Louisiana State University
LSU Digital Commons

LSU Doctoral Dissertations | Graduate School

2005

Associating molecular markers with phenotypes in sweetpotatoes and liriopogons using multivariate statistical modeling

Thomas Mwamburi Mcharo
Louisiana State University and Agricultural and Mechanical College, tmchar1@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation
Mcharo, Thomas Mwamburi, "Associating molecular markers with phenotypes in sweetpotatoes and liriopogons using multivariate statistical modeling" (2005). *LSU Doctoral Dissertations*. 3274.
https://digitalcommons.lsu.edu/gradschool_dissertations/3274

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contactgradetd@lsu.edu.

ASSOCIATING MOLECULAR MARKERS WITH PHENOTYPES IN SWEETPOTATOES AND LIRIOPOGONS USING MULTIVARIATE STATISTICAL MODELING

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Horticulture

By

Thomas Mwamburi Mcharo
B.Sc., University of Nairobi, 1992
M.Sc., University of Nairobi, 1995
May 2005

DEDICATION

To my parents Mwakio and Wakesho Mcharo, my sisters Maguwa, Mkawasi, Wakufwa and Mboli and my brother Mcharo. You all love me unconditionally, I couldn't ask for more. God bless you mightily.

ACKNOWLEDGEMENTS

I wish to express my profound gratitude to Dr Don R. LaBonte who has been my advisor, my mentor and my sponsor for my PhD studies. He took me up in faith and I've been honored to reciprocate to a man who in more ways than one has been a big brother to me during my stay in Louisiana State University. Thank you very much Dr LaBonte. I also highly appreciate the constructive advice of Drs David Picha, Gerald Myers and James Oard who have been my committee members. I can not forget Dr Mike Cannon and his staff at the sweetpotato research station at Chase, Louisiana, for sacrificially assisting me in my field work.

I am grateful to three wonderful ladies, Mary Bowen, Gloria McLure and Mary Hoy, who greatly assisted me in my lab work. Words are not enough to express my appreciation. You mean a lot to me. I wish to express my appreciation to Charalambos Kokkinos, Cecilia McGregor and Gabriel Aluko, my three fellow graduate students who guided me through the rigmarole of molecular genetics research and to my indefatigable student worker, Michael Cavalier.

This work would not have been complete without the generosity of Dr Regina Kapinga of International Potato Center (CIP), Africa region and Daniel Reynoso of CIP, Lima Peru, for supplying us with the East African and CIP test populations, respectively. Thank you very much.

Finally I give glory to my creator, the Almighty God, who chose to reveal to me all the knowledge generated through my studies. He strengthened me and has been my source of courage through my studies. I hope in some way this work will glorify His name.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

ABSTRACT ix

CHAPTER 1: INTRODUCTION 1
- 1.1 Literature Cited 11

CHAPTER 2: MOLECULAR AND MORPHOLOGICAL INVESTIGATION OF ORNAMENTAL LIRIOPOGONS 15
- 2.1 Introduction 15
- 2.2 Materials and Methods 16
- 2.3 Results and Discussion 19
- 2.4 Literature Cited 24

CHAPTER 3: LINKING QUANTITATIVE TRAITS WITH AFLP MARKERS IN SWEETPOTATOES USING DISCRIMINANT ANALYSIS 25
- 3.1 Introduction 25
- 3.2 Materials and Methods 27
- 3.3 Results and Discussion 31
- 3.4 Literature Cited 37

CHAPTER 4: ASSOCIATING MOLECULAR MARKERS WITH VIRUS RESISTANCE TO CLASSIFY SWEETPOTATO GENOTYPES 39
- 4.1 Introduction 39
- 4.2 Materials and Methods 42
- 4.3 Results and Discussion 44
- 4.4 Literature Cited 50

CHAPTER 5: MOLECULAR MARKER VARIABILITY FOR SOUTHERN ROOT-KNOT NEMATODE RESISTANCE IN SWEETPOTATO 53
- 5.1 Introduction 53
- 5.2 Materials and Methods 55
- 5.3 Results and Discussion 59
- 5.4 Literature Cited 71

CHAPTER 6: COMPARISON OF LOGISTIC REGRESSION AND DISCRIMINANT ANALYSIS IN ASSOCIATING AFLP MARKERS WITH TOTAL SUGAR AND ß-CAROTENE AMONG SWEETPOTATO CLONES 74
6.1 Introduction 74
6.2 Materials and Methods 75
6.3 Results and Discussion 80
6.4 Literature Cited 96

CHAPTER 7: SUMMARY AND CONCLUSIONS 99

APPENDIX 1: ADDITIONAL DATA TABLES 106

APPENDIX 2: DATA ANALYSIS PROTOCOLS 109

APPENDIX 3: PERMISSION LETTERS FROM JOURNAL PUBLISHERS 112

VITA 117

LIST OF TABLES

3.1 List of sweetpotato clones obtained from USDA/NPGS repository 28

3.2 AFLP predictor markers in sweetpotato as selected by the STEPDISC procedure 33

3.3 Rate of correct classification of 58 training clones of sweetpotato into dry matter group after cross-validation using nearest neighbor in discriminant analysis...35

3.4 Rate of correct classification of 10 sweetpotato test clones into dry matter group using discriminant analysis 35

4.1 QTL analysis for DNA markers associated with ≥10% resistance to sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato 47

4.2 STEPDISC selection for molecular markers associated with ≥10% resistance to sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato 47

4.3 Logistic regression selection for molecular markers associated with ≥10% resistance to sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato 49

4.4 Partition for the Hosmer and Lemeshow test for sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato...49

5.1 Descriptive statistics for gall rating, egg count and logarithm transformed egg count for the LSU population...60

5.2 Descriptive statistics for gall rating, egg count and logarithm transformed egg count for the LSU population...60

5.3 STEPDISC selection for DNA markers associated with resistance to Southern root-knot nematode in two sweetpotato populations.....................66

5.4 Logistic regression selection for DNA markers associated with resistance to Southern root-knot nematode in two sweetpotato populations.....................67

5.5 Rate of correct classification of sweetpotato clones into nematode resistance groups after cross-validation in discriminant analysis 67

5.6 Discriminant analysis marker selection statistics for a model that achieved 100% correct classification in the LSU population 69

5.7 Discriminant analysis marker selection statistics for a model that achieved 100% correct classification in the CIP population 70

6.1 Total sugar content and resultant phenotypic classification of sweetpotato clones from an F_1 polycross of 'Beauregard', 'Excel', 'L94-96', L'86-33', 'L89-110' and 'L96-117' 82

6.2 Logistic regression selection for DNA markers in sweetpotato associated with total sugar content 83

6.3 STEPDISC selection for DNA markers in sweetpotato associated with total sugar content 85

6.4 Rate of correct classification of 73 clones of sweetpotato into total sugar group after cross-validation in discriminant analysis 85

6.5 Rate of correct classification of 73 clones of sweetpotato into total sugar group and the AIC model fit statistic for logistic regression 85

6.6 Discriminant analysis marker selection statistics for a model that achieved 100% correct classification for the total sugar trait 86

6.7 Hunter color value *a* and phenotypic classification for ß–carotene content of sweetoptato clones from an F_1 polycross of 'Beauregard', 'Excel', 'L94-96', L'86-33', 'L89-110' and 'L96-117' 88

6.8 Logistic regression selection for DNA markers associated with β-carotene content in sweetpotato 89

6.9 Rate of correct classification of 73 clones of sweetpotato into ß -carotene group and the AIC model fit statistic for logistic regression 90

6.10 Probability levels and classification of sweetpotato clones into ß -carotene groups using 12 markers in logistic regression...93

6.11 STEPDISC selection for DNA markers associated with ß -carotene content in sweetpotato...94

6.12 Rate of correct classification of 73 clones of sweetpotato into ß -carotene group after cross-validation in discrimination analysis 94

6.13 Discriminant analysis marker selection statistics for a model that achieved 100% correct classification for the ß -carotene trait in the population 95

LIST OF FIGURES

2.1 Photomicrographs showing flower morphology of *Liriope* spp. and *Ophiopogon* spp. (a) ‘Royal Purple’ (*Liriope* spp.) showing a hypogenous ovary insertion. (b) ‘Aztecgrass’ (*Ophiopogon* spp.) showing a perigynous ovary insertion. Bar = 1.0 mm 21

2.2 UPGMA based dendrogram of 18 *Liriope* spp. and *Ophiopogon* spp. cultivars. The numerical scale indicates increasing genetic similarity 23

3.1 Dendrogram using the DIST coefficient and single-link clustering method between high and low root dry matter content sweetpotato groups. Upper cluster consists of high dry matter (NPGS accessions 1067 through 585093) and lower cluster consists of low dry matter clones (NPGS accessions 564157 through 344124) 32

5.1 Frequency of log total root-knot nematode eggs for the LSU sweetpotato genotypes 61

5.2 Frequency of log total root-knot nematode eggs for the CIP sweetpotato genotypes 62

ABSTRACT

Two horticultural crops, the ornamental liriopogon and the sweetpotato [*Ipomoea batatas* (L.) Lam.], were analyzed for morphological, quantitative and molecular marker variation using Amplified Fragment Length Polymorphism (AFLP) and various multivariate statistical techniques. Ornamental cultivars in genera *Liriope* and *Ophiopogon* were analyzed for relatedness using AFLP marker data and statistical clustering methods. Marker data did not substantiate the separation of these two genera. Greater than 95 % of the total genetic variability present was attributed to within group variation ($P \leq 0.05$). Trait-linked molecular markers were identified using Quantitative trait loci (QTL) analysis, logistic regression and discriminant analysis in the studies involving sweetpotato. The traits studied included dry matter content, virus disease resistance, root-knot nematode resistance, sugar content and β-carotene content. Analysis of molecular variance found significant ($P < 0.001$) differences between two phenotypic groups from unrelated genotypes for dry matter data. Using 14 markers selected through discriminant analysis the phenotypic grouping was validated with a zero error rate. Eighty-seven F_1 sweetpotato genotypes from a cross of 'Tanzania' and 'Wagabolige' landraces were used to generate AFLP and random amplified polymorphic DNA (RAPD) marker profiles for this study. One AFLP marker linked to sweetpotato chlorotic stunt closterovirus resistance and one RAPD marker linked to sweetpotato feathery mottle virus resistance previously identified by traditional mapping strategies were selected plus new markers. Two diverse F_1 populations of sweetpotato were used to identify and select markers suitable for identification of plants possessing a resistant reaction to southern root-knot nematode race 3 [*Meloidogyne incognita* (Kofoid and White) Chitwood]. Results for plant nematode resistance indicated a binomial distribution among the genotypes for population 1 and a normal distribution for population 2. A comparison of the

power of discriminant analysis models for southern root-knot nematode resistance class prediction achieved 88% classification efficiencies. An F_1 population of 73 clones consisting of parents and half-sibs was grouped into 2 phenotypic classes based on their sugar and ß-carotene content. Logistic regression and discriminant analysis selected meaningful markers that had significant associations with each of the traits. These results validated discriminant analysis and logistic regression as meaningful trait-linked marker selection methods.

CHAPTER 1: INTRODUCTION

Results presented in this study involve two horticultural crops, namely the ornamentals that are collectively called liriopogons and the sweetpotato. The multiple projects that fall under various chapters all involved the use of Amplified Fragment Length Polymorphism as the molecular marker generating system and different multivariate statistical modeling techniques. Molecular marker information was mined through the use of multivariate statistics to elucidate any underlying relationships between the various phenotypic groups and the DNA markers.

Liriopogons are the most important clonal ground covers sold by the nursery industry in the southeastern United States. The term liriopogon collectively represents ornamentals in the genera *Liriope* and *Ophiopogon*, a.k.a., monkeygrass, and denotes an underlying morphological similarity between the two genera. For example, the popular 'Aztecgrass' is commonly referred to as a *Liriope* sp., but in actuality it is identified taxonomically as an *Ophiopogon* sp. (Adams, 1989). This similarity has led to industry-wide misidentification of genera and putative clones possessing multiple trade names (Franz, 1994).

Molecular characterization is an approach useful in assigning cultivars or clones to genera and identifying unique banding profiles capable of fingerprinting cultivars for definitive identification. Single accessions of the aforementioned genera were included in a broader molecular and morphological systematic investigation of the family Convallariaceae by Rudall et al. (2000). Although Rudall et al. (2000) found significant morphological differentiation between the two genera, they did not find significant molecular differences. Cutler (1992) and Rudall et al. (2000) also suggested that *Liriope* and *Ophiopogon* did not differ taxonomically and hence belong in the tribe Ophiopogonae.

The objectives of this portion of the study were to use a combination of morphological descriptors and DNA fingerprinting to differentiate common industry cultivars to the proper genera and to describe the extent of genetic similarities among these cultivars. I was also interested in optimizing molecular techniques that would be useful for the larger portion of the thesis research involving sweetpotatoes.

The sweetpotato [*Ipomoea batatas* (L.) Lam] is a dicotyledonous plant, which belongs to the family Convolvulaceae and it is usually considered the only species of *Ipomoea* of major economic importance (Hall and Phatak, 1993). A highly heterozygous crop, sweetpotato is a natural hexaploid (2n=6x=90) occurring in the Convolvulaceae (Jones et al., 1986). Domestication and artificial hybridization and selection by man, natural hybridization and mutations over time have all resulted in a large number of cultivars. Woolfe (1992) reports that the level of diversity in the sweetpotato is higher than in root crops like cassava and yam. Cultivars differ from one another in the color of the root skin or flesh, in the size and shape of the roots and leaves, in the depth of rooting, the time to maturity, texture of cooked roots and resistance to biotic and abiotic stress.

To incorporate these attributes, various breeding techniques have been applied since antiquity. In centers of diversity, natural cross-pollination may have occurred to contribute to a wide array of genotypes in one location. Sweetpotato clones also naturally mutate for traits like root and skin color, and leaf and vine characteristics (Villordon and Labonte, 1996). The most desirable genotypes among these were selected by farmers and used until new ones arose over the centuries (Hall and Phatak, 1993). This mix of directed and undirected breeding has resulted in a large diversity of traits. Despite the existing diversity, the complicated nature of sweetpotato genetics did not lend itself to easy manipulation by early breeders. The net effect has been that

controlled crossing programs were slower to develop for sweetpotato than for crops with simpler systems.

It has been recognized that the inheritance of most traits in sweetpotato is quantitative and reliable genetic conclusions have therefore been derivatives of variate deviations within a population (Jones, 1986). The first systematic effort for genetic recombination in sweetpotato within the United States was initiated at Louisiana State University in the late 1930s using pedigree breeding (Miller, 1937). However this method results in inbreeding with its attendant debilitating effects on traits of interest i.e. increasingly inferior clones in subsequent generations. In 1965, Jones proposed mass selection for sweetpotato improvement as described by Jones et al. (1986). This method has the advantage of rapidly aggregating desirable genes. Collins (1992) agrees with this and states, "sweetpotato improvement through traditional breeding methods has been successful in the recent past, mainly due to development of a quantitative genetic methodology of selection suited for a hexaploid, highly heterozygous crop like sweetpotato".

Regardless of the breeding method and statistical procedures applied, high selection intensity at each stage of the breeding scheme is necessary to ensure progress. Furthermore, genetic gain is more likely to be significant if the diversity and level of genetic variability of desirable traits is sufficient. Highly heritable traits are easier to select for than those with low levels of heritability. For example Collins (1977) and Jones (1986) have reported that resistance to Fusarium wilt is a highly heritable character ($h^2 = 0.85$ to 0.90) with high levels of resistance available and evaluation is consistent and reliable from test to test. Consequently, gains in resistance to Fusarium wilt have been made in commercial cultivars.

In contrast, a trait like yield is highly influenced by the environment and hence yield evaluation for advanced clones must be done over several seasons and locations before release.

In assessing yield levels, the significant genotype-environment interaction reduces the reliability of yield stability estimates hence a minimum number of locations and seasons is necessary for conclusions to be made with confidence. Even though knowledge of the magnitude of genotype by environment interaction assists the breeder in making breeding progress, the progress has so far been slow and expensive. This is partly because classical breeding technologies in sweetpotato have yet to be supplanted with molecular approaches. In addition, its polyploid nature, genetic incompatibility and the high levels of mutation within the species have complicated conventional breeding of the sweetpotato (La Bonte, 2002).

Recent advances in plant biotechnology are offering novel technologies that may greatly reduce breeding costs and the time needed to develop a variety. Collins (1992) postulated that these techniques could provide sweetpotato breeders with very practical and accurate evaluation methodologies. Such technologies could prove to be more precise and easier to use in tagging specific genes or introducing genetic variation into germplasm with a narrow genetic base. They could also be used to introduce new traits directly into commercial cutivars (Connolly et al., 1994.).

Many of these technologies involve DNA characterization, in particular DNA fingerprinting using genetic molecular markers. Genetic markers provide a valuable tool for genetic analysis and breeding research in crops and they facilitate studies of genome organization, mating systems, genetic diversity and phylogenetic analysis (Arcade et al., 2000). For example Random Amplified Polymorphic DNA (RAPD) provided a significant advance in molecular marker studies (Williams et al., 1990). However, the RAPD technique has the disadvantage of lack of reproducibility over time (Jones et al., 1997).

The advent of the Amplified Fragment Length Polymorphism (AFLP) technique (Vos et al., 1995) provided a new class of highly polymorphic markers combining both Restricted Fragment Length Polymorphism (RFLP) and Polymerase Chain Reaction (PCR) strategies. Its advantage over RAPDs is its high reproducibility (Jones et al., 1997) and generation of more polymorphic loci per primer. It is particularly amenable for the construction of high density genetic maps. Recent studies have used AFLPs in genetic assays of vegetatively propagated crops like yams, cassava and now the sweetpotato. This study will therefore use the AFLP technique in conjunction with discriminant analysis and logistic regression to link molecular markers with phenotypic traits in order to delineate markers of interest.

In order to identify variation at the molecular level an analysis of molecular variance (AMOVA) is used to detect statistical differences among pre-defined phenotypic groups (Excoiffer et al., 1992). This procedure uses the square distance between molecular marker profiles as data to perform a permutation analysis of variance. Groups that are significantly different are subsequently analyzed using discriminant analysis to determine which markers discriminate between phenotypic groups using molecular marker information. Application of discriminant analysis to a molecular marker data set enables one to determine which markers contribute most to discriminate between groups and then use that information to predict group membership. The researcher can then test the validity of groups based on actual data, to test groups that have been created, or to assign lines into groups. There are two complementary approaches for integration of molecular and agronomic data:

1. Analysis of variance using molecular data to test differences among groups defined according to agronomic data using an AMOVA based algorithm.

2. Selection of informative markers and computation of a classification model for agronomic group membership using discriminant analysis and logistic regression.

Phenotypic grouping for subsequent AMOVA, QTL, discriminant and logistic regression analysis may be based on morphological observations or statistical techniques like cluster analysis. Hair et al. (1998) define cluster analysis as being a group of multivariate techniques that classify genotypes so that each genotype is very similar to others in the cluster with respect to some predetermined selection criterion (variate). The cluster variate is the set of variables representing the characteristics used to compare genotypes in the cluster analysis. The cluster variate then determines the 'character' of the genotypes. Cluster analysis does not estimate the variate empirically but instead uses the variate as specified by the researcher. Cluster analysis seeks to compare genotypes based on the variate, not on the estimation of the variate itself. The resulting clusters then exhibit high within-cluster homogeneity and high between-cluster heterogeneity.

Discriminant analysis (DA) is a multivariate statistical technique that can identify differences among groups of individuals (or treatments) and improve understanding of the relationships among the variables measured within those groups (Cruz-Castillo et al., 1994). Fisher (1936) was the first one to use the technique also known as Fisher's discriminant analysis (Mardia et al., 1979) or canonical variate analysis (Everitt, 1978). DA determines how best to separate or discriminate two or more groups of individuals given measurements of several individuals through simultaneous analysis of several variables. DA finds linear functions of variables that maximally separate two or more groups of individuals while keeping variation within groups as small as possible. This approach distinguishes several uncorrelated discriminant

functions (DF) or canonical variables. DFs are linear combinations of the original variables that best separate the means of groups of observations relative to within group variation (Rencher, 1992).

DA may be used when it is important to separate known groups or *a priori* groupings, and to identify major sources of difference between groups. Such groups may be determined through a statistical procedure like molecular marker profiling or through other clustering techniques. The discriminant analysis will therefore provide information on the discriminatory power of each variable or marker thus determining whether a particular set of variables or markers is useful in separating previously delineated groups (Rojas et al., 2000). It then follows that DA requires *a priori* groupings and the net result is to maximize variation between the groups of individuals and minimize within-group variation of the original variables. While the objective of both DA and multivariate analysis of variance (MANOVA) is to analyze between-group differences of multivariate data from designed experiments, the overall effect of the variables, rather than individual response is of innate value in MANOVA. In contrast DA generates functions that yield relative information on each variable in distinguishing between groups. While DA illustrates visual description of differences between the groups, MANOVA does not (Cruz-Castillo et al., 1994).

DA also serves as a predictive tool, assigning phenotypes in a previously unclustered population into predefined groups. DA may perform this by calculating the Mahalanobis' distance (D^2) which is a squared distance between two groups defined as $D^2_{ij} = (x_i - x_j)'W^{-1}(x_i - x_j)$, where x_i and x_j are the centroids (vectors of means) of the groups. W is the combined estimate of the within group variance-covariance matrix. The best strategy will produce the largest distance between groups, that is, more compact and better defined groups in relation to

the quantitative variables (Franco et al., 1997). The first individual selected into a group is the nearest neighbor to the centroid based on its distance. The procedure is iterated to select the nearest neighbors for each expanding group until the all the individuals have been classified into the predefined groups. The distance of any individual is calculated as an individual's mean value for each of the traits during the season (Woeste et al., 1998). When between population comparisons are made, distances are $\log_{10}$ transformations thus making them conform more closely to the assumptions of ANOVA (Fernandez, 1992).

Discriminant analysis may be used for either parametric or nonparametric traits. In the present study no assumption is made about the marker distribution pattern and hence the nonparametric method is used. Work on discriminant analysis has included investigations into quantitative traits like drought tolerance in Kentucky blue grass (Ebdon et al., 1998) and nutrition content and geographical origin of tea (Fernández-Cáceres et al., 2001). Recent work on molecular genetic analysis using discriminant analysis include that of Capdevielle et al. (2000) who associated microsatellites and agronomic traits in rice. Fahima et al. (2002) who investigated microsatellite polymorphism in wheat also used discriminant analysis in their characterization. Aluko (2003) associated micro-satellites in rice with resistance to sheath blight. Most of the work has involved descriptive discriminant analysis, which is using molecular markers to validate the classification of individuals that were previously classified phenotypically. Predictive discriminant analysis using molecular markers is more challenging in the sense that a reliable model has to be developed based on previously available data and the model has to be efficient in classifying future genotypes of unknown description. Most of the experimental material for such studies is limited to tens or at most a few hundreds of genotypes due to the complexity of obtaining sufficient progeny from crosses. Such limited requisite

material is likely to produce models that are informative but with limited application thus rendering development of predictive models even more difficult. The situation is compounded by the fact that present statistical software do not produce interpretable models for nonparametric analysis. The third chapter on dry matter analysis in this study investigates the possible development of a predictive model using a test sample and a training sample of sweetpotatoes.

Discriminant analysis as a statistical tool for molecular marker selection in agricultural crop was introduced in Louisiana State University by researchers working in Dr James Oard's laboratory within the department of agronomy. These workers, who included Dr James Oard, Monica Balzarini, Fabian Capdevielle and Gabriel Aluko, used discriminant analysis in selecting informative microsatellites for various agronomic traits in rice (Capdevielle, 2000). Monica Balzarini originated the concept of discriminant analysis use in trait linked marker association genetics in crop plants. They also investigated the effect of population structure on genotype classification. Pritchard et al. (1999, 2000) developed a method that uses molecular markers at the allelic level to detect population structure and to associate the markers with phenotypes that define the sub-populations. However it was not possible to use his method because the AFLP markers that I worked with did not have any alleles.

Hair et al. (1998) describe logistic regression as a form of regression in which the dependent variable is a non-metric binary variable. This method analyzes proportions based on the binary event and the proportions are then transformed into odds ratios (Ostir and Uchida, 2000). Odds is the ratio of the probability of an event occurring to the probability of the event not occurring while an odds ratio is a ratio of two odds. The event may group into either one or the other phenotypic group. During the analysis the odds undergo a logit transformation to obtain log odds for a given variable or DNA marker. The log odds for each marker selected by logistic

regression are then included in the group classification model. The ultimate result from logistic analysis in our study was the probability of inclusion into a phenotypic group. According to Ostir and Uchida (2000) the main strength of logistic regression is its ability to handle categorical and continuous explanatory variables simultaneously. However it is limited by the requirements that (i) all genotypes must be in the study the same length of time and (ii) estimates based on small sample sizes are imprecise.

In a study involving boars (*Sus scrofa* L.), Thurston et al. (2002) identified sixteen candidate genetic markers ($P<0.005$) by comparing the AFLP profile with semen freezability using logistic regression analysis. Dunsmuir et al. (2000) also used logistic regression modeling to identify 3 immunohistochemistry markers that were statistically significant predictors of the metastatic status (M-stage, bone metastasis vs no bone metastasis) of prostrate cancer in men. To my knowledge there are no studies involving logistic regression to select molecular markers for traits in sweepotatoes or in plant species.

The section of the study involving sweetpotatoes was thus conceived with the following objectives:

1. Classify the genotypes under study into phenotypic groups and determine the significance of each of the predefined clusters using analysis of molecular variance.
2. Identify the most important DNA markers contributing to variation among the established phenotypic groups and determine the degree of association between the markers and groups using discriminant analysis and logistic regression.
3. Verify the group prediction ability of the developed discriminant and logistic classification functions through use of cross-validation and test populations.
4. To validate markers selected through traditional map making strategies.

1.1 Literature Cited

Adams, G. 1989. Great ground covers. Amer. Nurseryman 170:83-91.

Aluko, G.K. 2003. Genetic Mapping of Agronomic Traits from The Interspecific Cross of *Oryza sativa* L. and *Oryza glaberrima* Steud. PhD dissertation. Louisiana State Univ., Baton Rouge, PhD Diss. ETD-1110103-143019.

Arcade, A., F. Anselin, P. Faivre Rampant, M.C. Lesage, L.E. Pâques, and D. Prat. 2000. Application of AFLP, RAPD and ISSR markers to genetic mapping of European and Japanese Larch. Theor. Appl. Genet. 100:299-307.

Capdevielle, F.M., G.K. Aluko, M. Balzarini, and J.H. Oard. 2000. Application of molecular markers and discriminant analysis to identify rice lines with contrasting phenotypes for agronomic traits, p.216. In: G.S. Khush, D.S. Brar and B. Hardy (eds.) Proceedings of the Fourth International Rice Genetics Symposium, International Rice Research Institute, Los Banos, Philippines (abstr.).

Collins, W.W. 1977. Diallel analysis of sweetpotatoes for resistance to Fusarium wilt. J. Amer. Soc. Hort. Sci. 102:109-111.

Collins, W.W. 1992. Practical applications of biotechnology for sweetpotato improvements, p 514-520, In: Hill, W.A, C.K. Bonsi, and P.A. Loretan (eds). Sweetpotato technology for the 21st century. Tuskegee University, Tuskegee, Alabama.

Connolly, A.G., I.D. Godwin, M. Cooper, and I.H. DeLacy. 1994. Interpretation of randomly amplfied polymorphic DNA marker data for fingerprinting sweetpotato (*Ipomoea batatas* L.) genotypes. Theor. Appl. Genet. 88:332-336.

Cruz-Castillo, J.G., S., Ganeshanandam, B.R. MacKay, G.S. Lawes, C.R.O. Lawoko, and D.J. Woolley. 1994. Applications of discriminant analysis in horticultural research. HortScience 29:1115-1119.

Cutler, D.F. 1992. Vegetative anatomy of Ophiopogoneae (Convallariaceae). Bot. J. Linnean Soc. 110:385-419.

Dunsmuir, W.D., C.E. Gillett, L.C. Meyer, M.P. Young, C. Corbishley, R.A. Eeles, and R.S. Kirby. 2000. Molecular markers for predicting prostate cancer stage and survival. BJU International. 86: 869-878.

Ebdon, J.S., Petrovic, A.M. and Schwager, S.J. 1998. Evaluation of discriminant analysis in identification of low and high water use Kentucky Bluegrass cultivars. Crop Science 38:152-157.

Everitt, B. 1978. Graphical techniques for multivariate data. Heinemann, London.

Excoiffer, L., P.E. Smouse, and J.M. Quattiro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131:479-491.

Fahima, T., M.S. Roder, K. Wendehake, V.M. Kirzhner, and E. Nevo. 2002. Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor. Appl. Genet. 104:17-29.

Fernandez, G.C.J. 1992. Residual analysis and data transformations: Important tools in statistical analysis. HortScience 27:297-300.

Fernández-Cáceres, P.L., M. J. Martín, F. Pablos, and A. G. Gonzáles. 2001. Differentiation of Tea (*Camellia sinensis*) varieties and their geographical origin according to their metal content. J. Agric. Food Chem. 49: 4775-4779.

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics 7:179-188.

Franco, J., J. Crossa, J. Díaz, S. Taba J. Villaseńor, and S.A. Eberhart. 1997. A sequential clustering strategy for classifying gene bank accessions. Crop Sci. 37:1656-1662.

Franz, P.R.1994. A taxonomic research update of cultivated Liriopogons. HortTech. 4:46-48.

Hair, J.F., R.E. Anderson, R.L. Tatham, and W.C. Black. 1998. Multivariate data analysis (5th ed). Prentice Hall. New Jersey. MacMillan Publishing Company, New York, USA.

Hall, M. R and S.C Phatak. 1993. Sweetpotato *Ipomoea batatas* (L.) Lam., p.693-708 In: Kalloo G and B.O Bergh (eds) Genetic Improvement of vegetable crops. Pergamon Press.

Jones, A. 1986. Sweetpotato heritability estimates and their use in breeding. HortScience 21:14-17.

Jones, A., P.D. Dukes, and J.M. Schalk. 1986. Sweetpotato breeding, p 1-35. In: Breeding vegetable crops. Basset M.J. (ed). AVI publishing company, Westport, Connecticut.

Jones, C.J., K.J. Edwards, S. Castaglione, M.O. Winfield, F. Sala, C. van de Wiel, G. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Valzquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. 3:381-390.

La Bonte, D.R. 2002. Molecular biology in sweetpotato genetics: a means of progress, p. 21-29. In: T. Ames (ed.). Proceedings of the first international conference on sweetpotato: Food and health for the future, Lima, Peru. Acta Horticulturae 583.

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

Mardia, K.V., J.T. Kent, and J.M. Bibby. 1979. Multivariate analysis. Academic, London.

Miller, J.C. 1937. Inducing the sweetpotato to bloom and set seed. J. Hered. 28:347-349.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors using a complete RFLP linkage map. Nature 335:721-726.

Pritchard, J.K and N.A. Rosenberg. 1999. Use of unlinked genetic markers to detect population stratification in association studies. Am. J. Hum. Genet. 65:220-228.

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

Rencher, A.C. 1992. Interpretation of discriminant functions, canonical variates, and principal components. Amer. Statistician 46:217-225.

Rojas, W., P. Barriga, and H. Figueroa. 2000. Multivariate analysis of Bolivian quinua germplasm. Plant genetic resources newsletter. No 122:16-23.

Rudall, P.J., J.G. Conran, and M.W. Chase. 2000. Systematics of Ruscaceae/Convallariaceae: a combined morphological and molecular investigation. Bot. J. Linnean Soc. 134:73-92.

Soller, M. and J.S. Beckmann. 1983. Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67: 25-33.

Thurston L.M., K. Siggins, A.J. Mileham, P.F. Watson, and W.V. Holt. 2002. Identification of amplified restriction fragment length polymorphism markers linked to genes controlling boar sperm viability following cryopreservation. Biology of Reproduction 66: 545-554.

Villordon, A.Q. and La Bonte, D.R. 1996. Genetic variation among sweetpotatoes propgated through nodal and adventitious sprouts. J. Amer. Soc. Hort. Sci. 112:170-174.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Williams, J.G.K., A.R. Rubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Woeste, K., G. McGranaham and R. Bernatzky. 1998. Low correlation between genomic and morphologic introgression estimates in a walnut backcross. J. Amer. Soc. Hort. Sci. 123:258-263.

Woolfe, J. A. 1992. Sweetpotato: An untapped food resource. Cambridge University Press, Cambridge, UK. 643p.

CHAPTER 2: MOLECULAR AND MORPHOLOGICAL INVESTIGATION OF ORNAMENTAL LIRIOPOGONS[1]

2.1 Introduction

Liriopogons are the most important clonal ground covers sold by the nursery industry in the southeastern United States. The term liriopogon collectively represents ornamentals in the genera *Liriope* and *Ophiopogon*, a.k.a., monkeygrass, and denotes an underlying morphological similarity between the two genera. For example, the popular 'Aztecgrass' is commonly referred to as a *Liriope* sp., but in actuality it is identified taxonomically as an *Ophiopogon* sp. (Adams, 1989). This similarity has led to industry-wide misidentification of genera and putative clones possessing multiple trade names (Franz, 1994).

The industry widely regards cultivars in the genera *Liriope* as possessing evergreen grass-like leaves, erect flowers with a superior ovary (Adams, 1989), and berrylike, fleshy black seeds (Bailey and Bailey, 1976). Species in the genera *Ophiopogon* possess narrow, evergreen grasslike leaves, drooping flowers with an inferior ovary (Adams, 1989), and a blue, berrylike fruit (Bailey and Bailey, 1976). A recent taxonomic treatment of the Convallariaceae family differentiates the genera based mostly on flower part insertion (Conran and Tamura, 1998). Flowers of *Liriope* Lour. are hypogynous, i.e., the petals are inserted round the base of the gynoecium, slightly zygomorphic, i.e., bilateral symmetry, and produce blackish seeds. Flowers of *Ophiopogon* Ker Gawler are perigynous, i.e., the petals are attached to a cup-like hypanthium which arises round the base of the gynoecium, actinomorphic, i.e., regular symmetry, and produce blue seeds (Fig. 2.1). In contrast, Rudall et al., (2000) uses hemi-epigynous to describe the flowers. The genus *Peliosanthes* Andrews also shares these later characteristics and

[1] Reprinted by permission of "Journal of the American Society for Horticultural Science"

represents a putative third genus in the Tribe Ophiopogoneae Endl. *Peliosanthes* spp. have a corolla, i.e., inner set of floral leaves, whereas *Ophiopogon* spp. do not. Further refinement to species, e.g., *Liriope spicata* Lour., is more daunting, requiring mostly cellular level characterization (Cutler, 1992). Cultivar identification is equally difficult and confounded by inconsistent naming.

Molecular characterization is an approach useful in assigning cultivars or clones to genera and identifying unique banding profiles capable of fingerprinting cultivars for definitive identification. To date no reports exist in the literature using a molecular approach for these genera, notwithstanding a systematic investigation of the Convallariaceae by Rudall et al. (2000). Single accessions of the aforementioned genera were included in broader molecular and morphological analyses of the family Convallariaceae. Rudall et al. (2000) suggested just *Liriope* and *Ophiopogon* belong in the tribe Ophiopogonae.

The objective of the present study was to use a combination of morphological descriptors and DNA fingerprinting to differentiate common industry cultivars to the proper genera, to describe the extent of genetic similarities among these cultivars, and to identify possible duplicate cultivars.

2.2 Materials and Methods

2.2.1 Plant Material

Clones of 18 common industry cultivars (‘Variegata’, ‘Samantha’, ‘Spicata’, ‘Densiflora’, ‘Monroe white’, ‘Evergreen giant’, ‘Royal purple’, ‘John Birch’, ‘Silver dragon’, ‘King Felix’, ‘Big blue’, ‘Christmas tree’, ‘Silver midget’, ‘Variegated mondo’, ‘Mini mondo’, ‘Black mondo’, ‘Mondograss’, ‘Aztecgrass’) were collected from nurserymen in Louisiana or from an *in situ* collection at Louisiana State University Agricultural Center, Burden Research

Station, Baton Rouge, Louisiana for inclusion in the study. During the period of June through August of 2002, inflorescences were collected and preserved in AFE (ethanol 63 %, acetic acid 5 %, formaldehyde 2%) solution for later microscopic characterization. Voucher specimens were also collected, dried, and retained in the Louisiana State University herbarium.

2.2.2 Plant DNA Extraction

Young leaves were harvested, freeze-dried, and stored at -20°C until needed. Total DNA was isolated from 100 mg of leaf tissue using the Genelute™ Plant Genome Kit (Sigma-Aldrich Inc., St. Louis, Mo.).

2.2.3 AFLP Analysis

Techniques for AFLP fingerprinting were adapted from those described by Vos et al. (1995). A total of 0.08 µg of total genomic DNA from each sample was double-digested with 1.25 units each of a mixture of *Eco*RI and *Mse*I (AFLP Core Reagent Kit, Gibco BRL, Gaithersburg, Md) in 5X reaction buffer (50 mM Tris-HCl (pH 7.5), 50 mM Mg-acetate, 250 mM K-acetate) in a final volume of 6.25 µl. The reaction was incubated at 37°C for 3 h. Six µl of a mixture containing *Eco*RI and *Mse*I adapters (*Eco*RI/*Mse*I adapters, 0.4 mM ATP, 10 mM Tris-HCl (pH 7.5), 10 mM Mg-acetate, 50 mM K-acetate, 0.25 units of T4 DNA ligase) were added and the reactions were incubated at 18°C overnight. Samples were subsequently diluted 1:10 with TE buffer and stored at -20°C.

The pre-selective PCR amplifications were performed in 13 µl reaction volumes each of which contained 10 µl pre-amplification primer mix, 1.5 µl diluted adapter-ligated genomic DNA, 1.25 unit *Taq* DNA polymerase (Promega, Madison, Wis.) and 10X PCR buffer (100 mM Tris-HCl (pH 8.0), 15 mM MgCl2, 500 mM KCl). Pre-selective PCR amplifications (28 cycles) were performed as follows; 30 s at 94°C, 30 s at 56°C, and 60 s at 72°C.

The selective PCR amplifications were performed in 20 µl reaction volumes each of which contained 2.5 µl diluted (1:10) pre-amplified DNA template, 4.4 µl *Mse*I primer (6.7 ng/µl), dNTPs (200 uM each), 0.4 µl (0.5 pmol) of IRD700-label *Eco*RI primer (LI-COR, Lincoln, Neb.), and 1 unit of *Taq* polymerase, in 10X PCR buffer (100 mM Tris-HCl (pH 8.0), 15 mM $MgCl_2$, 500 mM KCl). The conditions for selective PCR amplifications were as follows; 12 cycles of 30 s denaturation at 94^oC, 30 s at 65^oC until reaching an annealing temperature of 56^oC, and extension for 60 s at 72^oC. This was followed by an additional 28 cycles of 30 s at 94^oC, 30 s at 56^oC, and 60 s at 72^oC.

All PCR reactions were conducted in a Perkin-Elmer GeneAmp PCR System 9600 thermal cycler (Perkin-Elmer Corp., Foster City, Calif.). PCR amplification fragments were separated by electrophoresis in 6.5 % polyacrylamide using a LI-COR Long ReadIRTM 4200L-1 DNA Sequencer (LI-COR protocols, http://www.bio.licor.com\pubs\biopub.htm; LI-COR, Lincoln, Neb.).

2.2.4 Data Analysis

Gel images were recorded during electrophoresis and analyzed with Gene ImagIR software (LI-COR, version 3.55). AFLP markers were scored for the presence (1) or absence (0) of bands. Data were analyzed (NTSYS-PC software version 2.0; Rolf, 1993) in two steps. Similarity values were calculated using Dice's coefficients of similarity (Sneath and Sokal, 1973). A dendrogram was then constructed based on the matrix of the similarities using the Unweighted Paired Group Method (UPGMA) (Sneath and Sokal, 1973). An Analysis of Molecular Variance (AMOVA) was also used to estimate variance attributed to among and within genotypic groups (Huff et al., 1993). Genetic distances for the AMOVA analysis were estimated using the Euclidean metric distance of Excoffier et al. (1992) defined as:

$E = n [1- 2n_{xy}/2n]$, where $2n_{xy}$ is the number of markers shared by two individuals, and n is the total number of polymorphic markers. The AMOVA analysis was performed using WINAMOVA 1.55 software (Excoffier et al., 1992).

2.3 Results and Discussion

2.3.1 Morphological Characterization

Hypogynous flowers were found on ‘Variegata’, ‘Samantha’, ‘Spicata’, ‘Densiflora’, ‘Monroe white’, ‘Evergreen giant’,’Royal purple’, ‘John Birch’, ‘Silver dragon’, ‘King Felix’, ‘Big blue’, ‘Christmas tree’, and ‘Silver midget’, affirming their inclusion in the genus *Liriope* (Fig. 2.1a). Perigynous or hemi-epigynous flowers were found on ‘Aztecgrass’ and ‘Black mondo’, affirming their inclusion in *Ophiopogon* (Fig. 2.1b). Three of the 18 cultivars seldom produced flowers. Undeveloped flowers of ‘Variegated mondo’ were found on one plant and are best described as hemi-epigynous. No flowers were found on ‘Mini mondo’; ‘Mondograss’ had some undeveloped inflorescences, but no flowers were found during our evaluation period.

2.3.2 Molecular Characterization

A total of 344 polymorphic AFLP markers (65 bp to 160 bp) were scored from four primer combinations (E-AAG/M-CTT, E-AAG/M-CTG, E-AAG/M-CAG, E-AAG/M-CTA); discrimination between AFLP markers was poor outside of this size range. The number of AFLP markers per genotype ranged from 64 (Christmas tree) to 94 (Samantha and Silver Dragon) with a mean of 78.

The 344 polymorphic AFLP markers were sufficient to differentiate all 18 genotypes (Fig. 2.2). Similarity values ranged from 0.23 to 0.57. The highest level of similarity (0.57) was between two groups of *Liriope* spp. cultivars, ‘Big blue’ / ‘Silver dragon’ and ‘Samantha’/

‘Variegata’, and the lowest was a group of two mondo grasses (Black mondo and Mini mondo) and all other cultivars (0.23), including other mondo grasses.

The AFLP marker-based UPGMA dendrogram (Fig. 2.2) depicts the relationships between individual plants. The goodness-of-fit was good ($r = 0.83$). *Ophiopogon* spp.cultivars mostly showed similarity. *Liriope* spp. cultivars were poorly grouped, exemplified by ‘Variegated mondo’.

A number of different groupings were analyzed using AMOVA, e.g., *Liriope* spp. cultivars with and without ‘Variegated mondo’, and variance components were estimated. The majority of the total genetic variability present was attributed to within group effects and was estimated to be >95 %. The among group variance component accounted for < 5.0 % of the total variability using various grouping strategies, but none showed significance ($p \leq 0.05$).

Floral structure mostly affirmed previous industry perceptions that ‘Variegata’, ‘Samantha’, ‘Spicata’, ‘Densiflora’, ‘Monroe white’, ‘Evergreen giant’, ‘Royal purple’, ‘John Birch’, ‘Silver dragon’, ‘King Felix’, ‘Big blue’, ‘Christmas tree’, and ‘Silver midget’ belong to the genus *Liriope* based on the Conran and Tamura (1998) Convallariaceae generic key (Fig. 2.1a). Flowers were less common on putative *Ophiopogon* spp. cultivars. Flowers were absent on ‘Mini mondo’ and ‘Mondograss’. Only ‘Aztecgrass’ (Fig. 2.1b) and ‘Black mondo’ had perigynous flowers. ‘Variegated mondo’ flowers were best described as hemi-epigynous, i.e., somewhat intermediate between the two classes in morphology.

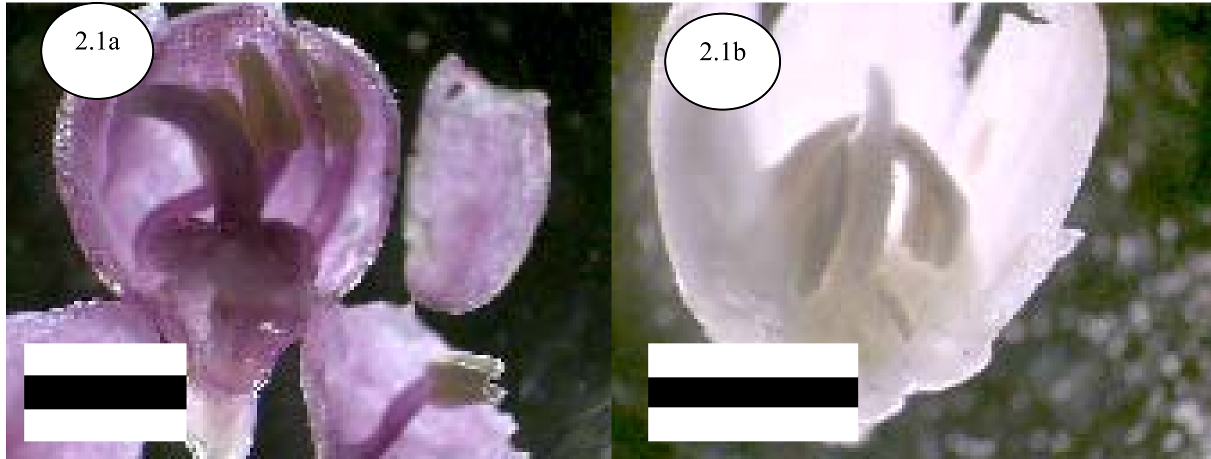

Fig. 2.1. Photomicrographs showing flower morphology of *Liriope* spp. and *Ophiopogon* spp. (a) ‘Royal Purple’ (*Liriope* spp.) showing a hypogenous ovary insertion. (b) ‘Aztecgrass’ (*Ophiopogon* spp.) showing a perigynous ovary insertion. Bar = 1.0 mm.

All cultivars used in this study were genetically different from one another based on polymorphic molecular marker profiles (Fig. 2.2). Morphological approaches alone are incapable of differentiating the ~100 named cultivars (Franz, 1993), underscoring the value of a molecular marker approach for identification.

Genetic distances estimated from the molecular marker data showed little differentiation between cultivars identified as *Liriope* spp. and *Ophiopogon* spp. (Fig. 2.2). Although trends are observable, i.e., cultivars tend to group based on genera, no significant groups existed based on AMOVA ($P \leq 0.05$). Most notable amongst the cultivars is 'Variegated mondo'. It tends to group more closely with *Liriope* spp. cultivars than to *Ophiopogon* spp. cultivars. Morphological data also showed 'Variegated mondo' has an intermediate floral structure between the genera. This cultivar may represent an inter-generic hybrid, if a distinction actually exists between these two genera. Cutler (1992) used various morphological characteristics to differentiate among five species of *Liriope* and twelve species of *Ophiopogon*. Using mostly crystal types, cuticular sculpting on cells adjacent to stomata, and leaf width, Cutler was able to assign various voucher materials to species. Yet, Cutler (1992) stated, "There appear to be no anatomical reasons for regarding *Liriope* and *Ophiopogon* as separate taxa". Rudall et al. (2000) found morphological differentiation between the two genera significant, but did not find differentiation based on molecular data, substantiating, in part, the findings of our current study. Rudall et al. (2000) also concluded that these two genera "are clearly closely related to each other".

In conclusion, we found molecular marker data clearly differentiated the cultivars from one another and showed that no duplicates existed in the samples of popular Liriopogon cultivars investigated.

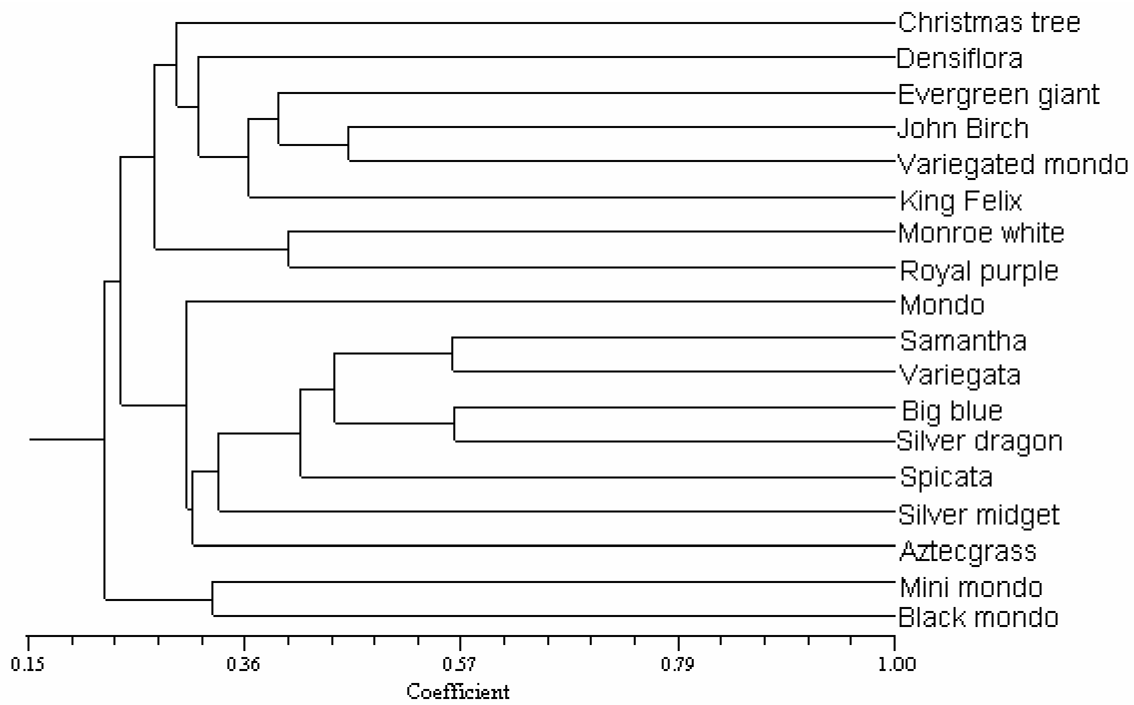

Fig. 2.2. UPGMA based dendrogram of 18 *Liriope* spp. and *Ophiopogon* spp. cultivars. The numerical scale indicates increasing genetic similarity.

Floral anatomy easily assigned cultivars belonging to *Liriope* spp., but was less successful on *Ophiopogon* spp. cultivars. Flowers were rare and ovary position less discernable on the latter. Most importantly, molecular marker data did not substantiate the existence of two genera. This preliminary data suggests close genetic affinity exists among representative *Liriope* and *Ophiopogon* spp. cultivars and supports a view that these genera are possibly one.

2.4 Literature Cited

Adams, G. 1989. Great ground covers. Amer. Nurseryman 170:83-91.

Bailey, L. H. and E. Z. Bailey. 1976. Hortus Third: A concise dictionary of plants cultivated in the United States and Canada. Macmillan Publishing Co. New York, New York, United States.

Conran, J.G., and M.M. Tamura. 1998. Convallariaceae, p. 186-198. In: J.G. Conran and M.M. Tamura. The families and genera of vascular plants. Springer-Verlag, Berlin.

Cutler, D.F. 1992. Vegetative anatomy of Ophiopogoneae (Convallariaceae). Bot. J. Linnean Soc. 110:385-419.

Excoffier, L., P.E. Smouse & J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Franz, P.R. 1993. Taxonomic problems in cultivated Liriopogons. HortTech. 3:146-149.

Franz, P.R.1994. A taxonomic research update of cultivated Liriopogons. HortTech. 4:46-48.

Huff , D.R., R. Peakall & P.E. Smouse. 1993. RAPD variation within and among natural-populations of outcrossing buffalograss (*Buchloe dactyloides* (Nutt) Engelm). Theor. Appl. Genet. 86:927-934.

Rudall, P.J., J.G. Conran, and M.W. Chase. 2000. Systematics of Ruscaceae/Convallariaceae: a combined morphological and molecular investigation. Bot. J. Linnean Soc. 134:73-92.

Sneath, P.H.A. & R.R. Sokal. 1973. Numerical taxonomy. The principles and practice of numerical classification. W.H. Freeman, San Francisco.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T.van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper & M. Zabeua. 1995. AFLP; a new technique for DNA fingerprinting. Nucleic Acids Red. 23:4407-4414.

CHAPTER 3: LINKING QUANTITATIVE TRAITS WITH AFLP MARKERS IN SWEETPOTATOES USING DISCRIMINANT ANALYSIS[1]

3.1 Introduction

Many of the agronomically important traits in sweetpotatoes [*Ipomoea batatas* (L.) Lam.] like root dry matter content are quantitative. Mcharo et al. (2001) suggest that root dry matter content is especially an important trait depending on the market being targeted. While some communities prefer low dry matter coupled with high sugar content others prefer starchy varieties with high sugar content. However, breeding for quantitative traits like root dry matter content in hexaploid sweetpotatoes has been inhibited by the significant genotype by environment interaction and by the complex polyploid genome of the sweetpotato (Jones et al., 1986). The net effect of these hurdles has been that controlled crossing programs were slower to develop for sweetpotato than for crops with simpler genetic systems. Due to the important role of the crop in the diets of many communities in the developing countries (FAO, 2000) there is a need to fast track the development of the sweetpotato but this will require use of novel breeding technologies. In recent years the development of marker assisted selection (MAS) protocols has attracted a lot of attention from plant breeders (Young, 1999). Young (1999) goes on to suggest that the development of marker systems which involve DNA manipulation has offered promise in construction of highly accurate DNA marker maps and quantitative trait loci (QTL) analysis.

The lack of complete genetic maps for the sweetpotato (Labonte et al., 1997) has presented greater challenges in identification of QTLs for the crop's improvement. Although polymerase chain reaction (PCR) based marker systems like random amplified polymorphic DNAs (RAPDS) (Williams et al., 1990), amplified fragment length polymorphisms (AFLPs) (Vos et al., 1995) and simple sequence repeats (Akkaya et al., 1992) have provided easier

---

[1] Published by Permission of "Acta Hort. (ISHS) 637:285-293; http://www.actahort.org/books/637/637_35.htm"

protocols for genome analysis, efficient techniques for linking quantitative traits with markers in a crop like the sweetpotato are yet to be developed. A study on the development of techniques for trait-marker linkage reveals two major breakthroughs, which though they have been useful, are plagued with daunting disadvantages. These protocols include the use of near isogenic lines (NILs) (Young et al., 1988) and bulked segregant analysis (BSA) (Michelmore et al., 1991). Analysis of NILs is a classical genetics approach that requires numerous resource consuming backcrosses. According to Michelmore et al., (1991) a further disadvantage to the NILs approach is that only half the targeted loci may be mapped after five backcrosses. Although bulked segregant analysis may overcome the shortcomings associated with NILs, the technique requires new DNA bulks for every loci targeted. The challenge of assessing purity for homozygosity in F_2 and the fact that there is no guarantee of detecting differences between bulks further compound the problem in BSA. A groundbreaking paper on the application of BSA in sweetpotatoes by Ukoskit et al., (1997) clearly illustrates these problems. Out of 760 primers screened for root-knot nematode resistance in sweetpotato only 1 out of the 9 polymorphic primers could be used to detect linkage between DNA markers and the root-knot nematode resistance gene.

Our study therefore sought to develop a protocol that can elucidate many molecular marker profiles with minimal primers. Due to the dearth of genome maps in sweetpotato, we seek to develop a method that does not require *a priori* genetic maps and also one that is capable of analyzing multiple traits or variables simultaneously. The multiple variables or markers in our case as generated by AFLP required a robust statistical technique to analyze them simultaneously hence the use of discriminant analysis. Discriminant analysis (DA) is a multivariate statistical technique that can identify differences among groups of individuals (or treatments) and improve the understanding of relationships among the variables measured within those groups (Cruz-

Castillo et al., 1994). DA determines how best to separate or discriminate two or more groups of individuals, given quantitative measurements of several individuals, through simultaneous analysis of several variables (Rencher, 1992). DA finds linear functions of quantitative variables that maximally separate two or more groups of individuals while keeping variation within groups as small as possible. DA may be used when it is important to separate known groups or *a priori* groupings, and to identify major sources of difference between groups. Such groups may be determined through a statistical procedure like molecular marker profiling or through cluster analysis techniques. The discriminant analysis will therefore provide information on the discriminatory power of each variable or marker thus determining whether a particular set of variables or markers is useful in separating previously delineated groups (Rojas et al., 2000). The study presented in this paper involves various dry matter groups of sweetpotatoes.

3.2 Materials and Methods

3.2.1 Experimental Material

DNA samples were obtained from sixty-eight sweetpotato clones. These clones were selected from an original USDA population of 394 clones on the basis of their root dry matter content (http://www.ars-grin.gov/npgs/). Root dry matter content is dry weight as a percentage of fresh root weight. The 68 clones were divided into two populations the first one consisting of 34 clones with high dry matter content (36.67% to 43.48%) and another 34 clones with low dry matter content (12.20% to 21.99%) (Table 3.1). A training sample consisting of 29 high dry matter and 29 low dry matter clones from the original 68 was used for the development of a phenotypic group prediction model. A second group consisting of the remaining 5 high dry matter (40.48% to 41.61%) and 5 low dry matter (16.22% to 17.91%) clones was used as a test population to validate the model.

Table 3.1. List of sweetpotato clones obtained from USDA/NPGS repository.

No.	Accession No.	Dry matter content (%)	Dry matter group	No.	Accession No.	Dry matter content (%)	Dry matter group
1	1067	43.48	high	35	564157	21.99	low
2	573309	43.35	high	36	320446	21.82	low
3	566633	41.89	high	37	585055	21.79	low
4	564152	41.82	high	38	556935	21.76	low
5	1096	41.80	high	39	1149	21.69	low
6	564114	41.69	high	40	566613	21.53	low
7	508508	41.61	high	41	531116	21.27	low
8	556940	41.58	high	42	538285	21.14	low
9	153905	41.14	high	43	573324	20.95	low
10	531113	40.79	high	44	1168	20.83	low
11	1227	40.48	high	45	1146	20.79	low
12	531097	39.45	high	46	585073	20.76	low
13	564112	39.16	high	47	1143	20.67	low
14	585087	38.86	high	48	1239	20.59	low
15	531093	38.85	high	49	531114	20.50	low
16	1039	38.76	high	50	1134	20.48	low
17	1054	38.68	high	51	1229	20.48	low
18	531131	38.58	high	52	1145	19.92	low
19	585100	38.23	high	53	1223	19.62	low
20	564151	38.20	high	54	556941	19.45	low
21	564770	38.06	high	55	1150	18.75	low
22	573298	37.79	high	56	556944	18.65	low
23	1068	37.72	high	57	538284	18.55	low
24	508518	37.69	high	58	1158	18.51	low
25	564149	37.63	high	59	12500	18.44	low
26	556946	37.60	high	60	585063	17.91	low
27	376945	37.58	high	61	1236	17.63	low
28	508530	37.53	high	62	531141	16.87	low
29	1063	37.40	high	63	1228	16.70	low
30	585068	37.38	high	64	1125	16.22	low
31	564106	37.18	high	65	566627	14.72	low
32	566662	37.12	high	66	585065	14.54	low
33	508521	37.03	high	67	564109	13.69	low
34	585093	36.67	high	68	344124	12.20	low

3.2.2 AFLP Analysis

The DNA samples were amplified in a three-step process using a GeneAmp PCR system 9600 thermocycler (Perkin Elmer, Fulerton, CA). Reagents for AFLP™ were obtained from Invitrogen™ (AFLP starter primer kit, Cat No. 10483-014) and LI-COR Inc. (Lincoln, NE, Cat. No. 420032). Genomic DNA (80ng/µl) was digested using an *Eco*RI/ *Mse*I restriction enzyme mix at 37 ^{0}C for 3 hours. The enzymes were then inactivated by incubating the mix at 70 ^{0}C for 10 minutes. Double stranded adaptors were then ligated to the restricted DNA fragments resulting in template DNA which was used for pre-amplification. Diluted template DNA (1.5µl) was added to 10µl pre-amp primer mix, 1.25 units Taq DNA polymerase (Invitrogen™) and 1.25µl RedTaq™ PCR reaction buffer 10x with $MgCl_2$ (Sigma-Aldrich™) to make a 13µl reaction volume. The pre-amplification conditions were 20 cycles each of 94 ^{0}C for 30s, 56 ^{0}C for 60s, 72 ^{0}C for 60s and a final hold at 4 ^{0}C. Pre-amplification of the final product was verified by running on a 1.2% agarose gel.

The reaction volume for selective amplification consisted of 3.0 µl pre-amplified diluted DNA (where 3.0 µl were insufficient, the volume was raised to 3.5 µl), 0.4 µl of *Eco*RI (fluorescently labeled) primer (AAG), 4.4 µl of *Mse*I (unlabelled) primer, 2.0µl RedTaq™ PCR reaction buffer 10x with $MgCl_2$ (Sigma-Aldrich™), 1.4ul $MgCl_2$, 1 unit Taq DNA polymerase (Invitrogen™) and 6.38 µl of double distilled or AFLP grade water. Four primer pairs as identified by Farjado (2000) were used for selective amplification (CAG, CTA, CTG, CTT). The conditions for selective PCR amplifications were as reported in chapter 2. Blue stop solution (2.0µl) (LI-COR, Lincoln, NE) was added onto each amplified DNA sample. The amplified DNA was then denatured at 95 ^{0}C for 3 minutes and thereafter covered in aluminium foil and placed in a freezer at -20 ^{0}C for 10 minutes to prevent annealing of complementary fragments,

before loading onto a 25cm acrlyamide gel. PCR amplification fragments were separated by 6.5% acrylamide gel electrophoresis using a LI-COR Global IR^2 sequencer (LI-COR, Lincoln, NE) for 3 hours. The AFLP fragments were automatically detected and recorded during electrophoresis using the LI-COR $SAGA^{GT}$ v 2.1 software. Data were collected and presence (=1) or absence (=0) of bands scored using Gene ImagIR v 4.0 software (LI-COR, Lincoln, NE).

3.2.3 Statistical Analysis

Phenotypic grouping of the test genotypes was done by cluster analysis using the NTSYS v 2.0 software (Rohlf, 1998). Analysis of molecular variance (AMOVA) on the genotypes with AFLP marker profiles was used to test genotypic variability based on molecular marker information (Excoiffer et al., 1992). Due to the numerous markers generated, it was necessary to use a variable reduction technique to select the most discriminating ones. Step discriminant analysis using the STEPDISC procedure (SAS, 2001) was used to select the most informative markers from the original set of markers. The stepwise and forward selection option in STEPDISC were used to select markers to be included in the classification model. A significance level of 0.1 of an F test from an analysis of covariance was imposed to choose the most discriminating markers (SAS, 1990). The level of significance was based on a study by Costanza and Afifi (1979). Wilk's lambda was used as the criterion to determine the classification efficiency with the entry of each marker. The selected markers were then used in a discriminant analysis, DISCRIM option (SAS, 2001), to develop and validate a phenotypic group prediction model and to predict group membership of the test genotypes. The categorical nature of the data could not allow assumption of normal distribution to be made hence a nonparametric method; the k-nearest neighbor method (Rosenblatt, 1956) was used to estimate the group specific densities that produce a classification criterion. The performance of the discriminant criterion was

evaluated by posterior probability error rate and group specific error count estimates during cross-validation. These two estimators give the proportion of misclassified observations in each group.

3.3 Results and Discussion

3.3.1 Cluster Analysis

Cluster analysis, whose objective was to statistically establish the phenotypic groups, classified 34 of the 68 clones as high dry matter clones and 34 as low dry matter clones (Fig 3.1). From these clones 58 were selected as a training sample to develop a classification criterion. The training sample proportions were equal (29 out of 58 clones) for both dry matter groups and hence the prior probabilities of group membership were assumed to be equal. Lack of normality (results not reported) for dry matter data obviates the need for ANOVA. STEPDISC analysis using the stepwise selection option was used to reduce the number of polymorphic markers generated by the four primer combinations from an initial 903 to a more manageable 224. STEPDISC was further used to reduce the number of markers to form classification models of up to 14 markers (Table 3.2). An AMOVA found significant ($P < 0.001$) genetic variation between the two phenotypic groups of the original 68 clone population using the 224 selected markers. AMOVA also found significant differences ($P < 0.001$) for the 68 clones using six to fourteen of the most informative markers (data not shown). The significant result using fewer markers suggests that the STEPDISC procedure is useful in selecting a critical subset of markers. Concentrating on the selected markers could reduce the resources needed in investigating trait-marker relationships without compromising on the information gained.

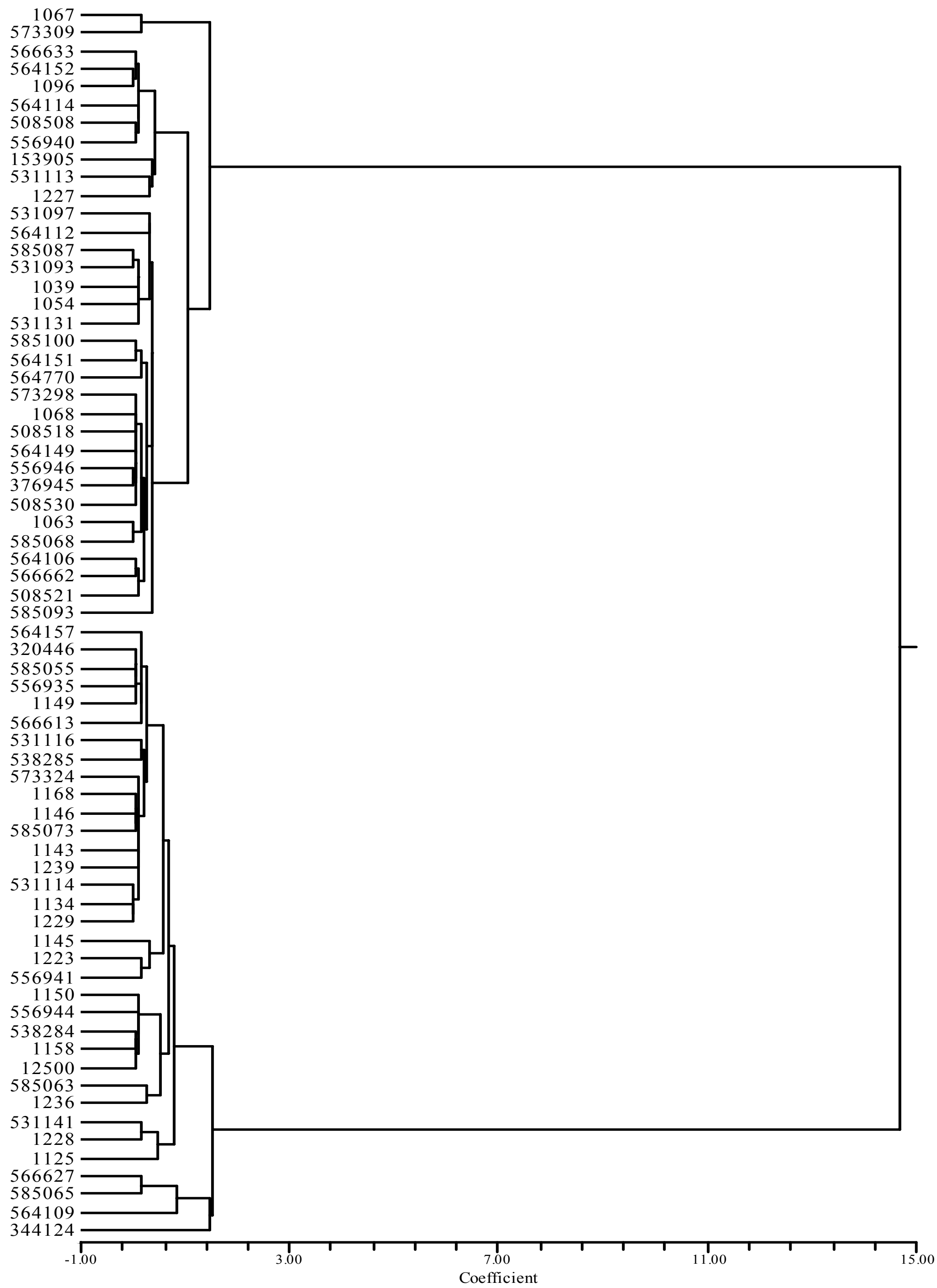

Fig 3.1. Dendrogram using the DIST coefficient and single-link clustering method between high and low root dry matter content sweetpotato groups. Upper cluster consists of high dry matter (NPGS accessions 1067 through 585093) and lower cluster consists of low dry matter clones (NPGS accessions 564157 through 344124).

Table 3.2. AFLP predictor markers in sweetpotato as selected by the STEPDISC procedure.

Marker[1]	Entry step	Wilk's Lambda	Pr < Lambda
cta084	1	0.84	0.0019
cag185	2	0.69	< 0.001
cag235	3	0.51	< 0.001
cag148	4	0.42	< 0.001
cta212	5	0.36	< 0.001
ctt183	6	0.28	< 0.001
cta265	7	0.23	< 0.001
cag271	8	0.17	< 0.001
ctt241	9	0.11	< 0.001
cta076	10	0.08	< 0.001
cta235	11	0.07	< 0.001
cta254	12	0.06	< 0.001
ctg251	13	0.05	< 0.001
cag273	14	0.04	< 0.001

[1]The numbers beside the primer trinucleotide code indicate the molecular weight of the marker in base pairs.

3.3.2 Discriminant Analysis

The main objective of discriminant analysis in this study was to identify the best combination of molecular markers that could be used to assign individuals to pre-defined groups. As part of the study it was important to verify the predictive power of the selected combination of markers or model. Table 3.3 shows the rates of correct classification of genotypes into dry matter groups as achieved by the various models after cross-validation. Our prior probabilities of group membership were equal at 0.5. Consequently any model that achieved a correct classification that was significantly greater than 50% was acceptable. From our studies the higher the number of markers used, up to 14, the greater the level of accuracy in classification. During evaluation by cross-validation no more than 14 markers were required to achieve a 100% correct classification and any number of markers selected beyond 14 achieved 100% correct classification. As the number of predictor markers decreased, an increase in misclassification arose, e.g. six markers misclassified 6 clones out of 58 (Table 3.3). The error rates were

calculated using the misclassified clones. The total error rate is the mean of the group error rates. A negative error rate (-0.03 in the low dry matter group) is theoretically due to incongruence between prior group membership probabilities and group sizes.

Marker populations selected by discriminant analysis were used to assess correct classification of a test set of 10 genotypes. Although a test set of 10 genotypes may be considered small, the results give us an indication of the accuracy of the models developed. This test set which consists of two phenotypic groups also serves to illustrate the limitation of nonparametric discriminant functions. Not all the test clones could be correctly classified using models with either 6 or 10 markers (Table 3.4). Increasing the number of markers in the model improved the accuracy of prediction of the test clones. Any number of markers in the model beyond 12 achieved 100% correct classification.

As these results indicate it is necessary to strike a good balance between sufficient number of markers and the amount of information desired. Selecting too few markers may limit the researcher's progress if those markers are not found in a particular set of genotypes. It is also important to use as many genotypes as possible during the model calibration stage to determine the best combination of markers that may have a wide application. Table 3.2 shows the 14 most powerful markers as determined by Wilk's lambda and Pr < lambda. As the selection of markers progresses the influence of the unselected markers is not taken into account. The power of the selected marker is determined by the already selected markers which act as covariates during the analysis. Consequently the model developed will not necessarily be the best possible. However as our results indicate careful examination of the data combined with the statistical analyses is likely to yield the desirable results (Tables 3.3 and 3.4).

Table 3.3. Rate of correct classification of 58 training clones of sweetpotato into dry matter group after cross-validation using nearest neighbor in discriminant analysis.

Number of predictor markers	Error type	High dry matter group error rate	Low dry matter group error rate	Total
6	PPER[1]	0.17 (5)[3]	-0.04 (1)	0.07
	APER[2]	0.17	0.03	0.10
10	PPER	0.03 (2)	0.00 (1)	0.02
	APER	0.07	0.03	0.05
12	PPER	0.03 (1)	-0.03	0.00
	APER	0.03	0.00	0.02
14	PPER	0.00	0.00	0.00
	APER	0.00	0.00	0.00

[1]Posterior probability error rate estimates.
[2]Apparent error rate estimates.
[3]Number in parenthesis is the number of misclassified clones in group.

Table 3.4. Rate of correct classification of 10 sweetpotato test clones into dry matter group using discriminant analysis.

Number of predictor markers	Error type	High dry matter group error rate	Low dry matter group error rate	Total
6	PPER[1]	0.40 (2)[3]	0.20 (1)	0.30
	APER[2]	0.40	0.20	0.30
10	PPER	0.20 (1)	0.20 (1)	0.20
	APER	0.20	0.20	0.20
12	PPER	0.00	0.00	0.00
	APER	0.00	0.00	0.00
14	PPER	0.00	0.00	0.00
	APER	0.00	0.00	0.00

[1]Posterior probability error rate estimates.
[2]Apparent error rate estimates.
[3]Number in parenthesis is the number of misclassified clones in group.

The two error rates used to evaluate the effectiveness of the developed models give us different results with the apparent error count estimate giving a consistently higher figure with the exception of the 14 marker model where the error rates were both zero. The apparent error count estimate, also referred to as an apparent error rate has been criticized due to its tendency to

have an optimistic bias. Indeed Ebdon et al. (1998) obtained results which would seem to support this view. However these workers used quantitative and normally distributed data. Our work involves categorical non-normally distributed data and for almost every model evaluated the apparent error rate was more conservative than the posterior probability error rate. According to the results presented in Table 3.3 this translates into the apparent error rate having at least 2% higher level of total misclassification rate, compared to the posterior probability error rate, except for the 14-marker model.

Marker assisted selection especially for the purpose of germplasm improvement has come a long way. Although great strides have been made the reality of nonparametric and categorical data is a hurdle that will require more innovative techniques in order to overcome it. Young (1999) argues on the need for caution when approaching crop improvement through marker assistance and more so through QTL analysis. Dry matter as a quantitative trait would be influenced by several loci. Our results suggest that there are dominant AFLP markers associated with both high root dry matter content and low dry matter content. Although this is not a gene mapping study, the markers identified are likely to be closely associated with QTLs responsible for expression of this trait. This suggestion is supported by Capdevielle (2001) who investigated the linkage between marker assisted classification and differential response of rice to sheath blight disease. For a crop whose genome has not been mapped, novel techniques like discriminant functions combined with molecular marker data are likely to herald an era of faster progress in breeding. Applications of such protocols include screening of large germplasm collections for desired quantitative traits and also phenotypic class identification and verification of clones that have been assigned particular classes. Our results are encouraging but a wider study with more clones from a wider diversity of agro-ecological zones and progeny populations

needs to be done to build on the present progress. With further research we hope to develop a marker assisted selection protocol for important traits in sweetpotato based on discriminant analysis.

3.4 Literature Cited

Akkaya, M.S., A.A. Bhagwatt, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Capdevielle, F.M. 2001. Evaluation of a discriminant analysis procedure combining agronomic and molecular marker information for germplasm improvement in rice. MS Thesis. Louisiana State University, Baton Rouge.

Costanza, M.C. and A.A. Afifi. 1979. Comparison of stopping rules in forward stepwise discriminant analysis. Journal of the American Statistical Association. 74:777-785.

Cruz-Castillo, J.G., S. Ganeshanandam, B.R. MacKay, G.S. Lawes, G.S, C.R.O. Lawoko, and D.J. Woolley. 1994. Applications of canonical discriminant analysis in horticultural research. Hortscience 29:1115-1119.

Ebdon, J.S., A.M. Petrovic, and S.J. Schwager. 1998. Evaluation of discriminant analysis in identification of low and high water use Kentucky Bluegrass cultivars. Crop Science 38:152-157.

Excoiffer, L., P.E. Smouse, and J.M. Quattiro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131:479-491.

Fajardo, D.S. 2000. Genetic diversity in Papua New Guinea sweetpotato [*Ipomoea batatas* (L.) Poir] germplasm. MSc thesis, Louisiana State University, Baton Rouge.

FAO. 2000. FAO yearbook, Rome.

Jones, A., P.D. Dukes, and J.M. Schalk. 1986. Sweetpotato breeding. p 1-35. In: Basset M.J. (ed.), Breeding vegetable crops. AVI publishing company, Westport, Connecticut.

La Bonte, D.R., M.I. Buteler, and A.Q. Villordon. 1997. Molecular markers in sweetpotato p 181-191. In: D.R. LaBonte, M. Yamashita and H. Mochida (eds.), Proceedings of International Workshop on Sweetpotato Production Systems towards the 21st Century, Miyakonojo, Miyazaki, Japan.

Mcharo, M., E.E. Carey, and S.T. Gichuki. 2001. Performance of selected sweetpotato varieties in Kenya. African Crop Science Journal. 9:49-57.

Michelmore, R.W., I. Paran, and R.V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulk segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. USA 88:9829-9832.

Rencher, A.C. 1992. Interpretation of canonical discriminant functions, canonical variates, and principal components. Amer. Statistician 46:217-225.

Rohlf, F.J. 1998. Numerical taxonomy and multivariate analysis system v2 user guide. Applied Biostatistics Inc. Setauket, New York.

Rojas, W., P. Barriga, and H. Figueroa. 2000. Multivariate analysis of Bolivian quinua germplasm. Plant Genetic Resources Newsletter. No 122:16-23.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics. 27:832-837.

SAS Institute Inc. 1990. SAS/STAT User's guide. SAS Institute Inc. Cary, North Carolina.

SAS Institute Inc. 2001. The SAS system for windows v8. SAS Institute Inc. Cary, North Carolina.

Ukoskit, K., P.G. Thompson, C.E. Watson Jr., and G.W. Lawrence. 1997. Identifying a randomly amplified polymorphic DNA (RAPD) marker linked to a gene for root-knot nematode resistance in sweetpotato. J. Amer. Soc. Hort. Sci. 122:818-821.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Williams, J.G.K., A.R. Rubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Young, N. D. 1999. A cautiously optimistic vision for marker-assisted breeding. Molecular Breeding 5:505-510.

Young, N.D., D. Zamir, M.W. Ganal, and S.D. Tanksley. 1988. Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the Tm-za gene in tomato. Genetics. Baltimore, Md.: Genetics Society of America. 120:579-585.

CHAPTER 4: ASSOCIATING MOLECULAR MARKERS WITH VIRUS RESISTANCE TO CLASSIFY SWEETPOTATO GENOTYPES

4.1 Introduction

The most widely used approach to germplasm improvement in sweetpotato [*Ipomoea batatas* (L.) Lam.] is mass selection (Jones, 1986). This fundamentally simple technique lends itself well to an allohexaploid crop that possesses few, if any, simply inherited traits (Buteler et al., 1999; Jones, 1986; Zhang et al., 2001). Unfortunately, self-incompatibility prevents quick introgression of desirable traits from landraces or exotics into adapted material. The preponderance of quantitatively inherited traits in sweetpotato (Jones, 1986), environmentally unstable trait expression, and the encumbrance of two year breeding cycles underlie recent attempts to use trait-linked DNA molecular markers in breeding programs.

The most compelling reason to develop trait-linked DNA markers in sweetpotatoes is to effectively breed for virus resistance. Viral diseases have been identified to cause up to 56 to 98% yield reduction in sweetpotatoes (Gibson et al., 1997; Ngeve, 1990). According to Geddes (1990) the sweetpotato virus disease (SPVD), a result of the co-infection of SPCSV and SPFMV, is the most destructive of the viral diseases within the African continent (Gibson et al., 1997; Karyeija et al., 1998). Hahn et al. (1981) and Mwanga et al. (2002a) estimated high broad-sense heritability for resistance to SPVD suggesting that breeding for resistance is possible.

Ukoskit and Thompson (1997) and later Kriegner et al. (2003) demonstrated the feasibility of constructing a genetic linkage map for sweetpotato. Mapping was based on the single-dose restriction fragment (SDRF) method described by Wu et al. (1992). Progeny segregating in a 1 present :1 absent marker ratio arise from a simplex x homozygous recessive test cross, e.g., aaaaaa x Aaaaaa. Progeny segregating in 4:1 ratio that generate duplex markers AAaaaa were also used. A map by Kriegner et al. (2003) contained 1067 AFLP markers ordered

in ~90 linkage groups with genome coverage suitable for detection of QTL. Mwanga et al. (2002b) used linkage maps to associate DNA RAPD and AFLP markers to SPFMV and SPCSV resistance. Attempts to identify markers associated with SPVD were unsuccessful because few progeny in the defined population were resistant. Despite such efforts, QTL analysis is still a challenge in breeding programs for a crop like the sweetpotato since a complete genetic map is still unavailable.

Faced with the complications associated with QTL analysis this study seeks to determine whether the results of molecular marker and phenotype association for genotype characterization and disease resistance prediction using multivariate statistical procedures like discriminant analysis and logistic regression are comparable to those of QTL analysis. Both permit assemblage of test populations, i.e., resistant and susceptible classes, without regard to genetic relationships.

Discriminant analysis is a multivariate technique that uses multiple predictor variables to associate an individual with a descriptive class (Fisher, 1936). In our genetic studies of sweetpotatoes, discriminant analysis would involve relating a genotype, as described by its molecular marker profile, with its disease resistance class or phenotypic group. Using the markers as the multiple variables the method would develop an optimum classification function that maximizes differences between the resistant and susceptible classes while minimizing differences among members of those two classes. The efficiency of the function is evaluated by cross-validation as described by Lachenbruch and Mickey (1968). Cross-validation is a leave-one-out method that aims to reduce the optimistic bias in calculating classification error-rate. If the assumption of multivariate normality is not met or when no assumptions are valid about the distribution within each phenotypic group, a nonparametric procedure like *k*-nearest-neighbor

(Rosenblatt, 1956) is used for classification. With the *k*-nearest-neighbor method, a Mahalanobis distance is calculated between group means and this distance is then used to determine an individual's closeness to the mean. All *n* observations in the dataset are used to calculate distances; however the observation to be classified is left out from the *k* nearest neighbors of that observation during function development.

Recent work on genetic analysis using discriminant analysis include that of Capdevielle et al. (2000) who associated microsatellites and agronomic traits in rice. Fahima et al. (2002) who investigated microsatellite polymorphism in wheat also used discriminant analysis in their characterization.

Logistic regression is a statistical technique that is frequently used to associate explanatory variables with a binary outcome (Ostir and Uchida, 2000). When there are more than two response classes discriminant analysis is more reliable than logistic regression. Models developed using logistic regression have the advantage of summarizing and interpreting the data and hence the researcher can make inferences on factors being investigated. In our studies, the process involved relating molecular markers as the independent factors and presence or absence of viral infection as the dichotomous response. To my knowledge there is no work on logistic regression for molecular marker selection in sweetpotatoes. However, field disease prediction studies have extensively used logistic regression tools. Mukasa et al. (2003) used logistic regression to determine the odds ratios to detect single- or multiple-virus infections in sweetpotato genotypes.

The specific objectives of this study were to compare marker – trait association using QTL analysis, discriminant analysis and logistic regression for DNA markers derived from a

population of sweetpotato genotypes and to develop and validate genotype classification and resistance class prediction models for SPCSV and SPFMV.

4.2 Materials and Methods

Molecular marker detection and phenotypic data acquisition for the population was previously described by Mwanga et al. (2002b). The data consisted of an array of molecular markers and phenotypic information for 87 F_1 genotypes derived from a cross between 'Tanzania' and 'Wagabolige' cultivars. Both 'Tanzania' and 'Wagabolige' are resistant to SPVD although 'Wagabolige' is more resistant to SPVD than 'Tanzania' under field SPVD inoculum pressure (Mwanga et al., 2002b). Detection and confirmation of virus presence or absence in the genotypes involved using the triple-antibody sandwich enzyme-linked immunosorbent assay (TAS-ELISA) test for SPCSV and the nitrocellulose membrane enzyme-linked immunosorbent assay (NCM-ELISA) test for SPFMV (International Potato Center, 1990). Resistant genotypes tested negative while susceptible genotypes tested positive for virus presence (Mwanga et al., 2002b). 'Tanzania' is widely grown in Sub-Sahara Africa while 'Wagabolige' is more confined to Uganda. The 87 F_1 genotypes had 269 dominant molecular markers, 232 of them being AFLP fragments and 37 being RAPD fragments. These were 2 markers less than what Mwanga et al. (2002b) used. Fifty genotypes were classified as resistant and 37 susceptible to SPFMV. Forty two were classified as resistant and 45 susceptible to SPCSV, respectively. Phenotypic data for SPCSV and SPFMV scores were obtained using ELISA tests and were coded as a one (1) for positive virus infection or zero (0) for negative results.

Statistical analyses were done on SPFMV and SPCSV only because only 3 out of 87 genotypes were resistant to SPVD. These were too few for the statistical procedures to converge and give a solution.

QTL analysis was done using QGENE (Nelson, 1997). Single-point analysis of variance (ANOVA), or regression on a marker (Soller et al., 1976), was conducted to reveal significant ($P<0.05$) associations between markers and virus resistance. Markers were considered independent variables and virus resistance score as the dependent variable. Proportion of the total phenotypic variance accounted for by the markers was estimated by the coefficient of determination (R^2) obtained from simple linear regression (Nelson, 1997; Soller et al., 1976).

Step discriminant analysis using the STEPDISC procedure (SAS, 2001) was used to select the most informative markers from the original set of markers. The progeny population was divided into two phenotypic groups or classes based on resistance – susceptible reactions, similar to QTL analysis. The forward selection option in STEPDISC was used to select markers to be included in the classification model. As described in SAS (1999) the forward selection process commences with no markers in the model. Entry-significance levels of $P\leq0.03$ and $P\leq0.02$ for SPCSV and SPFMV, respectively, of the chi-square score for entering an effect or marker into the model to achieve at least 95% prediction accuracy was imposed to choose the most discriminating markers (SAS, 1999). At each step the procedure enters the marker that contributes most to the discriminatory power of the model as measured by Wilks' Lambda, the likelihood ratio criterion. The process stops when all unselected markers fail to meet the entry condition. The selected markers were then used in a nonparametric discriminant analysis ($k = 1$), DISCRIM option (SAS, 2001), to construct and validate a class prediction function and to predict group membership, resistant or susceptible, of the test. Since the categorical nature of the data could not allow assumption of normal distribution to be made a nonparametric method, the *k*-nearest neighbor method (Rosenblatt, 1956), was used to estimate the group-specific densities that produce a classification criterion. The performance of the discriminant criterion was

evaluated by group-specific error-count estimates during cross-validation. The error estimator gives the proportion of misclassified observations in each group. Total error, from which we derive percent correct classification, is the weighted mean of error estimates of the two phenotypic groups.

PROC LOGISTIC (SAS, 2001) was used to perform logistic regression with the logit link to select markers that significantly accounted for phenotypic variation, with the forward selection option used for marker selection. With forward selection, PROC LOGISTIC first estimates parameters for effects, markers in my case, forced into the model. These effects are the intercept and the first *n* explanatory effects in the model, *n* is zero by default. Next, the procedure computes the score chi-square statistic for each effect not in the model and examines the largest of these statistics. If it is significant at a preset probability level, the corresponding effect is added to the model. Once an effect is entered in the model, it is never removed from the model. The process is repeated until none of the remaining effects meet the specified level for entry. For my study, entry significance levels of $P \leq 0.05$ and $P \leq 0.02$ for SPCSV and SPFMV respectively to enter marker into model were set. The phenotype was a binary outcome as either resistant or susceptible. The Hosmer and Lemeshow Goodness of Fit test was used to determine model efficiency (Hosmer et al., 1991).

4.3 Results and Discussion

4.3.1 Sweetpotato Chlorotic Stunt Virus

All three statistical procedures selected two common markers (e41m33.a and e38m36.u) that had a significant association with resistance to SPCSV (Tables 4.1, 4.2 and 4.3). These results are in concordance with those of Mwanga et al. (2002b). Marker e38m36.u was selected by discriminant and logistic but not QTL procedures as one of the informative markers. Marker

e41m33.a was the most informative, regardless of the method of selection used, accounting for 66.75% of the variation in SPCSV resistance according to QTL analysis. Our results suggest the possibility of using these multivariate methods for identifying strong QTLs. With the six markers selected by discriminant analysis to create a classification model, we achieved 92% correct classification for SPCSV in the whole population. Discriminant analysis misclassified one genotype out of 42 that were phenotypically classified as being resistant. This model also misclassified into the resistant class three genotypes out of 45 that were visually described as susceptible. The model developed using logistic regression fitted the data well according to the Hosmer and Lemeshow (Hosmer et al., 1991) goodness of fit test ($X^2 = 0.082$, df = 3, P = 0.9939). Table 4.4 shows observed vs expected class values for SPCSV and SPFMV.

4.3.2 Sweetpotato Feathery Mottle Virus

Marker S13.1130 was identified by all three selection procedures and accounted for 64.62% of the variation in SPFMV resistance (Tables 4.1, 4.2 and 4.3) according to QTL analysis. Marker e39m32.f was the second most powerful marker selected by QTL analysis but neither by discriminant nor logistic analyses. Markers e41m37.a and e44m36.d were selected as being informative by discriminant and logistic analyses only (Tables 4.2 and 4.3) but in differing order of importance. The discriminant analysis model constructed using the four selected markers achieved a 96% correct classification rate. Four clones out of 50 that were phenotypically classified as being resistant were classified by discriminant analysis as being susceptible. These are the clones that discriminant analysis misclassified and they give us the total error rate of the model developed. This model achieved a highly accurate classification rate and with only four variables was a desirable result because we aim to have as few markers as possible to create an efficient class prediction model. Results of Hosmer and Lemeshow model evaluation indicate

that the logistic model fitted the data well (X^2 = 0.071, df = 2, P = 0.9951) and the proximity of expected values to observed values (Table 4.4) confirm this.

QTL analysis, while providing useful results, is limited by the fact that it can only be used in a population of individuals that are closely related. Furthermore, results from QTL analysis have in the past been inapplicable in other populations. Lubberstedt et al. (1998) in their work on maize reported that QTL results were not consistent among crosses, within the flint heterotic pool, suggesting that prior to marker-assisted selection, QTL mapping must be performed separately for each population. In a review of QTL methods in plants Kearsey and Farquhar (1998) showed that analytical methods locate QTL with poor precision (10-30 centimorgans), unless the heritability of an individual QTL is high. They further stated that this unreliability of QTL location may suggest false candidate genes. In his review of QTL analysis Young (1999) stated that phenotypic mis-scoring of even a few individuals can totally confound QTL discovery and placement. He also suggested that nonparametric categorical data may not be as amenable to QTL analysis as normally distributed quantitative data thus underscoring the difficulties associated with using QTL for marker assisted selection.

Mwanga et al. (2002b) reported studies combining both quantitative genetics and molecular marker investigations in an attempt to describe inheritance of resistance to SPCSV, SPFMV, and SPVD. However these workers were faced with the undesirable difficulty of constructing genetic maps using null alleles and having insufficient plant numbers in the SPVD resistant category. In their QTL studies on molecular markers, Mwanga et al. (2002b) found that marker e41m33.a was responsible for 70% of the variation in SPCSV resistance. They also found that marker S13.1130 accounted for 72% of the resistance to SPFMV. Their conclusions were

Table 4.1. QTL analysis for DNA markers associated with ≥10% resistance to sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato.

Disease	Marker	F	R^2	Prob
SPCSV				
	e41m33.a	170.66	0.67	<0.0001
	e40m34.c	111.16	0.57	<0.0001
	e38m36.u	19.24	0.18	<0.0001
	e36m49.a	14.42	0.14	0.0003
	e35m49.d	12.58	0.13	0.0006
	R9.650	9.41	0.10	0.0029
	e39m40.e	9.23	0.10	0.0032
SPFMV				
	S13.1130	155.28	0.65	<0.0001
	e39m32.f	52.49	0.38	<0.0001
	e36m59.a	10.53	0.11	0.0017
	e33m59.a	9.76	0.10	0.0024

Table 4.2. STEPDISC marker selection for DNA markers associated with ≥10% resistance to sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato.

Disease	Marker	Wilk's lambda	Pr < lambda
SPCSV	e41m33.a	0.33	<0.001
	e38m36.u	0.28	<0.001
	K2.650	0.26	<0.001
	AL3.1300	0.24	<0.001
	e44m.41.j	0.23	<0.001
	e39m.41f	0.22	<0.001
SPFMV			
	S13.1130	0.34	<0.001
	e41m37.a	0.31	<0.001
	e40m36.d	0.29	<0.001
	e44m36.d	0.26	<0.001

based on single point ANOVA and multiple-regression analyses with two markers in the model. Unfortunately, a lack of progeny in the SPVD resistant class prevented disclosure of markers linked to SPVD resistance.

I attribute the slight differences between our QTL results and those of Mwanga et al. (2002b) to the fact that the data set that we used contained 2 markers less. Nonetheless my conclusions concur with theirs and do confirm the power of markers e41m33.a and S13.1130 to account for variation in resistance to SPCSV and SPFMV, respectively.

Discriminant and logistic regression analyses have greater commonality with each other than with QTL analysis as far as markers selected is concerned possibly because the former two are both multivariate techniques. These multivariate techniques determine the power of a marker while other markers are still in the model hence giving a more accurate comparative analysis compared to the single point ANOVA for QTL analysis. Single point ANOVA is also referred to as regression on a marker and it assumes the presence of a single QTL on the genome (Broman, 2001). The ANOVA calculates the variance at marker loci without considering the effect of the other markers when they act together hence the value obtained may not be the true effect of the QTL. In his study on genetic mapping of agronomic traits from the interspecific cross of *Oryza sativa* L. and *Oryza glaberrima* Steud., Aluko (2003) concluded that discriminant analysis was superior to interval mapping in selecting markers that could be used in grouping doubled haploid lines of rice into pre-defined groups.

I suggest that logistic regression is an increasingly valuable tool in molecular marker selection protocols in a sweetpotato breeding program. Investigators in the medical sciences have exploited the power of logistic regression to select important markers. In a study involving boars

Table 4.3. Logistic regression selection for DNA markers associated with ≥10% resistance to sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato.

Disease	Marker selected	χ^2 score	Pr>χ^2
SPCSV			
	e41m33.a	58.49	<0.0001
	e38m36.u	11.36	0.0007
	H14.700	7.44	0.0064
	e44m41.j	6.61	0.0101
	e40m41.c	5.71	0.0169
	B7.1130	5.47	0.0194
	e32m60.g	5.31	0.0211
SPFMV			
	S13.1130	57.38	<0.0001
	e40m38.b	8.90	0.0029
	e44m36.d	8.56	0.0034
	e41m37.a	7.56	0.0060
	e39m45.c	6.98	0.0083
	e39m39.e	6.00	0.0143

Table 4.4. Partition for the Hosmer and Lemeshow test for sweetpotato chlorotic stunt closterovirus (SPCSV) and sweetpotato feathery mottle virus (SPFMV) in sweetpotato.

Disease	Observed number of clones	Expected number of clones	Disease resistance group.
SPCSV	45	44.98	1 (present)
SPCSV	42	42.02	0 (absent)
SPFMV	37	36.98	1 (present)
SPFMV	50	50.02	0 (absent)

(*Sus scrofa* L.), Thurston et al. (2002) identified sixteen candidate genetic markers (P<0.005) by comparing the AFLP profile with a semen freezability trait using logistic regression analysis. These findings support the hypothesis that there is a genetic basis for variation in postthaw semen quality between individuals, and that AFLP technology may be able to identify molecular markers linked to genes influencing this variation. Dunsmuir et al. (2000) used logistic

regression modeling to identify 3 immunohistochemistry markers that were statistically significant predictors of the metastatic status (M-stage, bone metastasis vs no bone metastasis) of prostrate cancer in men.

Results from my work suggest that it is possible to use logistic and discriminant multivariate techniques to select powerful markers that may be useful to breeders. While multivariate techniques may not replace QTL analysis, the flexibility of assembling test populations, regardless of relatedness, is particularly useful when progeny from a controlled cross rarely arise as resistant, e.g., SPVD progeny. Our results demonstrated that multivariate techniques can achieve results that are comparable to those of QTL analysis even in the absence of a genetic map and identify markers not disclosed through QTL analysis.

4.4 Literature Cited

Aluko, G.K. 2003. Genetic Mapping of Agronomic Traits from The Interspecific Cross of *Oryza sativa* L. and *Oryza glaberrima* Steud. PhD dissertation. Louisiana State Univ., Baton Rouge, PhD Diss. ETD-1110103-143019.

Broman, K. 2001. Review of statistical methods for QTL mapping in experimental crosses. Lab Animal 30: 44-52.

Buteler, M.I., R.L. Jarret, and D.R. La Bonte. 1999. Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*s. Theor. Appl. Genet. 99:123-132.

Capdevielle, F.M., G.K. Aluko, M. Balzarini, and J.H. Oard. 2000. Application of molecular markers and discriminant analysis to identify rice lines with contrasting phenotypes for agronomic traits. p.216. In: G.S. Khush, D.S. Brar and B. Hardy (eds.) Proceedings of the Fourth International Rice Genetics Symposium, International Rice Research Institute, Los Banos, Philippines (abstr.).

Dunsmuir, W.D., C.E. Gillett, L.C. Meyer, M.P. Young, C. Corbishley, R.A. Eeles, and R.S. Kirby. 2000. Molecular markers for predicting prostate cancer stage and survival. BJU International. 86: 869-878.

Fahima, T., M.S. Roder, K. Wendehake, V.M. Kirzhner, and E. Nevo. 2002. Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor. Appl. Genet. 104:17-29.

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics 7:179-188.
Geddes, A.M.W. 1990. The relative importance of crop pests in Sub-Saharan Africa. Natural Resources Institute Bulletin No. 36. NRI, Kent, UK.

Gibson, R.W., R.O.M. Mwanga, S. Kasule,, I. Mpembe, and E.E. Carey. 1997. Apparent absence of viruses in most symptomless field grown sweetpotato in Uganda. Ann. Appl. Biol. 130:481-490.

Hahn, S.K., E.R. Terry and K. Leuschener. 1981. Resistance of sweetpotato to virus complex. Hortscience. 16:535-537.

Hosmer, D.W., S. Taber, and S. Lemeshow. 1991. The importance of assessing the fit of logistic regression models: A case study. American Journal of Public Health 81(12): 1630-1635.

International Potato Center. 1990. Annual report for 1990. Intl. Potato Center, Lima, Peru.

Jones, A. 1986. Sweet potato heritability estimates and their use in breeding. Hortscience. 21: 14-17.

Karyeija, F., R.W. Gibson, and J.P.T Valkonen. 1998. Resistance to sweetpotato virus disease (SPVD) in wild East African *Ipomoea*. Ann. Appl. Biol. 133:39-44.

Kearsey, M.J. and A.G.L. Farquhar. 1998. QTL analysis in plants; where are we now? Heredity 80: 137-142.

Kriegner, A., J.C. Cervantes, K. Burg, R.O.M. Mwanga, and D. Zhang. 2003. A genetic linkage map of sweetpotato [*Ipomoea batatas* (L.) Lam.] based on AFLP markers. Molecular Breeding 11:169-185.

Lachenbruch, P.A. and M.A. Mickey. 1968. Estimation of error rates in discriminant analysis. Technometrics 10:1-10.

Lubberstedt T., A.E. Melchinger, S. Fahr, D. Klein, A. Dally, and P. Westhoff. 1998. QTL mapping in testcrosses of flint lines of maize: III. Comparison across populations for forage traits. Crop Sci. 38: 1278-1289.

Mukasa, S.B., P.R. Rubaihayo, and J.P.T. Valkonen. 2003. Incidence of viruses and viruslike diseases of sweetpotato in Uganda. Plant Dis. 87:329-335.

Mwanga, R.O.M., G.C. Yencho, and J.W. Moyer. 2002a. Diallel analysis of sweetpotatoes for resistance to sweetpotato virus disease. Euphytica 128:237-248.

Mwanga, R.O.M., A. Kreigner, J.C. Cervantes-Flores, D.P. Zhang, J.W. Moyer and G.C Yencho. 2002b. Resistance to sweetpotato chlorotic stunt virus and sweetpotato feathery mottle virus is mediated by two separate recessive genes in sweetpotato. J. Amer. Soc. Hort. Sci. 127:798-806.

Nelson, J.C. 1997. QGENE: Software for marker-based genomic analysis and Breeding. Molecular Breeding 3:239-245.

Ngeve, J.M. 1990. Yield stability and yield depression in sweetpotato cultivars susceptible to the sweetpotato virus disease. J. Hort. Sci. 65:225-230.

Ostir, G.V. and T. Uchida. 2000. Logistic regression: A non-technical review. American Journal of Physical Medicine and Rehabilitation. 79:565-572.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics. 27:832-837.

SAS Institute Inc. 1999. SAS/STAT User's guide. SAS Institute Inc. Cary, North Carolina.

SAS Institute Inc. 2001. The SAS system for windows v8. SAS Institute Inc. Cary, North Carolina.

Soller, M., A. Genizi, and T. Brody. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47:35-39.

Thurston L.M., K. Siggins, A.J. Mileham, P.F. Watson, and W.V. Holt. 2002. Identification of amplified restriction fragment length polymorphism markers linked to genes controlling boar sperm viability following cryopreservation. Biology of Reproduction 66: 545-554.

Ukoskit, K. and P.G. Thompson. 1997. Autoploidy vs. allopolyploidy and low-density randomly amplified polymorphic DNA linkage maps of sweetpotato. J. Am. Soc. Hort. Sci. 122:822-828.

Wu, K.K., W. Burnquish, M.E. Sorrells, T.L. Pew, P.H. Moore, and S.D. Tanksley. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83:294-300.

Young, N. D. 1999. A cautiously optimistic vision for marker-assisted breeding. Molecular Breeding 5:505-510.

Zhang, D.P., D. Carbajulca, L. Ojeda, G. Rossel, S. Milla, and C. Herrera, and M. Ghislain. 2001. Microsatellite analysis of genetic diversity in sweetpotato varieties from Latin America, pp.295-301.In: Program Report 1999-2000, International Potato Center, Lima, Peru.

CHAPTER 5: MOLECULAR MARKER VARIABILITY FOR SOUTHERN ROOT-KNOT NEMATODE RESISTANCE IN SWEETPOTATO

5.1 Introduction

Among the more than fifty described species of plant parasistic nematodes, root-knot nematodes (*Meloidogyne* spp.) are the leading cause of crop loss (Roberts, 1995). Agriculturally important species include the southern root-knot nematode, *M. incognita*, which is a major pest of sweetpotatoes. Other species that have been known to affect sweetpotato production are *M. arenaria* and *M. javanica* (Giamalva et al., 1963). Jones et al. (1986) reported success in finding resistance to southern root-knot nematode using mass selection techniques. The advantage of mass selection is the possibility of exploiting a wide gene base. Using mass selection, Jones et al. (1991) released two sweetpotato populations designated as I/13 and J/8 to provide a wide genetic base for use with introductions and exotic materials in order to develop enhanced germplasm.

In recent investigations Cervantes-Flores et al. (2002) found differential susceptibility (none, low, medium and high infection levels) of the genotypes they tested to different *Meloidogyne* populations thus suggesting that root-knot nematode resistance in sweetpotato may, in some cases, be quantitatively controlled. The challenge for breeders has so far been to design effective quantitative gene identification protocols through classical breeding methods like mass selection. This effort has however been hampered by the self-incompatibility nature of the sweetpotato (Jones, 1986). Previous studies have shown that identification of suitable genes is an important prerequisite for the success of a breeding program aiming to develop sweetpotatoes with resistance to root-knot nematode (Cervantes-Flores et al., 2002). This view is supported by Ukoskit et al. (1997) who argued that since more than one race is capable of infecting sweetpotatoes, the nematode population being tested must be identified in order to specify the type of resistance from a given source. Furthermore Lawrence (1984) showed that for some

populations, there was greater reproduction of nematodes on the resistant ‘Jewel’ and ‘Jasper’ than on the susceptible ‘Centennial’. However the differences did not correlate with race as determined by the North Carolina race differential tests. Molecular marker techniques, as suggested by Barker and Koenning (1998), would therefore be an important consideration in combining markers for parasitism (virulence) within different nematode populations and host-resistance genes for faster breeding advances.

Ukoskit et al. (1997) evaluated the inheritance of genes for root-knot nematode (race 3) resistance by classical Mendelian procedures based on the distribution of progenies. They used bulk segregant analysis (BSA) to identify a Random amplified polymorphic DNA (RAPD) marker linked to a root-knot nematode resistance gene and also estimated degree of linkage between the resistance gene and the RAPD marker. Out of 760 primers they screened for root-knot nematode resistance in sweetpotato only 1 out of the 9 resultant polymorphic primers could be used to detect linkage between DNA markers and the root-knot nematode resistance gene. Michelmore et al. (1991) stated that BSA requires two DNA bulks derived from a segregating population of F_2 individuals that are homozygous but contrasting for the trait of interest therefore a new cross has to be made and new DNA bulks obtained for each trait to be studied. Previous studies suggest that there is no guarantee of detecting differences between bulks (Ukoskit et al., 1997; Horejsi et al., 2000). In addition in BSA an individual marker is correlated with the trait of interest one at a time. We therefore can not identify groups of markers simultaneously nor can we work with populations of unrelated progeny. Above factors are further indications of the inefficiency of BSA.

A recent DNA marker system, AFLP, has been found to be highly polymorphic. AFLP combined with discriminant analysis has been shown to identify important markers that would

otherwise be difficult to identify. Discriminant analysis applications in molecular marker selection has been extensively discussed by Capdevielle et al. (2000) and Aluko (2003) who associated microsatellites and agronomic traits in rice. Fahima et al. (2002) who investigated microsatellite polymorphism in wheat also used discriminant analysis in their characterization. In sweetpotatoes, Mcharo et al. (2004) used discriminant analysis to select useful AFLP markers that identified variability in dry matter in a USDA sweetpotato collection. Logistic regression has also been used to select AFLP and RAPD markers associated with virus resistance in sweetpotatoes and the results were in agreement with those from a traditional QTL mapping procedure.

In this study I evaluated the capabilities of discriminant and logistic regression analysis to identify AFLP markers that are associated with southern root-knot nematode resistance in two sweetpotato populations. I also compared the similarity or differences among groups of selected markers between the populations.

5.2 Materials and Methods

5.2.1 Planting Material

Two sweetpotato [*Ipomoea batatas* (L.) Lam] F_1 populations were used for southern root-knot nematode race 3 (*Meloidogyne incognita*) resistance expression. Population one consisted of 48 half-sib genotypes developed at the Louisiana State University AgCenter. Maternal clones used to obtain the open pollinated F_1 half-sibs were 'Beauregard', 'Excel', 'L94-96', 'L89-110', 'L86-33' and 'L96-117'. Approximately ten progeny were randomly selected from each parent. The second population consisted of 54 full-sibs developed by the National Agricultural Research Organization, Kampala Uganda, and International Potato Center, Lima Peru, sweetpotato breeding program. The crosses from which the second population was derived were 'Beauregard

x Wagabolige', 'Kyukei No. 63 x Jonathan W218', 'Jonathan W154 x Wagabolige', 'CN1732-4 x Jonathan W218', 'Tanzania x Wagabolige' and open pollinated 'Tanzania'. Approximately eleven progeny were randomly selected from each parent. 'Beauregard' was included as a susceptible control and 'L94-96' was used as a resistant control.

A root-knot nematode population was increased on Bell pepper (*Capsicum annum* L.) cv 'Yolo Wonder' in a greenhouse. Nematode eggs were then extracted from the roots of the 2-month-old seedlings with 0.6% sodium hypochlorite for 4 minutes. The sodium hypochlorite with the eggs was poured through an 80-mesh sieve to remove root and leaf debris then onto a 500-mesh sieve to collect the eggs. The eggs were then washed under running water and suspended in water in standard volumes containing 5000 eggs. Fresh sweetpotato cuttings were planted in 4 inch clay pots in a sterilized 1 sand: 1 soil mixture (v/v). Each of the cuttings was inoculated with 5000 eggs 4 days after planting for population 1 and at planting for population 2. The experiment was laid out as a randomized complete block design with three replicates in a greenhouse and watered as necessary. Observations were made 8 weeks after planting for population 1 and 6 weeks after planting for population 2. Rating for resistance or susceptibility was on the scale of number of egg masses per plant as follows: 0 (0); 1(1-3); 2 (4-10); 3 (11-30); 4 (31-100); 5(> 100), with 0 being the most resistant and 5 the most susceptible.

Nematode eggs were extracted from vigorous fresh sweetpotato plants by dipping the combined roots of the three replications in 0.6% sodium hypochlorite for 10 minutes. As above, the resulting suspension was poured through an 80-mesh sieve and then onto a 500-mesh sieve to collect the eggs. The 500-mesh sieve was backwashed into a beaker, the volume adjusted to 20 ml and diluted as necessary to count the total number of eggs per plant. The eggs per plant were

computed as a check on the egg mass rating and also for determining resistance level frequency distribution in the populations.

5.2.2 DNA Extraction

DNA extraction was previously described by Mcharo et al. (2004). Young leaves were harvested and stored at -40^0 C until needed. Total DNA was isolated from 100mg of fresh leaf tissue using the Genelute plant genome kit (Sigma-Aldrich Inc., St. Louis, Mo).

5.2.3 AFLP Analysis

The DNA samples were amplified in a three-step process using a GeneAmp PCR system 9600 thermocycler (Perkin Elmer, Fulerton, CA). Reagents for AFLP were obtained from Invitrogen™ (AFLP starter primer kit, Cat No. 10483-014) and LI-COR Inc. (Lincoln, NE, Cat. No. 420032). Genomic DNA (120ng/µl) was digested using an *Eco*RI/ *Mse*I restriction enzyme mix at 37 ^{0}C for 3 hours. The enzymes were then inactivated by incubating the mix at 70 ^{0}C for 10 minutes. Double stranded adaptors were then ligated to the restricted DNA fragments resulting in template DNA which was used for pre-amplification. Diluted template DNA (1.5µl) was added onto 10µl pre-amp primer mix, 1.25 units Taq DNA polymerase (Invitrogen™) and 1.25µl RedTaq™ PCR reaction buffer 10x with $MgCl_2$ (Sigma-Aldrich™) to make a 13µl reaction volume. The pre-amplification conditions were 20 cycles each of 94 ^{0}C for 30s, 56 ^{0}C for 60s, 72 ^{0}C for 60s and a final hold at 4 ^{0}C.

The reaction volume for selective amplification consisted of 3.0 µl pre-amplified diluted DNA, 0.4 µl of *Eco*RI (fluorescently labeled) primer (AAG), 4.4µl of *Mse*I (unlabelled) primer, 2.0µl RedTaq™ PCR reaction buffer 10x with $MgCl_2$ (Sigma-Aldrich™), 1.35ul $MgCl_2$ (where 1.35µl were insufficient, the volume was raised to 1.4µl), 1 unit Taq DNA polymerase (Invitrogen™) and 6.38µl of double distilled or AFLP grade water Four primer pairs as

identified by Fajardo (2000) were used for selective amplification (CAG, CTA, CTG, CTT). Blue stop solution (3.0µl) (LI-COR, Lincoln, NE) was added onto each amplified DNA sample. The amplified DNA was then denatured at 95 ^{0}C for 3 minutes and thereafter covered in aluminium foil and placed in a freezer at -20 ^{0}C for 10 minutes to prevent annealing of complementary fragments, before loading onto a 25cm acrylamide gel. PCR amplification fragments were separated by 6.5% acrylamide gel electrophoresis using a LI-COR Global IR^2 sequencer (LI-COR, Lincoln, NE) for 3 hours. The AFLP fragments were automatically detected and recorded during electrophoresis using the LI-COR $SAGA^{MX}$ v 3.1.0 software. Data were collected and presence (=1) or absence (=0) of bands scored using LI-COR $SAGA^{MX}$ v 3.1.0 software. The markers were named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.

5.2.3 Statistical Analysis

Number of eggs per plant was correlated with egg mass rating to determine the level of confidence with which the data could be used for analysis. The mean egg number per genotype was log transformed and the frequency distribution plotted to describe the distribution of resistant and susceptible progeny within each population. Discriminant analysis previously described by Mcharo et al. (2004) was used to select informative molecular markers that are linked to root-knot nematode resistance in the two populations. Logistic regression (Hosmer and Lemeshow, 1989) was also used to model molecular markers as variables describing the resistance trait. Significance level to include a marker for both discriminant analysis and logistic regression was set at P=0.03 for the LSU population. Values for the CIP population were set at P=0.05 (discriminant analysis) and P=0.03 (logistic regression). These P values ensured we selected an optimum set of markers without compromising on model efficiency due to too few

markers or over-fitting due to too many markers selected. A slightly higher P value was used for discriminant analysis on the CIP population to include markers that increased the efficiency of the model developed. Logistic regression and discriminant analysis were done using SAS (1999, 2001). AMOVA on the selected AFLP markers was used to test genotypic variability between the resistant and susceptible groups (Excoiffer et al., 1992).

5.3 Results and Discussion

5.3.1 Response to Inoculation

A correlation analysis of egg mass rating with eggs per plant showed positive significant ($P<0.0001$) associations in both CIP ($r=0.654$) and LSU ($r=0.674$) populations. Descriptive statistics for the variables measured are presented in Tables 5.1 and 5.2. Results obtained from plotting the logarithm of total number of eggs (Figures 5.1 and 5.2) suggest that resistance to root-knot nematode may be qualitatively as well as quantitatively controlled. The LSU population showed a bimodal response to inoculation indicating that a major gene may be controlling the resistance trait with 18 clones assessed as being susceptible and 30 resistant. I found no obvious trend that resistant parents, e.g., ‘Excel’, ‘L94-96’, ‘L89-110’, ‘L86-33’ and ‘L96-117’ produced progeny with greater propensity for resistant reactions. Ukoskit et al. (1997) also obtained a bimodal response among the genotypes they worked on. The cross from which they obtained their F_1 genotypes involved ‘Vardaman’ and ‘Regal’ as the parents. Material used in the United States breeding programs has narrow genetic base (La Bonte, personal communication) and this may partially explain similar qualitative responses between our population and the population used by Ukoskit et al. (1997).

Table 5.1. Descriptive statistics for gall rating, egg count and logarithm transformed egg count for the LSU population.

Variable	Upper 95% confidence interval	Lower 95% confidence interval	Number of observations	Mean	Std. Deviation	Std. Error	Shapiro Wilk test for normality (W)	Pr<W	CV%
Gall rating	1.88	3.03	48	2.46	1.97	0.28	0.84	<0.0001	80.08
Egg count	12311.12	394680.84	48	259245.98	466422.09	67322.23	0.62	<0.0001	179.91
Log Eggs	3.22	4.09	48	3.66	1.49	0.22	0.86	<0.0001	40.78

Table 5.2. Descriptive statistics for gall rating, egg count and logarithm transformed egg count for the CIP population.

Variable	Upper 95% confidence interval	Lower 95% confidence interval	Number of observations	Mean	Std. Deviation	Std. Error	Shapiro Wilk test for normality (W)	Pr<W	CV%
Gall rating	1.28	2.32	54	1.80	1.90	0.26	0.51	<0.0001	219.15
Egg count	3367.60	133393.29	54	8380.44	18365.60	2499.24	0.81	<0.0001	106.17
Log Eggs	2.45	3.17	54	2.81	1.31	0.18	0.95	0.03	46.55

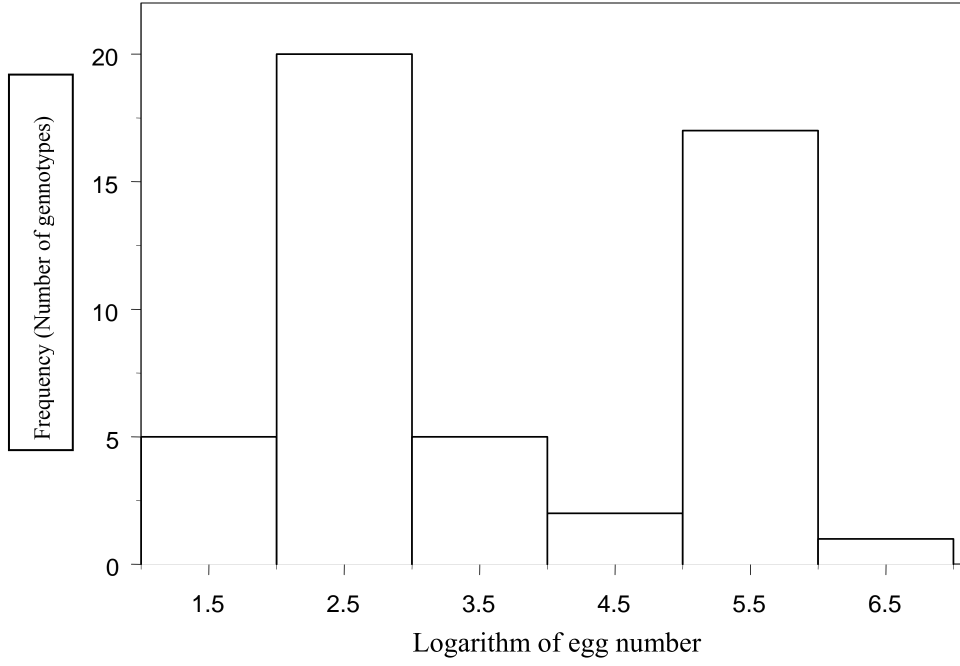

Figure 5.1. Frequency of log total root-knot nematode eggs for the LSU sweetpotato genotypes.

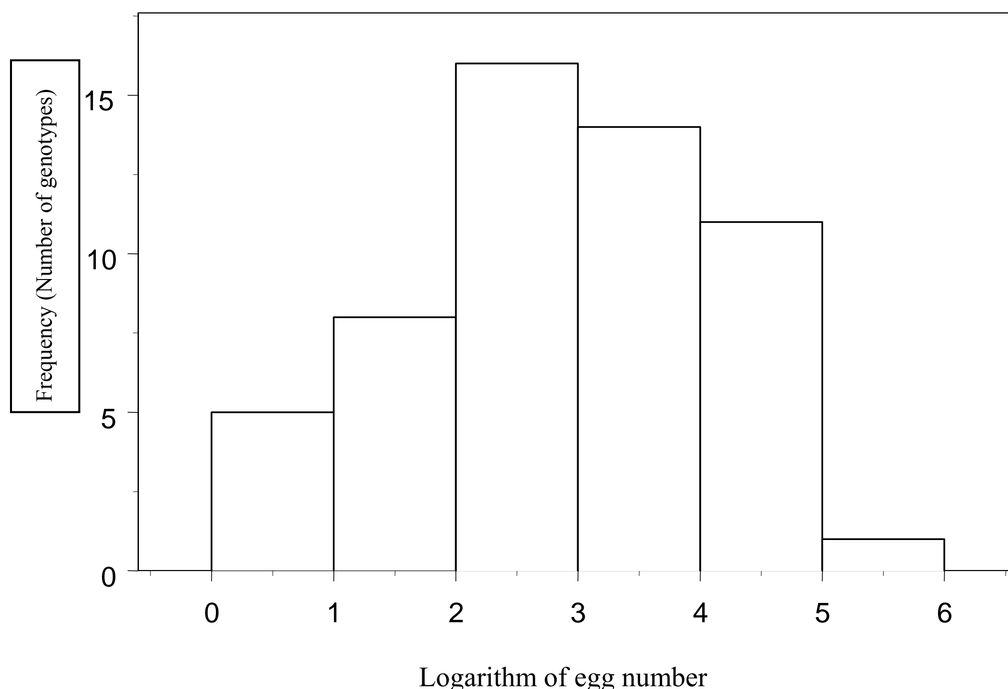

Figure 5.2. Frequency of log total root-knot nematode eggs for the CIP sweetpotato genotypes.

In contrast, genotypes from the CIP population exhibited a quantitative response suggesting that a few major genes may be involved in inheritance of nematode resistance with 14 clones being considered susceptible and 40 resistant. I also noted that the progeny from various CIP parents did not exhibit any differential patterns for resistance. These results agree with those of Cervantes-Flores et al. (2002) who suggested that under some conditions resistance to root-knot nematode may be quantitative. These workers also found that 'Porto Rico' and 'Pelican Processor' had different reactions to the *M. incognita* populations, regardless of the host race. These responses suggested that different genes could be involved in the resistance of sweetpotato to root-knot nematodes.

Early work on sweetpotato resistance to the root-knot nematode includes that of Davide and Struble (1966), Giamalva et al. (1963), and Struble et al. (1966). Resistance to root-knot nematode has previously been explained to be either qualitatively or quantitatively controlled by these and other investigators. In their study on inheritance to resistance, Struble et al. (1966) suggested a multigenic control on inheritance of resistance. They found varying degrees of resistance among the tested progenies from different varieties expressing differing levels of resistance. They also noted differential reactions of the same variety to different populations of *M. incognita,* but no inheritance was determined. Furthermore it is possible that their observations may have been affected by environmental factors like temperature. Lawrence and Clark (1986) also noted that different populations of *M. incognita* varied in their virulence on sweetpotato with some of the populations capable of overcoming resistance previously exhibited by cultivars in the study and some differential interactions. Lawrence et al. (1986) concurred with Struble et al. (1966) when they concluded that the benefits of a resistant cultivar may be affected by field conditions.

5.3.2 Molecular Marker Variation

From my statistical analysis, we did not obtain informative markers that were common to both populations at the set significance levels (Tables 5.3 and 5.4). This may partly be explained by the fact that the two populations exhibit different modes of resistance gene inheritance. This may suggest that there is need for a more extensive study involving different populations with different trait distribution patterns to define an array of markers that may be universally applicable with a certain level of confidence. Due to observed differential interactions by other workers (Cervantes-Flores et al., 2002; Lawrence, 1984) an array of markers may also be needed to account for differing genes for resistance to different populations of *M. incognita.* Lack of

universal applicability of markers is the same challenge that is faced by traditional QTL mapping studies. The quest for marker assisted selection in breeding for resistance to root-knot nematode is a result of complications associated with tedious field observation experiments. For example in a heritability study of resistance to two Meloidogyne species, *M. incognita* and *M. javanica*, Jones and Dukes (1980) found high heritability estimates (0.57 to 0.78) for reactions of sweetpotato parental lines to both species. Although they concluded that development of resistant cultivars is possible, these workers also postulated that determining resistance levels with a high degree of confidence would require experimental analysis using egg mass, gall indices and root necrosis concurrently. This is because a variety may exhibit resistance based on one evaluation while exhibiting susceptibility using another method. Results from our study show that egg mass rating and egg counts were positively and significantly correlated. It is thus expected that marker assisted selection will obviate the need for screening with multiple populations of the pathogen and the tedious phenotypic tests to select progeny carrying the desired allele (Bent and Yu, 1999).

5.3.3 LSU Genotypes

An important aspect of marker assisted selection is the selection of as few informative markers as possible without losing out on the prediction efficiency. Out of a total of 229 polymorphic markers that were generated, five (ctg218, ctg227, cta098, cag267 and ctg232) were selected by both discriminant analysis and logistic regression as having a significant effect on resistance variation (Tables 5.3 and 5.4). AMOVA found significant differences ($P < 0.001$) between resistant and susceptible groups using the five markers for the LSU population. Markers ctg218 and ctg227 were consistently selected by the two statistical procedures as being strongly associated with the resistance trait. We suggest that future studies focus on the two markers

strongly associated with resistance in an effort towards identifying markers that would be useful for breeding against root-knot nematode. Using the seven significant markers selected by discriminant analysis we achieved a population prediction efficiency of 88.78% with 6 out of 48 genotypes misclassified by the model. Two of the 6 were susceptible according to the nematode count but were classified as resistant using molecular markers. The other four were phenotypically classified as resistant with three of them being of intermediate resistance but all four were classified as susceptible by the molecular marker method. Further model expansion resulted in a 97.22% correct population classification rate using 14 markers and 100% using 18 markers (cag279, cag267, cag259, cag108, cta213, cta155, cta144, cta098, ctg284, ctg232, ctg227, ctg218, ctg188, ctg146, ctg088, ctg079, ctt113 and ctt081). Consequently a gain of only 12% in a selection program using an extra 12 markers may not be very cost effective.

Efficiencies of various discriminant analysis models with differing number of markers are presented in Table 5.5. There does not seem to be any linear relationship between number of markers added into the model and the decrease in classification errors. The increase in error rate for the CIP population using a model with 10 markers may be due to addition of markers that have a strong but negative association with resistance.

5.3.4 CIP Genotypes

Compared to the LSU population there were four common markers (ctg228, cag118, cag108 and cta148) selected by both logistic regression and discriminant analysis out of a total of 220 polymorphic markers generated (Tables 5.3 and 5.4). AMOVA found significant differences (P <0.001) between resistant and susceptible groups in the CIP population using the four markers. Consequently it is recommended that further research involve these four markers from

Table 5.3. STEPDISC marker selection for DNA markers associated with resistance to southern root-knot nematode in two sweetpotato populations.

Population	Marker[1]	Partial R-square[2]	Wilks' lambda[3]	Pr < lambda
LSU	ctg218	0.133	0.867	0.0107
	ctg227	0.110	0.771	0.0029
	cta098	0.111	0.685	0.0008
	ctg232	0.164	0.573	<0.0001
	cag267	0.185	0.467	<0.0001
	cag259	0.190	0.378	<0.0001
	ctt113	0.112	0.336	<0.0001
CIP				
	ctg228	0.247	0.753	0.0001
	ctt126	0.239	0.573	<0.0001
	cag118	0.149	0.488	<0.0001
	cag108	0.081	0.448	<0.0001
	cta148	0.104	0.401	<0.0001

[1]Markers are named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.
[2]Partial R-square is the marginal variability accounted for by a variable when all others are already included in the model.
[3]Wilks' lambda is the likelihood ratio measure of a marker's contribution to the discriminatory power of the model.

the CIP population. A group prediction model created using the five significant markers selected by discriminant analysis (Table 5.3) achieved 88.04% prediction efficiency with 4 out of 55 clones misclassified in the whole population. Out of the 4 clones, 3 were previously classified as being highly susceptible using nematode egg count while the fourth had intermediate resistance. Further investigations revealed that gains from additional markers in the model were minimal with 15 markers giving 93.86% correct classification in the population and 16 markers (cag268, cag217, cag213, cag195, cag118, cag116, cag108, cta309, cta211, cta148, cta124, cta081, cta071, ctg228, ctg110 and ctt126) resulting in 100% correct classification in the population.

Table 5.4. Logistic regression selection for DNA markers associated with resistance to southern root-knot nematode in two sweetpotato populations.

Population	Marker selected[1]	χ^2 score[2]	Pr>χ^2
LSU			
	ctg218	6.400	0.0114
	ctg227	5.534	0.0186
	cag198	5.596	0.0180
	cta098	5.127	0.0236
	cta183	8.008	0.0047
	cag267	13.792	0.0002
	cag232	8.000	0.0047
CIP			
	ctg228	13.335	0.0003
	cag118	12.015	0.0005
	cta148	4.999	0.0253
	cag108	6.790	0.0092
	cta081	6.300	0.0121
	cta237	9.473	0.0021
	cta172	19.001	<0.0001

[1]Markers are named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.
[2]χ^2 score is the largest significant score for maker not in model to be included in the model.

Table 5.5. Rate of correct classification of sweetpotato clones into nematode resistance groups after cross-validation in discriminant analysis.

Population	Number of predictor markers	Resistant group error rate	Susceptible group error rate	Total error rate
LSU	7	0.133	0.111	0.122
	8	0.100	0.111	0.106
	14	0	0.056	0.028
	17	0	0.056	0.028
	18	0	0	0
CIP				
	5	0.025	0.214	0.120
	6	0.075	0.071	0.073
	10	0.100	0.143	0.122
	15	0	0.143	0.072
	16	0	0	0

Investigations by Ukoskit et al. (1997) found a low level of linkage (0.2421) between the identified marker and the resistance gene. These workers further recommended that it is important to find more molecular markers associated with the resistance trait to increase the efficiency of screening seedlings. The multiple markers identified for both populations in my study shows that such markers exist. The high levels of classification efficiency provide further proof that there are significant gains to be achieved in using multiple markers for progeny selection. Use of only the top ranked marker (ctg218) in discriminant analysis for the LSU population resulted in a lower cross-validated population classification efficiency to 67.78% while use of only the top ranked marker for the CIP population reduced the population classification efficiency to 73.69%. We therefore recommend that future investigations for nematode resistance in sweetpotatoes involve use of multiple markers whether the genotype frequency distribution suggests qualitative or quantitative inheritance. Tables 5.6 and 5.7 present statistics used in discriminant analysis marker selection. The partial R-Square value for each marker is low thus suggesting that each has a minimal effect on the power of the model.

In their review, Bent and Yu (1999) showed that in disease resistance investigations, molecular markers have been primarily used to select for single genes that have a clear, major and dependable effect on phenotype. Such selections ensure that the phenotype of interest will most likely be advanced through breeding lines. Breeding efficiency may be further improved by determining the type of linkage because if a resistance locus is linked in repulsion to other desirable loci, marker-based selection can greatly reduce the time and space needed to generate the desired allelic combinations. Marker assisted selection will likely play an important role in identifying sources of nematode resistance since there is still debate on the mode of resistance

Table 5.6. Discriminant analysis marker selection statistics for a model that achieved 100% correct classification in the LSU population.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	ctg218	0.133	7.080	0.011	0.867	0.011
2	ctg227	0.110	5.580	0.023	0.771	0.003
3	cta098	0.111	5.500	0.024	0.685	0.001
4	ctg232	0.164	8.420	0.006	0.573	<.0001
5	cag267	0.185	9.520	0.004	0.467	<.0001
6	cag259	0.190	9.640	0.004	0.378	<.0001
7	ctt113	0.112	5.060	0.030	0.336	<.0001
8	cta213	0.108	4.730	0.036	0.299	<.0001
9	ctg079	0.132	5.770	0.021	0.260	<.0001
10	ctg188	0.085	3.420	0.072	0.238	<.0001
11	ctg088	0.094	3.750	0.061	0.216	<.0001
12	cag108	0.111	4.380	0.044	0.192	<.0001
13	ctg146	0.197	8.360	0.007	0.154	<.0001
14	cta144	0.091	3.320	0.078	0.140	<.0001
15	cag279	0.103	3.680	0.064	0.125	<.0001
16	ctg284	0.081	2.720	0.110	0.115	<.0001
17	cta155	0.101	3.360	0.077	0.104	<.0001
18	ctt081	0.104	3.360	0.077	0.093	<.0001

inheritance; quantitative, qualitative or a mixture of the two. In potatoes the gene $H_{1,}$ which confers a high level of resistance to the golden nematode (a cyst nematode) has been bred into several potato cultivars (Brodie, 1999). Mapping of the H_1 gene has led to the development of amolecular marker to screen segregating populations for resistance to the golden nematode. According to Brodie (1999) limited success has been realized in control of root-knot nematodes in potatoes although sources of resistance have been identified. It is possible that the golden nematode is an introduced species that has had little time to evolve in the US. Compared to the root-knot nematode there is a possibility that the golden nematode represents a much more genetically homogeneous pathogen (Clark, personal communication).

Table 5.7. Discriminant analysis marker selection statistics for a model that achieved 100% correct classification in the CIP population.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	ctg228	0.247	17.050	0.000	0.753	0.000
2	ctt126	0.239	16.020	0.000	0.573	<.0001
3	cag118	0.149	8.750	0.005	0.488	<.0001
4	cag108	0.081	4.340	0.042	0.448	<.0001
5	cta148	0.104	5.570	0.022	0.401	<.0001
6	cta124	0.077	3.910	0.054	0.371	<.0001
7	cag268	0.083	4.180	0.047	0.340	<.0001
8	cag195	0.113	5.700	0.021	0.301	<.0001
9	cta211	0.103	5.050	0.030	0.270	<.0001
10	ctg110	0.117	5.710	0.021	0.239	<.0001
11	cta081	0.108	5.070	0.030	0.213	<.0001
12	cta071	0.118	5.500	0.024	0.188	<.0001
13	cag116	0.102	4.560	0.039	0.169	<.0001
14	cta309	0.147	6.720	0.013	0.144	<.0001
15	cag213	0.216	10.470	0.003	0.113	<.0001
16	cag217	0.201	9.280	0.004	0.090	<.0001

Models based on large samples are expected to be more reliable. Our samples were about 50 clones in each group and hence may not have provided sufficiently large numbers for efficient marker selection. This may partly explain why one single marker with a large effect could not be found for the LSU population that has a bimodal distribution as in the case of Ukoskit et al. (1997). Large samples are expected to have a wider range of markers from which to choose and more clones which provide greater variability within the disease tolerance response variable. However our results suggest that there is utility in using small sample sizes not only for development of protocols but also for studying important traits that may occur in small proportions within a population. According to Cruz-Castillo et al. (1994) the combination of too many variables and a small sample size may cause unreliable estimates. They further suggest that discriminant analysis may not provide meaningful results where there are too many markers

selected even with large samples. However, according to them a reliable approach would be to use a sample size that is about 10 times the number of markers selected per phenotypic group. In addition samples sizes smaller than the corresponding number of markers should be avoided.

A fundamental difference between the multivariate approach and other marker selection techniques like bulked segregant analysis or QTL analysis is that while the latter two seek markers that may be linked to the gene of interest, multivariate analysis selects an array of markers that can be used to predict a clone of unknown resistance status into a predefined resistance group. Therefore in the complicated sweetpotato genome where mapping is very resource consuming, multivariate techniques seem to present a better approach for trait-linked marker selection. Our results indicate that whether studies on resistance to root-knot nematode resistance involve segregating populations (Ukoskit et al., 1997) or non-segregating populations, the mystery of mode of resistance inheritance still needs to be unraveled, regardless of the marker generation technique used.

5.4 Literature Cited

Aluko, G.K. 2003. Genetic Mapping of Agronomic Traits from The Interspecific Cross of *Oryza sativa* L. and *Oryza glaberrima* Steud. PhD dissertation. Louisiana State Univ., Baton Rouge, PhD Diss. ETD-1110103-143019.

Barker, K.R. and S.R. Koenning. 1998. Developing sustainable systems for nematode mangament. Annual Review of Phytopathology. 36: 165-205.

Bent, A.F. and I.C. Yu. 1999. Applications of molecular biology to plant disease and insect resistance. Advances in Agronomy, 66:251-298.

Brodie, B.B. 1999. Classical and molecular approaches for managing nematodes affecting potato. Canadian Journal of Plant Pathology, 21:222-230.

Capdevielle, F.M., G.K. Aluko, M. Balzarini, and J.H. Oard. 2000. Application of molecular markers and discriminant analysis to identify rice lines with contrasting phenotypes for agronomic traits. p.216. In: G.S. Khush, D.S. Brar and B. Hardy (eds.) Proceedings of the Fourth International Rice Genetics Symposium, International Rice Research Institute, Los Banos, Philippines (abstr.).

Cervantes-Flores, J.C., G.C. Yencho, and E.L. Davis.2002. Host reactions of sweetpotato genotypes to root-knot nematodes and variation in virulence of *Meloidogyne incognita* populations. Hortscience, 37:1112-1116.

Cruz-Castillo, J.G., Ganeshanandam, S., MacKay, B.R., Lawes, G.S., Lawoko, C.R.O. and Woolley, D.J. 1994. Applications of canonical discriminant analysis in horticultural research. Hortscience 29:1115-1119.

Davide, R.G. and F.B. Struble. 1966. Selection from a field population for variability in *Meloidogyne incognita* on sweetpotato. Philipine Agriculturist 50:15-29.

Excoiffer, L., P.E. Smouse, and J.M. Quattiro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131:479-491.

Fahima, T., M.S. Roder, K. Wendehake, V.M. Kirzhner, and E. Nevo. 2002. Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor. Appl. Genet. 104:17-29.

Fajardo, D.S. 2000. Genetic diversity in Papua New Guinea sweetpotato [*Ipomoea batatas* (L.) Poir] germplasm. MSc thesis, Louisiana State University, Baton Rouge.

Giamalva, M.J., W.J. Martin, and T.P. Hernandez. 1963. Sweetpotato varietal reaction to species and races of root-knot nematodes. Phytopathology 53:1187-1189.

Horejsi, T., J.E. Staub, and C. Thomas. 2000. Linkage of random amplified polymorphic DNA markers to downy mildew resistance in cucumber (*Cucumis sativus* L.). Euphytica 115:105-113.

Jones, A. 1986. Sweet potato heritability estimates and their use in breeding. Hortscience. 21: 14-17.

Jones, A. and P.D. Dukes. 1980. Heritabilities of sweetpotato resistances to root-knot caused by *Meloidogyne incognita* and *M. javanica*. J. Amer. Soc. Hort. Sci. 105:154-156.

Jones, A., P.D. Dukes, J.M. Schalk and M.G. Hamilton. 1991. I/13 and J/8 sweetpotato mass selection populations. Hortscience 26:929-930.

Jones. A, P.D. Dukes, and J.M. Schalk. 1986. Sweetpotato breeding, pp. 1-35. In: M.J. Basset (Ed.). Breeding vegetable crops. AVI, Westport, Conn.

Lawrence, G. W. 1984. The Role of Sweet Potato Resistance in the Population Dynamics and Development of *Meloidogyne incognita*. PhD dissertation. Louisiana State University, Baton Rouge.

Lawrence, G.W and C.A. Clark. 1986. Identification, race determination, and pathogenicity of root-knot nematodes to resistant and susceptible sweetpotatoes. J. Nematol. 18:617 (Abstr.).

Lawrence, G.W., C.A. Clark, and V.L. Wright. 1986. Influence of Meloidogyne incognita on resistant and susceptible sweetpotato cultivars. J. Nematol. 18:59-65.

Mcharo, M., E. Bush, D. La Bonte, C. Broussard, and L. Urbatsch. 2003. Molecular and morphological investigation of ornamental Liriopogons . J. Amer. Soc. Hort. Sci. 128: 575-577.

Mcharo, M., D.R. Labonte, J.H. Oard, S.J. Kays, and W.J. McLaurin. 2004. Linking quantitative traits with AFLP markers in sweetpotatoes using discriminant analysis. Acta Hort. 637:285-293

Michelmore, R.W., Paran, I. and R.V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulk segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. USA 88:9829-9832.

Roberts, P.A. 1995. Conceptual and practical aspects of variability in root-knot nematodes related host-plant resistance. Annual Review of Phytopathology, 33:199-221.

SAS Institute Inc. 1999. SAS/STAT User’s guide. SAS Institute Inc. Cary, North Carolina.

SAS Institute Inc. 2001. The SAS system for windows v8. SAS Institute Inc. Cary, North Carolina.

Struble, F.B., L.S. Morrison, and H.B. Cordner. 1966. Inheritance of resistance to stem rot and root-knot nematode in sweetopotato. Phytopathology 56:1217-1219.

Ukoskit, K., P.G. Thompson, C.E. Watson Jr., and G.W. Lawrence. 1997. Identifying a randomly amplified polymorphic DNA (RAPD) marker linked to a gene for root-knot nematode resistance in sweetpotato. J. Amer. Soc. Hort. Sci. 122:818-821.

CHAPTER 6: COMPARISON OF LOGISTIC REGRESSION AND DISCRIMINANT ANALYSIS IN ASSOCIATING AFLP MARKERS WITH TOTAL SUGAR AND ß-CAROTENE AMONG SWEETPOTATO CLONES

6.1 Introduction

The importance of the nutritive value of the sweetpotato as human food and animal feed has been well documented by Woolfe (1992). Rubatzky and Yamaguchi (1997) state that as a food, the roots may be roasted, baked, fried or prepared in various combination dishes, providing a rich source of carbohydrates and Pro-vitamin A. Pro-vitamin A, or ß-carotene, and total sugar content have been found to be important consumer traits for various markets (Woolfe, 1992). ß-carotene has been associated with prevention of night blindness in children (West, 1994) and orange fleshed sweetpotato cultivars being low input crops are therefore attractive sources for ß-carotene for low income societies.

Recent studies on root total sugar and ß-carotene content have focused on physiological assessment and little on breeding strategies (Bushway, 1986; Katayama et. al., 1996; K'Osambo et. al., 1998; Labonte et. al., 2000; Lu and Sheng, 1990 and Picha, 1987). Takahata et al. (1993) studied the relationship between β-carotene content and Hunter color values in sweetpotato cultivars. They found that color value *a* had the highest correlation coefficient (0.891) with β-carotene and concluded that the color value *a* could be used in rapidly estimating the β-carotene content in breeding programs.

Most breeding programs aiming to raise the ß-carotene content and alter the sugar content in cultivars have used a classical approach of mass selection (Jones, 1965; 1986), however genetic advance in breeding sweetpotatoes with desirable traits using this approach has been slow and resource consuming. The slow breeding progress may be attributed partly to the low success in inter-specific crosses thus precluding use of genetically related species for sweetpotato

improvement (Hall and Phatak, 1993). Alternative procedures to supplant classical approaches include use of molecular marker assisted selection. However, published reports on breeding methods that incorporate molecular marker approaches for organoleptic traits are uncommon. Molecular techniques that would effect inter-specific gene transfer could also greatly improve quantitative breeding programs (Hall and Phatak, 1993).

Molecular marker assisted selection in sweetpotatoes began with the use of bulked segregant analysis (BSA) (Michelmore et. al., 1991; Ukoskit et. al., 1997) to detect linkage between DNA markers and the root-knot nematode resistance gene. Another approach, described by Mcharo et al. (2004), is the use of multivariate statistical techniques like discriminant analysis to associate molecular markers with quantitative traits. Due to the sparsely saturated genome maps in sweetpotato, I seek to develop a selection method that does not require *a priori* genetic maps and also one that is capable of analyzing multiple traits (variables) simultaneously. The multiple variables or molecular markers in my case were generated by Amplified Fragment Length Polymorphism (AFLP) (Vos et. al., 1995). I propose to use cluster analysis, discriminant and logistic regression analyses in an AFLP marker-trait association study with these objectives.

1. Identify trait-linked AFLP molecular markers for total sugar content.
2. Identify trait-linked AFLP molecular markers for ß-carotene content.
3. Compare the effectiveness of discriminant analysis with that of logistic regression in classifying genotypes into predefined phenotypic classes.

6.2 Materials and Methods

6.2.1 Planting Material

One sweetpotato F_1 population was used for β-carotene and total sugar assessment. The population consisted of 73 half-sib genotypes developed at the Louisiana State University

AgCenter. Maternal clones used to obtain the open pollinated F_1 half-sibs were 'Beauregard', 'Excel', 'L94-96', 'L89-110', 'L86-33' and 'L96-117'. Approximately twelve progeny were randomly selected from each parent. 'Beauregard' and 'L94-96' were used as controls. F_1 'Beauregard' progeny were coded as 'BX' while 'Excel' F_1 progeny were coded as 'EL'. The rest of the F_1 progeny were given the codes '96', '110', '33' and '117' to reflect parentage. Field trials were conducted during the summer at the LSU AgCenter Sweetpotato Research Station, Chase, Louisiana in 2001 and at the LSU AgCenter Burden Research Station, Baton Rouge, Louisiana in summer 2002. Each experiment was laid out as a randomized complete block trial with 3 replications. Each plot consisted of 5 plants with an inter-row spacing of 60 cm and a within row spacing of 30 cm. Field management followed commercial sweetpotato cultural practices recommended by Boudreaux (1994). The crop was harvested after 4 months and 4 to 5 roots from all the plants in a plot were combined for post-harvest laboratory analysis. The roots were cured at 29^0 C and 90 to 98% relative humidity for about 10 days. Thereafter the roots were stored at 15^0C and 90% relative humidity.

6.2.2 DNA Analysis

DNA extraction and AFLP marker generation and analysis were previously described in chapter 5.

6.2.3 Total Sugar Analysis

Analyses of both total sugar and ß-carotene content were done on raw roots stored for 4 to 10 weeks after curing, during which sugar content changes are minimal (Picha, 1987). According to Picha (1987) carbohydrates, mainly starch and sugars, constitute most of the dry matter in sweetpotatoes. Sucrose, glucose and fructose are the main sugars in raw sweetpotatoes and sweeter baked sweetpotatoes are usually more acceptable to consumers. The three separate

sugars were quantified using high performance liquid chromatography (HPLC) procedures for quantitative analysis of sugars in raw sweetpotatoes as described by Picha (1985, 1987) using a SUPELCOSIL™ LC-NH_2 column (SUPELCO, Bellefonte, PA). Unpeeled roots were halved longitudinally and uniformly grated over the entire surface to a depth of about 3 mm. The grated tissue from each of the 4 roots per replication was combined and 10.0g was homogenized in 80% ethanol for 1 min at high speed using a Brinkman homogenizer (Brinkman Instruments, Westbury, NY). The resulting slurry was immediately boiled for 15 min, cooled, and filtered through Whatman #4 paper. The residue and original container were washed with 80% ethanol and the filtrate was made to a final volume of 100ml. 20µl of the sample were then injected into the HPLC machine for analysis. Total sugar values were obtained by summing fructose, glucose and sucrose values for every clone.

6.2.4 ß-carotene Analysis

Hunter color value *a* (Hunter, 1958) was used as an initial estimate of the ß-carotene content and thereafter transformed using the following linear regression model (Takahata et al., 1993):

$$\text{Carotene content} = 0.864a - 8.68$$

These workers linked ß-carotene content and the intensity of the orange flesh color of the sweetpotato using a linear model that results in positive ß-carotene content values and zero or negative values when the β-carotene content is negligible or undetectable. The orange color intensity was measured using the Hunter color system (Hunter, 1958). Hunter color value measurements were taken on the flesh of representative cured roots from each plot using a Minolta spectrophotometer cm 3500d (Minolta Co., Osaka, Japan). This color system is based on *L, a* and *b* measurements where *L*=lightness, *a*=green-red scale, *b*=blue-yellow scale. This

method measures color by using a positive and negative number scale. For the color value *a*, a positive value is perceived as approaching red, a negative value as approaching green. For the color value *b*, a positive value is perceived as tending to yellow, a negative value as tending to blue. Lightness (*L*) is measured on a scale of 0-100 where 0 = black and 100 = white. Flesh color measurements were taken by measuring a cross-section of the interior of the root.

6.2.5 Statistical Analysis

Cluster analysis according to Ward's method (SAS, 1999) was used to define *a priori* groupings to be used in discriminant analysis. A constraint of two groups was applied in the model to limit the output to two groups; a high value and a low value group for total sugar and β-carotene content separately. From the original array of AFLP generated molecular markers the most informative markers were selected using step discriminant analysis, STEPDISC procedure, and logistic regression using PROC LOGISTIC (SAS, 2001). A significance level to include a marker for both selection methods was set at P=0.03. Higher probability values were inefficient and resulted in models that had more molecular markers yet were less efficient in classifying the genotypes (data not presented). The selected markers were then used in a discriminant analysis, DISCRIM option (SAS, 2001; Mcharo et al., 2004), to develop and validate a phenotypic group prediction model and to predict group membership of the test genotypes. The performance of the discriminant criterion was evaluated by group specific error count estimates during cross-validation. The error rate estimator gives the proportion of misclassified observations in each group. The model was formulated using the markers selected based on the 73 previously clustered clones. Since the primary goal is to breed for high β-carotene and low sugar content we were interested in efficient classification into each of these categories.

PROC LOGISTIC with the logit link was used to perform logistic regression analysis to select markers that were significantly associated with the phenotype. The phenotype was a binary outcome as either a high or low group value. This being a multiple regression and hence complicated, the prediction model was developed on a spreadsheet using the generated parameter coefficients. The phenotypic group probabilities obtained were used to classify the genotypes to belong to either high or low total sugar or β-carotene group as presented in the results section. The logistic model was of the form:

$$\text{Phenotypic group probability} = \frac{\exp(\alpha + \beta_1 * m_1 + \beta_2 * m_2 + \ldots + \beta_i * m_i)}{[1 + \exp(\alpha + \beta_1 * m_1 + \beta_2 * m_2 + \ldots + \beta_i * m_i)]}$$

where:

α is the intercept of the model and β_i is the increment in log odds for selected marker m_i

Phenotypic group probability is a probability value approaching 0 or 1.

The Akaike Information Criterion (AIC) model fit statistics for logistic regression were computed using the following formula as described in SAS (1999):

$AIC = -2\text{Log } L + 2(k + s)$ where:

k is the number of response levels (phenotype group) minus 1 and s is the number of explanatory variables (molecular markers) and:

$-2\text{Log } L = -2 \sum_j w_j f_j \log(p_j)$.

For the jth observation (genotype), p_j is the estimated probability of the observed response (phenotypic group) and w_j and f_j are weight and frequency values for the jth genotype.

Analysis of molecular variance on the selected AFLP markers was used to test genotypic variability between the low and high groups (Excoiffer et al., 1992).

6.3 Results and Discussion

6.3.1 Total Sugar Content

Phenotypic grouping of the clones for total sugar was based on cluster analyses. Total sugar content ranged from 0.59 µg/gm fresh-weight of sweetpotato in the clone 96-9 to 1.10µg/gm fresh-weight for sweetpotato clone 33-23 (Table 6.1). We found no obvious trend that parents, 'Beauregard', 'Excel', 'L94-96', 'L89-110', 'L86-33' and 'L96-117' produced progeny with greater propensity for high or low sugar or β-carotene content. Total sugar is a sum of fructose, glucose and sucrose content. Each of these sugars has a different magnitude of contribution of sweetness to the overall flavor of the sweetpotato (Wang and Kays, 2003) and their relative concentrations vary among genotypes (Picha, 1987). Since a primary objective of this study was to compare models developed by discriminant analysis and logistic regression, separate models were developed using each of the statistical procedures. The probability of a clone being in a low total sugar group as calculated by logistic regression is presented in Table 6.1. All the clones classified as having high sugar content (H) had a probability of having low total sugar content as 0 or approaching 0 and all the clones that were classified as having low total sugar content (L) had a probability of 1 or close to 1 to belong to this group. Under an ideal model the sum of probabilities for all the clones would equal the sum of probabilities for the low sugar group or the number of clones in the low sugar group. The sum of probabilities for the high sugar group would then be zero. Our results indicate that the sum of probabilities for the low sugar group is 36.912 which is very close to 37. The sum of probabilities for the high sugar group is 0.088.

A clone that had a probability of ≤ 0.50 was considered to belong to the low sugar group while one that had a probability of >0.50 was classified as belonging to the high sugar group.

Using these criteria the model developed by logistic regression achieved 100% correct classification. The molecular markers used in the model, their coefficients and the chi-square criterion used to select markers are presented in Table 6.2. A total of 8 markers were selected using step discriminant analysis (Table 6.3). The low partial R^2 for markers selected by discriminant analysis values suggested that each of the markers accounts for just a small percentage of the variation for total sugar content. This is consistent with what would be expected of a quantitative trait like total sugar. The correct population prediction rate for the discriminant analysis model after cross-validation was 80.82%. Seven clones (117-3, 94-96, 96-10, 96-17, 110-13, 33-19, and 33-23) that were phenotypically classified as having high total sugar content were reclassified into the low sugar group by discriminant analysis. Seven clones (33-7, 110-4, 96-8, BX-17, 110-15, 110-30 and BX-25) previously in the low sugar group were also reclassified into the high sugar group. Comparison of these models is based on the significance level of P=0.03. When a significant level of P=0.05 was imposed, the number of markers selected by step discriminant analysis were too many to deal with efficiently.

In most societies, sweetpotatoes are processed before consumption and the post-harvest procedures include baking, boiling and roasting. Processing and especially heat related processing has the effect of transforming the stored starch to maltose, which is the main sugar in baked sweetpotatoes (Picha, 1985). It is therefore expected that different phenotypic groups with different clones as members would have been obtained if classification was based on processed sweetpotatoes. However, it is not certain what markers would be selected with maltose content as the grouping variable. Studies involving maltose as a classification variable would therefore give useful information to a breeder for a more successful breeding program. The present study involved F_1 genotypes which are highly heterogeneous for root size, and root size affects

uniformity of baking. Investigations on marker-sugar association in baked roots may therefore yield better results if established cultivars that provide uniform roots are used. Multivariate tools overcome the problem of unrelated genotypes encountered by QTL analysis and therefore useful results will be realized from study involving established cultivars that may be unrelated. I therefore suggest that future studies focus on maltose as one of the main criteria for genotype classification and model development.

Table 6.1. Total sugar content (fresh weight basis) and resultant phenotypic classification of sweetpotato clones from an F_1 polycross of 'Beauregard', 'Excel', 'L94-96', 'L86-33', 'L89-110' and '96-117'.

Clone[1]	Total sugar (µg/g fresh weight)	Probability[2]	Group[3]	Clone	Total sugar (µg/g fresh weight)	Probability	Group
33-23	1.10	0.006	H	BX-23	0.76	1	L
96-20	1.06	0.001	H	EL-4	0.76	0.996	L
96-10	0.97	0.001	H	110-29	0.76	1	L
BX-16	0.93	0	H	96-2	0.76	0.996	L
EL-24	0.93	0	H	110-21	0.75	1	L
EL-15	0.92	0.005	H	EL-3	0.75	1	L
96-18	0.90	0	H	110-15	0.75	1	L
117-3	0.90	0	H	BX-20	0.75	1	L
EL-27	0.90	0	H	110-23	0.74	1	L
96p	0.88	0.021	H	96-5	0.74	0.996	L
BX-15	0.88	0	H	117-28	0.74	1	L
96-7	0.87	0	H	96-15	0.74	0.986	L
BX-5	0.87	0	H	117-8	0.74	1	L
BX-2	0.86	0.013	H	BX-25	0.73	0.986	L
96-27	0.86	0	H	117p	0.73	1	L
33-18	0.85	0	H	33-7	0.73	0.999	L
EL-7	0.85	0	H	110-24	0.73	1	L
BX-13	0.85	0	H	BX-1	0.73	1	L
EL-20	0.84	0	H	EL-14	0.73	0.996	L
EL-8	0.84	0	H	96-8	0.73	0.997	L
BX-29	0.84	0	H	EL-5	0.72	0.996	L
BX-18	0.84	0	H	110-18	0.71	1	L
96-11	0.84	0	H	117-2	0.70	1	L
BX-8	0.84	0.001	H	110-14	0.69	1	L
33-9	0.84	0.005	H	BX-17	0.69	1	L
110-13	0.84	0	H	117-6	0.68	0.996	L
96-17	0.83	0	H	BX-26	0.68	1	L

(table continued)

110-25	0.82	0	H	33-30	0.66	1	L
33-1	0.82	0.005	H	EL-16	0.66	1	L
110-26	0.81	0.001	H	BXp	0.66	0.979	L
BX-24	0.81	0	H	33-11	0.66	1	L
BX-22	0.81	0	H	117-25	0.65	1	L
BX-6	0.81	0	H	110-22	0.64	1	L
33-19	0.80	0.024	H	33-16	0.63	1	L
110-1	0.78	0.005	H	110-4	0.61	0.989	L
EL-11	0.78	0	H	110-30	0.61	1	L
				96-9	0.59	1	L
Sum		0.088	36			36.912	37

[1]Clones followed by the letter 'p' are parental clones
[2]Probability of being in a low total sugar group.
[3]H – High sugar clones and L – Low sugar clones as determined by HPLC analysis.

Table 6.2. Logistic regression selection for DNA markers in sweetpotato associated with total sugar content.

Marker entry step	Marker[1]	Estimate (β)	Chi-square score[2]	Pr > ChiSq
0	Intercept	-30.733		
1	ctg134	-36.060	6.649	0.010
2	cta209	79.862	7.583	0.006
3	cta122	-10.712	6.791	0.009
4	cta223	36.403	4.842	0.028
5	cta168	-59.960	6.765	0.009
6	cag224	-35.169	6.554	0.011
7	ctt272	-84.002	7.357	0.007
8	ctg066	-67.694	9.155	0.003
9	ctt090	-18.635	8.367	0.004
10	cta211	-26.324	5.207	0.023
11	ctt207	-27.883	8.944	0.003
12	ctt326	-9.100	8.290	0.004
13	ctg201	12.250	5.004	0.025

[1]Markers are named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.
[2]χ^2 score is the largest significant score for maker not in model to be included in the model.

Performances of different sized models are presented in Table 6.4 for discriminant analysis and Table 6.5 for logistic regression. Modeling aims to have as few variables as possible in the model and from the results presented logistic regression outperforms discriminant analysis. Table 6.6 presents marker selection criteria used by discriminant analysis. Discriminant analysis is

traditionally conducted using an ‘ordinary least squares’ approach while logistic regression uses a ‘maximum likelihood estimate’ approach. Least squares is a mathematical approach for finding the best fitting curve to data by minimizing the sum of squares of the residuals (deviations) from the curve. Although discriminant analysis provides partial and average R^2 values for markers, no model R^2 values are given for comparing models. Maximum likelihood estimation begins with an algorithm known as the likelihood function of the data. The likelihood of a dataset is the probability of obtaining that particular set of data given the selected probability model. This model is composed of the unknown coefficients which in essence are weights for the molecular markers. The variable or marker coefficients that maximize the data likelihood are the maximum likelihood estimates. Maximum likelihood gives a model R^2 value but it is useful only for comparing a model with an intercept against an alternative model without an intercept. Other model fit statistics like the ‘Akaike Information Criterion’ (AIC) are automatically calculated for logistic regression and are more useful for comparing models with different variables. Newer versions of the SPSS® are accommodating more powerful algorithms for improved error rate calculations and are worth trying in future studies. While logistic regression was designed primarily for categorical data, discriminant analysis was designed for multivariate normal quantitative variables but it has been adapted for categorical variables using the nonparametric facility.

6.3.2 β-carotene Content

Investigations by Ameny and Wilson (1997), Hagenimana et al. (1998) and Takahata et al. (1993) revealed negligible levels of β-carotene in non-orange cultivars. The results for the Hunter color value *a* are presented in Table 6.7. An exploratory data analysis showed that

Table 6.3. STEPDISC selection for DNA markers in sweetpotato associated with total sugar content.

Marker entry step	Marker[1]	Partial R-Square[2]	Wilks' Lambda[3]	Pr < Lambda
1	ctg134	0.091	0.909	0.010
2	cta209	0.112	0.807	0.001
3	cta122	0.096	0.730	<.0001
4	cta223	0.072	0.678	<.0001
5	ctt165	0.081	0.623	<.0001
6	ctt272	0.094	0.564	<.0001
7	ctg066	0.099	0.508	<.0001
8	cta211	0.076	0.469	<.0001

[1]Markers are named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.
[2]Partial R-square is the marginal variability accounted for by a variable when all others are already included in the model.
[3]Wilks' Lambda is the likelihood ratio measure of a marker's contribution to the discriminatory power of the model.

Table 6.4. Rate of correct classification of 73 clones of sweetpotato into total sugar group after cross-validation in discriminant analysis.

Number of predictor markers	High total sugar group error rate	Low total sugar group error rate	Total error rate
8	0.194	0.189	0.192
9	0.194	0.162	0.178
12	0.083	0.081	0.082
20	0.056	0.108	0.082
27	0.028	0	0.014
28	0	0	0

Table 6.5. Rate of correct classification of 73 clones of sweetpotato into total sugar group and the AIC model fit statistic for logistic regression

Number of predictor markers	Probability level for marker entry	High sugar group error rate	Low sugar group error rate	Total error rate	Akaike Information Criterion (AIC)
11	≤ 0.030	0.028	0.027	0.027	35.434
12	≤ 0.030	0	0	0	31.011
13	≤ 0.030	0	0	0	28.362
14	0.669	0	0	0	30.162
15	0.776	0	0	0	32.161

Table 6.6. Dsicriminant analysis marker selection statistics for a model that achieved 100% correct classification for the total sugar trait.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	ctg134	0.091	7.110	0.010	0.909	0.010
2	cta209	0.112	8.810	0.004	0.807	0.001
3	cta122	0.096	7.300	0.009	0.730	<.0001
4	cta223	0.072	5.240	0.025	0.678	<.0001
5	ctt165	0.081	5.930	0.018	0.623	<.0001
6	ctt272	0.094	6.850	0.011	0.564	<.0001
7	ctg066	0.099	7.150	0.010	0.508	<.0001
8	cta211	0.076	5.290	0.025	0.469	<.0001
9	ctg232	0.068	4.570	0.036	0.438	<.0001
10	ctt090	0.102	7.050	0.010	0.393	<.0001
11	cag342	0.108	7.380	0.009	0.351	<.0001
12	ctt207	0.130	8.980	0.004	0.305	<.0001
13	ctg126	0.091	5.910	0.018	0.277	<.0001
14	cag250	0.074	4.630	0.036	0.257	<.0001
15	ctt140	0.058	3.510	0.066	0.242	<.0001
16	ctt104	0.070	4.220	0.045	0.225	<.0001
17	cta244	0.080	4.780	0.033	0.207	<.0001
18	cag143	0.084	4.980	0.030	0.189	<.0001
19	cta258	0.077	4.440	0.040	0.175	<.0001
20	ctt096	0.062	3.440	0.069	0.164	<.0001
21	ctt121	0.107	6.090	0.017	0.146	<.0001
22	cta237	0.121	6.850	0.012	0.129	<.0001
23	cta082	0.147	8.450	0.006	0.110	<.0001
24	cag224	0.112	6.070	0.017	0.098	<.0001
25	ctt338	0.082	4.180	0.047	0.090	<.0001
26	cta202	0.082	4.090	0.049	0.082	<.0001
27	cag096	0.089	4.410	0.041	0.075	<.0001
28	cta140	0.105	5.170	0.028	0.067	<.0001

clusters based on *L, a* and *b* color variables were similar to the clusters formed by color variable *a* alone (results not presented for *L* and *b*). Consequently the phenotypic grouping was based on cluster analysis based on color variable *a* with a constraint of 2 clusters being imposed. The first cluster consisted of 51 clones that were classified as having high β-carotene content while cluster two consisted of 22 clones that had low carotene content. The disproportionate numbers in the two clusters, 22 low and 51 high, violated the assumption of equal prior probabilities for the two

clusters for discriminant analysis purposes. That notwithstanding, our results suggest that both logistic regression and discriminant analysis were sufficiently robust to overcome the limitation of unequal and also small sized groups.

Clone 96-9 had the lowest *a* value (-0.385) and this color value is consistent with a white fleshed clone therefore clone 96-9 was not expected to contain any significant quantities of β-carotene (0 mg of β-carotene/100g fresh weight) as presented in Table 6.7. Clone 110-18 had a deep orange color and recorded the highest *a* value (34.847), which was equivalent to 21.4mg β-carotene/100g fresh weight. Low et al. (1997) suggested that sweetpotato cultivars with a minimum of 100 μg retinol equivalent (RE)/ 100g fresh weight may be a good source of β-carotene. One RE is equivalent to 6μg or 0.006mg of β-carotene and consequently any clone with β-carotene content above zero (Table 6.4) may be a good source of pro-vitamin A. However most sweetpotatoes are consumed after processing and post-harvest processing of sweetpotatoes denatures the β-carotene to varying degrees depending on the method of processing (Hagenimana et. al., 1998; K'Osambo et al., 1999). This would suggest that the higher the β-carotene content of a clone the more advantageous it is to the processor and consumer. This is especially so in resource poor communities where orange-fleshed sweetpotatoes have been suggested as an intervention strategy to overcome vitamin A deficiency in the diet (Low et al., 1997). An effective genotype selection and classification protocol that clearly delineates high β-carotene genotypes as suggested in this study would therefore enable breeders to select clones that would ultimately be of great benefit to a breeding program.

Logistic regression classification resulted in two clones being reclassified. Clone 33-9 which was phenotypically considered to have high carotene levels was reclassified as a low carotene level clone based on its probability value (> 0.676). Clone EL-14 that was previously

considered to have low carotene content was reclassified as a high carotene clone. This resulted in a sum of probabilities of the low group being 20.606 and the high group being 1.455. Models were built to select all markers that passed the P=0.03 significance test.

Table 6.7. Hunter color value *a* and phenotypic classification for ß–carotene content of sweetoptato clones from an F_1 polycross of 'Beauregard', 'Excel', 'L94-96', 'L86-33', 'L89-110' and 'L96-117'.

Clone[1]	Mean color *a*	Carotene quantity (mg/100g fresh weight)	Probability of being in low carotene group	Group[2]	Clone	Mean color *a*	Carotene quantity (mg/100g fresh weight)	Probability of being in low carotene group	Group
110-18	34.847	21.428	0	H	BXP	27.392	14.987	0	H
117P	34.446	21.081	0	H	BX-2	25.999	13.783	0	H
117-25	34.291	20.948	0	H	96-11	25.553	13.398	0	H
EL-27	34.193	20.862	0.001	H	BX-22	25.372	13.241	0.001	H
BX-26	33.872	20.586	0	H	33-9	25.341	13.214	0.676	H
EL-20	33.392	20.171	0	H	96-15	24.949	12.876	0	H
BX-13	32.718	19.588	0	H	EL-16	24.659	12.625	0	H
96-18	32.298	19.225	0	H	33-1	24.636	12.605	0	H
110-22	32.297	19.225	0	H	BX-6	24.456	12.450	0	H
EL-4	32.162	19.108	0	H	33-19	23.972	12.032	0	H
110-24	32.063	19.022	0	H	33-23	23.812	11.894	0	H
EL-15	31.962	18.935	0.001	H	110-13	22.298	10.585	0	H
BX-24	31.958	18.932	0	H	96-7	21.158	9.600	0	H
96P	31.717	18.724	0.258	H	EL-8	19.810	8.436	0	H
110-29	31.669	18.682	0	H	117-8	17.404	6.357	1	L
117-2	31.048	18.146	0	H	BX-16	15.251	4.496	0.676	L
96-8	31.008	18.111	0	H	110-21	14.780	4.090	0.998	L
BX-29	30.913	18.029	0	H	EL-11	14.621	3.952	1	L
117-28	30.710	17.853	0	H	96-2	10.297	0.217	1	L
BX-5	30.533	17.701	0.258	H	33-16	9.357	0	1	L
BX-8	30.439	17.619	0	H	110-23	8.274	0	1	L
EL-7	29.990	17.231	0	H	BX-20	8.038	0	1	L
BX-18	29.658	16.945	0	H	33-11	7.297	0	1	L
110-25	29.444	16.760	0.001	H	BX-1	5.606	0	1	L
33-18	29.302	16.637	0	H	33-30	5.537	0	1	L
110-1	29.287	16.624	0	H	BX-23	5.111	0	1	L
110-26	29.181	16.532	0	H	BX-17	4.210	0	1	L
96-5	29.117	16.477	0	H	BX-25	3.736	0	0.999	L
110-4	28.798	16.201	0.001	H	96-10	3.102	0	1	L
117-3	28.645	16.069	0.258	H	BX-15	2.737	0	1	L
EL-24	28.439	15.891	0	H	EL-14	1.539	0	0.258	L
117-6	28.307	15.777	0	H	EL-3	1.503	0	1	L
110-14	28.215	15.697	0	H	EL-5	1.261	0	0.676	L
96-27	28.177	15.665	0	H	110-30	0.070	0	1	L
96-17	28.157	15.647	0	H	110-15	-0.240	0	1	L
96-20	28.137	15.631	0	H	96-9	-0.385	0	0.999	L
33-7	27.982	15.496	0	H					

[1]Clones followed by the letter 'p' are parental clones.
[2]H – High carotene clones and L – Low carotene clones as determined by the spectrophotometer.

Table 6.8. Logistic regression selection for DNA markers associated with β-carotene content in sweetpotato.

Marker entry step	Marker[1]	Estimate (β)	Chi-square score[2]	Pr > ChiSq
0	Intercept	-95.758		
1	ctt240	102.800	6.635	0.010
2	ctt347	69.913	6.317	0.012
3	cag224	-56.940	7.802	0.005
4	cta067	62.667	6.769	0.009
5	ctg058	-64.939	8.041	0.005
6	ctg186	41.461	11.105	0.001
7	cta155	-26.830	8.958	0.003
8	ctt229	-22.377	10.107	0.002
9	ctg213	14.795	6.481	0.011
10	cag170	13.003	5.926	0.015

[1]Markers are named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.
[2]χ^2 score is the largest significant score for maker not in model to be included in the model.

The logistic model for the β-carotene phenotype, though effective, was not as efficient as the logistic model for total sugar content. Logistic regression selected 10 molecular markers linked to β-carotene content (Table 6.8). Logistic regression models with varying numbers of markers are presented in Table 6.9. Table 6.10 presents classification results from an improved logistic regression model having 12 markers. This was the best model with the lowest AIC value (Table 6.9).

A predictive model for genotype classification purposes using discriminant analysis (P=0.03) consisted of 16 markers (Table 6.11). When tested by cross-validation the discriminant model achieved a population prediction accuracy of 87.97%. Three clones (110-25, 117-25 and EL-6) that were previously considered to have high carotene content were reclassified into the low ß-carotene group. Four other clones (96-9, 110-21, 110-30 and BX-25) that were classified as belonging to the low carotene class were reclassified into the high carotene group. Clone 96-9 had the lowest carotene content and was not expected to be reclassified. This suggests a need for

more tests with more populations before a reliable model can be built using either logistic regression or discriminant analysis. When logistic regression models that achieved 100% correct classification for β-carotene content (Table 6.9) were compared with the most efficient discriminant analysis models (Table 6.12), logistic regression models were found to have fewer markers. Table 6.13 presents marker selection statistics used by discriminant analysis.

Table 6.9. Rate of correct classification of 73 clones of sweetpotato into β-carotene group and the AIC model fit statistic for logistic regression.

Number of predictor markers	Probability level for marker entry	High carotene group error rate	Low carotene group error rate	Total error rate	Akaike Information Criterion (AIC)
10	≤ 0.030	0.020	0.045	0.033	30.345
11	≤ 0.050	0.020	0	0.010	26.795
12	0.156	0	0	0	26.560
13	0.595	0	0	0	28.482
16	0.623	0	0	0	34.471

AMOVA results from testing the genotypic variability between the low and high groups showed that the total sugar groups were better defined than the carotene groups. Results from the total sugar analysis revealed that the 13 markers selected by logistic regression and the 8 selected by discriminant analysis (Tables 6.2 and 6.3) showed significant genotypic differences between the low and high total sugar groups ($P<0.001$). After analyzing the β-carotene data, the 10 markers selected by logistic regression (Table 6.8) showed significant genotypic differences between high and low carotene groups ($P=0.006$) while the 16 markers selected by discriminant analysis (Table 6.11) resulted in a less stringent significant difference ($P=0.032$).

My results indicate the possibility of using molecular markers to predict phenotypic grouping of new genotypes of unknown nutrient composition. We found promising results through the use of the discriminant analysis and logistic regression techniques for quick clone

selection during the early stages of sweetpotato breeding. The AIC values for both total sugar and β-carotene models decrease with increasing model efficiency (Tables 6.5 and 6.9). The model with the lowest AIC value achieves 100% correct group and population classification. The AIC value increases as the number of markers in the models increases beyond the number in the model with the lowest AIC value. Although such models also achieve 100% correct classification the model with the lowest AIC value is the most desirable because it has the fewest markers. Although logistic regression seems to produce models that are easily visualized, with the contribution of each marker being expressed by its coefficient, nonparametric discriminant analysis is just as effective when phenotypic classes are clearly defined. More effective models are likely to be developed when a larger array of markers from many varied populations are used. Pohar et al. (2004) used simulation studies in an attempt to compare the performance of logistic regression and linear discriminant analysis. These workers also outlined different model comparison criteria other than error rates. They concluded that results for linear discriminant analysis and logistic regression are close when the assumption of multivariate normality is not too badly violated. They further concluded that use of linear discriminant analysis when explanatory variables are not normally distributed is theoretically wrong. My study using molecular markers which have a bimodal distribution violates this assumption without any recourse to remedy the violation through data transformation.

My studies involved marker variables that differed between discriminant analysis and logistic regression. The functional basis for discriminant analysis and logistic regression may be similar as far as classification is concerned but the two methods have structural differences that can not be ignored if they are to be compared. In contrast to the present study both methods as investigated by Pohar et al. (2004) used identical explanatory variables that were quantitative and

that were generated through simulations. From a plant breeding perspective what may be needed for effective marker selection is (1) classification strategies that assist in describing phenotypic classes that have a meaningful physical interpretation and that produce optimum prediction models and (2) efficient and simple predictive models that are applicable across a wide range of populations.

Table 6.10. Probability levels and classification of sweetpotato clones into β-carotene groups using 12 markers in logistic regression.

Clone[1]	Probability of being in low carotene group	Phenotypic group for β-carotene[2]	Clone	Probability of being in low carotene group	Phenotypic group for β-carotene
33-1	0	H	177-28	0	H
110-1	0	H	BX-22	0.009	H
177-3	0.023	H	BX-29	0.006	H
96P	0.023	H	EL-20	0	H
33-7	0	H	BX-24	0	H
96-5	0	H	EL-24	0	H
110-4	0.001	H	EL-15	0.012	H
96-7	0	H	33-18	0	H
177-6	0	H	BX-26	0	H
33-9	0.005	H	EL-16	0	H
96-8	0	H	EL-27	0.009	H
96-11	0	H	33-19	0.002	H
BX-8	0	H	33-23	0	H
EL-4	0	H	96-15	0	H
EL-8	0.006	H	96-2	1	L
BX-2	0.006	H	96-9	0.985	L
BX-5	0.023	H	177-8	1	L
BX-6	0	H	33-11	1	L
EL-7	0	H	96-10	1	L
177P	0.006	H	BX-1	1	L
BXP	0	H	EL-5	0.961	L
110-24	0	H	BX-15	1	L
96-17	0	H	EL-11	0.998	L
110-13	0	H	BX-16	1	L
110-25	0.001	H	EL-14	0.965	L
96-18	0.001	H	BX-17	1	L
BX-13	0	H	110-15	1	L
110-14	0	H	110-30	1	L
110-26	0.001	H	110-21	0.977	L
96-20	0.019	H	BX-20	1	L
110-29	0	H	110-23	1	L
96-27	0	H	BX-23	1	L
BX-18	0	H	EL-3	1	L
110-18	0.001	H	33-16	1	L
177-2	0	H	BX-25	0.988	L
110-22	0	H	33-30	1	L
177-25	0	H			

[1]Clones followed by the letter 'p' are parental clones.
[2]H – High carotene clones and L – Low carotene clones as determined by the spectrophotometer.

Table 6.11. STEPDISC selection for DNA markers associated with β-carotene content in sweetpotato.

Marker entry step	Marker[1]	Partial R-Square[2]	Wilks' Lambda[3]	Pr < Lambda
1	ctt240	0.091	0.909	0.010
2	ctt347	0.079	0.837	0.002
3	cag224	0.108	0.747	0.000
4	ctg213	0.090	0.679	<.0001
5	cta067	0.074	0.629	<.0001
6	ctt092	0.088	0.573	<.0001
7	cag170	0.094	0.520	<.0001
8	cta258	0.093	0.471	<.0001
9	cta221	0.127	0.412	<.0001
10	ctt106	0.083	0.377	<.0001
11	ctg086	0.088	0.344	<.0001
12	ctg117	0.097	0.311	<.0001
13	cta126	0.089	0.283	<.0001
14	cag099	0.148	0.241	<.0001
15	ctg105	0.118	0.213	<.0001
16	ctg137	0.117	0.188	<.0001

[1]Markers are named starting with the three letters coding for the primer followed by the molecular weight of the marker in base pairs.
[2]Partial R-square is the marginal variability accounted for by a variable when all others are already included in the model.
[3]Wilks' Lambda is the likelihood ratio measure of a marker's contribution to the discriminatory power of the model.

Table 6.12. Rate of correct classification of 73 clones of sweetpotato into β-carotene group after cross-validation in discriminant analysis.

Number of predictor markers	High carotene group error rate	Low carotene group error rate	Total error rate
16	0.059	0.182	0.121
17	0.059	0.182	0.121
19	0.039	0.136	0.088
20	0.020	0.136	0.078
21	0.020	0.136	0.078
22	0	0	0

Table 6.13. Discriminant analysis marker selection statistics for a model that achieved 100% correct classification for the β-carotene trait in the population.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	ctt240	0.091	7.100	0.010	0.909	0.010
2	ctt347	0.079	6.040	0.017	0.837	0.002
3	cag224	0.108	8.340	0.005	0.747	0.000
4	ctg213	0.090	6.760	0.012	0.679	<.0001
5	cta067	0.074	5.350	0.024	0.629	<.0001
6	ctt092	0.088	6.390	0.014	0.573	<.0001
7	cag170	0.094	6.740	0.012	0.520	<.0001
8	cta258	0.093	6.540	0.013	0.471	<.0001
9	cta221	0.127	9.180	0.004	0.412	<.0001
10	ctt106	0.083	5.600	0.021	0.377	<.0001
11	ctg086	0.088	5.900	0.018	0.344	<.0001
12	ctg177	0.097	6.450	0.014	0.311	<.0001
13	cta126	0.089	5.750	0.020	0.283	<.0001
14	cag099	0.148	10.060	0.002	0.241	<.0001
15	ctg105	0.118	7.660	0.008	0.213	<.0001
16	ctg137	0.117	7.420	0.009	0.188	<.0001
17	ctg066	0.079	4.680	0.035	0.173	<.0001
18	ctt272	0.101	6.030	0.017	0.156	<.0001
19	ctg274	0.103	6.070	0.017	0.140	<.0001
20	cta178	0.111	6.470	0.014	0.124	<.0001
21	cag250	0.126	7.340	0.009	0.109	<.0001
22	cag301	0.207	13.020	0.001	0.086	<.0001

6.4 Literature Cited

Ameny, M.A., and P.W. Wilson. 1997. Relationship between Hunter color and ß – carotene content in white-fleshed African sweetpotato. J. Sci. Food Agric. 73:301-306.

Boudreaux, J.E. 1994. Commercial vegetable production recommendations. La. Coop. Ext. Serv. Publ. 2433.

Bushway, R. J. 1986. Determination of α and ß - carotenoids in fruits and vegetables by high performance liquid chromatography. J. Agri. Food Chem. 34:409-412.

Excoiffer, L., P.E. Smouse, and J.M. Quattiro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131:479-491.

Hagenimana, V., L.M. K'Osambo, and E.E. Carey. 1998. Potential of sweetpotato in reducing vitamin A deficiency in Africa. CIP Program Report, Lima, Peru. p 287-294.

Hall, M. R and S.C Phatak. 1993. Sweetpotato [*Ipomoea batatas* (L.) Lam.], p 693-708 In: Genetic Improvement of vegetable crops. Kalloo G and B.O Bergh (eds). Pergamon Press.

Hunter, P.S. 1958. Photoelectric color difference meter. Journal Optical Society of America 48:485-995.

Jones, A. 1965. Cytological observations and fertility measurements of sweetpotato [*Ipomoea batatas* (L.) Lam.], Proc. Am. Soc. Hort. Sci. 86:527-537.

Jones, A. 1986. Sweetpotato heritability estimates and their use in breeding. HortScience 21:14-17

Katayama, K., K. Komaki, and S. Tamiya. 1996. Prediction of starch, moisture, and sugar in sweetpotato by near infrared transmittance. HortScience 31:1003-1006.

K'osambo, L.M., E.E. Carey, A.K. Misra, J. Wilkes, and V. Hagenimana. 1998. Influence of age, farming site and boiling on pro-vitamin A content in sweetpotato [*Ipomoea batatas* (L.) Lam.] storage roots. J. Food Composition and Analysis 11: 305-321.

Labonte, D.R., D.H. Picha, and H.A. Johnson. 2000. Carbohydrate-related changes in sweetpotato storage roots during development. J. Amer. Soc. Hort. Sci. 125:200-204.

Low, J., P. Kinyae, S. Gichuki, M.A. Oyunga, V. Hagenimana, and J. Kabira. 1997. Combating vitamin A deficiency through the use of sweetpotato. Results from phase I of an action research project in South Nyanza, Kenya. International Potato Center, Lima, Peru. 110p.

Lu, G.Q. and J.L. Sheng. 1990. Application of near infrared reflectance spectroscopy (NIRS) in sweetpotato quality breeding. Scientia Agricultura Sinica 23:76-81.

Mcharo, M., D.R. Labonte, J.H. Oard, S.J. Kays, and W.J. McLaurin. 2004. Linking quantitative traits with AFLP markers in sweetpotatoes using discriminant analysis. Acta Hort. 637:285-293.

Mcharo, M., E. Bush, D. La Bonte, C. Broussard, and L. Urbatsch. 2003. Molecular and Morphological Investigation of Ornamental Liriopogons J. Amer. Soc. Hort. Sci. 128: 575-577.

Michelmore, R.W., I. Paran, and R.V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulk segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. USA 88:9829-9832.

Picha, D.H. 1985. HPLC determination of sugars in raw and baked sweetpotatoes. J. Food Sci. 50:1189-1190.

Picha, D.H. 1987. Carbohydrate changes in sweetpotatoes during curing and storage. J. Amer. Soc. Hort. Sci. 112:89-92.

Pohar, M., B. Mateja and S. Turk. 2004. Comparison of logistic regression and linear discriminant analysis: A simulation study. Metodološki zvezki, 1:143-161.

Rubatzky, V.E and M. Yamaguchi, M. 1997. Sweetpotato, p 130-146. In: World vegetables, principles, production and nutritive values. AVI. Van Nostrand Reinhold Company, New York.

SAS Institute Inc. 1999. SAS/STAT User's guide. SAS Institute Inc. Cary, North Carolina.

SAS Institute Inc. 2001. The SAS system for windows v8. SAS Institute Inc. Cary, North Carolina.

Simonne, A.H., S.J. Kays, P.E. Koehler and R.R. Eitenmiller. 1993. Assessment of ß-carotene content in sweetpotato breeding lines in relation to dietary requirements. Journal of food composition and analysis 6:336-345.

Takahata, Y., T. Noda, and T. Nagata. 1993. HPLC determination of ß – carotene of sweetpotato cultivars and its relationship with color values. Japanese J. Breed 43: 421-427.

Ukoskit, K., P.G. Thompson, C.E. Watson Jr., and G.W. Lawrence. 1997. Identifying a randomly amplified polymorphic DNA (RAPD) marker linked to a gene for root-knot nematode resistance in sweetpotato. J. Amer. Soc. Hort. Sci. 122:818-821.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Wang, Y. and S.J. Kays. 2003. Analytically directed flavor selection in breeding food crops. J. Amer. Soc. Hort. Sci. 128: 711-720.

West, C.E. 1994. Vitamin A nutrition. Handout for course in Food Science and Nutrition. Wageningen Agricultural University, Wageningen, Holland. p 13.

Woolfe, A.J. 1992. Sweetpotato: An untapped food resource. Cambridge University Press, Cambridge, U.K. 643p.

CHAPTER 7: SUMMARY AND CONCLUSIONS

The research efforts and results in this study covered a wide range of traits of commercial importance in liriopogon ornamentals and sweetpotatoes. In the liriopogon ornamental studies I found that molecular marker data distinguished the cultivars from one another and showed that no duplicates existed in the samples of popular liriopogon cultivars investigated. Molecular marker data did not substantiate the existence of two genera, Liriope and Ophiopogon These results are consistent and extend results by other workers in previous studies. In liriopogons floral morphology has great aesthetic value in the floriculture industry and has consequently been used to classify putative cultivars in the past. Since morphological characterization has been shown to be inadequate in characterizing these ornamental cultivars I suggest that there is a significant role that molecular characterization can play in certifying cultivars as genetically distinct for commercial purposes. It is therefore recommended that molecular marker studies be used more extensively for taxonomic studies to clarify uncertainty in classification and subsequent naming of cultivars.

Another objective of the liriopogon study was to use it as a training platform to gain skills in the techniques that I would later use in the larger sweetpotato project. The skills included sample tissue preservation for future DNA extraction, DNA extraction, DNA amplification and electrophoresis, molecular marker scoring and data analysis. The utility of clustering, discriminant analysis and logistic regression as data analysis techniques in associating molecular markers with genotypes was investigated with a primary objective of either confirming or discounting prevailing phenotype based classification of genotypes.

The sweetpotato aspect of my study covered some of the major traits that affect field production and post-harvest consumption. The first project on clone classification based on root

dry matter content provided an initial opportunity to test discriminant analysis on clearly defined groups for a quantitative trait. The dry matter data and raw DNA were obtained from the USDA repository hence my laboratory analysis was limited to gel electrophoresis and data analysis. The dry matter study also provided an opportunity to use a population of unrelated clones which can not be analyzed using traditional QTL techniques. Success in using discriminant analysis in the first project provided the impetus to further test this analytical tool and compare it with traditional QTL analysis for selection of AFLP and RAPD markers associated with virus resistance in the second project. Logistic regression was also introduced at this stage as an alternative multivariate marker selection tool. Both multivariate techniques were comparable to the univariate QTL analysis method because they also selected the most informative markers that were selected by QTL analysis.

Validation of the usefulness of discriminant analysis and logistic regression in associating molecular markers with virus resistance suggested possible use of these methods on other equally important traits like resistance to root-knot nematode. The study involving root-knot nematode resistance also provided an opportunity to investigate the effectiveness of these multivariate techniques on analyzing unequal phenotypic groups and small sized populations. Although the methods have different optimizing techniques, as described below, they selected a number of identical markers suggesting that the markers selected had a strong association with the resistance trait. The study also provided a chance to investigate the possible effect of different genetic origins of the two populations on markers selected for the same trait. It was observed that no common markers were selected for the two populations using the two modeling techniques.

The efficiency of discriminant analysis and logistic regression were then compared using sugar content and β-carotene content as post-harvest response traits. This study found that logistic regression was more efficient than discriminant analysis in classifying clones into phenotypic groups. Discriminant analysis automatically provides error rates as part of the results output and this is an advantage over logistic regression which does not provide error rates. The error rate decreases as the classification model constructed by discriminant analysis improves. In logistic regression the AIC decreases as the model improves until an optimum model is obtained. When more variables or markers are added onto the model beyond the optimum point, the AIC increases suggesting a poorer model. Calculating error rates in logistic is a tedious and complicated process because for every marker added or removed from the model, the variable coefficients change and the whole calculation matrix in the spreadsheet has to be changed. Newer versions of software like SPSS® do give summary results for error rates in a contingency table and are making calculations of error rates easier. Consequently I suggest use of the AIC for model selection in logistic regression as an alternative to the error rate criterion.

Discriminant analysis and logistic regression models for genotype classification and unknown genotype prediction were constructed. The variables used for model construction were binary in nature and hence nonparametric discriminant analysis had to be used for classification. The drawback to nonparametric analysis is that, unlike parametric analysis that produces an interpretable model, the software could not produce a model that could be visualized and interpreted. This is a challenge if a breeder intends to classify a genotype of an unknown class. Currently the option available within the SAS® software is to include the unknown genotype as part of the test population but leave the class cell empty. The software would then go ahead to assign a class to the new genotype. Conversely multiple logistic regression produces models,

though complex can be interpreted by the breeder. i.e the markers are weighted with coefficients that have a biological interpretation. In situations where more than two phenotypic classes exist, discriminant analysis or logistic regression for a polychotomous response may still be used for modeling.

There is need to use well defined biological classes that have clear physical interpretation for modeling. Such classes may show clear differences for example in sweetpotato flesh color or distinct variations in nutrient content. The advantage of such clear classification is that the models that are constructed are likely to have low error rates during classification. A discriminant analysis model constructed with such populations is also likely to classify an unknown genotype with greater precision because the nearest neighbor distance used for grouping is less ambiguous compared to using populations with fuzzy descriptive boundaries. Multivariate statistical procedures combined with molecular marker information have shown that they are just as good as traditional QTL mapping or morphological classification as tools for genotype identification. However it was also noted that populations with different distributions may produce different models that may not have utility beyond the population from which they were created.

Studies for all the response traits had dichotomous explanatory variables that were molecular markers scored either as present or absent. By their very nature dichotomous variables will always have bimodal distribution regardless of the sample size. The response variables were also dichotomous. Such binary definition of the response variable may not be adequate since the traits that were dealt with, namely dry matter, disease resistance and nutrient content, are inherently quantitative. Individuals in large populations tend to be normally distributed for these traits. The binary nature of both explanatory and response variables is a limitation for methods

like discriminant analysis that have a basic assumption of multivariate normality. However logistic regression is sufficiently robust to handle dichotomous variables. When the explanatory variables have more than two categories the assumption of normality may be satisfied with increasing sample size. Multiple categories are easily handled by either discriminant analysis or multiple logistic regression.

The issue of using either a test sample or cross-validation as methods for determining model efficiency is pertinent for developing good prediction models. For small datasets like those used in this study cross-validation is a better approach due to lower variability in the results. Use of training and test samples is more appropriate when dealing with large sample sizes. In the virus resistance, root knot nematode resistance and the sugar and β-carotene data analyses, all the genotypes were used to create training samples for marker selection by both discriminant analysis and logistic regression. Thereafter the same genotypes were used during cross-validation to test the efficiency of models obtained. It is possible that if the selected markers were used on a test population whose genotypes were not part of the training sample, rates of correct classification may have been different and consequently the conclusions made would be different. The implications of such a result are significant because they would signify uncertainty in obtaining identical markers from different populations. A natural next step would then be to focus on larger populations that could be subdivided into training and test samples and also use of unrelated populations as training and test samples in the same study.

To my knowledge there is no literature on prediction and classification of plant genotypes using logistic regression as a method and molecular markers as explanatory variables. Software that give detailed prediction and classification results for logistic regression, other than summarized contingency tables, are also not readily available. Consequently computer

programming skills among breeders will become increasingly important in order to create custom made programs that may not have a wide client base but are pertinent in the area of marker assisted selection. In conclusion trait linked marker association is still a nascent field in plant population studies and as cheaper and more reliable marker generation techniques become available plant breeders will have greater leeway in exploiting the power of statistical techniques available in their selection programs. These techniques have been used with great success in fields of human and animal genetics and I recommend that they be incorporated into marker assisted selection plant breeding programs.

The following are specific recommendations for future research based on results obtained from this study.

1. Future classification studies need to investigate use of graphics to display results. Visual displays provide a much clearer picture from which conclusions can be made. Discriminant analysis does provide an option for graphic displays.
2. The greatest benefit from these techniques may be derived at early stages of the selection process hence they need to be incorporated early as part of a marker assisted selection approach. This is because genotypes with extreme values of traits of interest are easily classified and hence the breeder can discard most of the unnecessary average genotypes.
3. Multivariate techniques will especially be useful in dealing with populations of unrelated clones, e.g. selecting from landrace genotypes, and therefore further research also needs to be geared to test the utility of these techniques for important traits in such heterogeneous populations. QTL analysis is unsuited for such populations.

4. Breeders are encouraged to use large sized populations as much as possible in search of informative markers. Large populations with a wide genetic base are likely to provide a greater array of markers from which to select the most informative ones.
5. Generation of reliable molecular markers is a challenge that will hopefully be overcome as gel electrophoresis technology improves. While I used an AFLP marker system for my studies, it is disadvantageous in that:
    i. The markers generated were dominant thus providing less information compared to co-dominant markers.
    ii. The automatic marker scoring system was not efficient thus requiring a lot of extra work in visual scoring. Possible use of micro-arrays and newer technologies like capillary electrophoresis may provide more reliable results in future studies.

APPENDIX 1: ADDITIONAL DATA TABLES

Discriminant analysis marker selection statistics for a model that achieved 100% correct classification for the dry matter trait in the USDA population.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	cta084	0.160	10.670	0.002	0.840	0.002
2	cag185	0.184	12.370	0.001	0.686	<.0001
3	cag235	0.263	19.250	<.0001	0.506	<.0001
4	cag148	0.160	10.110	0.003	0.425	<.0001
5	cta212	0.154	9.490	0.003	0.359	<.0001
6	ctt183	0.212	13.730	0.001	0.283	<.0001
7	cta265	0.202	12.660	0.001	0.226	<.0001
8	cag271	0.248	16.110	0.000	0.170	<.0001
9	ctt241	0.353	26.180	<.0001	0.110	<.0001
10	cta076	0.230	14.010	0.001	0.085	<.0001
11	cta235	0.184	10.400	0.002	0.069	<.0001
12	cta254	0.196	10.950	0.002	0.056	<.0001
13	ctg251	0.171	9.050	0.004	0.046	<.0001
14	cag273	0.128	6.330	0.016	0.040	<.0001

Rate of correct classification of sweetpotato clones into SPCSV and SPFMV resistance groups after cross-validation in discriminant analysis.

Disease	Number of predictor markers	Resistant group error rate	Susceptible group error rate	Total error rate
SPCSV	6	0.071	0.089	0.080
	10	0.048	0.067	0.058
	15	0.024	0.067	0.046
	20	0.024	0	0.012
	21	0	0	0
SPFMV				
	4	0.080	0	0.040
	15	0.020	0.054	0.037
	20	0.02	0.024	0.037
	25	0	0.027	0.014
	26	0	0	0

Discriminant analysis marker selection statistics for a model that achieved 100% correct classification for the SPCSV trait.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	e41m33.a	0.672	174.350	<.0001	0.328	<.0001
2	e38m36.u	0.129	12.380	0.001	0.286	<.0001
3	K2.650	0.078	7.000	0.010	0.263	<.0001
4	AL3.1300	0.078	6.930	0.010	0.243	<.0001
5	e44m41.j	0.058	4.950	0.029	0.229	<.0001
6	e39m41.f	0.058	4.900	0.030	0.216	<.0001
7	e39m33.e	0.101	8.850	0.004	0.194	<.0001
8	e38m36.m	0.063	5.250	0.025	0.182	<.0001
9	e39m39.e	0.076	6.330	0.014	0.168	<.0001
10	e39m39.a	0.056	4.530	0.037	0.158	<.0001
11	e40m39.a	0.051	4.050	0.048	0.150	<.0001
12	B13.370	0.045	3.460	0.067	0.144	<.0001
13	e39m44.a	0.052	4.040	0.048	0.136	<.0001
14	e32m62.b	0.067	5.150	0.026	0.127	<.0001
15	e40m41.a	0.054	4.030	0.049	0.120	<.0001
16	e41m45.c	0.056	4.120	0.046	0.114	<.0001
17	e39m40.e(c)	0.069	5.100	0.027	0.106	<.0001
18	R9.650	0.144	11.440	0.001	0.090	<.0001
19	Z15.550	0.120	9.170	0.004	0.080	<.0001
20	e39m34.g	0.101	7.420	0.008	0.072	<.0001
21	e36m49.c	0.122	9.020	0.004	0.063	<.0001

Discriminant analysis marker selection statistics for a model that achieved 100% correct classification for the SPFMV trait.

Marker entry step	Marker	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	S13.1130	0.660	164.680	<.0001	0.340	<.0001
2	e41m37.a	0.089	8.190	0.005	0.310	<.0001
3	e40m36.d	0.074	6.590	0.012	0.287	<.0001
4	e44m36.d	0.076	6.720	0.011	0.266	<.0001
5	e39m45.L	0.064	5.570	0.021	0.249	<.0001
6	e32m62.g	0.067	5.710	0.019	0.232	<.0001
7	G2.1700	0.087	7.520	0.008	0.212	<.0001
8	e41m36.b	0.072	6.090	0.016	0.196	<.0001
9	e40m44.d	0.063	5.220	0.025	0.184	<.0001
10	e40m46.d	0.054	4.370	0.040	0.174	<.0001
11	e40m33.c	0.065	5.200	0.026	0.163	<.0001
12	e33m48.b	0.062	4.860	0.031	0.153	<.0001
13	e39m39.c	0.051	3.910	0.052	0.145	<.0001
14	e32m37.a	0.066	5.110	0.027	0.135	<.0001
15	e35m62.a	0.060	4.490	0.038	0.127	<.0001
16	e33m61.c	0.076	5.790	0.019	0.118	<.0001
17	e40m41.a	0.066	4.850	0.031	0.110	<.0001
18	e41m62.b	0.056	4.030	0.049	0.104	<.0001
19	e36m37.d	0.075	5.430	0.023	0.096	<.0001
20	e35m62.b	0.066	4.670	0.034	0.090	<.0001
21	e35m62.e	0.069	4.790	0.032	0.083	<.0001
22	e40m38.b	0.085	5.910	0.018	0.076	<.0001
23	e32m60.h	0.081	5.540	0.022	0.070	<.0001
24	e41m33.a	0.061	4.060	0.048	0.066	<.0001
25	e39m33.f	0.070	4.610	0.036	0.061	<.0001
26	e41m45.c	0.083	5.460	0.023	0.056	<.0001

APPENDIX 2: DATA ANALYSIS PROTOCOLS

Data set 1: Data with 256 or less columns;

```
dm'log;clear;output;clear';
data LSUsugar;
input clone $   cag342............ctt066  ctt062  group;
cards;
33-1    0……..1         0         1
96-2    1……..1         1         1
96-15   0……..0         1         0
run;
```

Data set 2: Data with more than 256 columns;

```
dm'log;clear;output;clear';
data LSUsugar;
input clone $   cag342 cag329 cag323...........
ctt059  ctt056 ctt066   ctt062  group;
cards;
33-1    0         1         0………
1       0         0         1         1
96-2    1         1         1………
0       0         0         0         1
96-15   0         0         1………
1       1         1         0         0
run;
```

```
Title1 'sugar level in sweetpotatoes';
Title2 'Stepwise discriminant marker selection';
proc stepdisc data=LSUsugar method=forward slentry=0.03;
class group;
var  cag342............ctt066       ctt062;
run;

Title2 'Discriminant analysis on selected markers';
proc discrim data=LSUsugar testdata=LSUsugar method=npar k=1 crosslist crossvalidate
testlist;
testid clone;
class group;
var   ……selected markers………;
run;

Title2 'Logistic regression for marker selection and model testing';
proc logistic data=LSUsugar descending outest=LSUsugarone;
class clone;
model group= cag342............ctt066   ctt062 / selection=forward slentry=0.03 lackfit
clparm=wald;
output out=LSUsugartwo predprobs=(individual crossvalidate);
run;
quit;
```

Notes:

1. The above codes are a summary of the information that is readily available in "SAS Help" or in SAS manuals, consequently the appendix will point out critical codes that may need to be included in the SAS statements.

2. Data set 2 consists of markers that are too numerous to fit onto a standard SAS data sheet. The input line and the corresponding marker lines are therefore edited and truncated in a spreadsheet e.g MICROSOFT EXCEL so that they can fit the data sheet.
3. The length of the input line, proc stepdisc line and proc logistic line may not exceed the length of the marker data lines otherwise the SAS program is likely to hang and fail to compute.
4. In the 'proc discrim' and 'proc logistic' statements the selection method may be changed from "selection=forward" to selection=backward in which case "slentry=p-value" is changed to "slstay=p-value". The option "stop=n" may also be used instead of slentry or slstay in order to instruct the selection procedure to select a maximum of n markers. The selection method may also be changed to "selection=stepwise" to select markers without imposing a preset p-value.
5. The "method=npar" option instructs SAS to use the nonparametric procedure for classification purposes. The "k=1" option instructs SAS to use 1 nearest neighbor for classification purposes. The "crossvalidate" option instructs SAS to conduct a cross-validation procedure using the markers selected.
6. The group value of 1 or 0 is the phenotypic group coding. The coding may also be 1 or 2 or in case of n groups where n is more than two groups, values may have the range 1,2,3….n.
7. The selected markers in 'proc discrim' will be found in the output of 'proc stepdisc'.
8. An analysis of molecular variance (AMOVA) may be conducted on the original array of molecular markers or on the selected array of markers. I found it more beneficial and more informative to conduct an AMOVA on selected markers because the selected markers provide the information that I'm looking for.

APPENDIX 3: PERMISSION LETTERS FROM JOURNAL PUBLISHERS

From: "Don La Bonte" <dlabonte@agctr.lsu.edu>
To: "Mwamburi Mcharo" <mwamburim@yahoo.com>, ebush@agctr.lsu.edu
Subject:Fw: ASHS
Date: Thu, 14 Oct 2004 18:33:24 -0500

----- Original Message -----
From: Joan Herto
To: Don La Bonte
Sent: Thursday, October 14, 2004 11:05 AM
Subject: Re: ASHS

The article belongs to you even though the entire publication belongs to ASHS so you can use it in his dissertation.
Joan Herto

On Oct 14, 2004, at 11:50 AM, Don La Bonte wrote:

Could you send me a pdf file of a paper I published in 2003? The senior author (my student) needs this for a part of his dissertation. We also need permission to include this in his dissertation. Thanks, Don La Bonte 225-578-1024

JASHS 128(4):575-577. "Molecular and Morphological Investigation of Ornamental Liriopogons".

<Don La Bonte.vcf>

From: "Peter Vanderborght" <peter.vanderborght@ishs.org>
To: "'Mwamburi Mcharo'" <mwamburim@yahoo.com>
CC: dlabonte@agctr.lsu.edu
Subject:RE: Permission to use published paper
Date: Thu, 21 Oct 2004 11:43:18 +0200

Dear Mwamburi Mcharo,

Thank you for contacting the ISHS. Please consider this message as our formal permission to use the below stated article, strictly for non-commercial purposes, as part of your dissertation, provided a clear and full reference to the original article is included. We thank you for your interest in Acta Horticulturae(r) and hope the above answers your question. Good luck with your dissertation !

Should you have any further question, please do not hesitate to contact us again.

Sincerely yours,

Peter Vanderborght

International Society for Horticultural Science
PO Box 500 - 3001 Leuven 1 - Belgium
Phone: +32 16229427 Fax: +32 16229450
info@ishs.org

Visit our website www.ishs.org or www.actahort.org

The ISHS, originated in 1864, formally established in 1959 and with members from over 128 countries, is the world's leading - independent - organization of horticultural scientists.

The contents of this email and any files transmitted with it are confidential, proprietary and may be legally privileged. They are intended solely for the use of the individual or entity to whom they are addressed. If you have received this email in error please notify the sender. If you are not the intended recipient you may not use, disclose, distribute, copy, print or rely on this email. The sender is not responsible for any changes made to any part of this email after transmission. Any views or opinions presented are solely those of the author and do not necessarily represent those of the Society. Although this email and any attachments are believed to be free from any virus or other defects which might affect any computer or IT system into which they are received, no responsibility is accepted by the Society for any loss or damage arising in any way from the receipt or use thereof. An interest in horticultural science is all you need to enjoy being a member of the ISHS. For more information please visit http://www.ishs.org

-----Original Message-----
From: Mwamburi Mcharo [mailto:mwamburim@yahoo.com]
Sent: woensdag 20 oktober 2004 17:51
To: Peter Vanderborght
Cc: dlabonte@agctr.lsu.edu
Subject: Permission to use published paper

Dear Dr Vanderborght,
I wish to request for permission to use the publication cited below as part of my dissertation. I am a PhD student in Louisiana State University, USA working under Dr. Don Labonte and I am the senior author of the paper. Thank you very much.
Regards

Mcharo, M., Labonte, D.R., Oard, J.H., Kays, S.J. and McLaurin, W.J. 2004.

LINKING QUANTITATIVE TRAITS WITH AFLP MARKERS IN SWEETPOTATOES USING DISCRIMINANT ANALYSIS. Acta Hort. (ISHS) 637:285-293
http://www.actahort.org/books/637/637_35.htm

From: "Don La Bonte" <dlabonte@agctr.lsu.edu>
To: "Mwamburi Mcharo" <mwamburim@yahoo.com>
Subject:Fw: JASHS-00375R accepted for publication in Journal of ASHS
Date: Mon, 11 Oct 2004 09:37:17 -0500

----- Original Message -----
From: <pubs@ashs.org>
To: <dlabonte@agctr.lsu.edu>
Sent: Saturday, October 09, 2004 2:55 PM
Subject: JASHS-00375R accepted for publication in Journal of ASHS

> 9 Oct. 2004
>
> Dear Dr. LaBonte:
>
> The review of your revised manuscript "Associating molecular markers
with
virus
> resistance to classify sweetpotato genotypes" has been completed. I
am
pleased to
> inform you that this manuscript has been accepted for publication.
>
> Please incorporate the minor changes detailed below in a revised
electronic copy. Most
> of these changes relate to minor deviations from ASHS style. Refer to
the
ASHS
> Publications Style Manual at
<http://www.ashs.org/authors/stylemanual.html> for further
> details.
>
> 1) If you cite more than one reference by the same author or group of
authors, and the
> year is also identical (e.g., Mwanga et al., 2002), insert lowercase
letters (in alphabetical
> order) according to the sequence in which they are cited in the text.
>
> 2) Soller and Beckmann, 1983 was not found cited in the manuscript.
>
> 3) Insert the city of publication in the Literature Cited listing for
Capdevielle et al., 2000.
> You report only one page number; is this an abstract? If so, indicate
this
fact in the
> Literature Cited.
>
> Please e-mail the final version of your paper in Microsoft Word to me
at <
> ndevos@got.net> within 2 weeks of the acceptance date at the top of
this

letter. If you use
> another word processing program, please include an ASCII or rich text
file.
>
> Prior to sending the requested items, please review your paper very
carefully for factual
> errors, incomplete references, and general flow/readability of text.
Please use double-
> spacing throughout the paper including the abstract, Literature
Cited, and
tables.
>
> When your paper nears publication, you will be sent galley proofs.
The
purpose of the
> galley proofs is to allow authors to check for possible typesetting
errors. Major author
> revisions and changes to galley proofs will result in charges to the
author of $50 per
> change, in addition to any applicable page charges and color charges.
>
> Please contact me if you have any questions, or if I can provide
additional assistance.
>
> Thank you.
>
> Sincerely,
>
> Neal De Vos
>
> Neal E. De Vos, PhD, CPH
> Editor in Chief
>
> <ndevos@got.net>

VITA

Thomas Mwamburi Mcharo was born in 1969 in Taita-Taveta District of the Republic of Kenya. After completing his primary education in various parts of Kenya he pursued high school in Nairobi, Kenya. He was accepted in the University of Nairobi in 1988 for a bachelor of science degree program in agriculture and after completing his studies was offered a scholarship to pursue a master's program in plant breeding in the same institution in 1991. After completing his studies he pursued a career in agronomy and crop research for six years before commencing his doctoral studies in horticulture at Louisiana State University in 2001. He will receive the degree of Doctor of Philosophy in horticulture during the Spring Commencement 2005.