082	6H	71.5	437,175,681-437,178,395	1564754	Resembles gene encoding *VTE4*
12_30296	7H	1.1	2,219,656	178733	SNP Marker
11_21201	7H	95.0	Unknown	45924	SNP Marker
11_20311	7H	95.7	Unknown	1566790	SNP Marker
11_21209	7H	136.0	570,924,522	1579096	SNP Marker
MLOC_37476	7H	138.4	575,423,330-575,428,610	25484480	Resembles genes encoding enzymes upstream of geranylgeranyl diphosphate, and VTE2-1
11_10861	7H	138.8	Unknown	7405	SNP Marker
11_10797	7H	141.4	577,196,196	354235	SNP Marker
12_10973	7H	142.4	578,753,062	77635	SNP Marker
MLOC_12567	7H	145.1	584,322,177-584,325,723	1563577	HGGT
11_10885	7H	145.2	584,352,965	2547604	SNP Marker

Table 4. Significant SNPs associated with tocochromanols, and annotated sequences known or predicted to be associated with the tocochromanol biosynthesis pathway that occurred within 2.5 cM of a significant marker.

*Linkage map positions (Muñoz-Amatriaín et al. 2011)

**Genome sequence positions, (International Barley Genome Sequencing Consortium 2012)

Marker	Chromosome	Position	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_20021	1H	109.8	0.50	1.83	0.010	0.09	0.002	0.062	-0.12	0.36	2.26	0.51	3.02
11_10586	1H	127.6	-0.04	-0.16	0.047	0.44	0.008	-0.007	0.08	-0.01	-0.15	0.02	0.15
12_30802	6H	58.3	-0.38	0.42	-0.038	0.13	-0.045	0.018	-0.36	0.12	0.61	-0.46	-0.44
12_30637	6H	70.6	0.03	0.10	0.002	0.17	0.013	0.045	0.13	0.50	0.63	0.04	1.09
12_30296	7H	1.1	-0.07	0.57	0.006	0.19	-0.011	0.048	0.03	0.44	1.05	-0.07	1.19
11_21201	7H	95.0	-0.28	-0.24	-0.053	0.08	-0.029	0.013	-0.06	0.00	-0.18	-0.36	-0.84
11_21209	7H	136.0	0.08	0.53	0.018	0.58	0.050	0.034	0.45	0.01	0.56	0.15	1.30
11_10861	7H	138.8	-0.33	-1.18	-0.020	-0.26	-0.026	-0.024	-0.05	-0.22	-1.42	-0.38	-1.99
11_10797	7H	141.4	-0.07	-0.91	-0.004	-0.28	-0.043	-0.009	-0.20	-0.14	-1.06	-0.11	-1.74
12_10973	7H	142.4	-0.18	0.90	0.011	0.91	0.038	0.084	0.24	0.26	1.22	-0.13	1.84
11_10885	7H	145.2	0.07	-0.76	-0.006	-0.85	0.002	-0.097	0.01	-0.32	-1.17	0.06	-1.56
11_20311	7H	95.7	0.62	1.04	0.043	0.01	0.041	-0.024	0.24	0.09	1.03	0.70	2.44
12_31511	Unknown	Unknown	0.19	0.02	-0.003	-0.06	0.032	-0.016	0.15	-0.07	-0.01	0.21	0.24

Table 5. Best Linear Unbiased Estimators (BLUEs) for concentration (mg/kg) of all tocochromanol forms, fractions, and Total Tocochromanol (TTC). Positive values indicate that individuals with the "A" allele has a higher tocochromanol concentration, and negative values indicates that genotypes with the "B" allele have a higher tocochromanol concentration. Highlighted values show significant marker-trait associations.

Marker		AB	BA	MN	MT	N2	N6	UT	WA
	Total Individuals	190	186	191	192	183	189	132	190
11_20021	A	104	107	10	192	183	1	32	159
	B	86	79	181	0	0	188	87	31
	Heterozygous	0	0	0	0	0	0	4	0
11_10586	A	77	75	1	52	79	21	45	53
	B	112	111	189	140	102	168	81	134
	Heterozygous	1	0	1	0	1	0	4	3
12_30802	A	173	149	163	155	182	189	118	90
	B	17	37	28	37	1	0	13	98
	Heterozygous	0	0	0	0	0	0	1	2
12_30637	A	105	112	6	189	137	0	91	168
	B	85	74	185	3	40	189	40	22
	Heterozygous	0	0	0	0	6	0	1	0
12_30296	A	64	47	0	119	21	1	57	86
	B	122	138	191	71	161	187	75	85
	Heterozygous	2	0	0	0	1	1	0	0
11_21201	A	152	167	166	143	161	189	113	186
	B	35	19	24	49	20	0	11	4
	Heterozygous	3	0	1	0	2	0	1	0
11_20311	A	32	15	0	26	2	0	11	1
	B	155	171	191	166	180	189	116	189
	Heterozygous	3	0	0	0	1	0	3	0
11_21209	A	109	77	175	42	71	145	115	92
	B	79	109	16	150	107	43	15	93
	Heterozygous	2	0	0	0	5	0	0	5
11_10861	A	32	98	14	109	74	33	8	68
	B	154	88	177	83	105	155	124	119
	Heterozygous	4	0	0	0	4	1	0	3
11_10797	A	169	171	191	143	140	189	68	110
	B	19	15	0	49	39	0	61	74
	Heterozygous	2	0	0	0	3	0	2	5
12_10973	A	17	2	0	58	10	0	62	62
	B	172	184	191	134	173	189	70	122
	Heterozygous	1	0	0	0	0	0	0	6
11_10885	A	48	78	0	68	61	0	46	46
	B	140	108	191	123	116	189	83	136
	Heterozygous	2	0	0	1	6	0	0	6
12_31511	A	2	7	0	36	0	0	31	77
	B	188	179	191	156	183	189	101	109
	Heterozygous	0	0	0	0	0	0	0	3

Table 6. Distribution of significant markers associated for tocochromanols for each of eight breeding program.

Figures
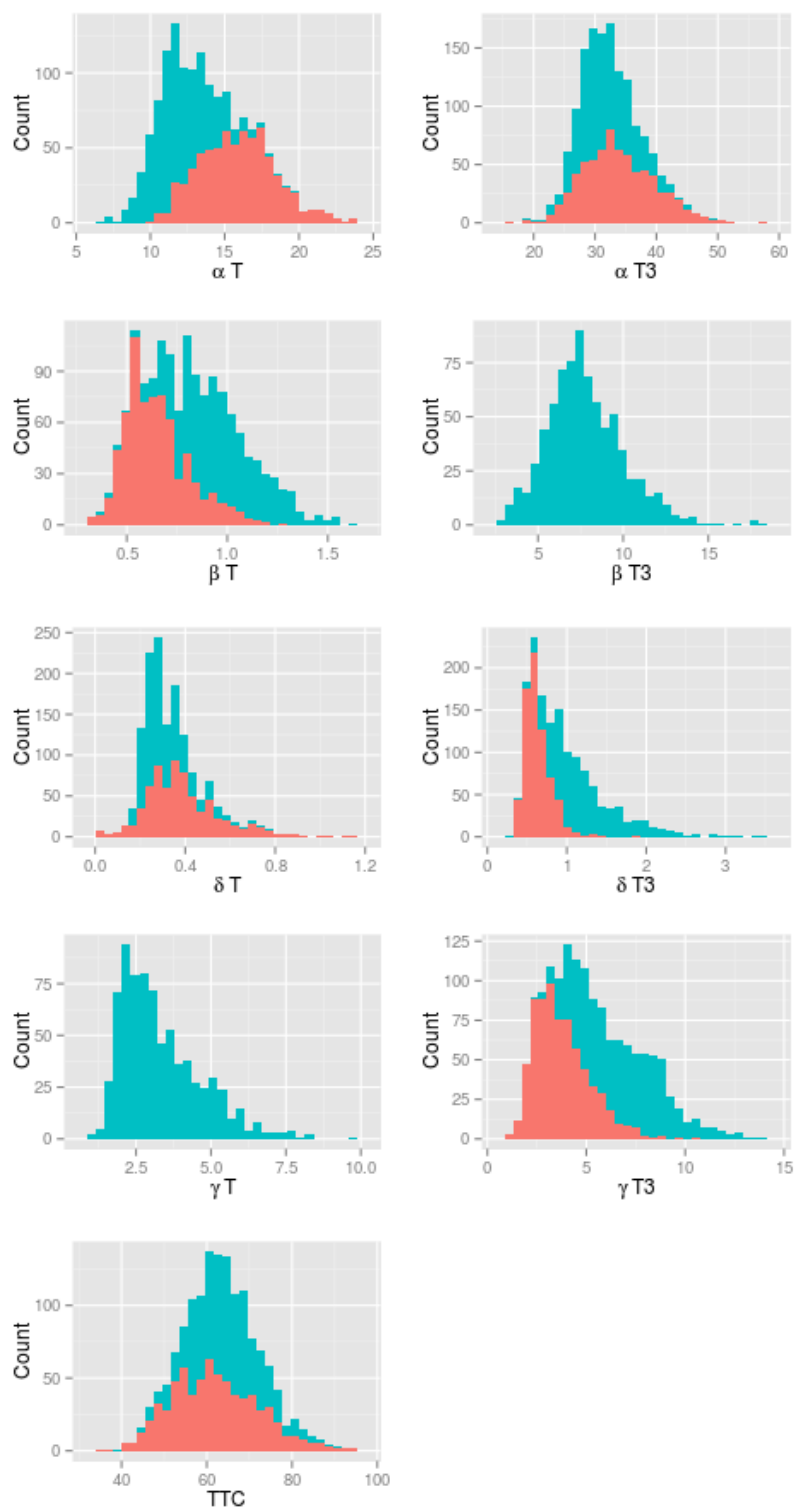
Figure 1. Distributions of concentrations of all tocochromanol forms and Total Tocochromanol (TTC). Red represents 2006 and blue represents 2007. Reliable data for βT3 and γT are not available for 2006.
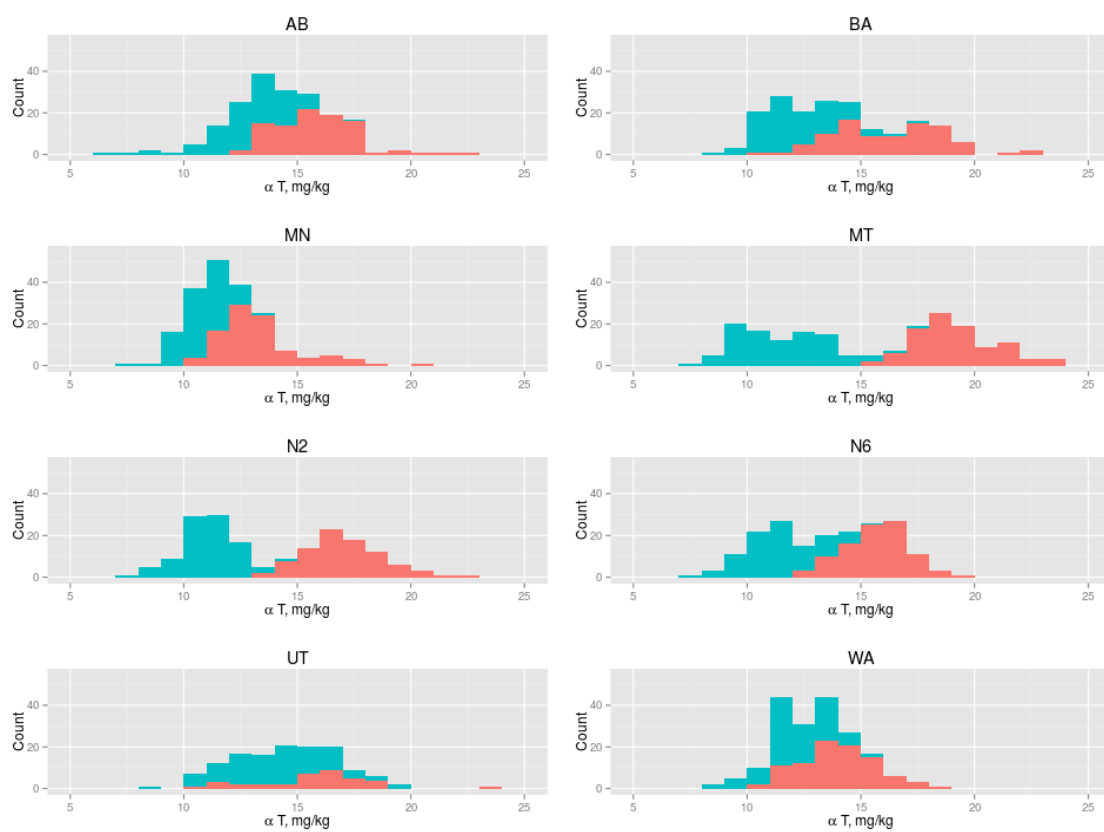
Figure 2. Distributions of α-tocopherol (αT) across breeding programs. Red represents 2006 and blue represents 2007.
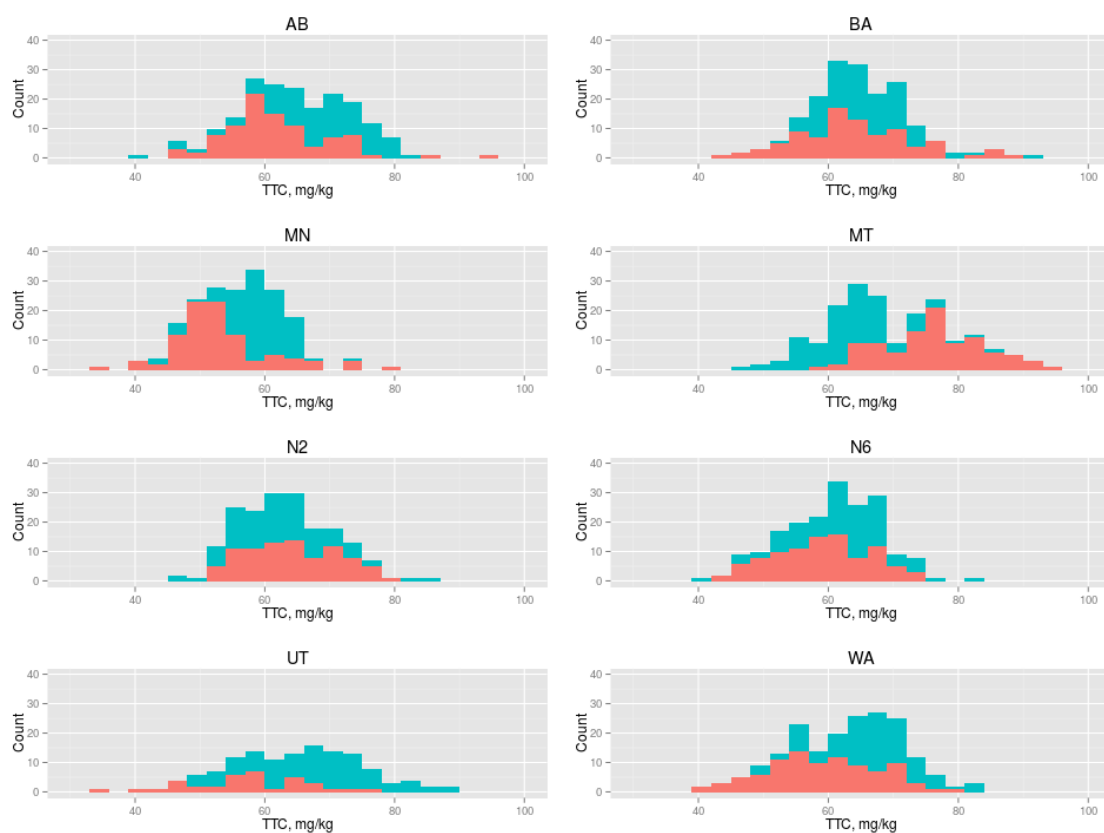
Figure 3. Distributions of α-tocopherol (αT) and Total Tocochromanol (TTC) across breeding programs. Red represents 2006 and blue represents 2007.

Figure 4. Fraction of genetic variance for this population explained by each principal component.
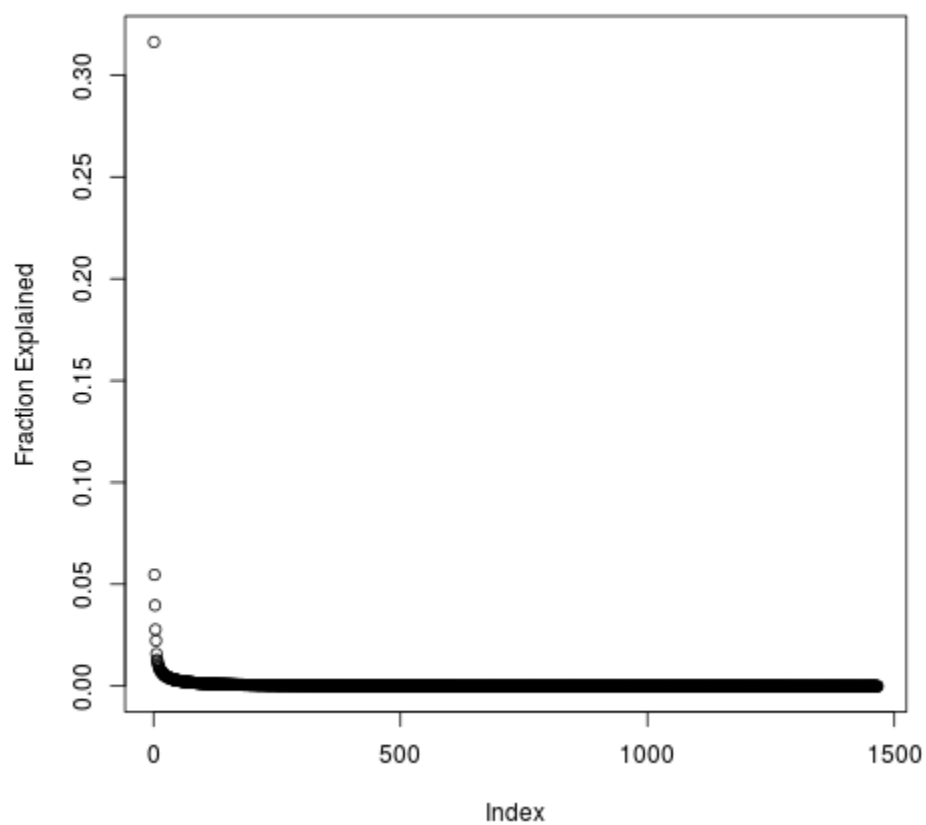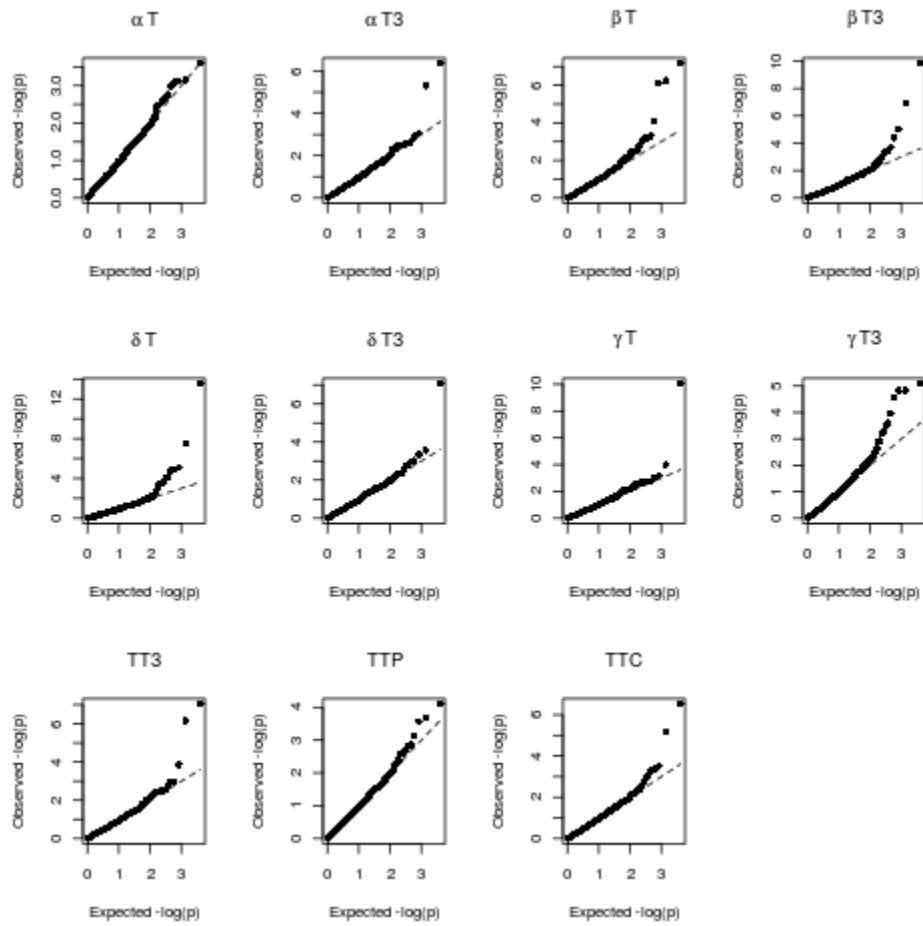
Figure 5. Q-Q plots showing the distribution of p-values, plotted against the expected distribution of p-values, for each analysis in this study.
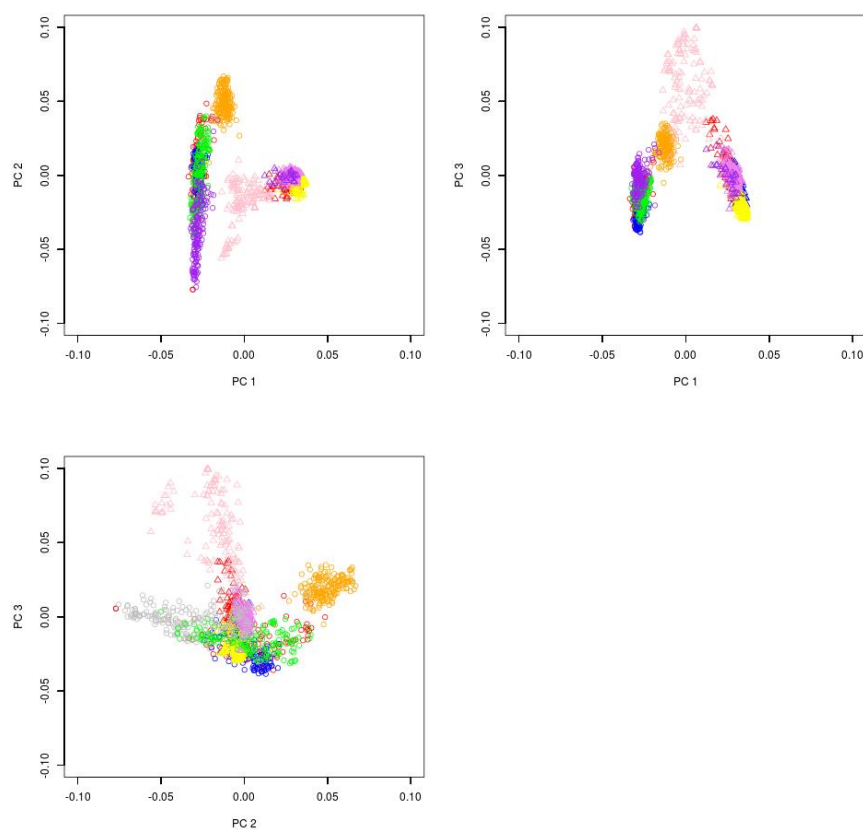
Figure 6. Principal component analysis of all three combinations of the first, second, and third principal components for this set of germplasm. Two-row accessions are depicted by circles, and six-row accessions are depicted by triangles. The program of origin for each accessions is color-coded, as follows: AB (red), BA (blue), the MN (yellow), MT (green), N2 (orange), N6 (violet), UT (pink), and WA (grey).
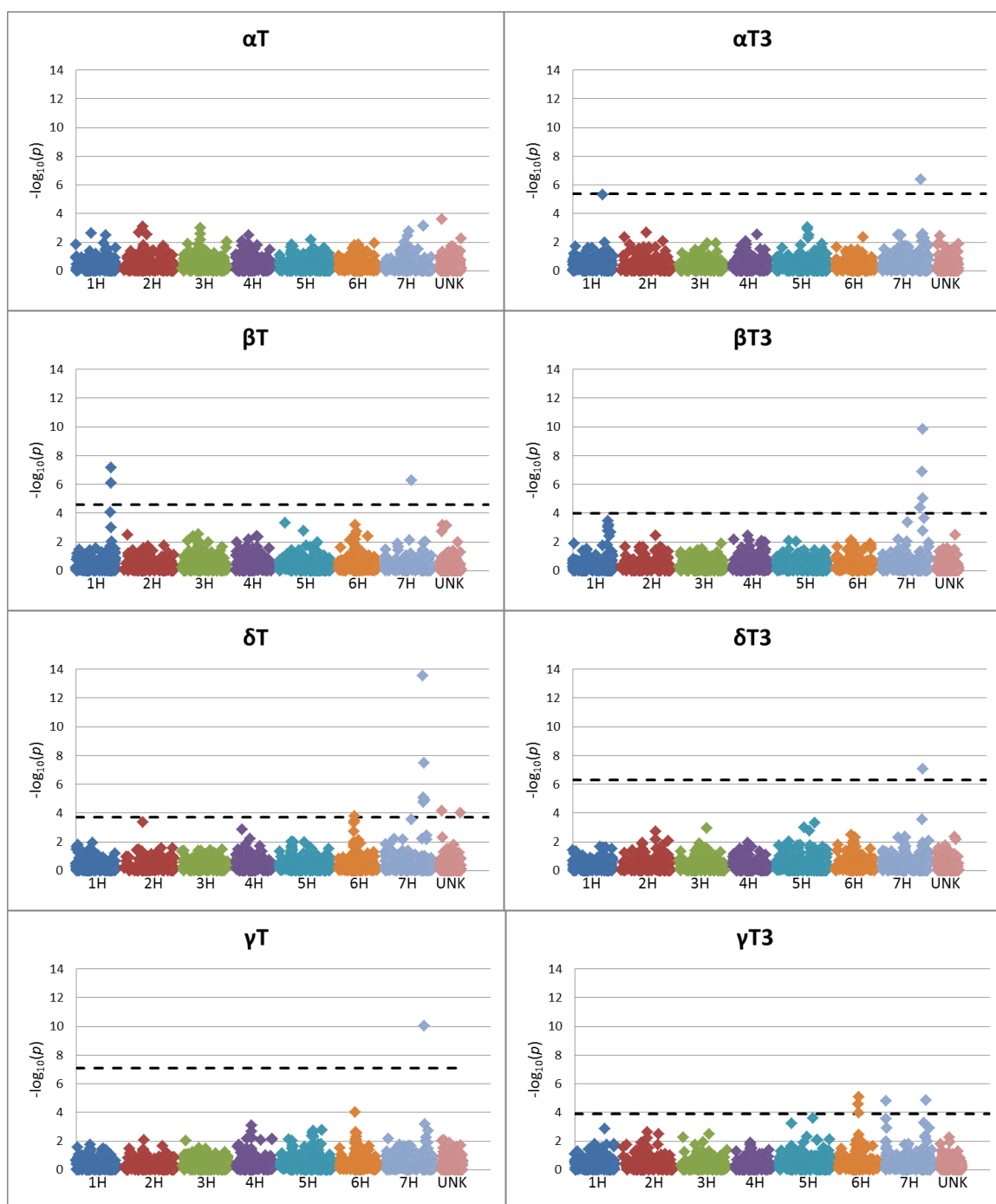
Figure 7 (Part One)
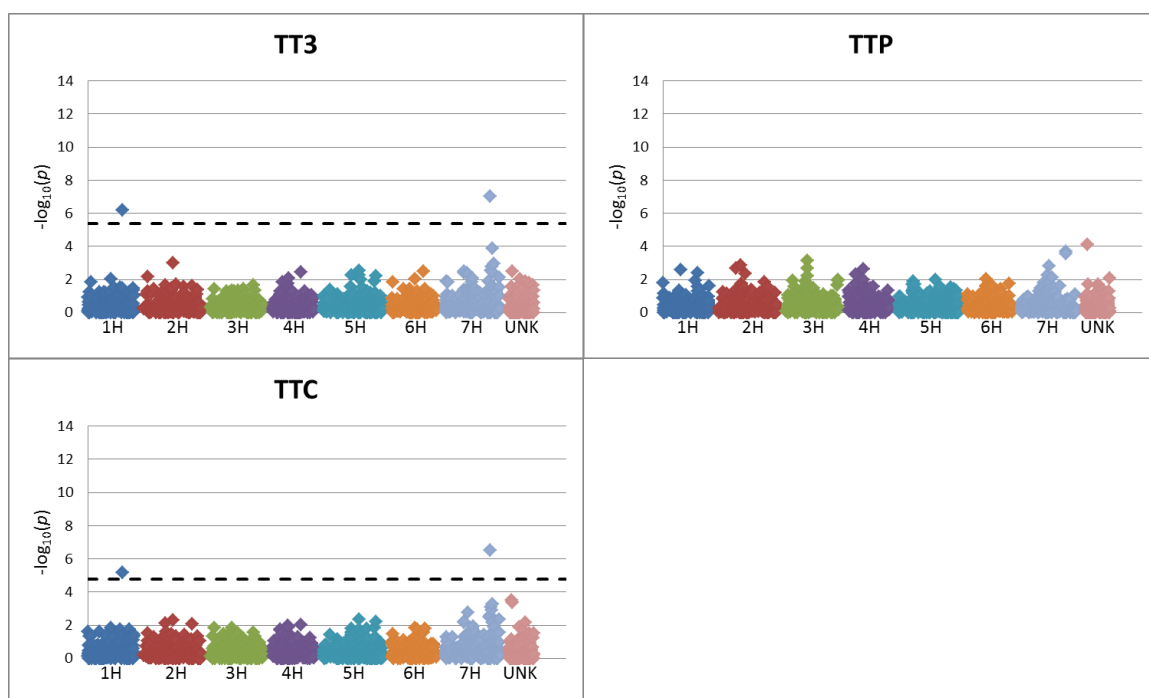
Figure 7 (continued). Manhattan plots showing results of GWAS for concentrations of all tocochromanol forms, fractions, and Total Tocochromanol (TTC). In analyses where one or more markers met the significance threshold determined by a false-discovery rate adjustment, a dotted line shows the significance threshold. Points in pink, adjacent to chromosome 7H, represent unmapped markers.

Registration of 'Alba' barley

Ryan C. Graebner, Alfonso Cuesta-Marcos, Scott Fisk, Brook O. Brouwer, Stephen S. Jones and Patrick M. Hayes*

R.C. Graebner, A. Cuesta-Marcos, S. Fisk and P.M. Hayes, Dep. of Crop and Soil Science, Oregon State Univ., Crop Science Building, 3050 SW Campus Way, Corvallis, OR 97331; B.O. Brouwer and S.S. Jones Dep. of Crop and Soil Science, Northwestern Washington Research and Extension Center, Washington State Univ., Mount Vernon, WA 98273. *Corresponding author (patrick.m.hayes@oregonstate.edu)

Abstract

'Alba' (Reg. No. CV-355, PI 672535) is a winter, six-row barley released by the Oregon Agricultural Experiment Station in 2012. The name "Alba" was chosen due to the bright and attractive appearance of the crop at maturity. In high rainfall environments, it has a notable yield advantage over check varieties and maintains excellent test weight and kernel plumpness. These advantages are, in part, attributable to resistance to barley stripe rust (incited by *Puccinia striiformis* f. sp. *hordei*) and scald (incited by *Rhynchosporium commune*). The principal end-use of Alba grain is as feed, but it could also be used for food (after pearling) and preliminary tests show that it can be successfully malted and used in craft brewing.

Introduction

'Alba' (Reg. No. CV-355, PI 672535) is a winter, six-row barley released by the Oregon Agricultural Experiment Station in 2012. The name "Alba" was chosen due to the bright and attractive appearance of the crop at maturity. "Alba" is derived from the Latin word for "white," and means dawn in Spanish. Prior to being named, Alba was tested under the experimental designations 'OR77' and 'TCFW6-002.' In high rainfall environments, it has a notable yield advantage over some varieties, and it maintains excellent test weight and kernel size. These advantages are, in part, attributable to

resistance to barley stripe rust (incited by *Puccinia striiformis* f. sp. *hordei*) and scald (incited by *Rhynchosporium commune*). The principal end-use of Alba grain is as feed, but it could also be used for food (after pearling), and preliminary tests show that Alba can be successfully malted and used in craft brewing.

Alba was derived from a cross made in 1997 between 'Strider' (released by the Oregon Agricultural Experiment Station in 1997), and 'Orca' (Hayes et al., 2000). Strider is a six-row, compact spike, winter growth habit, feed variety with low temperature tolerance comparable or superior to other commercially available varieties. Strider requires vernalization and long days to transition from the vegetative to the reproductive state. Strider has exceptionally poor malting quality, with almost no detectable enzyme activity in malt (Filichkin et al., 2010). Strider is resistant to stripe rust, moderately resistant to scald, and can show severe symptoms of Barley Yellow Dwarf Virus (BYDV). Orca is a two-row, erect spike, spring growth habit feed cultivar. Orca does not require vernalization to transition from the vegetative to the reproductive state (Hayes et al., 2000). Orca is resistant to stripe rust with mapped adult plant resistance QTL on chromosomes 4H and 5H, moderate resistance to scald, and it has the *Ryd2* gene for resistance to BYDV (Hayes et al., 2000).

Methods

Generation Development and Line Selection

The cross between Strider and Orca was made in 1997. From the F1 generation until head-row purification, all generations were fall-planted under field conditions at the Oregon State University Hyslop Farm, near Corvallis, OR USA. The F2 was planted as a bulk population of approximately 2000 plants. Selected F2 heads were threshed and bulked and grown as an F3 population. Selected F3 heads were grown as F4 head rows. Selected F4 head rows were harvested in bulk and advanced to a preliminary yield trial. Selections moved through subsequent cycles of replicated, multi-environment yield testing in Oregon and in the fall of 2004 one of the selected Strider/Orca sibling lines (F7) was designated as OR77 and tested regionally in replicated yield trials.

Seed Purification and Increase

Five hundred F10 heads were selected from OR77 plots and planted for head row purification and increase in the fall of 2007. Seed from one head from one row (F11) was used for Single Nucleotide Polymorphism (SNP) genotyping under the auspices of the USDA-NIFA Triticeae Coordinated Agricultural Project (http://www.triticeaecap.org/), and these data are available at the T3 database (Triticeae Coordinated Agricultural Project, 2014). In the T3 database Alba is designated as TCFW6-002. One thousand F11 heads were harvested from selected rows, threshed individually and transferred to the Washington Crop Improvement Association for production of F12 Breeder's seed.

Statistical Analyses

All statistical analyses were conducted using R version 3.0.1 (R Core Team, 2014). Alba was compared with 'Maja', Strider, 'Eight-Twelve' (Wesenberg et al., 1992), and 'Charles' (Obert et al., 2006) for agronomic traits in 33 environments over the years 2008-2012, although not all traits were measured in all years. At each location, plot sizes, nutrient management, weed control, and irrigation (if applied) were in accordance with local practice. Varieties were replicated either three or four times at each location, also in accordance with local practice, although only the mean value from each environment was used in this analysis. The same five varieties were tested for disease resistance in Corvallis, Oregon over the course of five years. Alba, Maja, and Strider were tested for low temperature tolerance in controlled freeze tests at the Martonvasar Research Institute (MRI; Hungary) in 2006 and 2008, as described by Skinner et al. (2006). Eight-Twelve and Charles were tested for low temperature tolerance in controlled freeze tests for one year each at the MRI, in 2006 and 2008 respectively. Winter survival was recorded in eight field trials where differential survival occurred. In 2013 Alba, Maja, and 'Full Pint,' were evaluated for leaf rust resistance at the Northwestern Washington Research and Extension Center, Washington State University, Mount Vernon, WA (WSU Mount Vernon). The malt quality of composite samples from Alba, Maja, and Charles was compared in 10 environments over the years 2009-2011, using the methods described by Budde et al. (2008). For the purposes of this report, two trials grown at the same location

but in different years, or under substantially different growing conditions, are considered as different environments. The mean of measurements collected from each environment for each cultivar were used for mean separation based on LSD ($p = 0.05$) except for the leaf rust data, where there was a single replicated experiment.

Characteristics

Botanical Description

Phenotypic selection for agronomic type and performance in the progeny of this wide cross (winter/spring and two-row/six-row define the principal germplasm groups of barley) resulted in a six-row barley with a lax spike. Alba has a winter growth habit. Alba has grain with adhering hulls, a white aleurone, short rachilla hairs, and rough awns.

Agronomic Performance

Across all 33 environments, Alba demonstrated a significantly higher yield than Charles. Grain from Alba had significantly higher test weight than all varieties except Maja, and higher plumpness than Maja and Eight-Twelve. Alba was significantly taller than Charles. There were no statistically significant differences in heading date or lodging between varieties (Table 1).

In high-rainfall environments (Brownsville, Corvallis, and Junction City, OR), where the average rainfall is greater than 800 mm year^{-1} (Western Regional Climate Center), Alba had a significantly higher yield than Eight-Twelve and Charles and a similar yield to Maja and Strider. Alba had a significantly higher test weight than Eight-Twelve, but a similar test weight to other varieties. Alba had significantly higher kernel plumpness than Maja, Strider, and Eight-Twelve (all six-rows), but was not significantly different from Charles (a two-row). Alba was significantly taller than all varieties in this trial except for Eight-Twelve, and later maturing than all varieties in this subset of environments. In the limited number of trials where lodging was observed, no significant difference in lodging was detected between these varieties. Variable within-trial lodging at the limited number of sites where lodging occurred precludes a robust statistical comparison of means (Table 2).

Pendleton, OR and Pullman, WA are classified as dryland locations because no irrigation is applied and the long-term rainfall averages are 420 mm year[-1] and 540 mm year[-1] (Western Regional Climate Center). These environments are typical of optimum dryland environments in the Pacific Northwest of the US and results cannot be extended to truly dry areas (e.g. the summer–fallow zones). At the irrigated locations (Hermiston, OR; Aberdeen, Burley, Filer and Kimberly, ID; and Fort Collins, CO) supplemental irrigation is routinely applied in accordance with local practice since average annual rainfall is below 400 mm. Under dryland (Table 3) and irrigated (Table 4) environments, there were no significant differences between varieties for yield. The test weight advantage of Alba over the other varieties was not as apparent under dryland or irrigated conditions as it was under high rainfall. Under dryland and irrigated conditions, there were no significant differences in terms of kernel plumpness. Alba was significantly taller than Charles in irrigated environments. No statistically significant differences in heading date were detected among these varieties in either dryland or irrigated environments. Lodging percentages for the varieties were variable and non-significant, in part reflecting the variability of this trait within environments.

In eight field environments, the winter survival of Alba was not statistically different from that of the other four varieties. Differential winter survival data are very difficult to obtain. The Corvallis location rarely experiences sufficiently low temperatures to cause winter injury in varieties with some level of cold tolerance. The high survival values in field trials that experienced differential winter survival over four years of testing indicate that Alba has a level of winter survival at least comparable to that of other commercially available winter barley varieties (Table 5). In addition to field data, we present the results from two controlled freeze tests (Table 6). While controlled freeze test data can only approximate field conditions, they do provide a meaningful ranking of cultivar performance.

Disease Resistance

For scald, plants were rated using a 1 (resistant) to 9 (susceptible) disease reaction score. For stripe rust, plants were rated based on percent of leaf area that was covered by

lesions at anthesis. Alba displayed significantly better resistance to scald than Maja, Eight-Twelve, and Charles, but not Strider, in five years of testing under intense natural epidemic conditions at Corvallis, OR (Table 7). Alba was significantly more resistant to stripe rust than Eight-Twelve but had similar resistance to other varieties in this trial. Alba, Maja, and Strider were all developed at Oregon State University and selected for maximum levels of adult plant resistance to stripe rust. Alba is susceptible to leaf rust (incited by *Puccinia hordei*) (Table 8).

Malt Quality

Alba was included in malting quality tests in 10 environments (Table 9). Eight-Twelve and Strider were not included in the malting quality analyses because they had previously been determined to have poor quality. Charles is currently the American Malting Barley (AMBA) winter barley check for malting quality. At the time these tests were conducted, Maja was a six-row facultative growth habit candidate cultivar for AMBA approval. Alba was included in the ten malt analyses because it is a parent of germplasm involved in genetic studies of malting quality.

Key malting parameters are grain protein, malt extract, the ratio of soluble/total protein, enzyme activity (as measured by alpha amylase and diastatic power) and wort beta glucan. There were no significant differences in grain protein. Alba was significantly lower than Charles for malt extract and soluble/total protein. Alba's diastatic power was significantly lower than that of both Maja and Charles and the level of alpha-amylase was significantly lower than that of Charles. The wort beta glucan of Alba was significantly higher than that of either Maja or Charles. Therefore, Alba does not meet current standards for malting barley as established by the AMBA (AMBA, 2010). However, lower soluble protein and enzyme levels have also been noted by AMBA as a priority for the Craft Brewing industry. Preliminary results from tests involving changes to malting and brewing protocols indicate that Alba grain, produced under high rainfall conditions can produce excellent malt and beer (M. Doehnel and W. Carpenter, Skagit Malting and Brewing, personal communication, 2014). In high rainfall environments, the significantly higher yield of Alba, as compared to Charles, may warrant the additional

effort required to produce malt and beer.

Feed, Forage and Food Quality

The limited data that are available for Alba feed, forage, and food quality can be found at "http://barleyworld.org/breeding-genetics/data".

Availability

Breeder seed is maintained by the Barley Project at Oregon State University, Corvallis, OR 97331. Seed for research purposes will be available on request from the corresponding author for at least 5 years. It is requested that appropriate recognition of source be given when this cultivar contributes to development of new germplasm or cultivars. Alba is a public release without Plant Variety Protection (PVP) and no licensing restrictions.

Acknowledgements

References

American Malting Barley Association. 2010. Malting barley breeding guidelines: Ideal
    commercial malt criteria. Available at: http://ambainc.org/content/53/production-and-

quality-information (verified Mar. 28, 2014).

Budde, A.D., C. Martens, and Staff. 2008. Mississippi valley regional spring barley nursery 2008 crop preliminary quality report. Available at http://www.ars.usda.gov/SP2UserFiles/Place/36551000/barleyreports/2008MVN.pdf (verified 23 June 2014). USDA Research Services, Midwest Area, Cereal Crop Research Unit, Madison, WI.

Filichkin, T.P., M.A. Finge, A.D. Budde, A.E. Corey, S.H. Duke, L. Gallagher, J. Helgesson, C.A. Henson, D.E. Obert, J.B. Ohm, S.E. Petrie, A.S. Ross, and P.M. Hayes. 2010. Phenotypic variation for diastatic power, β-amylase thermostability vs. allelic variation at the *BMY1* locus in a sample of North American barley germplasm. Crop Sci. 50:826-834.

Hayes, P.M., A.E. Corey, R. Dovel, R. Karow, C. Mundt, K. Rhinart, and H Vivar. 2000. Registration of 'Orca' barley. Crop Sci. 40:849-851.

Obert, D.E., D.M. Wesenberg, D.E. Burrup, B.L. Jones, and C.A. Erickson. 2006. Registration of 'Charles' barley. Crop Sci. 46:468-469.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/ (verified Dec. 8, 2013).

Skinner, J.S., P. Szűcs, J. von Zitzewitz, L. Marquez-Cedillo, T. Filichkin, E.J. Stockinger, M.F. Thomashow, T.H.H. Chen, and P.M. Hayes. 2006. Mapping of barley homologs to genes that regulate low temperature tolerance in Arabidopsis. Theor. Appl. Genet. 112:832-842.

Triticeae Coordinated Agricultural Project. 2014. The triticeae toolbox. USDA-NIFA. http://triticeaetoolbox.org/barley/ (verified Dec. 8, 2013).

Wesenberg, D.M., D.E. Burrup, M.F. Kolding, B.D. Brown, and C.R. Burnett. 1992. Registration of 'Eight-Twelve' barley. Crop Sci. 32:828-828.

Western Regional Climate Center. SOD USA Climate Archive. Available at

    http://www.wrcc.dri.edu/summary/ (verified Mar. 25, 2014).

Tables

Cultivar	Yield	Test Weight	Plump	Plant Height	Heading Date	Lodging
	kg ha^{-1}	g L^{-1}	%	cm	Julian Days	%
Alba	7412.2	669	90.3	102.8	149.3	16.4
Maja	6858.9	663	75.6	96.4	145.2	16.2
Strider	7522.7	644	82.4	96.3	146.1	21.0
Eight-Twelve	6804.7	630	74.1	95.3	146.0	21.0
Charles	6262.6	639	92.0	85.4	144.9	26.2
# of trials	33	28	28	28	23	20
LSD (p=0.05)	871.7	21	10.1	7.8	7.0	18.4

Table 1. Agronomic performance of Alba and check cultivars across 33 environments (7 high rainfall, 7 dryland, and 19 irrigated).[t]
[t]Brownsville, Corvallis, Hermiston, Junction City and Pendleton, OR; Pullman, WA; Aberdeen, Burley, Filer and Kimberly, ID; and Fort Collins, CO.

Cultivar	Yield	Test Weight	Plump	Plant Height	Heading Date	Lodging
	kg ha^{-1}	g L^{-1}	%	cm	Julian Days	%
Alba	6846	666	90.6	121.7	137.8	28.0
Maja	5961	640	56.1	106.3	125.6	50.0
Strider	6435	610	67.1	106.1	129.6	48.0
Eight-Twelve	4156	562	42.7	110.0	127.6	50.0
Charles	5008	603	86.6	89.6	120.2	63.0
# of trials	7	7	7	7	5	2
LSD (p=0.05)	1566	57	22.9	12.8	4.1	150.6

Table 2. Agronomic performance of Alba and check cultivars across 7 high rainfall environments.[t]
[t]Brownsville, Corvallis and Junction City, OR.

Cultivar	Yield	Test Weight	Plump	Plant Height	Heading Date	Lodging
	kg ha^{-1}	g L^{-1}	%	cm	Julian Days	%
Alba	6806	673	84.7	101.6	152.3	21.5
Maja	6290	679	76.0	94.3	149.3	19.8
Strider	7084	662	81.3	100.5	150.7	19.5
Eight-Twelve	6204	650	73.1	95.1	149.7	31.5
Charles	6245	645	91.7	87.1	150.0	27.2
# of trials	7	7	7	7	3	6
LSD (p=0.05)	1059	28	20.3	17.1	4.3	38.4

Table 3. Agronomic performance of Alba and check cultivars across 7 dryland environments.[t]
[t]Pendleton, OR and Pullman, WA.

Cultivar	Yield	Test Weight	Plump	Plant Height	Heading Date	Lodging
	kg ha^{-1}	g L^{-1}	%	cm	Julian Days	%
Alba	7844	666	92.9	93.9	152.6	11.8
Maja	7399	668	85.2	92.1	150.9	8.8
Strider	8085	652	90.5	88.8	150.7	17.3
Eight-Twelve	8002	653	90.3	88.1	151.5	10.8
Charles	6731	655	94.8	82.4	152.1	19.6
# of trials	19	14	14	14	15	12
LSD (*p*=0.05)	1163	20	8.5	8.9	5.7	17.8

Table 4. Agronomic performance of Alba and check cultivars across 19 irrigated environments.[t]
[t]Hermiston, OR; Aberdeen, Burley, Filer and Kimberly, ID; and Fort Collins, CO.

Cultivar	Low Temperature Tolerance
	% survival
Alba	78.3 (12-96)
Maja	76.1 (33-100)
Strider	73.1 (16-100)
Eight-Twelve	72.5 (23-100)
Charles	60.9 (9-93)
# of trials	8
LSD (*p*=0.05)	29.43

Table 5. Low temperature tolerance at Aberdeen, ID; Pullman, WA; Hermiston and Pendleton, OR; St. Paul, MN; and Bozeman, MT over the years 2008-2011. Ranges are listed in parentheses. Low temperature tolerance was only recorded in environments where differential survival was observed.

Cultivar	MRI '06	MRI '08
	% survival	% survival
Alba	85	87
Maja	78	75
Strider	58	98
Eight-Twelve	82	-
Charles	-	31

Table 6. Percent survival in controlled freeze tests at the Martonvasar Research Institute (Hungary), in 2006 and 2008. Charles and Eight-Twelve were not included in the 2006 and 2008 tests, respectively.

Cultivar	Scald	Stripe Rust
	1-9[t]	%
Alba	1.6 (1-3)	0.4 (0-2)
Maja	4.8 (1-7)	2 .0(0-7)
Strider	2.0 (1-4)	0.0 (0-0)
Eight-Twelve	5.4 (3-7)	59.6 (0-97)
Charles	7.4 (4-9)	17.6 (0-63)
# of trials	5	5
LSD (*p*=0.05)	2.3	31.2

Table 7. Disease ratings for scald (*Rhynchosporium commune*) and percent severity for barley stripe rust (*Puccinia striiformis* f. sp. *hordei*) at Corvallis, Oregon, 2008-2012. Ranges are listed in parentheses. [t]1 = most resistant, 9 = most susceptible

Cultivar	Leaf Rust
	%
Alba	75.0 (75-75)
Maja	93.3 (90-95)
Full Pint (BCD47)	0.3 (0-1)
# of replicates	3
p-value	0.002

Table 8. Percent severity for barley leaf rust (*Puccinia hordei*) from a single replicated trial at WSU Mount Vernon, WA, 2013. Ranges are listed in parentheses.

Cultivar	Malt extract	Barley protein	Wort protein	Soluble/total protein	Diastatic power	α-amylase	β-glucan
	%	%	%	%	°ASBC	D.U.	mg L^{-1}
Alba	78.3	10.5	3.6	36.4	91.4	43.3	466.4
	(72.7-81.7)	(8.3-13.8)	(3.1-4.2)	(26.8-48.8)	(67.0-126.7)	(34.6-70.6)	(111.0-720.3)
Maja	79.4	11.0	4.1	40.5	130.3	51.6	102.9
	(77.0-81.7)	(8.7-14.8)	(3.6-4.7)	(33.4-47.8)	(109.0-164.6)	(38.3-67.9)	(43.1-179.6)
Charles	81.7	11.3	4.9	47.0	125.5	84.1	165.2
	(79.3-83.6)	(9.8-13.2)	(4.6-5.5)	(38.3-57.7)	(92.8-159.9)	(73.0-97.0)	(51.5-310.1)
# of trials	10	10	10	10	10	10	10
LSD (*p*=0.05)	1.7	1.4	0.3	5.0	16.5	8.4	119.9

Table 9. Malting quality profile of Alba compared to other cultivars grown at Corvallis, Hermiston and Pendleton, OR; Aberdeen, ID; and Pullman, WA; over the years 2009-2011. Ranges are listed in parentheses.

Conclusion

In this thesis, I developed a program that can identify favorable subsets for rare-trait discovery and Genome-Wide Association Studies (GWAS), I conducted GWAS for tocochromanol concentrations in barley, and I described the barley variety "Alba" in a germplasm release.

The R program "GeneticSubsetter" which was presented in this thesis, uses genetic data to create subsets of germplasm collections. This program depends on two criteria that can be used to assign a value to a subset: Polymorphism Information Content (PIC), and the Mean of Transformed Kinships (MTK). The function SubsetterPIC removes one genotype at a time on the basis of which genotype's removal will result in the subset with the highest PIC, until only two genotype (or only genotypes that the user specified could not be removed) are left, creating a full ranking of genotype's contributions to PIC. This function returns a list identical to an excel macro published by Muñoz-Amatriaín et al. (2014), but SubsetterPIC uses a more efficient algorithm to arrive at this list, in some cases decreasing the computation time needed by more than an order of magnitude. A similar function, SubsetterMTK, uses the same single-genotype elimination approach to identify a subset, except using MTK as the criteria to judge subsets.

When tested, SubsetterPIC was able to identify subsets that contained on average 24 percent more rare traits than random subsets, but was unable to identify a subset that was significantly better for GWAS than random subsets. Alternatively, SubsetterMTK was able to identify a subset that performed better than 199 of 200 random subsets for GWAS, but there was no significant difference between subsets identified by SubsetterMTK and random subsets in terms of rare-trait discovery.

These results, and the accompanying R package, hold the potential to substantially reduce the resource requirements for a wide range of studies, with little effort on the part of the researcher. Currently, this package is only designed to accommodate inbred crops. However, with minor modifications, it could also be applied to heterozygous organisms. This could extend its utility into clonally propagated crops, and possibly human genetics.

On a more theoretical level, the fact that the benefits of these functions are limited to either rare-trait discovery or GWAS highlights the different germplasm requirements needed to be successful for the two tasks.

GWAS of a collection 1466 of elite barley accessions revealed 13 SNP markers that were significantly associated with the concentration of one or more tocochromanol forms. These 13 markers occurred in seven linkage groups across the barley genome: two on 1H, two on 6H, and three on 7H. Of these groups, one group on 6H and two groups were co-located with QTL for tocochromanol concentrations identified in a mapping population (Oliver et al. 2014). Eight markers in this study were within 2.5 cM of one or more sequences that showed homology to genes previously shown to encode enzymes important to tocotrienol production, including a sequence previously shown to encode HGGT in barley (Cahoon et al. 2003).

The characterization of tocochromanol concentrations in elite lines, in tangent with the QTL for tocochromanol concentrations identified via GWAS, could serve as a foundation for breeding for increased tocochromanol concentrations, should that become a viable objective.

Additionally, this study demonstrated the ability of GWAS to be used together with genome sequence data to identify candidate genes, which using traditional methods would generally require additional genotypic and/or phenotypic data. The availability of two studies that have previously identified QTL and candidate genes for tocochromanol concentrations in barley helps to validate the statistical methods used to account for genetic structure and differences between environments in this analysis, while demonstrating that GWAS can match or exceed the power of alternative approaches.

In the germplasm release, the barley variety 'Alba' was shown to be comparable to the otherwise leading barley variety, 'Strider.' However, in high-rainfall environments, Alba did demonstrate significant improvements over Strider in terms of grain plumpness, and non-significant improvements over Strider in terms of yield and test-weight. While significant differences between Alba and Strider were not detected, the fact that Alba was trending towards higher yield indicates that it may produce the best yield of the two varieties. The release of the barley variety Alba helps to make advanced germplasm

available to farmers, diversifying the available varieties. Also, the in-depth characterization of Alba provides important information were the variety used as a parent in future breeding efforts.

　　　　While these three projects were conducted simultaneously, they are intended to facilitate or contribute to key sequential steps in germplasm improvement, contributing to the effort to produce improved barley varieties for the Pacific Northwest.

Bibliography

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25-29.

Behall KM, Scholfield DJ, Hallfrisch J (1997) Effect of beta-glucan level in oat fiber extracts on blood lipids in men and women. J Am Coll Nutr 16: 46-51.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practival and powerful approach to multiple testing. J R Stat Soc 57:289-300.Berger GL, Liu S, Hall MD, Brooks WS, Chao S et al. (2012) Marker-trait associations in Virginia Tech winter barley identified using genome-wide mapping. Theor Appl Genet 126:693-710.

Blake VC, Kling JG, Hayes PM, Jannink JL, Jillella SR, et al. (2012) The hordeum toolbox: The barley coordinated agricultural project genotype and phenotype resource. Plant Genome 5:81-91.

Bruno RS, Leonard SW, Atkinson J, Montine TJ, Ramakrishnan R, Bray TM and Traber MG (2005) Faster plasmid vitamin E disappearance in smokers is normalized by vitamin C supplementation. Free Radic Biol Med 40:555-556.

Burke CC, Wildung MR, Croteau R (1999) Geranyl diphosphate synthesis: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. Proc Natl Acad Sci USA 96:13062-13067.

Cahoon EB, Hall SE, Ripp KG, Ganzke TS, Hitz WD and Coughlan SJ (2003) Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. Nat Biotechnol 21:1082:1087.

Cantor RM, Lange K and Sinsheimer JS (2010) Prioritizing GWAS Results: A review of statistical methods and recommendations for their application. Am J Hum Genet 86:6-22.

Chutimanitsakun Y, Cuesta-Marcos A, Chao S, Corey A, Filichkin T, et al. (2013) Application of marker-assisted selection and genome-wide asociation scanning to the development of winter food barley germplasm resources. Plant Breeding 132:563-570.

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, et al. (2009) Development and

implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582.

Collakova E, DellaPenna D (2001) Isolation and functional analysis of homogentisate phytyltransferase from *Synechocytis* sp. PCC 6803 and Arabidopsis. Plant Physiol 127:1113-1124.

Cooney RV, Franke AA, Harwood PJ, Hatch-Pigott V, Custer LM, Mordan LJ (1993). Gamma-tocopherol detoxification of nitrogen dioxide: superiority to alpha-tocopherol. Proc Natl Acad Sci 90:1771-1775

Cordero Z, Drogan D, Weikert C, Boeing H (2010) Vitamin E and risk of cardiovascular diseases: A review of epidemiological and clinical trial studies. Crit Rev Food Sci 50:420-440.

Deng W, Nickle DC, Learn GH, Maust B, Mullins JI (2007) ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics 23:2334-2336.

Desel C, Krupinska K (2005). The impact of tocochromanols on early seedling development and NO release. J Plant Physiol 162:771-776. Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250-255.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250-255.

Evangelou E, Ioannidis JPA (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14:379-389.

Filichkin TP, Finge MA, Budde AD, Corey AE, Duke SH, et al. (2010) Phenotypic variation for diastatic power, β-amylase thermostability vs. allelic variation at the *BMY1* locus in a sample of North American barley germplasm. Crop Sci 50:826-834.

Frank J (2004) Beyond vitamin E supplementation: An alternative strategy to improve vitamin E status. J of Plant Physiology 162: 834-843.

Fratianni A, Caboni MF, Irano M, Panfili G (2002) A critical comparison between traditional methods supercritical carbon dioxide extraction for the determination of tocochromanols in cereals. Eur Food Res Technol 215: 353-358.

Gey FK, Puska P, Jordan P, Ulrich MK (1991) Inverse correlation between plasm and vitamin E and mortality from ischemic heart disease in cross-cultural epidemiology. American J Clinical Nutrition: 326S-334S.

Handelman GJ, Machlin LJ, Fitch AK, Weiter JJ, Dratz EA (1985) Oral α-tocopherol supplements decrease plasma γ-tocopherol levels in humans. J Nutr 115:807-813.

Hayes PM, Corey AE, Dovel R, Karow R, Mundt C, Rhinart K, Vivar H (2000) Registration of 'Orca' barley. Crop Sci 40:849-851.

Holbrook CC, Anderson WF and Pittman RN (1993) Selection of a core collection from the U.S. germplasm collection of peanut. Crop Sci 33:859-861.

Huang HY, Appel LJ (2003) Supplementation of diets with α-tocopherol reduces serum concentrations γ- and δ-tocopherol in humans. J Nutr 133:3137-3140.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, et al. (2011). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res gkr948.

Jaiswal P (2011) Gramene database: a hub for comparative plant genomics. Methods Mol Biol 678:247–275.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348-354.

Kent WJ (2002) BLAT—The BLAST-Like Alignment Tool. Genome Res 12:656–664.

Khanna S, Roy S, Ryu H, Bahadduri P, Swaan PW, Ratan RR, Sen CK (2003) Molecular basis of vitamin E action: tocotrienol modulates 12-lipoxygenase, a key mediator of glutamate-induced neurodegeneration. J Biol Chem 278:43508-43515.

Klein EA, Thompson IM, Tangen CM, Crowley JJ, Lucia MS et al. (2011) Vitamin E and the risk of prostate cancer: the selenium and vitamin E cancer prevention trial (SELECT). JAMA 306:1549-1556.

Lee IM, Cook NR, Gaziano JM, Gordon D, Ridker PM et al. (2005). Vitamin E in the primary prevention of cardiovascular disease and cancer: the women's health study: a randomized controlled trial. JAMA 294:56-65.

Liu K, Muse SV. 2005. PowerMarker: an integrated analysis environment for genetic

marker analysis. Bioinformatics 21:2128-2129.

Maeda H, Song W, Sage TL, DellaPenna D (2006) Tocopherols play a crucial role in low-temperature adaptation and phloem loading in arabidopsis. Plant Cell 18:2710-2732.

Mahajan RK, Bisht IS, Agrawal RC and Rana RS (1996) Studies on South Asian okra collection: Methodology for establishing a representative core set using characterization data. Genet Resour Crop Evol 43:249-255.

Matsushita Y, Kang WK, Charlwood BV (1996) Cloning and analysis of a cDNA encoding farnesyl diphosphate synthase from Artemisia annua. Gene 172:207-209.

Mather DE, Tinker NA, LaBerge DE, Edney BL, Jones BL et al. (1997) Regions of the genome that affect grain and malt quality in a North American two-row barley cross. Crop Sci 37:544- 554.

Mohammadi M, Endelman JB, Nair S, Chao S, Jones SS et al. (2014) Association mapping of grain hardness, polyphenol oxidase, total phenolics, amylose content, and β-glucan in US barley breeding germplasm. Mol Breeding 34:1229-1243.

Munné-Bosch S (2005) The role of α-tocopherol in plant stress tolerance. Plant Physiol 162:743-748.

Muñoz-Amatriaín M, Cuesta-Marcos A, Endelman JB, Comadran J, Bonman JM (2014) The USDA barley core collection: genetic diversity, population structure, and potential for genome-wide association studies. PloS One 9:e94688.

Muñoz-Amatriaín M, Cuesta-Marcos A, Hayes PM, Meuhlbauer GJ (2014) Barley genetic variation: implications for crop improvement. Brief Funct Genomics 13:341-350.

Muñoz-Amatriaín M, Moscou MJ, Bhat PR, Svensson JT, Bartoš J et al. (2011) An improved consensus linkage map of barley based on flow-sorted chromosomes and single-nucleotide polymorphism markers. Plant Genome 4:238-249.

Obert DE, Wesenberg DM, Burrup DE, Jones BL, Erickson CA (2006) Registration of 'Charles' barley. Crop Sci 46:468-469.

Oliver RE, Islamovic E, Obert DE, Wise ML, Herrin LL et al. (2014) Comparative systems biology reveals allelic variation modulating tocochromanol profiles in barley (*Hordeum vulgare* L.). Plos One 9:e96276.

Qureshi AA, Burger WC, Peterson DM, Elson CE (1986) The structure of an inhibitor of cholesterol biosynthesis isolated from barley. J Biol Chem 261:10544-10550.

Qureshi AA, Qureshi N, Wright JJK, Shen Z, Kramer G et al. (1991) Lowering of serum cholesterol in hypercholesterolemic humans by tocotrienols (palmvitee). Am J Clin Nutr 53:1021S-1026S.

Roberts JM, Myatt L, Spong CY, Thom EA, Hauth JC et al. (2010) Vitamins C and E to prevent complications of pregnancy-associated hypertension. N Engl J Med 362:1282-1291.

Sadre R, Gruber J, Frentzen M (2006) Characterization of homogentisate prenyltransferases involved in plastoquinine-9 and tocochromanol biosynthesis. Febs Lett 580:5357-5362.

Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM et al. (2010) Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med 362:1675-1685.

Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, DellaPenna D (2004) Vitamin E Is Essential for Seed Longevity and for Preventing Lipid Peroxidation during Germination. Plant Cell 16:1419-1432

Sen CK, Khanna S, Roy R (2007) Tocotrienols in health and disease: The other half of the natural vitamin E family. Mol Aspects Med 28:692-728.

Shintani D, DellaPenna D (1998) Elevating the vitamin E content of plants through metabolic engineering. Science 282:2098-2100.

Skinner JS, P. Szűcs, von Zitzewitz J, Marquez-Cedillo L, Filichkin T, et al. (2006) Mapping of barley homologs to genes that regulate low temperature tolerance in Arabidopsis. Theor Appl Genet 112:832-842.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ et al. (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays L.*): comparisons with data from RFLPS and pedigree. Theor Appl Genet 95:163-173.

Szűcs P, Blake VC, Bhat PR, Chao S, Close TJ, et al. (2009) An integrated resource for barley linkage map and malting quality QTL alignment. Plant Genome 2:134-140.

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711-716.

Theriault A, Chao JT, Wang Q, Gapor A, Adeli K (1999) Tocotrienol: a review of its therapeutic potential. Clin Biochem 32:309-319.

Thompson JN, Hatina G (1979) Determination of tocopherols and tocotrienols in foods and tissues by high performance liquid chromatography. J Liquid Chrom 2: 327-344.

Traber MG, Arai H (1999) Molecular mechanisms of vitamin E transport. Annu Rev Nutr 19:343-55.

Traber MG, Burton GW, Huges L, Ingold KU, Hidaka H et al. (1992) Discrimination between forms of vitamin E by humans with and without genetic abnormalities of lipoprotein metabolism. J Lipid Res 33:1171-1182.

Upadhyaya HD, Bramel PJ and Singh S (2001) Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci 41:206-210.

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG, Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci 49:1769:1780.

Watts EJ, Shen Y, Lansky EP, Nevo E, Bobe G, Traber MG (2014) High environmental stress yields greater tocotrienol content while changing vitamin E profiles of wild emmer wheat seeds. J Med Food.

Wesenberg DM, Burrup DE, Kolding MF, Brown BD, Burnett CR (1992) Registration of 'Eight-Twelve' barley. Crop Sci 32:828-828.

Yang JL, Kim YH, Lee MS, Moon YK (2003) Barley beta-glucan lowers serum cholesterol based on the up-regulation of cholesterol 7alpha-hydroxylase activity and mRNA abundance in cholesterol-fed rats. J Nutr Sci Vitaminol 49:381-387.

Zewdie Y, Tong N, Bosland P (2004) Establishing a core collection of Capsicum using a cluster analysis with enlightened selection of accessions. Genet Resour Crop Evol 51:147-151.

APPENDICES

Appendix A- Supplemental GWAS analyses for tocochromanol concentration in barley

Introduction

In the manuscript titled "Quantitative trait loci associated with the tocochromanol (vitamin E) pathway in barley," 1,534 spring barley accessions were grown in Bozeman, Montana, and their grain was characterized for the concentration each of eight tocochromanol forms. In 2006 and 2007, in the same years as the spring-habit accessions were grown in Bozeman, Montana, each of 318 winter-habit barley accessions were grown in either Corvallis, Oregon or Blacksburg, Virginia, based on each accession's origin.

This appendix contains five additional analyses: one where only winter-habit barley accessions are considered, one where only two-rowed spring-habit accessions are considered, one where only six-rowed spring-habit accessions are considered, one where only accessions grown in Montana in 2006 are considered, and one where only accessions grown in Montana in 2007 are considered.

Methods

The analysis for the winter-habit accessions presented in this supplementary file was based on 318 winter-habit barley accessions from the Barley Coordinated Agriculture Project (Barley CAP), a predecessor to the Triticeae Coordinated Agricultural Project (TCAP; http://www.triticeaecap.org/, verified 26 October 2014).

Phenotypic data and genotypic data for these accessions were collected using the same methods as described in the main paper. GWAS was also conducted using the same methods as described in the main paper.

Results and Discussion

Winter-habit accessions

Phenotypic data for the winter lines was comparable to phenotypic data for the spring lines for each tocochromanol form (Figure 1).

GWAS for the winter-habit accessions did not identify any of the QTL that were identified in the main paper. However, it did identify five markers that were significantly associated with BT: two on chromosome 2H, one on chromosome 5H, and two on chromosome 7H (Figures 2, 3, 4; Tables 1, 2, 3). However, several factors make the QTL identified using these winter-habit accessions may be less reliable than those identified using the spring-habit accessions. First, we were not able to link these QTL to QTL identified using the spring-habit accessions, QTL that had been identified by Oliver et al., or to candidate genes that were clearly associated with tocochromanol biosynthesis. Also, a relatively small number of winter-habit accessions were used, which can increase the probability that a seemingly significant marker is in fact a false-positive, as described by Wang et al. (2012). However, even with these considerations in mind, this analysis did identify marker-trait associations that had a significance level that was well above the FDR threshold, which suggests that these markers are linked to real QTL for BT.

Two-row accessions

The analysis of two-row spring-habit accessions revealed generally the same groups of significant marker as seen in the overall analysis of spring-habit accessions (Figures 5, 6, 7; Tables 4, 5). One marker was marginally significant on chromosome 2H that was not close to any significant markers detected in the overall analysis.

Six-row accessions

The analysis of six-row spring-habit accessions revealed generally the same groups of significant marker as seen in the overall analysis of spring-habit accessions (Figures 8, 9, 10; Tables 6, 7). One marker was significant on chromosome 5H that was not close to any significant markers detected in the overall analysis.

Accessions grown in 2006

The analysis of spring-habit accessions grown in 2006 revealed generally the same groups of significant marker as seen in the overall analysis of spring-habit accessions (Figures 11, 12, 13; Tables 8, 9).

Accessions grown in 2007

The analysis of spring-habit accessions grown in 2007 revealed generally the same groups of significant marker as seen in the overall analysis of spring-habit accessions (Figures 14, 15, 16; Tables 10,11).

Conclusion

The analysis of the winter-habit accessions revealed a different set of significant markers than the analyses of the spring-habit accessions. This could be because different regions of the genome tend to control tocochromanol concentrations as a result of different conditions, or a different genetic background, or because the QTL detected for the winter-habit accessions are in fact false positives, as would be suggested by the failure to detect plausible candidate genes in these new regions.

In general, the smaller analyses of spring-habit accessions revealed fewer significant markers than the overall analysis. Few markers fell outside of the groups of markers identified in the overall analysis of spring-habit accessions, with the exceptions being of one marker on chromosome 2H, and one marker on chromosome 5H.

Tables

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_21037	2H	101.3	0.27	0.11	7.84	0.86	0.35	0.03	0.39	0.07	0.08	0.76	0.28
11_20080	2H	107.0	0.23	0.36	4.15	0.28	0.22	0.20	0.11	0.52	0.43	0.38	0.46
12_20350	5H	61.9	2.85	1.11	5.08	0.32	0.99	0.21	0.90	0.31	1.00	3.80	1.43
11_20750	7H	61.7	0.24	0.18	6.42	0.39	0.21	0.56	0.16	0.50	0.26	0.02	0.21
12_10888	7H	157.9	0.30	1.32	4.21	0.43	0.44	0.59	0.93	0.25	1.04	0.77	1.05

Table 1. Significance ($-\log_{10}(p)$) of markers for concentration of all tocochromanol forms, fractions, and Total Tocochromanol (TTC). Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_21037	2H	101.3	0.24	0.44	0.29	0.67	0.02	-0.01	0.18	-0.10	0.38	0.57	1.57
11_20080	2H	107.0	0.16	0.87	0.16	0.20	0.01	-0.04	0.04	0.39	1.24	0.26	1.75
12_20350	5H	61.9	0.90	1.93	0.19	-0.30	0.03	0.04	-0.31	0.26	2.23	1.10	3.70
11_20750	7H	61.7	-0.19	0.58	0.24	-0.35	0.01	-0.10	-0.09	0.46	0.99	0.02	1.08
12_10888	7H	157.9	0.21	2.34	0.16	0.31	0.02	-0.09	0.27	0.24	2.48	0.44	3.28

Table 2. Best Linear Unbiased Estimators (BLUEs) for concentration (mg/kg) of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for winter-habit barley accessions. Positive values indicate that individuals with the "A" allele has a higher tocochromanol concentration, and negative values indicates that genotypes with the "B" allele have a higher tocochromanol concentration. Highlighted values show significant marker-trait associations.

SNP Marker or Annotated Genome Sequence	Chromsome	Position (cM)*	Position (bp)**	Sequence Annotation
11_21037	2H	101.3	534833906	SNP Marker
11_20080	2H	107.0	543098152	SNP Marker
MLOC_14023	2H	107.2	554948963-554953749	Unknown Terpene Synthase-like Enzyme
MLOC_38010	2H	107.0	554881424-554883134	Unknown Terpene Synthase-like Enzyme
MLOC_56110	2H	106.9	554721317-554726288	Unknown Terpene Synthase-like Enzyme
12_20350	5H	61.9	399268402	SNP Marker
11_20750	7H	61.7	90238942	SNP Marker
12_10888	7H	157.9	594325148	SNP Marker

Table 3. Significant SNPs associated with tocochromanols, and annotated sequences known or predicted to be associated with the tocochromanol biosynthesis pathway that occurred within 2.5 cM of a significant marker.

*Linkage map positions (Muñoz-Amatriaín et al. 2011)

**Genome sequence positions, (International Barley Genome Sequencing Consortium 2012)

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_10586	1H	127.6	0.06	0.20	6.74	1.05	0.52	0.08	1.25	0.12	0.08	0.12	0.16
12_30781	2H	13.4	4.77	2.86	0.04	0.00	0.29	0.14	0.00	0.85	2.77	4.35	2.44
11_20710	7H	2.5	0.40	3.23	0.69	0.48	0.36	1.01	0.09	4.99	4.43	0.47	3.66
11_21209	7H	136.0	0.56	1.95	0.46	2.66	9.95	0.51	9.05	0.55	1.68	0.84	2.64
11_10861	7H	138.8	3.24	6.15	1.45	1.41	1.48	1.25	0.50	3.70	6.75	3.48	5.25
11_21104	7H	138.8	4.26	2.12	1.30	0.39	5.18	0.18	3.05	0.86	2.07	4.71	3.55
11_10797	7H	141.4	0.03	1.56	0.19	0.49	5.36	0.09	1.33	0.36	1.39	0.11	1.60
12_10973	7H	142.4	0.45	2.27	0.29	5.58	4.16	3.58	2.52	2.75	3.01	0.28	2.96
12_30380	7H	142.4	0.73	0.66	0.12	2.22	4.58	0.41	1.12	0.11	0.57	0.55	0.83
11_10885	7H	145.2	0.25	3.58	0.50	9.39	0.23	6.55	0.03	4.12	5.05	0.20	4.17

Table 4. Significance (-$\log_{10}(p)$) of markers for concentration of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for two-row spring-habit barley accessions. Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_10586	1H	127.6	-0.02	-0.14	0.05	0.27	0.01	0.00	0.16	0.03	-0.07	0.04	0.20
12_30781	2H	13.4	-0.90	-1.47	0.00	0.00	0.01	-0.01	0.00	-0.20	-1.69	-0.90	-2.34
11_20710	7H	2.5	0.15	1.36	0.02	0.22	0.01	0.05	-0.03	0.51	1.92	0.18	2.58
11_21209	7H	136.0	0.15	0.74	0.01	0.53	0.05	0.02	0.54	0.09	0.79	0.20	1.56
11_10861	7H	138.8	-0.39	-1.22	-0.02	-0.32	-0.01	-0.03	0.08	-0.27	-1.51	-0.43	-1.97
11_21104	7H	138.8	-0.46	-0.66	-0.02	-0.12	-0.03	-0.01	-0.24	-0.11	-0.77	-0.50	-1.57
11_10797	7H	141.4	-0.01	-0.68	0.01	-0.17	-0.04	-0.01	-0.18	-0.07	-0.74	-0.04	-1.22
12_10973	7H	142.4	-0.14	0.94	0.01	0.87	0.04	0.09	0.29	0.31	1.30	-0.10	1.93
12_30380	7H	142.4	0.18	-0.36	0.00	-0.45	-0.03	-0.02	-0.15	-0.02	-0.38	0.15	-0.74
11_10885	7H	145.2	0.07	-0.97	-0.01	-0.90	0.00	-0.10	-0.01	-0.31	-1.39	0.06	-1.88

Table 5. Best Linear Unbiased Estimators (BLUEs) for concentration (mg/kg) of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for two-row spring-habit barley accessions. Positive values indicate that individuals with the "A" allele has a higher tocochromanol concentration, and negative values indicates that genotypes with the "B" allele have a higher tocochromanol concentration. Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_10586	1H	127.6	0.09	0.24	2.82	5.53	0.30	0.56	0.16	0.20	0.26	0.03	0.88
11_10363	5H	138.4	0.69	0.11	0.87	1.45	0.87	4.48	5.78	2.61	0.26	0.80	1.15
12_10811	6H	54.7	0.24	1.13	0.31	0.99	2.11	1.20	4.68	0.73	1.17	0.27	0.44
11_10954	6H	58.9	0.12	0.06	4.84	0.99	1.26	0.25	0.63	0.64	0.02	0.23	0.08
11_21209	7H	136.0	0.05	0.58	1.12	3.09	5.22	1.66	1.77	0.16	0.60	0.12	1.14
11_10861	7H	138.8	0.29	2.18	0.99	0.73	7.18	0.41	2.58	1.26	2.26	0.55	2.59
12_31282	7H	145.2	0.04	4.70	0.28	4.62	0.40	3.42	1.25	2.90	5.58	0.01	4.48

Table 6. Significance (-$\log_{10}(p)$) of markers for concentration of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for six-row spring-habit barley accessions. Highlighted values show significant marker-trait associations.

81

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_10586	1H	127.6	-0.03	0.21	0.04	0.89	0.01	0.03	-0.04	0.05	0.27	0.01	0.96
11_10363	5H	138.4	-0.33	0.19	-0.04	-0.65	-0.02	-0.19	-0.67	-0.54	-0.45	-0.38	-1.97
12_10811	6H	54.7	-0.13	1.03	-0.01	0.49	-0.04	0.08	-0.57	0.22	1.25	-0.15	0.89
11_10954	6H	58.9	-0.07	-0.10	-0.09	-0.50	0.03	-0.03	0.17	0.21	0.03	-0.13	-0.21
11_21209	7H	136.0	-0.02	0.41	0.02	0.72	0.04	0.06	0.24	0.04	0.49	0.04	1.09
11_10861	7H	138.8	-0.08	-0.88	-0.02	-0.27	-0.04	-0.02	-0.28	-0.17	-1.05	-0.14	-1.63
12_31282	7H	145.2	-0.01	1.50	0.01	0.88	0.01	0.09	0.18	0.31	1.91	0.00	2.41

Table 7. Best Linear Unbiased Estimators (BLUEs) for concentration (mg/kg) of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for six-row spring-habit barley accessions. Positive values indicate that individuals with the "A" allele has a higher tocochromanol concentration, and negative values indicates that genotypes with the "B" allele have a higher tocochromanol concentration. Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	δT	δT3	γT3	TT3	TTP	TTC
11_20021	1H	109.8	3.45	6.26	1.30	0.23	1.67	2.35	6.20	3.54	5.80
11_10586	1H	127.6	1.06	0.50	6.09	0.13	0.12	0.09	0.39	0.71	0.29
11_20709	6H	72.8	0.14	0.26	0.30	1.10	1.55	4.34	0.71	0.10	0.67
12_30296	7H	1.1	0.36	2.30	0.05	0.39	1.25	4.71	3.00	0.32	2.43
11_21201	7H	95.0	0.17	0.14	5.99	4.04	0.03	0.10	0.19	0.49	0.23
11_21209	7H	136.0	0.14	1.61	0.84	9.28	1.63	0.91	1.66	0.39	2.13
11_10861	7H	138.8	1.82	6.84	3.08	5.12	2.21	3.58	7.16	2.36	6.74
11_10797	7H	141.4	0.23	1.44	0.13	4.40	0.03	0.58	1.42	0.38	2.02

Table 8. Significance ($-\log_{10}(p)$) of markers for concentration of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for spring-habit barley accessions grown in 2006. Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	δT	δT3	γT3	TT3	TTP	TTC
11_20021	1H	109.8	0.91	3.12	0.04	0.01	0.05	0.39	3.53	0.97	4.82
11_10586	1H	127.6	-0.24	-0.35	0.05	0.00	0.00	0.02	-0.32	-0.19	-0.36
11_20709	6H	72.8	-0.07	0.31	-0.01	0.02	0.03	0.46	0.75	-0.06	1.02
12_30296	7H	1.1	0.15	1.32	0.00	-0.01	0.03	0.45	1.76	0.14	2.20
11_21201	7H	95.0	-0.07	0.14	-0.06	-0.04	0.00	-0.02	0.21	-0.17	-0.34
11_21209	7H	136.0	0.05	0.74	0.01	0.06	0.02	0.11	0.86	0.12	1.41
11_10861	7H	138.8	-0.28	-1.49	-0.03	-0.04	-0.02	-0.23	-1.74	-0.34	-2.38
11_10797	7H	141.4	-0.09	-0.86	0.00	-0.05	0.00	-0.10	-0.97	-0.14	-1.71

Table 9. Best Linear Unbiased Estimators (BLUEs) for concentration (mg/kg) of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for spring-habit barley accessions grown in 2006. Positive values indicate that individuals with the "A" allele has a higher tocochromanol concentration, and negative values indicates that genotypes with the "B" allele have a higher tocochromanol concentration. Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_10586	1H	127.6	1.11	0.34	5.24	2.96	0.99	0.04	0.66	0.07	0.29	1.56	1.41
12_10811	6H	54.7	0.55	1.07	0.00	1.13	4.29	0.99	4.01	0.19	0.75	0.64	0.24
11_10954	6H	58.9	0.18	0.41	4.35	1.04	0.12	0.03	0.25	0.18	0.24	0.38	0.49
11_21209	7H	136.0	1.30	2.12	1.40	4.21	15.06	1.47	9.99	0.14	1.95	1.81	4.84
12_10973	7H	142.4	0.63	1.42	0.85	6.87	5.71	4.96	2.35	2.68	2.67	0.37	3.66
11_10885	7H	145.2	0.05	1.86	0.64	9.61	0.13	8.96	0.03	4.00	3.56	0.01	4.43

Table 10. Significance ($-\log_{10}(p)$) of markers for concentration of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for spring-habit barley accessions grown in 2007. Highlighted values show significant marker-trait associations.

Marker	Chromosome	Position (cM)	αT	αT3	βT	βT3	δT	δT3	γT	γT3	TT3	TTP	TTC
11_10586	1H	127.6	0.18	0.18	0.05	0.43	0.01	0.00	0.08	0.02	0.20	0.24	0.99
12_10811	6H	54.7	-0.19	0.71	0.00	0.39	-0.04	0.08	-0.43	-0.08	0.70	-0.22	0.46
11_10954	6H	58.9	-0.08	-0.37	-0.07	-0.41	0.00	0.00	0.07	0.08	-0.31	-0.15	-0.83
11_21209	7H	136.0	0.23	0.72	0.03	0.57	0.05	0.07	0.45	0.04	0.87	0.30	2.26
12_10973	7H	142.4	-0.17	0.69	0.02	0.91	0.04	0.17	0.24	0.40	1.27	-0.12	2.35
11_10885	7H	145.2	0.01	-0.62	-0.01	-0.84	0.00	-0.18	0.01	-0.39	-1.15	0.00	-2.01

Table 11. Best Linear Unbiased Estimators (BLUEs) for concentration (mg/kg) of all tocochromanol forms, fractions, and Total Tocochromanol (TTC) for spring-habit barley accessions grown in 2007. Positive values indicate that individuals with the "A" allele has a higher tocochromanol concentration, and negative values indicates that genotypes with the "B" allele have a higher tocochromanol concentration. Highlighted values show significant marker-trait associations.
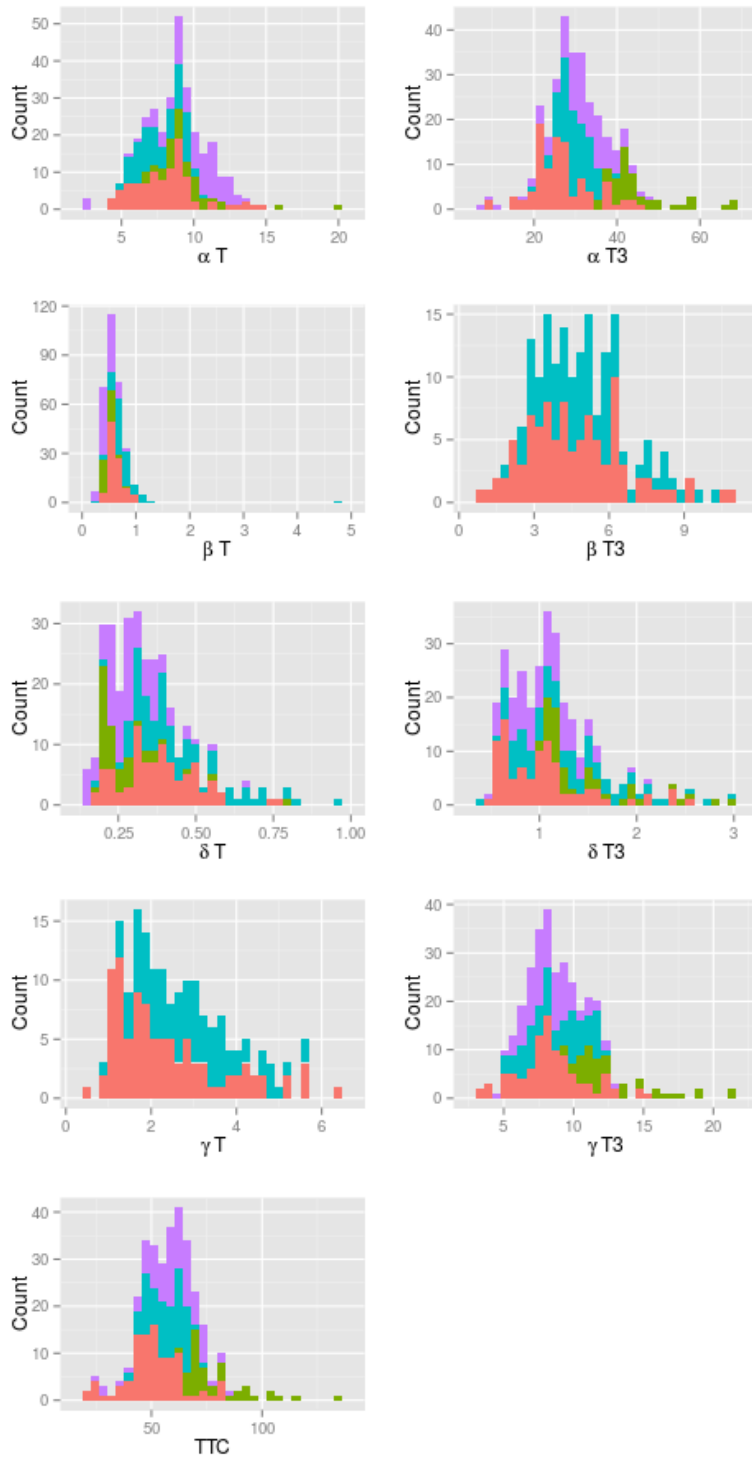
Figures

Figure 1. Distributions of concentrations of all tocochromanol forms and Total Tocochromanol (TTC) in winter-habit barley accessions. Red represents the accessions grown in Oregon in 2007, green represents the accessions grown in Oregon in 2006, blue represents the accessions grown in Virginia in 2007, and purple represents the accessions grown in Virginia in 2006. Reliable data for βT3 and γT are not available for 2006.
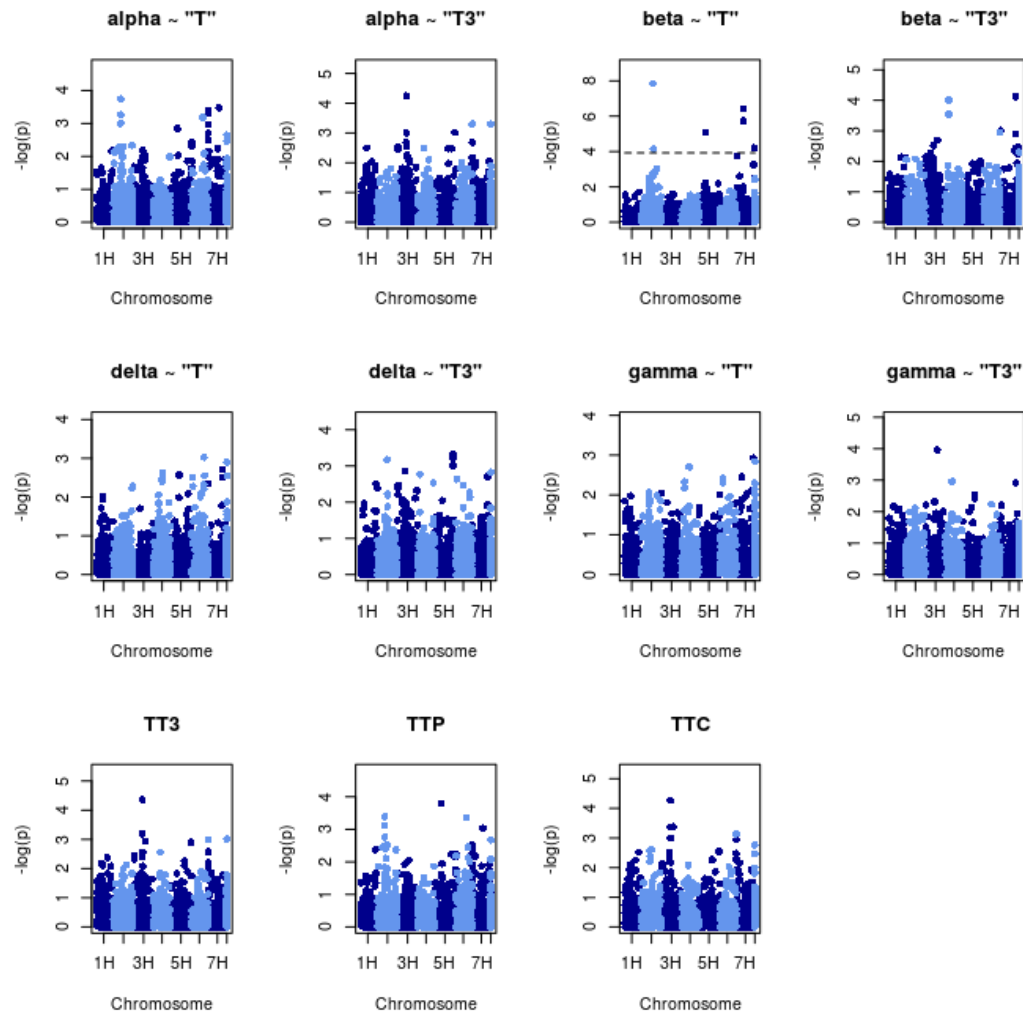
Figure 2. Manhattan plots showing results of GWAS for concentrations of all tocochromanol forms, fractions, and Total Tocochromanol (TTC), for winter-habit accessions. In analyses where one or more markers met the significance threshold determined by a false-discovery rate adjustment, a dotted line shows the significance threshold. Points in light blue, to the right of chromosome 7H, represent unmapped markers.
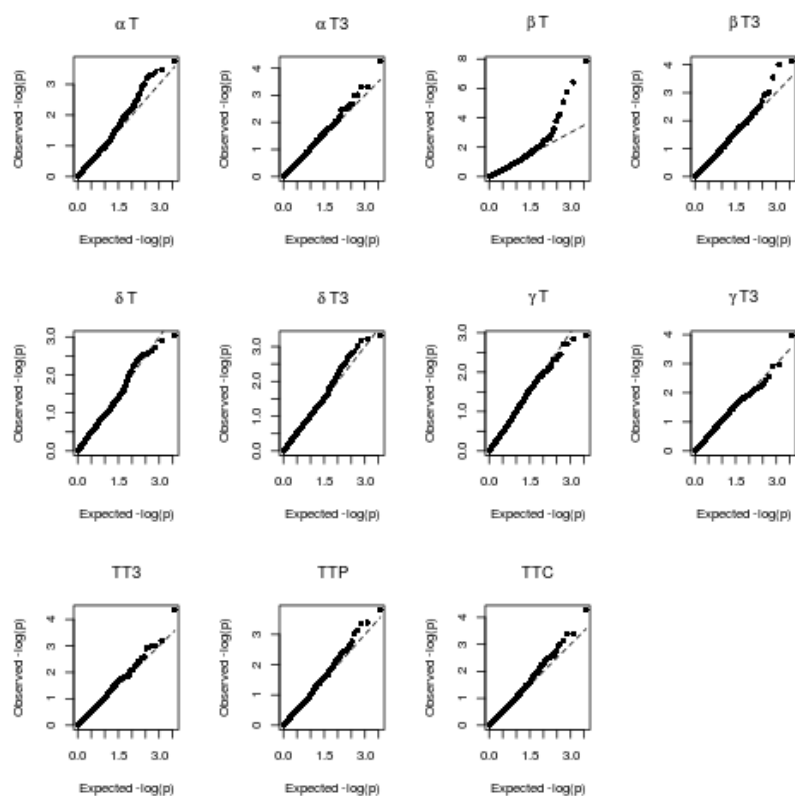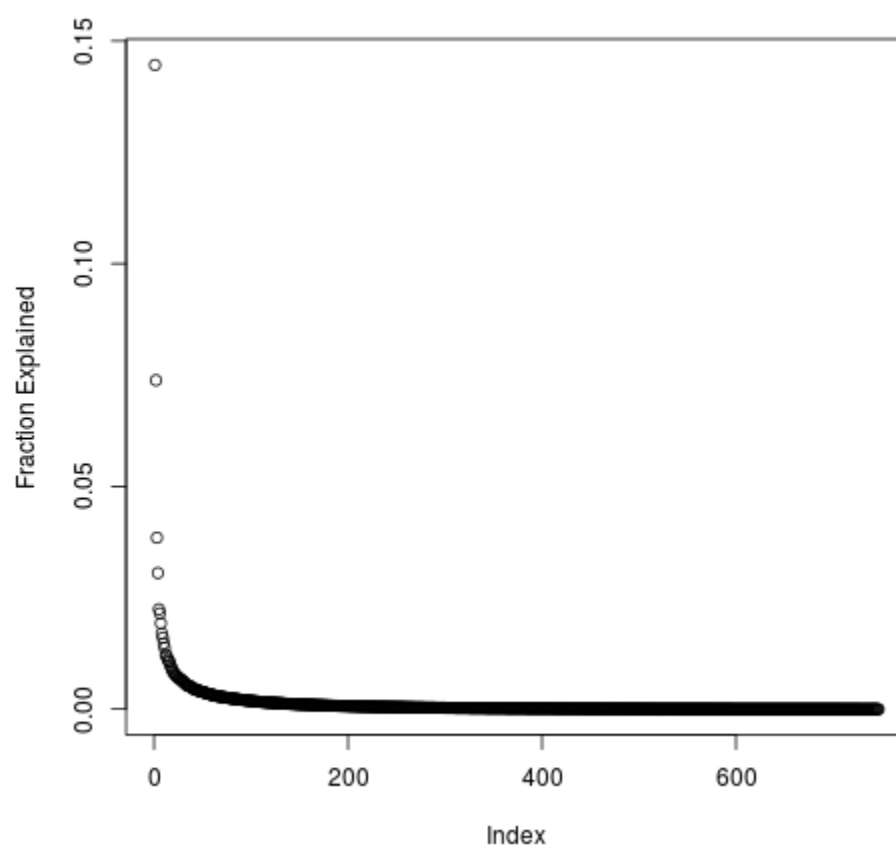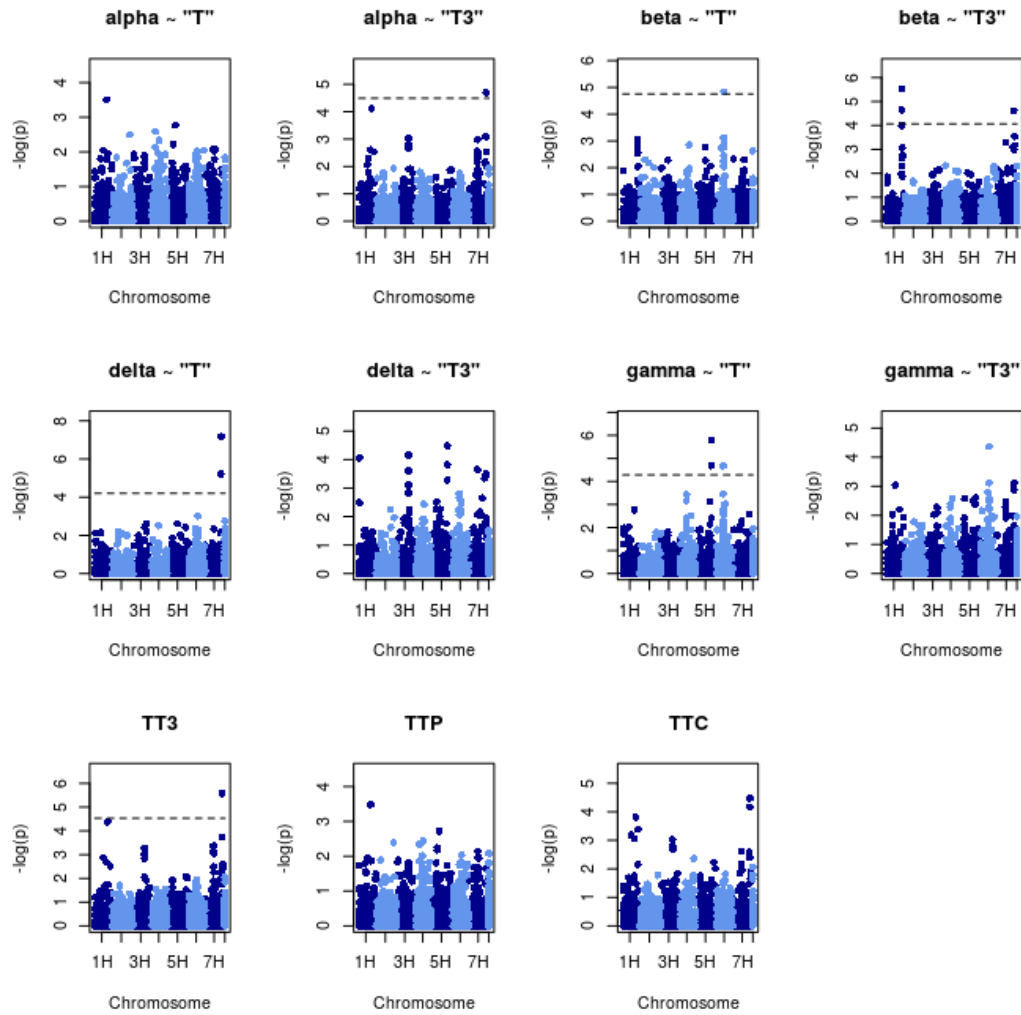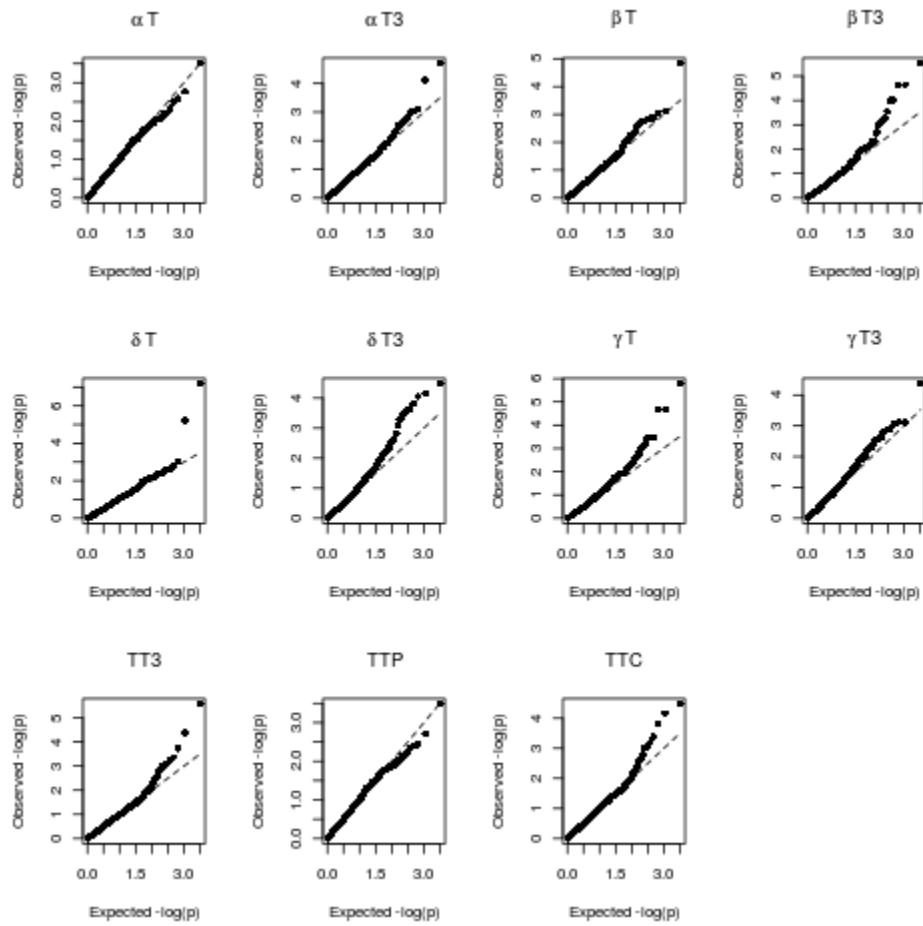
Figure 3. Q-Q plots showing the distribution of p-values, plotted against the expected distribution of p-values, for winter-habit accessions.
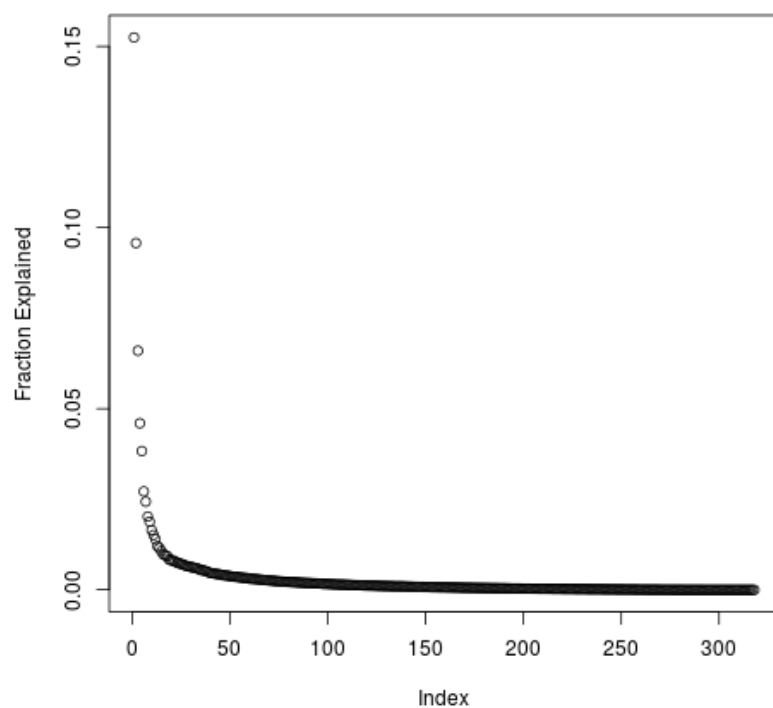
Figure 4. Fraction of genetic variance for this population explained by each principal component for winter-habit accessions.
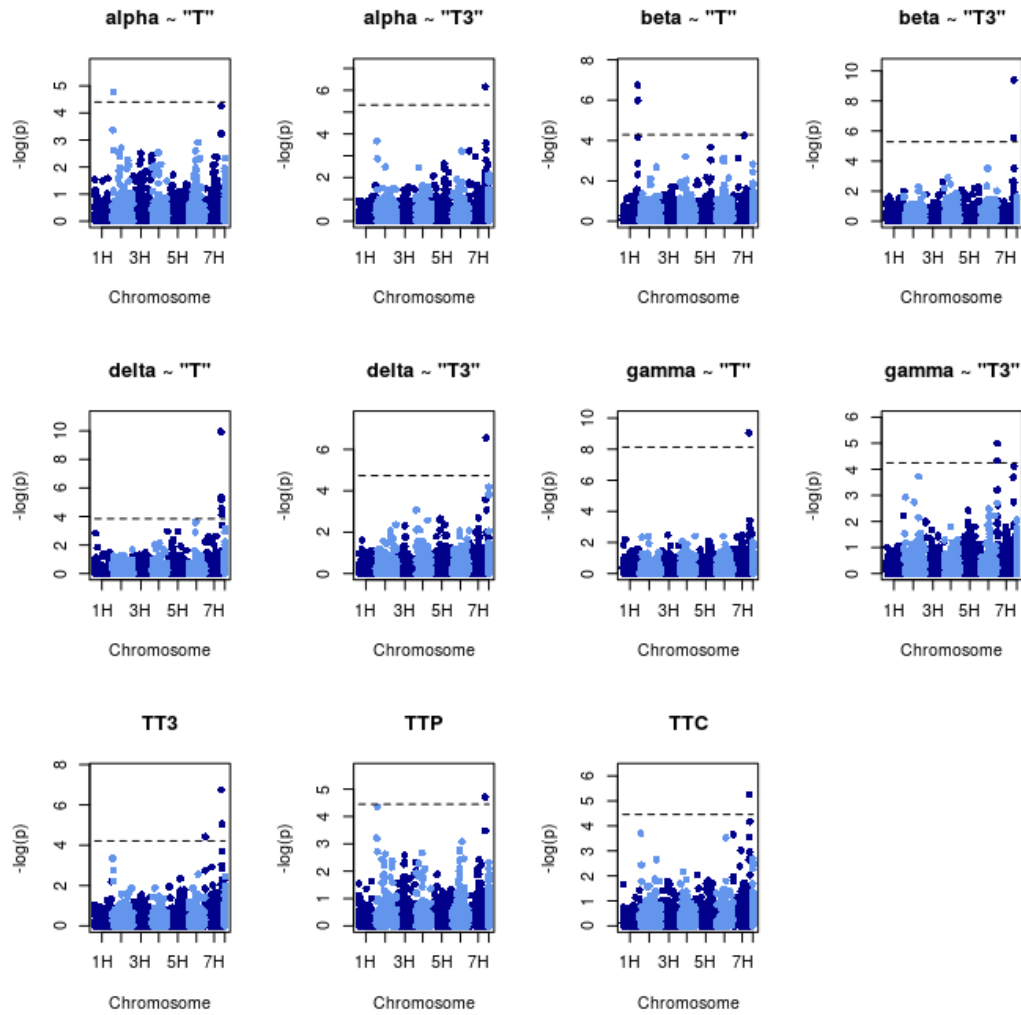
Figure 5. Manhattan plots showing results of GWAS for concentrations of all tocochromanol forms, fractions, and Total Tocochromanol (TTC), for two-row spring-habit accessions. In analyses where one or more markers met the significance threshold determined by a false-discovery rate adjustment, a dotted line shows the significance threshold. Points in light blue, to the right of chromosome 7H, represent unmapped markers.
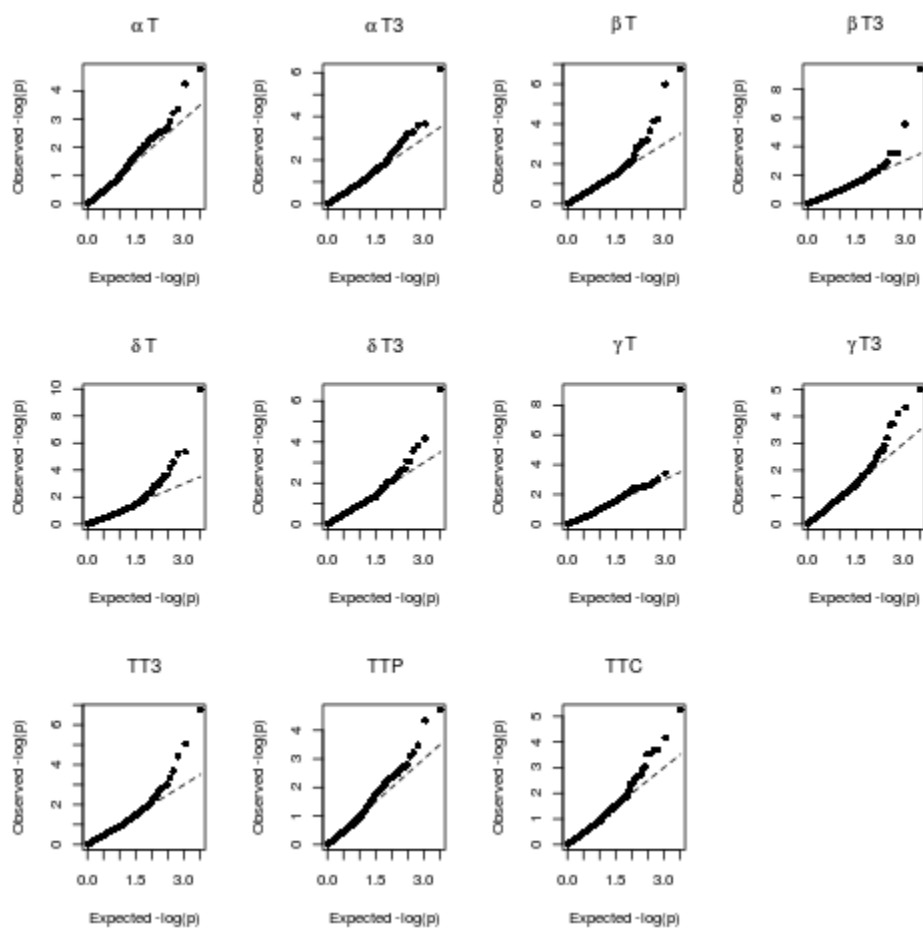
Figure 6. Q-Q plots showing the distribution of p-values, plotted against the expected distribution of p-values, for two-row spring-habit accessions.
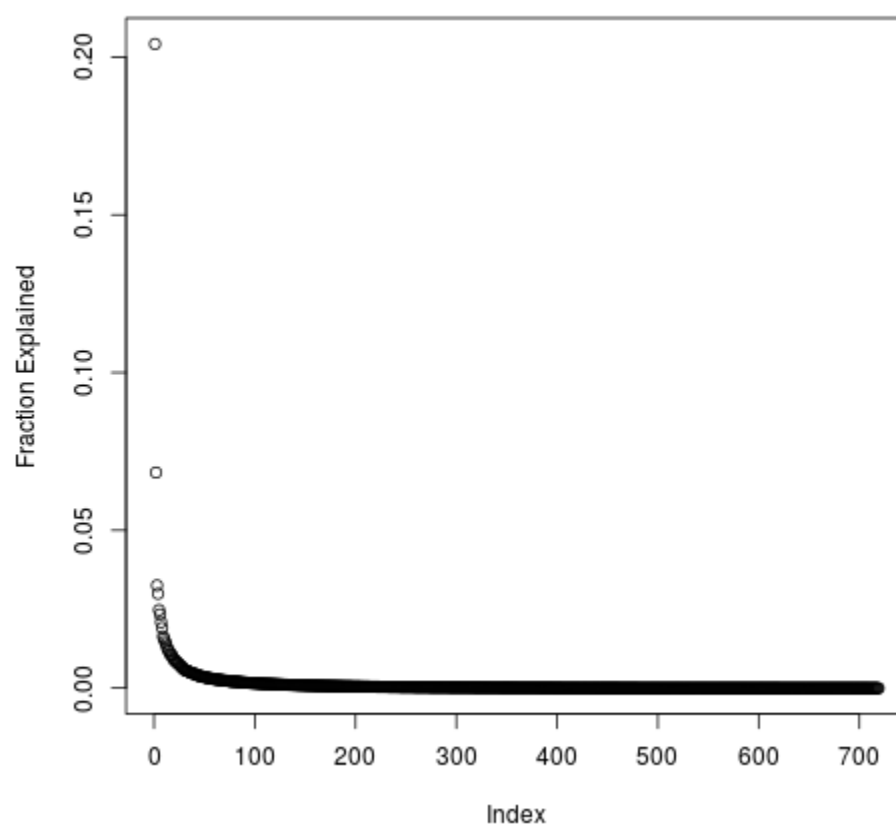
Figure 7. Fraction of genetic variance for this population explained by each principal component, for two-row spring-habit accessions.
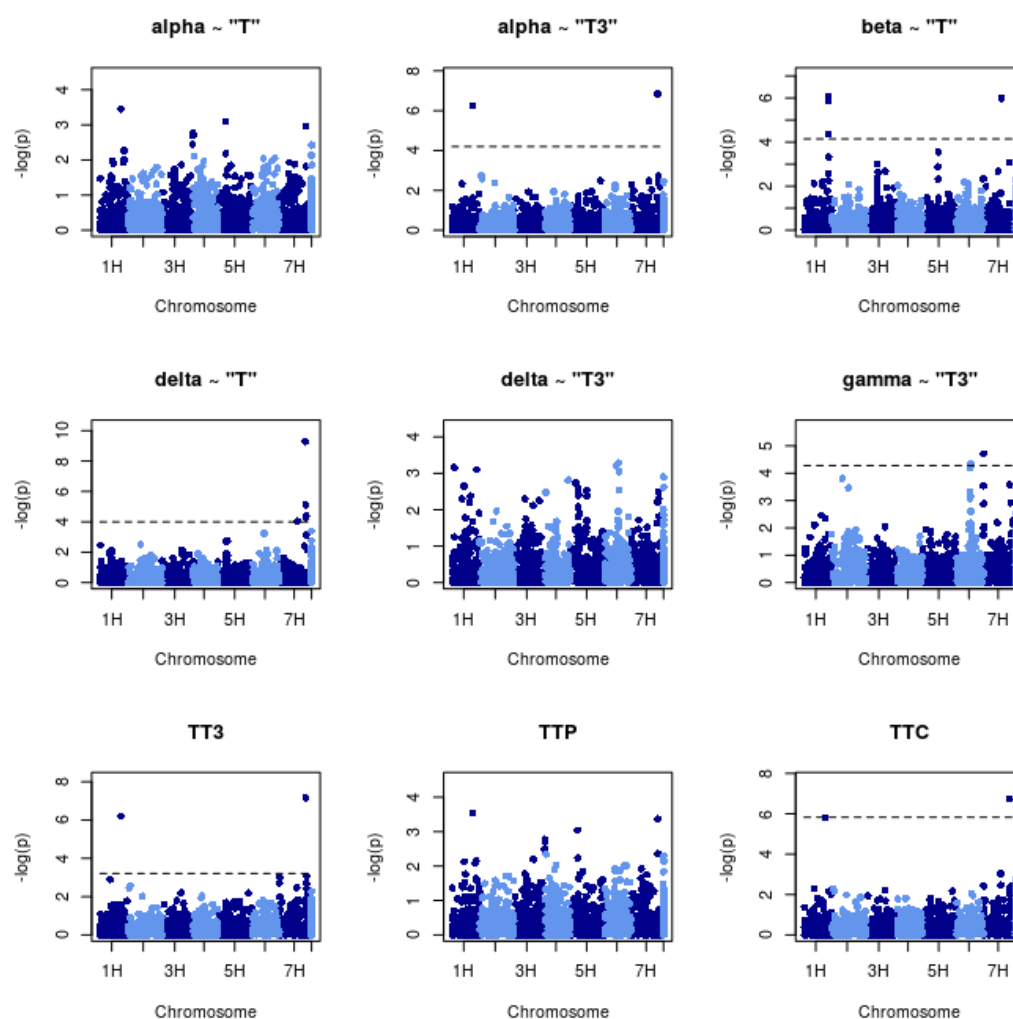
Figure 8. Manhattan plots showing results of GWAS for concentrations of all tocochromanol forms, fractions, and Total Tocochromanol (TTC), for six-row spring-habit accessions. In analyses where one or more markers met the significance threshold determined by a false-discovery rate adjustment, a dotted line shows the significance threshold. Points in light blue, to the right of chromosome 7H, represent unmapped markers.
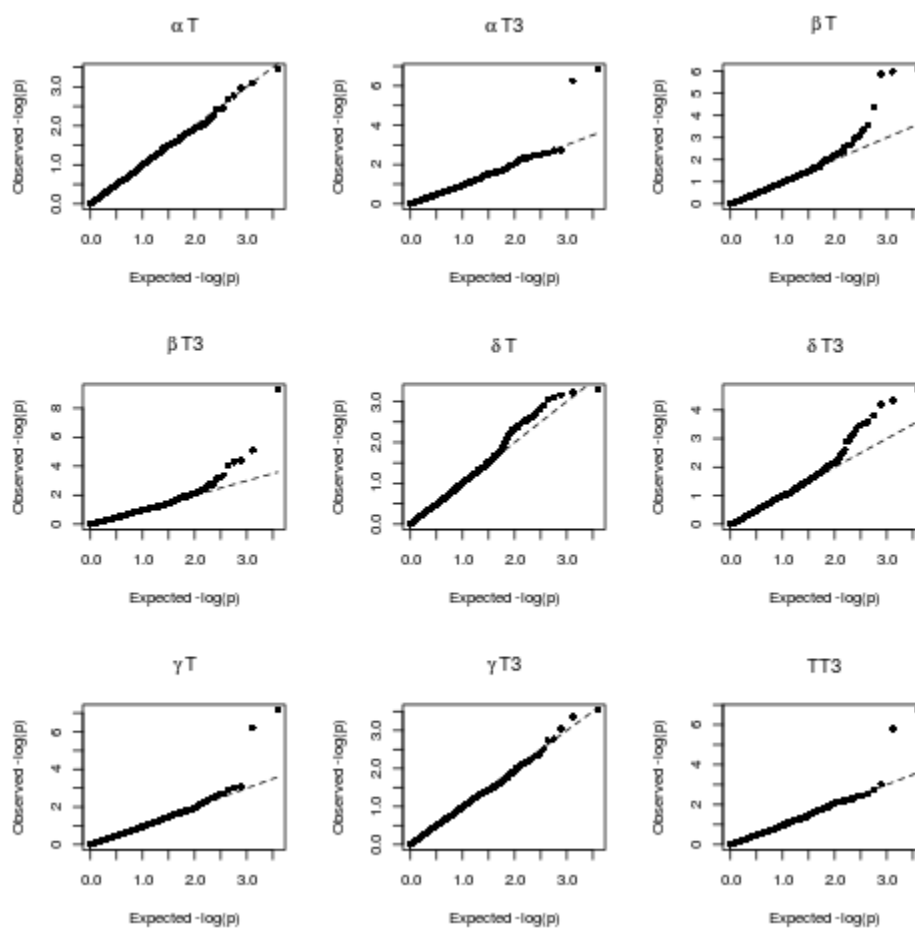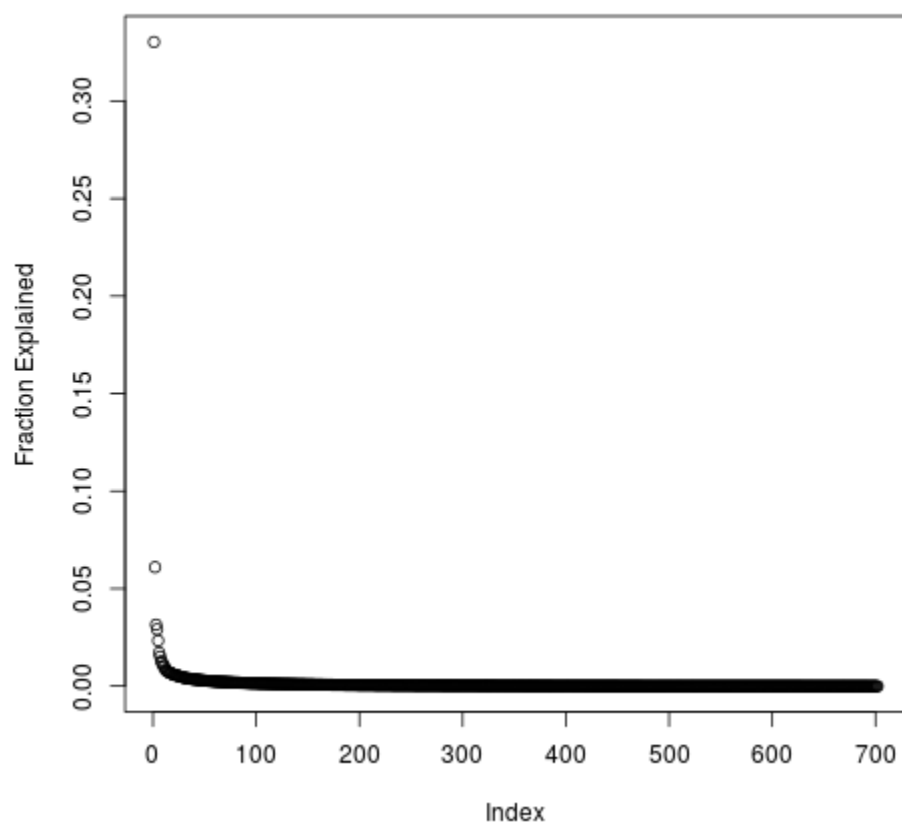
Figure 9. Q-Q plots showing the distribution of p-values, plotted against the expected distribution of p-values, for six-row spring-habit accessions.

Figure 10. Fraction of genetic variance for this population explained by each principal component, for six-row spring-habit accessions.
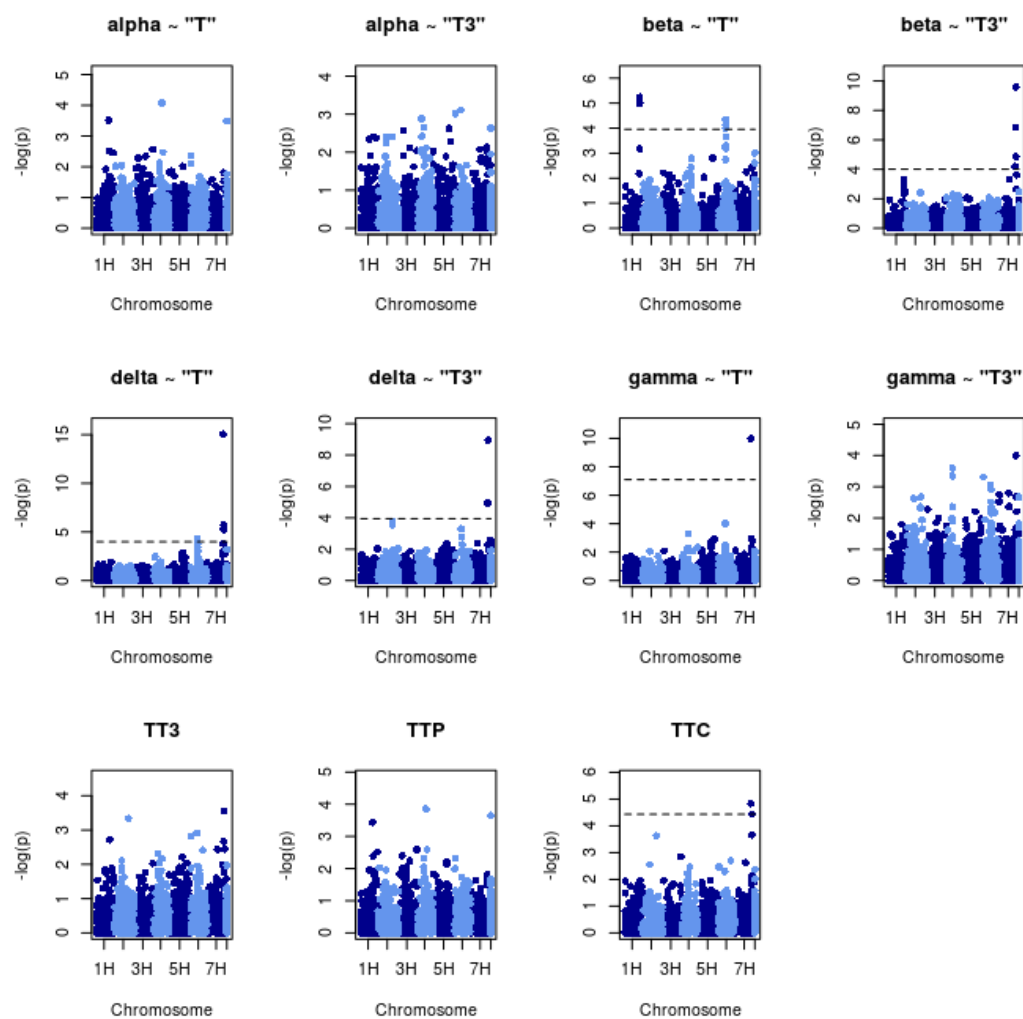
Figure 11. Manhattan plots showing results of GWAS for concentrations of all tocochromanol forms, fractions, and Total Tocochromanol (TTC), for spring-habit accessions grown in Montana in 2006. In analyses where one or more markers met the significance threshold determined by a false-discovery rate adjustment, a dotted line shows the significance threshold. Points in light blue, to the right of chromosome 7H, represent unmapped markers.
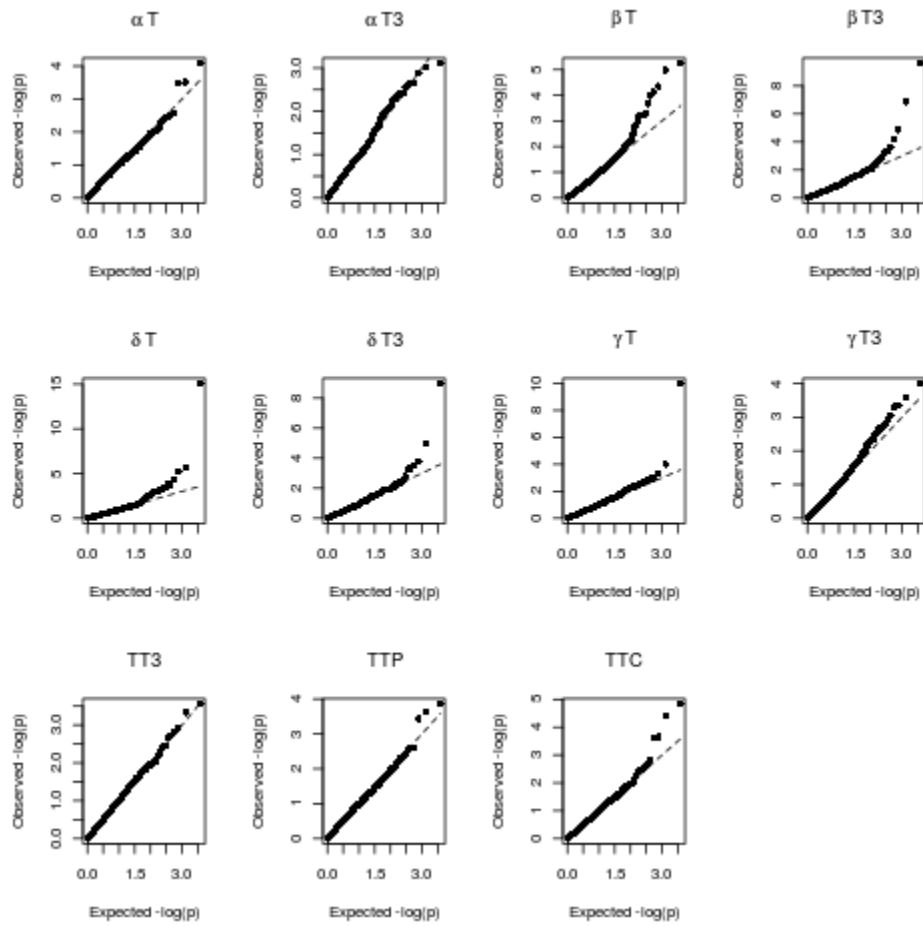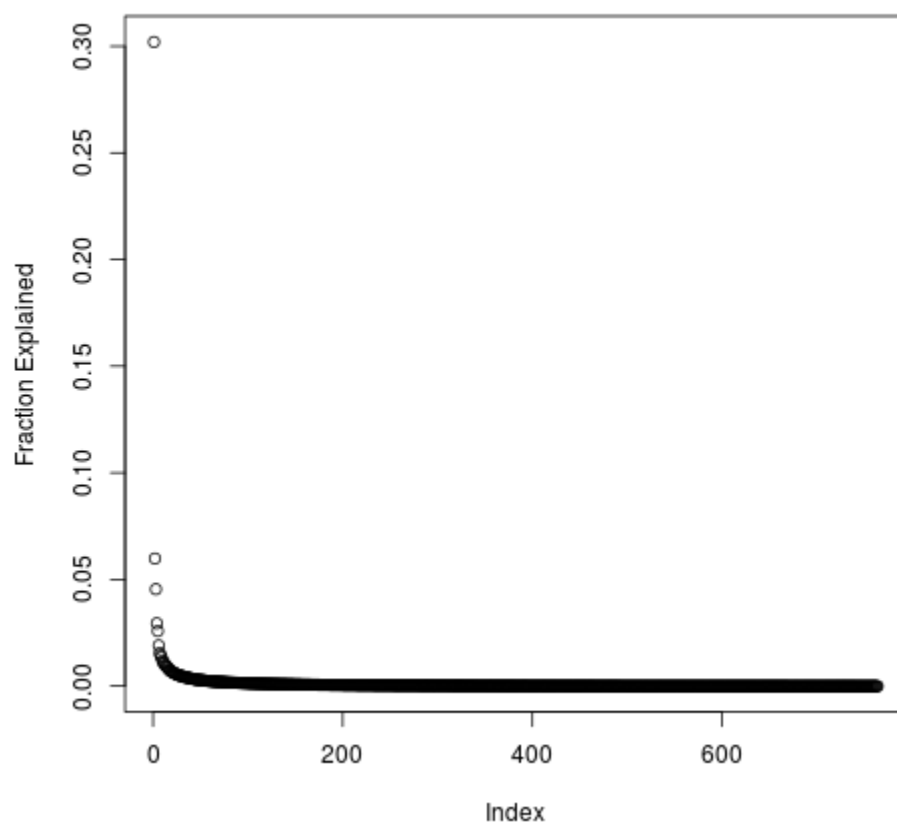
Figure 12. Q-Q plots showing the distribution of p-values, plotted against the expected distribution of p-values, for spring-habit accessions grown in Montana in 2006.

Figure 13. Fraction of genetic variance for this population explained by each principal component, for spring-habit accessions grown in Montana in 2006.

Figure 14. Manhattan plots showing results of GWAS for concentrations of all tocochromanol forms, fractions, and Total Tocochromanol (TTC), for spring-habit accessions grown in Montana in 2007. In analyses where one or more markers met the significance threshold determined by a false-discovery rate adjustment, a dotted line shows the significance threshold. Points in light blue, to the right of chromosome 7H, represent unmapped markers.

Figure 15. Q-Q plots showing the distribution of p-values, plotted against the expected distribution of p-values, for spring-habit accessions grown in Montana in 2007.

Figure 16. Fraction of genetic variance for this population explained by each principal component, for spring-habit accessions grown in Montana in 2007.

Appendix B- Linkage Disequilibrium

Linkage Disequilibrium
on Significant Regions
of Chromosome 1H

Linkage Disequilibrium
on Significant Regions
of Chromosome 6H

Supplementary Figure 1 (part one)

Linkage Disequilibrium on First Significant Region of Chromosome 7H
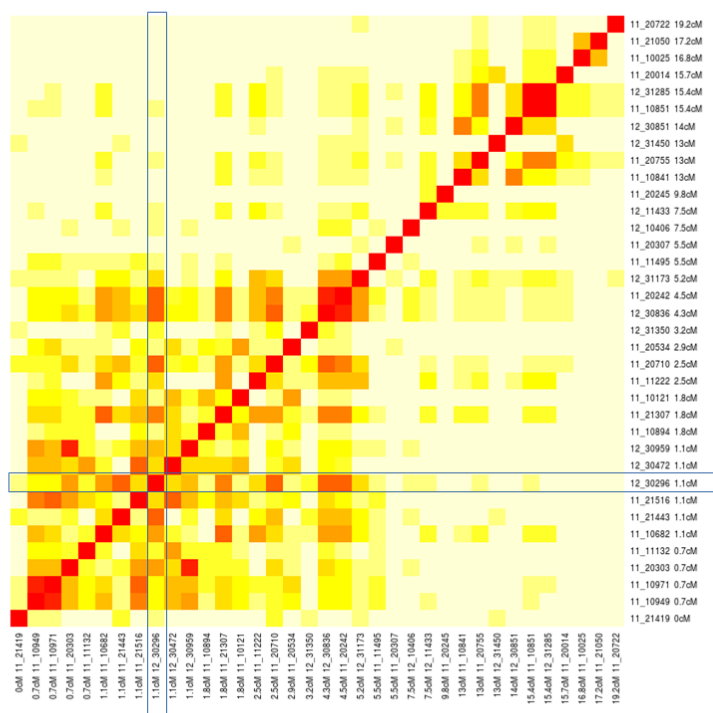
Linkage Disequilibrium on Second Significant Region of Chromosome 7H
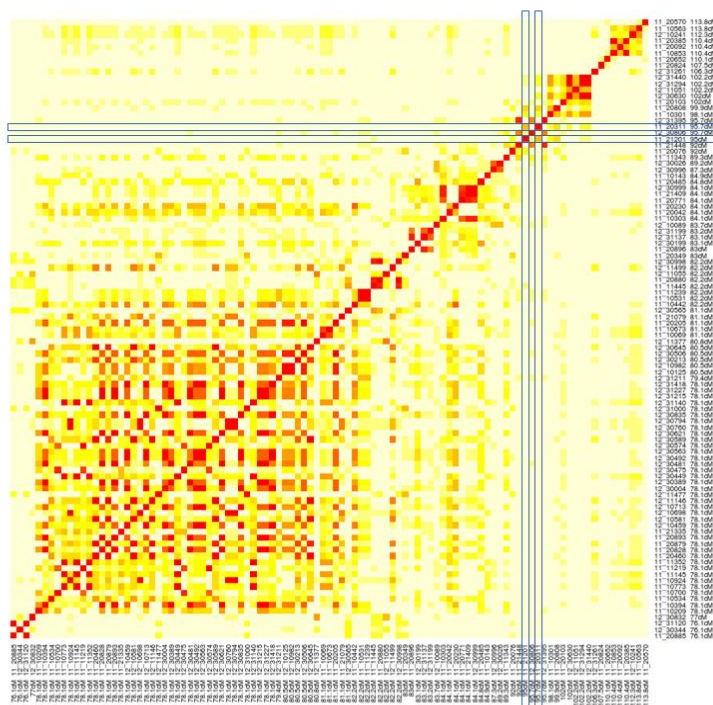
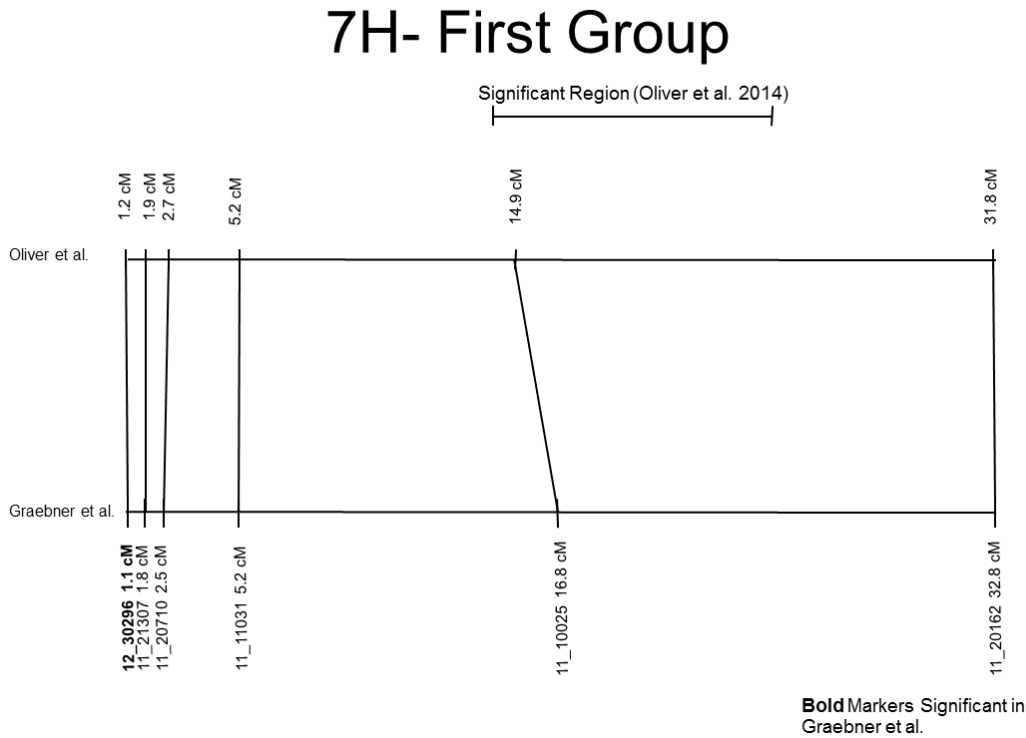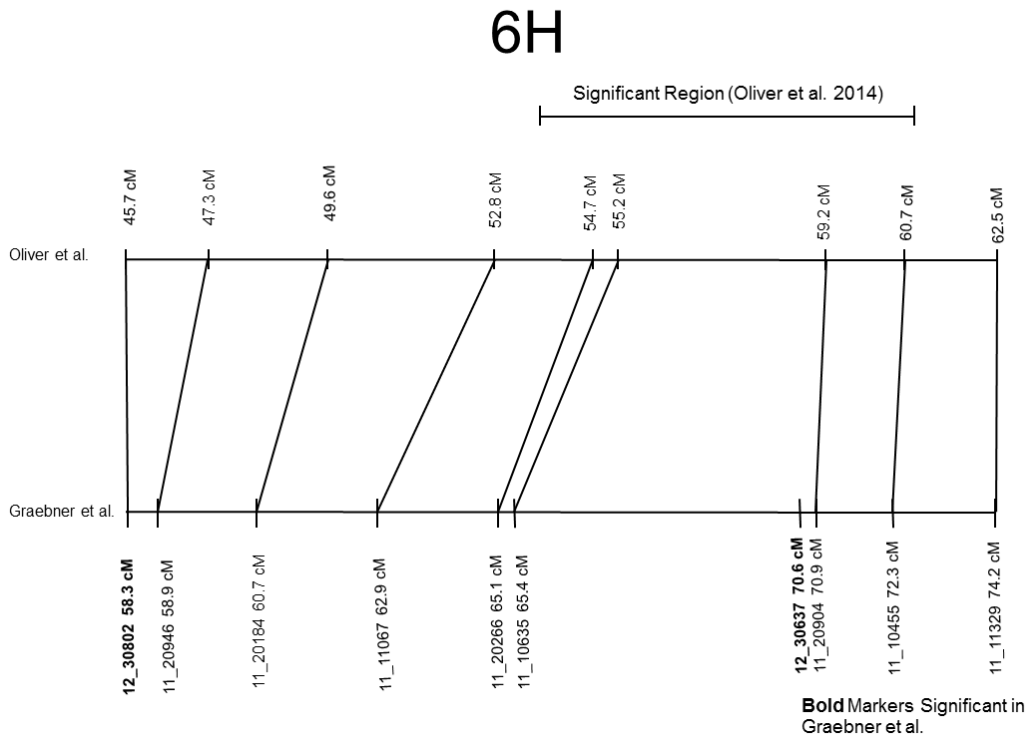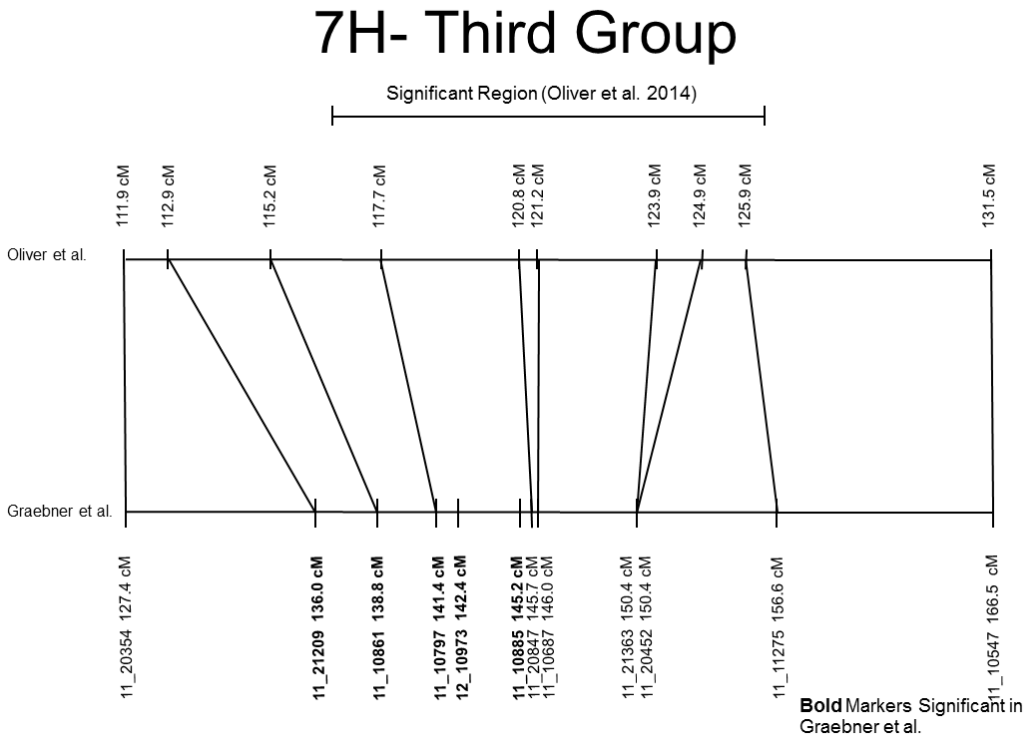Supplementary Figure 1 (part two)
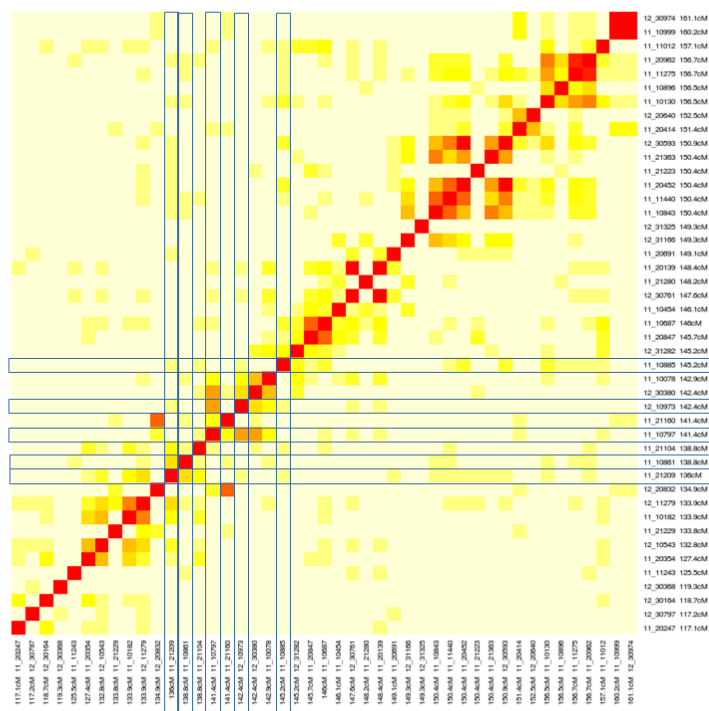
Supplementary Figure 1 (part three). Linkage disequilibrium in regions of the barley genome that demonstrate significant associations with tocochromanols.

Appendix B- Alignment of Linkage Maps

6H

7H- First Group

Supplementary Figure 2 (part one)

Supplementary Figure 2 (part two). Alignments of the linkage maps used by Oliver et al. (2014) and by this study.