AN ABSTRACT OF THE DISSERTATION OF

Thomas B. Parker for the degree of Doctor of Philosophy in Crop and Soil Science presented on June 20, 2007.

Title: Investigation of Hop Downy Mildew through Association Mapping and Observations of the Oospore.

Abstract approved:

John A. Henning

Hop downy mildew is a devastating disease affecting hop requiring expensive fungicide applications throughout the growing season. Plant resistance is highly desirable and theorized as being decidedly quantitative with dominance and epistasis involved in resistance. An association mapping approach using a mixed-model was used to identify AFLP markers associated with the hop downy mildew resistant phenotype. New protocols for the extraction, isolation and recovery of oospores from plant tissue and soil were developed to aid in the study of the hop downy mildew oospore. In addition, logic regression was used in a mathematical model in order to test the statistical procedure's potential for modeling epistasis. Our results suggest the hop downy mildew resistant phenotype has a broad-sense heritability of 76% with an estimated narrow-sense heritability of 49%. Mixed-model results revealed 9% of the AFLP markers to be associated with the hop downy mildew resistant phenotype. The association mapping results suggest resistance to hop downy mildew is quantitatively inherited with moderate heritability, which can be successfully investigated using mixed-models. The concentration of soil-borne oospores was 14 oospores/g soil. Germination of oospores occurred between two and eight weeks after preparation of the slides. Observations of the oospore suggested the sexual stage of this disease may play a role in over-wintering as oospores were capable of *in vitro* germination. In

addition, MTT as a stain for downy mildew spore viability comes into question due to the possibility of a chemical reduction in the presence of NADH. Logic regression correctly identified the model which best describes the epistatic interaction of *VRN-H1* and *VRN-H2* and the model was synonymous with the hypothesized genetic model for vernalization in barley. The mathematical simulation using logic regression software suggested Boolean logic may be more robust when compared to general linear modeling for the identification and modeling of epistasis.

©Copyright by Thomas B. Parker
June 20, 2007
All Rights Reserved

Investigation of Hop Downy Mildew through Association Mapping and Observations of the Oospore

by

Thomas B. Parker

A DISSERTATION

submitted to

Oregon State University

In partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 20, 2007
Commencement June 2008

Doctor of Philosophy dissertation of Thomas B. Parker
presented on June 20, 2007.

APPROVED:

Major Professor, representing Crop and Soil Science

Head of the Department of Crop and Soil Science

Dean of the Graduate School

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

Thomas B. Parker, Author

ACKNOWLEDGEMENTS

I would like first like to thank my graduate committee for all their help and encouragement throughout my growth and development as a scientist. In addition, I would like to thank three people for helping me through all the ups and downs on my journey to becoming a better scientist. First, I would like to thank Dr. Ross Koning from Eastern Connecticut State University in Willimantic, Connecticut. Ross was the first professor to take an interest in my development as a scientist. He was a true mentor for young people and a great teacher of science. Second, I would like to thank Dr. Gary Janssen from Miami University in Oxford, Ohio. Without question, I would not have made it through my doctoral program without him. Gary was always there to listen to me during all my ups and downs. He gave me sound advice and encouragement when I needed it the most. He is a true scholar and an excellent friend. Third, I would like to thank Dr. Walt Mahaffee. Ross introduced me to the world of science, Gary kept me in the game of science and Walt showed me how to do science. Every discovery I made resulted from Walt advising me on what was required to put myself in a position to make scientific discoveries. As Walt told me many times, “you have to put yourself in a position to make the discovery.” I did what I was told and, well, I made some interesting discoveries. Thanks Walt, I will continue using your advice and maybe someday I will be able to share your insights with other young scientists.

CONTRIBUTION OF AUTHORS

I want to thank Dr. M. Shaun Townsend who performed the AFLP extraction (Townsend and Henning 2005) which was used for my research in chapter 2. I want to thank Dr. John Henning for contributing his knowledge and experience in helping me better understand the variance components in chapter 2. I want to thank Dr. Walt Mahaffee and his lab for allowing me to use their microscope and image analysis software needed for my research in chapter 3. I want to thank Dr. Walter Mahaffee, Dr. Christopher Mundt and Dr. David Gent contributed to the general editing of chapter 3. I want to thank Dr. Patrick Hayes who gave me the barley vernalization sequence data to test my hypothesis in chapter 4. I want to thank Dr. Peter Szűcs for performing the genotyping and phenotyping of the barley vernalization data used in chapter 4. I would like to thank Dr. Walt Mahaffee for suggesting I add the coefficient of variation analysis to chapter 4. I would like to thank Dr. John A. Henning, Dr. Peter Szűcs, Dr. Patrick M. Hayes and Dr. Walter F. Mahaffee contributed to the general editing of chapter 4. I want to thank Dr. Peter Szűcs for helping me format my dissertation.

TABLE OF CONTENTS

	Page
1 GENERAL INTRODUCTION	1
Discovery of hop downy mildew in the United States………………………………....	1
Disease importance and breeding for resistance….....	1
Marker-assisted selection………..…………………	1
Linkage analysis vs. association analysis…………..	5
Association mapping and population structure……..	6
Genomic control……………………………………	6
Structured association………………………………	7
Mixed-models………………………………………	8
F-Statistic theory on kinship estimates with dominant markers……………………………..	8
Bayesian inference and F-statistics………..………..	9
Spurious association revisited………………………	12
Problems inherent in the statistical modeling of epistasis………………………………..	12
Alternative modeling approach……………………..	13
Tree-building algorithms……………………………	13
Logic regression, Boolean expressions and the search space………………………………….	14
The null model test………………………………...	15
Logic regression and identification of QTL interaction………..……………………………..	16
VRN-H1/VRN-H2 and epistasis in barley…….……...	16
Hop downy mildew epidemiology…………….……..	17
The hop downy mildew oospore……………………...	18
Controversy concerning over-wintering.…………….……	19
Dissertation goal……………………………………...	20

TABLE OF CONTENTS (Continued)

Page

2 Association Mapping Analysis in *Humulus lupulus* L. to Identify AFLP Markers Associated with the Hop Downy Mildew Resistant Phenotype………….….. 21

ABSTRACT……………………………………… 22

INTRODUCTION…………………………………… 23

MATERIALS AND METHODS…………………..... 26

Plant population…………………………….. 26

Plant material……………………………….. 26

Hop downy mildew resistant phenotype…… 26

AFLP analysis………………………………. 27

Data analysis………………………………… 27

RESULTS………………………………………… 29

DISCUSSION…………………………………….. 31

Variance components and heritability………. 31

Association mapping using mixed linear models…………..……………… 33

Conclusions………………………………….. 34

REFERENCES……………………………………. 36

3 The Extraction, Tetrazolium Staining and Germination of the Oospore of *Pseudoperonospora humuli* Miyabe and Tak. (Wil.)……………………………… 48

ABSTRACT………………………………….…. 49

INTRODUCTION…………………………….…. 50

MATERIALS AND METHODS……………….……. 53

Production of oospores from plant material….. 53

Extraction of oospores from plant tissue…..…. 53

TABLE OF CONTENTS (Continued)

Page

Extraction of oospores from soil……….…….. 54

MTT stain of oospores…………...……….….. 54

Data analysis…………………………………. 55

Germination of oospores……………………... 55

RESULTS……………………………………………. 56

Quantification of oospores…………………..... 56

Oospore germination……………………….…. 56

DISCUSSION……………………………………….. 57

Oospore extraction and quantification……..… 57

MTT staining of oospores…………………… 58

Oospore germination………………………... 59

Chestnut-brown oospore…………………….. 59

Conclusion…………………………………… 60

REFERENCES………………………………………. 63

4 Statistical Modeling of Epistasis and

Linkage Decay using Logic Regression………………….. 70

ABSTRACT………………………………………… 71

INTRODUCTION…………………………………... 72

MATERIAL AND METHODS……………………... 75

Plant material, phenotype and data set………. 75

Logic regression analysis……………………. 75

TASSEL analysis……………………………. 76

SAS-GLM…………………………………… 76

Linkage decay data set………………………. 77

RESULTS…………………………………………… 78

VRN-H1/VRN-H2 model selection…………... 78

TABLE OF CONTENTS (Continued)

	Page
TASSEL analysis………………………………	78
General linear modeling analysis of variance…………………………………..	79
Linkage decay……………………………….	79
DISCUSSION……………………………………….	81
ACKNOWLEDGMENTS…………………………..	85
REFERENCES……………………………………...	86
5 GENERAL CONCLUSION………………………………	97
BIBLIOGRAPHY…………………………………………..	102
APPENDICES………………………………………………	118
APPENDIX A………………………………………	119
HOP GENOTYPE PEDIGREE INFORMATION……………………………	119
APPENDIX B……………………………………….	123
Disease Note………………………………....	123
APPENDIX C……………………………………….	124
Chapter 4 Supplemental Material……….…...	124
APPENDIX D……………………………………....	131
General thoughts and comments………….....	131

LIST OF FIGURES

Figure Page

2.1 Frequency distribution for the
hop downy mildew resistant phenotype………..41

2.2 Breeding Population Jaccard Distance Ward Clustering…………………42

3.1 Oospores and zoospores in the process of germination…….…………….68

4.1 The relationship between model score and linkage decay……………….89

4.2 The coefficients of variation for 100 single-fit model scores
for the individual single-fit model data sets…………….……………….90

LIST OF TABLES

Table Page

2.1 List of hop genotypes arranged by pedigree based on cluster analysis of AFLP molecular marker data………………43

2.2 Analysis of variance results for broad-sense heritability of the transformed hop downy mildew phenotype……………………..…44

2.3 Analysis of variance and estimated variance components used in calculating narrow-sense heritability of the transformed hop downy mildew resistant phenotype…………………………………...……………..45

2.4A Mixed linear model results for primer pairs A-C listing the AFLP markers associated with the hop downy mildew resistant phenotype, *p*-values and marker R^2 values………….………46

2.4B Mixed linear model results for primer pairs D-F listing the AFLP markers associated with the hop downy mildew resistant phenotype, *p*-values and marker R^2 values…….……………….47

3.1 Kruskal-Wallis test for soil-borne oospore data…………………………67

4.1 Null model tests for the 'Dicktoo' x 'Calicuchima'-sib[A] and 'Dicktoo' x 'OWB-D'[B] data used to check for signal in the data…………………..……...………………….…….……..………91

4.2 Cross-validation tests for the 'Dicktoo' x 'Calicuchima'-sib[A] and 'Dicktoo' x OWB-D'[B] data used to identify the size of the logic tree model with the best predictive capability……………………..…92

4.3 Permutation tests for the 'Dicktoo' x 'Calicuchima'-sib[A] and 'Dicktoo' x 'OWB-D'[B] data…………………..……..……..……....93

LIST OF TABLES (Continued)

Table Page

4.4 TASSEL results identifying markers associated with the days to flowering phenotype…………………………………..……….94

4.5 Analysis of variance results for the 'Dicktoo' x "Calicuchima'-sib data full model with markers which were found to be associated with the days to flowering phenotype in TASSEL……………………......95

4.6 Analysis of variance results for the 'Dicktoo' x 'OWB-D' data full model with markers which were found to be associated with the days to flowering phenotype in TASSEL……………………….96

DEDICATION

I would like to thank my Uncle Mike for listening to me. I don't have any family except my Uncle Mike. Although he told me many times that I should have been born rich to do science and probably wished I picked a more lucrative career, he never gave up hope that I would finish my studies and make a contribution to science and the world. Thanks Uncle Mike for being there and listening to me.

GENERAL INTRODUCTION

Discovery of hop downy mildew in the United States

Downy mildew [(*Pseudoperonospora humuli* Miyabe and Tak. (Wil.))] of hop was first observed in North America in Wisconsin (Davis 1910). The first authentication of downy mildew of hop in Oregon was April 10, 1930 when infected 'spikes' were found at Mount Angel in Marion County (Hoerner 1939). Hop downy mildew is now problematic in all areas where hops are grown in the United States.

Disease importance and breeding for resistance

Hop downy mildew is a disease of major importance in the hop industry. Plants infected with the disease have inferior cones with reduced yields and, depending on the level of resistance, can die within a few years. Prior attempts at breeding for increased levels of resistance have not revealed a discernable pattern of inheritance (Kenny 1991; Klein 1995). Partial resistance to downy mildew has been identified in cultivars like 'Cascade' (Neve 1991) and 'Newport' (Henning and Haunold. 2003), while appreciable levels of resistance have existed for decades in cultivars coming out of the Hop Breeding Program at Wye College in Britain. Cultivars such as 'Yeoman' (Neve 1991; Neve and Darby 1983) are highly prized for their strong resistance to this disease. Although breeding for resistance has shown some success, resistant varieties are not grown in large acreage due to their limited demand by the brewing industry. Therefore, it is a continuing goal of hop breeding programs to use these resistant cultivars as breeding stock in developing new cultivars. One breeding technique that has become increasingly popular is marker-assisted selection.

Marker-assisted selection

Sax (1923) first described the theory of quantitative trait loci mapping (QTL mapping) when he observed that the complex trait of bean size was associated with the

simple trait of seed coat color. Thoday (1961) suggested that that it might be possible to map and characterize all QTLs involved in complex traits if the segregation of simply inherited monogenes could be used to detect linked QTLs. Modern QTL mapping uses that premise with the innovation that defined sequences of DNA act as the linked monogenic markers (Young 1996). In a QTL analysis, scientists attempt to identify associations between quantitative traits and marker alleles within a segregating population (Lander and Bostein 1989; Weller et al. 1990) to identify the genomic locations of loci contributing to complex traits, the contribution of each and the interaction between loci.

Many economically important characters in plants are controlled by quantitative trait loci (QTL). Quantitative traits show continuous distributions and are assumed to be controlled by many minor genes that exert varying effects. However, since the expression of these genes is strongly influenced by the environment and many quantitative traits have a low heritability, high individual performance is no guarantee that the selection has desirable genes at the relevant QTL. Therefore, marker-assisted selection is useful in improving the efficiency of selection early in the breeding cycle helping to improve characters with low heritability. By studying the segregation of each marker locus, researchers can estimate the effect of the linked polygene and learn more about genetic control of the trait. Factors that affect the ability to detect a QTL include the heritability of the trait, the number of individuals in the experimental population, the accepted significance level for type I errors (false-positives), experimental design, method of statistical analysis, genome size, phenotypic effects of the QTL and genetic distance between the QTL and the marker loci (Beavis 1998; Darvasi et al. 1993; Lander and Botstein 1989; Weller 1992).

Linkage mapping involves the ordered assembly of genetic markers within linkage groups based on the recombination (Lander et al. 1987; Ritter et al. 1990). If two markers are close together, recombination between the markers will be rare. When the markers are very close together, they never recombine and are assumed to

be linked. For QTL analysis, it is not necessary to produce an overly dense linkage map. Modeling and empirical evidence have shown that marker densities of less than 20cM offer little improvement in the precision of QTL identification and location (Darvasi et al. 1993; Kearsey and Farquhar 1998).

In the simplest form, QTL identification is associations between individual molecular markers and a phenotype by linear regression analysis (Kearsey and Hyne 1994; Hyne and Kearsey 1995). When a particular marker is associated with a statistically significant phenotypic mean, it is reasonable to conclude that there is a QTL for that trait associated with that marker (Lander and Botstein 1989). For a simple dominant marker, comparisons are made between the mean phenotypic scores for all individuals with the marker and the mean scores of those individuals without the marker. Darvasi et al. (1993) reported that in cases where marker density is low, associations between markers and QTL can be identified using simple linear regression or least squares. However, these tests are limited to identifying QTL with reasonably tight linkage to the marker (van Ooijen 1992). In addition, simple association cannot identify the position of the QTL relative to the marker so it is impossible to differentiate a QTL with a large effect that is poorly linked to a marker from a QTL with a small effect that is tightly linked to a marker (Lander and Botstein 1989; Weller 1992).

Interval mapping overcomes some of the shortcomings of simple linear regression and provides greater statistical power. Interval mapping requires a linkage map with markers spaced at 1 to 20cM intervals (Darvasi et al. 1993; Darvasi and Soller 1994; Darvasi and Soller 1997). Interval mapping examines the effect of a QTL located in the interval between two markers for every interval in the genome (Lander and Botstein 1989; Luo and Kearsey 1992). The advantages of interval mapping are that it indicates the presence of a QTL, the effect of that QTL on the trait and its approximate location (confidence interval) (Knott and Haley 1992; Lander and Botstein 1989). The presence of the QTL is reported as an odds ratio and is based on the likelihood of the presence of the genotypic data with no linked QTL. If the odds

ratio is larger than an accepted threshold, it is assumed that there is a QTL in that interval. The threshold level is normally determined empirically (Lander and Kruglyak 1995) and is dependent upon the acceptable rate of type I errors, the number of individuals in the analysis and the number of markers in the linkage map (Lander and Botstein 1989; van Ooijen 1992). An estimation of the effect of the QTL on the trait of interest can be estimated as a proportion of phenotypic variance or as a proportion of the genotypic variance when the heritability is known (Lande and Thompson 1990; Lander and Botstein 1989; Melchinger et al. 1998; Weller 1992). Composite interval mapping (CIM) is an extension of interval mapping and is used to detect more than one QTL on a chromosome where the effect for a likely QTL is fixed and the data is used to map the likelihood of another QTL (Chmielewicz and Manly 2002; Haley and Knott 1992; Knapp 1991; Lander and Botstein 1989). Linkage mapping is an iterative process that uses all the available information and will yield sensible estimates that can be applied to both linear and non-linear models. However, the method is parametric so the type of data distribution needs to be known (Staub et al. 1992; Weller 1992).

Although QTL analysis may appear to be a panacea, there are some drawbacks. It can be very time-consuming, expensive and the information gained may be limited. In a typical QTL study, only one cross is made to form a recombinant population and therefore, extrapolations to other populations and individuals may be spurious (Verhoeven et al. 2006). In addition, although the degree of dominance can be estimated in the F_2 progenies (for clonal crops use F_1 seedlings), replication of the F_2 means that the cross needs to be cloned and until quite recently (Ruczinski et al. 2003) epistatic interactions could not be realistically studied. Finally, Hyne and Kearsey (1995) showed that it is difficult to locate more than 12 QTLs in a population at one time and bias may creep into the analysis. Therefore, only significant effects are reported resulting in an underestimation of the true number of QTLs and exaggerating their additive and dominance effects (Kearsey and Farquhar 1998).

Linkage analysis vs. association analysis

During the last few decades linkage analysis has been a popular experimental design to study the genetic basis of inherited diseases in humans (Carlson et al. 2004). Linkage analyses search for regions in the genome that have a higher than expected number of shared alleles within affected individuals within a family (Carlson et al. 2004). Because closely related individuals tend to share large regions of the genome, fewer polymorphic markers need to be genotyped to detect linked regions. However, because there are few recombination events within most families, it is often difficult to narrow the region of interest below several megabases (Carlson et al. 2001).

Alternatively, association mapping (AM) attempts to find a statistical association between the genetic markers and the quantitative trait. Unlike traditional QTL mapping, AM is performed at the population level where a collection of genotypic molecular markers and the corresponding phenotypic trait are determined in a set of unrelated or distantly-related individuals. AM relies on linkage disequilibrium between the molecular markers and the causative polymorphism in the linked gene. Linkage disequilibrium focuses on population-based differences in alleles, whereas linkage focuses on differences at a single locus. Therefore, AM looks for ancestral recombination events viewed as a measure of co-segregation within a population. It is these population-based recombination events and their statistical associations with a phenotype which are the basis for AM analysis.

Linkage disequilibrium (LD) usually extends only short distances in out-crossing species (less than 1500bp in maize) (Gaut and Long 2003), whereas in inbreeding species like *Arabidopsis*, LD can range from 1 to 50cM or more (Nordborg et al. 2002). Therefore, success in association mapping requires the candidate gene to have a measurable effect on the phenotype and candidate markers must fall within or directly upstream/downstream of the candidate gene. The markers which are identified as associated with the trait are broad-based rather than cross-specific. Presently, linkage analysis is more powerful than association analysis for identifying rare high-risk alleles, but association analysis is expected to be more powerful for

detection of common disease alleles that confer modest disease risk (Carlson et al. 2001).

One advantage of association mapping is that it is easier to recruit large numbers of unrelated affected individuals than to collect large amounts of pedigree information. Unfortunately, because the region around the marker shared identical by decent in unrelated, affected individuals will be much smaller than the shared region for related individuals, marker density needs to be much higher than for traditional linkage analysis (Carlson et al. 2001).

Association mapping and population structure

Association analysis has been investigated as a possible tool for identification of population-wide polymorphisms associated with specific phenotypic traits (Cardon and Bell 2001). However, population structure/stratification may lead to spurious associations (associations without linkage) between a candidate marker and phenotype (Lander and Schork 1994; Bacanu et al. 2000; Pritchard et al. 2000a and 2000b; Devlin et al. 2001; Yu et al. 2006). Transmission/disequilibrium test (TDT) (Spielman et al. 1993) using family-based tests of association has been implemented to deal with problems associated with population structure. Although this method has been used successfully to identify QTL by association, the analysis comes at a considerable cost as the DNA needs to be collected from close relatives of affected individuals.

Genomic control

Other statistical controls have been implemented to help adjust for sources of experimental error. One such method, genomic control (GC) (Bacanu et al. 2000; Devlin et al. 2001) touts the robustness of family-based designs on population data. Devlin et al. (2001) reported GC exploits population substructure generated 'overdispersion' of statistics to assess association. By testing multiple polymorphisms throughout the genome, only some of which are pertinent to the disease of interest, the degree of overdispersion generated by population substructure can be estimated and

taken into account (Devlin et al. 2001). With GC, random markers are used to estimate and adjust for inflation of the test statistics generated by population structure (Yu et al. 2006).

Structured association

In response to GC, Pritchard et al. (2000a) proposed an alternative method called structured association (SA). Structure association uses prior information and Bayesian clustering to form a posterior probabilistic distribution. In SA, X denotes genotypes of sampled individuals and Z denotes the (unknown) populations of origin of the individuals and P denotes the (unknown) allele frequencies in all populations (Pritchard et al. 2000a). Using Hardy-Weinberg equilibrium within populations along with complete linkage equilibrium among loci within populations as model assumptions, each allele at each locus is independent within a frequency distribution giving us the probabilistic distribution Pr(X/Z, P) (Pritchard et al. 2000a). Since Z and P are inferred quantities, Bayesian statistics is used to specify model posterior probabilities. These inferred quantities are in essence random probabilities since the 'unknown' populations are assumed to be in Hardy-Weinberg equilibrium with complete linkage equilibrium. Structured association uses marker loci unlinked to the candidate genes under study to estimate population structure (Yu et al. 2006). It then incorporates this information into further statistical analysis such as modeling through association analysis. The use of random unlinked markers to estimate population structure using Markov chain Monte Carlo (MCMC) methods is a non-parametric randomization test resulting in a frequency distribution used to estimate the percentage of random error (Pritchard et al. 2000a). The resulting *p*-value is the proportion of randomized unlinked markers associated with the phenotype by chance. If the proportion is low, the chance of incorporating random markers in the model is low. A large *p*-value suggests a large amount of random error being incorporated in the data set (possibly resulting from population structure) which may lead to a high proportion of type I errors (false-positives). However, populations showing high levels of linkage

disequilibrium may cause invalid results as the model assumptions are violated, because structured association assumes the population is in linkage equilibrium.

Mixed-models

Recently, Yu et al. (2006), developed a unified mixed-model method of data analysis which can simultaneously account for multiple levels of gross population structure (Q) and finer scale relative kinship (K). This new software provides a powerful complement to the current methods for association mapping. Yu et al. (2006) suggest this software is superior due to a novel method that controls both type I and type II error rates. The use of pedigree structure in genetic models has shown some promise in identifying molecular markers associated with phenotypes in both humans (Morley et al. 2004) and plants (Thornsberry et al. 2001; Yu et al. 2006). In addition, using kinship estimates in a mixed-model format, reasonable estimates of narrow-sense heritability (h^2) can be obtained. However, the use of a mixed-model within the software TASSEL© requires knowledge of population structure and kinship estimates. Therefore, researchers must obtain this information prior to running the analysis.

F-Statistic theory on kinship estimates with dominant markers

Although many pair-wise marker-based estimates of kinship have been developed for co-dominant markers (Queller and Goodnight 1989; Loiselle et al. 1995; Ritland 1996; Lynch and Ritland 1999), a similar estimator for use with dominant markers has only recently been developed (Hardy 2003). Kinship coefficients are also called co-ancestry coefficients and are based on the probability of identity of alleles for two homologous loci sampled in a particular way (Hardy and Vekemans 2006). In the case of a kinship coefficient between two individuals, two loci are randomly sampled between each of the two individuals. (Hardy and Vekemans 2006). The kinship coefficient between two individuals (F_{ij}) has often been defined as the identity by descent (IBD or Θ) of the loci being compared (Hardy 2003). However, estimators based on genetic markers actually estimate the 'relative kinship' (Rousset 2002; Hardy

2003). Therefore, it is best to define kinship in individuals from a population under isolation-by-distance, as a ratio of differences of probabilities of identity-in-state (IIS) between homologous loci (Rousset 2002). These definitions result in the following model: $F_{ij} = Q_{ij} - \overline{Q} / 1 - \overline{Q}$ (where F_{ij} are the pair-wise kinship coefficients, Q_{ij} is the probability of IIS of random loci *i* and *j* and $\overline{Q}$ is the probability of IIS between random loci from the reference population or reference sample) (Hardy 2003). When relatedness is assessed from genetic markers, the reference population is usually a sample of individuals from the marker population ($\overline{Q}$). Alternatively, IBD can be defined in terms of IIS by assuming mutation rates are negligible or by redefining the equation as a ratio of differences of probabilities of IBD, replacing Q by Θ which results in an approximation of IBD: $F_{ij} = \Theta_{ij} - \overline{\Theta} / (1 - \overline{\Theta})$.

Hardy (2003) reported the stepwise derivation of kinship estimates for dominant markers. The derivation was a 15 equation process that resulted in the estimation of Fd_{ij} (the kinship coefficient for dominant markers). Kinship can now be estimated for dominant markers by identifying all the possible differences in alleles and mining for differences that are identified within the marker data set. In order to obtain reliable estimates of kinship from dominant markers, researchers must have access to an ‘assumed’ inbreeding coefficient since departures from Hardy-Weinberg equilibrium must be known. Hardy (2003) suggested using Bayesian inference where the inbreeding coefficient would be given uniform prior probabilities between 0 and 1. Holsinger et al. (2002) used Bayesian prior probability statistics to estimate the inbreeding coefficient (F_{IS}) in AFLP data and it is now possible to obtain reliable estimates of kinship from dominant molecular markers.

Bayesian inference and F-statistics

Bayesian inference (a statistical inference where prior observations are used to update a newly inferred probability in which a hypothesis may be true) has been developed for use in population genetics (Beaumont et al. 2002; Holsinger et al 2002;

Pritchard et al. 2000a). Since biological diversity is inherently hierarchical (eg. closely related species are part of a genus, closely related genera groups to a family etc), re-sampling and updating the likelihood based on a prior probability for an expected allele frequency may increase the precision of the likelihood for a given allele frequency. The fixation index F_{st} is the correlation between random gametes within subpopulations (s) relative gametes in the total sample (t), the 'relative' nature of the statistic makes the value a probability (Wright 1969) for which Bayesian inference is well-suited using updated likelihood estimates to converge at the most likely probability for a given allele frequency.

Holsinger (1999) reported that allele frequencies in samples may be different than actual frequencies in those populations. In addition, a second source of error comes from the possibility that the population in the study represent only a portion of the total population. Attempts to correct this sampling error have been implemented (Nei and Chesser 1983; Weir and Cockerman 1984). Using fixed effects models, Nei and Chesser (1983) implemented a bias correction resulting in G_{st}, while Weir and Cockerman (1984) created a set of indicator variables (1, 0: for 1= allele A_i and 0 ≠ allele A_i). By running an analysis of variance on the indicator variables partitioned into within population and among populations, the model became a function of the co-ancestry coefficients (θ) which are equivalent to Wright's statistics (Holsinger 1999). However, G_{st} and θ show little difference when the population sample is very large.

It would be useful to have a statistical method to partition the within and among population components that does not depend on any evolutionary model (Holsinger 1999). Most likelihood estimates (Barton et al. 1983; Wehrhahn and Powell 1987; Slatkin and Barton 1989; Wehrhahn 1989; Rannala and Hartigan 1995) can be used without making any inference as to an evolutionary model. However, these models are implicitly random-effects models of population sampling. Holsinger (1999) extends the existing Bayesian approach to F-statistics to allow for likelihood estimates for multiple levels within the population sampling hierarchy and allowing for both fixed and random effects. Many statisticians have used the Bayesian

inference approach for the analysis of hierarchical data and there is a comprehensive review on Bayesian methods for analysis of complex data consisting of multiple levels of order (Goldstein 1995).

In the Bayesian approach, when given two alleles A_1 and A_2 at locus *I*, with individuals (*i*) from populations (*k*), where a_{ik} is the number of A_1 alleles and n_{ik} is the sample size, the following model can be formed:

$L(a_{ik}|n_{ik}, p_{ik}) = \prod_{i=1}^{I}\prod_{k=1}^{K}\binom{nik}{aik}p_{ik}^{aik}(1-pik)^{nik-aik}$ (Holsinger 1999). The maximum likelihood estimates of for p_{ik} are those values which maximize $L(a_{ik}|n_{ik}, p_{ik})$.

Because Bayesian estimates for p_{ik} also depend on the prior probability of p_{ik} [($\Phi(p_{ik})$)], this must be taken into account as a multiple of the likelihood equation:

$P(p_{ik}|n_{ik}, p_{ik}) \infty \prod_{i=1}^{I}\prod_{k=1}^{K}\binom{nik}{aik}p_{ik}^{aik}(1-p_{ik})^{nik-aik}\phi(p_{ik})$. In the absence of any prior information about allele frequencies, a uniform distribution based on (0,1) should be chosen assuming that any allele frequency is equally likely. If the mean of the posterior distribution is used as the point estimate for p_{ik}, then $\hat{p}_{ik} = (a_{ik}+1)/n_{ik}+2)$.

Going one step further, a hierarchical component can be added to the equation such as partitioning between and among populations. By defining the allele frequency distribution of p_{ik} in all populations (including those not sampled) as a beta distribution with parameters a_i and b_i, $Be(a_i, b_i)$, defines $a_i = (1-\theta/\theta)x_i$ and $b_i = (1-\theta/\theta)(1-x_i)$, where x_i corresponds to the mean allele frequency at locus *i* (averaged across all populations) and θ corresponds to F_{st} (calculated from the frequency distribution of alleles across all populations) (Holsinger 1999). By defining $X(x_i)$ as the prior distribution for x_i and $T(\theta)$ as the prior distribution of θ, the posterior probability distribution for p_{ik} and θ is given by the following equation:

$P(p_{ik}, \theta|n_{ik}, a_{ik}) \infty \prod_{i=1}^{I}\prod_{k=1}^{K}\binom{nik}{pik}p_{ik}^{aik}(1-p_{ik})^{nik-aik} * Be(a_i, b_i)X(x_i)T(\theta)$. The Bayesian methods outlined above have been recently extended to incorporate dominant markers

such as AFLPs so that estimates of the hierarchical inbreeding coefficient (F_{is}) can be determined for this class of molecular markers (Holsinger et al. 2002).

Spurious association revisited

Cardon and Bell (2001) suggested the effects of population structure may be overstated. For example, there are few examples to support the assumption that differences in allele frequencies among populations lead to spurious association, suggesting the problem is overemphasized and that other factors may be playing more important roles in the spurious associations. It is believed that over-interpreting marginal findings and the publication bias has been underemphasized and attempts to reduce effects of population stratification will not lower this type of error (Cardon and Bell 2001), while another source of error involves the small effects of many genetic factors that contribute to disease such as environmental risks in which the relative affect of many, small impact genetic factors contribute to the disease. Also, modest sample size tends to overestimate the size of the genetic effect (Cardon and Bell 2001). In genome-wide association analysis studies, sampling concerns shift from statistical power to the inflation of false-positives caused by the testing of very large numbers of markers (Cardon and Bell 2001). Fine mapping of suggestive linkage peaks by association makes sense only when simulations show clear suggestive evidence for an excess of suggestive linkage across the genome or when the 'cost' of doing so is acceptable without compelling evidence of linkage (Carlson et al. 2004). Therefore, although most researchers agree that population structure and relatedness leads to spurious association, there are other factors which may play a greater role in spurious associations.

Problems inherent in the statistical modeling of epistasis

Box (1979) stated, "All models are wrong but some are useful." Are the problems inherent in the model design or are the problems inherently due to the fact that our current model paradigm cannot correctly model living systems? For example,

interval mapping and CIM have been used successfully to identify QTL associated with specific phenotypes which has eventually led to the identification of statistically significant interactions among QTL (epistasis). However, the use of linear modeling to test for statistically significant interaction between QTL makes the assumption that all genetic marker interactions are the result of an interaction of additive genetic effects. This assumption may or may not be true depending on the system. Is there another way to statistically model by association which can identify potential epistasis and/or additive genetic effects and identify the correct model without the implicit assumptions of linear modeling?

Alternative modeling approach

A recently discovered statistical technique entitled logic regression is a tree-based model system which creates Boolean (binary) expressions in the form of logic models. These tree-based logic models consist of Boolean combinations of binary covariates (Ruczinski et al. 2003). Logic regression may allow scientists to identify dominant and dominant/suppression forms of epistasis in dominantly scored genetic data sets.

Tree-building algorithms

There are many search algorithms (i.e. CART, MARS, Logic Regression) used in predicting continuous variables or categorical variables from a set of continuous predictors and/or categorical factor effects. Two of these algorithms (CART and Logic Regression) use tree-based logic to determine a set of ‘if-then’, ‘yes-no’, ‘on-off’ logical conditions that permit accurate prediction or classification within a system (Ruczinski et al. 2003).

Tree classification can produce accurate predictions or predicted classifications based on few logical operators and have a number of advantages over many other modeling techniques. One advantage is that the interpretation of results summarized in a tree is very simple. Search algorithms like CART and logic regression can often

yield much simpler models for explaining the relationship between the predictors and the response. The results of using tree-based methods for classification or regression can be summarized in models with very few nodes.

In using tree-based algorithms, there are no implicit assumptions for the underlying relationships between the predictor variables and the dependent variable. The model may be linear, non-linear, bimodal etc. Therefore, tree-based modeling methods are particularly well-suited for data mining tasks, where there is often little *a priori* knowledge regarding which variables are related and how (Ruczinski et al. 2003). Given this supposition, population and kinship estimates are not required for the analysis.

Preliminary evidence suggests that logic regression may be a better alternative for identifying molecular marker associations with a phenotype (Ruczinski et al. 2003). A comparison of logic regression with mixed-models, multivariate adaptive regression splines (MARS) and classification and regression trees (CART) indicated logic regression was more robust in identifying the correct genetic model which best described the association of molecular marker data with the disease phenotype without the use of pedigree structure (Ruczinski et al. 2003). However, it remains to be seen whether the inclusion of pedigree information into logic models will make them more robust.

Logic regression, Boolean expressions and the search space

Logic regression can be used in the analysis of dominant-scored genetic data sets. The software utilizes a tree-based simulated annealing algorithm which searches binary data sets and constructs mathematical models consisting of Boolean combinations of binary covariates. With $X_1 \ldots X_k$ as binary predictors and Y as the response, logic regression will fit regression models in the form $g(E[Y]) = \beta_0 + \sum_{j=1}^{t} \beta_j L_j$, where L_j is a Boolean expression of the predictors X_i (Ruczinski et al. 2003). Logic regression searches for combinations of binary variables that have high

predictive power for the response variable. The software is unique in that it utilizes Boolean algebra as a basis for building Boolean expressions called logic models.

Boolean algebra uses values of 0 and 1 to represent a specific state (true and false, on and off, yes and no,…). Variables (X_1, X_2, X_3,….) are represented as either of these two values. Logic operators: $\wedge$ (AND), $\vee$ (OR) and ^{c} (NOT or the conjugate of X) are used to combine values and variables to build logic (Boolean) expressions such as the form: $Y = (X_1 \wedge X_2^c)$. By following a set of specified rules for logic trees and tree-growing, the logic regression simulated annealing algorithm moves in the search space (S) by defining the neighbor of the logic tree to be those trees that can be reached by a single move. Logic regression compares the new state model score with the previous state model score. If the score of the new state is better than the score of the old state, the move is accepted. If the score of the new state is worse than the score of the old state, the move is accepted conditioned on a specific probability. The acceptance probability depends on the score of the two states under consideration and a parameter that reflects a time point in the annealing chain (this parameter is referred to as temperature) (Ruczinski et al. 2003). For any pair of scores, the further ahead in the annealing scheme, the lower the acceptance probability, if the proposed state has a score worse than the score of the old state. As a result, when the search algorithm reaches the end of state space, the result is generally a good scoring state (Ruczinski et al. 2003).

The null model test

The null test in logic regression is used to provide information concerning strength of signal to the researcher. In evaluating models of varying sizes, logic regression looks for signal vs. noise in the data set. In statistical modeling, signal is identified by asking whether the slope (b) is equal to zero or not. Therefore, is X associated with Y? This is considered 'signal'. When additional covariates are added to the model and have no relationship with Y, this is considered 'noise'. By evaluating models of various sizes for signal vs. noise, researchers can determine the

level of ‘over-fitting’ (noise) in each ‘model-size’ class. In doing so, search parameters can be set to identify the correct model in a search space containing the least amount of noise. In addition, potentially troublesome data sets with ‘unacceptable’ levels of noise are quickly identified so that no further time is wasted in the analysis of potentially problematic data.

Logic regression and identification of QTL interaction

The authors of logic regression have created a search algorithm that appears to have the ability to identify dominant and dominant/recessive genetic interactions in binary genetic data and present the interaction(s) in a regression format. The use of logic operators may allow scientists to predict the precise genetic model governing the interaction because the logic operators account for the dominant and recessive condition in binary data sets. The inclusion of the null model test allows researchers to identify potentially troublesome data sets where the levels of statistical signal are very low. Therefore, given all these attributes, logic regression may be useful in identifying and modeling dominant, dominant/suppression as well as additive forms of genetic epistasis associated with specific phenotypic traits.

VRN-H1/VRN-H2 and epistasis in barley

In order to test logic regression to see if it can correctly identify dominant and/or dominant suppression epistasis, a model system with a known statistical interaction is needed. The epistatic interaction of alleles at the *VRN-H1*, *VRN-H2* and *VRN-H3* locus is the hypothesized determinant for vernalization-sensitivity in cultivated barley (*Hordeum vulgare* subsp. *vulgare*) (Takahashi and Yasuda 1971). There is no allelic variation at *VRN-H3* in most cultivated barley genotypes, reducing the genetic model to a two-locus epistatic model (Takahashi and Yasuda 1971). *VRN-H2* encodes a dominant flowering repressor (ZCCT-H) down-regulated by vernalization (Yan et al. 2004). *VRN-H1* is a MADS-box floral meristem identity gene (*HvBM5A*) (Danyluk et al. 2003; Yan et al. 2003) and large deletions within the

first intron result in a dominant *VRN-H1* allele and spring growth habit (Fu et al. 2005; von Zitzewitz et al. 2005). A molecular model has been recently proposed to explain the *VRN-H2/VRN-H1* epistatic interaction where dominant *VRN-H2* inhibits the expression of recessive *VRN-H1* alleles (Yan et al. 2004). Based on this model, genotypes with *VRN-H2_/vrn-H1vrn-H1/vrn-H3vrn-H3* allelic architecture flower late in the absence of vernalization (vernalization-sensitive) and all other allelic configurations lead to a lack of significant vernalization-sensitivity. This well validated epistatic interaction (reviewed in Szűcs et al. 2007) was used as a model system to test the ability of logic regression in identifying epistasis in binary molecular data.

Hop downy mildew epidemiology

Ware (1926) studied hop downy mildew and the mycelial invasion of the host plant. The early observations by Ware (1926 and 1929), followed by the work of Coley-Smith (1960), has led British researchers to believe that zoospores, originating from asexual zoosporangia, produce mycelium which over-winters in the crown of the hop plant. Thus, it was concluded that the pathogen over-winters as mycelium in the dormant crown buds and oospores do not play a significant role in the disease epidemiology (Ware 1929). These conclusions were later confirmed by Skotland (1961). However, Skotland and Johnson (1983) reported that the over-wintering role of oospores remained unclear and in wet environments, where resistant cultivars are grown, the germinating oospore may be an important source of primary inoculum. Royle and Kremheller (1981) agreed with Ware (1929) and suggested that the importance of oospores seems to have been exaggerated and imply that this belief is at least in part due to repeated failures in England and Poland to induce germination under laboratory conditions.

Recent research indicates that downy mildew may over-winter as living mycelium in onion, wild sorghum and rose (Populer 1981; Ryley 2001; Aegerter 2002). However, oospores have been documented in all these species and molecular

research suggests that oospores may be the primary over-wintering in *Peronospora sparsa* (syn. *P. rubi*) (Lindqvist-Kreuze et al. 2002). In addition, other research suggests oospores may play a significant role in disease epidemiology (Rooms of Diaz and Polanco 1984; Gowda and Bhat 1986; Kennelly et al. 2007).

Although conclusive evidence showing downy mildew mycelium within the 'dormant' crown tissue in hop is to date lacking, it has been accepted that hop downy mildew over-winters within the dormant buds of the hop crown and that oospores play no role in the disease cycle (Royle and Kremheller 1981), despite overwhelming evidence suggesting oospores play a major role in the epidemiology in most downy mildews (Populer 1981). Interestingly, oospores have been detected within the pith and parenchyma tissue in dormant buds, stems, leaves and cones of infected hop plants, (Royle and Kremheller 1981) making it difficult to assess their potential role in pathogenesis.

The hop downy mildew oospore

Arens (1929) completed extensive observations of hop downy mildew. He outlined a detailed life cycle of the pathogen and provided researchers with pertinent information needed to study the fungus in depth. Most importantly, he provided researchers with detailed environmental factors necessary for propagation of the fungus under laboratory conditions. In addition, Arens (1929) was the first researcher to germinate the oospore of hop downy mildew in the laboratory. Jones (1932b) successfully produced oospores of hop downy mildew in abundance by sowing 'Late Cluster' seedlings and inoculating the cotyledons with minute portions of infected leaves obtained from basal spikes. Bressman and Nichols (1933) reported they were able to get oospores of hop downy mildew to germinate in their laboratory. They used inoculum from two-year old herbarium specimens but were unable to confirm whether or not a dormancy period was involved in the oospore lifecycle. Unfortunately, none of these studies included a visual record to show that oospore germination had occurred.

Controversy concerning over-wintering

Research by Ware (1926 and 1929), Arens (1929), Hoerner (1949) and Coley-Smith (1960) suggest there were differences in opinion among American, Continental Europe and British hop researchers concerning the over-wintering mechanism in hop downy mildew. Arens (1929) reported that he was able to get hop downy mildew oospores to germinate in his laboratory in Germany and he believed that the oospore was the main over-wintering mechanism. In addition, Hoerner (1949) believed that hop downy mildew over-wintered as oospores. He reported there was ample inoculum in the form of infected leaves and shattered cones bearing innumerable oospores and these winter spores are capable of initiating the disease the following spring. Although the work by Ware (1929) and Coley-Smith (1960) substantiated mycelium in the non-dormant crown, researchers have never been able to provide visual evidence of downy mildew in the crown being connected to mycelium in the basal spikes. Furthermore, the timing and methods of observation were not adequate to rule out the potential contribution of the oospore. Therefore, the over-wintering mycelium hypothesis is only supported by circumstantial evidence by suggesting the over-wintering mycelium enter the dormant bud thereby being the primary source of inoculum from year to year.

Clearly, more work needs to be done to learn more about the ‘potential’ role of the hop downy mildew oospore in the disease cycle. Although some research has been done detailing the germination of oospores (Morgan 1978; Gowda and Bhat 1986; Van der Gaag and Frinking 1996; Spring and Zipper 2000), very little is known about oospore germination in oomycetes. A major factor contributing to this lack of information is that it is very difficult to collect the number of oospores required for observational studies on germination. In many infected plant tissues, oospore production is rare or non-existent. By investigating the oospore of hop downy mildew, we will gain further knowledge about sexual stage of this disease and hopefully gain some important information about its potential role in pathogenesis in hop.

Dissertation goal

The goal of this dissertation was to use the new F-statistics and Bayesian inference developed for dominant markers to help identify AFLP markers associated with the hop downy mildew resistant phenotype. Logic regression was a new statistical method which needed a well documented epistatic interaction in order to test the statistical procedure's ability to identify complex genetic interaction. These statistical analyses together with the observations of the hop downy mildew oospore will provide hop researchers and breeders with preliminary heritability values for the hop downy mildew resistant phenotype along with valuable information concerning the downy mildew oospore. It is hoped that this work will help increase our understanding of hop downy mildew as well as suggest some novel approaches to statistical modeling that can be used to learn more about this devastating disease.

Association Mapping Analysis in *Humulus lupulus* L. to Identify AFLP Markers Associated with the Hop Downy Mildew Resistant Phenotype

Thomas B. Parker[1], M. Shaun Townsend[1] and John A. Henning[2]*.

[1]Department of Crop and Soil Science, Oregon State University, Corvallis, OR, 97331, USA.

[2]USDA-ARS National Forage Seed Processing Research Center, Corvallis, OR, 97331, USA.

*Corresponding author: E-mail: John.Henning@oregonstate.edu.

ABSTRACT

Little is known about plant resistance to downy mildew [*Pseudoperonospora humuli* Miyabe and Tak.(Wil.)] in hop (*Humulus lupulus* L.), although previous research suggests a quantitative rather than a qualitative basis. The objective of our research was to use association mapping to identify potential molecular markers for future use in marker-assisted selection and to estimate heritability for this trait. We describe an association mapping approach using a population of 99 elite hop genotypes with varying levels of susceptibility to hop downy mildew. Plants were inoculated with sporangia for use in a leaf sampling assay for phenotype. In total, 492 AFLP markers were used in the analysis. Three subpopulations were identified within our data set. This population structure data was used to obtain an average estimate of the inbreeding coefficient for the three subpopulations as well as kinship estimates. These kinship and population estimates were subsequently used in a mixed-model approach for association mapping. Hop downy mildew resistant phenotype has a broad-sense heritability of 76% with an estimated narrow-sense heritability of 49%. Mixed-model results revealed 9% of the AFLP markers to be associated with the hop downy mildew resistant phenotype. The amount of phenotypic variation explained by the individual markers ranged from 4% to 11%. AFLP markers A2, D75 and E44 were the only three 'diagnostic' markers associated with the most susceptible genotypes. Our results suggest that fixation indices coupled with Bayesian inference can be valuable tools for use in association mapping within dioecious, clonal species like hop.

AFLP, association mapping, Bayesian, F-statistics, heritability, hop, *Humulus*, kinship, mixed-model, null allele, *Pseudoperonospora* and QTL

Abbreviations: AM for association mapping

INTRODUCTION

Modern QTL analysis and linkage mapping involves segregating populations derived from crosses with parents of contrasting phenotypes. Recombination frequencies between the markers and the genes of interest are estimated from co-segregation patterns. Although this method of analysis has a good track record for success, there are some limitations. QTL mapping can be costly and time-consuming because mapping requires a recombinant population and this population needs to be large to achieve high-resolution mapping. Furthermore, in a typical QTL study, only one cross from a population is made to form a recombinant population and therefore extrapolations to other populations and individuals may be spurious (Verhoeven et al. 2006).

Association mapping (AM) is utilized to find a statistical association between molecular markers and a quantitative trait. AM differs from QTL analysis in that it is performed at the population level where a collection of genotypic molecular markers and the corresponding phenotypic trait are determined in a set of unrelated or distantly-related individuals sampled from a population. The theory behind AM presumes that populations have had sufficient time to break most gene linkages and only those genes that are tightly linked remain present in the population. Therefore, AM relies on linkage disequilibrium between the molecular markers and the QTLs responsible for expression of the phenotypic trait.

In natural out-crossing populations like maize, linkage disequilibrium extends from a few hundred to less than 2000bp (Palaisa et al. 2001; Remington et al. 2001; Tenaillon et al. 2001; Gaut and Long 2003). Therefore, success in association mapping requires the candidate gene to have a measurable effect on the phenotype and candidate markers must fall within or directly upstream/downstream of the candidate gene. Markers identified as being in association with the trait are broad-based rather than cross-specific. In addition, some scientists believe that population structure/stratification can lead to spurious associations (associations without linkage)

between a candidate marker and phenotype (Lander and Schork 1994; Bacanu et al. 2000; Pritchard et al. 2000a and 2000b; Devlin et al. 2001; Yu et al. 2006) and that this type of structure needs to be accounted for in the data analysis. This is because individual subpopulations can have different allele frequencies than that of the whole population. Therefore, it is believed that unidentified population structure may lead to spurious association due to these unaccounted-for differences in allele frequencies leading to false association.

AM has been used in a number of plant species to identify molecular markers and genetic loci associated with phenotypic characters ranging from salt tolerance and eco-geographical traits in barley (Pakniyat et al. 1997; Ivandic et al. 2002; Ivandic et al. 2003) to cold tolerance in perennial ryegrass and morphological traits in rice (Virk et al. 1996). With an accurate phenotype and high numbers of polymorphic loci, AM has been shown to work. Unfortunately, generating high levels of polymorphic loci can be costly and time-consuming especially with co-dominant markers.

Until recently, dominant molecular markers (AFLP and RAPD) have been deemed less informative when compared to their co-dominant counterparts, due in large part, to the null allele. Heterozygote genotypes possessing the expressed band cannot be directly distinguished from homozygotes possessing the same band. Multi-locus dominant DNA markers, such as AFLP, pose a problem in the estimation of frequencies of null alleles. Recent advances in Bayesian statistical methods and F-statistics in population genetics have given researchers the statistical tools necessary for satisfactory estimates of allele frequencies leading to expected heterozygosities, genetic distances and F-statistics (Zhivotovsky 1999; Krauss 2000; Hardy and Vekemans 2002; Holsinger et al. 2002; Falush et al. 2007). These techniques were developed by population geneticists studying randomly sampled natural populations. However, it follows that these techniques can be extended for use in crop species. These statistical methods have been made available in software packages and many of the analysis techniques can be used for AM using dominant markers such as AFLP.

Trait analysis by association, evolution and linkage (TASSEL) has implemented a mixed-model approach in their AM software (Yu et al. 2006). Their model includes population and kinship estimates which are included to help reduce type I and type II errors. The use of AFLP data in programs such as TASSEL shows promise in a cost to benefit ratio where hundreds of dominant markers can be generated relative to the cost of co-dominant markers and, with recently added statistical procedures, have relatively equal informative value.

Hop downy mildew is one of the most economically important diseases in all but a few of the worlds hop producing regions (Neve 1991). In severe cases, complete yield loss and plant death can occur. Management of hop downy mildew requires numerous applications of fungicides throughout the growing season and often at times when soil moisture makes tractor movement in the field difficult. Effective management is further hindered by the development of fungicide resistance (Nelson et al. 2004). Therefore, development of resistant varieties is a major component of the U.S. public research and breeding facilities.

Partial resistance to downy mildew has been identified in some American cultivars including ‘Cascade’ (Neve 1991) and more recently ‘Newport’ (Henning and Haunold 2003), while partial levels of resistance are present in cultivars from the Hop Breeding Program at Wye College, England. Cultivars such as ‘Yeoman’ (Neve 1991; Neve and Darby 1984b) were initially used for production due to their strong resistance to this disease. Nevertheless, plant resistance to downy mildew in England eventually succumbed as new strains of the pathogen evolved. Despite decades of partial success in breeding for downy mildew resistance in both England and USA, no discernable pattern of inheritance for the hop downy mildew resistant phenotype has yet been identified.

The two objectives of this research were to determine the heritability of resistance to downy mildew in hop and to identify potential molecular markers associated with resistance using AFLP in AM format.

MATERIAL AND METHODS

Plant population

Ninety nine hop accessions representing a broad range of genetic backgrounds were evaluated. The accessions are part of the USDA-ARS hop breeding and genetics program's collection. The test population consisted of male (N=73) and female (N=26) individuals from wild American and European ancestry (Table 2.1).

Plant material

The male and female genotypes were growing in the USDA-ARS hop breeding and genetics research yard near Corvallis, OR. All plants received the same cultural practices and were grown under the same soil, Chehalis silty clay loam fertilized with urea at a rate of 168 kg ha^{-1} with regular irrigation applications. The fields were managed using practices similar to those used in commercial hop yards in the Pacific Northwest.

Hop downy mildew resistant phenotype

All genotypes were inoculated with a sporangial suspension (20,000 sporangia/ml) prepared for use in a backpack sprayer. Determination of sporangia concentrations were made prior to inoculation using a haemacytometer. Genotypes were inoculated in late April and early May on overcast days with light, misting rain. Multiple leaf samples per genotype were harvested following a six day incubation period. The samples were placed in separately labeled plastic bags containing a moist paper towel to induce sporulation. Upon sporulation, five leaf samples/genotype/year were selected arbitrarily for phenotypic analysis.

Each leaf was scanned into digital form and opened into GNU Image Manipulation Program (GIMP v 1.2©) (Free Software Foundation, Boston MA). The images were set to a blue background and the percent leaf area was determined using ASSESS© (APS Press, St. Paul, MO). Resulting percent leaf area infection

distributions exhibited a non-normal distribution. As a result, the percent leaf area diseased was normalized with an arcsine/square root transformation prior to running the ANOVA. The mean/year phenotypic values were used in the ANOVA and a three year mean was used in the AM analysis.

AFLP analysis

AFLP DNA fingerprints for all 99 individuals was performed by Townsend and Henning (2005). Approximately 100 to 600g of freeze dried leaf tissue was used in the DNA extraction protocol of Kidwell and Osborn (1992). The primer pair sequences for selective amplification and PCR conditions are described in Townsend and Henning (2005). Fluorescently labeled AFLP bands were detected on an ABI 377 DNA sequencer (Applied Biosystems, Foster City, CA).

Data analysis

AFLP gels were scored using Genographer© software (Benham et al. 1999). A binary data matrix was created based on the scored gel. The AFLP analysis of the 99 genotypes of *Humulus lupulus* L. resulted in the detection of 571 fragments with six primer combinations of which 492 were found to be polymorphic. The 79 monomorphic markers were removed from the data set for all subsequent analyses. Identification of population structure was made using a Jaccard's distance matrix with Ward's clustering in R© (R Foundation for Statistical Computing, Vienna, Austria). The cluster analysis was used for identification of population substructure and to identify hierarchical associations among sub-populations. In the course of our investigation, we analyzed population structure using other clustering methods such as UPGMA and Structure v 2.1© (Pritchard et al. 2000a). However, UPGMA produced a cluster plot that did not match well with prior cluster data (data not shown). Structure was unable to identify the three hop subpopulations currently accepted by hop researchers. In the few runs where Structure identified three clusters, the clusters did not match those identified using Ward's method (data not shown). Based on the three

published cluster analyses (Murakami 2000; Seefelder et al. 2000; Jakse et al. 2001) coupled with inconsistencies in the Structure analyses, we decided to use the Jaccard's distance/Ward's cluster results for use as binary covariates in the mixed linear model AM analysis. Using the population cluster information, an average estimate of the coefficient of inbreeding (F_{IS}) was obtained using the software Hickory v1.0© (Holsinger et al. 2002). The full model was chosen for the estimate of F_{IS} based on the Deviance Information Criterion (DIC). Estimates for kinship coefficients were obtained using the software Spagedi v1.2© (Hardy and Vekemans 2002) using the methods developed by Hardy (2003). The kinship estimates were normalized (negative values set to zero) and sorted for ANOVA in SAS Version 9.1© (SAS Institute, Cary, NC). The ANOVA was used in the calculation of estimates of phenotypic additive, environmental and g * e variance components, which were subsequently used to determine broad-sense heritability (*H*) and narrow-sense heritability (h^2) for the downy mildew resistant phenotype. The model for *H* was a type III random effects model with years treated as environments and genotypes and years as random effects. The model for h^2 was a type III random effects model with years treated as environments and populations and years as random effects. Kinship estimates obtained from Spagedi were sorted by population and used to calculate average coefficients of co-ancestry for each population. The average coefficient of co-ancestry was used in combination with estimated variance components to determine the additive variance and the resulting estimate of h^2. The kinship and population cluster matrices were used in an AM format using the mixed-model method in the software TASSEL© (Yu et al. 2006). The mixed-model computed the R^2 values for the individual markers to determine the amount of phenotypic variation explained by those markers found to be significantly associated with the hop downy mildew resistance phenotype.

RESULTS

The percent leaf area diseased showed a strong right tail (Figure 2.1A), so we normalized the data set with an arcsine/square root transformation (Figure 2.1B). Jaccard's distance and Ward's clustering showed two large subpopulations which were further divided into three smaller subpopulations containing n=35, n=28 and n=36 individuals, respectively (Figure 2.2). The cluster analysis showed that the genotypes were grouped according to their region of origin with one group primarily Continental European, the second group primarily English/American/European hybrids (A) and the third cluster, Brewers Gold/wild American/European hybrids (B) (Table 2.1). A small cluster of resistant female genotypes was revealed within the European cluster (Figure 2.2). Running the AFLP marker data according to subpopulation divisions resulted in a mean estimation for the coefficient of inbreeding (F_{IS}) of 0.2014. The DIC was 6153 units for the full model, 6182 units for the $f = 0$ model and 9105 units for the $\Theta^B = 0$ model.

The ANOVA for broad-sense heritability revealed statistically significant differences in the means for all the variables tested (Table 2.2). There was also a statistically significant interaction among genotypes*years (Table 2.2). The ANOVA used to generate estimates of genetic variances for narrow-sense heritability revealed statistically significant differences in means among populations and also years with no significant interaction among population*years (Table 2.3). The ANOVA used to generate estimates of overall population variance resulted in significant differences among the individuals in the population. Broad-sense heritability estimate of the downy mildew resistance phenotype was 76% and a narrow sense-heritability of the phenotype equaled 49%.

The mixed linear modeling analysis in TASSEL revealed 43 AFLP markers (9% of the total) associated with the hop downy mildew resistant phenotype with p-values < 0.05 (Table 2.4A and B). The amount of phenotypic variation explained by the individual markers ranged from 4% to 11% (Table 2.4A and B). Only three of the

forty-three markers were diagnostic with the most susceptible genotypes. Markers A2, D75 and E44 were absent in 7 highly resistant genotypes (Perle, Yeoman, Challenger, Omega, Orion and Wye Viking and M64037) while present in 7 highly susceptible genotypes (Comet, Galena VF, Wye Target, M21420, M21329, M21313 and M21339).

DISCUSSION

We observed strong evidence suggestive of population structure based upon hierarchical clustering (Figure 2.2). Our results suggesting three subpopulation clusters in hop agree with those previously reported (Murakami 2000; Seefelder et al. 2000; Jakse et al. 2001). Analysis of our data in Hickory revealed a large difference in the DIC between the full and the $\Theta^B = 0$ model which suggest there is compelling evidence for genetic differences among the populations (a difference of 2952 units) (Spiegelhalter et al. 2002). In addition, the 29 unit difference in DICs between the full and the $f = 0$ model, suggest there is some level of inbreeding within subpopulations. The data showed a moderately high level of inbreeding within the subpopulations with a mean $F_{IS} = 0.2014$. The high level of inbreeding within populations may be due to the narrow genetic base of cultivated hop since many cultivars share kinship amongst three or four cultivars (Henning et al. 2004). These results suggest our data set deviated from Hardy-Weinberg equilibrium and this deviation needed to be accounted for when deriving kinship estimates using the model proposed by Hardy (2003).

Variance components and heritability

The variance among genotypes was significant (Table 2.2) suggesting that there is genetic variation for resistance to hop downy mildew among genotypes present in the population under observation. Certainly, the distribution of resistance scores (Figure 2.1B) implies quantitative control over the expression of resistance. The ANOVA for broad-sense heritability also showed that there was significant interaction among genotypes*years demonstrating that multiple environments must be utilized when examining this trait in order to effectively identify genotypes possessing resistance (Table 2.2).

Variance components estimated from the ANOVA for narrow-sense heritability also showed significant ($p<0.05$) variation among populations (Table 2.3). This demonstrates that there are true differences in levels of resistance among the

populations and that these differences in populations should prove responsive to selection. We did not observe any significant interaction for population*year and therefore the variance component for this interaction was not estimated (Table 2.3). It was interesting that the genotype*year interaction proved significant but the population*year was not significant. Causes for this difference in significance may lie in the different genetic components estimated by these two tests. In the case of genotypes*year, the genetic component estimated from the ANOVA consists of both additive and dominance genetic components. In the case of the population*year interaction, the genetic component of population should theoretically consist of additive genetic variance. Thus, the difference between both estimates of interaction may lie in the presence of a strong dominance effect upon this trait.

Our estimates of broad-sense (H^2 = 76%) and narrow-sense heritability (h^2 = 49%) for the downy mildew resistant phenotype also suggest a means whereby estimates of dominance or epistasis are possible. Theoretically, broad-sense heritability represents total genetic variation relative to phenotypic variation. As a result, the estimate of variance components for total genetic variance consists of additive, or selectable variation, and dominance/epistasis, or non-selectable deviations from expectations. On the other hand, narrow-sense heritability estimates represent the ratio of selectable genetic variation, or additive genetic variance, relative to total phenotypic variation. Thus, comparing between the two estimates of heritability suggests, but does not prove, the presence of a strong dominance component. If true in this case, the presence of significant dominance variation could prove deleterious to response to selection if phenotypic or mass selection was the primary mode of selection—as it has proven to be in hop (Henning 2006). The only means to overcome this effect would be the use of some means of genotypic recurrent selection or by means of molecular markers linked to downy mildew resistance or susceptibility. As genotypic recurrent selection for this particular trait would prove highly time-consuming and space-consuming, use of marker-assisted selection should prove highly advantageous.

Association mapping using mixed linear models

Analysis of the mixed-model results from TASSEL showed a large number of AFLP markers (43) associated with the downy mildew resistant phenotype. The percentage of phenotypic variation explained by the individual markers was quite low ranging from 4% to 11% of the total variation (Table 2.4A and B). These low estimates suggest some of the variation was a result of environmental variance. Repeated study on the incidence of hop downy mildew have shown that prevailing weather conditions conducive to its development and spread may play a large role in year to year variation in the disease (Hoerner 1939; Pejml and Petrlik 1964; Pejml and Petrlik 1967; Royle 1970; Royle 1973; Skotland and Johnson 1983; Skotland and Romanko 1964). After reexamining the phenotype between years, it became apparent that this could be one explanation for the low R^2 values.

Young (1996) reported that most resistant phenotypes are measured quantitatively, so they are known as quantitative resistant characters and the genetic loci associated with them are called quantitative resistant loci (QRL). Although polygenic suggests that many minor genes have equal effect on the phenotype, QTL mapping suggests this is not true (Young 1996). It has been shown that for many diseases, most often a few QTL are involved in quantitative resistance. Compared with prior work on quantitative resistant loci (QRL), forty-three markers appear to be a very large number to be associated with a disease resistance phenotype. In barley powder mildew (*Erysiphe graminis*) there are two QRL, common bean bacterial blight (*Xanthomonas campestris*) there are seven QRL, Mungbean powdery mildew (*Erysiphe polygoni*) there are three QRL, soybean cyst nematode (*Heterodera glycines*) there are three QRL in addition to many other similar examples (Young 1996).

However, 43 loci may not be very many loci to be associated with the hop downy mildew resistant phenotype if they loci are clustered into few, large QRL or the trait itself is truly highly quantitative in nature similar to such traits as ‘yield’. Recent research by Dodds et al. (2006), suggest that the type of host response to pathogen will dictate the level of complexity involved in host resistance. In true ‘gene for gene’

resistance mechanisms such as that observed by Flor (1942), simple R-Avr protein interactions should result in qualitatively inherited traits. In host-pathogen interactions such as *Arabidopsis* and *Psuedomonas syringae* the mechanism of resistance is more complex and the host response is dependent upon detecting modifications of disease response proteins (Dodds et al. 2006). In the latter case, it is quite easy to see how a mechanism for resistance could be significantly more quantitative than that for a simple R-Avr protein interaction. Nevertheless, this does not discount the presence of multiple overlapping markers found within major regions of association with downy mildew resistance nor does it account for prior evidence suggesting few QRL. Without actually mapping the location of markers via QTL analysis neither explanation can be proven from our data.

Although the mixed-model analysis may have resulted in type I and/or type II errors, there is no way to know for sure as there is no mapping population for the hop downy mildew resistant phenotype. After data mining our results, we discovered three AFLP markers (markers A2, D75 and E44) to be 'diagnostic' for the most susceptible hop genotypes. However, by definition, 'all' 43 markers are QRL. The data mining can only suggest which markers 'appear' to have a strong association with the resistant or susceptible phenotype, whether they are the true QRL or not. Currently, we are in the process of running bi-parental 'pseudo testcrosses' to help determine mapping locations and presence of major regions of influence.

Conclusions

The AFLP analysis of *H. lupulus* for resistance to hop downy mildew showed that association mapping techniques may prove useful in the dissection of complex traits in dioecious, clonal crop species like hop and in the identification of useful molecular markers of potential use in a marker-assisted selection breeding program. Our work suggests that the presence of dominance controlling the expression of resistance may be one factor explaining the difficulty breeders have experienced over the years when practicing phenotypic or mass selection. Other factors that must be

taken into account are the effects of environment upon screening and the possibility that resistance to downy mildew in hop may be controlled by numerous QTL, each with minor effects upon expression. Certainly, our results demonstrate the need for further examination over the expression of resistance to downy mildew in hop and that either marker-assisted selection or some means of genotypic recurrent selection should be practiced if expected genetic gain due to selection is hoped for. Finally, the ANOVA results of hop downy mildew resistance suggest the new methods in Bayesian inference used in the derivation of F-statistics for AFLP marker data can be applied to dioecious, clonal crop species like hop. We believe these new methods to be robust and applicable to out-crossing crop species like hop making dominant markers such as AFLP just as informative as co-dominant markers. Because AFLP markers are less expensive to generate on a ‘per band basis’ in terms of time and money, many studies in minor crops can now be realistically investigated.

REFERENCES

Bacanu S-A, Devlin B and Roeder K (2000) The power of genomic control. Am J Hum Genet 66: 1933-1944

Benham J, Jeung J, Jasieniuk M, Kanazin V and Blake T (1999) Genographer: A graphical tool for automated fluorescent AFLP and microsatellite analysis. http://hordem.oscs.montana.edu/genographer/ (Checked 5/26/2007)

Devlin B, Roeder K and Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Pop Biol 60: 155-166

Dodds P N, Lawrence G J, Catanzariti A, Teh T, Wang C A, Ayliffe M A, Kobe B and Ellis J G (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc Nat Acad Sci USA 103: 8888–8893

Falush D, Stevens M and Pritchard J K (2007) Inferences of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecology Notes doi: 10.1111/j.147-8286.2007.01758.x

Flor H H (1942) Inheritance of pathogenicity in *Melampsora lini*. Phytopath 32: 653-669

Gaut B S and Long A D (2003) The lowdown on linkage disequilibrium. Plant Cell 15: 1502-1506

Hardy O J (2003) Estimation between individuals and characterization of isolation-by-distance processes using dominant genetic markers. Mol Ecology 12: 1577-1588

Hardy O J and Vekemans X (2002) SPAGeDI: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecology Notes 2: 618-620

Henning J A (2006) The breeding of hop. In: Bamworth C (ed) Brewing: New Technologies. Woodhead Publishing Limited Cambridge, England, pp. 102-122

Henning J and Haunold A (2003) Notice of release of high-yielding, multi-disease resistant, high bittering acid hop variety, 'Newport'. USDA-ARS, Washington, DC

Henning J A, Townsend M S and Kenny S (2004) Potential heterotic crosses in hops as estimated by AFLP-based genetic diversity and coefficient of coancestry. J Am Soc Brew Chem 62: 63-70

Hoerner G R (1939) The relation of the climatology of western Oregon to the incidence and control of downy mildew of hops. Plant Dis Rpt 23: 361-366

Holsinger K E, Lewis P O and Dey D K (2002) A Bayesian approach to inferring population structure for dominant markers. Mol Ecology 11: 1157-1164

Ivandic V, Hackett C A, Nevo E, Keith R, Thomas W T B and Forster B P (2002) Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. Plant Mol Biol 48: 511–527

Ivandic V, Thomas W T B, Nevo E, Zhang Z and Forster B P (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. Plant Breed 122: 300–304

Jakse J, Kindlhofer K and Javornik B (2001) Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. Genome 44: 773-782

Kidwell K K and Osborn T C (1992) Simple plant DNA isolation procedures. In: Plant genomes. Methods for genetic and physical mapping. Kluwer Academic Publ, the Netherlands, pp. 1-13

Krauss S L (2000) Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. Mol Ecology 9: 1241–1245

Lander E S and Schork N J (1994). Genetic dissection of complex traits. Science 265: 2037–2048

Murakami A (2000) Hop variety classification using genetic distance based on RADP. J Inst Brew 106: 157-161

Nelson M E, Eastwell K C, Grove G G, Barbour J D, Ocamb C M and Alldredge J R (2004) Sensitivity of *Pseudoperonospora humuli* (the casual agent of hop downy mildew) from Oregon, Idaho, and Washington to fosetyl-Al (Aliette). Online. Plant Health Progress doi: 1094/PHP-2004-0811-01-RS

Neve R A (1991) Hops. Chapman and Hall, New York

Neve R A and Darby P (1984b) Yoeman and Zernith: two new varieties. Rep Dept Hop Res Wye Coll for 1983, pp. 7-9

Pakniyat H, Powell W, Baird E, Handley L L, Robinson D, Scrimgeur C M, Nevo E, Hackett C A, Caligari P D S and Forster B P (1997) AFLP variation in wild barley (*Hordeum spontaneum* C Koch) with reference to salt tolerance and associated ecogeography. Genome 40: 332–341

Palaisa K A, Morgante M, Williams M and Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15: 1795–1806

Pejml K and Petrlik Z (1964) The influence of weather conditions on the spread of downy mildew. Chmelarstvi 37: 73-73

Pejml K and Petrlik Z (1967) Typization of characteristic weather for spreading of downy mildew in hops. Ochr Rost 5: 109-116

Pritchard J K, Stephens M and Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Pritchard J K, Stevens M and Donelly P (2000b) Association mapping in structured populations. A J Hum Genet 67: 170-181

Remington D L, Thornsberry J M, Matsuoka Y, Wilson L M, Whitt S R, Doebley J, Kresovich S, Goodman M M and Buckler IV E S (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98: 11479–11484

Royle D J (1970) Infection periods in relation to the natural development of hop downy mildew (*Pseudoperonospora humuli*). Ann App Biol 66: 281-291

Royle D J (1973) Quantitative relationships between infection by hop downy mildew pathogen, *Pseudoperonospora humuli*, and weather and inoculum factors. Ann Appl Biol 73: 19-30

Skotland C B and Johnson D A (1983) Control of downy mildew of hops. Plant Dis 67: 1183-1186

Skotland C B and Romanko R R (1964) Life history of the hop downy mildew fungus. Bull Idaho Agric Exp Stn, no. 424

Seefelder S, Ehrmaier H, Schweizer G and Seigner E (2000) Genetic diversity and phylogenetic relationships among accessions of hop, *Humulus lupulus*, as determined by amplified fragment length polymorphism and fingerprinting compared with pedigree data. Plant Breed 119: 257-263

Spiegelhalter D J, Best N G, Carlin B P and van der Linde A (2002) Bayesian measures and model complexity and fit. J Royal Stat Soc Ser B 64: 483-689

Tenaillon M I, Sawkins M C, Long A D, Gaut R L, Doebley J F and Gaut B S (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). Proc Natl Acad Sci USA 98: 9161–9166

Townsend M S and Henning J A (2005) Potential heterotic groups in hop as determined by AFLP analysis. Crop Science 45: 1901-1907

Vekemans X (2002) AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium

Verhoeven K J F, Jannink J-L, McIntyre L M (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96: 139-149

Virk P S, Ford-Lloyd B V, Jackson M T, Pooni H S, Clemeno T P and Newbury H J (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76: 296–304

Young N D (1996) QTL mapping and quantitative disease resistance in plants. Ann Rev Phytopath 34: 479-501

Yu J, Pressoir G, Briggs W H, Bi I V, Yamasaki M, Doebley J F, McMullen M D, Gaut B S, Nielsen D M, Holland J B, Kresovich S and Buckler E S (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38: 203-208

Zhivotovsky L A (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecology 8: 907–913

Figure 2.1A and B. Frequency distribution for the hop downy mildew resistant phenotype. The percent leaf area diseased (A) and the arcsine/square root transformation (B) for three years of data are displayed for the 99 genotypes of *Humulus lupulus* L. on the x-axis as categorical bins in relation to the frequency response (y-axis).

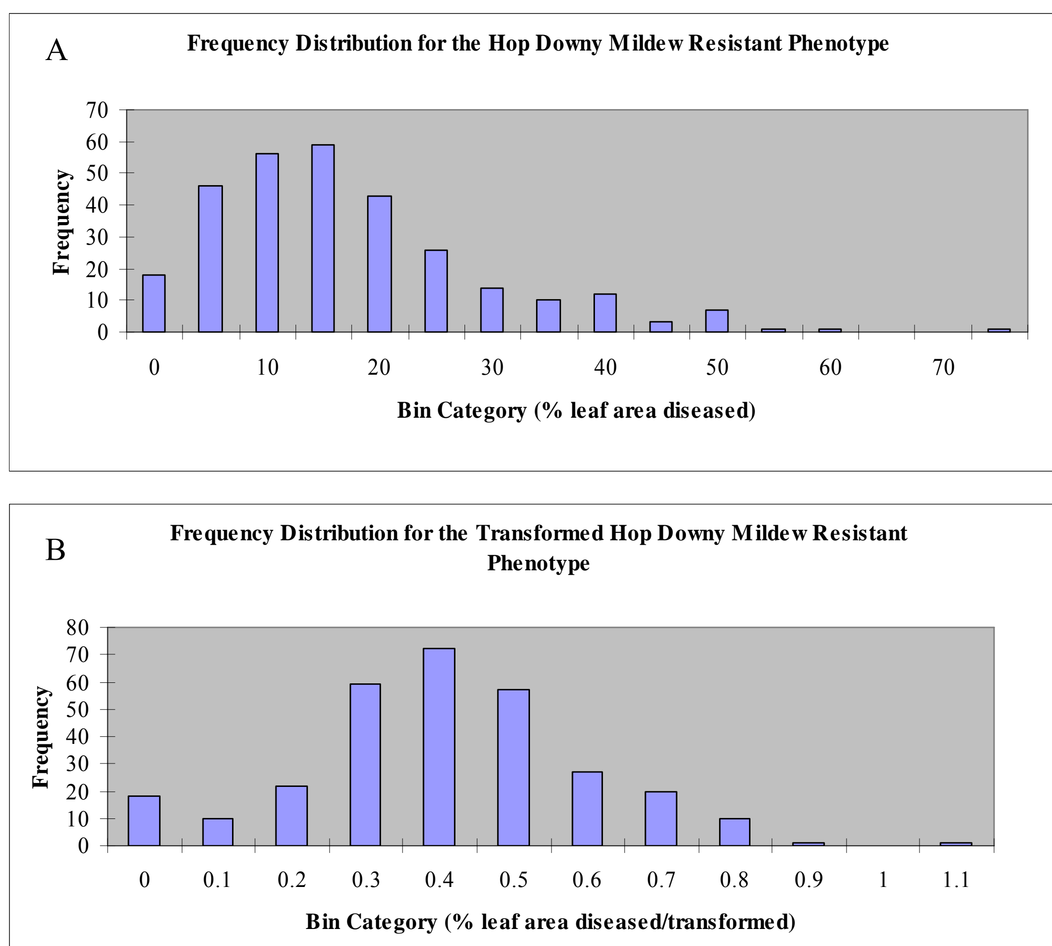


Figure 2.2. Phenogram of Ward's cluster analysis of AFLP molecular marker data from 99 genotypes of *Humulus lupulus* L. with the three subpopulations identified and the corresponding three year mean percent leaf area disease listed next to the genotype.

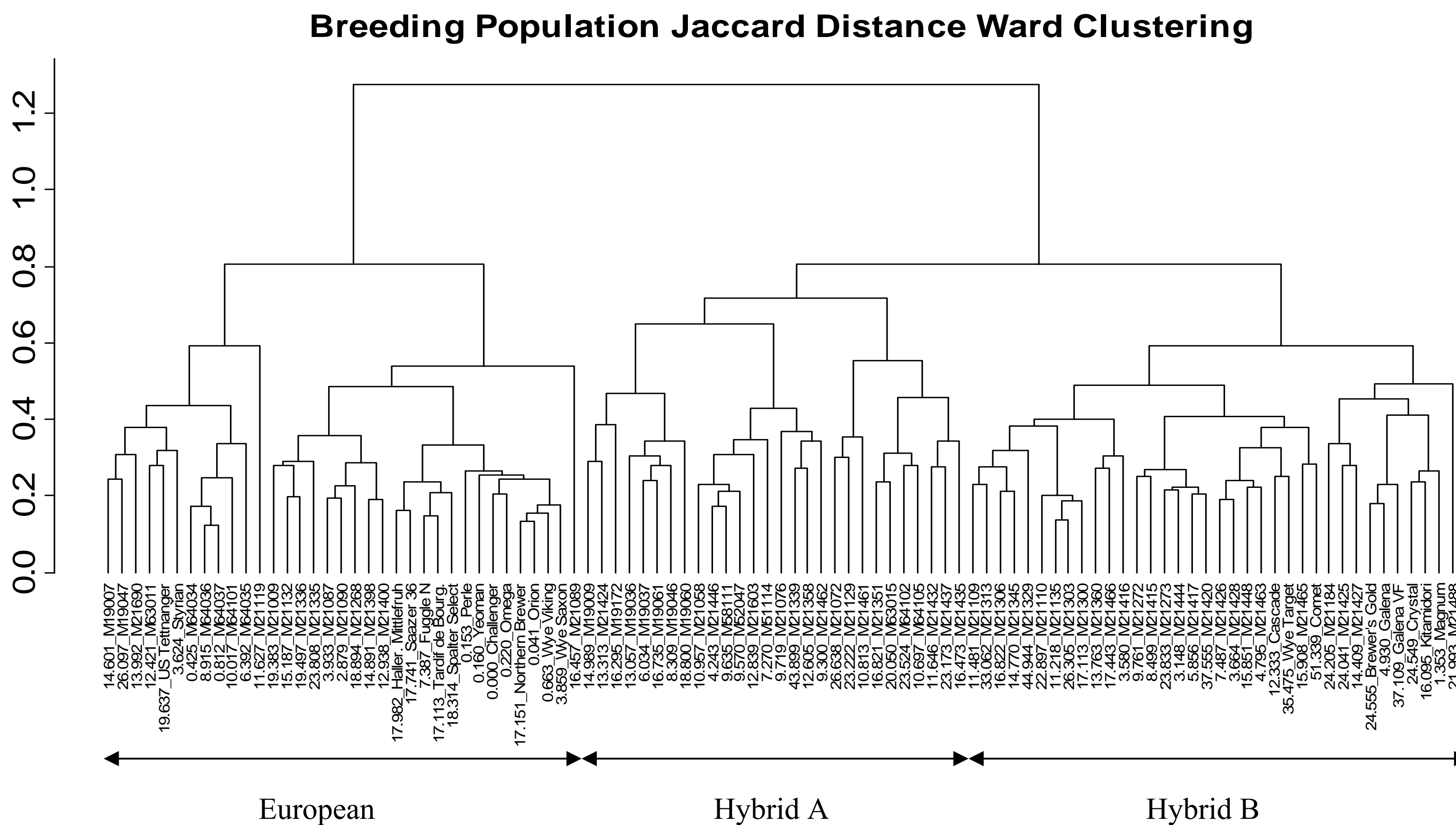


Table 2.1. List of hop genotypes arranged by pedigree based on cluster analysis of AFLP molecular marker data.

Primarily European	Primarily Wild American/European Hybrids (A)	Primarily Wild American/European Hybrids (B)
M19007	M19009	M21109
M19047	M21424	M21313
M21690	M19172	M21306
M63011	M19036	M21345
US Tettnanger	M19037	M21329
Styrian	M19061	M21110
M64034	M19046	M21135
M64036	M19060	M21303
M64037	M21058	M21300
M64101	M21446	M21360
M64035	M58111	M21466
M21119	M52047	M21416
M21009	M21603	M21272
M21132	M51114	M21415
M21336	M21076	M21273
M21335	M21339	M21444
M21087	M21358	M21417
M21090	M21462	M21420
M21268	M21072	M21426
M21398	M21129	M21428
M21400	M21461	M21448
Hallertauer Mittlefruh	M21351	M21463
Saazer 36	M63015	Cascade
Fuggle N	M64102	Wye Target
Tardif de Bourgogne	M64105	M21465
Spalter Select	M21432	Comet
Perle	M21437	M21184
Yeoman	M21435	M21425
Challenger		M21427
Omega		Brewer’s Gold
Northern Brewer		Galena
Orion		Galena VF
Wye Viking		Crystal
Wye Saxon		Kitamidori
M21089		Magnum
		M21488

Table 2.2. Analysis of variance results for broad-sense heritability of the transformed hop downy mildew phenotype. Years were treated as environments in a type III random effects model.

Source	DF	Type III SS	Mean Square	F-value	*p*-value	Variance Components
Genotype	98	37.921	0.387	6.81	< 0.001*	0.022
Year	2	0.936	0.468	8.24	0.0004*	0.0008309
Error: MS(Genotype*Year)	196	11.140	0.057			
Genotype *Year	196	11.140	0.057	2.14	<0.001*	0.0060664
Error: MS(Error)	1188	31.486	0.027			

*Significant at the 0.001 level

Table 2.3. Analysis of variance and estimated variance components used in calculating narrow-sense heritability of the transformed hop downy mildew phenotype. Years were treated as environments in a type III random effects model.

Source	DF	Type III SS	Mean Square	F-value	*p*-value	Variance Components
Population	2	4.878	2.439	111.840	0.003*	0.0049118
Error: MS(Pop*Yr)	4	0.087	0.022			
Year	2	0.927	0.463	21.070	0.007*	0.000762
Error: 0.993*MS(Pop*Yr) + 0.007*MS(Error)	4.129	0.091	0.022			
Population*Year	4	0.087	0.022	0.022	0.790	-0.0001792
Error: MS(Error)	1476	72.602	0.051			

*Significant at the 0.01 level

Table 2.4A. Mixed linear model results for primer pairs A-C listing the AFLP markers associated with the hop downy mildew resistant phenotype, *p*-values and marker R^2 values.

Marker	*p*-value	R^2
A1	0.0497	0.0365
A2	0.0057	0.0705
A6	0.0018	0.0885
A12	0.0314	0.0438
A71	0.0236	0.0480
B11	0.0257	0.0465
B13	0.0301	0.0445
B20	0.0370	0.0415
B21	0.0239	0.0482
B58	0.0240	0.0480
B69	0.0151	0.0549
B70	<0.001	0.1023
B91	0.0191	0.0522
C8	0.0386	0.0403
C18	0.0406	0.0397
C49	0.0037	0.0785
C73	0.0240	0.0480
C75	0.0099	0.0627

Table 2.4B. Mixed linear model results for primer pairs D-F listing the AFLP markers associated with the hop downy mildew resistant phenotype, *p*-values and marker R^2 values.

Marker	*p*-value	R^2
D17	0.0468	0.0371
D18	0.0168	0.0532
D60	0.0303	0.0442
D75	0.0265	0.0464
D77	<0.001	0.1083
D79	0.0128	0.0586
D92	0.0449	0.0380
D96	<0.001	0.1037
E8	0.0224	0.0495
E28	0.0191	0.0523
E30	0.0085	0.0641
E40	0.0392	0.0401
E44	<0.001	0.0949
E47	0.0227	0.0485
E68	0.0440	0.0386
E79	0.0152	0.0558
E84	0.0338	0.0425
E87	0.0274	0.0459
E91	0.0018	0.0888
E95	0.0419	0.0389
E107	0.0049	0.0748
E111	0.0019	0.0881
F57	0.0068	0.0688
F59	0.0071	0.0673
F83	0.0407	0.0393

TITLE

The Extraction, Tetrazolium Staining and Germination of the Oospore of *Pseudoperonospora humuli* Miyabe and Tak. (Wil.)

T. B. Parker[1], J. A. Henning[2], D. Gent[2] and W. F. Mahaffee[3*].

[1]Department of Crop and Soil Science, Oregon State University, Corvallis, OR, 97331, USA.

[2]USDA-ARS National Forage Seed Processing Research Center, Corvallis, OR, 97331, USA.

[3]USDA-ARS Horticultural Crops Research Laboratory, Corvallis, OR, 97331, USA.

*Corresponding author: E-mail: mahaffew@science.oregonstate.edu

ABSTRACT

New protocols were developed to extract, isolate and recover oospores of [*Pseudoperonospora humuli* Miyabe and Tak. (Wil.)] from hopyard soil and cotyledon material of hop (*Humulus lupulus* L.). Oospores extracted from autoclaved-treated and untreated hopyard soil and those generated *in situ* in cotyledons were used for observational studies on germination and MTT staining. Oospores extracted from soil were washed through a series of wet sieves, collected, run through density centrifugation with a final wash and placed in de-ionized water. Oospores were produced in cotyledons by inoculating hop seedlings with zoosporangia. After six days of incubation, the cotyledons were collected and air dried. Oospores were extracted from cotyledons by repeatedly grinding wetted plant material in a mortar and pestle and processing the material through a series of sieves. The oospore inoculum was given a final wash and placed in de-ionized water. Oospore counts per g soil were calculated and analyzed with a Kruskal-Wallis rank sum test. There were significant differences in mean ranks between autoclaved and non-autoclaved samples and among MTT color categories while there was no significant difference in mean ranks among the samples. Germination of oospores occurred two to eight weeks after slide preparation. Our results indicate that oospores of *P. humuli* can germinate *in vitro* and, therefore, may play a role in the epidemiology of hop downy mildew. In addition, we suggest colored formazan production may result from a chemical reduction of NADH and, therefore, the use of MTT as a test for viability of oospores must be reexamined.

INTRODUCTION

Pseudoperonospora humuli Miyabe and Tak. (Wil.), the causal agent of hop downy mildew, has become a widespread pathogen affecting most of the hop growing regions of the world. It is an obligate pathogen specific to hop. If left unchecked, both cone quality and crop yield are affected and in severe cases, complete yield loss and plant death can occur. Although downy mildew of hop was first discovered in North America almost 100 years ago (Davis 1910), there are still numerous unanswered questions concerning the epidemiology of this disease.

Ware (1926) studied hop downy mildew and the mycelial invasion of the host plant. While oospores were observed within the pith of dormant buds (Ware 1929), numerous attempts failed to result in any observable germination. Thus, it was concluded that the pathogen over-winters as mycelium in the dormant crown buds (Ware 1929) and oospores do not play a significant role in the disease epidemiology. These conclusions were later confirmed by Skotland (1961). However, Skotland and Johnson (1983) reported that the over-wintering role of oospores remained unclear and in wet environments, where resistant cultivars are grown, the germinating oospore may be an important source of primary inoculum. Royle and Kremheller (1981) agreed with Ware (1929) and suggested that the importance of oospores seems to have been exaggerated and imply that this belief is at least in part due to repeated failures in England and Poland to induce germination under laboratory conditions.

However, oospores from *P. humuli* have been reported to germinate *in vitro* (Arens 1929; Bressman and Nichols 1933) and are formed in leaves (Hoerner 1949, Royle and Kremheller and 1981), shoots and cones (Royle and Kremheller and 1981), and in pith tissue within crown and bud (Ware 1929; Royle and Kremheller and 1981). Thus, there is potential that oospores could serve as the inoculum source for basal spikes that originate in the spring.

Unfortunately, very little is known about oospore production and germination in *P. humuli*. A major factor contributing to this lack of information is that it is very

difficult to collect sufficient quantities of oospores from the field for observational studies on germination. In warmer locations such as Washington, USA , oospore production is rare or non-existent (Chee et al. 2006). Oospores of *P. humuli* have been produced in abundance by sowing 'Late Cluster' seedlings and inoculating the cotyledons with minute portions of infected leaves obtained from basal spikes (Jones 1932b). Recently, Chee and Klein (1998) reported on the production of oospores in the laboratory as a function of temperature. Therefore, this information may prove useful in future investigations aimed at learning the precise role (if any) that the oospore plays in the downy mildew disease cycle.

Tetrazolium salts have been used for many years in oxidative-reduction histochemistry (Altman 1974; Altman 1976). During this time, tetrazolium salts have been used as a vitality assay in seeds (MacKay 1972; Van Waes and Debergh 1986), pollen (Binder et al. 1974) and oospores (Sutherland and Cohen 1983; Cohen 1984; El-Hamalawi and Erwin 1986; Jiang and Erwin 1990; Van der Gaag 1994). MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide) is reduced within the electron transport chain at the Co-enzyme Q/Cytochrome b site in plant tissue (Gahan and Kalina 1968). Within living cells, MTT is reduced to form insoluble colored formazans by dehydrogenase enzymes. The formation of different colored formazans is partly dependent on concentrations of chelating ions (copper, cobalt, silver and nickel) and solvent polarity, however, the formation of colored formazans never indicates a negative reaction (Altman 1974).

A major problem with MTT as a viability stain for oospores is that color is often seen in the control spores (Sutherland and Cohen 1983; Meier and Charvat 1993; Van der Gaag 1994; Medina and Platt 1999). In addition, the interpretation of color varies among researchers. Sutherland and Cohen (1983) and El-Hamalawi and Erwin (1986) reported that rose-colored oospores are dormant, active oospores are blue and non-viable oospores are black or unstained. Bowers et al. (1990) reported red-stained oospores were considered viable while black and non-stained were nonviable. Similar to Bowers' report, Van der Gaag and Frinking (1997) make no mention of

blue-stained oospores and assume red-stained to be viable with black and clear non-viable.

The objectives of this research were to develop techniques for collection of large quantities of oospores and examine the potential for oospore germination. To accomplish these objectives, oospores from several sources were isolated and incubated for several weeks under conditions suitable for germination and stained with MTT viability stain.

MATERIALS AND METHODS

Production of oospores from plant material

Oospores used in the studies were obtained by inoculating hop cotyledons as described by Jones (1932b) with some modifications. Hop seeds were stratified similar to Haunold and Zimmerman (1972). The seeds were surface sterilized in 5% sodium hypochlorite, washed and stratified in autoclaved moist peat moss for eight weeks at 4°C. Seeds were sown in flats of Sunshine™ BX-Mix and placed on a greenhouse bench and watered as needed. When the first primary leaf was apparent, the seedlings were sprayed with a 2×10^4 zoosporangia/ml suspension that was obtained by washing sporulating hop leaves collected at the USDA-ARS Hop Research Station outside Corvallis OR in de-ionized (DI) water. Zoosporangia concentration was estimated using a haemacytometer. The flats of seedlings were bagged overnight then incubated on a greenhouse bench with the bags removed at 26°C for 6 days prior to inducing sporulation. Sporulation was induced by misting the flats with DI water and bagging them overnight. The cotyledons showing signs and symptoms of downy mildew were harvested and allowed to air dry for two weeks.

Extraction of oospores from plant tissue

Dried cotyledon material was wetted with DI water and gently ground using a mortar and pestle. The suspension was passed through a series of sieves (250, 90, 75, 50, 35 and 25µ mesh) with the oospores being concentrated on the 35µ and 25µ sieves. Material remaining on the 250, 90, 75 and 50µ sieves was re-collected, ground and passed through the sieve series for further collection. This process was repeated until less than 10% of the original material remained. Once the grinding and sieving steps were complete, the oospore/tissue mixture was suspended in 0.05M citrate buffer pH 4.6. An equal volume of a buffered solution of cellulase (4mg/ml in 0.05M citrate buffer, pH 4.6) was mixed with the oospore suspension and incubated for 2 hours at 20°C. After incubation, the macerated suspension was placed in the 25µ sieve and

rinsed with DI water for 5 minutes. The oospores were collected in a small beaker and brought up to 10ml with DI water amended with 200ug/ml ampicillin to inhibit the growth of bacteria.

Extraction of oospores from soil

Eight soil samples were collected from the top 6cm of soil and duff above the hop crowns at the USDA Hop Research farm outside Corvallis, OR. The soil samples were dried and 500g were used for subsequent analysis. The treatment samples (4) were autoclaved for 60 minutes allowed to cool and then re-autoclaved an additional 60 minutes. The control samples (4) were only air-dried before weighing. The samples were washed through a series of sieves (225, 75, 53 and 25µ). The material remaining in the 25µ sieve was washed into a 500ml centrifuge tube to which 300ml of 70% sucrose solution and DI water was added for a final volume of 500ml. The suspension was centrifuged at 3500rpm for 5 minutes. The resulting supernatant was strained through a 25µ sieve and the remaining material was collected with 1mM potassium phosphate buffer pH 6.3 to a volume of 10ml. Viability of the oospore populations was estimated using tetrazolium bromide (MTT) (Spring and Zipper 2000; Van Der Gaag and Frinking 1997). Number of oospores/g soil was calculated by counting the number of oospores in four 75µl subsamples/soil sample.

MTT stain of oospores

Viability of oospores was estimated with 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) according to Van der Gaag (1994). Oospores extracted from soil were allowed to incubate for two weeks in DI water containing ampicillin prior to staining with MTT in an attempt to catch and stain a germination event in progress. After the initial incubation period, both the control and treatment oospores were incubated for 36h in a 0.1% MTT solution in a 1mM potassium phosphate buffer (pH 6.3) at 37°C. The MTT oospore suspensions were vortexed and a 50ul sample was transferred to a microscope slide, covered with a

coverslip, and edges sealed with nail polish to prevent evaporation. Violet/blue, red, chestnut-brown, clear and black soil-borne oospores were counted. Clear oospores were considered non-viable and chestnut-brown oospores were considered viable but dormant while violet/blue and red-stained oospores were considered viable.

Data analysis

A Kruskal-Wallis test was used to test the null hypothesis (HO) that the mean ranks of samples were the same between the control and autoclaved samples and among the color categories and samples (Table 3.1) using SAS Version 9.1© for windows as the count data were non-normally distributed (SAS Institute, Cary, NC).

Germination of oospores

An oospore suspension was mixed and 50µl aliquots were transferred to (25x75x1mm) microscope slides and covered with (22x40mm No.1) cover-slips. Nail polish was used to border the cover slips to prevent evaporation. Daily observations were made and upon completion of the day's observations, the slides were placed in petri dishes containing wetted filter paper and wrapped with Parafilm™. Observations continued over an eight-week or more period as necessary. Germination images were obtained using a Leica DMRB™ compound microscope at 400x magnification and SPOT RT Color™ digital camera model 2.2.1 (Diagnostic Instruments Inc., Sterling Heights, MI) processed with Image Pro™ 3.1 software (Leeds Precision Instruments, Minneapolis, MN) with some finishing in Photoshop v 8.0™ (Adobe Systems Inc, San Jose, CA).

RESULTS

Quantification of oospores

The average number of oospores/g soil in the non-autoclaved control was 14.75 ± 0.59 and was greater than the autoclaved treatment, 7.73 ± 0.37. Kruskal-Wallis test results suggested there was a statistically significant difference between mean ranks in the autoclaved versus the non-autoclaved counts (p-value < 0.001) (Table 3.1). There were also statistically significant differences among the mean ranks in the four color categories (p-value < 0.001) while initial grouping of samples into two replicates per treatment showed no statistically significant differences in mean ranks among the replicates (p-value = 0.7860) (Table 3.1).

Oospore germination

Oospores were often observed with distinct nuclei and ooplast (Figure 3.1A and B) and ranged from 25μ to 50μ in diameter. Initial signs of oospore germination began approximately two weeks after preparation of the wet mounts. Germination occurred in the presence or absence of the oogonium wall and the germ tubes varied in size, branching patterns and thickness (Figure 3.1C, D, E and G). One oospore showed swelling at the tip of what appeared to be a germ tube with dense refractory elements forming with the oogonium wall intact (Figure 3.1C). After approximately four weeks of incubation, we observed a pair of fused oospores with two thick germ tubes which contained dense cytoplasmic material (Figure 3.1D). After approximately six weeks of incubation, the oospores appeared to have split open and expelled large circular ‘sporangium-like’ structures attached to a short, thick germ tube (Figure 3.1E and F). Once the oospores split open, we observed small ‘zoospore-like’ bodies that were approximately 15μ in diameter which began to germinate (Figure 3.1L and M).

DISCUSSION

Oospore extraction and quantification

Effective techniques were developed for the formation, extraction, recovery and germination of oospores of *P. humuli*. These techniques allowed for quick and prolific production in addition to precise soil-borne estimates. We learned an outstanding way to obtain prolific amounts of oospores quickly by inoculating hop cotyledons. The original technique reported by Jones (1932b) used 'Late Cluster' seedlings because that was one of the few cultivated varieties available in North America at that time in addition to it being very susceptible to hop downy mildew. After numerous experiments involving seeds derived from crosses with wild-American varieties, we learned that oospore production in hop cotyledons appeared to be a general cotyledon-based phenomenon and not restricted to the downy mildew susceptible Late Cluster (Unpublished data, APPENDIX B).

Mortar and pestle as tools to extract oospores from plant tissue proved to be an excellent method to finely grind plant tissue that resulted in the consistent exaction of thousands of oospores from the plant tissue. In addition, the further inclusion of a cellulase maceration step helped to release even more oospores from the plant tissue. Using a modified isolation technique based on those described in Van der Gaag and Frinking (1996), we isolated large amounts of oospores from *P. humuli* using a series of sieves. Using these techniques, we found no need for further isolation techniques as there was very little plant tissue left in suspension following the sieving procedures. These combined steps allow for the generation of large supplies of oospores from which further studies on hop downy mildew can be pursued.

The extraction, isolation and recovery of oospores from soil samples were also successful. Although 500g samples were large and it was very time-consuming process, the results showed that the samples were uniform as the ranks among sample counts were statistically similar (Table 3.1). It is unclear why there were differences between ranks for the autoclaved and non-autoclaved samples (Table 3.1). One

possible explanation may be the autoclaving procedure itself. Autoclaving oospores for two hours may have compromised the cellular integrity of the oospores. In doing so, the compromised oospores may have leaked their contents, lost buoyancy and settled out during centrifugation. Additional research is required to test this hypothesis.

MTT staining of oospores

The significant difference in mean ranks among the MTT color categories suggests *nothing*. Most scientists agree that black oospores are dead (Van der Gaag and Frinking 1997; Medina and Platt 1999; Spring and Zipper 2000). Because MTT has been reported as an enzymatically induced stain, theoretically, dead cells should not stain. It has been hypothesized that the formation of colored formazans in the autoclaved samples may be the result of dead oospores being parasitized by mycoparasites (Nelson and Olsen 1967; Van der Gaag and Frinking 1997). However, there may be another explanation.

Altman (1974 and 1976) indicated that MTT can undergo a chemical reduction in the presence of NADH. In addition, several other non-enzymatic compounds have been identified as having the ability to reduce MTT under elevated temperatures (Hamed 2004). If non-enzymatic, chemical reduction of tetrazolium salts takes place within the oospores, there would be no way to accurately differentiate between viable and dead cells. An incubation temperature of 37°C for 36h in an MTT solution may result in the chemical reduction of MTT thereby causing a false-positive, non-dehydrogenase-specific staining reaction. If Altman (1974 and 1976) was correct, MTT should be used with caution as a test for viability within oomycota or other living systems. Certainly, this potential for false-positive staining could result in the possibility that a number of oospores were alive prior to MTT staining and an unknown amount of the observed black-stained spores, or false-positives, were actually viable spores.

Oospore germination

Although germination of the hop downy mildew oospore was infrequent, there was visual evidence to suggest that germination had occurred on the *in vitro* wet mounts. The thick germ tubes which appeared to contain dense cytoplasm are indicative of germination (Figure 3.1D, E, F and G). In addition, the large 'sporangium-like' structures closely resemble those reported by Rooms of Diaz and Polanco (1984). The 'zoospore-like' bodies (Figure 3.1L and M) were larger than those reported for hop downy mildew (~15µ vs. 10µ for zoospores coming from asexual zoosporangia), these 'zoospore-like' bodies appeared to have originated from the oospore. With little documented evidence of oospore germination, it is unclear whether zoospores derived from oospore germination would be similar in size as those derived from asexual zoosporangia.

Chestnut-brown oospore

During oospore extraction, we discovered a chestnut-brown color variant in all of our soil samples. Reviewing oospore literature revealed this color variant had not been previously reported. The brown color was observed in all the soil-borne extraction samples, however, none of our cotyledon-borne oospore samples showed this distinctive color phase. Therefore, we hypothesize the chestnut-brown color may be a sign of dormancy. Dormancy may be a common survival technique in *P. humuli* because most of the observed oospores extracted from the soil were chestnut-brown.

Although some cotyledon-borne oospores appear to have some shade of brown within the cytoplasm and parts of the inner oogonium wall (Figure 3.1C, D, E and F), none were like the soil-borne oospores which showed a complete chestnut-brown staining throughout the entire cell wall similar to those in the MTT staining experiment (Figure 3.1J and K). In addition, observations of the MTT staining results suggest the stain was unable to enter the 'dormant' cells. Brown formazan (to our knowledge) has not been reported and the chestnut-brown oospores in the MTT samples appeared identical to dormant oospores observed during extraction. Although

this does not prove the brown color variants produced melanin, are dormant or if the MTT contributed to the brown stain, it does suggest the possibility that a high molecular weight compound may be formed in the cells which may prevent the uptake of MTT into the cells.

The onset of dormancy is a costly process for the cell and we suggest this process begins weeks, if not months after the oospores are initially produced in the plant tissue. This would help explain the chestnut-brown color variant observed in all the soil-borne samples while being absent in the cotyledon-borne samples. In addition, melanization would help explain why the MTT-stained samples with brown oospores appeared identical to our observations on unstained oospores. Although color is a very subjective analysis, the observed differences between cotyledon-borne and soil-borne oospores were quite distinct and it will be interesting to see if others discover this color variant within their systems in the future.

Conclusion

With the aid of these new extraction and isolation protocols, we studied the potential viability and germination capacity in order to learn more about the sexual spore of *P. humuli*. This study was initiated in part due to lack of published information concerning the hop downy mildew oospore. In the course of our investigation, we discovered previously reported information which may help explain the inconsistent results obtained by studying the potential viability of soil-borne hop downy mildew oospores.

The observations and hypotheses reported by Altman (1974 and 1976) suggesting that MTT can undergo both enzymatic and non-enzymatic reduction are appealing given our results. We hypothesize that a temperature induced chemical reduction of tetrazolium salts by NADH could potentially have resulted in some non-enzymatic reduction of MTT in the autoclaved samples. However, our studies were not designed to test this hypothesis and simply suggest that chemical reduction might have occurred. Undoubtedly, scientists should reconsider the use of MTT, or the

methodology of use, as a viability stain in living systems. We propose that the oospore of *P. humuli* has the ability germinate *in vitro*, which is in agreement with (Arens 1929; Bressman and Nichols 1933). It is not clear whether oospore germination has a significant role in the epidemiology of hop downy mildew. Our observations suggest that *P. humuli* oospores germinate and therefore, may play a role in the epidemiology of this disease.

There continues to be disagreement as to whether hop downy mildew over-winters as mycelium in dormant hop crowns or as oospores. Numerous macro-and microscopic observations (Salmon and Ware 1925; Ware 1926; Ware 1929; Coley-Smith 1960; Skotland 1961) led to the conclusion that hop downy mildew over-winters in the dormant buds and tissue within the crown of the hop plant. However, it doesn't appear that these investigations were performed on infected crowns harvested during winter or tests run to determine if observed mycelium were viable.

However, Arens (1929) made detailed observations of the mycelium within the hop crowns and strap cutting and he reported that the mycelium were dead due to the presence of shriveled fungal nuclei and the absence of staining with haematoxylin; thus suggesting that mycelium were not viable and could not serve as over-wintering structures. Further support for oospores serving as one of the over-wintering structures of *P. humuli* comes from inoculation studies of Coley-Smith (1960). When healthy hop plants were infected with oospore inoculum, 7% developed basal spikes (Coley-Smith 1960). Also, as mentioned earlier, oospores have been found to reside in the pith of dormant buds (Ware 1929). These data appear to suggest that oospores may play an important role in the over-wintering of *P. humuli* and the development of basal spikes in the spring. Based on the previous work by Arens (1929) and Bressman and Nichols (1933) in conjunction with our own observations, we agree with the hypothesis that *P. humuli* oospores have the potential to over-winter inside the dormant buds and contribute to infection during the breaking of dormancy. Once dormancy in hop is broken in the spring, the oospores germinate resulting in the initiation of the disease symptoms.

Preliminary molecular data (Chee et al. 2006) suggests that there is a large degree of genetic heterogeneity among the *P. humuli* population in Oregon and to a lesser extent in Washington. At the same time, Neve (1991) reports that plant resistance to *P. humuli* is under polygenic control and that no major gene action has been observed. These two observations suggest that sexual recombination is occurring in the field and that oospores may play a more important role in the epidemiology of hop downy mildew than previously thought. It is clear that further investigation of the role of oospores in the epidemiology of hop downy mildew is warranted in order to determine the role of oospores in over-wintering and disease development and what effect this has on plant breeding efforts.

REFERENCES

Altman F P (1974) Studies on the reduction of tetrazolium salts. III. The products of chemical and enzymatic reduction. Histochemie 38: 155-171

Altman F P (1976) Tetrazolium salts and formazans. Progr Histochem Cytochem 9: 1-55

Arens K (1929) Untersuchungen über *Pseudoperonospora humuli* (Miyabe u. Takah.), den Erreger der neuen Hopfenkrankheit. Phyto Zeitschr 1: 169-193

Binder W D, Mitchell G M and Ballantyne D J (1974) Pollen viability testing, storage and related physiology: review of literature with emphasis on gymnosperm pollen. Canadian Forestry Service, Victoria, BC, Canada

Bowers J H, Papavizas G C and Johnston S A (1990) Effect of soil temperature and soil-matrix potential on the survival of *Phytophthora capsici* in natural soil. Plant Dis 74: 771-777

Bressman E M and Nichols A A (1933) Germination of the oospores of *Pseudoperonospora humuli*. Phytopath 23: 485-487

Chee H Y and Klein R E (1998) Laboratory production of oospores in *Pseudoperonospora humuli*. Korean J Plant Pathol 14: 618-621

Chee H Y, Nelson M E, Grove G G, Eastwell K C, Kenny S T and Klein R E (2006) Population biology of *Pseudoperonospora humuli* in Oregon and Washington. Plant Dis 90: 1283-1286

Cohen S D (1984) Detection of mycelium and oospores of *Phytophthora megasperma* f. sp. *glycinea* by vital stains in soil. Mycologia 76: 34-39

Coley-Smith J R (1960) Overwintering of hop downy mildew *Pseudoperonospora humuli* (Miy. and Tak.) Wilson. Rep Dep Hop Res Wye College for 1959: 107-114

Davis J J (1910) A new hops mildew. Science 31: 752

El-Hamalawi Z A and Erwin D C (1986) Physical, enzymatic and chemical factors affecting viability and germination of oospores of *Phytophthora megasperma* f. sp. *medicaginis*. Phytopath 76: 503-507

Gahan P B and Kalina M (1968) The use of tetrazolium salts in the histochemical demonstration of succinic dehydrogenase activity in plant tissues. Histochemie 14: 81-88

Hamed S H (2004) Efficacy and mechanism of action of a new tyrosinase inhibitory agent. PhD Diss, Univ of Cincinnati, OH. http://www.ohiolink.edu/etd/send-pdf.cgi?ucin1085566152 (verified 07/23/2007)

Haunold A and Zimmermann C E (1974) Pollen collection, crossing and seed germination of hop. Crop Science 14: 774-776

Hoerner G R (1949) Hop diseases in the United States. The Brewers Digest 24: 45-51

Jiang J and Erwin D C (1990) Morphology, plasmolysis and tetrazolium bromide stain as a criteria for determining viability of *Phytophthora* oospores. Mycologia 82: 107-113

Jones W (1932b) A new technique for obtaining oospores of hop downy mildew by inoculating cotyledons. Science 75: 108

MacKay D B (1972) The measurements of viability. In Roberts E H (ed) Viability of seeds. Chapman and Hall, London, United Kingdom, pp. 172-208

Medina M V and Platt H W (1999) Viability of oospores of *Phytophthora infestans* under field conditions in northeastern North America. Can J Plant Path 21: 137-143

Meier R and Charvat I (1993) Reassessment of tetrazolium bromide as a viability stain for spores of vesicular-arbuscular mycorrhizal fungi. Am J of Bot 80: 1007-1015

Nelson G A and Olsen O A (1967) Staining reactions of resting sporangia of *Synchytrium endobioticum* with a tetrazolium compound. Phytopath 57: 965-968

Neve R A (1991) Hops. Chapman and Hall, New York

Rooms of Diaz G and Polanco C D (1984) Oospores of the parasitic fungus *Peronosclerospora sorghi* under conditions of flood. Tropical Agronomy 34: 87-94

Royle D J and Kremheller H Th (1981) Downy mildew of hop. In: Spencer D M (ed) The Downy Mildews. Academic Press, London, New York, San Francisco, pp. 395-419

Salmon E S and Ware E M (1925) The downy mildew of hop and its epidemic occurance in 1924. Ann Biol 12: 121-151

Skotland C B (1961) Infection of hop crowns and roots by *Pseudoperonospora humuli* and its relation to crown and root rot and overwintering of the pathogen. Phytopath 51: 241-244

Skotland C B and Johnson D A (1983) Control of downy mildew of hops. Plant Dis 67: 1183-1186

Spring O and Zipper R (2000) Isolation of oospores of sunflower downy mildew, *Plasmopara halstedii*, and microscopic studies of oospore germination. J Phytopath 148: 227-231

Sutherland D E and Cohen S D (1983) Evaluation of tetrazolium bromide as a vital stain for fungal oospores. Phytopath 73: 1532-1535

Van der Gaag D J (1994) The effect of pH on staining of oospores of *Peronospora viciae* with tetrazolium bromide. Mycologia 86: 454-457

Van der Gaag D J and Frinking H D (1996) Extraction from plant tissue and germination of oospores of *Peronospora viciae* f.sp. *pisi*. J Phytopath 144: 57-62

Van der Gaag D J and Frinking H D (1997) Extraction of oospores of *Peronospora viciae* from soil. Plant Path 46: 675-679

Van Waes J M and Debergh P C (1986) Adaptation of the tetrazolium method for testing the seed viability, and scanning electron microscopy of some Western European orchids. Physiol Plantar 66: 435-442

Ware W M (1926) *Pseudoperonospora humuli* and its mycelial invasion of the host plant. Trans Brit Mycol Soc 11: 91-107

Ware W M (1929) Experiments on the production of diseased shoots by the hop downy mildew, *Pseudoperonospora humuli* (Miy. et Takah.), Wils. Ann of Bot 43: 683-711

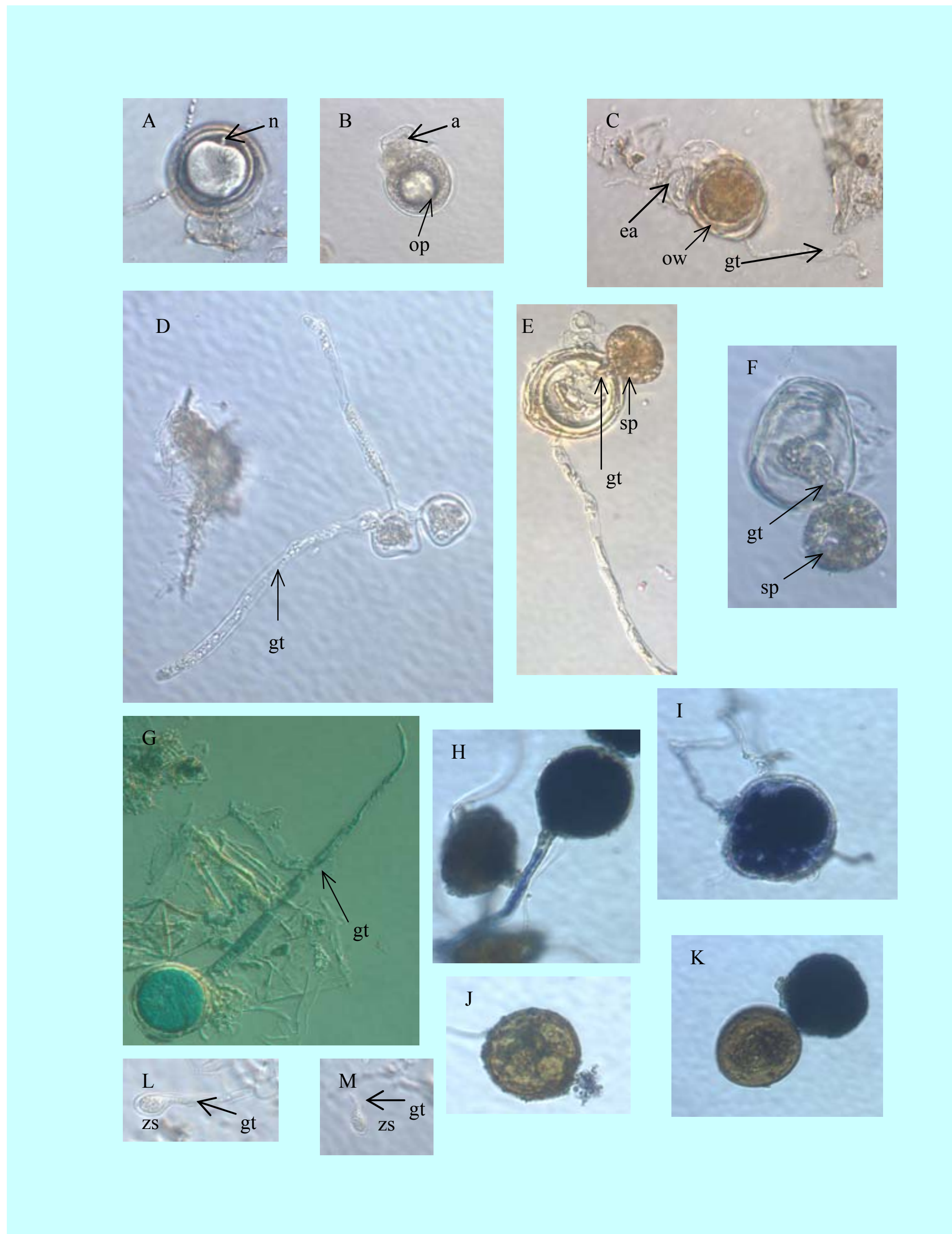
Table 3.1. Kruskal-Wallis test for soil-borne oospore data. Spore counts were analyzed to test for similarity between mean rank scores and were performed according to treatment, color category and sample number with respective *p*-values.

	N	Sum of Scores	Expected Under HO	Std. Dev. Under HO	Mean Score	Chi-Sq.	DF	Pr >Chi Sq.
Treatment								
Control	64	5112.50	4128.0	209.37	79.88	22.11	1	<0.0001*
Auto	64	3143.50	4128.0	209.37	49.12			
Color								
Black	32	3261.0	2064.0	181.32	101.91	57.98	3	<0.0001*
Chestnut	32	2305.0	2064.0	181.32	72.03			
Violet	32	1253.0	2064.0	181.32	39.16			
Clear	32	1437.0	2064.0	181.32	44.91			
Sample								
1	32	2002.0	2064.0	181.32	62.56	1.06	3	0.7860
2	32	2173.0	2064.0	181.32	67.91			
3	32	2162.5	2064.0	181.32	67.57			
4	32	1918.5	2064.0	181.32	59.95			

*Significant at the 0.0001 level

Figure 3.1. Oospores and zoospores in the process of germination. The figure shows both unstained and MTT stained oospores using the following key: A-G) Cotyledon-borne oospores, H-K) Soil-borne oospores, L-M) 'Zoospore-like' bodies from cotyledon-borne oospores, A) Oospore with well formed nuclei (n), B) Oospore with distinct ooplast (op) and antheridia (a), C) Oospore germinating with presence of oogonium wall (ow), empty antheridia (ea) and a germ tube (gt), D) Two fused oospores with two germ tubes (gt), E) Oospore cracked open and expelling a 'sporangium-like' (sp) body on top of a short, thick germ tube (gt), F) Second oospore expelling a 'sporangium-like' (sp) body on top of a thick germ tube (gt), G) MTT-stained oospore showing blue formazan production and a well defined germ tube (gt), H) Black formazan production, I) Uneven black/blue formazan production within an oospore, J-K) Chestnut-brown oospores, L-M) 'Zoospore-like' (zs) bodies which are in the process of germination.

Figure 3.1. Oospores and zoospores in the process of germination.



Statistical Modeling of Epistasis and Linkage Decay using Logic Regression

T. B. Parker[1], P. Szűcs[1,2], W. F. Mahaffee[4] and J. A. Henning[3]*

---

[1]Department of Crop and Soil Science, Oregon State University, Corvallis, OR, 97331, USA.

[2]Agricultural Research Institute of the Hungarian Academy of Sciences, H-2462, Martonvásár, Hungary.

[3]USDA-ARS National Forage Seed Processing Research Center, Corvallis, OR, 97331, USA.

[4]USDA-ARS Horticultural Crops Research Laboratory, Corvallis, OR, 97331, USA.

*Corresponding author: E-mail: John.Henning@oregonstate.edu.

ABSTRACT

Epistatic interactions are an important but often overlooked component of genetic models. Traditional linear methods in QTL analysis identify model epistasis and potentially result in the supposition that all observed genetic interactions are due to additive genetic effects. Logic regression has been identified as a tool that can identify both additive interactions as well as complex, non-additive genetic interactions. We propose that logic regression is an alternative and robust method for identification and classification of genetic models involving non-additive epistatic interaction. Logic regression, TASSEL-GLM and SAS-GLM were compared for analytical precision using a previously characterized model system to identify the best genetic model explaining epistatic interaction for vernalization-sensitivity in barley. We also analyzed logic regression's precision at identifying models in the presence of partially-linked randomly generated markers. Logic regression identified the correct epistatic model containing the two molecular markers identified in vernalization response in barley. TASSEL-GLM and SAS-GLM both identified *VRN-H1* and *VRN-H2* as being associated with the days to flowering phenotype, however, they also identified random generated markers as being associated with the phenotype. When the significant markers were analyzed in full models, a spurious interaction was falsely identified. Linkage decay results demonstrated that randomly generated, less-tightly linked markers were correctly identified down to 50% linkage. Based on these results, logic regression may be a useful tool for identification and modeling non-additive forms of epistasis.

INTRODUCTION

Epistatic interactions have been long thought to play a vital role in the evolutionary diversification of species (Wright 1931). It is currently believed that epistasis plays a central role in the maintenance of genetic variation (Weinig et al. 2003) and in the evolution of species (Orr 1995). Recently, researchers identified epistatic interactions in the *FRI* and *FLC* flowering time genes in *Arabidopsis*. These genes are thought to play a role in the generation of a latitude cline in the species (Caicedo et al. 2004). In addition, quantitative trait loci (QTL) analysis identified epistatic interactions that resulted in natural phenotypic variation in *Arabidopsis* (Ungerer et al. 2002; Weinig et al. 2003), *Drosophila* (Dilda and Mackay 2002) and *Caenorhabditis elegans* (Shook and Johnson 1999).

Quantitative trait loci present a greater challenge in identification and mapping than simple Mendelian traits. In the simplest form, QTL identification is performed by individual associations identified between a molecular marker and a phenotype by linear regression analysis (Kearsey and Hyne 1994; Hyne and Kearsey 1995). When a particular marker is associated with a statistically significant phenotypic mean, it is reasonable to conclude that there is a QTL for that trait tightly linked with that marker. In situations where marker density is low, associations between molecular markers and a QTL can be identified using simple linear regression or least squares (Darvasi et al. 1993). However, these tests are limited to identifying QTLs with reasonably tight linkage to the markers (van Ooijen 1992) and epistatic interactions between markers cannot be identified.

Linkage analysis involving interval mapping and composite interval mapping has been used to overcome some of the shortcomings of simple linear regression by providing greater statistical power. Although modern linkage analysis can identify statistically significant QTL between two flanking markers and account for the effects of additional QTL at other loci, there are some drawbacks to this approach. Linkage analysis is time-consuming, expensive and the information gained may be of limited

use as only one cross from a population is made to form a recombinant population. Therefore, extrapolations to other individuals and populations may be spurious (Verhoeven et al. 2006).

Recent advances in adaptive regression methodology have been developed to explore high-order interactions in genomic data (Kooperberg et al. 2001; Ruczinski et al. 2003; Ruczinski et al. 2004). One such technique, logic regression, utilizes a simulated annealing algorithm to identify statistical models for binary data sets. Logic regression constructs models consisting of Boolean combinations of binary covariates (Ruczinski et al. 2003). With $X_1 \ldots X_k$ as binary predictors and Y as the response, logic regression will fit regression models in the form $g(E[Y]) = \beta_0 + \sum_{j=1}^{t} \beta_j L_j$, where L_j is a Boolean expression of the predictors X_i (Ruczinski et al. 2003). These are collectively called logic models. In evaluating models of varying sizes, logic regression looks for signal vs. noise in the data set. In statistical modeling, signal is identified by asking whether the slope (b) is equal to zero or not equal to zero. *Signal* is when X is associated with Y. When additional covariates not associated with Y are added to the model, this is considered *noise*. By evaluating models of various sizes for signal vs. noise, researchers can determine the level of over-fitting (noise) in each *model-size* class. In addition, potentially troublesome data sets where there are unacceptable levels of noise are quickly identified so that no further time is wasted in the analysis of these problematic data sets.

Logic regression has many potential benefits when compared to traditional QTL linkage mapping. This methodology does not require data obtained from time-consuming and expensive recombinant inbred line progeny although one can use data from such populations. Logic regression offers numerous scoring functions for linear regression (residual sums of square), logistic regression (deviance), classification (misclassification) and proportional harzards models (partial likelihood). In addition, the software allows for inclusion of binary or non-binary additive predictors in the model. Furthermore, by creating statistical models consisting of Boolean

combinations of binary covariates, this statistical methodology shows promise in the identification of dominant forms of epistatic interactions between molecular markers.

The epistatic interaction of alleles at the *VRN-H1*, *VRN-H2* and *VRN-H3* locus is the hypothesized determinant for vernalization-sensitivity in cultivated barley (*Hordeum vulgare* subsp. *vulgare*) (Takahashi and Yasuda 1971). There is no allelic variation at *VRN-H3* in most cultivated barley genotypes, reducing the genetic model to a two-locus epistatic model (Takahashi and Yasuda 1971). *VRN-H2* encodes a dominant flowering repressor (ZCCT-H) down-regulated by vernalization (Yan et al. 2004). *VRN-H1* is a MADS-box floral meristem identity gene (*HvBM5A*) (Danyluk et al. 2003; Yan et al. 2003) and large deletions within the first intron result in a dominant *VRN-H1* allele and spring growth habit (Fu et al. 2005; von Zitzewitz et al. 2005). A molecular model has been recently proposed to explain the *VRN-H2/VRN-H1* epistatic interaction where dominant *VRN-H2* inhibits the expression of recessive *VRN-H1* alleles (Yan et al. 2004). Based on this model, genotypes with *VRN-H2_/vrn-H1vrn-H1/vrn-H3vrn-H3* allelic architecture flower late in the absence of vernalization (vernalization-sensitive) and all other allelic configurations lead to a lack of significant vernalization-sensitivity. This well validated epistatic interaction (reviewed in Szűcs et al. 2007) was used as a model system to test the ability of logic regression in identifying epistasis in binary molecular data.

The objective of this work was to determine whether logic regression can be used to identify the interaction between molecular markers associated with the days to flowering phenotype in barley with little or no spurious associations and to compare logic regression with traditional modeling techniques. In addition, we wanted to determine logic regression's capabilities at identifying spurious associations using a linkage decay series.

MATERIALS AND METHODS

Plant material, phenotype and data set

'Dicktoo' (*vrn-H2vrn-H2*/*vrn-H1vrn-H1*), 'Calicuchima'-sib (*Vrn-H2Vrn-H2*/*Vrn-H1Vrn-H1*) and the 'Oregon Wolf Barley Dominant' genetic stock (hereafter referred to as 'OWB-D') (*Vrn-H2Vrn-H2*/*Vrn-H1Vrn-H1*) are vernalization-insensitive barley genotypes (Szűcs et al. 2007). 'Dicktoo' was crossed with 'Calicuchima'-sib and 'OWB-D' and two F_2 populations were established (Szűcs et al. 2007). Flowering time was measured for all unvernalized F_2 plants grown under long-day greenhouse conditions with supplemental lighting and constant temperature according to Szűcs et al. (2007). Previously reported gene-specific primers were used to assign *VRN-H2* and *VRN-H1* allele-types for each F_2 individual (Szűcs et al. 2007). We sequenced the recently cloned *VRN-H3* gene (Yan et al. 2006) from the three parents and confirmed that 'Calicuchima'-sib (EU007825), 'Dicktoo' (EU007827), and OWB-D (EU007829) have the recessive allele.

The *VRN-H1* and *VRN-H2* molecular markers were coded as binary. The heterozygotes were bulked with the homozygous dominants and scored as 1 while the homozygous recessives were scored as 0. In addition to the actual molecular markers, *VRN-H1* and *VRN-H2,* we created 100 randomized binary markers (simulated data) for a total of 102 binary markers.

Logic regression analysis

To test the null hypothesis that logic regression cannot identify the epistatic interaction in barley, the datasets from the two F_2 populations were modeled with Logic Regression© using the linear regression scoring function (Kooperberg and Ruczinski 2005). Days to flowering was used as the continuous response variable and the molecular marker data were used as binary predictors. Initially, logic regression was allowed to choose the high and low temperatures for the simulated annealing algorithm using a single-fit selection with one tree. Once the program chose the

annealing algorithm parameters, the high and low temperatures were optimized according to the author's instructions (Kooperberg and Ruczinski 2005) for selection of a single-fit model. After analyzing the single-fit model data, multiple-fit model selection was performed for use in model selection. When the results warranted further investigation, we performed null model tests to test for statistical signal vs. noise in the data. Upon verification of a strong statistical signal with little noise, we ran a cross-validation test to identify the logic trees with the best predictive capability. In the final step, permutation tests were run to confirm the results of the search algorithm so that we could positively identify the best model that describes the association between predictors and response.

TASSEL analysis

The two F_2 datasets were analyzed using the association mapping software TASSEL©-GLM (Trait analysis by association evolution and linkage) (Zhang et al. 2006). The binary coded two vernalization markers and the 100 randomly generated markers were imported into TASSEL along with the phenotypic matrix. A population structure matrix called the Q-matrix was designed to suggest a single population for our data. The general linear model function was selected for analysis.

SAS-GLM

Analysis of variance was performed on both F_2 datasets using the general linear model (GLM) of SAS Version 9.1© (SAS Institute, Cary, NC). The individual markers which were identified as being significantly associated with the phenotype in TASSEL were analyzed in SAS using a type III fixed effects model analysis to confirm the single marker association results in TASSEL. A type III fixed effects full model containing all the significantly associated markers was performed in SAS to identify marker interactions.

Linkage decay data set

Spurious associations between trait and randomly generated markers were tested with logic regression using a linkage decay series to determine the point at which logic regression could no longer make valid associations between truly linked markers and random noise. Two randomly generated sets of linkage decay markers were created each set based upon one of the F_2 populations in our study. Both sets of linkage decay markers had 90%, 80%, 70%... 0% similarity to *VRN-H1*. Our goal was to create randomly generated markers that would decay in a predictable pattern as the signal in the data became progressively weaker as the similarity to the original vernalization marker decreased (Figure 4.1). The decay series data was created by randomly changing 10% of the 1's to zeros and using this *new* linkage decay marker as the basis for creating the next marker in the decay series. Original *VRN-H1* and *VRN-H2* markers were removed from the analysis as they interfered with the analysis of the decay series due to their strength of association with the phenotype. This procedure created a linkage decay series where the model association became progressively weaker as the linkage to the phenotype decayed resulting in a smooth logarithmic response (Figure 4.1).

RESULTS

VRN-H1/VRN-H2 model selection

Logic regression correctly identified the genetic model explaining the epistatic interaction of the vernalization alleles in both data sets. The search resulted in a model with a score (residual sums of square) of 12.47 and the equation [+74.9 * (*VRN-H2* and (not *VRN-H1*))] for the in the 'Dicktoo' x 'Calicuchima'-sib data and a model score 8.825 and the *conjugate* equation [-85.4 * (*VRNH1* or (not *VRNH2*))] for the 'Dicktoo' x 'OWB-D' data. The single-fit models were repeated 100 times and it was found that the scores (Figure 4.1) and the coefficients of the selected models never changed. In addition, the null model tests suggested that there was a strong signal in the data with very little noise because 0% of the model scores were better than the best score (Table 4.1A and B).

The cross validation and the 1000 randomization permutation tests on the multiple-fit model analyses confirmed the results of the single-fit model search. Cross-validation tests indicated the optimum model to be model two with one tree and two leaves as it had the lowest cross-validation test average (Table 4.2A and B). The permutation tests identified the same model with one tree and two leaves as being the optimum sized and correct model for the data set as that was the point where the mean of the randomization scores stopped decreasing as the model size increased (Table 4.3A and B).

TASSEL analysis

TASSEL-GLM results showed *VRN-H1* and *VRN-H2* as being associated with the days to flowering phenotype in the 'Dicktoo' x 'Calicuchima'-sib and 'Dicktoo' x 'OWB-D' data (Table 4.4). TASSEL-GLM also identified the randomly generated marker RANDOM 70 as being associated with the days to flowering phenotype in the 'Dicktoo' x 'Calicuchima'-sib data and randomly generated markers RANDOM 46

and RANDOM 58 as being associated with the phenotype in the 'Dicktoo' x 'OWB-D' data (Table 4.4).

General linear model analysis of variance

The type III fixed effects full model for the 'Dicktoo' x 'Calicuchima'-sib data revealed a significant interaction between *VRN-H1* and *VRN-H2*, but there were no significant singular effects or interactions with the randomly generated marker RANDOM 70 (Table 4.5). The type III fixed effects full model for the 'Dicktoo' x 'OWB-D' data revealed a significant interaction between *VRN-H1* and *VRN-H2*, but there were no significant singular effects with either marker RANDOM 46 or marker RANDOM 58 (Table 4.6). GLM identified a spurious interaction between *VRN-H1* and RANDOM 58 (Table 4.6).

Linkage decay

The linkage decay results for the two data sets showed that they were quite different in how they responded in a controlled decay simulation. The 'Dicktoo' x 'Calicuchima'-sib data showed less overall variation in single-fit model scores when compared with the 'Dicktoo' x 'OWB-D' data (Figures 4.1 and 4.2). Closer examination of the 'Dicktoo' x 'Calicuchima'-sib data revealed a large increase in variation (CV) within the single-fit model selection scores when linkage decay reached 40% similar to *VRN-H1* (Figure 4.2), which corresponded where logic regression could no longer distinguish between linkage decay markers and the simulated markers (APPENDIX C). Also, there were large variations in the single-fit model scores for *VRN-H2* over multiple runs which resulted in extremely large CVs at the initiation of the decay series (Figure 4.2).

The 'Dicktoo' x 'OWB-D' data showed more overall variation in single-fit model scores, when compared with the 'Dicktoo' x 'Calicuchima'-sib data (Figure 4.1). A stable single-fit regression model was *only* identified when both the markers appeared in the data set (Figure 4.1). Furthermore, large increases in the CV were

observed when *VRN-H1* was modeled in the linkage decay series (Figure 4.2). It is interesting to note that none of the individual markers in the decay series for this cross had a CV of less than 6% (Figure 4.2). This is in direct contrast with the 'Dicktoo' x 'Calicuchima'-sib data where *all* of the individual markers up to 40% similar to *VRN-H1* (the point where logic regression could no longer distinguish between decay and dummy markers) had CVs of less than 6% (Figure 4.2).

DISCUSSION

Logic regression correctly identified the previously published epistatic interaction of *VRN-H1* and *VRN-H2* (Szűcs et al. 2007) in both the 'Dicktoo' x 'Calicuchima'-sib and 'Dicktoo' x 'OWB-D' F_2 data. Logic regression indicated there was an epistatic interaction between the two alleles and this interaction was best explained by a dominant/recessive epistatic model. Vernalization in barley is hypothesized to be an interaction where dominant *VRN-H2* inhibits the expression of recessive *VRN-H1* alleles (Yan et al. 2004). Because a dominant/suppression form of epistasis has been hypothesized to govern the vernalization response in barley, we suggest that logic regression correctly identified this proposed genetic model [+74.9 * (*VRN-H2* and (not *VRN-H1*))]. Numerous QTL and genic studies have identified multiple loci involved in the expression of a single trait (Carlborg and Haley 2004). However, in all these cases, additional studies were required to ascertain the actual genetic model defining the interaction of the various loci. The use of logic regression appears to address both problems simultaneously.

Interval mapping and composite interval mapping have been used successfully to identify QTL associated with specific phenotypes that led to the identification of statistically significant interactions, or epistasis among QTL (Lefebvre et al. 2003; Ma et al. 2006; Shook and Johnson 1999). However, one can argue that linear modeling testing for significant interactions between QTL appears to make assumptions that *all* genetic marker interactions are the result of an interaction of additive genetic effects. For example, when the 'Dicktoo' x 'Calicuchima'-sib *VRN-H1/VRN-H2* model data set was run in TASSEL-GLM and SAS-GLM, both programs identified *VRN-H1* and *VRN-H2* as being associated with days to flowering with p-values < 0.001 (Table 4.4). Therefore, analysis using general linear modeling resulted in the following statistical model: full linear model [days to flowering ~ *VRN-H1* + *VRN-H2* + (*VRN-H1**VRN-H2*)] vs. the reduced or additive model (days to flowering ~ *VRN-H1* + *VRN-H2*). An F-test for the interaction between *VRN-H1* and *VRN-H2* determined there was a

statistically significant interaction. If there was no interaction, the full model would be rejected in favor of the *additive* model. Erroneous modeling resulted from the use of both TASSEL-GLM and SAS-GLM, regardless of the model chosen (Tables 4.4 and 4.6).

Our results suggest logic regression better identified and modeled dominant/suppression epistasis when compared with traditional general linear modeling. Unlike traditional linear modeling, logic regression creates Boolean expressions in the form of logic models. In the formation of logic groups, there are no full and reduced models and no assumption of interaction of additive effects. In addition, because logic regression utilizes logic operators and is not limited in model assumptions, the best model identified by logic regression should theoretically have greater precision in matching the true genetic model in unknown interactions than traditional modeling analyses.

Although TASSEL-GLM and SAS-GLM both identified *VRN-H1* and *VRN-H2* as being associated with the days to flowering phenotype, both programs also identified random markers as being associated when run with default settings (Tables 4.4 and 4.6). It wasn't until we ran a type III fixed effects full model with all three associated markers that RANDOM 70 was rejected from association with the phenotype in the 'Dicktoo' x 'Calicuchima'-sib data (Table 4.5). However, even when a type III fixed effects full model was run on the 'Dicktoo' x 'OWB-D' data, a spurious interaction remained (Table 4.6). These results suggest that a GLM single marker QTL analysis using the default settings for association to be less precise than logic regression. One possible explanation for the difference between the two modeling approaches may be how the models are generated. Logic regression creates Boolean logic groups prior to performing regression analysis. Because the logic groups are formed prior to regression, individual associations are ignored, in favor of the *most likely* combination of binary predictors which best explain the association to the response. Only when combinations of binary predictors are determined as insignificant does logic regression focus on independent predictors for association

with the phenotype. In GLM, the individual markers are identified as being in association with the phenotypic response prior to identification of possible interaction.

Modeling linkage decay helped demonstrate the power of logic regression to accurately model data sets where linkage between markers may be incomplete or spurious. Although the value and standard deviation in single-fit model scores increased as the decay series progressed (Figure 4.1), these values only show the smooth progression within the decay series itself. The coefficient of variation (CV) is a dimensionless value used to quantify uncontrolled experimental error (Patel et al. 2001). The coefficient of variation results suggest there were measurable differences between the two crosses, however, this data alone does not provide any diagnostic information about acceptable levels of variation within the data sets.

When the CV data was compared with TASSEL-GLM (Table 4.4) and SAS-GLM (Table 4.6) output, we discovered that the CV may be diagnostic in the identification of potentially troublesome data sets. As the decay series proceeded further from actual data (Figure 4.1), there was an increase in noise in the 'Dicktoo' x 'Calicuchima'-sib data represented by a *jump* in the CV at 40% similar to *VRN-H1* (Figure 4.2). This jump in CV within the decay series suggests there may be a limit in predictive capability. Our results suggest the *limit of predictability* threshold may be where there this substantial increase in CV was observed (Figure 4.2). Comparing CV values with the results from the SAS-GLM suggest CVs above 6% may result in the modeling of spurious associations. This was the precise point in the decay series where logic regression could no longer differentiate between the decay marker and randomly generated markers (APPENDIX C).

SAS-GLM identified a spurious interaction *before* running the linkage decay series for the 'Dicktoo' x 'OWB-D' data (Table 4.6). Therefore, the 'Dicktoo' x 'OWB-D' data may be problematic right from the start due to lack of signal. The ANOVA suggest a lack of signal and the CV analysis on the single-fit model scores suggest variation above 6% may lead to spurious association (Figure 4.2 and APPENDIX C). In support of the hypothesized threshold, Ruczinski et al. (2003)

reported when there are large variations in single-fit model scores during initial model identification, there may be problems with the data set. Unfortunately, it is unclear where that cutoff might be. This was a concern for us and it became one of the major reasons for performing the linkage decay series. Based on our results, we suggest any data set that has a single-fit model selection CV of 6% or less should prove a very good data set and identify real associations.

Although logic regression has many strong points, there are some limitations inherent in the program. First, the program handles binary data for markers. This means interactions have some dominant and/or recessive interaction. However, if the goal is to correctly identify and model all forms of epistasis and do so simultaneously, logic regression appears to be a very robust statistical process. Another limitation is data analysis for multiple model sizes requires extensive computing capabilities in order to run through the required thousands of permutations. Users should be cautioned to perform logic regression on a cluster computer or higher-end mainframe computer having the capability to perform simultaneous permutations rather than use of a desktop computer. Despite these limitations, logic regression should prove itself useful anywhere logic models are needed to identify complex genetic interaction.

Our results suggests logic regression works in accurate identification of epistatic interaction and that the model building algorithm appears to be more robust and accurate when compared with traditional general linear modeling in QTL analysis. From a theoretical point of view, logic regression may use the more appropriate approach for modeling epistasis by forming logic groups prior to running the regression analysis. More work needs to be done using other documented examples of epistasis before definitive statements on the usefulness of the program be made. However, it does appear that logic regression is a useful tool in data mining applications and provides researchers with a complement to traditional QTL identification.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Patrick M. Hayes for his critical suggestions on the data analysis and manuscript preparation.

REFERENCES

Caicedo A L, Stinchcombe J R, Olsen K M, Schmitt J and Purugganan M D (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitude cline in a life history trait. Proc Natl Acad Sci USA 101: 15670-15675

Carlborg Ö and Haley C S (2004) Epistasis: too often neglected in complex trait studies. Nature 5: 618-625

Danyluk J, Kane N D, Breton G, Limin A E, Fowler D B and Sarhan F (2003) *TaVRT-1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol 132: 1849–1860

Darvasi A, Weinreb A, Minke V, Weller J I and Soller M (1993) Detecting marker QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943-951

Dilda C L and Mackay T F C (2002) The genetic architecture of *Drosophila* sensory bristle number. Genetics 162: 1655-1674

Fu D, Szűcs P, Yan L, Helguera M, Skinner J S, von Zitzewitz J, Hayes P M and Dubcovsky J (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. Mol Genet Genom 273: 54–65

Hyne V and Kearsey M J (1995) QTL analysis: further uses of marker regression. Theor Appl Genet 91: 471-476

Kearsey M J and Hyne V (1994) QTL analysis: a simple 'marker-regression' approach. Theor Appl Genet 89: 698-702

Kooperberg C and Ruczinski I (2005) The logic regression package. In: Contributed Packages. R Project. . http://cran.r project.org/src/contrib/PACKAGES.html

Kooperberg C, Ruczinski I, LeBlanc M L and Hsu L (2001) Sequence analysis using logic regression. Genetic Epidemiology 21 (Suppl. 1): S626-S631

Lefebvre V, Daubèze A–M, Rouppe van der Voort J, Peleman J, Bardin M and Palloix A (2003) QTLs for resistance to powdery mildew in pepper under natural and artificial infections. Theor Appl Genet 107: 661-666

Ma H–X, Bai G–H, Zhang X and Lu W–Z (2006) Main effects, epistasis, and environmental interactions of quantitative trait loci for Fusarium head blight resistance in a recombinant inbred population. Phytopath 96: 534-541

Orr H A (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139: 1805-1813

Patel J K, Patel N M and Shiyani R L (2001) Coefficient of variation in field experiments and yardstick thereof-An empirical study. Current Science 81: 1163-1164

Ruczinski I, Kooperberg C and LeBlanc M (2003) Logic regression. J Comput Graph Stat 12: 475–511

Ruczinski I, Kooperberg C and LeBlanc M L (2004) Exploring interactions in high-dimensional genomic data: an overview of logic regression with applications. J Multi Analysis 90: 178-195

Shook D R and Johnson T E (1999) Quantitative trait loci affecting survival and fertility related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. Genetics 153: 1233-1243

Szűcs P, Skinner J S, Karsai I, Cuesta-Marcos A, Haggard K G, Corey A E, Chen T H H and Hayes P M (2007) Validation of the *VRN-H2/VRN-H1* epistatic model in barley reveals that intron length variation in *VRN-H1* may account for the continuum of vernalization sensitivity. Mol Genet Genom 277: 249-261

Takahashi R and Yasuda S (1971) Genetics of earliness and growth habit in barley. In: Nilan R A (ed) Barley Genetics II. Proceedings of the Second International Barley Genetics Symposium. Washington State University Press Pullman, pp. 388–408

Ungerer M C, Halldorsdottir S S, Modliszewski J L, Mackay T F C and Purugganan M D (2002) Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. Genetics 160: 1133-1151

van Ooijen J W (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84: 803-811

Verhoeven K J F, Jannink J-L and McIntyre L M (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96: 139-149

von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Pecchioni N, Francia E., Casas A, Chen T H H, Hayes P M and Skinner J S (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59: 449–467

Weinig C, Dorn L A, Kane N C, German Z M, Halldorsdottir S S, Ungerer M, ToyonagaY, Mackay T F C, Purugganan M D and Schmitt J (2003) Heterogenous selection at specific loci in natural environments in *Arabidopsis thaliana*. Genetics 165: 321-329

Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97-159

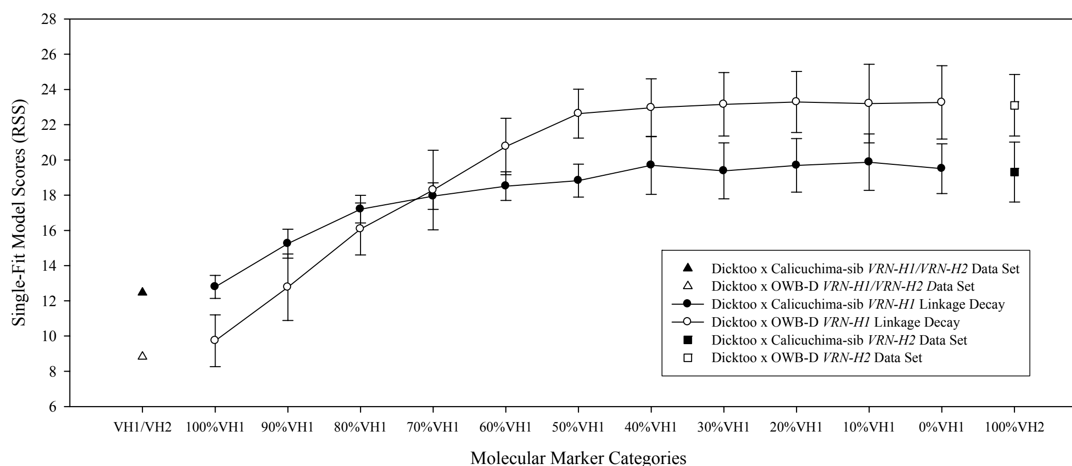
Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci USA 103: 19581-19586

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen J L, Echenique V and Dubcovsky J (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303: 1640–1644

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T and Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100: 6263–6268

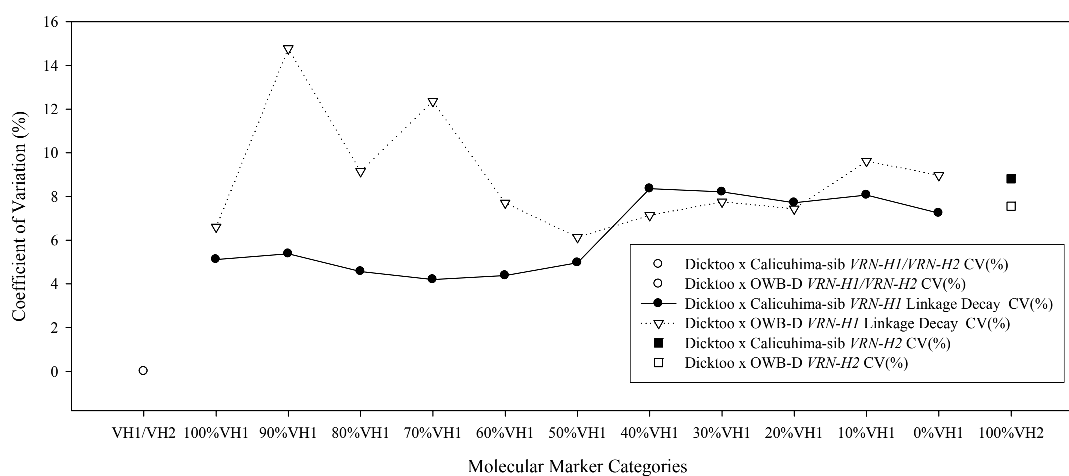
Zhang Z, Bradbury P J, Kroon D E, Casstevens T M and Buckler E S (2006) "TASSEL 2.0: a software package for association and diversity analyses in plants and animals", Plant & Animal Genomes XIV Conference

Figure 4.1. The relationship between model score and linkage decay. The mean (100 replicates) single-fit model scores (RSS) for 11 categories of linkage decay markers. RSS is the residual sums of square with error bars indicating the standard deviation from the mean.



VH1=*VRN-H1*, VH2=*VRN-H2*
VH1/VH2=Both vernalization markers in the data set showing no deviation in single-fit model scores.
VH2=*VRN-H2* showing initial large standard deviation indicating they may contribute little to strength of signal.

Figure 4.2. The coefficients of variation for 100 single-fit model scores for the individual single-fit model data sets.



VH1=*VRN-H1*, VH2=*VRN-H2*
VH1/VH2=Both vernalization markers in the data set showing no deviation in single-fit model scores.
VH2=*VRN-H2* showing initial large standard deviation indicating they may contribute little to strength of signal.

Table 4.1A and B. Null model tests for the 'Dicktoo' x 'Calicuchima'-sib[A] and 'Dicktoo' x 'OWB-D'[B] data used to check for signal in the data. Summary data where the response of the best-fit model for the data was randomly permuted 1000 times. Scores represent the residual sums of square (RSS) for the randomized data.

A

Summary Statistics	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Score (RSS)	18.87	25.24	25.95	25.79	26.68	28.19

Null Score 31.03 ; Best Score 12.47
0 randomized scores (0 %) are better than the best score

B

Summary Statistics	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Score (RSS)	25.76	31.63	32.24	32.14	32.80	34.03

Null Score 36.95 ; Best Score 8.83
0 randomized scores (0 %) are better than the best score

Table 4.2A and B. Cross-validation tests for the 'Dicktoo' x 'Calicuchima'-sib[A] and 'Dicktoo' x 'OWB-D'[B] data used to identify the size of the logic tree model with the best predictive capability. Ten cross-validation steps were performed for each model size. Training average and c/v test scores represent the residual sums of square (RSS) calculated during the cross-validation test.

A

Step #	ntree	nleaf	train.ave	train.sd	cv/test	cv/test.sd
10	1	1	22.21	1.02	24.69	10.88
20	1	2	12.43	0.66	13.68	6.14
30	1	3	12.43	0.66	13.68	6.14
40	1	4	12.43	0.66	13.68	6.14
50	1	5	12.43	0.66	13.68	6.14

B

Step #	ntree	nleaf	train.ave	train.sd	cv/test	cv/test.sd
10	1	1	20.43	0.82	21.73	8.61
20	1	2	8.83	0.22	9.69	2.23
30	1	3	8.83	0.22	9.69	2.23
40	1	4	8.83	0.22	9.69	2.23
50	1	5	8.83	0.22	9.69	2.23

Table 4.3A and B. Permutation tests for the 'Dicktoo' x 'Calicuchima'-sib[A] and 'Dicktoo' x 'OWB-D'[B] data. Summary statistics where the response of the best-fit for each model size was randomly permuted 1000 times. Scores represent the residual sums of square (RSS) for the randomized data.

A

Model Size	No. trees	No. leaves	Null	Start	Best Rand.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	% < Best
1	1	1	31.03	22.30	12.47	12.47	14.06	15.03	15.02	16.04	19.23	3.7
2	1	2	31.03	12.47	12.47	12.47	12.47	12.47	12.47	12.47	12.47	53.5
3	1	3	31.03	12.47	12.47	12.47	12.47	12.47	12.47	12.47	12.47	52.8
4	1	4	31.03	12.47	12.47	12.47	12.47	12.47	12.47	12.47	12.47	51.8
5	1	5	31.03	12.47	12.47	12.47	12.47	12.47	12.47	12.47	12.47	53.7

B

Model Size	No. trees	No. leaves	Null	Start	Best Rand.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	% < Best
1	1	1	36.95	20.45	8.83	8.83	10.99	12.81	12.11	14.20	16.75	14.1
2	1	2	36.95	8.83	8.83	8.83	8.83	8.83	8.83	8.83	8.83	64.6
3	1	3	36.95	8.83	8.83	8.83	8.83	8.83	8.83	8.83	8.83	65.7
4	1	4	36.95	8.83	8.83	8.83	8.83	8.83	8.83	8.83	8.83	63.6
5	1	5	36.95	8.83	8.83	8.83	8.83	8.83	8.83	8.83	8.83	66.7

Table 4.4. TASSEL results identifying markers associated with the days to flowering phenotype. Markers are listed with corresponding *p*-values and R^2 values.

	'Dicktoo' x 'Calicuchima'-sib	
Marker	*p*-value	R^2 (%)
VRN-H1	<0.001***	49
VRN-H2	<0.001***	12
RANDOM 70	0.0070**	8
	'Dicktoo' x 'OWB-D'	
VRN-H1	<0.001***	70
VRN-H2	<0.001***	13
RANDOM 58	<0.001***	12
RANDOM 46	0.04*	4

*Significant at the 0.05 level, ** Significant at the 0.01 level,
*** Significant at the 0.001 level

Table 4.5. Analysis of variance results for the 'Dicktoo' x 'Calicuchima'-sib data full model with markers which were found to be associated with the days to flowering phenotype in TASSEL. A type III fixed effects model with corresponding *p*-values.

Source	DF	Type III SS	Mean Square	F-value	*p*-value
VRN-H1	1	14374.4	14374.4	101.4	<0.001**
VRN-H2	1	14615.0	14615.0	103.1	<0.001
RANDOM 70	1	83.3	83.3	0.6	0.445
*VRN-H1*VRN-H2*	1	6367.3	6367.3	44.9	<0.001**
*VRN-H1**RANDOM 70	1	200.1	200.1	1.4	0.238
*VRN-H2**RANDOM 70	1	1.4	1.4	0.01	0.921
*VRN-H1*VRN-H2**RANDOM 70	1	26.3	26.3	0.2	0.668

** Significant at the 0.001 level

Table 4.6. Analysis of variance results for the 'Dicktoo' x 'OWB-D' data full model with markers which were found to be associated with the days to flowering phenotype in TASSEL. A type III fixed effects model with corresponding *p*-values.

Source	DF	Type III SS	Mean Square	F-value	*p*-value
VRN-H1	1	9236.0	9236.0	136.42	<0.001**
VRN-H2	1	8188.2	8188.2	120.94	<0.001**
RANDOM 46	1	53.0	53.0	0.78	0.379
RANDOM 58	1	169.4	169.4	2.50	0.118
*VRN-H1***VRN-H2*	1	5705.0	5705.0	84.26	<0.001**
*VRN-H1**RANDOM 46	1	23.8	23.8	0.35	0.555
*VRN-H1**RANDOM 58	1	301.1	301.1	4.45	0.038*
*VRN-H2**RANDOM 46	1	10.6	10.6	0.16	0.694
*VRN-H2**RANDOM 58	1	208.9	208.9	3.09	0.083
*VRN-H1***VRN-H2**RANDOM 46	1	17.9	17.9	0.26	0.609
*VRN-H1***VRN-H2**RANDOM 58	1	185.6	185.6	2.74	0.1017

*Significant at the 0.05 level, ** Significant at the 0.001 level

GENERAL CONCLUSION

The results from the association mapping analysis suggest there are three subpopulation clusters in hop which agree with those previously reported (Murakami 2000; Seefelder et al. 2000; Jakse et al. 2001). The subpopulations showed a moderately high level of inbreeding with a mean F_{IS} = 0.2014. The high level of inbreeding within populations may be due to the narrow genetic base of cultivated hop since many cultivars share kinship amongst three or four cultivars (Henning et al. 2004).

The variance among genotypes was significant (Table 2.2) suggesting that there is genetic variation for resistance to hop downy mildew among genotypes present in the population under observation. The distribution of resistance scores (Figure 2.1B) implies quantitative control over the expression of resistance. The ANOVA for broad-sense heritability also showed that there was significant interaction between genotypes and years demonstrating that multiple environments must be utilized when examining this trait in order to effectively identify genotypes possessing resistance (Table 2.2). Variance components estimated from the ANOVA for narrow-sense heritability also showed significant ($p<0.05$) variation among populations (Table 2.3). This demonstrates that there are true differences in levels of resistance among the populations and that these differences in populations should prove responsive to selection. There was not a significant interaction for population*year and therefore the variance component for this interaction was not estimated (Table 2.3). It was interesting that the genotype*year interaction proved significant but the population*year was not significant. Causes for this difference in significance may lie in the different genetic components estimated by these two tests. In the case of genotypes*year, the genetic component estimated from the ANOVA consists of both additive and dominance genetic components. In the case of the population*year interaction, the genetic component of population should theoretically consist of additive genetic variance. Thus, the difference between both estimates of interaction

may lie in the presence of a strong dominance effect upon this interaction. The estimates of broad-sense (H^2 = 76%) and narrow-sense heritability (h^2 = 49%) for the downy mildew resistant phenotype also suggest dominance or epistasis may be the cause of some of the differences among variance components. Broad-sense heritability represents total genetic variation relative to phenotypic variation. Therefore, the estimate of variance components for total genetic variance consists of additive, or selectable variation, and dominance/epistasis, or non-selectable deviations from expectations. By comparison, narrow-sense heritability estimates represent the ratio of selectable genetic variation, or additive genetic variance, relative to total phenotypic variation. Comparisons between the two estimates of heritability suggest the presence of a strong dominance component. Methods to increase the efficiency of selection for this trait would be the use of some means of genotypic recurrent selection or by means of molecular markers linked to downy mildew resistance or susceptibility. As genotypic recurrent selection for this particular trait would prove highly time-consuming and space-consuming, use of marker-assisted selection should prove highly advantageous.

Analysis of the mixed-model results from TASSEL showed 43 AFLP markers associated with the downy mildew resistant phenotype. The percentage of phenotypic variation explained by the individual markers was quite low ranging from 4% to 11% of the total variation (Table 2.4A and B). These low estimates suggest some of the variation was a result of environmental variance. Repeated study on the incidence of hop downy mildew have shown that prevailing weather conditions conducive to its development and spread may play a large role in year to year variation in the disease (Hoerner 1939; Pejml and Petrlik 1964; Pejml and Petrlik 1967; Royle 1970; Royle 1973; Skotland and Johnson 1983; Skotland and Romanko 1964). After reexamining the phenotype between years, it became apparent that this could be one explanation for the low R^2 values.

The results from the investigation of the hop downy mildew oospore demonstrated excellent extraction, isolation and recovery techniques for obtaining

oospores from plant tissue and soil work. Using a mortar and pestle on rewetted cotyledon material resulted in finely ground tissue which resulted in the extraction of thousands of oospores for analysis. In addition, the extraction, isolation and recovery of oospores from soil samples were a success. Although 500g samples were large and it was very time-consuming process to sieve through the samples, the results showed that the samples were uniform as the ranks among sample counts were statistically similar (Table 3.1). It is unclear why there were differences between ranks for the autoclaved and non-autoclaved samples (Table 3.1). However, the results suggest the reason for the difference may be the autoclaving procedure. Autoclaving oospores for two hours may have compromised the cellular integrity of the oospores. In doing so, the compromised oospores may have leaked their contents, lost buoyancy and settled out during centrifugation.

Although the MTT staining showed significant differences in rank among the color category counts and between the autoclaved and non-autoclaved counts, this cannot be used as a measure of viability. Altman (1974 and 1976) reported that MTT reduction to colored formazans can occur through a chemical reduction of MTT in the presence of NADH. Other compounds have been shown to do the same (Hamed 2004). Therefore, although there are significant differences in our MTT stained ranks, this data cannot be used as an objective test for viability.

Although germination of the hop downy mildew oospore was infrequent, there was some visual evidence to suggest that germination had occurred on the *in vitro* wet mounts. The thick germ tubes which contained dense cytoplasm are indicative of germination (Figure 3.1D, E, F and G). In addition, the large 'sporangium-like' structures closely resemble those reported by Rooms of Diaz and Polanco (1984). The 'zoospore-like' bodies (Figure 3.1L and M) were larger than those reported for hop downy mildew (~15µ vs. 10 µ for zoospores coming from asexual zoosporangia), these 'zoospore-like' bodies appeared to have come from the oospore as no other structures were nearby. With little documented evidence of oospore germination, it is

unclear whether zoospores derived from oospore germination would be similar in size as those derived from asexual zoosporangia.

The results from the logic regression analysis suggest Boolean logic used in a linear regression format to be more robust in the statistical modeling of epistasis when compared with general linear modeling (GLM). Logic regression correctly identified the genetic interaction of the *VRN-H1/VRN-H2* locus in a model using the days to flowering phenotype without the inclusion of spurious markers. In contrast, GLM identified randomly generated markers as being associated with the phenotype and in the 'Dicktoo' x 'OWB-D' data, GLM identified not just a spurious association, but a spurious interaction between a randomly generated marker (RANDOM 58) and marker *VRN-H1* (Table 4.6).

The linkage decay (LD) results suggest logic regression may be able to identify 'potentially troublesome' data sets, which when modeled, may lead to type I errors. Logic regression will model most data sets, however it is crucial to identify the precise point where statistical signal falls short resulting in the modeling of nonsensical data. The LD results suggest coefficients of variation (CVs) for single-fit models of less than 6% result in precise modeling with no spurious association (Figure 4.2, APPENDIX C). The results also suggest that single-fit models with CVs greater than 6% may result in the modeling of nonsensical data (Figure 4.2, APPENDIX C).

Logic regression also revealed one unexpected result. When logic regression identified the model used to explain the days to flowering phenotype in barley, it created a Boolean logic group that consisting of *VRN-H2* and 'not' *VRN-H1*. The word 'not' suggests the conjugate form of *VRN-H1* to be epistatic to *VRN-H2* in the logic model. Barley vernalization is one of the best-documented epistatic interactions in the world (Takahashi and Yasuda 1971; Yan et al. 2004; Szűcs et al. 2007). It's a classic dominant/suppression interaction where the product of gene one normally suppresses a second locus (gene two), unless a mutation at the second locus prevents binding of the repressor allowing for the expression of gene two. In barley, dominant *VRN-H2* normally inhibits the expression of recessive *VRN-H1* alleles resulting in

vernalization sensitivity unless a mutation (at the first intron of *VRN-H1*) removes the repressor binding site resulting in a dominant *VRN-H1* which is vernalization insensitive. The Boolean combination of *VRN-H2* and 'not' *VRN-H1* (or the conjugate model) suggests a dominant *VRN-H2* and a recessive *VRN-H1* allele. This is the hypothesized genetic model governing the vernalization response in barley (Yan et al. 2004; Szűcs et al. 2007).

BIBLIOGRAPHY

Aegerter B J, Nuñez J J and Davis R M (2002) Detection of management of downy mildew in rose rootstock. Plant Dis 86: 1363-1368

Altman F P (1974) Studies on the reduction of tetrazolium salts. III. The products of chemical and enzymatic reduction. Histochemie 38: 155-171

Altman F P (1976) Tetrazolium salts and formazans. Progr Histochem Cytochem 9: 1-55

Arens K (1929) Untersuchungen über *Pseudoperonospora humuli* (Miyabe u. Takah.), den Erreger der neuen Hopfenkrankheit. Phyto Zeitschr 1: 169-193

Bacanu S-A, Devlin B and Roeder K (2000) The power of genomic control. Am J Hum Genet 66: 1933-1944

Barton N H, Halliday R B and Hewitt G B (1983) Rare electrophoretic variants in a hybrid zone. Heredity 50: 139-146

Beaumont M A, Zhang W and Balding D J (2002) Approximate Bayesian computation in population genetics. Genetics 162: 2025-2035

Beavis W D (1998) QTL analysis: power precision and accuracy. In: Paterson A H (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, pp. 145-162

Benham, J, Jeung J, Jasieniuk M, Kanazin V and Blake T (1999) Genographer: A graphical tool for automated fluorescent AFLP and microsatellite analysis. http://hordem.oscs.montana.edu/genographer/ (Checked 5/26/2007)

Binder W D, Mitchell G M and Ballantyne D J (1974) Pollen viability testing, storage and related physiology: review of literature with emphasis on gymnosperm pollen. Canadian Forestry Service, Victoria, BC, Canada

Bowers J H, Papavizas G C and Johnston S A (1990) Effect of soil temperature and soil-matrix potential on the survival of *Phytophthora capsici* in natural soil. Plant Dis 74: 771-777

Box G E P (1979) Robustness in the strategy of scientific model building. In: Launer R L and Wilkinson G N (eds) Robustness in statistics. Academic Press, New York, pp. 201-236

Bressman E M and Nichols A A (1933) Germination of the oospores of *Pseudoperonospora humuli*. Phytopath 23: 485-487

Caicedo A L, Stinchcombe J R, Olsen K M, Schmitt J and Purugganan M D (2004) Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitude cline in a life history trait. Proc Natl Acad Sci USA 101: 15670-15675

Cardon L R and Bell J I (2001) Association study designs for complex diseases. Nature Reviews (Genetics) 2: 91-99

Carlborg Ö and Haley C S (2004) Epistasis: too often neglected in complex trait studies. Nature 5: 618-625

Carlson C S, Eberle M A, Kruglyak L and Nickerson D A (2004) Mapping complex disease loci in whole-genome association studies. Nature 429: 446-452

Carlson C S, Newman T L and Nickerson D A (2001) SNPing in the human genome. Curr Op in Chem Biol 5: 78-85

Chee H Y and Klein R E (1998) Laboratory production of oospores in *Pseudoperonospora humuli*. Korean J Plant Pathol 14: 618-621

Chee H Y, Nelson M E, Grove G G, Eastwell K C, Kenny S T and Klein R E (2006) Population biology of *Pseudoperonospora humuli* in Oregon and Washington. Plant Dis 90: 1283-1286

Chesler, E J, Rodriguez-Zas S L and Mogil J S (2001) In silico mapping of mouse quantitative trait loci. Science 294: 2423a

Chmielewicz K M and Manly K F (2002) User manual for Mapmanager QTX. Roswell Park Cancer Institute, New York

Cohen S D (1984) Detection of mycelium and oospores of *Phytophthora megasperma* f. sp. *glycinea* by vital stains in soil. Mycologia 76: 34-39

Coley-Smith J R (1960) Overwintering of hop downy mildew *Pseudoperonospora humuli* (Miy. and Tak.) Wilson. Rep Dep Hop Res Wye College for 1959: 107-114

Danyluk J, Kane N D, Breton G, Limin A E, Fowler D B and Sarhan F (2003) *TaVRT1*, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol 132: 1849–1860

Darvasi A (2001) In silico mapping of mouse quantitative trait loci. Science 294: 2423a

Darvasi A and Soller M (1994) Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor Appl Genet 89: 351-357

Darvasi A and Soller M (1997) A simple method to calculate resolving power and confidence interval of QTL map location. Behaviour Genetics 27: 125-132

Darvasi A, Weinreb A, Minke V, Weller J I and Soller M (1993) Detecting marker QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943-951

Davis J J (1910) A new hop mildew. Science 31: 752

Devlin B, Roeder K and Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Pop Biol 60: 155-166

Dilda C L and Mackay T F C (2002) The genetic architecture of *Drosophila* sensory bristle number. Genetics 162: 1655-1674

Dodds P N, Lawrence G J, Catanzariti A, Teh T, Wang C A, Ayliffe M A, Kobe B and Ellis J G (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. Proc Nat Acad Sci USA 103: 8888–8893

El-Hamalawi Z A and Erwin D C (1986) Physical, enzymatic and chemical factors affecting viability and germination of oospores of *Phytophthora megasperma* f. sp. *medicaginis*. Phytopath 76: 503-507

Falush D, Stevens M and Pritchard J K (2007) Inferences of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecology Notes doi: 10.1111/j.147-8286.2007.01758.x

Flor H H (1942) Inheritance of pathogenicity in *Melampsora lini*. Phytopath 32: 653-669

Fu D, Szűcs P, Yan L, Helguera M, Skinner J S, von Zitzewitz J, Hayes P M and Dubcovsky J (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. Mol Genet Genom 273: 54–65

Gahan P B and Kalina M (1968) The use of tetrazolium salts in the histochemical demonstration of succinic dehydrogenase activity in plant tissues. Histochemie 14: 81-88

Gaut B S and Long A D (2003) The lowdown on linkage disequilibrium. Plant Cell 15: 1502-1506

Goldstein J (1995) Multilevel statistical models. Kendall's Library of Statistics. Edward Arnold, London, England

Gowda P S B and Bhat S S (1986) Germination of oospores of *Peronosclerospora sorghi*. Trans Br Mycol Soc 87: 653-655

Grupe A, Germer S, Usuka J, Aud D, Belknap J K, Klein R F, Ahluwalia M, Higuchi R and Peltz G (2001) In silico mapping of complex disease-related traits in mice. Science 292: 1915-1918

Haley C S and Knott S A (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324

Hamed S H (2004) Efficacy and mechanism of action of a new tyrosinase inhibitory agent. PhD Diss, Univ of Cincinnati, OH. http://www.ohiolink.edu/etd/send-pdf.cgi?ucin1085566152 (verified 07/23/2007)

Hardy O J (2003) Estimation between individuals and characterization of isolation-by -distance processes using dominant genetic markers. Mol Ecology 12: 1577-1588

Hardy O J and Vekemans X (2002) SPAGeDI: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecology Notes 2: 618-620

Hardy O and Vekemans X (2006) SPAGeDi 1.2 user's manual. Laboratoire Eco-éthologie Evolutive, Bruxelles, Belgium

Haunold A and Zimmermann C E (1974) Pollen collection, crossing and seed germination of hop. Crop Science 14: 774-776

Henning J A (2006) The breeding of hop. In: Bamworth C (ed) Brewing: New Technologies. Woodhead Publishing Limited Cambridge, England, pp. 102-122

Henning J and Haunold A (2003) Notice of release of high-yielding, multi-disease resistant, high bittering acid hop variety, 'Newport'. USDA-ARS, Washington, DC

Henning J A, Townsend M S and Kenny S (2004) Potential heterotic crosses in hops as estimated by AFLP-based genetic diversity and coefficient of coancestry. J Am Soc Brew Chem 62:63-70

Hoerner G R (1939) The relation of climatology of western Oregon to the incidence and control of downy mildew of hops. Plant Dis Rpt 23: 361-366

Hoerner G R (1949) Hop diseases in the United States. The Brewers Digest 24: 45-51

Holsinger K E (1999) Analysis of genetic diversity in geographically structured populations: A Bayesian approach. Hereitas 130: 245-255.

Holsinger K E, Lewis P O and Dey D K (2002) A Bayesian approach to inferring population structure for dominant markers. Mol Ecology 11: 1157-1164

Hyne V and Kearsey M J (1995) QTL analysis: further uses of marker regression. Theor Appl Genet 91: 471-476

Ivandic V, Hackett C A, Nevo E, Keith R, Thomas W T B and Forster B P (2002) Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. Plant Mol Biol 48: 511–527

Ivandic V, Thomas W T B, Nevo E, Zhang Z and Forster B P (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. Plant Breed 122: 300–304

Jakse J, Kindlhofer K and Javornik B (2001) Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. Genome 44: 773-782

Jiang J and Erwin D C (1990) Morphology, plasmolysis and tetrazolium bromide stain as a criteria for determining viability of *Phytophthora* oospores. Mycologia 82: 107-113

Jones W (1932b) A new technique for obtaining oospores of hop downy mildew by inoculating cotyledons. Science 75: 108

Kearsey M J and Farquhar G L (1998) QTL analysis in plants: where are we now? Heredity 80: 137-142

Kearsey M J and Hyne V (1994) QTL analysis: a simple 'marker-regression' approach. Theor Appl Genet 89: 698-702

Kennelly M M, Gadoury D M, Wilcox, W F, Magarey P A and Seem R C (2007) Primary infection, lesion productivity, and survival of sporangia in the grapevine downy mildew pathogen *Plasmopara viticola*. Phytopath 97: 512-522

Kenny S T (1991) Registration of five hop germplasms for hop downy mildew research. Crop Science 31: 1391-1392

Kidwell K K and Osborn T C (1992) Simple plant DNA isolation procedures. In: Plant genomes. Methods for genetic and physical mapping. Kluwer Academic Publ, the Netherlands, pp. 1-13

Klein R E (1995) Hop diseases and their control. HRC Reports, 41-45

Knapp S J (1991) Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny. Theor Appl Genet 81: 333-338

Knott S A and Haley C S (1992) Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. Genetical Research 60: 139-151

Kooperberg C and Ruczinski I (2005) The logic regression package. In: Contributed Packages. R Project. http://cran.r project.org/src/contrib/PACKAGES.html

Kooperberg C, Ruczinski I, LeBlanc M L and Hsu L (2001) Sequence analysis using logic regression. Genetic Epidemiology 21 (Suppl. 1): S626-S631

Krauss S L (2000) Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. Mol Ecology 9: 1241–1245

Lande R and Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743-756

Lander E S and Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199

Lander E S, Green P, Abrahamson J, Barlow A, Daly M J, Lincoln S E and Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181

Lander E S and Kruglyak L (1995) Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genetics 11: 241-247

Lander E S and Schork N J (1994). Genetic dissection of complex traits. Science 265: 2037–2048

Lefebvre V, Daubèze A–M, Rouppe van der Voort J, Peleman J, Bardin M and Palloix A (2003) QTLs for resistance to powdery mildew in pepper under natural and artificial infections. Theor Appl Genet 107: 661-666

Lindqvist-Kreuze H, Koponen H and Valkonen J P T (2002) Variability of *Peronospora sparsa* (syn. *P. rubi*) in Finland as measured by amplified fragment length polymorphism. Eur J of Plant Path 108: 327-335

Loiselle B A, Sork V L, Nason J and Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82: 1420-1425

Luo Z W and Kearsey M J (1992) Interval mapping of quantitative trait loci in an F_2 population. Heredity 69: 236-242

Lynch M and Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genetics 152: 1753-1766

Ma H–X, Bai G–H, Zhang X and Lu W–Z (2006) Main effects, epistasis, and environmental interactions of quantitative trait loci for Fusarium head blight resistance in a recombinant inbred population. Phytopath 96: 534-541

MacKay D B (1972) The measurements of viability. In Roberts E H (ed) Viability of seeds. Chapman and Hall, London, United Kingdom, pp. 172-208

Medina M V and Platt H W (1999) Viability of oospores of *Phytophthora infestans* under field conditions in northeastern North America. Can J Plant Path 21: 137-143

Meier R and Charvat I (1993) Reassessment of tetrazolium bromide as a viability stain for spores of vesicular-arbuscular mycorrhizal fungi. Am J of Bot 80: 1007-1015

Melchinger A E, Utz H F and Schön C C (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics. 149: 383-403

Morgan W M (1978) Germination of *Bremia lactucae* oospores. Trans Br Mycol Soc 71: 337-340

Morley M, Cliona M M, Weber T M, Devlin J L, Ewens K G, Spielman R S and Cheung V G (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430: 743-747

Murakami A (2000) Hop variety classification using genetic distance based on RADP. J Inst Brew 106: 157-161

Nei M and Chesser R K (1983) Estimation of fixation indices and gene diversities. Ann Hum Gen 47: 253-259

Nelson G A and Olsen O A (1967) Staining reactions of resting sporangia of *Synchytrium endobioticum* with a tetrazolium compound. Phytopath 57: 965-968

Nelson M E, Eastwell K C, Grove G G, Barbour J D, Ocamb C M and Alldredge J R (2004) Sensitivity of *Pseudoperonospora humuli* (the casual agent of hop downy mildew) from Oregon, Idaho, and Washington to fosetyl-Al (Aliette). Online. Plant Health Progress doi: 1094/PHP-2004-0811-01-RS

Neve R A (1991) Hops. 1st edition, Chapman and Hall, New York

Neve R A and Darby P (1983) Plant breeding. Rep Dept Hop Res Wye Coll for 1982, pp. 8-11

Neve R A and Darby P (1984b) Yoeman and Zernith: two new varieties. Rep Dept Hop Res Wye Coll for 1983, pp. 7-9

Nordborg M, Borevitz J O, Bergelson J, Berry C C, Chory J Hagenblad J, Kreitman M, Maloof J N, Noyes T, Oefner, P J, Stahl E A and Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30: 190-193

Orr H A (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139: 1805-1813

Pakniyat H, Powell W, Baird E, Handley L L, Robinson D, Scrimgeur C M, Nevo E, Hackett C A, Caligari P D S and Forster B P (1997) AFLP variation in wild barley (*Hordeum spontaneum* C Koch) with reference to salt tolerance and associated ecogeography. Genome 40: 332–341

Palaisa K A, Morgante M, Williams M and Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15: 1795–1806

Patel J K, Patel N M and Shiyani R L (2001) Coefficient of variation in field experiments and yardstick thereof-An empirical study. Current Science 81: 1163-1164

Pejml K and Petrlik Z (1964) The influence of weather conditions on the spread of downy mildew. Chmelarstvi 37: 73-73

Pejml K and Petrlik Z (1967) Typization of characteristic weather for spreading of downy mildew in hops. Ochr Rost 5: 109-116

Populer C (1981) Epidemiology of downy mildews. In Spencer D M (ed) The Downy Mildews. Academic Press, London, New York, San Francisco pp. 57-105

Pritchard J K, Stephens M and Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Pritchard J K, Stevens M and Donelly P (2000b) Association mapping in structured populations. A J Hum Genet 67: 170-181

Queller D and Goodnight K F (1989) Estimating relatedness using genetic markers. Evolution 43: 258-275

Rannala B and Hartigan J A (1995) Identity by descent in island-mainland populations. Genetics 139: 429-437

Remington D L, Thornsberry J M, Matsuoka Y, Wilson L M, Whitt S R, Doebley J, Kresovich S, Goodman M M and Buckler IV E S (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98: 11479–11484

Ritland K (1996) A marker-based method for inferences about quantitative inheritance in natural populations. Evolution 50: 1062-1073

Ritter E, Gebhardt C and Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125: 645-654

Rooms of Diaz G and Polanco C D (1984) Oospores of the parasitic fungus *Peronosclerospora sorghi* under conditions of flood. Tropical Agronomy 34: 87-94

Rousset F (2002) Inbreeding and relatedness coefficients: what do they mean? Heredity 88: 371-380

Royle D J (1970) Infection periods in relation to the natural development of hop downy mildew (*Pseudoperonospora humuli*). Ann App Biol 66: 281-291

Royle D J (1973) Quantitative relationships between infection by hop downy mildew pathogen, *Pseudoperonospora humuli*, and weather and inoculum factors. Ann Appl Biol 73: 19-30

Royle D J and Kremheller H Th (1981) Downy mildew of hop. In: Spencer D M (ed) The Downy Mildews. Academic Press, London, New York, San Francisco, pp. 395-419

Ruczinski I, Kooperberg C and LeBlanc M (2003) Logic regression. J Comput Graph Stat 12: 475–511

Ruczinski I, Kooperberg C and LeBlanc M L (2004) Exploring interactions in high-dimensional genomic data: an overview of logic regression with applications. J Multi Analysis 90: 178-195

Ryley M J (2001) Location and activity of hyphae of downy mildew, *Peronosclerospora noblei* (Family Peronosporaceae), and its relationship to symptom expression on wild sorghum (*Sorghum leiocladum*). Aust J Bot 49: 487-492

Salmon E S and Ware E M (1925) The downy mildew of hop and its epidemic occurance in 1924. Ann Biol 12: 121-151

Sax K, (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8: 552-560

Seefelder S, Ehrmaier H, Schweizer G and Seigner E (2000) Genetic diversity and phylogenetic relationships among accessions of hop, *Humulus lupulus*, as determined by amplified fragment length polymorphism and fingerprinting compared with pedigree data. Plant Breed 119: 257-263

Shook D R and Johnson T E (1999) Quantitative trait loci affecting survival and fertility related traits in *Caenorhabditis elegans* show genotype-environment interactions, pleiotropy and epistasis. Genetics 153: 1233-1243

Skotland C B (1961) Infection of hop crowns and roots by *Pseudoperonospora humuli* and its relation to crown and root rot and overwintering of the pathogen. Phytopath 51: 241-244

Skotland C B and Johnson D A (1983) Control of downy mildew of hops. Plant Dis 67: 1183-1186

Skotland C B and Romanko R R (1964) Life history of the hop downy mildew fungus. Bull Idaho Agric Exp Stn, no. 424

Slatkin M and Barton N H (1989) A comparison of three indirect methods for estimation average levels of gene flow. Evolution 43: 1349-1368

Spiegelhalter D J, Best N G, Carlin B P and van der Linde A (2002) Bayesian measures and model complexity and fit. J Royal Stat Soc Ser B 64: 483-689

Spielman R S, McGinnis R E and Ewens W J (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506-513

Spring O and Zipper R (2000) Isolation of oospores of sunflower downy mildew, *Plasmopara halstedii*, and microscopic studies of oospore germination. J Phytopath 148: 227-231

Staub J E, Serquen F C and Gupta M (1996) Genetic markers, map construction, and their application in plant breeding. Hortscience 31: 729-740

Sutherland D E and Cohen S D (1983) Evaluation of tetrazolium bromide as a vital stain for fungal oospores. Phytopath 73: 1532-1535

Szűcs P, Skinner J S, Karsai I, Cuesta-Marcos A, Haggard K G, Corey A E, Chen T H H and Hayes P M (2007) Validation of the *VRN-H2/VRN-H1* epistatic model in barley reveals that intron length variation in *VRN-H1* may account for the continuum of vernalization sensitivity. Mol Genet Genom 277: 249-261

Takahashi R and Yasuda S (1971) Genetics of earliness and growth habit in barley. In: Nilan RA (ed) Barley Genetics II. Proceedings of the Second International Barley Genetics Symposium. Washington State University Press Pullman, pp. 388–408

Tenaillon M I, Sawkins M C, Long A D, Gaut R L, Doebley J F and Gaut B S (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). Proc Natl Acad Sci USA 98: 9161–9166

Thoday J M (1961) Location of polygenes. Nature 191: 368-370

Thornsberry J M, Goodman M M, Doebley J, Kresovich S, Nielsen D and Buckler IV E S (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genetics 28: 286-289

Townsend M S and Henning J A (2005) Potential heterotic groups in hop as determined by AFLP analysis. Crop Science 45: 1901-1907

Ungerer M C, Halldorsdottir S S, Modliszewski J L, Mackay T F C and Purugganan M D (2002) Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. Genetics 160: 1133-1151

Van der Gaag D J (1994) The effect of pH on staining of oospores of *Peronospora viciae* with tetrazolium bromide. Mycologia 86: 454-457

Van der Gaag D J and Frinking H D (1996) Extraction from plant tissue and germination of oospores of *Peronospora viciae* f. sp. *pisi*. Phytopath 144: 57-62

Van der Gaag D J and Frinking H D (1997) Extraction of oospores of *Peronospora viciae* from soil. Plant Path 46: 675-679

van Ooijen J W (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84: 803-811

Van Waes J M and Debergh P C (1986) Adaptation of the tetrazolium method for testing the seed viability, and scanning electron microscopy of some western European orchids. Physiol Plantar 66: 435-442

Vekemans X (2002) AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium

Verhoeven K J F, Jannink J-L and McIntyre L M (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96: 139-149

Virk P S, Ford-Lloyd B V, Jackson M T, Pooni H S, Clemeno T P and Newbury H J (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76: 296–304

von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Pecchioni N, Francia E., Casas A, Chen T H H, Hayes P M and Skinner J S (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59: 449–467

Ware W M (1926) *Pseudoperonospora humuli* and its mycelial invasion of the host plant. Trans Brit Mycol Soc 11: 91-107

Ware W M (1929) Experiments on the production of diseased shoots by the hop downy mildew, *Pseudoperonospora humuli* (Miy. et Takah.), Wils. Ann of Bot 43: 683-711

Wehrhahn C F and Powell R (1987) Electrophoretic variation, regional differences, gene flow in the coho salmon (*Onchorhyncus kisutch*) of southern British Columbia. Can J Fish and Aquat Sci 44: 822-831

Wehrhahn C R (1989) Proceedings of the ecological genetics workshop. Genome 31: 1098-1099

Weinig C, Dorn L A, Kane N C, German Z M, Halldorsdottir S S, Ungerer M, Toyonaga Y, Mackay T F C, Purugganan M D and Schmitt J (2003) Heterogenous selection at specific loci in natural environments in *Arabidopsis thaliana*. Genetics 165: 321-329

Weir B S and Cockerman C C (1984) Estimating F-statistics for analysis of population structure. Evolution 38: 1358-1370

Weller J I (1992) Statistical methodologies for mapping and analysis of quantitative trait loci. In: Beckmann J S and Osborn T C (eds) Plant genomes: methods for genetic and physical mapping. Klewer Academic Publishers, Norwell, MA pp. 181-207

Weller J I, Kashi Y and Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J of Dairy Science 73: 2525-2537

Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97-159

Wright S (1969) Evolution and the genetics of populations, Volume 2. The theory of gene frequencies. Un of Chicago Press, Chicago, IL

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S and Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci USA 103: 19581-19586

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen J L, Echenique V and Dubcovsky J (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303: 1640–1644

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T and Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100: 6263–6268

Young N D (1996) QTL mapping and quantitative disease in plants. Ann Rev Phytopath 34: 479-501

Yu J, Pressoir G, Briggs W H, Bi I V, Yamasaki M, Doebley J F, McMullen M D, Gaut B S, Nielsen D M, Holland J B, Kresovich S and Buckler E S (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38: 203-208

Zhang Z, Bradbury P J, Kroon D E, Casstevens T M and Buckler E S (2006) "TASSEL 2.0: a software package for association and diversity analyses in plants and animals", Plant & Animal Genomes XIV Conference

Zhivotovsky L A (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecol 8: 907–913

APPENDICES

APPENDIX A

HOP GENOTYPE PEDIGREE INFORMATION (Townsend and Henning 2005).

Primarily European ancestry:

M19007 Brewer's Favorite-s

M19047 Elsasser/Fuggle-s

M21690 Late Grape-s//Fuggle/Fuggle-s/3/Late Cluster-s/Fuggle-s/4/(Late Grape-s//Fuggle/Fuggle-s/3/Late Cluster-s/Fuggle-s

M63011 Late Grape/Fuggle-s//Early Green/Unknown-s

U.S. Tettnanger (F21015) Fuggle

Styrian (F21049) Yugoslavian selection from Fuggle

M64034 Zattler-s

M64036 Zattler-s

M64037 Zattler-s

M64101 Unknown

M64035 Zattler-s

M21119 Late Grape/OP/3/Fuggle//Fuggle/OP/4/Late Cluster/OP//Fuggle/OP

M21009 Sunshine-s/3/Utah-523-4//Early Green/Unknown-s

M21132 Yakima Cluster/Zattler-s

M21336 Northern Brewer//Bullion/Zattler-s

M21335 Northern Brewer//Bullion/Zattler-s

M21087 Yugoslavia Selection 3-3

M21090 Yugoslavia Selection 12-17

M21268 Northern Brewer/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

M21398 Native Yugoslavian Male 01P04

M21400 Native Yugoslavian Male 20P09

Hallertauer Mittlefruh (F21014) German landrace

Saazer 36 (F21521) Clonal selection from Saazer

Fuggle N (F21016) Clonal selection from Fuggle

Tardif de Bourgogne (F21169) France landrace

Spalter Select (F21674) German 76-18-80/German 71-16-07

Perle (F21227) Northern Brewer/German 63-5-27

Yeoman (F21498) Wye 43/69/17 _ Wye 25/68/173

Challenger (F21043) Zattler/open pollinated//Northern Brewer/Wye 22-56

Omega (F21667) Wye Challenger/English male

Northern Brewer (F21093) Brewer's Gold/OY1//Canterbury Golding

Orion (F21675) Perle/German 70-10-15

Wye Viking (F21283) Svalof//Bramling Cross/Wye 1-63-42

Wye Saxon (F21282) Svalof//Bramling Cross/Wye 1-63-42

M21089 Yugoslavia selection 5-10

Primarily Wild American/Europe Hybrids (A):

M19009 Fuggle/Fuggle-s

M21424 Cascade/Late Cluster-s

M19172 Cat's Tail//Fuggle/Fuggle-s

M19036 Late Cluster/Fuggle-s

M19037 Fuggle-s/Fuggle-s

M19061 Late Grape/Fuggle-s

M19046 Late Cluster-s/Fuggle-s

M19060 East Kent Golding/Bavarian-s

M21058 Fuggle//Striesselspalt/Late Cluster-s

M21446 Northern Brewer/3/Brewer's Gold//East Kent Golding/Bavarian-s

M58111 Brewer's Gold//Belgian-31-s/Belgian-31/3/Late Grape/Fuggle-s

M52047 Striesselspalt//Early Green/Unknown-s/3/Striesselspalt/Late Cluster-s

M21603 Cascade//Semsch-s/8-3B yrd

M51114 Landhopfen-s//Golden Cluster/Fuggle-s/3/Semsch-s/8-2B Yrd

M21076 Comet/3/Brewer's Gold//Fuggle/Colorado-2-1

M21339	Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s
M21358	Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Brewer's Gold//East Kent Golding/Bavarian-s
M21462	Cascade//Fuggle/Fuggle-s
M21072	Brewer's Gold/Arizona-1-2
M21129	Late Grape-s//Fuggle/Fuggle-s/3/Brewer's Gold/Utah-526-4
M21461	Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/(Brewer's Gold//Early Green/Unknown-s)/3/Zattler-s/5/Comet//Bullion/Zattler-s
M21351	Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Bullion/Zattler-s
M63015	Brewer's Gold//East Kent Golding/Bavarian-s
M64102	Wild American/open pollinated
M64105	Fuggle//Wild American/open pollinated
M21432	Cascade/4/Late Grape-s//Fuggle/Fuggle-s/3/Early Green/Unknown-s
M21437	Fuggle/open pollinated
M21435	Cascade/Colorado-1-1

<u>Primarily Wild American/European Hybrids (B)</u>:

M21109	Brewer's Gold//Early Green/Unknown-s/3/Zattler-s
M21313	Comet/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s
M21306	Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s/5/Comet//Bullion/Zattler-s
M21345	Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s
M21329	Comet//Bullion/Zattler-s
M21110	Bullion/Zattler-s
M21135	Brewer's Gold//Early Green/Unknown-s/3/Zattler-s
M21303	Bullion/Zattler-s
M21300	Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

M21360 Cascade/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

M21466 Comet/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

M21416 Bullion/Zattler-s

M21272 Northern Brewer//Bullion/Zattler-s

M21415 Brewer's Gold//Early Green/Unknown-s/3/Late Cluster-s/Fuggle-s

M21273 Comet/3/Brewer's Gold//Fuggle/Colorado-2-1

M21444 Comet/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

M21417 Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s/5/Comet//Bullion/Zattler-s

M21420 Comet/3/Brewer's Gold//Fuggle/Colorado-2-1/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s/4/Northern Brewer//Bullion/Zattler-s

M21426 Cascade//Fuggle/Fuggle-s

M21428 Cascade//Fuggle-s/Fuggle-s

M21448 Cascade/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

M21463 Cascade/Yugoslavian 3-3

Cascade (F21092) Fuggle//Serebrianca/Fuggle-s/3/open pollinated

Wye Target (F21112) Northern Brewer/Wye 22-56//Eastwell Golding/OB79

M21465 Comet/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

Comet (F62013) Sunshine-s/Utah 524-2

M21184 Unknown

M21425 Cascade//Semsch-s/8-2B yrd

M21427 Cascade//Red Vine/Fuggle-s

Brewer's Gold (F21116) BB1/open pollinated

Galena (F21182) Brewer's Gold/open pollinated

Galena VF (F21699) Meristem-tip culture from Galena

Crystal (F21490) 4xHall Mittlefruh //Cascade/USDA 65009M

Kitamidori (F21677) Japan C79-27-01/Japan C79-64-110

Magnum (F21670) Galena/German 75/5/3

M21488 Cascade/4/Brewer's Gold//Early Green/Unknown-s/3/Zattler-s

APPENDIX B

Disease Note

The Oospore of *Pseudoperonospora humuli* Produced in Abundance by Inoculating Hop Cotyledons. T. B. Parker, Department of Crop and Soil Science, Oregon State University, Corvallis, OR; 97331. J. A. Henning, Hop Research Unit, USDA-ARS National Forage Seed Processing Research Center, Corvallis, OR; 97331. W. Mahaffee, USDA-ARS Horticultural Crops Research Laboratory, Corvallis, OR, 97331.

Downy mildew is a major disease affecting hop (*Humulus lupulus* L.) in Oregon and around the world. Jones (1932b) successfully produced oospores of hop downy mildew in abundance by sowing 'Late Cluster' hop seedlings and inoculating the cotyledons with minute portions of infected leaves obtained from basal spikes. We wanted to determine if oospores could be produced in abundance within any cotyledon tissue or whether this phenomenon was genotype specific to Late Cluster. We repeated the work of Jones (1932b), except we inoculated flats of 'Montana II' x Open Pollinated and 'Wyoming' x Open Pollinated F_1 seedlings with zoosporangia collected from host plants grown in a mist chamber. The flats were bagged, misted daily and maintained at 16°C for six days. When the cotyledons showed obvious signs of downy mildew infection, they were removed. Leaf peels of the cotyledons revealed oospores in abundance inside the plant tissue. The results indicate oospores are easily produced in hop cotyledons and that their production is not restricted to the highly susceptible Late Cluster variety. Additional work is being carried out to determine the viability of the oospores by staining with MTT.

References: (1) W. Jones. Science 75: 108, 1932b.

APPENDIX C

Chapter 4. Supplemental Material: The Dicktoo' x 'OWB-D' data barley vernalization data showing the single-fit and multiple-fit model scores. The 'Dicktoo' x 'Calicuchima'-sib data linkage decay series.

Single-fit and multiple-fit model selections for the 'Dicktoo' x 'OWB-D' data barley vernalization data:

Single-fit model:

```
> fitA
score 8.825
 -85.4 * (VRNH1 or (not VRNH2))
```

Multiple-fit model for the 'Dicktoo' x 'OWB-D' barley:

```
> fitB
1 trees with 1 leaves: score is 20.45
 +68.4 * (not VRNH1)
1 trees with 2 leaves: score is 8.825
 +85.4 * ((not VRNH1) and VRNH2)
1 trees with 3 leaves: score is 8.825
 -85.4 * (VRNH1 or (not VRNH2))
1 trees with 4 leaves: score is 8.825
 +85.4 * ((not VRNH1) and VRNH2)
1 trees with 5 leaves: score is 8.825
 -85.4 * (VRNH1 or (not VRNH2))
```

‘Dicktoo’ x ‘Calicuchima’-sib data linkage decay data for the individual decay series of *VRN-H1* and *VRN-H2* shown in Figures 4.1 and 4.2. **Note** ‘Dicktoo’ x ‘OWB-D’ linkage decay series was not analyzed as Figure 4.2 suggested the signal was too weak to obtain any meaningful results.

VRN-H1 Linkage Decay Series

Multiple-fit model of the ‘Dicktoo’ x ‘Calicuchima’-sib data for *VRN-H1*:

```
> fitB
1 trees with 1 leaves: score is 22.298
 -49.3 * VRNH1
1 trees with 2 leaves: score is 12.472
 +74.9 * ((not VRNH1) and VRNH2)
1 trees with 3 leaves: score is 12.472
 +74.9 * (VRNH2 and (not VRNH1))
1 trees with 4 leaves: score is 12.472
 +74.9 * ((not VRNH1) and VRNH2)
1 trees with 5 leaves: score is 12.472
 +74.9 * (VRNH2 and (not VRNH1))
1 trees with 6 leaves: score is 12.472
+74.9 * (VRNH2 and (not VRNH1))
```

Multiple-fit model for 90% similar to *VRN-H1*:

```
> fitB
1 trees with 1 leaves: score is 25.974
 -36.3 * Ninety
1 trees with 2 leaves: score is 24.531
 +45.6 * (RANDOM83 and (not Ninety))
1 trees with 3 leaves: score is 22.313
 +60.3 * (((not Ninety) and RANDOM80) and RANDOM83)
```

1 trees with 4 leaves: score is 19.629

+62.1 * ((not Ninety) and ((RANDOM83 and RANDOM80) or (not RANDOM29)))

1 trees with 5 leaves: score is 18.594

+62.7 * ((not Ninety) and ((RANDOM83 and RANDOM80) or ((not RANDOM29) or (not RANDOM79))))

1 trees with 6 leaves: score is 16.371

-71.4 * (((RANDOM40 and RANDOM93) and RANDOM78) or (Ninety or (RANDOM28 and (not RANDOM57))))

Linkage decay multiple-fit model for 80% similar to *VRN-H1*:

> fitB

1 trees with 1 leaves: score is 27.448

+29.8 * (not Eighty)

1 trees with 2 leaves: score is 26.39

+36.7 * (RANDOM83 and (not Eighty))

1 trees with 3 leaves: score is 24.384

+47.7 * ((not Eighty) and ((not RANDOM85) or (not RANDOM63)))

1 trees with 4 leaves: score is 21.64

+62.2 * ((not Eighty) and (((not RANDOM70) or (not RANDOM85)) and RANDOM83))

1 trees with 5 leaves: score is 19.961

-84.6 * ((Eighty or (not RANDOM48)) or ((RANDOM74 and RANDOM66) or (not RANDOM84)))

1 trees with 6 leaves: score is 17.103

-77 * (((RANDOM63 and RANDOM96) or ((not RANDOM84) and RANDOM25)) or ((not RANDOM35) or Eighty))

Linkage decay multiple-fit model for 70% similar to *VRN-H1*:

> fitB

1 trees with 1 leaves: score is 28.289

-26 * Seventy

1 trees with 2 leaves: score is 27.246

+48.5 * ((not Seventy) and (not RANDOM66))

1 trees with 3 leaves: score is 25.307

+80 * ((not RANDOM63) and ((not RANDOM70) and RANDOM83))

1 trees with 4 leaves: score is 23.222

-83.9 * (((Seventy or (not RANDOM7)) or RANDOM66) or (not RANDOM84))

1 trees with 5 leaves: score is 20.434

-72.2 * (((Seventy or (not RANDOM35)) or (RANDOM63 and RANDOM96)) or (not RANDOM84))

1 trees with 6 leaves: score is 20.283

-83.6 * (((RANDOM63 and RANDOM96) or Seventy) or ((RANDOM78 and RANDOM33) or (not RANDOM79)))

Linkage decay multiple-fit model for 60% similar to *VRN-H1*:

> fitB

1 trees with 1 leaves: score is 29.004

-22.9 * Sixty

1 trees with 2 leaves: score is 27.33

-56.4 * (RANDOM63 or RANDOM70)

1 trees with 3 leaves: score is 25.307

-80 * (RANDOM63 or (RANDOM70 or (not RANDOM96)))

1 trees with 4 leaves: score is 24.156

+74 * (((not RANDOM70) and RANDOM96) and ((not RANDOM63) or (not RANDOM78)))

1 trees with 5 leaves: score is 20.596
 -82.7 * (((RANDOM63 or RANDOM70) and (RANDOM66 or RANDOM78)) or Sixty)
1 trees with 6 leaves: score is 19.296
 -82 * ((RANDOM66 and ((not RANDOM47) or RANDOM78)) or (Sixty or ((not RANDOM48) or (not RANDOM84))))

Linkage decay multiple-fit model for 50% similar to *VRN-H1*:
> fitB
1 trees with 1 leaves: score is 29.723
 -19.5 * Fifty
1 trees with 2 leaves: score is 27.33
 -56.4 * (RANDOM63 or RANDOM70)
1 trees with 3 leaves: score is 25.307
 +80 * (((not RANDOM70) and (not RANDOM63)) and RANDOM96)
1 trees with 4 leaves: score is 24.156
 +74 * ((RANDOM96 and (not RANDOM70)) and ((not RANDOM63) or (not RANDOM78)))
1 trees with 5 leaves: score is 21.83
 -63.6 * (((RANDOM78 and RANDOM63) or (RANDOM70 and RANDOM29)) or Fifty)
1 trees with 6 leaves: score is 19.755
 -71.2 * ((RANDOM70 and ((not RANDOM79) or RANDOM96)) or ((RANDOM78 and RANDOM33) or Fifty))

Linkage decay multiple-fit model for 40% similar to *VRN-H1*:
> fitB
1 trees with 1 leaves: score is 29.969
 -19.9 * RANDOM70

1 trees with 2 leaves: score is 27.33

-56.4 * (RANDOM70 or RANDOM63)

1 trees with 3 leaves: score is 25.307

-80 * (RANDOM70 or ((not RANDOM96) or RANDOM63))

1 trees with 4 leaves: score is 24.156

+74 * (((not RANDOM63) or (not RANDOM78)) and (RANDOM96 and (not RANDOM70)))

1 trees with 5 leaves: score is 23.918

-75.1 * ((RANDOM78 and (RANDOM60 or RANDOM34)) or ((not RANDOM96) or RANDOM70))

1 trees with 6 leaves: score is 20.827

+66.3 * ((((not RANDOM63) and (not Forty)) or (not RANDOM78)) and ((RANDOM44 and (not RANDOM29)) or (not RANDOM70)))

VRN-H2 Linkage Decay Series

Multiple-fit model of the 'Dicktoo' x 'Calicuchima'-sib data for *VRN-H2*:

> fitB

1 trees with 1 leaves: score is 29.347

-25.5 * (not VRNH2)

1 trees with 2 leaves: score is 27.33

-56.4 * (RANDOM70 or RANDOM63)

1 trees with 3 leaves: score is 25.307

-80 * ((RANDOM63 or RANDOM70) or (not RANDOM41))

1 trees with 4 leaves: score is 24.626

-77.1 * ((RANDOM63 or (RANDOM70 and RANDOM16)) or (not RANDOM96))

1 trees with 5 leaves: score is 22.787

-79.9 * ((((not RANDOM68) or (not RANDOM6)) or ((not RANDOM20) and RANDOM64)) or RANDOM78)

1 trees with 6 leaves: score is 21.38

-67 * (((RANDOM15 or RANDOM96) and RANDOM70) or ((RANDOM78 or (not RANDOM68)) and RANDOM34))

Multiple-fit model for 90% similar to *VRN-H2*:

1 trees with 1 leaves: score is 29.969

-19.9 * RANDOM70

1 trees with 2 leaves: score is 27.33

-56.4 * (RANDOM70 or RANDOM63)

1 trees with 3 leaves: score is 25.307

+80 * (((not RANDOM70) and (not RANDOM63)) and RANDOM41)

1 trees with 4 leaves: score is 23.222

-83.9 * (((not RANDOM79) or (not NinetyVRNH2)) or (RANDOM66 or (not RANDOM80)))

1 trees with 5 leaves: score is 23.329

-63.4 * ((RANDOM63 and RANDOM78) or ((RANDOM33 or RANDOM28) and RANDOM70))

1 trees with 6 leaves: score is 22.392

+69.4 * (((not RANDOM70) or ((not RANDOM8) and RANDOM48)) and ((not RANDOM78) or ((not RANDOM90) and RANDOM28)))

APPENDIX D

General thoughts and comments

I decided to approach this last part of the dissertation as an 'overview' of my thought processes which led to each part of my dissertation, thereby linking them together as pieces in a grand puzzle. I believe this to be an important part of the dissertation in that it documents (to the best of my knowledge) the outline of my discoveries. Bits of information that were either inappropriate for a publishable chapter or the general introduction are included below. It is hoped that this appendix will help tie the events of discovery together so that everyone reading can see how all the 'pieces' fit together in the finished work.

First, let me state, hop downy mildew was a difficult disease to study. The *Pseudoperonospora* pathogen was very difficult to work with because it is an obligate parasite. The pathogen needs living host tissue to survive and grow. Generating inoculum from oospores *was* not an option. People forgot about the oospore when interest waned because previous attempts to induce germination in the lab had failed (Royle and Kremheller 1981). Therefore, asexual inoculum must be maintained throughout the winter on susceptible plants which were then rotated in and out of mist chambers in an attempt to recreate a continuous infection cycle. The inoculum produced was used to infect hop cotyledons for the production of oospores for experiments during the winter months.

I decided early on to work with field plots. Field plots contain genotypes that are several years old which have many advantages over greenhouse clones. First, the field plots contain big plants. Big plants have many mature leaves giving the researcher many more potential points of infection than greenhouse clones. Second, the plants are maintained by common cultural practices so information on epidemiology is related to infectivity under 'field conditions' which really makes the most sense. Third, the environment is controlled by 'mother nature'. Some may consider that to be a negative. I took it to be a positive in that experimental error can

occur at many stages and levels within an experiment. We try to control potential sources of error as much as we can, hoping that the factors we cannot control do not come back to haunt us in the form of excessive experimental error. By utilizing the outdoor plots correctly (inoculating during light, continuous misting rain), I was able to collect three years of quantitative data for the hop downy mildew resistant phenotype.

Based on my initial inoculation experiments, I learned that I was dealing with a quantitative trait. There was no evidence to suggest qualitative differences among susceptible and resistant genotypes. The phenotype appeared to be continuous. Because it appeared that the hop downy mildew resistant phenotype was quantitative in nature, I was pointed in the direction of association mapping during a committee meeting. The initial method I investigated was an in-silico statistical analysis (Grupe et al. 2001). However, this method was criticized because the results were not reproducible (Chesler et al. 2001) and there were many questions concerning the experimental design (Darvasi 2001).

These concerns about the in-silico method appeared valid, so I decided to investigate other potential methods which could be used in my data analysis. After rereading my introductory statistics, I realized that logistic regression was a method of modeling binary data sets. However, based on the examples I investigated, I was lead to believe that logistic regression could only be run on data with a binary response and unfortunately, I had a continuous response variable. However, I new that I was heading down a productive path. Next, I performed a search in Google™ for logistic regression and stumbled onto something called logic regression.

Logic regression was a search and model-building algorithm written for R statistics. After reviewing the logic regression manual provided by R statistics, I realized this software allowed for the modeling of a non-binary response as a function of binary predictors. After reading Ruczinski et al. (2003), I realized he came up with a statistical algorithm that combines Boolean algebra with standard statistical modeling and that logic regression allowed for a continuous response variable.

I hypothesized that logic regression might (due to Boolean algebra and creation of logic groups) very well identify complex genetic interaction. While using logic regression on my hop data set, I quickly learned that the program forms logic groups prior to running the linear regression. These logic groups consist of Boolean combinations of binary covariates separated by logical operators ('and', 'or' and 'not') and these logic groups are independent of the model. Therefore, no model assumptions are implicit in the formation of these groups. By combining groups of Boolean expressions in a well-defined search space, the search algorithm steps through this space searching for the best combination of Boolean expressions which most accurately describes the relationship between the predictors and the response. I became really excited, but there was no way to prove that logic regression would work using my hop data set. I needed a well-characterized epistatic interaction to test my hypothesis.

Unfortunately, well-characterized epistatic interactions are quite rare. Fortunately for me, the barley research group here at Oregon State University had the perfect model system to test my hypothesis. The *VRN-H1/VRN-H2* epistatic interaction for vernalization response in barley was exactly what I needed to prove my hypothesis. Barley vernalization is one of the best-studied epistatic interactions in the world (Takahashi and Yasuda 1971; Yan et al. 2004; Szűcs et al. 2007). It's a classic dominant/suppression interaction where the product of gene one normally suppresses a second locus (gene two), unless a mutation at the second locus prevents binding of the repressor allowing for the expression of gene two. In barley, dominant *VRN-H2* normally inhibits the expression of recessive *VRN-H1* alleles resulting in vernalization sensitivity unless a mutation (at the first intron of *VRN-H1*) removes the repressor binding site resulting in a dominant *VRN-H1* which is vernalization-insensitive. Because my dissertation focus was hop downy mildew, I had to justify the inclusion of the logic regression study on barley vernalization. The justification for its inclusion was a simple as cracking open a bottle of beer. Barley and hops are two main

ingredients in beer. Therefore, I had an industry-based connection with beer and a purely mathematical/statistical connection for DNA data analysis in general.

After running the vernalization data in a logic model, I quickly realized that logic regression worked. The software was able to identify the *VRN-H1* and *VRN-H2* interaction as the most likely model with no spurious markers added to the model. In the course of running the data set, I realized that logic regression appeared to identify the hypothesized genetic model used to explain the interaction of *VRN-H1* and *VRN-H2* in barley (Yan et al. 2004; Szűcs et al. 2007). In the formation of logic groups, the inclusion of operators 'and', 'or' and 'not' is suggestive of dominant and recessive genetic terminology. When the 'Dicktoo' x 'Calicuchima'-sib data resulted in a model with a score of 12.47 and the equation: [+74.9 * (*VRN-H2* and (not *VRN-H1*))]. Not only did logic regression identify *VRN-H1* and *VRN-H2* as the two markers that best describe the interaction of predictors with response, it also nailed the proposed genetic model by suggesting that the true model consists of the epistatic interaction of *VRN-H2* and the 'conjugate' or recessive form *VRN-H1* ('not' *VRN-H1*). This revelation was totally unexpected! The results of this work showed that logic regression worked in the identification of high order epistatic interaction. In addition, because Boolean logic groups are formed prior to running the regression analysis, the interactions are not limited by model assumptions. Therefore, logic regression may be able to distinguish between dominant and dominant/suppression from additive components in a model.

After my defense during editing of the logic regression chapter, I realized that the impact of my discovery was not as great as I previously thought. I was told by Pat Hayes that there are other software packages out there that are able to identify epistasis and what I have discovered was *nothing new*. This was a cause of great concern for I knew that logic regression was the most correct method for modeling epistasis, but I needed a way to prove it. The simulated data set meant that I had very few comparative options. I decided to run my vernalization data sets in the association mapping software TASSEL-GLM. TASSLE-GLM uses simple linear regression to

identify potential marker associations with a phenotypic response. Because TASSEL-GLM required population structure covariates in their model, I tricked the software into thinking I had a single population by setting up a population Q-matrix consisting of Q_1=1 and Q_2=0. This allowed me to run TASSEL-GLM on my data sets.

When I analyzed the data, I was *shocked*, but not surprised to see that TASSEL identified three randomly generated markers as being associated with the days to flowering phenotype! I knew right then that this was going to be a BIG discovery. In order to confirm my results, I decided to run the data sets in SAS proc GLM. Sure enough, those results were the same as those from TASSEL. Next, I decided to perform a full model containing ALL the statistically significant markers and their interactions. The full model ANOVA for the 'Dicktoo' x 'Calicuchima'-sib data showed that although the single marker association of RANDOM 70 was significant, the marker was found to be insignificant in the full model containing interactions. However, the 'Dicktoo' x 'OWB-D' data revealed two markers (RANDOM 46 and RANDOM 58) to be associated with the days to flowering phenotype. When a full model was run and analyzed, RANDOM 58 was found to have statistically significant interaction with *VRN-H1*! I now had proof that logic regression worked better at identifying epistasis when compared to general linear modeling. This discovery should increase the impact factor of my paper, but only time will tell whether the scientific community will accept my position concerning the use of Boolean logic for identification of epistasis.

Unfortunately, logic regression has a few drawbacks. First, because logic regression uses Boolean algebra, the DNA must be dominantly coded. Second, logic regression is a fixed effects model. Therefore, the results pertain to only those individuals within the data set. Third, logic regression output for a QTL analysis can be difficult to interpret, especially when there are deviations from the expected response. When analyzing potentially 'noisy' data sets, it is up to the user to decide what constitutes an unacceptable data set (unacceptable in terms of noise). We can use the null model test to test for signal in the data and we can chart single-fit model

scores to get an idea of the coefficient of variation (CV) in the over-fit models, however, these statistical tests are merely tools to help the researcher make appropriate choices. Many more studies need to be performed using logic regression to determine whether the mean CV for single-fit models can be used as a diagnostic test in the identification of potentially troublesome data sets.

I rediscovered the hop downy mildew oospore and it was quite by accident. It happened while I was walking in one of the research plots at the hopyard. I looked closely at the leaves and they had numerous late-season downy lesions. These lesions looked very similar to those described by Hoerner (1949). Hoerner (1949) reported that there are oospores in the late-season downy mildew lesions. I took some material back to the lab and worked it with a mortar and pestle, put the material through a series of wet sieves and I found oospores in that material. After learning oospores could be produced in abundance in hop cotyledons (Jones 1932b), I began a year and a-half investigation of the oospore. During that time period, I was fortunate to capture multiple germination events on laboratory wet mounts.

Unfortunately, germination events showing a classic germ tube were admittedly quite rare. In addition, I was unable to obtain infection when I carried out an experiment on infective capability of hop seedlings using oospores as an inoculum source. I knew going into that experiment that it had a very high chance for failure with a very small chance for success. I knew this because the disease is over 100 years old and none of the researchers in all that time have claimed success in showing the oospore to be an infective agent for the disease. When the seedling experiment failed, all I had were observations of what appeared to be germination events and counts for oospore numbers in hopyard soil. Although I was disappointed that I could not ‘prove’ the oospore was an infective agent, I did have good evidence that they appeared quite capable of germination. With this information in hand, I went about reading the journals to see whether there was anything that researchers might have missed that would contribute to a better chapter on the oospore.

I knew that in addition to my germination pictures I had data on potential viability in my oospore soil samples. However, after counting the number of oospores in my samples, I realized that something was not quite right. I was getting too many false-positives and too few clear oospores. I immediately knew that there was something wrong with the MTT stain as a viability test. Most researchers stated they could not explain the reason for false-positive reactions and that this percentage needed to be subtracted out of the data (Nelson and Olsen 1967). Others suggested that the false-positives were the result of mycoparasites (Van der Gaag and Frinking 1997). However, this line of reasoning did not sit well with me. I told myself this is not correct and that there had to be a more simple reason. I turned to a review article on tetrazolium staining and believe that I found the answer I was looking for. Altman (1974 and 1976), produced two excellent reviews on the chemistry of tetrazolium salts. After reading the title of the first review article, I knew that all the pathologists missed something crucial. The title of Altman (1974) was, "Studies on the reduction of tetrazolium salts. III. The products of chemical and enzymatic reduction." Altman mentioned a chemical reduction of tetrazolium in addition to the well known enzymatic reduction. I thought to myself, "I bet NADH and/or NADPH is involved in this chemical reduction which Altman mentioned in the title of his paper." NADH and NADPH are two very common reducing agents within cells and because they are not a protein, should survive autoclaving intact. I went home and starting reading the review paper. After a little while, I found the section on chemical reduction. In that section of the review, Altman tells the readers that MTT should not be used as a diagnostic stain for viability because tetrazolium salts are readily reducible in the presence of the reducing agent NADH (Altman 1974). I felt vindicated after making this discovery. Here I discovered something that all the previous pathologists missed and the answer was right in the title of the publication cited in most of the prior studies on oospore viability!

Based on this new information, I was able to put together a better chapter on oospore germination and MTT staining. Although I was unable to prove infective

capability of the hop downy mildew oospore, I was still able to come up with new protocols for the extraction, isolation and recovery of oospores from soil and plant tissue, I captured multiple germination events and photographed them and I discovered something that all the previous reports missed (MTT cannot be used as a viability stain). Based on my observations, I believe that the hop downy mildew oospore is quite capable of germination. I hypothesize that the oospore may germinate within the plant at the break of winter dormancy. This scenario would help explain why an infected bud grows into a downy mildew diseased spike while a bud right next to an infected shoot develops normally. Early on, Ware (1929) was unable to provide sufficient evidence to support his claim but remained steadfast that the disease over-wintered as mycelium in the dormant crown of the hop plant. However, oospores which survive within the dormant buds have a huge selective advantage when compared to oospores in the soil. Oospores within the plant are protected from the direct onslaught of winter. Once the hop plant breaks dormancy, all the oospore has to do is germinate within the bud and the disease cycle can start anew.

With less than six months of funding remaining, I challenged myself to complete the hop downy mildew association mapping chapter of my dissertation and to make it a chapter worthy of publication. As mentioned earlier, I did not like the approach taken by Grupe et al. (2001). It was clear to me that there were too many problems and too few solutions with this approach to association mapping. Logic regression was a viable alternative, however, the software was still very new and I only just recently showed that it was a viable alternative for AFLP marker analysis, not to mention, I did not (as of yet) run the comparative analysis with general linear modeling. Therefore logic regression was still 'too new' and wouldn't help me in my association mapping chapter.

After going to numerous department seminars on association mapping, I learned the maize group at Cornell was using the software TASSEL in a mixed-model format for association mapping. I learned that TASSEL has many modeling options and one method, a mixed effects model, takes population structure and kinship into

account in an attempt to limit type I and type II errors (Yu et al. 2006). After contacting the Buckler group, I learned that TASSEL could be used with my AFLP data set by treating the AFLP data as a string length polymorphisms. Yu et al. (2006) reported that the software Spagedi was used to determine estimates of kinship and the software Structure was used to determine population structure both of which were subsequently added to the mixed-model in TASSEL. The Buckler lab suggested that I read Hardy (2003) as that paper reported on new F-statistics which were being used to derive kinship and relationship coefficients with dominant molecular markers such as AFLP. After reading the paper, I realized that it was now possible to perform a mixed-model analysis on my hop downy mildew AFLP data set. I quickly learned how to use the software Structure v2.1 and Spagedi v1.2 and Hickory v1.0. I knew that I had to first determine whether there was any population substructure in my AFLP data set. Previous reports (Murakami 2000; Seefelder et al. 2000; Jakse et al. 2001) suggested that there were two main clusters in hop that could be further broken down into 3 to 6 subclusters. However, when I ran Structure v2.1, I was unable to get similar clustering as that previously reported.

After running all my analyses and comparing the output, I decided to use a traditional genetic distance cluster analysis as these results made the most sense both biologically and with the pedigree information available on hop. With the population structure identified, I ran the AFLP data set in the genetic analysis software Hickory. Hardy (2003) reported that it was possible to obtain an inbreeding coefficient using Bayesian inference. I knew that I needed an 'assumed' inbreeding coefficient in order to obtain kinship estimates for my AFLP marker data set. Holsinger et al. (2002), reported on estimating population structure (F-statistics) with dominant markers like AFLP. After I read the paper, I quickly learned how to format my data set for the software Hickory. After 45 minutes, I got a value for the mean inbreeding coefficient necessary for calculating estimates of kinship using the Hardy (2003) method.

The kinship estimates allowed me to calculate the variance components for narrow-sense heritability of the downy mildew resistant phenotype in addition to their

use in a mixed-model approach to association mapping. The broad-sense heritability of 76% and a narrow-sense heritability of 49% suggested to me that the F-statistics developed by Hardy (2003) might actually work. If the narrow-sense heritability was greater than the broad-sense heritability, then the results would have been viewed with extreme caution. However, looking at my phenotypic distribution and the broad-sense heritability results, I realized the estimated narrow-sense heritability results were probably correct. Since these techniques are the only ones currently available for use with dominant markers and the initial results 'looked right', I continued on with my analysis.

I incorporated the population structure and kinship estimates into the mixed-model in TASSEL. The results revealed 43 AFLP markers to be statistically associated with the hop downy mildew resistant phenotype. That was approximately 9% of the 492 markers in my data set. The R^2 of the markers themselves was quite low ranging from less than 1% to 23%. The low R^2 of the markers makes perfect sense considering the hop downy mildew resistant phenotype is quantitative. Many molecular markers linked to genes, each contributing a little bit to the expressed phenotype. Based on the heritability estimates for the phenotype and the mixed-model results, some form of family-based genotypic recurrent selection or marker-assisted selection will be needed to help introduce many levels of resistance to this devastating pathogen.

Looking back at the process of discovery, it becomes apparent that each step in the process led directly to the next step. Often the steps seemed haphazard and disconnected at the time, but looking back it all fit together. Would I have discovered logic regression and its potential use in modeling of epistasis without first learning about the in-silico methods of Grupe et al. (2001) and being dissatisfied with that approach? I would have to say probably not. Would I have pushed myself to learn mixed-models for use in association mapping if I did not first learn about logic regression? I can safely say, maybe, but then again, maybe not. Learning logic regression gave me the confidence I needed so that I could learn the additional theory

and software necessary for a mixed-model approach. There was a good chance that I would not have learned mixed modeling without the prior confidence boost from logic regression. I know people are going to be curious (as I would be) and ask, "What were you thinking when you discovered Boolean logic and where did the idea that epistasis might best be modeled using Boolean logic? All I can say is that I did what I was told to do, solve the problem, whatever problem was in my path at the time. You cannot look at a problem from just one dimension or one area of thought and/or expertise. I was able to make these important discoveries because I was well prepared to make the discoveries. It took a lot of reading and a lot of coursework, but in the end, I was prepared to pick up bits of information that otherwise might have gone unnoticed. I always joked with a friend, Dr. Janssen, and I told him, "Yeah I have the gift of observation, but I wish I was better at math or computers instead." I always thought (out of all the gifts that people have in this world) I got one that stinks. It wasn't until Dr. Janssen told me, "Tom, you have a wonderful gift, many great scientists are really good at mathematics, many are good with computers and many are good in the lab, but many of these people cannot see the nose at the end of their face. You have the ability to see things that others miss and you have shown that you have this ability numerous times and that is a good gift." I will never forget what Dr. Janssen told me.

The rediscovery of the hop downy mildew oospore, well, that was just fortuitous on my part. This incident of 'dumb-luck' reminds me of the movie Forrest Gump. I only did what you told me to do. I remember Walt told me that I have to make my own luck and put myself in a position to make a discovery. How do you do that, again, you go to the library and read! When you think you are finished, you need to read some more. Every single discovery I made was a direct result of reading. I cannot overemphasize the importance of reading in scientific discovery. It is, in my opinion, the most important step in the whole process. I am no different than any other average scientist. I just really like to read A LOT! The discoveries are out there for everyone to find. The people that make the discoveries are most often the ones

that put themselves in the position to make those discoveries. Thank you Walt, I will never forget what you told me and I hope to teach others these time tested techniques 'of discovery' in the future.

I am glad to have had to opportunity to help the hop industry. Scientific achievement is not always readily apparent to those most affected by a problem. As you can see by reading these thoughts and comments, scientific discovery is an incremental process where each discovery leads to new discoveries. Will science ever be able to defeat hop downy mildew? I cannot answer that. However, I do know that without making an attempt, failure is a certainty.