Characterizing the Genetic Architecture of Fitness Related Traits in an Annual Grass, *Avena barbata*

by

Kyle Matthew Gardner

Submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
October, 2004

Your file Votre référence
ISBN: 0-494-00951-9
Our file Notre référence
ISBN: 0-494-00951-9

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages
- Examiners Signature Page (pii)
- Dalhousie Library Copyright Agreement (piii)

Appendices
- Copyright Releases (if applicable)

To my parents,
Brian and Paula Gardner

Table of Contents

LIST OF FIGURES ..... **vii**

LIST OF TABLES ..... **viii**

ABSTRACT ..... **x**

ACKNOWLEDGEMENTS ..... **xi**

CHAPTER ONE : **1**

INTRODUCTION 1

Introduction *1*

Biology and Natural History of Avena barbata *4*

Creation of the Recombinant Inbred Line Population *6*

Research Objectives *7*

CHAPTER TWO: **10**

CO-LOCALIZING QTL IN THE ESTIMATION OF GENETIC CORRELATION BETWEEN CONTINUOUS TRAITS 10

Introduction: *10*

Materials and Methods: *15*

Results: *21*

Discussion: *32*

CHAPTER THREE: **37**

HERITABLE VARIATION AND GENETIC CORRELATION AMONG FITNESS RELATED TRAITS IN *A. BARBATA* 37

Introduction: *37*

Materials and Methods: *40*

Results: *45*

Discussion: *58*

CHAPTER FOUR: **67**

CONSTRUCTION OF AN AMPLIFIED FRAGMENT LENGTH POLYMORPHISM GENETIC LINKAGE MAP IN *A. BARBATA* 67

Introduction: *67*

Materials and Methods: *70*

Results: *77*

Discussion: *84*

CHAPTER FIVE: **89**

IDENTIFICATION AND ANALYSIS OF QTL FOR FITNESS RELATED TRAITS 89

Introduction: *89*

Materials and Methods: *92*

Results: *99*

Discussion: *116*

CHAPTER SIX: **125**

GENETIC CORRELATIONS AND QTL ANALYSIS FOR FITNESS IN A RECIPROCAL TRANSPLANT EXPERIMENT 125

Introduction: *125*

Materials and Methods: *128*

Results: *134*

Discussion: *160*

CHAPTER SEVEN: **167**

SUMMARY AND CONCLUSIONS 167

REFERENCES **179**

APPENDIX ONE **189**

SINGLE MARKER ANALYSES 189

APPENDIX TWO **197**

SIGNIFICANCE THRESHOLDS AND LIKELIHOOD RATIO PROFILES FOR ALL GREENHOUSE AND FIELD TRAITS 197

List of Figures

FIGURE 2.1. HISTOGRAMS OF CORRELATION AND QTL DATA FOR SIGNIFICANTLY CORRELATED VERSUS UNCORRELATED TRAITS 23

FIGURE 2.2. SCATTER PLOT OF R_G VERSUS R_Q 27

FIGURE 3.2. BOXPLOTS OF COMPONENT SCORES ALONG THE MAJOR AXES OF VARIATION IN THE GREENHOUSE. 57

FIGURE 4.1. GENETIC LINKAGE MAP OF THE *A. BARBATA* GENOME CONSTRUCTED WITH AFLP MARKERS. 81

FIGURE 5.1. GENETIC LINKAGE MAP WITH POSITIONS OF QTL FOR TRAITS MEASURED IN THE GREENHOUSE. 103

FIGURE 6.1. BOXPLOTS OF COMPONENT SCORES ALONG THE MAJOR AXES OF VARIATION IN THE PARENTAL ENVIRONMENTS. 145

FIGURE 6.2. GENETIC LINKAGE MAP WITH POSITIONS OF QTL FOR TRAITS MEASURED IN TWO FIELD ENVIRONMENTS. 148

List of Tables

TABLE 2.1. STUDIES REPORTING GENETIC CORRELATIONS AND QTL ARCHITECTURE AMONG TRAITS. 16

TABLE 2.2. SUMMARY STATISTICS OF DATA BETWEEN SIGNIFICANTLY CORRELATED TRAITS VERSUS UNCORRELATED TRAITS. 22

TABLE 2.3. NUMBER AND PROPORTION OF SHARED QTL AMONG PAIRS OF TRAITS GROUPED BY THE SIGN OF R_G AND R_Q. 28

TABLE 2.4. MEAN DATA VALUES (AND STANDARD DEVIATIONS) FOR TRAITS THAT HAVE NEGATIVE PLEIOTROPIC QTL 31

TABLE 3.1. MEANS, RANGES, AND VARIANCE COMPONENTS AMONG TRAITS IN THE GREENHOUSE FOR XERICS, MESICS AND THE F6 GENERATION. 46

TABLE 3.2A) DIFFERENCES AMONG MESIC AND XERIC ECOTYPES TESTED USING NESTED MULTIVARIATE ANOVA. 49

TABLE 3.2B) DIFFERENCES AMONG F6 LINES TESTED USING MULTIVARIATE ANOVA. 50

TABLE 3.3. WEIGHTED LEAST SQUARES TESTS FOR DEPARTURES FROM AN ADDITIVE RELATIONSHIP AMONG GENERATION MEANS. 51

TABLE 3.4. GENETIC CORRELATIONS AMONG TRAITS MEASURED IN THE F6 LINES. 53

TABLE 3.5. PRINCIPAL COMPONENTS ANALYSIS ON MEAN TRAIT VALUES IN THE F6 LINES. 55

TABLE 4.1. AFLP SELECTIVE AMPLIFICATION PRIMER CODES ALONG WITH THEIR SELECTIVE NUCLEOTIDE SEQUENCES. 73

TABLE 4.2. NUMBER OF POLYMORPHIC MARKERS FROM THE 20 PRIMER PAIRS USED TO CONSTRUCT THE GENETIC LINKAGE MAP. 79

TABLE 4.3. AFLP MARKER LOCI EXHIBITING SEGREGATION DISTORTION SIGNIFICANT AT $P<0.01$ 80

TABLE 5.1. QTL IDENTIFIED FOR TRAITS MEASURED IN THE GREENHOUSE. 101

TABLE 5.2. NUMBER OF TWO-WAY MARKER INTERACTIONS FOR EACH TRAIT MEASURED IN THE GREENHOUSE. 115

TABLE 5.3. EPISTATICALLY INTERACTING PAIRS OF MARKER LOCI SIGNIFICANT AT THE BONFERRONI ADJUSTED 10% SIGNIFICANCE LEVEL. 115

TABLE 6.1. MEANS, RANGES AND VARIANCE COMPONENTS FOR TRAITS MEASURED IN THE FIELD ENVIRONMENTS. 135

TABLE 6.2A. DIFFERENCES BETWEEN ECOTYPES TESTED USING NESTED MULTIVARIATE ANOVA. 137

TABLE 6.2B. DIFFERENCES BETWEEN ECOTYPES TESTED USING A REDUCED MULTIVARIATE ANOVA MODEL. 137

TABLE 6.3. WEIGHTED LEAST SQUARES ESTIMATES OF GENERATION MEAN (M), ADDITIVE ([A]) AND EPISTATIC ([AA]) EFFECTS IN THE FIELD ENVIRONMENTS. 138

TABLE 6.4A. GENETIC CORRELATIONS DERIVED FROM VARIANCE COMPONENTS AMONG TRAITS MEASURED AT THE SIERRA FIELD SITE. 140

TABLE 6.4B. GENETIC CORRELATIONS DERIVED FROM VARIANCE COMPONENTS AMONG TRAITS MEASURED AT THE HOPLAND FIELD SITE. 140

TABLE 6.5. GENETIC CORRELATION ACROSS FIELD AND GREENHOUSE ENVIRONMENTS ESTIMATED FROM LINE MEANS. 142

TABLE 6.6. PRINCIPAL COMPONENTS ANALYSIS ON TRAIT DATA COLLECTED FROM THE GREENHOUSE AND FIELD SITES. 144

TABLE 6.8. NUMBER OF TWO-WAY INTERACTIONS AMONG MARKER LOCI DETECTED IN THE FIELD ENVIRONMENTS. 158

TABLE 6.9. EPISTATICALLY INTERACTING PAIRS OF MARKER LOCI SIGNIFICANT AT THE BONFERRONI ADJUSTED 10% SIGNIFICANCE LEVEL. 158

TABLE 7.1. COMPARISON OF GENETIC PARAMETERS MEASURED AT THE PHENOTYPIC AND GENETIC LEVEL 176

Abstract

I conducted a series of experiments designed to characterize the phenotypic differences between two genetically distinct ecotypes of *Avena barbata,* that inhabit mesic and xeric habitats, and their F6 inbred line progeny in both common garden greenhouse and natural environments. In the greenhouse environment, the two ecotypes differed for a suite of quantitative traits including fitness. These differences translated into a substantial release of variation among the F6 lines, with broadsense hertitabilities for most traits exceeding 50% and the range of phenotypes exceeding both parental ecotypes. There were constraints to the release of trait variation manifest as strong genetic correlations among traits, including flowering time and spikelet production. A genetic linkage map of the wild oat genome was constructed by genotyping 180 F6 inbred lines with 133 Amplified Fragment Length Polymorphism (AFLP) markers. The map spanned 640 cM, or approximately 40%, of the genome and consisted of 21 linkage groups. Quantitative Trait Loci (QTL) were identified for all traits in the greenhouse except one. Each trait had at least one QTL with additive effects in the opposite direction than expected, thus providing a mechanism for transgressive segregation. Clusters of QTL were found on several linkage groups indicating pleiotropic loci underlie the genetic correlations among traits. The QTL could account for approximately 50-60% of the genetic variation indicating a relatively few loci account for a substantial amount of the variation among traits. In the native environments, the effects of epistasis overshadowed the additive effects for fitness and plant height. Growth and fitness were poorly genetically correlated across field environments, and the greenhouse, indicating the presence of loci with environment specific effects. Few QTL were identified for fitness in the field, and they mapped to locations devoid of QTL for fitness in the greenhouse.

Acknowledgements

First and foremost I would like to thank my supervisor Dr. Bob Latta for his support over the years. I thank him for his input in all aspects of this project, for allowing me to expand my horizons beyond just the lab bench, and for the free trips to California. However, I do not thank Bob for agreeing to go to that Applebee's in Yuba City.

Thanks as well to all the fomer Latta Lab members, J. Johansen, J. MacKenzie, S. Yorke, for helpful discussions about work and life in general. I would like to extend a special thank you to Joanna MacKenzie for help with harvesting the greenhouse experiment, counting 300 000 seeds, and bagging 1000 plants despite weeklong 35 C temperatures and 100 % humidity in the summer of 2002. Thanks as well go to Carman Mills for advice with all things greenhouse related, and for keeping my experiments free of aphids.

On the lab side of things, I want to acknowledge Doug Cook for teaching me almost everything there is to know about running acrylamide gels on outdated technology, and Kristian Gordos for pestering sales representatives for free samples of comsumables and keeping me in gel solution through shipping delays and faculty strikes.

I am also grateful to my supervisory committee members, Drs. Mark Johnston and Christophe Herbinger. I thank them for their time, energy and comments every step of the way, from project proposals to preliminary exam to thesis defence.

Last and certainly not least I would like to thank my wife Audra for her love and support..... even if she'll be the *real* doctor someday. You are a constant source of inspiration (and amusement) and I am grateful for all that you bring to my life.... including not letting me take myself too seriously. I would also like to thank the people around the Biology department who have made the last few years interesting: F. Harper, J. Addison, W. MacDonald, A. Haigh, and N. Gorguy.

This research project was supported by a National Science and Engineering Research Council of Canada (NSERC) operating grant to Dr. Bob Latta and an NSERC PGS A scholarship to Kyle Gardner. Additional salary support was provided by the Faculty of Graduate Studies, through the Department of Biology, Dalhousie University.

Chapter One :

Introduction

Introduction

Traditionally the study of quantitative traits has taken a statistical approach, using means and variances between generations of an experimental population, to describe the inheritance of continuously varying traits (Kearsey and Pooni, 1996; Lynch and Walsh, 1998). Since very little can be inferred about the genotype at specific loci controlling a quantitative trait purely from its phenotype, early studies reported the cumulative effect of all genetic and non-genetic (i.e. environmental) sources of variation in a trait (Falconer and Mackay, 1996). It was generally assumed however that there were multiple genetic factors (eventually termed polygenes or quantitative trait loci (QTL)), all with equal but small effects, segregating for each trait, the combination of which produced the normal distribution of phenotypes characteristic of most quantitative traits (Tanksley, 1993; Kearsey and Farquhar, 1998). This assumption served to greatly reduce the complexity of some of the statistical models of polygenic inheritance, however its applicability to modeling evolutionary change in natural populations was questionable (Orr, 1999). Indeed most theory regarding the evolution of quantitative characters has in the past been based on unreasonable theoretical assumptions (e.g. infinite genes for each trait, no epistasis, equal gene effects) that have little relevance in biological systems (Lynch and Walsh, 1998).

With the development of DNA based molecular markers mapping individual QTL and estimating their effect has become feasible in many study systems (Kearsey and Pooni, 1996; Lynch and Walsh, 1998). The basic method, first proposed by Botstein *et al.* (1980), involves creating a genomic linkage map of markers that are dispersed randomly throughout the genome and using it to detect all marker-QTL linkage for a given trait. This was essentially a large-scale extension of using single gene (Mendelian segregating) marker-trait associations, first described by Sax (1923), to indirectly describe the complex patterns of quantitative trait inheritance. The term "genetic architecture" of quantitative traits has since been used to describe the numbers of QTL present, their individual effects on the trait, their location in the genome and their interactions with each other and the environment (Zeng *et al.*, 1999). Gaining an understanding of these details of the components quantitative trait variation will allow for increased insight into the connection between an individual's genotype and observable phenotype (Zeng *et al.*, 1999).

Much of the support for the map based QTL identification method has come from applied agricultural studies. This logically follows from the fact that identification of any of the components of the genetic architecture has immediate practical use to selective breeders, especially those simply interested in locating one or two QTL of large effect for an economically important trait, such as flowering time or milk yield, for incorporation into improved lines of crops or livestock (Kearsey and Farquhar, 1998). At present there are complete linkage maps available for most of the major crop plants including maize (Austin and Lee, 1996), oats (O'Donoughue *et al.*, 1995; Wight *et al.*, 2003) and barley (Mesfin *et al.*, 2003; Verhoeven *et al.*, 2004)) which have been used to map QTL for

yield related traits. QTL have also been mapped for a wide variety of other traits most notably factors for resistance to pathogens in wheat (Bai *et al.,* 1999), salt tolerance in tomato (Foolad *et al.*, 1997). In general, most mapping studies report from 1 to 16 QTL that can explain anywhere up to 95% of the variation in any one crossing experiment (Kearsey and Farquhar, 1998). It is questionable whether or not this represents the trend actually found in natural populations primarily due to the fact that the genotypes used in the agricultural mapping are probably not representative of their wild progenitors. This does however open the door to comparative mapping between cultivars and their wild relatives, which may yield insight into the evolutionary origins of domesticated species.

Based on the success in the applied breeding applications, QTL analysis has become an area of interest of evolutionary quantitative genetics. This has been due to the fact that most ecologically important (i.e. adaptive) traits, like most agronomic traits, are quantitative in nature (Tanksley, 1993). The dissection of the quantitative genetic architecture of these adaptive or "fitness" traits will allow for longstanding theoretical questions regarding evolutionary change in natural populations to be addressed in an empirical manner (Mitchell-Olds, 1995). Of particular interest will be examining the prevalence of QTL pleiotropic effects on correlated traits, which will lead to insight into mechanisms for constraint to evolutionary change and the genetic basis for ecological tradeoffs. In addition it will be possible to assess the importance of "major" genes and genotype by environment interaction in the process of adaptation to local environment (Mitchell-Olds, 1995; Hoffmann *et al.*, 1995). Perhaps the most significant overall result will be gained from identifying QTL for fitness. Certainly the number of fitness factors identified and their estimated additive effects will allow for refinement in our

understanding about potential response to natural selection and subsequently the maintenance of genetic variation over long periods of time (Mitchell-Olds, 1995; Merila and Sheldon, 1999; Barton and Keightley, 2002).

Biology and Natural History of *Avena barbata*

The slender wild oat is an annual, highly self fertilizing grass (>95% selfing) believed to have been derived from a polyploidization event following hybridization of *A. hirtula* and *A. westii* (Allard *et al.*, 1993; Garcia *et al.*, 1989). Cytological characterization of the *A. barbata* genome has shown that this species is an autotetraploid, comprised of the A and B oat homeologous genomes (Allard *et al.* 1993; Katsiotis *et al.*, 1997). *A. barbata*, like all other oat species (except *A. macrostachya)*, is only known to form bivalents during meiosis (i.e. AA or BB) indicating there is no exchange of genetic material between genomes. As a result *A. barbata* can be considered to act as a diploid species (2n=4x=28; Allard *et al.*, 1993). This has been confirmed by examining the patterns of allelic segregation of a select group of allozyme molecular markers in a F2 test population (Hakim-Elahi and Allard, 1983; Hutchinson *et al.*, 1983).

Ancestrally, *A. barbata* is known to have been distributed throughout the Mediterranean basin, from Spain to the more arid climates of the Middle East (Allard *et al.*, 1993). Throughout this range, *A. barbata* is considered the most successful colonizer among all other tetraploid oat species (Rajhathy and Thomas, 1974). As such it has become a significant component of fodder grass in livestock pastures and a weed of cultivated fields (Rajhathy and Thomas, 1974). The current distribution of this species

has also come to include areas where accidental introductions have allowed for establishment of large *A. barbata* populations (Allard *et al.*, 1972). One such event, approximately two hundred years ago, introduced the slender wild oat to the west coast of the United States during the Spanish colonization of California and Oregon (Clegg and Allard, 1972; Hamrick and Allard, 1972).

Early population level studies of the Californian oats have shown that there are two predominant ecotypes, each associated with a specific climatic region (Clegg and Allard, 1972). The "xeric" ecotype is strongly associated with the more arid region (simply referred to as Region I) and was originally characterized as having a distinct monomorphic multilocus allozyme genotype. In contrast, the "mesic" ecotype is associated with the cooler, more moist region (i.e. Region II) and has also been distinguished as having a monomorphic genotype at the same allozyme loci, however for the alternative alleles as in the xeric genotype (Clegg and Allard, 1972). When surveying the same populations with ribosomal DNA markers Cluster and Allard (1995) found a similar pattern of each dominant ecotype having a specific genotype.

Interestingly, the multilocus allozyme/rDNA genotypes that define the Californian ecotypes of *A. barbata* are completely absent in all the ancestral Spanish populations surveyed (Cluster and Allard, 1995). This would seem to indicate that the mesic and xeric types arose after introduction into California, 200 years ago. Therefore it is believed that the differentiation into the two ecotypes (and the resulting population structure) has most likely resulted from a combination of selection imposed by the local environmental conditions, and the maintenance of genetic homogeneity made possible by

the highly self-fertilizing mating system (Clegg and Allard, 1972; Hamrick and Allard, 1972).

This study system should provide an excellent means to characterize the genetic basis for adaptation to local environment for several reasons. First, due to the high selfing rate, the two ecotypes can essentially be considered naturally occurring inbred lines, fixed for alternate genotypes. Since the two ecotypes differ for many ecologically important traits, they are presumed to be adapted to different environmental extremes (Hamrick and Allard, 1976; Latta *et al.*, 2004). As a result, a cross between the two ecotypes (referred to as the parental genotypes) will provide QTL detection power similar to that obtained by using inbred laboratory strains.

Creation of the Recombinant Inbred Line Population

The seed stock used to create the recombinant inbred line population was derived from wild collected xeric and mesic specimens donated to Dr. Robert Latta by Dr. Pedro Garcia from collections made in California (see Garcia *et al.*, 1989). Prior to crossing, the parental ecotypes were propagated by selfing in a greenhouse to expand the number of parental seeds. Six reciprocal F1 individuals were created by crossing a mesic individual (pollen donor) with an emasculated xeric individual (pollen recipient), and *vice versa*, by hand pollination. One resultant F1, expected to be heterozygous at all loci that differed among the parental ecotypes, was sown in the McGill university phytotron in a 12 inch pot filled with soil, sand and peat that was well watered and fertilized. The F1 was grown

to senescence and naturally self fertilized producing of a large number of F2 seeds. Two hundred F2 individuals were randomly chosen to initiate the recombinant inbred line population. Each F2 was sown in an individual pot, grown to maturity and allowed to self fertilize producing F3 lines. A single F3 seed from each line was used to propagate, by self fertilization, the line to the F4 generation. To eliminate the influence of natural selection this process of single seed descent (SSD) with self-fertilization was carried out for several generations until the lines reached the F6 generation at which point individuals within lines are nearly genetically uniform (~97% homozygosity).

Research Objectives

In this thesis I address questions about the underlying genetics of covarying quantitative traits in order to characterize the adaptive population divergence of two ecotypes of *Avena barbata*. I make use of traditional biometric analyses coupled with molecular techniques to first quantify the amount of variation that can be released by crossing divergent genotypes and then attempt to identify the individual portions of the genome that contribute to this variation.

As a prelude to the experimental work I begin with a literature review (Chapter Two) aimed at identifying studies that report both genetic correlation and QTL estimates for suites of quantitative traits. The goal of the review was to determine if significantly correlated traits shared more QTL in common, and whether the estimates of QTL effects can explain the magnitude of the biometric genetic correlation. In addition I explore the QTL architecture of negative genetic correlations that may lead to tradeoffs among traits.

I then describe a common garden greenhouse experiment (Chapter Three) where the F6 recombinant inbred line population is grown alongside the parental ecotypes. This experiment was designed to obtain the phenotypic data needed for identifying the individual QTL underlying traits that are presumed to have been important in the differentiation of an ancestral *A. barbata* population into the two ecotypes observed in California. As well I examine the amount genetic variation and covariation released among traits in order to determine the role of recombination in generating the fuel for natural selection. I also comment on the possible role of epistasis, in the form of coadapted gene complexes, in the formation of the mesic and xeric ecotypes.

In Chapter Four I present efforts to generate a genetic linkage map of the *A. barbata* genome using Amplified Fragment Length Polymorphism (AFLP; Vos *et al.*, 1995) molecular markers. I also explore both the utility of these markers in differentiating between the parental ecotypes and illustrating the genetic uniformity of individuals within the parental ecotypes.

By combining the molecular map with the phenotypic data collected in the greenhouse I will map the individual QTL that control trait variation in morphological and fitness related traits in the F6 line population (Chapter Five). Both the number of loci and their individual positions and effects will be examined to try to provide an explanation for significantly genetically correlated traits and thus possible constraints to evolutionary change.

Using the results of the greenhouse study as a guide, I then analyze a field experiment where the F6 lines are grown in a common garden experiment, with the parental ecotypes, within each native parental environment (Chapter Six). The methods

used in Chapter Three and Chapter Five will be employed to determine how growth and fitness correlate across environments and whether common QTL control variation and covariation in more than one environment.

Finally, I provide a summary of the results of both the greenhouse and field studies, comparing the efficacy of QTL mapping in explaining the patterns of variation and covariation in this study system (Chapter Seven).

Chapter Two:

Co-localizing QTL in the Estimation of Genetic Correlation Between Continuous Traits

Introduction:

Genetic correlations among traits have the effect of preventing the correlated traits from independently responding to selection. Selection on one trait will produce a response in all of the correlated traits (Lande and Arnold, 1983; Reznick, 1985; Arnold, 1992), and selection on one trait toward its optimum value may tend to drag other correlated traits away from their optima. Alternatively, selection acting in opposite directions on two correlated traits will tend to cancel out, such that with balanced selective forces the net result is little or no change. Thus genetic correlations tend to constrain the response to selection, by limiting the number of trait combinations which can occur, and may place a major limitation on the evolution of multitrait phenotypes (Arnold, 1992).

Genetic correlation can result either from pleiotropy, (where the same locus or loci have effects on two or more separate traits) or from linkage disequilibrium between separate loci, each of which affects a single trait (Lynch and Walsh, 1998; Via and Hawthorne, 2002). The distinction is important because linkage disequilibrium can be broken up by recombination, breaking down the genetic correlation (Via and Hawthorne, 2002; Conner, 2002). Pleiotropy, by contrast, cannot be broken down by recombination (see Conner, 2002), but instead will persist until mutations accumulate which have allelic effects on the two traits orthogonal to the correlation (i.e. if the traits are positively

correlated, mutations with positive effects on both traits would tend to increase the correlation, while a mutation with opposite effects on the two traits would reduce the correlation). Genetic correlations caused by linkage disequilibrium thus present a less permanent constraint to selection than do those caused by pleiotropic genes (Via and Hawthorne, 2002).

Of particular interest to evolutionary biologists are negative genetic correlations leading to tradeoffs among fitness related traits. Because limited resources must be distributed among multiple different functions, we expect such tradeoffs to occur frequently for traits that are under selection to increase in value (e.g. life history traits). If such tradeoffs occur through antagonistic pleiotropy they represent a substantial evolutionary constraint, and may provide an important mechanism for the maintenance of genetic variation in natural populations (Roff, 2002). However, while trade-offs have been documented in some systems such as *Arabidopsis thaliana* (Mitchell-Olds, 1996a; Mitchell-Olds, 1996b), *Drosophila melanogaster* (reviewed in Roff, 2002) and *Gryllus firmus* (Stirling *et al.* 1999), documenting such trade-offs has proven difficult in many organisms because variation in overall 'vigour' tends to produce positive correlations among traits that mask the tradeoff (Houle, 1991; Fry, 1993). Other pleiotropic loci found elsewhere in the genome, that affect the traits in the same direction ("positive" pleiotropy), can produce enough positive covariance to essentially mask the effects of the negatively pleiotropic loci and make the traits have a net positive correlation (Houle, 1991; de Jong and van Noordwijk, 1992; Fry, 1993). Houle (1991) describes this situation in terms genes affecting resource acquisition (which tend to affect all traits in parallel and thus lead to positive correlations) masking trade-offs at genes affecting

resource allocation. Fry (1993, 1996) argued that this potential for masking of negative pleiotropic loci makes it impossible to disprove the presence of a tradeoff simply by measuring the genetic correlation. Instead, direct information on the individual pleiotropic loci is needed.

Quantitative Trait Locus (QTL) mapping techniques (reviewed in Tanksley,1993; Doerge *et al.*, 1997; Doerge, 2002) provide a means to identify, at least in part, the regions of a genome containing the genes underlying quantitative traits, and by extension to locate pleiotropic genes. If pleiotropy underlies genetic correlation, then the same QTL should underlie correlated traits. Furthermore, QTL methods provide estimates of the allelic effects of the QTL on each trait (Lynch and Walsh, 1998). The effect sizes are commonly used to estimate how much of the observed phenotypic variance in a trait is explained by the identified QTL (Falconer and MacKay, 1996; Lynch and Walsh, 1998). By extension, using the effects of the pleiotropic QTL it should be possible to determine how much of the covariance between traits can be explained.

In this paper, I analyze data from QTL mapping studies to ask (1) Do trait pairs showing significant genetic correlation map to the same QTL, and do they share more QTL than do uncorrelated traits?, (2) Does the strength of the genetic correlation expected from the QTL effect sizes match the reported value based on biometric trait data?, (3) Does antagonistic pleiotropy necessarily lead to negative correlations and how often are they hidden within positive correlations?

Predicting correlation from QTL effects.

To determine whether pleiotropic QTL are sufficient to explain the measured correlation between two traits, we need to predict how much correlation would be expected if the identified QTL were the only loci underlying the traits. Using allelic effects at a locus, a (defined as half the difference of trait value between the homozygotes (Falconer and MacKay (1996) Fig. 7.1), the additive genetic variance expected in a trait, Z, is found by summing the squared allelic effects (deviations) across all loci

$$V_A = \sum_i^n 2p_i q_i a_i^2 \qquad \text{Eq. 1}$$

where p and q are allele frequencies. By extension, the covariance between trait Z1 and Z2 can be obtained by taking the cross products of the allelic effects and summing across all loci:

$$COV_{A\,(Z1,Z2)} = \sum_i^n 2p_i q_i a_{i(Z1)} a_{i(Z2)} \qquad \text{Eq. 2}$$

where $a_{i(Z1)}$ is the additive effect of locus i on trait Z1. From here it is straightforward to get the genetic correlation expected if pleiotropic QTL fully account for the genetic correlation:

$$r_{Q\,(Z1,Z2)} = \frac{COV_{A\,(Z1/Z2)}}{\sqrt{V_{A\,(Z1)} \times V_{A\,(Z2)}}} = \frac{\sum_i^n 2p_i q_i a_{i(Z1)} a_{i(Z2)}}{\sqrt{(\sum_i^n 2p_i q_i a_{i(Z1)}^2) \times (\sum_i^n 2p_i q_i a_{i(Z2)}^2)}} \qquad \text{Eq. 3}$$

In the above, I assume that all loci have purely additive effects on each trait. The influence of dominance effects on the additive genetic variance depends on d(q-p) (Falconer and MacKay, 1996). However, QTL mapping designs based on F2, Doubled haploid or recombinant inbred line mapping populations will be somewhat robust to this assumption, because p=q=0.5 across all loci, thus d(q-p) = 0.

Materials and Methods:

I searched the literature for studies that report both the QTL architecture of individual quantitative traits and genetic correlation between pairs of such traits. I surveyed five genetics and applied genetics journals (*Genetics, Theoretical and Applied Genetics, Crop Science, Genome,* and *Heredity*) between 1994 and the spring of 2003. I included only those studies that reported QTL magnitude as the additive effect (a) of alleles. Numerous studies reported instead, the proportion of variance explained by individual QTL, but these were excluded because they do not permit calculation of r_Q.

Most studies meeting these criteria were on important agronomic crop species or model species, such as *Arabidopsis,* because organisms of this type are highly amenable to inbred line experimental designs, which are easier than other crossing designs (e.g. F2, backross) to analyze for both QTL and genetic correlation estimates. However, a few studies involving model zoological systems, including *Drosophila* and pea aphids, were included. In a few studies, trait expression in different years or environments, were treated as separate correlated traits (Falconer, 1952). In order to prevent bias, data for only one year/environment was included. Similarly, two studies were removed due to discrepancies between the values for QTL position and magnitude reported in the text of the papers versus the data illustrated in the referenced tables and figures.

In total, 238 traits giving 1150 trait pairs across 27 studies were included in the data set (Table 2.1). Of these, 1108 trait pairs had r_G and r_Q values available for

Table 2.1. Studies reporting genetic correlations and QTL architecture among traits. Studies are ordered by total number of traits examined. p-values refer to the result of the Mantel randomization test among r_Q and r_G . IM – interval mapping, CIM – composite interval mapping, JCIM – joint trait composite interval mapping, MMIM – mixed model interval mapping.

Species	Cross Design	QTL Method	Correlation Method	Number of Traits	p-value	Ref.
Capsicum annuum	F3	IM	Line means	26	0.001	BenChaim *et al.*(2001)
Lycopersicon spp.	RIL	CIM	Line means	26	0.001	Saliba-Columbani *et al.* (2001)
Hordeum vulgare	RIL	CIM	Line means	14	0.001	Mesfin *et al.*(2003)
Arabidopsis thaliana	RIL	CIM	Var. comp	14	0.001	Ungerer *et al.* (2002)*
Arabidopsis thaliana	RIL	CIM	Var. comp	14	0.001	Ungerer *et al.* (2002)*
Arabidopsis thaliana	RIL	IM	Var. comp	11	0.048	Mitchelolds and Pedersen (1998)
Oryza sativa	RIL	MMIM	Line means	10	0.001	Cui *et al.* (2002)
Drosophila melanogaster	RIL	JCIM	Line means	10	0.0002	Viera *et al.* (2000)
Arabidopsis thaliana	RIL	CIM	Var. comp	8	0.001	Juenger *et al.* (2000)
Zea mays	F3	IM	Line means	8	0.002	Veldbloom and Lee (1996)a
Oryza sativa	DH	IM	Var. comp	8	0.002	Yan *et al.* (1999)
Avena sativa	RIL	CIM	Line means	8	0.005	Zhu and Kaeppler (2003)
Mimulus guttatus	F2	JCIM	Subtract Env.	7	0.006	Fishman *et al.* (2001)
Helianthus annuus	RIL	CIM	Var. comp	7	0.005	Herve *et al.* (2001)
Oryza sativa	RIL	CIM	Line means	7	0.002	Tan *et al.* (2001)
Hordeum vulgare	DH	CIM	Line means	7	0.001	Tinker *et al.* (1996)
Lycopersicon pimpinellifolium	F2	IM	Subtract Env.	6	0.181	Georgiady *et al.* (2002)
Oryza sativa	RIL	MMIM	Line means	6	0.154	Zhuang *et al.* (2002)

Table 2.1 continued.

Species	Cross Design	QTL Method	Correlation Method	Number of Traits	p-value	Ref.
Zea mays	RIL	CIM	Line means	5	0.009	Austin and Lee (1998)
Triticum aestivum	RIL	CIM	Line means	5	0.445	Kato *et al.* (2000)
Zea mays	F3	CIM	Var. comp	5	0.01	Lubberstedt *et al.* (1997)
Zea mays	F3	IM	Line means	5	0.008	Veldbloom and Lee (1996)b
Zea spp.	F3	CIM	Line means	4	0.153	Bohn *et al.* (1996)
Avena sativa	RIL	CIM	Line means	4	0.481	Groh *et al.* (2001)
Triticum aestivum	RIL	CIM	Line means**	3	na	Igrejas *et al.* (2002)
Oryza spp.	RIL	MMIM	Line means	3	na	Kamoshita *et al.* (2002)
Oryza spp.	DH	JCIM	Line means	3	na	Zhou *et al.* (2001)

* (two mapping populations in one study)
** (correlations of line means not explicitly stated as "correlation among lines")

comparison. There was considerable variation in the information contribution among individual studies, ranging from less than five trait pairs to 300 trait pairs. Some studies did not report correlations for all pairs of traits. Within each study I recorded the cross type (e.g. F2, RIL etc.), reported genetic correlation estimate (r_G), statistical significance of this estimate for each pair of traits, and the position and effect of all QTL. While many studies reported point estimates of r_G separately from significance tests, many studies simply reported "NS" for non-significant r_G. Although it is unlikely that all nonsignificant correlation estimates are indeed zero, I treated all nonsignificant r_G as zero, to keep data analysis uniform across studies.

For each trait pair I determined the raw number of co-localized (shared) QTL, and total number of QTL (across both traits). Since there is no consensus that I am aware of as to the description of a co-local or pleiotropic QTL, I defined QTL that have overlapping 95% confidence intervals (on their chromosomal position) as mapping to the "same" position. The 95% confidence intervals reported by the authors were used when given. Most studies reported confidence intervals estimated by bootstrapping or the one or two LOD support interval method (Doerge, 1997; Lynch and Walsh, 1998). Both methods have been shown to give a similar result of a somewhat conservative (i.e. wide) confidence region (Doerge, 1997). In instances where there were no confidence intervals given for the position of a QTL, I applied an arbitrary 10 centiMorgan (cM) confidence interval centered on the reported point estimate of QTL position. The average reported confidence interval around a QTL position estimate was 15.6 cM (based on >200 identified QTL), making my choice of 10cM somewhat conservative in the chance of confidence limits overlapping.

To measure the proportion of QTL shared between traits, while taking into account the total number of QTL detected for both traits, I used a formula originally proposed to compare DNA sequences (Nei and Li, 1979). The proportion of shared QTL (PSQ) is defined as:

$$PSQ = \frac{2(\text{Number of co-local QTL})}{\text{Total number of QTL}} \qquad \text{Eq. 4}$$

The resulting scale is standardized between zero and one and allows for straightforward comparisons across multiple trait pairs.

All 1108 trait pairs were used to ask how well the QTL data (r_Q) agreed with the biometrical correlation estimates (r_G). Qualitatively, I asked how often significantly correlated traits ($r_G \neq 0$) had co-localizing QTL, and whether r_G and r_Q had the same sign. For cases where r_G and r_Q had the same sign, I measured the correlation between r_G and r_Q as a measure of precision. In addition, I used major axis regression of r_G on r_Q (Sokal and Rohlf, 1995) to determine the slope of the linear relationship between the correlation estimates. I also used the difference between r_G and r_Q as a measure of estimation accuracy to determine if there is any bias toward over or underestimating a genetic correlation when using QTL data, with $r_G - r_Q = 0$ the expectation under no bias.

To assess the significance of the correlation between r_G and r_Q, a separate Mantel test was carried out for each study that had complete r_Q and r_G correlation matrices. These were compared using the raw correlation estimates and correlation values transformed via Fisher's Z transformation (Sokal and Rohlf,1996). Only the results from the use of the raw correlation are reported, as use of the transformed values did not

change the outcome of any analyses. The sampling distributions of the Mantel test statistic (Z) for each study were derived from 9999 matrix randomizations using the program MANTEL (Liedloff, 1999). In order to include data from studies that did not report all the possible pairwise correlations between traits (i.e. incomplete and asymmetrical matrices unable to be used in the Mantel tests), the listwise pairs of r_Q and r_G were compared using the "Correlation via Randomization" feature in the RESAMPLING program package (Howell, 2000), again using 9999 randomizations. The p-values from each test were then combined across studies, using the method described in Sokal and Rohlf (1996), to give an overall level of significance to the test of association between r_Q and r_G.

Trade offs (negative genetic correlations) must result from at least one negatively pleiotropic locus. However, such negative pleiotropy can be swamped by other positively pleiotropic loci, giving a positive genetic correlation (Houle,1991; de Jong and van Noordwijk, 1992). I therefore identified all instances where there was at least one shared (i.e. pleiotropic) QTL that had antagonistic effects on two traits ('negative pleiotropy'). I divided these trait pairs into three groups based upon the reported value of r_G ($r_G < 0$, $r_G = 0$ and $r_G > 0$), measured the relative frequency of these three outcomes, and compared the QTL architecture between groups.

Results:

Qualitative patterns:

The mean of the absolute values of r_G (0.45) across all significantly correlated trait pairs in the data set was found to approximate the overall genetic correlation estimate across trait types (~ 0.48) reported by Roff (1996) in a survey of approximately 1800 correlation estimates. Seventy percent of correlated trait pairs had identified colocalizing QTL (median 1 shared QTL), in contrast with only 30% of uncorrelated pairs (median 0 shared QTL), in spite of QTL being found for each of those traits individually (Table 2.2, Fig 2.1). The correlated trait pairs share significantly higher proportion of their QTL (higher PSQ) than uncorrelated traits (t-test, t = 13.7, df = 1108, p<.0001) with 30% of QTL co-localizing for correlated traits and 11% for the uncorrelated traits (Table 2.2). However the raw number of co-local QTL is relatively low, with 1.9 QTL shared between the average correlated trait pair. While small, this number of QTL is larger than that for uncorrelated trait pairs which on average shared less than one (0.88) QTL . Likewise, correlated traits had significantly larger values of r_Q (t-test, t =16.0, df = 1108, p<.0001) than uncorrelated traits (Table 2.2).

There is good qualitative agreement between r_G and r_Q. Out of a total of 1108 trait pairs, 68% (N=751) had r_G and r_Q of the same sign (Table 2.3). This included those instances where both r_Q and r_G were zero. The majority of the remaining trait pairs (32%) consisted of cases where a non-zero r_G was paired with a zero r_Q (N=114); or a non-zero r_Q with a zero r_G (N=215). These two results are likely attributable to lack of power either

Table 2.2. Summary statistics of data between significantly correlated traits versus uncorrelated traits.

	Significance			
	Correlated Traits		Uncorrelated Traits	
	Mean (SD)	Median	Mean (SD)	Median
$\|r_G\|$	0.20 (0.47)	0.40	0.00	0.00
$\|r_Q\|$	0.11 (0.43)	0.31	0.002 (0.19)	0.00
PSQ	0.34 (0.29)	0.33	0.14 (0.23)	0.00
Co-localized QTL	2.01 (4.05)	1.00	0.86 (3.21)	0.00
Total QTL	8.69 (8.07)	7.00	6.51 (6.29)	5.00

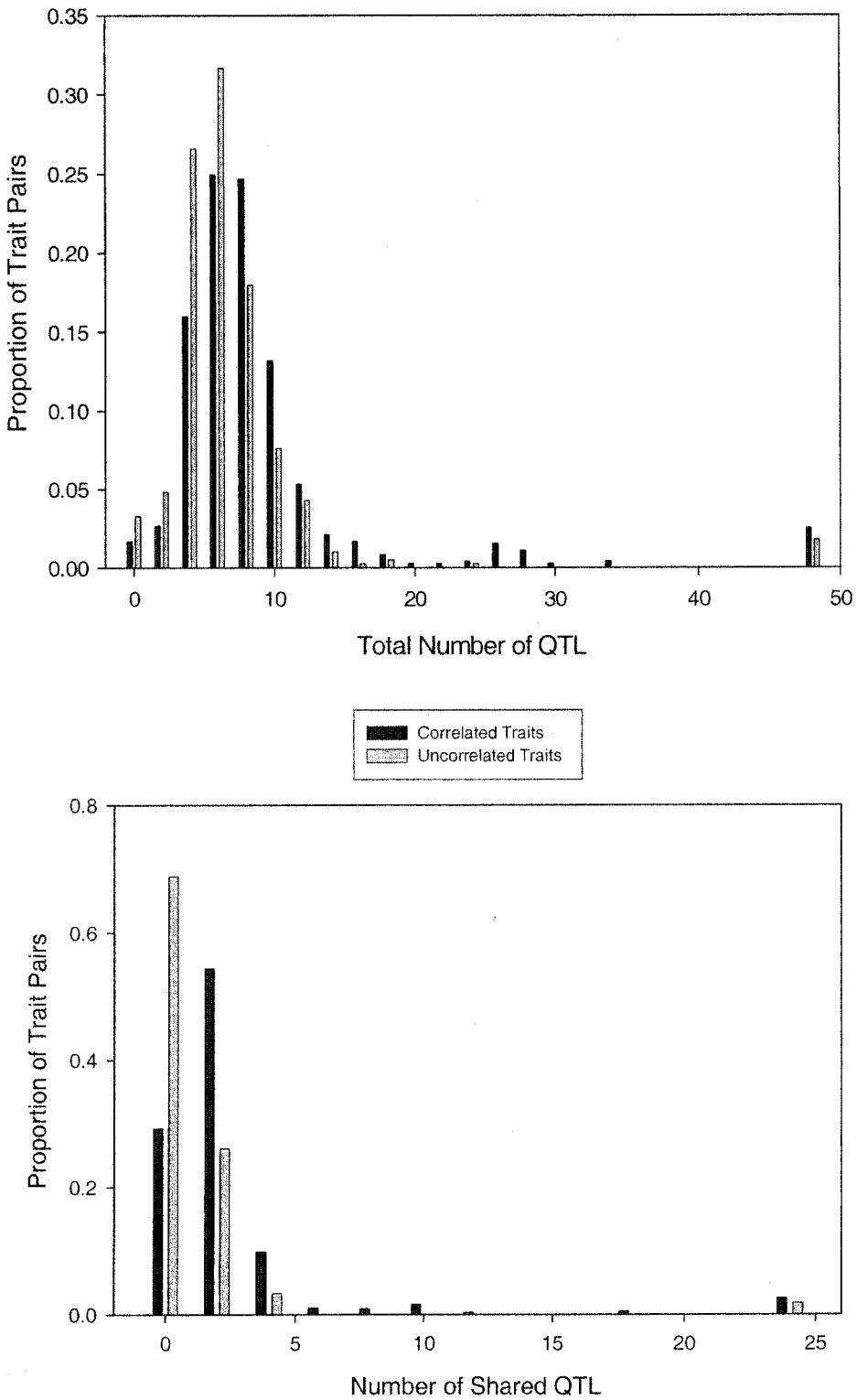


Figure 2.1. Histograms of correlation and QTL data for significantly correlated versus uncorrelated traits

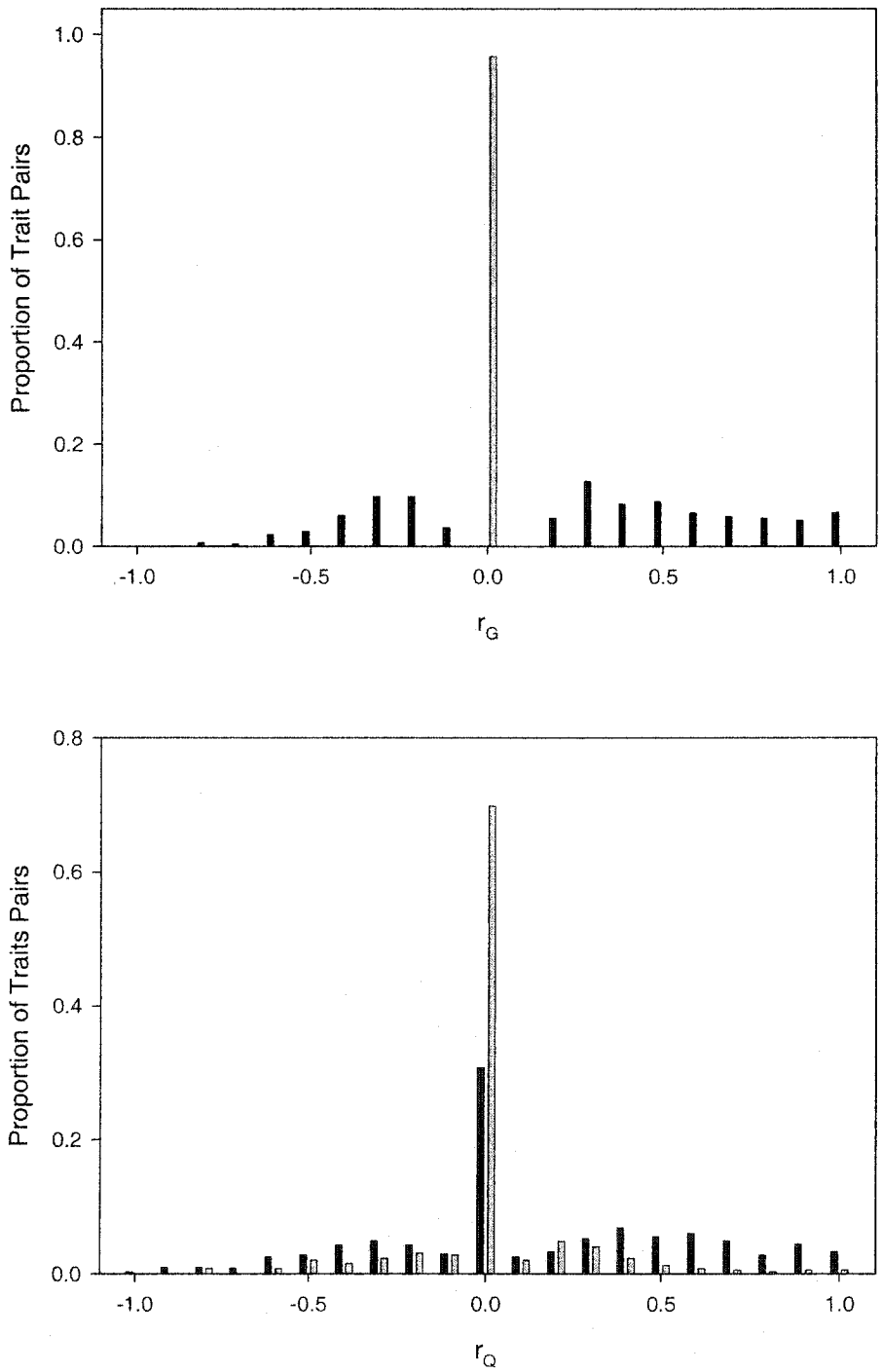
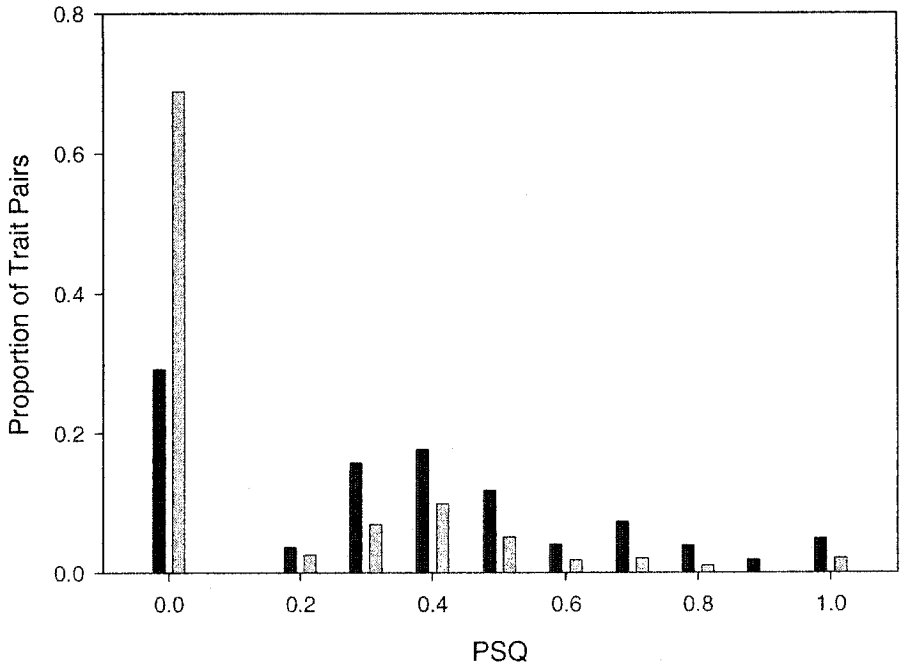


Figure 2.1 continued.



Figrure 2.1 continued.

to detect QTL or to detect correlations. In only 28 cases (2.5%) were r_G and r_Q of opposite signs.

Quantitative patterns : correlation of the correlations

There was a moderate linear correlation of 0.711 (Pearson product moment, N = 1108, p< .0001) between the r_Q and r_G estimates (Figure 2.2). The nonparametric Spearman rank correlation was 0.718 (N = 1108, p<.0001). For most studies Mantel tests revealed significant associations between the r_Q and r_G estimates (Table 2.1). The only studies that failed to reject the null hypothesis, of no association, were those that had only a few (<5) traits and thus only a small number of possible random permutations of the data in the Mantel test. The combined p-values across all Mantel tests show a highly significant association (χ^2 = 242.68, df = 42, p<0.00001) between the r_G and r_Q matrices, while controlling for the nonindependance within studies across the data set. Relatively few studies/trait pairs were excluded from the Mantel tests, and including the p-values from the correlation randomization for these studies (Table 2.1) does not alter this result (χ^2 = 319.24, df = 48, p<0.00001).

Of course, much of the overall correlation is driven by the qualitative agreement of sign of r_G and r_Q (Table 2.3). Where r_Q and r_G are both positive, variation in the positive r_Q estimates was found to explain approximately one quarter of the variation in the corresponding r_G estimates. Major axis regression (Type II) of r_G on r_Q yielded a slope not different from 1.00, indicating a relatively even spread of points around the line of unity (Figure 2.2). However, the mean difference between r_G and r_Q was positive, and

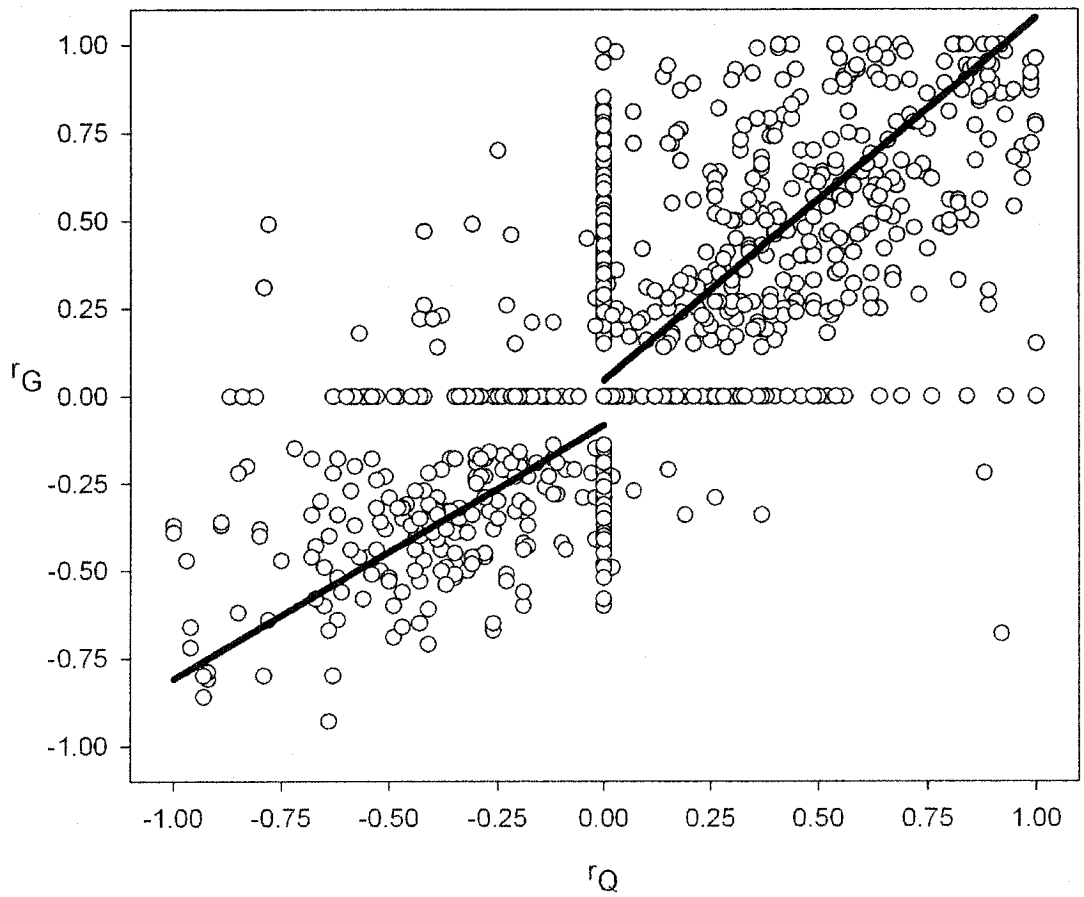


Figure 2.2. Scatter plot of r_G versus r_Q. Lines correspond to the slope of the major axis (Type II regression slope) of variation in the positive (b = 1.03) and negative (b = 0.73) quadrants respectively. Each data point represents one trait pair.

Table 2.3. Number and proportion of shared QTL among pairs of traits grouped by the sign of r_G and r_Q.

		r_Q			
		Negative	Zero	Positive	Total
r_G	Positive	N=19 PSQ:0.44 Shared QTL:4.10	N=136 PSQ:0 Shared QTL:0	N=315 PSQ:0.51 Shared QTL:3.35	470
	Zero	N=51 PSQ:0.40 Shared QTL:2.67	N=264 PSQ:0 Shared QTL:0	N=63 PSQ:0.44 Shared QTL:2.55	378
	Negative	N=172 PSQ:0.41 Shared QTL:1.76	N=79 PSQ:0 Shared QTL:0	N=9 PSQ:0.61 Shared QTL:4.33	260
	Total	242	479	387	1108

significantly different from zero ($r_G - r_Q = 0.051$, p=0.001) indicating there was a tendancy for r_Q to underestimate the magnitude of r_G (Figure 2.3). Trait pairs that shared a greater proportion of their QTL were generally more strongly correlated (r_G), however, the agreement between r_G and r_Q did not increase with a higher proportion of shared QTL.

A similar result was found in the negative quadrant (r_G and r_Q both negative) with variation in r_Q estimates explaining approximately 19% of the variation in r_G estimates (r^2=0.189, N=171). Interestingly the slope of the major axis regression for these data was significantly less than 1 (0.727). The mean difference between r_G and r_Q was again positive, and significantly different from zero (mean difference $r_G - r_Q = 0.034$, p=.044). Note that when both r_G and r_Q are negative, a positive difference means r_G is *less* negative than r_Q, indicating a slight tendency for r_Q to overestimate the strength of the genetic correlation for negatively correlated trait pairs. As with positively correlated trait pairs, a higher PSQ gives more strongly correlated traits, though this is not quite significant for r_G (p=0.055). Surprisingly, the agreement between r_G and r_Q *decreased* the higher the proportion of shared QTL.

Antagonistic QTL Pleiotropy and the Sign of Genetic Correlation:

Out of the 1108 trait pairs there were 276 cases (25%) where at least one antagonistically pleiotropic QTL was identified. The majority of the cases (177) led to a net negative biometrical genetic correlation (Table 2.4). In many fewer cases there was no correlation (45) or a net positive correlation (54) between the traits. These groups do not differ in the number of negative pleiotropic QTL (Table 2.4). Rather, the positively correlated trait

pairs have over twice as many shared QTL as the negatively correlated traits, and thus a greater proportion of the shared QTL have positive pleiotropy. Therefore antagonistic pleiotropy seems to have been swamped by positive pleiotropy in these cases.

Table 2.4. Mean data values (and standard deviations) for traits that have negative pleiotropic QTL. Trait classes are defined by the sign on the biometric estimate of genetic correlation.

Net Correlation Sign on r_G	r_Q	Number Negative Pleiotropic QTL	Total Number Co-local QTL	Total Number QTL	PSQ
Negative Correlation N=177	-0.405 (.237)	1.36 (.634)	1.51 (.782)	8.23 (3.50)	0.395 (.168)
Zero Correlation N=45	-0.197 (.132)	1.02 (.388)	1.48 (.661)	7.60 (3.16)	0.413 (.160)
Positive Correlation N=54	-0.020 (.395)	1.48 (.966)	3.52 (3.78)	12.55 (8.13)	0.509 (.190)

Discussion:

How well do QTL explain Genetic Correlations?

Genetically correlated traits are commonly interpreted to have some common genetic basis, in the form of either pleiotropic loci, or non-pleiotropic loci located in tightly linked parts of the genome. I found that correlated traits (r_G significantly different from zero) share a greater proportion (30%) of their QTL than uncorrelated traits (11%), and that traits which shared more QTL tended to be more tightly correlated (Table 2.1). However, the actual number of pleiotropic QTL was generally small –correlated traits typically share only one or two pleiotropic loci (mean 1.5). This seems to imply that a greater number of QTL in common will lead to larger values of genetic correlation, but sharing a few loci, most likely of moderate to large effect, can generate enough genetic covariance to lead to a net significant genetic correlation.

Using the additive effect of the pleiotropic QTL we can predict the genetic correlation. This prediction gave a good correlation with observed genetic correlations (Figure 2.2), and variation in r_Q explained approximately half of the variation in r_G. While this relationship can easily be seen from the points on the scatter plot of correlation values (Figure 2.2), what is perhaps equally apparent is that there are very few cases of r_Q predicting a value of opposite sign to r_G. This is rather compelling evidence that the error of estimating genetic correlation using QTL data is mainly in correlation magnitude rather than correlation sign. Bias in the prediction was slight and differs between positive and negative correlations – QTL data generally overestimated the strength of positive

associations, and underestimated the strength of negative correlations. Intriguingly, the agreement of r_G and r_Q increased with detection of additional pleiotropic QTL for positive correlations, but decreased with additional pleiotropic QTL for negative correlations.

Sources of error, and linkage

The observed scatter around the regression line results from the input of error from several sources. The most obvious error stems from the fact that both r_G and r_Q will be estimated with considerable error. Roff (1996) points out that genetic correlations have large confidence intervals at best. In addition, the large numbers of zero r_G and r_Q estimates are likewise candidates for error. Some of the time these two situations actually reflect the genetic independence of the traits, however more often they reflect the lack of statistical power to detect associations. In practice most of the zero r_G estimates would probably fall in the two gaps on either side of zero making the r_G distribution approximately normal (Figure 2.1) in shape. Similarly, a zero estimate of r_Q may reflect either a lack of pleiotropy or the inability to detect QTL. The sheer number of zero r_Q estimates (~500) is perhaps a good indicator of the inherent difficulty in finding QTL for a single trait, let alone finding QTL for more than one trait that map to the same position.

Furthermore, estimation bias in the detection of QTL, and the estimation of their effects, is a key problem to most QTL mapping methodology (Beavis, 1998). Bias in the detection of QTL arises due to limitations in statistical power to detect significant marker-trait associations and is essentially an issue of sample size (Orr, 2001). In most cases only the QTL of moderate to large effect are detected, leaving an unknown number

of small effect loci undetected and thus out of the analysis (see Otto and Jones (2000) for a method to "detect the undetected"). Such small loci may provide a plausible explanation for the overestimation of negative r_G values (Figure 2.2). If these undetected loci contribute positive trait covariance, they may pull the biometrical correlation estimate closer to zero and thus farther from the more negative r_Q .

A commonly voiced criticism of QTL studies is that multiple, closely linked loci will appear as a single QTL. I have operationally treated QTL with overlapping confidence limits as a single pleiotropic gene, though this will probably include some loci which are closely linked but distinct. I justify this decision in two ways. First, closely linked loci represent a small subset of locus pairs in linkage disequilibrium, many of which are unlinked. In natural populations, especially those with pronounced population structure, linkage disequilibrium between unlinked loci can be extensive and can contribute a substantial amount to the variance and covariance of traits (Hawthorne and Via, 2001). In QTL studies which I have surveyed here, crosses are performed to ensure that all unlinked loci have been independently assorted. Those loci which remain linked, thus represent a more substantial constraint on the response to selection. Moreover, colocal QTL underlying multiple traits in *Arabidopsis*, appear to represent truly pleiotropic genes (McKay *et al.*, 2003)

Second, while it is possible to take linked QTL into account when determining r_Q, and so create a more accurate prediction of expected r_G, many studies do not report the necessary information to do so. Two linked loci each affecting a different trait, will contribute to the covariance of traits as:

$$COV_{(Z1/Z2)} = 2\sum\sum(a_{j(Z1)})(a_{k(Z2)})D_{jk} \qquad \text{Eq. 5}$$

where the summation is over all pairs of loci j and k, and D_{jk} is the linkage *disequilibrium* and determines what fraction of the allelic covariance contributes to the overall trait covariance (note that Djk depends not only on the recombination distance between loci, but also on the crossing scheme imposed in the QTL study). This allows more of the additive covariation in a trait to be explained, so that the total covariance expected through both linkage and pleiotropy is:

$$COV_{Total\ (Z1/Z2)} = 2pq\sum(a_{i(Z1)})(a_{i(Z2)}) + 2\sum\sum(a_{j(Z1)})(a_{k(Z2)})D_{jk} \qquad \text{Eq. 6}$$

where the first term is the covariance due to pleiotropy following Equation 2, and the second term is that due to linkage disequilibrium. While this equation permits a relatively simple inclusion of linked QTL in the estimation of r_Q, many studies did not report point estimates for the QTL position, presenting instead only the confidence limits. Thus it was not possible to determine the distance between neighboring QTL making the calculation of D_{jk} impossible.

Negative Pleiotropy and Tradeoffs:

Negative pleiotropy (trade-offs) were found in about 25% of trait pairs. In most of these cases, negative correlations were the result. However, in 54 trait pairs, negative pleiotropy of one or more QTL was obscured by positive pleiotropy at other loci.

Surprisingly, these positively correlated trait pairs differed from the negatively correlated pairs NOT by having fewer negatively pleiotropic QTL, but rather in having MORE positively pleiotropic QTL in their genetic architecture. This may help to explain the difference observed between positive and negative correlations in the relationship between r_G and r_Q above. More importantly though, it provides direct evidence for an important hypothesis regarding the nature of genetically based trade-offs. Several authors have proposed that such masking of tradeoffs may occur (Fry, 1993; Houle, 1991; de Jong and van Noordwijk, 1992,), making it difficult to detect a negative correlation between traits involved in a tradeoff. Indeed, some authors have questioned the relevance of negative genetic correlations (Fry, 1993) noting that theory of fitness tradeoffs does not preclude the presence of genotypes with low values of both traits – the presence of such individuals would tend to induce positive correlations – rather theory predicts the absence of individuals with high values of both traits (MacKenzie, 1996).

I have made no attempt here to identify trait pairs that should be involved in trade-offs, since the majority of studies in the data set do not explicitly address tradeoffs. There is thus no *a priori* reason to expect negative correlations between many of the trait pairs in the data set. In fact, simply finding negative pleiotropic loci hidden with positive correlations among these studies implies these mechanisms of acquisition and allocation of resources exist. However, the relative rarity of negative pleiotropy (25% of trait pairs) should not be taken as evidence that tradeoffs are uncommon in nature (see Roff, 2002).

Chapter Three:

Heritable Variation and Genetic Correlation Among Fitness Related Traits in *A. barbata*

Introduction:

In order for a population to respond to selection, there must be a heritable genetic basis to variation in fitness and related traits among individuals. By growing individuals that are native to different habitats in a common environment it is possible to assign a genetic basis to any observed phenotypic differences. In this study I have grown two distinct ecotypes of *Avena barbata* , and their recombinant progeny, in a common greenhouse environment in order to characterize the genetic differences that are believed to have been important in the adaptation of the ecotypes to their respective habitats. Specifically, this experiment was conducted to collect the requisite phenotypic data to be used in QTL mapping the factors underlying trait divergence (Chapter Five). However, as a prelude to the analyses at the molecular level it is possible to not only quantify the amount of variation released in the cross between ecotypes, a useful indicator of QTL segregation, but also identify any constraints to this release manifest as genetic correlations among traits.

The amount of heritable variation that can be released by crossing species or ecotypes has significant bearing on the evolutionary fate of a population in a heterogeneous environment. If the individuals being crossed are locally adapted to different habitats, the collection of recombinant genotypes in their progeny will most likely be maladaptive in each parental environment (Nagy, 1997); In addition, any

coadapted gene complexes in the parents would have been broken up by recombination. Thus the mean fitness of the progeny would be lower than the better parent in each environment.

Nevertheless, depending on the allelic composition of the parental genotypes it is possible for individual recombinant hybrid progeny to have above or below average fitness in multiple environments. For example, if the parental genotypes predominantly carry alleles having effects of the same sign at the loci controlling the trait of interest (i.e. all the "plus" alleles in the high line) and this allelic combination works well, the range of genotypes and phenotypes in the progeny would be intermediate to the existing parental forms and would likely be displaced from the phenotypic optima of each environment. However, if the range of phenotypes observed in the progeny greatly exceed that of the parents, termed trangressive segregation, it is likely that each parent carries both plus and minus alleles for the trait (de Vincente and Tanksley, 1993; Rieseberg *et al.*, 1999). It is these trangressive genotypes that are of interest because they may be able to outperform the parental types within the native habitats and perhaps be able to flourish in a novel environment (Rieseberg *et al.*, 1999; Schwarzbach *et al.*, 2001).

The degree to which different traits are controlled by the same set of genes, estimated as the genetic correlation, can dictate how the overall phenotype can respond to selection (Falconer and MacKay, 1996). Genetically correlated traits cannot evolve independently from one another; selection on one trait causes a change in the correlated trait, which itself could be selectively neutral (Lande and Arnold, 1983; Arnold, 1992). Thus the genetic relationships among traits are a guide to what type of phenotyes are possible in natural populations. Furthermore, since recombination will break up the

contribution of linkage disequilibrium to genetic correlations (Lynch and Walsh, 1998), studying genetic correlations in a line cross population provides an opportunity to focus on the genetic effects of loci that constitute more permanent evolutionary constraints (i.e. pleiotropy and tightly linked loci), for it is these loci that are presumably important in shaping the evolutionary trajectory of diverging populations .

In this study I explore the genetic relationships among fitness related traits that may have had a role in the adaptation of the ecotypes to their local environments. Specifically, I ask (1) Do the parental ecotypes differ in fitness and fitness related traits in a greenhouse environment? (2) Is there evidence of within parental ecotype variation? (3) is there heritable variation for these traits in their line cross progeny? (4) What are the genetic correlations among the fitness traits in the line cross population?, and is there evidence for the existence of tradeoffs? (5) Are there any nonadditive genetic effects for fitness in the greenhouse?

Materials and Methods:

Common garden greenhouse experiment:

The experimental design was chosen to make efficient use of the available greenhouse space and consisted of three randomized complete blocks set up in one room of the Dalhousie University greenhouse. Within blocks, two seeds from each F6 recombinant inbred were sown in an individual cell, labeled with line number, within a 72 cell plastic tray filled with a 1:1:1 mixture of peat, black earth and sand. Two replicates of 15 mesic families, containing two seeds per family per replicate (60 seeds in total), were sown in a similar manner in the same trays. Likewise, three replicates of six xeric families, two seeds per family per replicate (36 seeds in total) were also allocated to the plastic trays. Trays were watered and placed in a growth chamber set to provide 12 hr light, at 20 C, and 12 hr dark, at 15 C, per day. Each block was examined daily during the first week for signs of shoot growth to assess germination time.

At 20 days the height of each seedling was measured, and the pair of seedlings from each cell were transplanted to five inch standard pots filled with 1:1:1 mixture of peat, black earth and sand and placed in the Dalhousie University greenhouse. Lighting in the greenhouse was generated by high pressure sodium bulbs and metal halide bulbs set to illuminate the pots for 12 hr form approximetaly 7 a.m. to 7 p.m. After 60 days one seedling from each pot was harvested and dried (at 55 C for 14 days) to assess early biomass accumulation. The remaining plants were watered every three and fertilized 14 days respectively and allowed to grow to senescence. Each plant was examined daily for the emergence of the first spikelet to determine days to first flower.

At the end of May 2002 the experiment was harvested. Each plant was first measured for maximum height achieved, tiller number, and total spikelet number, then the tillers were separated from the leaves and each was bagged and dried at 55 C for two weeks. Leaves and tillers were weighed separately to assess allocation to vegetative growth and reproduction respectively and summed to give an estimate of total mass.

Statistical Analysis:

Genetic differences between the parental ecotypes for all the traits were examined using multivariate analysis of variance (multivariate GLM function in SPSS version 11.5) with *ecotype* (Fixed), *block* (Random), and *family within ecotype* (Random) as main factors and *block by ecotype* and *block by family within ecotype* as interaction factors. Within mesic and xeric ecotypes, analysis was carried out to determine if there was genetic variation among the parental families. Univariate analysis of variance was used to partition the variation due to *family* and *block* and the *family by block* interaction.

Similar statistical analysis was carried out on the F6 generation in order to quantify the heritable variation in the measured traits. Univariate analysis of variance was performed for each trait separately with *line* (Random), and *block* (Random) as main factors. Broadsense heritabilities for each trait were calculated as the fraction of total variation attributed to differences among the F6 lines (i.e. the intraclass correlation coefficient);. since each F6 line is genetically uniform, the proportion of the total variation attributed to among lines is entirely genetic. Variance components for the main factors were calculated by using the restricted maximum likelihood (REML) estimation

function of SPSS. In cases where the REML method failed to find a maximum, variance components were calculated by equating the mean squares for *line* and *error* to their expectations (Falconer and MacKay, 1996).

Trait means from the parental and inbred generations were used to test the adequacy of an additive model in explaining the observed trait variation for each quantitative trait. A weighted least squares procedure (Kearsey and Pooni, 1996; Kearsey *et al.*, 2003) was used to estimate the population mean (m), the additive effect (a) and an epistatic effect (aa). In this case, the method involved regressing a 3x1 vector of mean trait values (**Y**), consisting of the mesic, xeric and the F6 mean, on a 3x3 design matrix (**C**) consisting of indicator variables specific to the coefficients of the three parameters of interest (Figure 3.1). In order to account for the difference in estimation accuracy for each generation mean (i.e. sample size of F6 RILs much larger than the parents) the design matrix is weighted (**W**) by the inverse of the corresponding standard error of each mean. The vector of parameter values (**M**) is obtained by solving the least squares equation: $\mathbf{M} = (\mathbf{C'WC})^{-1} \cdot (\mathbf{C'WY})$, where $\mathbf{C'}$ is the transpose of **C**, and $(\mathbf{C'WC})^{-1}$ indicates the inverse of the matrix $\mathbf{C'WC}$. This method is similar in principal to the linear contrasts employed by Mather and Jinks (1982) to estimate basic quantitative genetic parameters (e.g. compare F6 mean against midparent as a test for epistasis), but since the WLS takes into account the uncertainty of the estimate of each mean, the parameter estimates are unbiased (Kearsey and Pooni, 1996).

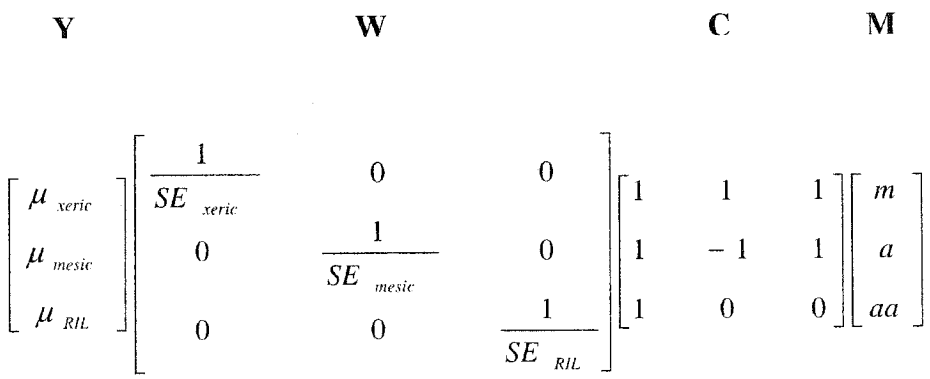


Figure 3.1. Components of the weighted least squares parameter estimation.

Relationships among traits were quantified by calculating estimates genetic correlation for all pairwise combinations of traits in the F6 lines. Genetic correlations were estimated by two methods. First, a simple approximation to the genetic correlations was estimated by taking the product moment correlation of the F6 line means for all traits. Statistical significance and confidence intervals for these correlations were calculated using the methods described in Sokal and Rohlf (1995). Second, a more thorough approach was used to estimate genetic correlations, consisting of partitioning the sums of squares and cross products (SSCP) attributed to among line differences, dividing by the degrees of freedom, and equating these values to their expected mean square and mean cross product respectively (Lynch and Walsh, 1998). The estimates of among line variances and covariances for the traits were then incorporated into the standard equation for determining correlation.

In order to assess the functional intergration among traits, principal component analysis was carried out to identify the main axes of variation in the F6 lines. Components were identified using the Factor function in SPSS on the mean trait values for each line. Factors having eigenvalues above one were retained in the model and the Varimax orthogonal rotation procedure was used to generate a component loading matrix. The trait means for the xeric and mesic families were then included in the principal components analysis in order to identify transgressive segregation in the F6 lines. The Factor function with the Varimax rotation was used as above to generate the component eigenvalues and loading matrix. Factor scores from each component were retained as variables using the default regression method of SPSS. The factor scores for each mesic and xeric family were then plotted along with the scores from each F6 line.

Results:

Genetic differences between and within Xeric and Mesic ecotypes:

Examining the means of the parental ecotypes reveals strong genetic differentiation for most of the traits (Table 3.1). Multivariate ANOVA analysis shows significant differences between the parental ecotypes (Wilk's lambda = 0.026, $p<0.0001$) while accounting for a significant block effect (Wilk's lambda = 0.118, $p<0.0001$) and significant ecotype by block interaction (Wilk's lambda = 0.553, $p= 0.026$). Tests of the individual traits indicate the parental ecotypes differ for all traits except for dry weight at 60 days (Table 3.2a). On average, the mesic parents flowered earlier, allocated more biomass to reproductive structures (i.e.tillers), producing near double the number of spikelets than the xeric parents. In contrast the xeric parents were on average larger, devoting more resources to vegetative growth as opposed to reproductive structures, and flowered over a month later than the mesics. The xerics produced fewer, albeit taller, tillers.

There were significant differences among families within the two parental ecotypes (Wilk's lambda = 0.017, $p<0.001$). Within ecotypes the amount of variation attributed to among family differences (i.e. broadsense heritiability (H^2)) was calculated (Table 3.1). Among xeric families there was significant variation found for flowering time ($H^2 = .277$) and dry mass of leaves ($H^2 = 0.205$) with the remaining traits, including fitness, generally having heritabilities less than 10-15% and lower confidence limits including zero. Similarly there was significant variation for flowering time among mesic

Table 3.1. Means, ranges, and variance components among traits in the greenhouse for xerics, mesics and the F6 generation. Variance components (and coefficients of variation, CV) are given for genetic, block, and environmental sources. (GM= germination time, HT20 = height at 20 days, DM60 = dry mass at 60 days, FL = flowering time, TN = tiller number, SN= spikelet number, TM= tiller mass, LM = leaf mass, TOTM = total mass.)

	GM	HT20	DM60	FL	MH	TN	SN	TM	LM	TOTM
Means										
Xeric	5.190	10.770	0.520	155.700	149.460	5.320	275.120	6.160	10.410	16.360
Mesic	5.120	13.650	0.500	120.100	135.320	15.260	504.430	9.150	5.010	14.190
F6 lines	5.270	12.220	0.390	135.810	137.600	11.830	425.080	7.790	7.380	15.170
Range										
Xeric	(4.0-6.0)	(8.0-12.8)	(0.05-1.71)	(134-177)	(32.5-191.0)	(0-15)	(0-742)	(0-16.8)	(4.23-18.10)	(9.17-34.89)
Mesic	(4.5-6.0)	(10.8-15.4)	(0.04-2.11)	(101-142)	(117.5-169.0)	(9-24)	(306-792)	(5.22-13.22)	(2.35-10.84)	(7.79-23.22)
F6 lines	(4.0-6.5)	(4.15-16.40)	(0.03-1.77)	(95-182)	(42.5-204.0)	(2-70)	(14-1038)	(.83-18.11)	(1.64-15.02)	(8.12-27.49)
Xeric families	Variance Components									
Var gen.	0.004	0.084	0.001	39.850	112.769	0.649	868.029	0.213	1.257	0.844
(CV gen)	0.012	0.027	0.061	0.041	0.071	0.151	0.107	0.075	0.108	0.056
Var block	0.022	0.000	0.143	69.890	121.801	3.516	7026.603	7.059	2.983	15.175
(CV block)	0.029	0.000	0.727	0.054	0.074	0.352	0.305	0.431	0.166	0.238
Var env.	0.276	0.962	0.076	103.030	558.604	6.358	16815.195	6.100	4.909	13.570
(CV env)	0.101	0.091	0.530	0.065	0.158	0.474	0.471	0.401	0.213	0.225
H^2 (Xeric)	0.014	0.080	0.013	0.279	0.168	0.093	0.049	0.034	0.204	0.059
95% conf.	-0.076, 0.381	-0.047, 0.520	-0.078, 0.371	0.045, 0.732	-0.012, 0.626	-0.042, 0.535	-0.052, 0.498	-0.050, 0.505	0.010, 0.674	-0.058, 0.475

Table 3.1 continued.

	GM	HT20	DM60	FL	MH	TN	SN	TM	LM	TOTM
Mesic families										
Var gen.	0.000	0.184	0.000	15.135	5.633	0.795	748.744	0.000	0.105	0.000
(CV gen)	0.000	0.031	0.000	0.032	0.018	0.058	0.054	0.000	0.065	0.000
Var block	0.010	0.138	0.090	16.278	26.849	0.708	1528.757	2.578	1.104	7.048
(CV block)	0.019	0.027	0.601	0.034	0.038	0.055	0.078	0.175	0.210	0.187
Var env.	0.082	0.565	0.086	43.531	98.892	5.243	6235.723	1.472	1.651	4.174
(CV env)	0.056	0.055	0.585	0.055	0.073	0.150	0.157	0.133	0.257	0.144
H^2 (Mesic)	0.000	0.246	0.000	0.258	0.054	0.132	0.107	0.000	0.060	0.000
95% conf.	-0.106, 0.177	0.056, 0.513	-0.107, 0.174	0.070, 0.531	-0.065, 0.287	-0.019, 0.387	-0.031, 0.362	-0.102, 0.188	-0.062, 0.294	-0.095, 0.208
F6 Lines										
Var gen.	0.046	0.706	0.003	81.739	233.690	15.813	15539.679	3.468	2.797	2.035
(CV gen)	0.041	0.069	0.139	0.067	0.111	0.336	0.293	0.239	0.227	0.094
Var block	0.041	0.110	0.053	21.205	69.664	0.306	1088.451	2.637	1.293	5.778
(CV block)	0.039	0.027	0.590	0.034	0.061	0.047	0.078	0.208	0.154	0.158
Var env.	0.257	1.103	0.031	57.960	169.773	13.159	10777.538	2.647	2.016	4.091
(CV env)	0.096	0.086	0.453	0.056	0.095	0.307	0.244	0.209	0.192	0.133
H^2 (F6)	0.151	0.390	0.086	0.585	0.579	0.546	0.590	0.567	0.581	0.332
95% conf.	0.246, 0.064	0.478, 0.298	-0.002, 0.174	0.659, 0.511	0.652, 0.502	0.621, 0.463	0.658, 0.510	0.641, 0.489	0.652, 0.502	0.421, 0.237

families ($H^2 = 0.260$), as well as height at 20 days ($H^2 = 0.238$). No heritable variation was detected for any of the remaining traits measured among mesic families (Table 3.1).

Broadsense heritabilities in the F6 lines:

Significant differences among the F6 lines (Table 3.2b) were detected using univariate ANOVA, while accounting for a significant block effect. For most traits, the among line component of variation accounted for 50%-60% of the total variation (Table 3.1). The exceptions were germination time, height at 20 days, and dry mass at 60 days, which had markedly lower heritabilities. The other notable exception is total mass, which has a hertitability of 33% despite the fact that its component parts (tiller mass and leaf mass) have heritabilities greater than 55%.

Adequacy of an additive model in explaining variation among generations:

Results of the weighted least squares analysis show that an additive model explains the variation for flowering time, height at 20 days, tiller mass, leaf mass, and total mass (Table 3.3) indicating that the F6 generation mean does not differ from the mean of the midparent. In contrast, departure from additivity was detected in the reproductive output traits, seed number and tiller number with significant epistatic effects causing a decrease of 38.2 spikelets and 1.5 tillers in the parental ecotypes. Similarly, significant epistatic effects were detected for germination time, with a trend for the parental types to germinate earlier than the F6 lines. The only epistatic effect that increased the trait value in the parental ecotypes was for maximum height, with the xeric

Table 3.2a) Differences among mesic and xeric ecotypes tested using nested multivariate ANOVA. Table shows multivariate between factor F ratio statistics and degrees of freedom. Items in bold are significant at $p<0.05$. Trait abreviations are the same as in Table 3.1.

Source	df	GM	HT20	DM60	FT	MH	TN	SN	TM	LM	TOTM
Ecotype	1	**11.59**	**314.20**	1.22	**472.68**	**16.51**	**435.96**	**159.04**	**69.71**	**232.98**	**10.90**
Family within Ecotype	19	0.89	**2.92**	**1.90**	**3.40**	1.20	**2.22**	**2.08**	0.87	**2.07**	0.60
Block	2	**15.54**	2.90	**50.28**	**19.19**	**7.72**	**4.87**	**8.35**	**26.83**	**12.24**	**26.06**
Ecotype by Block	2	**3.51**	2.02	**3.93**	**3.27**	1.40	**5.05**	**3.94**	**6.44**	0.02	2.87
(Family within Ecotype) by Block	37	1.31	1.23	1.50	1.21	0.41	0.92	1.14	0.65	0.91	0.40
Error	54										

Table 3.2b) Differences among F6 lines tested using multivariate ANOVA. Table shows the multivariate between factor F ratio statistics and degrees of freedom. Items in bold are significant at $p<0.05$. Trait abreviations are the same as in Table 3.1.

Source	df	GM	HT20	DM60	FT	MH	TN	SN	TM	LM	TOTM
Line	187	**1.49**	**3.44**	**1.38**	**4.77**	**4.89**	**4.25**	**4.92**	**4.40**	**4.80**	**2.12**
Block	2	**29.41**	**16.49**	**309.29**	**56.20**	**71.59**	**4.36**	**15.78**	**162.38**	**110.34**	**241.3**
Error	334										

Table 3.3. Weighted Least Squares tests for departures from an additive relationship among generation means. (m = estimate of trait mean; a = estimate of the additive effect of xeric genotype; aa = estimate of the epistatic effect due to having alleles of the same parent; SE = standard error). Estimates shown in bold are significant at $p<0.05$.

Trait	m	SE(m)	[a]	SE(a)	[aa]	SE(aa)
GM	**5.27**	0.02	0.03	0.04	**-0.11**	0.04
HT20	**12.22**	0.07	**-1.44**	0.08	-0.00	0.31
DM60	**0.39**	0.08	0.01	0.13	0.12	0.16
FL	**135.81**	0.83	**17.79**	0.83	2.08	1.17
MH	**137.60**	1.11	**7.07**	1.12	**4.79**	1.58
TN	**11.83**	0.10	**-4.97**	0.08	**-1.53**	0.41
SN	**424.70**	2.11	**-113.50**	2.63	**-38.20**	3.37
TM	**7.79**	0.38	**-1.49**	0.40	-0.13	0.56
LM	**7.37**	0.37	**2.69**	0.37	0.34	0.52
TOTM	**15.16**	0.36	**1.08**	0.50	0.11	0.62

and mesic genotypes growing approximately 4.7 cm taller than expected under additivity. Lastly, no genetic parameter, except the population mean, could be fit to dry mass at 60 days.

Genetic correlations among traits in the F6 generation:

Genetic correlation estimates for all pairwise combinations of traits in the F6 lines were calculated by variance components and by the line means method (Table 3.4). Several patterns among the correlations are evident. First, flowering time and overall allocation of resources to growth seem to be strongly associated with reproductive output in the greenhouse environment. Individuals that flowered earlier tended to allocate more biomass to reproductive structures (i.e. tillers) and produced more spikelets. In contrast, the later flowering individuals tended to produce more vegetative biomass and fewer tillers resulting in lower overall fitness. Likewise, the negative relationship ($r = -0.55$) between leaf mass and tiller mass also indicate a trade-off in resource allocation.

Second, the early growth traits do not correlate strongly with the later growth traits, including spikelet production (Table 3.4). Significant genetic correlations were detected in some instances, predominantly involving height at 20 days, however the variance in the later growth traits explained by these early traits was generally less than 5% (i.e. $R^2 < 0.05$). This indicates that germination and seedling establishment (in the absence of competition) do not affect fitness in the greenhouse environment. An exception to this conclusion may be the relationship between dry mass at 60 days and total mass.

Table 3.4. Genetic correlations among traits measured in the F6 lines. Family means correlation estimates are above the diagonal. Variance component genetic correlation estimates are below the diagonal. Estimates shown in bold are significant at $p<0.05$ after a Bonferroni correction for multiple testing.

	GM	HT20	DM60	FL	MH	TN	SN	TM	LM	TOTM
GM		0.177	-0.014	0.009	-0.087	0.007	0.004	0.055	0.036	0.097
HT20	0.085		0.197	0.190	0.179	-0.159	-0.235	-0.104	0.199	0.084
DM60	-0.061	**0.319**		-0.153	0.106	0.091	0.104	0.223	0.139	**0.388**
FL	0.009	0.221	-0.112		-0.225	**-0.747**	**-0.820**	**-0.807**	**0.639**	**-0.256**
MH	-0.184	0.202	**0.302**	**-0.310**		-0.114	0.211	**0.491**	**-0.386**	0.159
TN	0.091	**-0.263**	-0.070	**-0.839**	-0.101		**0.845**	**0.660**	**-0.447**	**0.285**
SN	0.056	**-0.355**	-0.058	**-0.916**	0.226	**0.899**		**0.851**	**-0.619**	**0.325**
TM	0.087	-0.199	**0.299**	**-0.919**	**0.548**	**0.676**	**0.854**		**-0.558**	**0.553**
LM	-0.044	**0.252**	0.167	**0.803**	**-0.432**	**-0.610**	**-0.806**	**-0.710**		**0.384**
TOTM	0.064	0.041	**0.616**	**-0.258**	0.213	0.166	0.166	**0.484**	**0.273**	

Factor analysis of the major axes of variation among traits in the F6 lines and the parental ecotypes:

Principal components analysis on F6 line means identified four main components that explain 82% of the variation among lines means in the traits examined (Table 3.5). The dominant principal component accounted for 41% of the total variation, with the remaining components explaining 18%, 13% and 10% of the total variation respectively. Analysis of the Varimax rotated loading matrix (Table 3.5) shows the dominant principal component is highly positively correlated with spikelet number (fitness), tiller number, and tiller mass. This fitness related component was also found to be highly negatively correlated with flowering time and leaf mass. The second component was correlated to dry weight at 60 days, leaf mass and total mass, but was not related to spikelet number. Similarly, the third component was correlated to the two height traits, and perhaps weakly to tiller and leaf mass, but was also not correlated to fitness. The last component was only correlated to the two earliest growth traits, germination time and height at 20 days.

The means from the parental families were analyzed along with the F6 lines in order to identify transgressive phenotypes in the RILs. The addition of the parental data did not change the number of identified components or the structure of the rotated factor loading matrix. Ten F6 lines were found to exceed the better parent, in this case the mesic, for PC1. In particular, F6 lines 16, 61, and 33 had factor scores on PC1 that were

Table 3.5. Principal components analysis on mean trait values in the F6 lines. Varimax rotated component loadings above 0.50 are shown in bold.

Trait	*Component*			
	1	2	3	4
Eigenvalue	4.08	1.78	1.31	1.06
GM	0.074	-0.039	-0.122	**0.922**
HT20	-0.306	0.255	0.428	**0.521**
DM60	0.053	**0.712**	0.163	-0.029
MH	0.165	0.073	**0.922**	-0.064
FL	**-0.908**	-0.064	-0.118	0.045
TN	**0.877**	0.114	-0.248	0.018
SN	**0.953**	0.091	0.047	-0.040
TM	**0.863**	0.295	0.328	0.018
LM	**-0.691**	**0.543**	-0.397	0.068
TOTM	0.267	**0.874**	-0.034	0.088

double the magnitude of the best (i.e. most fit) mesic family. The total range of factor scores among the F6 generation for PC1 encompassed nearly the entire range of parental scores, with some lines doing as poorly as the least fit xeric family (Figure 3.2a). There were no lines that exceeded the lower bound marked by the least fit xeric family.

The range of factor scores among the F6 lines for PC2 completetly encompassed the range of the parental families indicating positive and negative transgressive segregation (Figure 3.2b). Generally, the F6 lines that had high scores on PC1 were not the same as those for PC2. The notable exceptions were F6 lines 33 and 61 which had higher factor scores than the better parent on each axis of variation.

There was considerable overlap in factor scores for PC3 among the mesic and xeric families to the point where it would appear that the parents do not differ in scores along this axis of variation (Figure 3.2c). There was however a wide range of factor scores among the F6 lines. Similarly the range of factor scores among the F6 lines along PC4 spanned the entire range of the parental families indicating positive and negative transgression (Figure 3.2d). Unlike the pattern observed for the first to components, the F6 lines that outperformed the better parent on PC3 and PC4 were not the same.

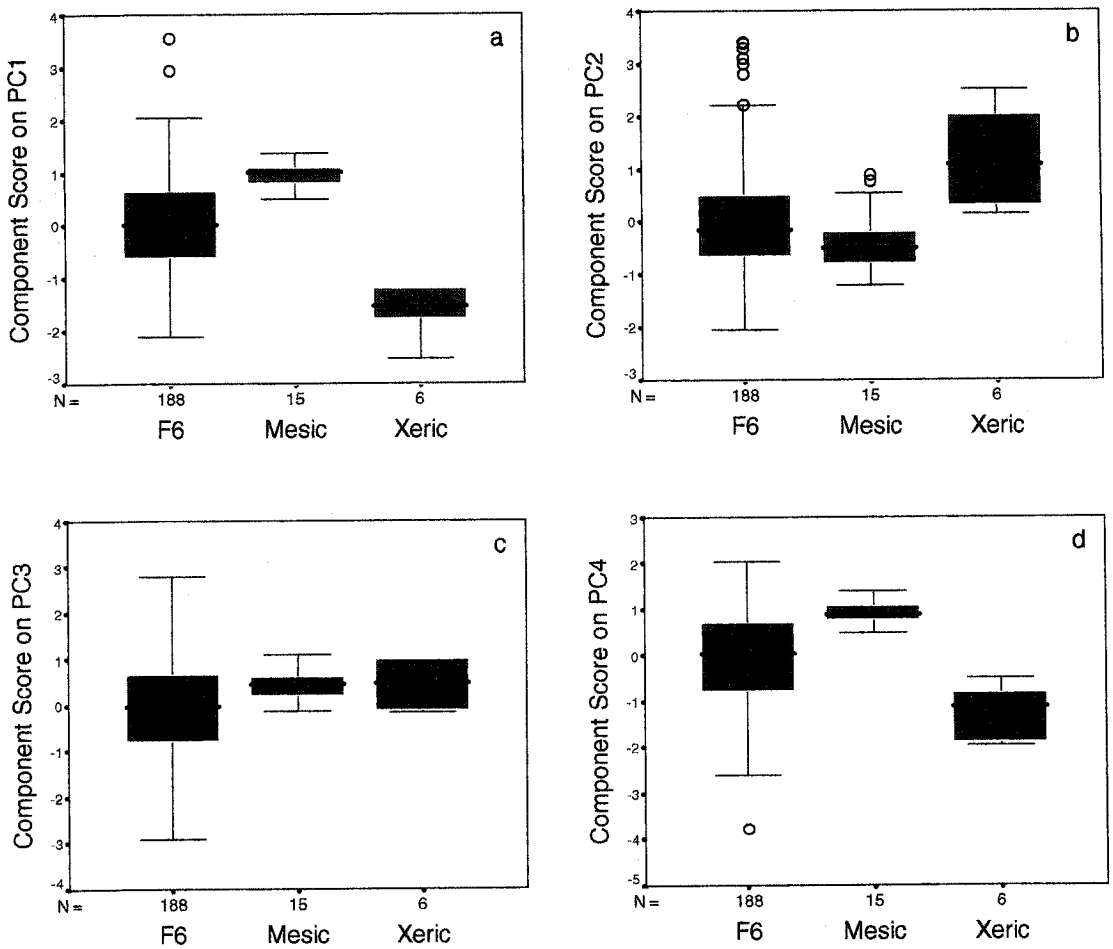


Figure 3.2. Boxplots of component scores along the major axes of variation in the greenhouse. a) PC1, b) PC2, c) PC3, d) PC4.

Discussion:

While the primary goal of this chapter was to collect trait data for use in QTL mapping, it is possible to use the common garden experiment to address several questions about the putative origins of the parental ecotypes and their adaptive divergence.

Genetic differentiation of the parental ecotypes:

If the xeric and mesic ecotypes arose after *Avena barbata* was introduced to California (Clegg and Allard, 1972), then they should exhibit genetic differences in the traits that were important to their adaptive divergence. The parental ecotypes were found to differ genetically for nearly all of the traits examined, including morphological and fitness related traits (Table 3.1). The mesic ecotype was the most fit in the greenhouse environment, flowering a month earlier, allocating more biomass to reproductive structures and producing on average double the number of spikelets than the xeric ecotype. This was perhaps not surprising given that the well watered greenhouse environment certainly represents a more mesic-like habitat. However, the same trend for the mesic genotype to flower early and have higher reproductive output was also observed over contrasting water and nutrient environments in a greenhouse study conducted by Johansen (2004). The xeric families were on average larger than the mesics, producing double the leaf mass as the mesic families, despite delayed flowering

and production of fewer tillers. A similar result was also observed by Johansen (2004) however in that case the difference among ecotypes was less pronounced.

Thus, there appears to be fundamental differences in life histories between the two ecotypes in relation to both reproductive timing and allocation of available resources to growth or reproduction. The tendency for the xeric ecotype to devote more resources to vegetative growth, and put off reproducing until later in the growing season, may be a mechanism for drought tolerance. This seems reasonable given that it has been shown that the xeric ecotype also has a higher root mass ratio and devotes more root mass to deeper roots, presumably mechanisms for increased water acquisition (Latta *et al.*, 2004). In addition, it has been shown that delayed flowering times are associated with greater water use efficiencies, which can aide in avoiding dehydration (MacKay *et al.*, 2003).

The mesic ecotype in contrast seems to increase its fitness through "drought escape" (MacKay *et al.*, 2003) by flowering early while water resources are plentiful and more mass can be allocated to reproductive structures. However, a caveat to this interpretation is the lack of competition among individuals in the experimental design. It has been demonstrated that the mesic ecotype is a superior competitor to both the xeric ecotype and some F3 generation hybrids (Latta *et al.*, 2004). Thus, whether early flowering or competitive exclusion is the major determinant of mesic fitness is unknown. This question will be addressed in future field experiments.

Variation within the xeric and mesic ecotypes:

Because *A. barbata* is predominantly selfing (Clegg and Allard, 1972), and therefore expected to be completely homozygous, it can be argued that the two predominant ecotypes represent naturally occurring inbred lines, fixed for the combination of alleles responsible for the adaptation to the respective local environments (Clegg and Allard 1972; Hamrick and Allard 1972). If this is the case, then there should be no genetic variation among individuals/families *within* either of the ecotypes and considerable variation *among* the inbred line progeny, the latter reflecting the genetic differentiation of the parental genotypes. I tested this assumption by partitioning the genetic variation within ecotypes and estimating the broadsense heritability for each trait. Within xeric and mesic ecotypes there was no significant genetic variation for reproductive output (number of spikelets and tillers) and most of the early and later growth traits (Table 3.1).

The lack of variation within parental ecotypes has significant bearing on the expected power of this system to detect the individual quantitative trait loci underlying each trait since crossing homozygous individuals should maximize the possible linkage disequilibrium among the QTL and molecular marker loci (Chapter Five). In effect, it should be possible to achieve experimental power similar to that observed in studies crossing highly inbred laboratory strains (i.e. *Arabidopsis* or *Drosophila*) but have results extendable to the natural populations of *A. barbata*. This latter point is also due in large part to allowing the inbred lines to self fertilize unaided, thus eliminating complicating factors that may arise due to forcing a non-natural mating system (i.e. sib mating in mice,

selfing an outcrossing plant) throughout the propagation of the recombinant inbred population.

Release of genetic variation and transgressive segregation in the RILs:

In the F6 generation, upwards of 55-60 % of the variation in fitness and related traits such as flowering time and tiller mass could be attributed to genetic differences among the recombinant inbred lines (Table 3.1). This substantial release of variation from divergent, but otherwise genetically uniform, ecotypes may implicate transient outcrossing as a major mechanism in initiating adaptive population divergence in *A. barbata*. While outcrossing events would be extremely rare, selection on the pool of variation coupled with the high self fertilization rate could result in the fixation of beneficial genotypes very quickly.

The range of genotypes that can be produced from an outcrossing or recombination event will depend on the allelic composition of the parental genomes involved. In this study substantial transgressive segregation was observed for virtually the entire suite of traits, with the F6 lines exhibiting phenotypes that were extreme in relation to the parental families. This is demonstrated both by the range of F6 individuals exceeding the mesics and xerics for fitness,flowering time, mass allocation, and perhaps more conclusively by the wide range of line mean principal component scores that summarize fitness and growth (i.e. PC1 and PC2 respectively) in the F6 generation. For example, there were F6 lines that outperformed the most fit mesic individuals by flowering nearly a week earlier and producing nearly 250 more spikelets. Other F6 lines did as poorly as the least fit xeric individuals, with some F6's flowering 5 days later than

the latest flowering xeric. Likewise, several F6 lines had factor scores from the first principal component that exceeded the most fit mesic families, while others scored as low as the xeric families. Interestingly, F6 lines 33 and 61 had component scores that were higher than the better parent on PC1 (i.e. mesic) and on PC2 (i.e. xeric) suggesting they have received favourable alleles from both parents. Thus it appears that complementary genes are responsible, at least in part, for the transgressive segregation in the F6 generation. The possible role of epistasis is discussed below.

Clearly it is possible to produce genotypes in the progeny, via recombining existing alleles, that have high fitness in a novel environment despite the mean of the lines being generally intermediate to the parental types. In this study, several of the F6 lines exhibiting extreme phenotypes could certainly outperform the parental forms and thus be selected for in the greenhouse environment. If a recombination/hybridization event were to take place in a natural population, where mesic and xeric genotypes co-occur, some of the extreme phenotypes could potentially thrive in a new habit that may be very different from, but not necessarily intermediate to, the parental habitats (Schwarzbach *et al.*, 2001). From an ecological point of view this could lead to niche separation and perhaps reproductive isolation between the new hybrid and the parental forms, provided the parental genotypes have lower fitness in the new niche (Schwarzbach *et al.*, 2001). From an evolutionary point of view the production of extreme phenotypes via recombination may provide a simple means to allow a population to traverse a fitness valley on an adaptive landscape (Phillips *et al.*, 2000). Either way, transgressive segregation is an "attractive mechanism" for population divergence (Rieseberg *et al.*, 2003).

Genetic correlations among fitness related traits and constraints to evolutionary change:

Regardless how much genetic variation exists for a given trait, and the range of possible phenotypes produced in a line cross, it is unlikely any one trait can evolve independently due to the existence of genetic correlations with other traits (Mitchell-Olds, 1996). In this study, strong genetic correlations in excess of -0.80 were found among reproductive output (spikelet number, tiller number, tiller mass) and flowering time (Table 3.4). This is perhaps not unexpected in a greenhouse environment where resources are plentiful and there is no benefit to later flowering. Therefore, despite the production of transgressive phenotypes for fitness in the RIL population, there are obviously constraints to what types of recombinant progeny are possible (i.e. a late flowering but highly fecund RIL is unlikely). Taken together, fitness, flowering time and the mass allocation traits were found to vary on one major axis of variation that could account for over 40 % of the variation among all traits in the greenhouse. This is evidence that these traits not only share some genetic control but are also functionally integrated. Following from Chapter Two, it is expected that co-localizing QTL must underlie this highly correlated group of traits. This will be addressed in Chapter Five.

In comparison, traits such as germination time, height at 20 days, dry mass at 60 days and maximum height are not as tightly genetically correlated to fitness and do not load heavily on the "fitness" axis of variation (PC1). Rather, they group together on other axes of variation (Table 3.5). For example, PC4 explains variation in the earliest

measured traits, germination time and height at 20 days. Neither trait seems to have any appreciable relationship with fitness, implying seedling establishment is not important in a competition and stress free environment.

The negative genetic correlation between leaf and tiller mass ($r = -0.71$) indicates that there is a tradeoff between growth and reproduction. This suggests that there must be antagonistically pleiotopic QTL underlying the variation in these two traits (Chapter Two). It has been suggested that these types of loci are actually involved in the allocation of some limited resource (e.g. mass or energy) to each trait, thus creating negative trait covariance (Houle, 1991; Mitchell-Olds, 1996). What is particularly interesting in this study is each mass trait by itself is positively genetically correlated to total mass despite the tradeoff. This suggests that there must be loci controlling variation in leaf mass independently of tiller mass, and *vice versa*, allowing each to increase total mass. Houle (1991) suggests that this type of locus may be involved in resource acquisition, which would have the effect of creating positive trait covariance. Overall, the constraint to evolutionary change in growth is manifest both as the tradeoff between the leaf and tiller mass and their combined effect on variation in total mass. The latter constraint is best demonstrated by the lower heritability for total mass ($H^2 = 0.33$) despite leaf and tiller mass having heritabilities greater than 55 %. Obviously the amount of genetic variation for total mass in the F6 generation is dominated by the negative covariance between leaf and tiller mass. This would have the net effect of reducing any possible response to selection on total mass.

Nonadditive genetic effects and coadapted gene complexes:

By comparing the mean performance of the parental ecotypes with their inbred progeny it is possible to observe the sum of the genetic contributions of all the factors affecting the phenotype and determine the major mode of inheritance for each trait. Using a weighted least squares procedure (Kearsey and Pooni, 1996; Kearsey *et al.*, 2003) I have shown that an additive model is sufficient to explain the variation in the observed generational means in flowering time and height at 20 days. In addition nonadditive effects were absent for growth traits including tiller, leaf and total mass (Table 3.4). In these instances the lack of detectable epistatic effects could reflect the absence of interacting genes, however since this method examines the sum of all epistatic effects it is possible some loci may cancel each other out. These "canceling out " epistatic effects are of interest because they represent another source of variation that can be released via recombination and may contribute to the production of transgressive segregants.

In contrast, departures from additivity were detected for spikelet number, tiller number, and maximum height in the greenhouse environment, indicating the presence of interacting genes. This is, of course, assuming the contribution of dominance to non-additive genetic effects is absent in the parental ecotypes and negligible in the F6 generation due to the multiple generation of self fertilization (Lynch and Walsh, 1998). What was surprising was that the epistatic effects on fitness were negative, with the parental ecotypes having fewer spikelets and tillers than expected under additivity. If there really are coadapted gene complexes, than the expectation would be for a fitness

increase to be observed in the parental ecotypes. However, given that the experiment employed a novel environment, there is no *a priori* reason to expect the gene combinations that work well in the field environments to impart any fitness advantage in the greenhouse. This will be addressed further in the field experiment detailed in Chapter Six.

Thus, the results of this portion of the study confirm the presence of epistatic interactions for fitness which, very simply, confirms the polygenic basis of these traits and the presence of putatively co-adapted gene complexes in the parental ecotypes. The among generation analysis presented here was by no means exhaustive; with only three generation means parameter estimation was limited. For a thorough treatment of the implications of dominance effects, along with epistasis, on fitness in *A. barbata* see Johansen (2004).

Chapter Four:

Construction of an Amplified Fragment Length Polymorphism Genetic Linkage Map in *A. barbata*

Introduction:

As a prelude to identifying and localizing the genomic regions controlling fitness related traits, it is necessary to construct a genetic linkage map that spans a considerable portion of an organism's genome. The rationale being if the genome is completely saturated with markers then all the loci having effects on a trait (i.e. QTL) should exhibit linkage to at least one or more markers (Thoday, 1961; Lynch and Walsh, 1998). Prior to the advent of molecular markers based on the polymerase chain reaction, the major limitation to identifying QTL was the availability of an inexpensive source of polymorphic markers, making thorough genetic analysis feasible only in the most economically important, or model species. Currently, several types of molecular markers, namely amplified fragment length polymorphism (AFLP, Vos *et al.*, 1995) and randomly amplified polymorphic DNA (RAPD, Williams *et al.*, 1990), are available that can be customized to any study system as they do not require any prior information about the genome under study. AFLP markers, in particular, have been used extensively to create linkage maps for a range of species for a fraction of the cost and effort of more traditional marker types such as restriction fragment length polymorphisms (RFLPs) and other PCR based markers such as microsatellites (Mueller and Wolfenbarger, 1999).

Aside from marker type an equally important consideration in linkage map construction is the crossing design used to create the mapping population. Most studies

attempting to map the genetic factors that underlie genetic and phenotypic divergence in natural populations rely on using established experimental designs to create an experimental population for analysis (Lynch and Walsh, 1998). In some instances the choice of design is severely limited due hybrid sterility (Lexer *et al.*, 2003), self incompatibility (e.g. *Raphanus raphanistrum*, Conner, 2003) or strong inbreeding depression associated with forced selfing in plants or sib mating in animals (e.g. Tani *et al.*, 2000; Remington and O'Malley, 2000). In this study I have the benefit of using an organism that is amenable to multiple rounds of selfing, an otherwise natural mating habit for wild oats, in order to produce recombinant inbred lines, which should allow for high QTL detection power not usually observed in naturally derived systems. Moreover, I have presented evidence at the phenotypic level (Chapter Three) that the parental ecotypes represent naturally occurring inbred lines. When these ecotypes are crossed, the pool of possible recombinants (informative meioses) and marker-QTL linkage disequilibrium should be maximized.

Currently, molecular maps, constructed using an array of different molecular markers (microsatellites, RAPD, RFLP, etc.), exist for several *Avena* species including varieties of the economically important hexaploid (2n = 6x = 42) oat *Avena sativa* (e.g. O'Donoughue *et al.*, 1995, updated in Wight *et al.*, 2003; Jin *et al.*, 2001) and several diploid (2n = 14) oat species including *Avena atlantica* X *Avena hirtula* (O'Donoughue *et al.* 1992) and *Avena strigosa* X *Avena wiestii* (Rayapati *et al.* 1994; updated in Kremer *et al.* 2001). To my knowledge, no linkage map currently exists for any of the tetraploid *Avena* species. Furthermore, most of the crosses that have been made among oat species were carried out to make mapping populations that segregated variation for specific yield

related traits and disease and pest resistance genes. While it may be possible in the future to compare the results of these existing maps with the *A. barbata* system, for example to identify syntenic chromosomal regions, in order to properly characterize the genetic factors responsible for the variation in the xeric and mesic ecotypes a linkage map specific to their progeny must be created.

In this chapter I make use of AFLP markers to (1) assess the level of genetic polymorphism within the mesic and xeric parental ecotypes, and (2) construct a genetic linkage map of the *A. barbata* genome by genotyping a population of recombinant inbred lines, (3) examine patterns of segregation distortion among mapped markers.

Materials and Methods:

Isolation of DNA from *Avena barbata*:

DNA was isolated from *A. barbata* by first grinding 0.5-1.0 g of fresh leaf tissue, frozen with liquid nitrogen, and dissolving the resulting powder in 5 mL of extraction buffer (10 mM Tris-HCL pH 8.0, 10 mM EDTA pH 8.0, 1% SDS, 14 mM 2-mercaptoethanol) in 15 mL Falcon tubes. Extraction tubes were incubated at 65 C, and vortexed at 5 minute intervals for 30 minutes, to ensure proper cell lysis. Upon cooling, extractions were centrifuged at 12000 rpm for 15 minutes and approximately 1 ml of supernatant was separated from the pellet of cellular debris and placed in 2 mL microcentrifuge tubes. An equal volume of 5 M potassium acetate (pH 7.2) was added and the tubes were quickly inverted several times and immediately placed on ice for 30 minutes. Tubes were centrifuged at 12000 rpm for 15 minutes and the supernatant was collected. An equal volume of isopropanol was added to each extraction and the tubes were mixed thoroughly and incubated at –20 C for 20 minutes to precipitate the DNA. Extractions were centrifuged at 8000 rpm for 5 minutes and the supernatants were discarded. DNA pellets were resuspended in 400 µl of TE (10 mM Tris-HCL pH 8.0, 50 mM EDTA pH 8.0) and precipitated again with an equal volume of isopropanol and a 1/10 volume of 3 M sodium acetate (pH 5.0). Extractions were then centrifuged at 12000 rpm for 20 minutes and the supernatants were discarded. Pelleted DNA was then washed twice with 70% ethanol and allowed to air dry. Each DNA pellet was resuspended with distilled water to a final volume of 150 µl. To remove contaminating RNAs 1 µl of

RNase A solution (Sambrook *et al.*, 1989) was added to the DNA preparations and each was incubated at 37 C for 30 minutes. An equal volume of chloroform was then added and the DNA preparations were vortexed for 30 s and immediately centrifuged at 12000 rpm for 15 minutes. The aqueous phase of each chloroform treated DNA preparation was recovered and was used for all subsequent molecular analyses.

Amplified Fragment Length Polymorphism (AFLP) marker assay:

DNA digestion and adapter ligation:

The generation of AFLP markers was carried out using a modified protocol following Vos *et al.* (1995). First, approximately 500 µg of total genomic DNA was digested for three hours at 37 C with 5 units of *Eco*R I, 7.5 units of *Mse* I, 4 µl 10X One-Phor-All buffer and dH_2O up to a total volume of 40 µl. Once the digest tubes were cooled, 20 ng of the *Eco*R I adapter, 50 ng of the *Mse* I adapter, 5 µl T4 DNA ligase buffer, 0.33 units T4 DNA ligase were added with enough dH_2O to make a total volume of 50 µl. The ligation reactions were carried out at 37 C for three hours followed by an overnight incubation (~ 18 hrs) at 18 C.

AFLP pre-amplification and selective amplification:

One µl of the ligation reaction was used as template DNA in a PCR containing 50 ng of *Eco*R I adapter specific primer with one selective nucleotide, 50 ng *Mse* I adapter specific primer with one selective nucleotide, 1.5 µl 10X PCR buffer,1.2 µl 25 mM

$MgCl_2$, 1.2 µl 2.5 mM dNTPs, 0.2 units *Taq* polymerase and dH_2O up to a total volume of 15 µl. The preamplification temperature profile consisted of 20 cycles of one minute at 94 C, one minute at 56 C and one minute at 72 C. This was immediately followed by a seven minute extension step at 72 C. The preamplification products were diluted 1/10 with dH_2O before being used as template DNA in the selective AFLP amplifications.

Selective AFLP reactions were constructed by adding one µl of diluted preamplification, 20 ng *Eco*R I adapter specific primer with three selective nucleotides (Table 4.1), 50 ng *Mse* I adapter specific primer with three selective nucleotides, 1.0 µl 10X PCR buffer, 0.6 µl 25 mM $MgCl_2$, 0.8 µl 2.5 mM dNTPs, 0.2 units *Taq* polymerase and dH_2O up to a total volume of 10 µl. The reaction was then covered with light mineral oil or Chill-Out wax prior to thermalcycling. The reaction temperature profile was slightly different from the premaplification with 12 cycles of 95 C for 30s, 65 C 30s – 0.6 C per cycle, one minute at 72 C. The remainder of the reaction consisted of 20 cycles of 95 C for 30s, 56 C for 30s and 72 C for one minute. The 32 cycle profile was then ended with a seven minute extension step at 72 C.

AFLP amplifications were separated by denaturing polyacrylamide gel electrophoresis and visualized by staining with SYBR gold (Molecular Probes). Gels were prepared by first treating the edges of the "long" glass plate with a bind silane solution (0.5 µl bind silane; 200 µl 95 % ethanol 5 % glacial acetic acid) and allowing it to sit for two minutes in a fume hood. The plate was then washed thoroughly several times with 95 % ethanol. The "short" glass plate was treated with 200 µl of SIGMAcote (Sigma) and allowed to sit for 1 minute in a fume hood. The gel mold apparatus was assembled by placing .45 mm spacers between the "long" and "short" plates and placing

Table 4.1. AFLP selective amplification primer codes along with their selective nucleotide sequences.

*Eco*R I primer code	Selective nucleotides (3' to 5')	*Mse* I primer code	Selective nucleotides (3' to 5')
e4	ACC	m2	CAC
e5	ACG	m3	CAG
e6	ACT	m4	CAT
e7	AGC	m6	CTC
e13	ATA	m7	CTG
e14	ATC	m8	CTT
e15	ATG	m11	CCG
e16	ATT		

the plates on a levelled GelSlider (CBS scientific). A 6 % acrylamide gel solution was prepared by adding 60 mL of Sequ-gel 6 acrylamide (National Diagnostics) to 15 mL of 5X Sequa-Gel TBE buffer containing TEMED (National Diagnostics) and mixing well in a 200 mL squirt bottle. Polymerization was initiated by adding 600 µl of 10 % (v/v) ammonium persulfate (Sigma). Once poured, gels were allowed to polymerize for at least 1 – 1.5 hrs. Prior to loading, the gels were placed on S2 gel rigs (BRL) and subjected to electrophoresis at 60 mA for 20 minutes to ensure even heating of the polyacryamide matrix.

Samples were prepared for loading by adding an equal volume of denaturing loading dye (99 % formamide, 1% EDTA pH 8, bromophenol blue) to each selective AFLP reaction, then denaturing at 95 C for 10 minutes. The standard loading protocol, used throughout the AFLP genotyping, involved pipetting 2.5 µl of denatured AFLP reaction into an individual well of the gel. During the gel loading process the samples were kept hot by partially immersing the PCR reaction tubes (or 96 well plate) in boiling water to delay renaturation. Normally there was enough room on each gel to separate two samples of a xeric and a mesic parent, and 30 recombinant inbred line samples. A 10 bp ladder (1 µl of a 1/50 dilution) was loaded as a size standard. All gels were subjected to electrophoresis at a constant 50 W for 2 – 2.5 hours.

Upon completion of the electrophoresis the gel assembly was removed from the gel rig, cooled under cold tap water, and disassembled by separating the glass plates and removing the spacers and comb. The long glass plate, with gel attached, was then treated with 30 ml SYBR Gold DNA stain (2 µl SYBR Gold stock, 11 µl 1 M Tris-HCl, 29.98 ml dH_2O). The stain was allowed to sit for a maximum of five minutes, then was drained

through a charcoal filter. The gel plate was dried thoroughly with paper towel, then placed into a FMBIO fluorescent scanner (Hitachi) where it was allowed to rest on two 0.40 mm plastic spacers to compensate for the gel plate thickness differences from the standard FMBIO plates. The gel was scanned using the 505 nm filter with the "Read Image" software image resolution set to 180 repeats. The digital image, output by the FMBIO scan software, was imported into Photoshop 6.0 (Adobe) for manual genotyping. In all cases, gel images were scored twice in an attempt to reduce incorporating genotyping errors in the mapping dataset.

Linkage Map Construction:

Genotype data from 180 recombinant inbred lines was used to construct a linkage map of the *A. barbata* genome. Raw genotype data was converted to the MAPMAKER format and the mapping analysis was carried out using a combination of MAPMAKER EXP 3.0 (Lander *et al.*, 1987; Lincoln and Lander, 1992) and Mapmanager QTX version b20 (Manly *et al.*, 2001). The genotype data at each marker locus was tested for segregation distortion, from the expected 1:1 ratio, using a χ^2 test with one degree of freedom. Highly significantly distorted markers ($p<0.01$) were removed from the data set in the initial stages of linkage group construction in order to avoid creating spurious groups. It should be noted that marginally significant segregation distortion ($\alpha \sim 0.05$) was tolerated in order to avoid removing too many markers, some of which may have been false positives. Furthermore, any loci that had less than 70 % of individuals genotyped were also removed from linkage analysis.

Linkage groups were determined by using Mapmaker's "group" command with a LOD (log_{10} likelihood ratio) threshold of 4.00 and a maximum linkage distance of 35 centiMorgans (cM). Within each linkage group the proper ordering of markers was estimated using the "order" command, using three point linkage data, with an initial threshold LOD of 3.00, indicating the order was 1000 times more likely than the next best order. A secondary threshold of LOD = 2.00 was then used to place any additional markers relative to the established order. Markers with segregation distortion were then reintroduced into the data set and tested for linkage to the established linkage groups using Mapmanager's "allow for segregation distortion" linkage function. Any distorted markers that caused an established linkage group to split into smaller pieces or join with another group were removed from the final map.

Results:

AFLP polymorphism between and within mesic and xeric ecotypes:

In total, 80 AFLP selective primer combinations were screened in the xeric and mesic ecotypes, initially using two families of each ecotype. The range of polymorphic markers produced by one primer combination was generally between 5 and 20, with majority of these bands being found in the 500 bp to 100 bp size range. When the ten "best" primer combinations (amplifying >100 marker loci) were screened on a larger panel of parental ecotypes, including five mesic and six xeric families, only one AFLP marker was found to be polymorphic within a parental ecotype. In this case one AFLP band was present in all xeric families and present in only one of the mesic families.

AFLP marker segregation in the F6 lines:

A total of 20 selective primer combinations were used to genotype 180 F6 RILs, producing 154 loci could be reliably scored in all lines (see below, Table 4.2). There was no significant difference in the number of "band present" alleles originating from either mesic or xeric ecotypes (χ^2 = .209 , 1 d.f., p = .621). Analysis of the pattern of allele segregation revealed 15 loci with segregation distortion at the 1% level (χ^2 > 6.63, 1 d.f., Table 4.3) and a further 12 loci with segregation distortion significant at the 5% level (χ^2 > 3.84, 1 d.f.). In most cases (20 out of 27) the distortion was toward an over-representation of the band present allele. Of the seven remaining distorted markers, five

had the band present allele in the mesic parent but over representation of the xeric allele in the F6 generation.

Genetic linkage map of the *A. barbata* genome:

Upon removing markers that exhibited ($p<0.01$) segregation distortion, there were 139 marker loci available for linkage map construction. These loci formed 21 linkage groups when grouped at a LOD of 4.00, with 12 markers left unlinked (Table 4.2). The 21 linkage groups, when ordered, were found to span approximately 644 cM of the wild oat genome with an average intermarker distance of 4.3 cM. Eleven of the linkage groups each covered more than 20 cM and consisted of more than three loci. The remaining groups were generally pairs or triplets of markers that covered between two to 19 cM (Figure 4.1). When the markers with strong segregation distortion were positioned relative to the existing map framework, nine of the fifteen markers were distributed over six linkage groups and the remaining markers clustered together, along with two non-distorted loci, into one unique linkage group (Table 4.3; Figure 4.1). These six loci were the only distorted markers that did not disrupt already established linkage groups and as such, were the only highly distorted markers retained in the final map. While this cluster of distorted markers (LG21) contained loci that had the band present allele in the xeric parent or mesic parent, all the loci were distorted toward an over representation of the alleles from the xeric parent.

Table 4.2. Number of polymorphic markers from the 20 primer pairs used to construct the genetic linkage map. The list does not include the unlinked markers or any distorted locus not associated with LG21.

Primer pair	Number of polymorphic markers
e16m11	10
e5m8	10
e6m7	10
e5m2	8
e6m2	8
e7m3	8
e13m7	7
e14m6	7
e14m8	7
e15m11	7
e4m11	7
e5m3	7
e5m7	7
e6m4	7
e6m6	7
e5m11	5
e14m11	4
e13m6	3
e5m6	3
e15m2	2
Avg.	6.7

Table 4.3. AFLP marker loci exhibiting segregation distortion significant at p<0.01. The parental origin of the "band present" allele is given as the last letter of the marker name (m = mesic, x = xeric). The direction of distortion indicates which parent's allele is over represented among the F6 lines.

Marker	Putative linkage group	Chi-square	p-value	Direction of distortion
e13m7-360x	8	8.00	0.004	xeric
e5m2-106m	8	21.44	<0.0001	mesic
e16m11-280m	8	48.64	<0.0001	mesic
e14m11-270m	9	6.92	0.008	mesic
e16m11-134m	9	6.64	0.01	mesic
e14m8-600m	10	24.04	<0.0001	mesic
e14m8-360m	11	10.02	0.001	mesic
e16m11-175m	11	6.64	0.01	mesic
e7m3-299x	19	8.29	0.004	mesic
e14m8-550x	21	10.29	0.001	xeric
e5m8-240m	21	10.56	0.001	xeric
e5m11-232m	21	20.26	<0.0001	xeric
e6m6-153m	21	28.8	<0.0001	xeric
e6m2-190x	21	68.61	<0.0001	xeric
e6m7-170x	21	72.25	<0.0001	xeric

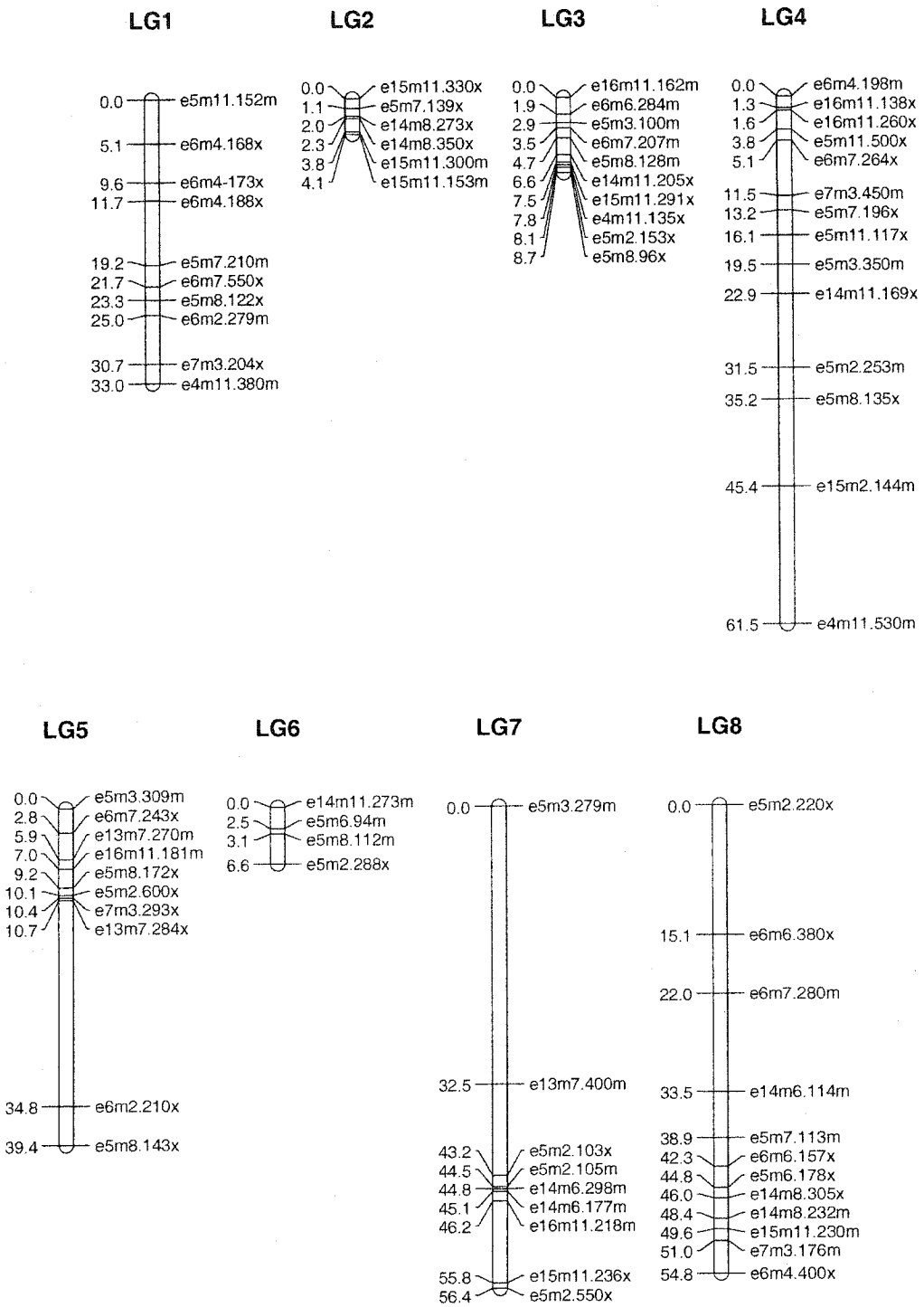


Figure 4.1. Genetic linkage map of the *A. barbata* genome constructed with AFLP markers. Cumulative map distances in centiMorgans (Haldane) are shown on the left of each group. Marker names are shown to the right of each group.

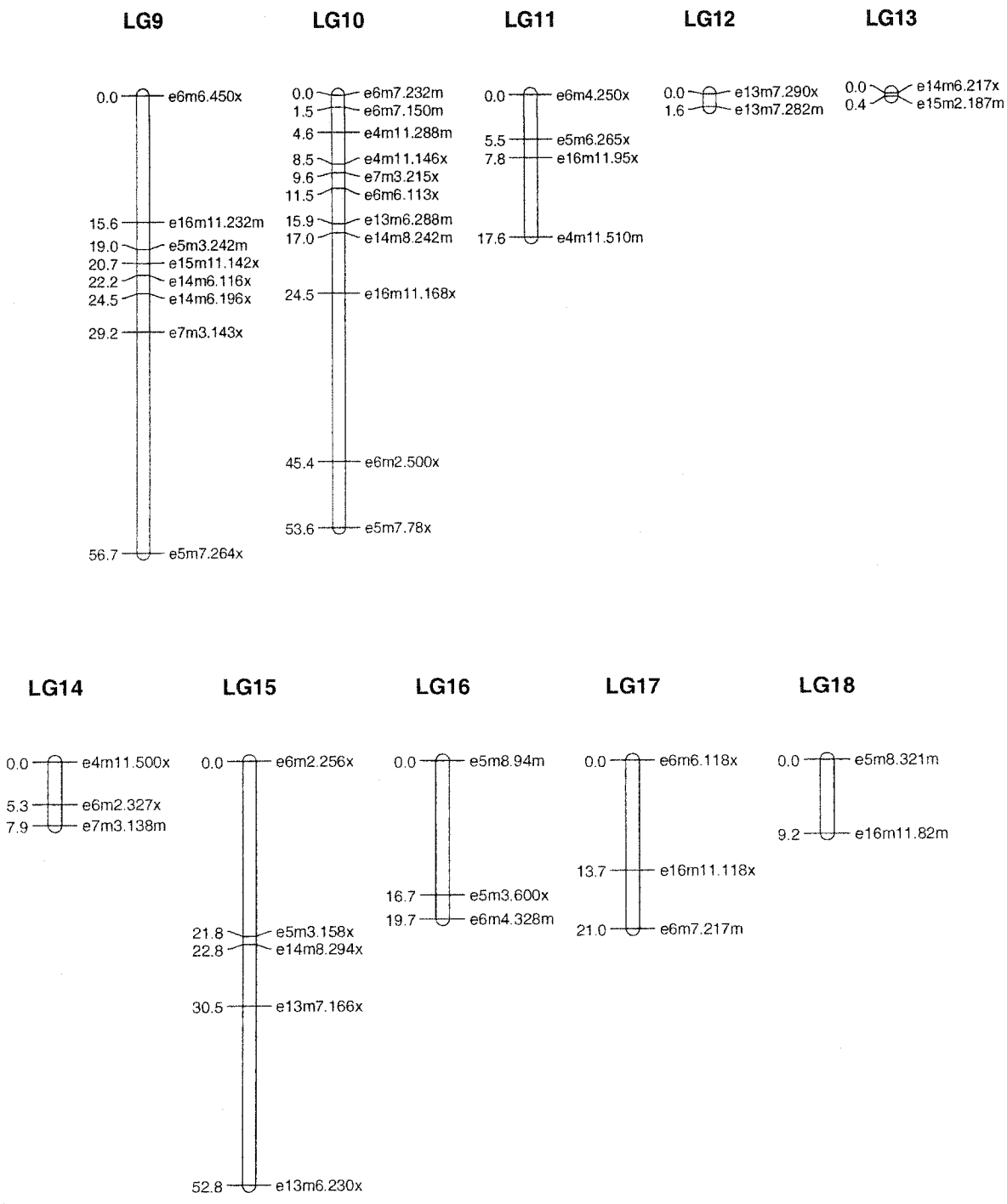


Figure 4.1 continued.

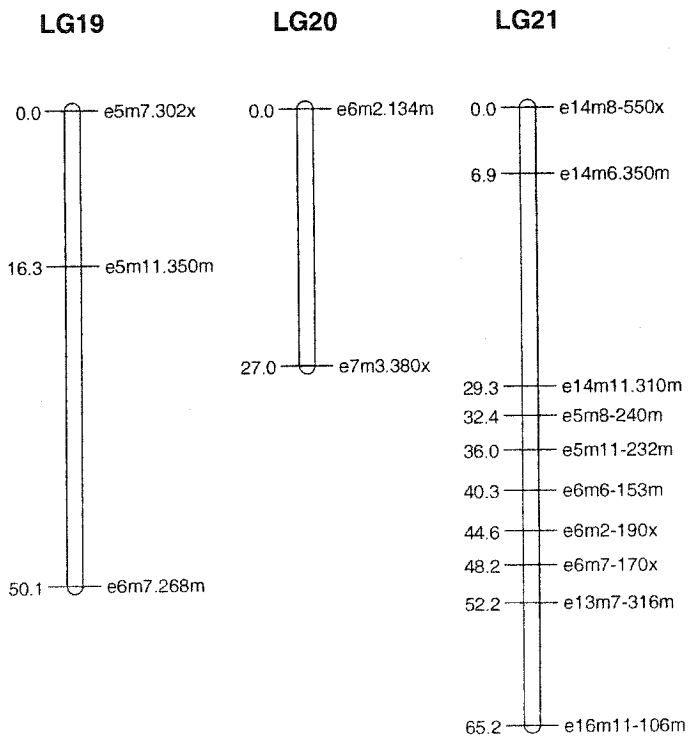


Figure 4.1 continued.

Discussion:

Throughout this thesis I have hypothesized that the wild collected parental ecotypes represent naturally occurring inbred lines, essentially completely homozygous at all loci. The F1 generation resulting from a cross between ecotypes would be expected to be heterozygous at all loci, QTL included, that differ among the parents. Ideally, this serves to maximize the number of possible recombinant gametes that will unite to form the beginnings of the recombinant line population. Evidence in support of this hypothesis is twofold. At the level of the phenotype I have presented evidence in a previous chapter (Chapter Three) that the mesic families and xeric families are genetically uniform within ecotypes, with regard to fitness and related traits, suggesting there are no QTL segregating among families within ecotypes. In this chapter I have genotyped a subset of the available parental families at >100 AFLP loci, found to be polymorphic between ecotypes, and did not find any significant number of polymorphic markers within ecotypes. While this survey was not exhaustive, it does however provide direct evidence at the molecular level for the genetic uniformity of the parental forms.

The estimated genomic distance covered by the genetic map was 644 cM (adjusted to account for multiple generations of recombination) across 21 linkage groups or approximately 40% – 45% of the expected total genome size of 1385-1600 cM. The expectation was calculated as the midpoint between the latest estimates of genome map coverage in hexaploid oat (1890 cM, Wight *et al.*, 2003; 2351 cM, Jin *et al.*, 2000) and diploid oat (880 cM, Kremer *et al.*, 2001). This is based on the assumption that genome size in centiMorgans for a tetraploid (fully diploidized in this instance; Hakim-Elahi and Allard, 1983) linearly increases with ploidy. In this case I expect the tetraploid species to

contain one more haploid genome than the diploid species and one less haploid genome from the hexaploid species.

The most straightforward explanation for the lower marker coverage in *A. barbata* is simply the lack of identifying polymorphic markers along certain parts of some of the oat chromosomes. These unidentified markers could be necessary to link together some of the smaller linkage groups. The prescence of markers that remain unlinked to any other markers, and therefore located in isolated portion of a chromosome, seem to support this conclusion. However, in adding the marker loci from the last three AFLP primer pairs all new loci were found to be linked to an existing marker and fell *within* the previously mapped portion of the genome. This would seem to imply that most of the AFLP markers generated from the standard method (Vos *et al.*, 1995) target similar portions of the genome. This result prompted me to cease adding more *Eco/Mse* AFLP markers as there was the potential for diminishing returns on additional map coverage.

It is possible the AFLP protocol utilizing the methylation sensitive restriction endonuclease *Eco*R I, as the "rare cutter" enzyme (Vos *et al.*,1995), fails to generate markers in portions of the genome where there is significant methylation of the DNA. It has been documented that if the cytosine base in the recognition sequence of *Eco*R I is methylated the restricting activity of this enzyme is severly inhibited (Sambrook *et al,* 1989). Future efforts to add more AFLP markers to the map could probe this possibility by altering the AFLP protocol to use a non methylation sensitive "rare cutter" enzyme such as *Pst* I, which has been reported to aid in generating randomly distributed markers in large complex genomes (e.g. Castiglioni *et al.*, 1999; Pradhan *et al.,* 2003)

A more intriguing possibility is that the xeric and mesic ecotypes contain expansive chromosomal regions that are nearly identical, as a result of relatively recent shared ancestry. This follows from the hypothesis that the extant ecotypes were derived from an ancestral pool of variation (Allard *et al.*, 1972) produced when an ancestral Spanish genotype underwent a bout of outcrossing. It is possible that the ancestral genotypes, in this pool of variation, that gave rise to the extant ecotypes each received identical chromosomal blocks in some areas of the genome while the rest of the genome was recombinant. If this were the case, it can be argued that only these recombinant areas of the oat genome segregated in the subsequent generations leading up to the two locally adapted forms observed in California. This becomes significant because it would only be necessary to map a smaller portion of the genome in order to find all the genes that underlie the phenotypic differences between the extant ecotypes.

To test this hypothesis it will be necessary to saturate the genetic map with more markers, ideally of different types, in order to determine how much genome coverage can be added. However, it may be possible to infer the completeness of the map, and therefore the chromosomal arrangement, by mapping the QTL for the fitness traits with the map generated in this study. If the map really only covers 40% of the total genome then, on average, we would have the opportunity to identify 40% of the QTL that explain roughly 40% of the trait variation. This of course assumes the genes controlling the trait variation are distributed randomly throughout the genome with approximately equal additive effects. However, if we can consistently identify numerous QTL for each trait, within the confines of the present map, that explain a large portion of the trait variation, then we can provide evidence that this mapped portion of the *A. barbata* genome, at the

very least, was important in contributing to underlying segregating genetic differences that led up to the local adaptations of the two dominant ecotypes.

Perhaps the most powerful test of this hypothesis would be to backcross the mesic and xeric ecotypes to Spanish asscessions of *A. barbata* and observe patterns of AFLP segregation at a variety of loci. The presence of AFLP alleles that appear to be polymorphic between the ancestral Spanish genotype and each ecotype but are not polymorphic between the ecotypes would provide strong evidence that the parents of our RILs share identical chromosomal regions. The challenge of testing this experimentally however would be to find a suitable ancestral genotype from the extant Spanish wild oat population and generate enough backcross progeny to yield a large experimental population.

While the primary objective of thesis chapter was the creation of a linkage map of the wild oat genome, it was also possible to utilize marker segregation patterns to make inferences about genome similarity between the parental strains used to create the mapping population. It is reasonable to assume that the more time that has elapsed since a crossed species pair shared a common ancestor the more likely some segregation distortion mechanisms could accumulate. Several possible causes of segregation distortion have been documented, including specific segregation distorter genes, aberrant meiotic pairing due to major chromosomal rearrangements, self incompatibility loci, and differential progeny viability (reviewed in Jenczewski *et al.*, 1997). In the *A. barbata* system, the upper limit on the time since introduction to California is approximately 200-400 years ago. The amount of marker segregation distortion that was observed in this study (16.2%) was comparable to that seen in hexaploid oat (9.8%, Wight *et al.*, 2003)

and diploid oat (16%, Yu and Wise, 2000) mapping populations but was less than that observed in studies that crossed highly divergent species (~50% in *Mimulus*; Fishman *et al.*, 2001). One linkage group in particular stood out in the segregation analysis for having loci that were all distorted toward the xeric parent, regardless of which parent possessed the "band present" AFLP allele. Because of this latter point this result can not simply be reconciled with a bias in overestimating the numbers of "band present" alleles due to marker scoring by eye. Instead, a possible explanation lies in the influence of selection during the early stages of recombinant line propagation.

Ideally, the single seed descent method of producing inbred lines should minimize the effects of selection, with each line being propagated to the next generation by one seed thereby ensuring each line has equal fitness. However, if there are loci causing, for example, differential pollen viability then any marker alleles linked to the more successful pollen allele will be passed on to the progeny at an increased proportion. If this were to occur in the early stages of RIL propagation (i.e. F1 to F2), before the lines were independently propagated, then a highly non-Mendelian segregation pattern would occur with the segregation ratio at all marker loci in the vicinity of the offending locus appearing skewed toward one of the parental genotypes. This scenario seems to fit the pattern of distortion identified on linkage group 21 where it would appear that the pollen carrying xeric alleles was the more successful type (Table 4.3). The pollen viability locus would appear to be located near the midpoint of the linkage group as the segregation distortion becomes more pronounced in this region and seems to drop off near the proximal and distal ends of the group.

Chapter Five:

Identification and Analysis of QTL for Fitness Related Traits

Introduction:

Understanding the genetic basis of variation in fitness related traits is a central goal in contemporary evolutionary biology. Using quantitative genetic analysis on common garden and reciprocal transplant experiments it has been shown that substantial genetic variation exists among species and among divergent locally adapted populations (Nagy, 1997). Documenting the individual genetic factors responsible for this population divergence has proven difficult however, due to the fact that most genes having a major effect on fitness will be fixed, as a result of strong selection in a given environment (Orr, 2001; Barton and Keightley, 2002). By making crosses of related morphs or ecotypes, within a species, it is possible to produce progeny that contain novel recombinations of the parental alleles (i.e. major genes included) that segregate in a Mendelian manner within a randomized genetic background (Lexer *et al.*, 2003). By analyzing a large population of these recombinant progeny it should be possible to identify the individual loci and thus each single locus contribution to the variance in the overall phenotype.

In this chapter I explore the genetic architecture underlying the adaptive population divergence among the mesic and xeric ecotypes of *A. barbata*, utilizing the method of QTL mapping (reviewed in Tanksley, 1993; Doerge *et al.*, 1997; Lynch and Walsh, 1998). The complexity of the spatial arrangement of QTL in the wild oat genome, will be of particular interest because it allows insight into how much of the genome may be a target of selection in a novel environment. Moreover, given the correlated

relationship among fitness and related traits in *A. barbata* demonstrated in Chapter Three, and the ubiquity of pleiotropic QTL underlying correlated traits reviewed in Chapter Two, it is also relevant to focus on how many loci underlie multivariate complexes of traits, for it is these loci that must have played a role in the divergence of the parental ecotypes. If these pleiotropic QTL become targets of selection then it is possible to produce change in multiple traits simultaneously.

There is still debate as to the importance of epistasis in adaptive evolution. In Sewall Wright's shifting balance theory, epistasis is thought to result in many different fitness peaks separated by fitness valleys along an adaptive landscape (Wright, 1931). Subpopulations occupying heterogeneous environments may begin to diverge if they climb different adaptive peaks. Kim and Rieseberg (2001) point out that the importance of epistatic interactions in population differentiation depends on whether the magnitude of the epistatic effects among loci rivals that of the additive main effects of individual loci. To this end, they concluded that epistasis is a relatively unimportant contributor to the variation among morphological traits in a backcross population of annual sunflowers (*Helianthus annuus* X *H. debilis*); instead favouring "Fisherian" selection on main effect loci as a more plausible mode of trait divergence in this system (Kim and Rieseberg, 2001). While they did find evidence of a few strong epistatic effects on pollen viability, the number of significant interactions observed for most other traits was close to that expected by chance (Kim and Rieseberg, 2001). Similarly, Ungerer *et al.* (2002) found only five interacting loci, despite finding >60 main effect QTL, in a broad survey of inflorescence traits in *Arabidopsis thaliana*. Interestingly, detection of the interactions was found to be dependant on genetic background (i.e. cross specific), and when

detected, the epistatic effects were generally overshadowed by the additive effects of the individual QTL (Ungerer *et al.*, 2002).

Here I present the results of efforts to identify the individual genomic regions underlying variation and covariation in fitness related traits as measured in a common garden greenhouse environment. The questions addressed in this chapter are (1) Can QTL for fitness and related traits be detected in the greenhouse?, (2) Do the signs on the QTL addititve effects explain the transgressive segregation in the F6 lines?, (3) Do QTL for genetically correlated traits co-localize to the same regions of the genome?, (4) Is there any evidence for epistatically interacting loci?, and are their magnitudes as large as the additive QTL effects?, (5) Does the hybrid index correlate with fitness?

Materials and Methods:

Single marker tests for marker-trait association:

Marker data in the form of a molecular map (Chapter Four) and trait data (Chapter Three) were imported into QTL Cartographer (Windows version 2, Wang *et al.*, 2004) for single marker analysis and composite interval mapping of QTL (see below). In order to identify significant marker trait associations each marker locus was fit to the individual quantitative traits in a simple linear regression model:

$$y_i = b_0 + b_1 x_i + \varepsilon$$

where y_i is the the trait value of the ith RIL, b_0 is the overall trait mean, x_i is an indicator variable based on the genotype of the ith RIL at a marker locus, b_1 is the effect of the marker allele, and ε is the model error. In all cases those RILs homozygous for the xeric allele were coded as 1, and those homozygous for the mesic allele as -1. A positive regression slope therefore indicates the xeric allele at the marker locus was associated with an increased trait value. Statistical analysis was carried out using the single marker analysis function of QTL Cartographer (Wang *et al.*, 2004) which tests the regression slope against the null hypothesis of $b_1 = 0$, with one degree of freedom. No adjustment for multiple tests was applied, as the single marker tests were used simply as a heuristic to identify markers with putative QTL linkage.

Genome composition and measures of fitness:

The proportion of xeric alleles across all the marker loci was calculated for each RIL as a measure of genome composition. The proportion of "xerism" for each line was then correlated to the mean trait value, for that line, for each trait. Significance of correlations was determined using randomization tests in order to account for the nonindependance (i.e. linkage) of the marker data incorporated into the measure of genome composition.

QTL localization via Composite Interval Mapping (CIM):

Composite interval mapping is an extension of the basic interval mapping (IM) method proposed by Lander and Botstein (1989) that scans the linkage map at regular positions (e.g. every 2 cM within each marker interval) and fits the molecular marker data at the test position to a multiple regression model. The power of the CIM method is derived from adding some marker loci as marker cofactors in an attempt to control the "background" effects of QTL outside the position being tested:

$$y_j = b_0 + b^* x_j^* + \sum_{k \neq i,i+1} b_k x_{jk} + \varepsilon_j \qquad \text{(Zeng, 1994)}$$

where y_j is the trait value of the jth individual, b_0 is the trait mean, b^* is an estimate of QTL effect at the test position, x_j^* is an indicator variable based on marker genotype of individual j at the markers flanking the test position (in Zeng's (1994) notation the

flanking markers are denoted *i* and *i+1*), b_k is the effect of a marker cofactor, x_{jk} is and indicator variable of marker genotype of individual j at a marker cofactor, and ε_j is the model error. The presence of a QTL is tested by comparing the fit of the multiple regression model including a QTL effect (b^*) at the test position, with a reduced model containing a QTL effect equal to zero ($b^* = 0$) at the test position (Zeng, 1994)

The challenge of CIM is finding the proper number and location of marker loci to act as cofactors (Doerge, 2002). Too few markers and the power of CIM is not realized; too many markers and the statistical power of the analysis is compromised (i.e. each added marker uses up one degree of freedom). Preferably, the smallest number of markers that can account for a sizable portion of the background variation would be chosen, thus limiting the number of extra parameters needed in the multiple regression model. These markers would be ideally located outside a window extending around the test position, near other QTL controlling the trait of interest, thereby accounting for some of the effects of these background QTL without absorbing effects of QTL near the position being tested (Zeng,1993; Doerge, 2002). Several methods exist to automate the cofactor selection process including stepwise regression with forward selection, backward elimination, or a mixture of forward selection and backward elimination. The goal of each method is the same; identify marker loci that have the largest effect on the trait (via regression of trait value on marker genotype), quantified by the R^2 change in the regression model by adding or dropping the marker locus from the selection process. As soon as no more markers add a significant change (significance dictated by the experimentor) to the regression model the selection process is stopped and the collection of markers, remaining in the model, are the best marker cofactor choices. In fact, Broman

and Speed (2002) suggest these same stepwise regression methods can actually be used to identify QTL rather than simply yield cofactors to be used in CIM.

In this study CIM was carried out for each trait individually, in order to find the best markers to use as cofactors in each analysis. In all cases the stepwise regression procedure implemented in QTL Cartographer, with forward selection and backward elimination, was used to identify marker cofactors. Out of the total of markers chosen to be cofactors, five (the default in QTL Cartographer) were chosen to control the background variation. The intervals of the linkage map were then scanned every 2 centiMorgans (cM) for the presence of a QTL, with the cofactor window size set to 15 cM. At each position the likelihood ratio (LR) of a present QTL tested against the null hypothesis of no QTL (or an unlinked QTL), was calculated and plotted against map location (Appendix Two). A QTL was declared significant if the LR peaked above the 5% genome wide error threshold. A suggestive QTL was identified when the LR peaked above the approximate chromosome wide 5% error threshold.

The significance threshold for each trait was determined using the permutation method of Churchill and Doerge (1994). Briefly, this method randomizes the trait data relative to the map data and then scans the map for significant QTL. The largest test statistic across the randomized QTL analysis is recorded and the randomization/QTL analysis procedure is repeated. After 1000 randomizations, the 1000 test statistics that have been retained are sorted by size and the 950^{th} value is taken as an estimate of the threshold needed to maintain a 5% genome wide Type I error rate (see Appendix Two for genome wide thresholds for each trait). The chromosome wide threshold was set at a LR

of 9.42 (a LOD of two) across all traits which was a conservative threshold when compared to the chromosome wide permutation test thresholds.

Marker Interactions and Epistatic QTL:

In order to identify epistatically interacting portions of the genome, the molecular marker and trait data from the F6 generation were imported into the Epistat computer package (Chase *et al.*, 1997). This program quickly tests all pairwise combinations of markers for departures from an additive relationship with each trait. The null hypothesis assumes that the sum of the trait means of the two parental genotype classes (i.e. AABB and aabb) at each pair of loci is equal to the sum of the trait means of the recombinant genotype classes (i.e. AAbb and aaBB). The null hypothesis is rejected when the log-likelihood ratio statistic for departure from additivity exceeds a given threshold, in a manner similar to interval mapping main effect QTL.

Determining the appropriate significance threshold for declaring significant epistasis is slightly more complicated than for mapping main effect QTL. Because all pairwise combinations of markers are tested, the problem of observing multiple false positives due to multiple testing arises very quickly. In this study I have used 133 loci to map QTL which translates into 8778 [(n*(n-1))/2] statistical tests for two-way epistatic interactions. Using a standard Bonferroni correction (assuming an overall error rate of 5% is desired), any single test would need to have a p-value less than 0.05/8778 = 5.7×10^{-6}, making detection of most interactions unlikely. However, as Cheverud (2000) points out not all of the tests are independent, given that we know some of the markers

are linked into groups, making the "effective" number of independent tests much less than the total. Cheverud (2000) suggests determining the effective number of independent markers (M_{eff}) on each linkage group by performing a factor analysis on the genotype scores at the markers of each group. This method involves relating the observed variance among the eigenvalues of the marker genotypes in one linkage group to the maximum possible variance among eigenvalues for that linkage group:

$$M_{eff} = M\left[1 - (V(\lambda)(M-1)/M^2)\right] \qquad \text{Cheverud (2000)}$$

where M is the maximum eigenvalue variance (equal to the number of markers in the group) and $V(\lambda)$ is the observed variance among marker eigenvalues. Once the effective number of tests is calculated for each linkage group, the effective number of independant two-way interaction tests across the entire map can be determined and used in the Bonferroni correction (Cheverud, 2000). Using this method the effective number of independant tests for the mapped portion of the oat genome was found to be 2624.

Because of the extra complexity in identifying epistasis at the molecular level I chose to analyze the data in two parts. First, I used Epistat to identify all the marker interactions that were significant at an arbitrary significance threshold of $p < 0.0015$ for each trait. This corresponded to the suggested default setting of a log-likelihood ratio of five (Chase *et al.*, 1997). I then compared the number of significant interactions to the number expected to arise solely due to chance, which for 8778 tests and an uncorrected p-value of 0.0015 is approximately 13. Thus, any number of interactions over and above this number would imply significant epistasic interactions were present for that trait.

Second, I used the MNTECRLO program of Epistat to conduct randomization tests of all significant interactions in order to generate a distribution assuming no interaction, for each marker pair for each trait. The data were randomized and analyzed for 250 000 permutations and the probability of observing an interaction as significant (or more significant) than the experimental value was recorded for each marker pair. This probability was then used to determine if individual marker interactions were significant by comparing it to the M_{eff} –Bonferroni corrected significance threshold ($p = 3.8 \times 10^{-5}$ for a corrected $\alpha = 0.10$; $p = 1.9 \times 10^{-5}$ for a corrected $\alpha = 0.05$). Any marker interactions significant at the Bonferroni adjusted level were then fit to a two-way fixed effects ANOVA model in order to estimate the epistatic effect.

Results:

In a preliminary "single marker"analysis using simple linear regression it was possible to determine marker-trait associations and the direction of the xeric additive effect for the markers incorporated in the linkage map and the markers that were unlinked (Appendix One). Among the mapped markers there were over 250 marker-trait associations significant at the 5% level, with the xeric alleles at different loci, in most instances, having increasing and decreasing additive effects on the same trait. None of the unlinked markers showed any association with any of the traits measured in the greenhouse. Furthermore, no significant association was found between the proportion of xeric alleles (i.e. the hybrid index) and the line means for each trait, using randomization tests to account for nonindependance of the marker data. Overall, the QTL analysis was able to detect at least two significant or suggestive QTL for each trait, with the exception of dry mass at 60 days. While the identified QTL were spread across 14 of the 21 linkage groups, there were large linkage groups (~300 cM total) that were completely devoid of QTL.

Composite interval mapping QTL for traits measured in the greenhouse:

Principal component 1 and fitness: flowering time, tiller number, spikelet number, and allocation to leaf and tiller mass.

A combined total of 23 QTL were identified for the traits known to load heavily on principal component one (PC1) in the greenhouse environment, including flowering

time, tiller and spikelet number, and allocation to tiller and leaf mass (Figure 5.1; Table 5.1). Out of this total, QTL affecting flowering time were located on LG8, LG10, LG15, LG17, and LG21. The QTL with the largest effect was located on LG 15 with the xeric allele increasing the flowing time by 3.5 days (14 % of the variation). Likewise, the xeric effect of the QTLs on groups 8, 17 and 21 all increased flowering by 2-3 days and accounted for 6.7%, 8.8%, and 4.6% of the variation in flowering respectively.

Five QTL were identified for fitness (i.e. spikelet production), however three of these loci mapped to the same position as the flowering time QTL on LG10, LG15 and LG 17 indicating possible pleiotropy. The xeric effect of the fitness QTL, that co-localized with the flowering time locus on LG10, increased spikelet production by almost 37 spikelets. This pattern of opposite effect signs was also observed for the fitness QTL that co-localized with the two increasing flowering time QTL on LG15 and LG17, which decrease spikelet production by 52 and 41 spikelets respectively. In total, the three spikelet number QTL accounted for nearly 30% of the variation in fitness among family means. The remaining two spikelet QTL were located on LG1 and LG6 and could

Table 5.1. QTL identified for traits measured in the greenhouse. QTL position is given as the number of the first marker in the interval (Marker), and as cumulative map distance from the first marker of the linkage group (Position). Significant QTL are shown in bold (see Appendix Two for significance values). Additive effects are given as the effect of the xeric allele. %Var is the variation among line means explained by the QTL.

Trait	QTL #	Group	Marker	Position	LR	Additive	% Var
GM	1	6	3	5	11.6	0.085	5.8
	2	**7**	**8**	**55**	**12.3**	**0.086**	**6**
	3	15	3	29	11.99	-0.109	20
HT20	**1**	**8**	**6**	**44**	**17.33**	**-0.294**	**8**
	2	10	4	8	9.67	-0.214	4.1
	3	11	2	5	11.6	-0.237	5
	4	21	5	40	9.46	0.232	4.1
FL	**1**	**8**	**1**	**14**	**16.03**	**2.5**	**6.7**
	2	**10**	**6**	**11.5**	**14.44**	**-2.26**	**5.4**
	3	**15**	**1**	**6**	**25.74**	**3.69**	**14.4**
	4	**17**	**1**	**12**	**19.95**	**2.87**	**8.8**
	5	21	2	27	10.59	2.11	4.6
MH	1	3	9	8	10.94	-3.69	4.7
	2	5	9	39	10.46	3.62	4.6
	3	**7**	**7**	**50**	**16**	**-4.92**	**8.4**
TN	**1**	**6**	**1**	**0**	**30.35**	**1.52**	**11**
	2	**8**	**2**	**19**	**12.02**	**-0.99**	**4.6**
	3	**15**	**1**	**16**	**31.38**	**-1.81**	**15**
	4	**12**	**1**	**12**	**18.91**	**-1.28**	**7.9**
SN	1	1	1	7	10.16	28.91	4.2
	2	**6**	**1**	**0**	**39.88**	**54.57**	**15.3**
	3	**10**	**6**	**11.5**	**18.74**	**36.92**	**6.8**
	4	**15**	**1**	**10**	**25.23**	**-52.92**	**14.3**
	5	**17**	**1**	**10**	**18.13**	**-41.8**	**9.1**

Table 5.1. continued

Trait	QTL #	Group	Marker	Position	LR	Additive	% Var
TM	**1**	**7**	**5**	**45**	**15.34**	**-0.54**	**6.2**
	2	**10**	**6**	**11.5**	**14.95**	**0.526**	**6**
	3	**15**	**1**	**10**	**37.11**	**-1.11**	**27.5**
	4	**17**	**1**	**12**	**18.42**	**-0.635**	**8.9**
LM	**1**	**3**	**1**	**0**	**37.81**	**0.723**	**14.7**
	2	**6**	**1**	**0**	**23.34**	**-0.547**	**8.5**
	3	**10**	**9**	**28.5**	**14.78**	**-0.492**	**7.2**
	4	**14**	**1**	**0**	**15.08**	**0.426**	**5.4**
	5	**17**	**2**	**19.7**	**22.72**	**0.555**	**9.1**
TOTM	**1**	**3**	**2**	**2**	**19.16**	**0.573**	**8.4**
	2	**15**	**1**	**16**	**22.49**	**-0.694**	**13.4**
PC1	**1**	**6**	**1**	**0**	**31.56**	**0.343**	**11.7**
	2	8	2	17	11.38	-0.208	4.3
	3	**10**	**6**	**11.5**	**16.14**	**0.242**	**5.8**
	4	**15**	**1**	**10**	**33.25**	**-0.461**	**20.9**
	5	**17**	**1**	**10**	**23.9**	**-0.335**	**11.2**
	6	21	3	31.2	11.09	-0.216	4.5
PC2	1	1	8	27	9.64	0.207	4.3
	2	2	5	3.8	9.42	-0.206	4.2
	3	3	1	1	11.17	0.228	5
	4	6	3	5	11.87	-0.235	5.4
	5	8	1	3	11.43	-0.239	5.9
PC3	**1**	**3**	**9**	**8**	**14.5**	**-0.246**	**5.9**
	2	**6**	**3**	**5**	**14.98**	**-0.27**	**7.2**
	3	**7**	**7**	**50.2**	**22.44**	**-0.356**	**12.5**
PC4	1	8	7	44.8	11.75	-0.238	5.5
	2	15	4	30.4	10.96	-0.235	5.4

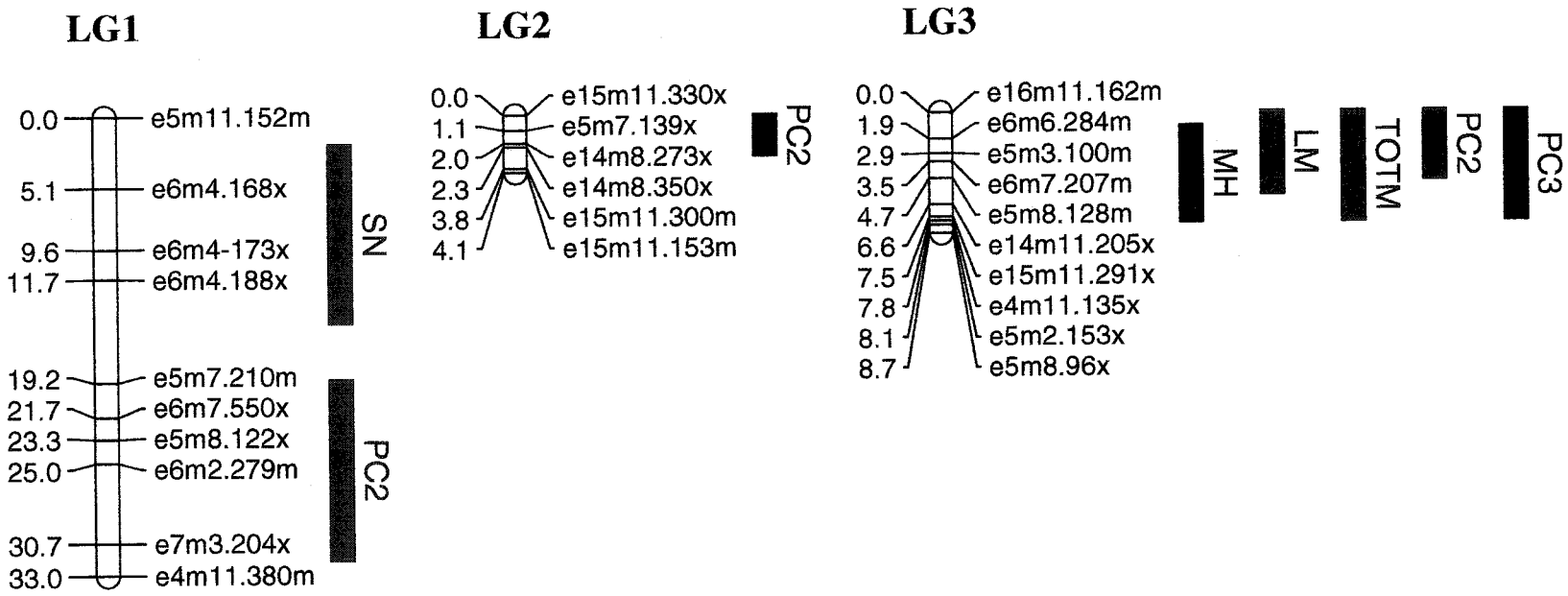


Figure 5.1. Genetic linkage map with positions of QTL for traits measured in the greenhouse. All QTL that are significant at the chromosome wide threshold are shown (see Appendix Two for further details). One LOD confidence intervals around the QTL position are shown as boxes. Red shading indicates the xeric allele increases the trait value. Blue shading indicates the xeric allele decreases the trait value.

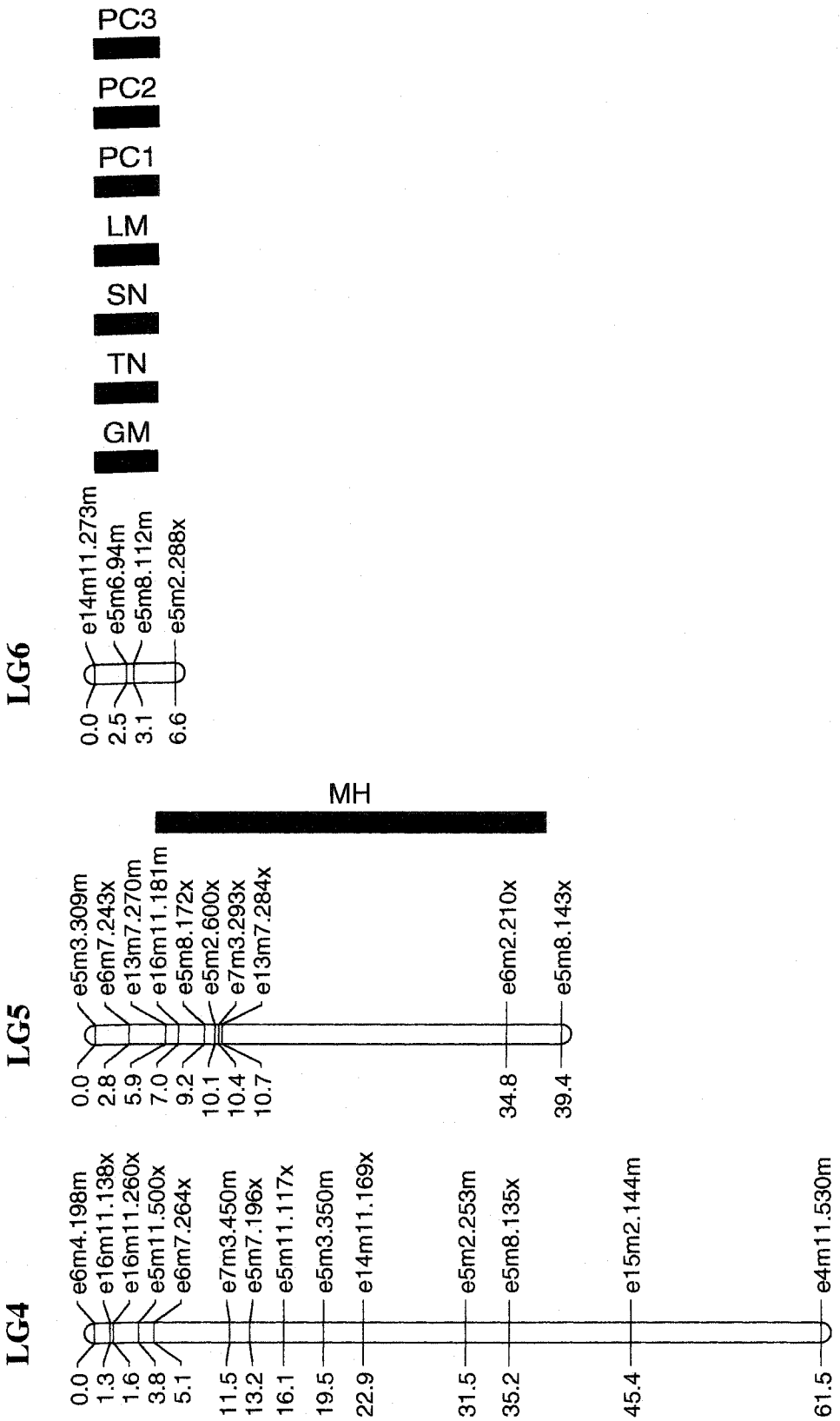


Figure 5.1. continued.

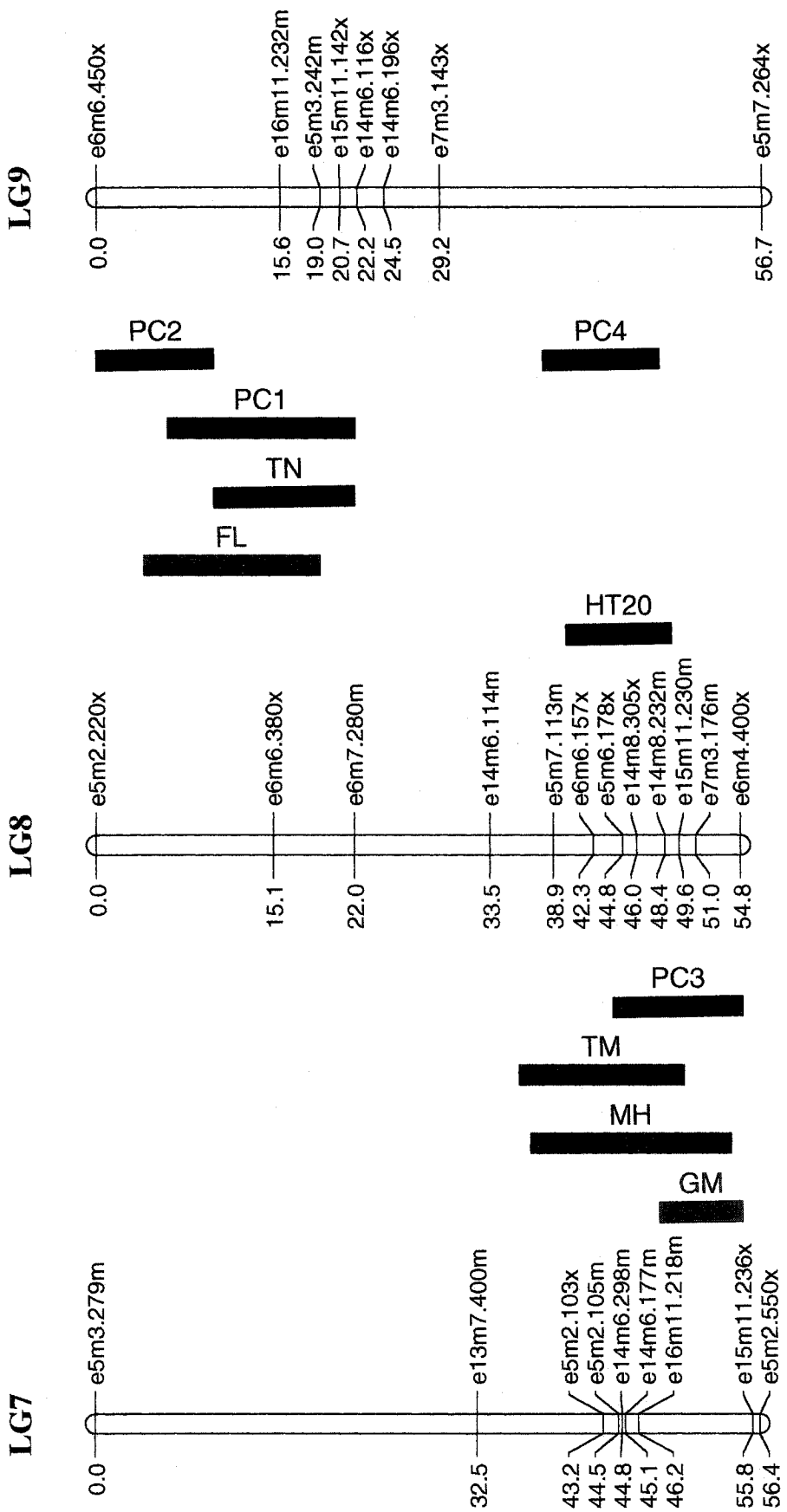


Figure 5.1 continued.

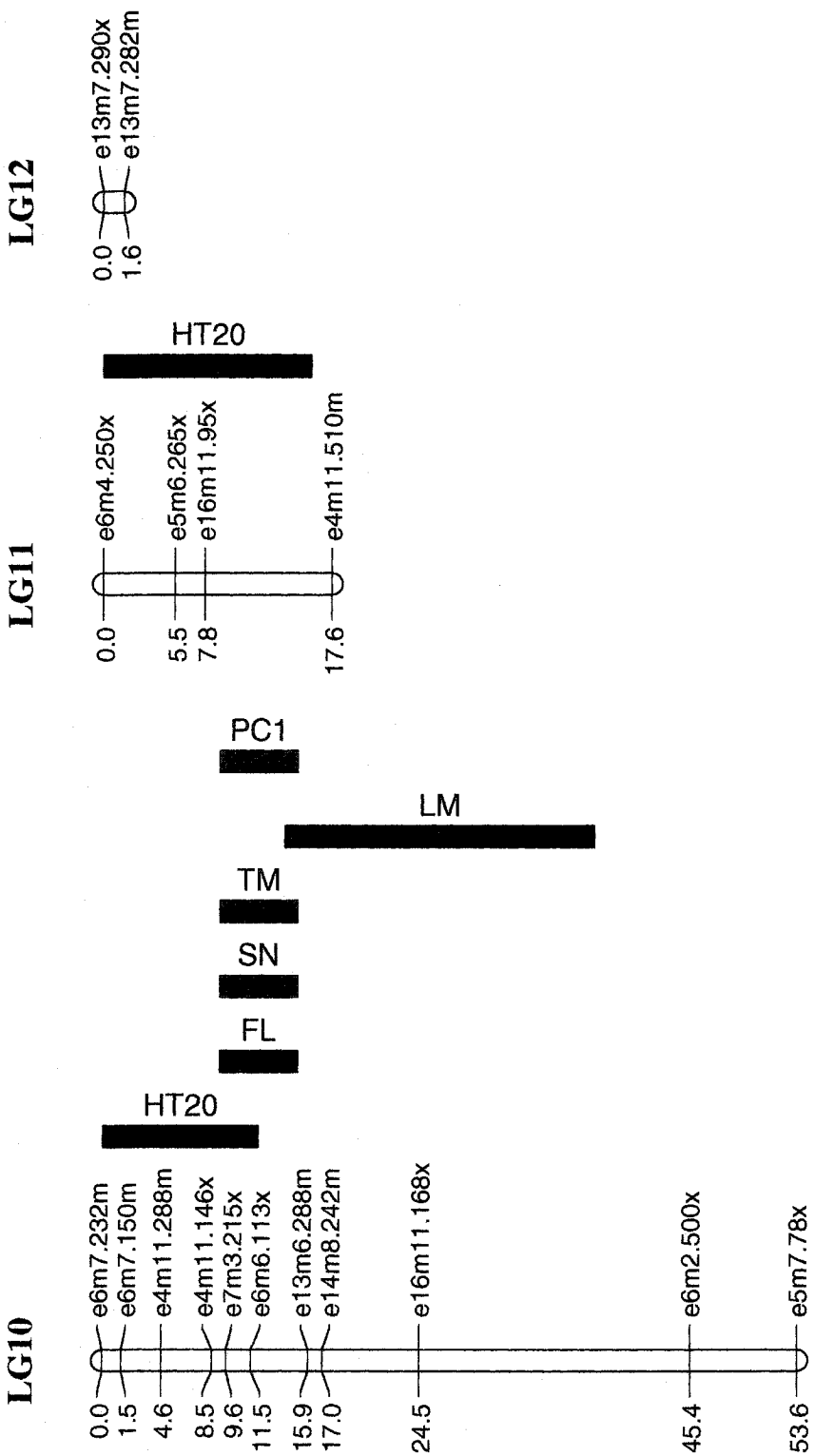


Figure 5.1 continued.

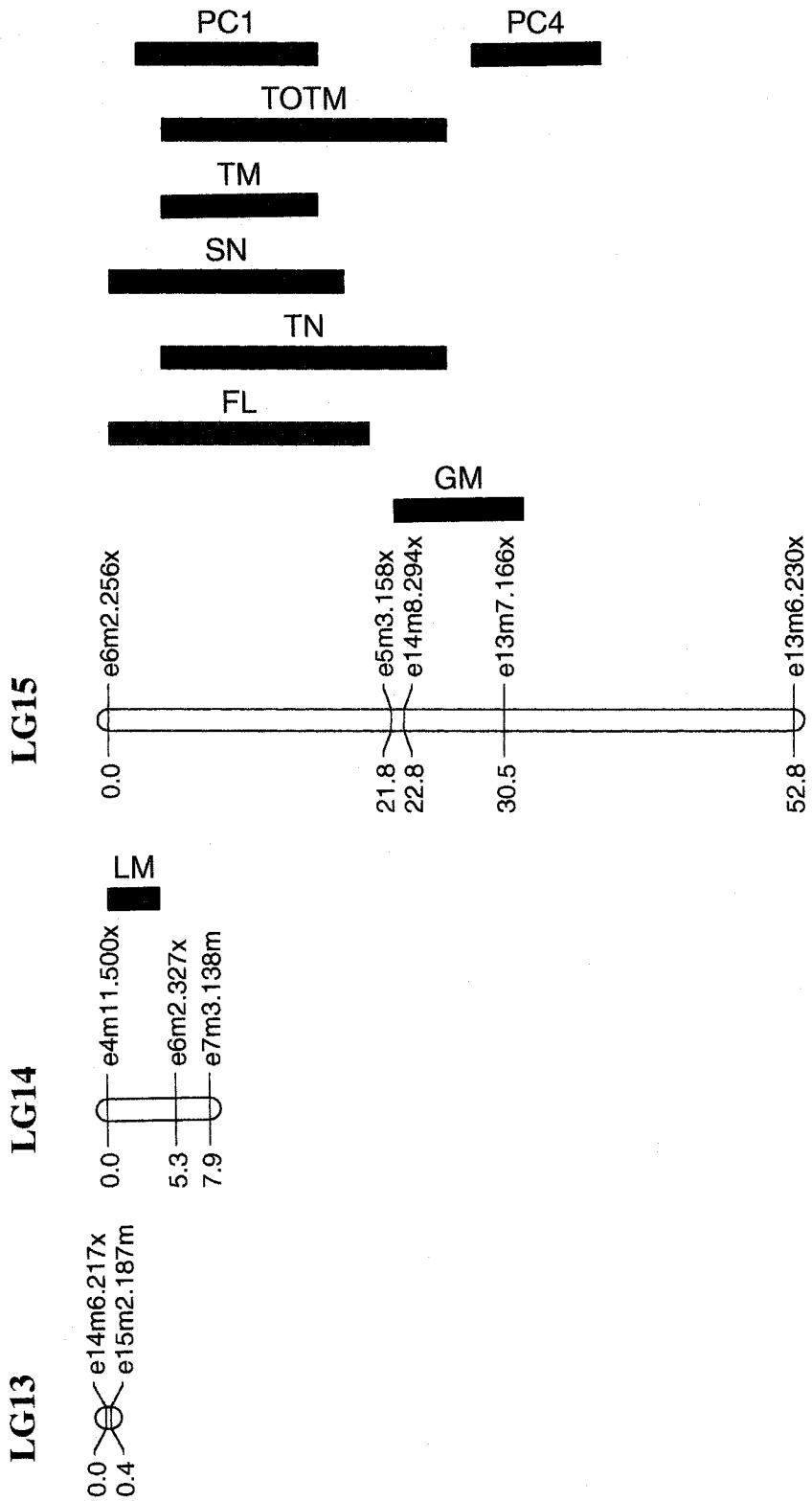


Figure 5.1 continued.

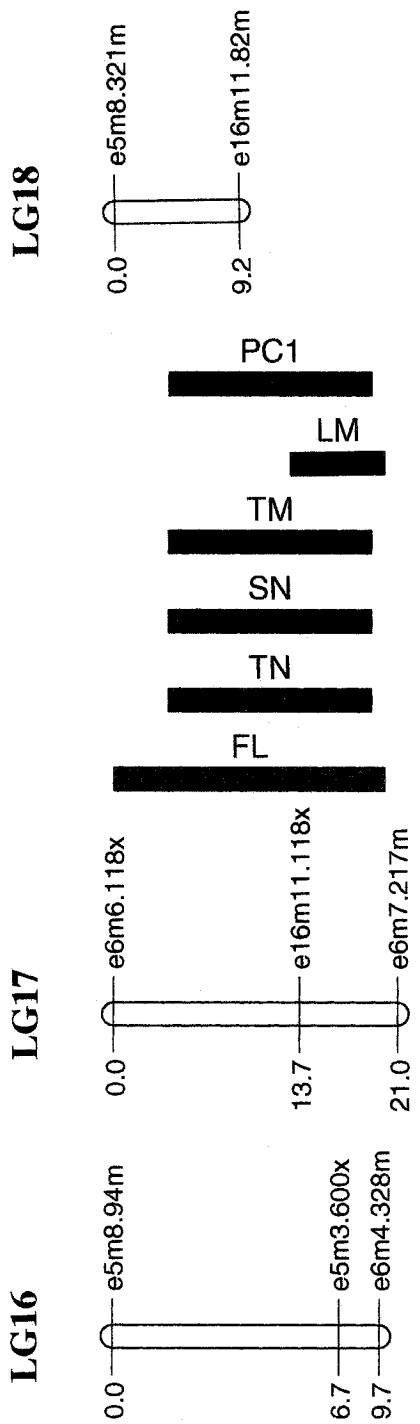


Figure 5.1 continued.

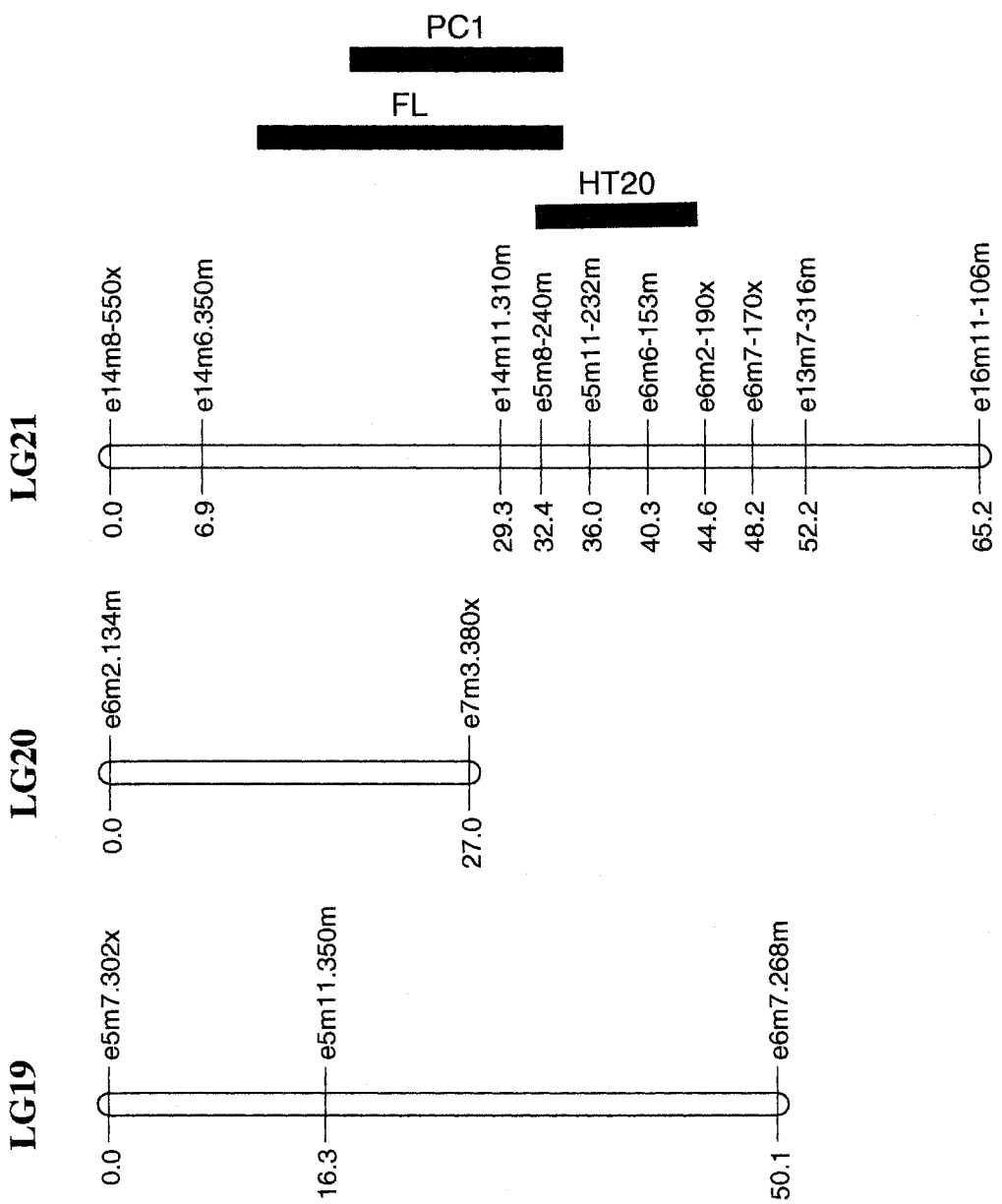


Figure 5.1 continued.

account for 4.2% and 15% of the variation in fitness respectively. The effect of the xeric alleles at these QTL was to increase spikelet production, with an increase of 28 on LG1 and an increase of 54 on LG6. Approximately 45% of the variation in fitness among family means could be explained by this five QTL model.

Significant QTL controlling the number of tillers produced were identified on LG6, LG15, and LG17 essentially in the same locations as the QTL for spikelet production (Figure 5.1). The direction of the additive effects of these three QTL were the same for both traits, with the xeric allele of the LG6 QTL increasing the number of tillers and spikelets, the QTL on LG15 and LG17 decreasing tillers and spikelets. An additional QTL for tiller number, not shared with spikelet number, was found on LG8 and decreased the number of tillers by one thereby explaining 4.6% of the variation in tiller number. However, this decreasing tiller QTL was found to co-localize with a QTL identified to increase flowering time (Figure 5.1).

Four QTL were detected for tiller mass mapping to LG7, LG10, and not surprisingly, to the same positions on LG15 and LG17 that have been identified to carry QTL for tiller and seed number. The direction of the additive effect of the xeric allele at each of these QTL was to decrease the mass of the tillers, with a decrease of 1.1 grams for the QTL on LG15 and 0.63 grams for the QTL on LG17. It should be noted that the QTL on LG 15, by itself, was found to explain 27.5% of the variation in tiller mass, making it the most influential QTL found in the entire greenhouse analysis. The tiller mass QTL on LG15 explained 9% of the variation in tiller mass. The locus detected on LG10 mapped to the same position as QTL already identified for a variety of traits including spikelet number, flowering time, and leaf mass (see below). The xeric allele at

this locus was found to increase the mass of the tillers by approximately 0.5 grams. In contrast, the remaining tiller mass QTL located on group 7 was not shared with any of the other traits related to principal component one. Instead this locus, which decreased tiller mass by 0.5 grams, was shared with germination time and maximum height (see below). The CIM model including the four tiller mass QTL explained 42% of the variation among line means for tiller mass.

A total of five QTL were identified for leaf mass, explaining 42% of the variation among line means. Two of the QTL, located on LG10 and LG17, were shared with tiller mass but had additive effects in opposing directions to those for tiller mass, fitting well with the known negative relationship between these traits. In addition, the "decreasing" leaf mass locus on LG6 had antagonistic additive effects on spikelet number, tiller number, and germination time (Figure 5.1). The next QTL, located on LG3, was only shared with maximum height and had antagonistic additive effects, with the xeric allele increasing leaf mass by 0.7 grams and decreasing maximum height by 3.7 centimeters. The last leaf mass QTL was mapped to a section of LG 14, a group that otherwise does not harbor QTL for any other trait.

Principal component two and mass: total mass, and dry mass at 60 days.

Only three traits, dry mass at 60 days, leaf mass and total mass, were identified to have loadings greater than 0.50 on principal component two (Chapter Three). Interestingly, only two QTL could be identified for the composite trait total mass, despite

numerous QTL being detected for its two component traits. However, the total mass QTL did map to positions estimated to have QTL for leaf mass (LG3) and tiller mass (LG15). The additive effects of the xeric alleles on total mass at these loci are in the same direction as the QTL effects for the two other mass traits, implying an overall positive relationship between total mass and its constituent parts. Even with only two QTL the model was found to explain nearly 23% of the variation in total mass among the F6 line means.

No significant QTL could be detected for dry weight at 60 days despite identifying several significant single marker associations (Appendix One).

Principal component three and height: maximum height

The trait that loaded most heavily on principal component three (PC3), maximum height (Chapter Three), was found to share one QTL with tiller mass and germination time on LG7, and one QTL with leaf mass on LG3. The QTL on LG7 had a xeric additive effect of decreasing maximum height by 4.9 centimeters, which was in the same direction as the additive effect on tiller mass, but in the opposite direction of the additive effect for germination time. Similarly, the xeric effect at the QTL on LG3 was to decrease maximum height by 3.6 centimeters but was found to increase leaf mass. The other QTL identified for maximum height was located on LG5 and was not shared with any of the other greenhouse traits. The R^2 of the three QTL maximum height QTL model was estimated to be 0.21.

The only other trait that had a loading on principal component three greater than 0.40 was height at 20 days, however as shown below it was known to be more tightly correlated with principal component four and did not share any QTL with maximum height.

Principal component four and early growth: Germination time, Height at 20 days:

Only the two traits associated with early growth were identified to load heavily on the last principal component, component four (PC4), in the greenhouse environment. Three putative QTL for germination time were identified, located on LG6, LG7 and LG15 respectively (Figure 5.1; Table 5.1). The additive effect of the xeric alleles of the loci on group six and seven were found to increase germination time by 0.10 days, with each QTL explaining approximately 6% of the variation among family means for germination time. The remaining QTL on LG15 was found to decrease germination time by 0.05 days and explained 3% of the variation in germination time. Overall, the linear model fitting these three QTL could explain 14% ($R^2 = 0.137$) of the variance among family means for germination time.

Loci having significant effects on height at 20 days were identified on LG8, LG10, LG11, and LG21 (Figure 5.1). The effect of the xeric allele of the most significant of these QTL, located on LG8, was estimated to reduce early height by 0.3 centimeters thereby accounting for 8% of the variation in early height. Similarly, the xeric effect of the QTL on group 10 and the QTL on group 11 was to reduce height at 20 days, however

each could only account for 4-5 % of the variation in early height (Table 5.1). The total R^2 of the QTL model for early height indicated these four QTL can explain approximately 20-25 % of the variance among F6 family means.

Additive X Additive epistasis for fitness related traits:

Significant marker interactions were identified for all of the traits measured in the greenhouse (Table 5.2). However, the number of interactions for germination time (8 interactions) and flowering time (12 interactions) were both less than the number expected to arise due to chance indicating they may be false positives. Similarly, the number of interactions for height at 20 days, spikelet number, PC1 and PC4 were marginally above the chance level. In contrast, the traits that load heavily on PC2 (mass related) and PC3 (height) had at least double the number of significant interactions expected by chance. Likewise, the fitness trait number of tillers had 20 more interactions than expected by chance alone.

Out of the 416 significant marker interactions only three were deemed significant at the Bonferroni adjusted 10% significance level (Table 5.3). Only one of these three was still significant at the adjusted 5% level. Each of these interactions was related to total mass and/or the mass principle component PC2. All three marker pairs describe the same interacting portions of LG10 and LG14. The effect (aa) of having alleles from the same parent at both loci (i.e. AABB and aabb) was to increase total mass by 0.7 grams and increase the component score on PC2 by 0.317.

Table 5.2. Number of two-way marker interactions for each trait measured in the greenhouse. Number of significant interactions and number expected by chance are not corrected for nonindependance of statistical tests. Bonferroni adjusted thresholds were calculated as in Cheverud (2000).

Trait	Number of Sig. Interactions	Number expected by chance at p~0.0015	Number Sig. at p< 3.8*10-5 (Bonferroni adjusted alpha = 0.1)	Number Sig. at p< 1.9*10-5 (Bonferroni adjusted alpha = 0.05)
GM	8	13	0	0
HT20	15	13	0	0
DM60	28	13	0	0
FL	12	13	0	0
MH	33	13	0	0
TN	33	13	0	0
SN	18	13	0	0
TM	28	13	0	0
LM	38	13	0	0
TOTM	62	13	2	1
PC1	15	13	0	0
PC2	61	13	1	0
PC3	50	13	0	0
PC4	15	13	0	0

Table 5.3. Epistatically interacting pairs of marker loci significant at the Bonferroni adjusted 10% significance level. LR is the likelihood ratio of an epistaitc model versus additivity. Probability value of significant interaction effect based on 250 000 permutations. aa = effect of having the parental genotypes at the marker loci.

Trait	Locus A (group-marker)	Locus B	LR	p-value	aa
TOTM	e13m6.288m (10-7)	e6m2.327x (14-2)	13.3965	0.000008	0.702
	e14m8.242m (10-8)	e6m2.327x (14-2)	10.51564	0.000024	0.661
PC2	e6m2.500x (10-10)	e4m11.500x (14-1)	10.26191	0.000028	0.317

Discussion:

The purpose of this chapter was to combine trait and genotype data in order to map main effect and epistatic QTL for fitness related traits in the greenhouse environment. In particular, the main motivation for this study was to determine if QTL for genetically correlated traits actually colocalize to the same portions of the genome and whether or not the QTL effects can explain the genetic covariation among pairs of traits.

Single marker analysis and the Hybrid index:

Preliminary single marker analysis, carried out largely before the molecular map was constructed, using simple linear regression identified significant marker trait associations for all traits spread more or less across all linkage groups (Appendix One). This suggested that the portion of the genome that has been mapped contained at least some of the genes responsible for the known genetic differences among the F6 recombinant inbred line population. The sign of the regression slope of some of the single marker tests also indicated that alleles originating from the same parent could have plus or minus effects on the same trait. This was the first indicator that there was the presence of some complementary gene action in this system, as predicted from the quantitative trait distributions observed in Chapter Three.

In addition, no association was found between fitness (or any of the other traits) and the proportion of xeric alleles (i.e. the hybrid index) carried by each F6 line. This may indicate there are relatively few loci underlying most traits. However, this could also

mean that performance was not simply a function of how much of the more fit parent's genome that a line contained, expected to be the mesic parent's genome in the well watered greenhouse environment (Chapter Three), but rather the specific combination of genomic regions inherited from both parents.

QTL for fitness related traits and the genetic basis of transgressive segregation in the F6 generation:

Significant and/or suggestive QTL were found for every trait except dry mass at 60 days. Generally more QTL were detected for the traits with the higher broadsense heritabilities (e.g. spikelet number, leaf mass) however the mean number of QTL per trait (3.5 QTL, excluding principal components as traits) fell short of that observed in the studies reviewed in Chapter Two (an average of 8.7 QTL for significantly correlated traits). This is perhaps not unexpected since it is known that molecular map does not include the entire genome and likely excludes some loci from detection. However, given that there was a rather large portion of the map devoid of any QTL, it is likely that most of the loci affecting trait variation that are situated in the mapped portion of the genome have been identified. Similarly, the use of principal component scores to condense the multivariate trait data into a few major axes did not result in detecting additional QTL. For a more detailed discussion of map completion in relation to how well the QTL analysis explains the trait variation see Chapter Seven.

Overall, a qualitative assessment of the signs of the additive effects of the identified QTL provides an explanation for the large amount of transgressive segregation observed within the F6 population. At least one QTL for each trait was found to have additive effects in the opposite direction (i.e. antagonistic) than that expected from the difference between the parental trait means (Chapter Three). Thus, the parental ecotypes contain complementary genes and by recombining the existing alleles at these loci it is possible to create novel genotypes that exhibit more extreme phenotypes than either parent. For example, by combining the increasing alleles from the mesic parent at the QTL for spikelet number on LG15 and LG17 with the xeric alleles at the other spikelet QTL, a shift in phenotype of approximately 90 spikelets is possible. Clearly a rather minor rearrangement of alleles following a hybridization/recombination event in this system can accomplish a substantial change in fitness, and related traits simultaneously (see below), thus demonstrating the potential importance of such a mechanism in adaptive evolution. However, having a mix of plus and minus alleles within one parental genotype is by no means unique to this study. In a review of over 3000 QTL from 96 studies, above and below the level of species, Rieseberg *et al.* (2003) found that upwards of 63% of traits had at least one opposite sign QTL, indicating the genetic architecture responsible for producing transgressive segregants may be quite common.

Co-localizing QTL for fitness:

Following the results of Chapters Two and Three, there was an expectation that the traits that were highly correlated with fitness in the greenhouse should have QTL that map to the same genomic regions. As in Chapter Two I refer to colocalizing QTL as pleiotropic loci. While they might not be pleiotropic in the strict sense, I justify using the term here for two reasons. First, because the F6 inbred line population was derived from a cross between the mesic and xeric ecotypes, all linkage disequlibrium from unlinked loci was broken up by recombination, it can be inferred that any genetic covariance among traits in the F6 generation must be due to true pleiotropy or linkage disequilibrium from linked loci (Lynch and Walsh, 1998). Second, given the multiple generations (i.e. opportunities for recombination) needed to propagate the RILs, it is also likely that some trait covariance due loosely linked QTL will have been broken up by recombination, leaving only tightly linked loci contributing to the genetic correlation among traits (see Haldane and Waddington, 1930).

Several linkage groups were identified to carry pleiotropic QTL for the major fitness related traits, known to load heavily on principal component one (PC1), including flowering time, tiller number, and spikelet number. In particular, linkage groups 15 and 17 were found to have clusters of QTL, with overlapping confidence limits on their genomic positions, that clearly show the negative relationship between increased flowering time and decreased reproductive output in spikelets and tillers (Figure 5.1). A similar pattern can be seen on LG8 with a QTL for increased flowering time associated with a decrease in fitness through a reduction in tiller number. Interestingly, this negative

relationship was also observed on LG10 but in the opposite direction, with the xeric alleles decreasing flowering time and increasing fitness through spikelet number and tiller mass. This is evidence that even though alleles from each parent can be arranged into transgressive phenotypes, a negative relationship among traits, and hence an evolutionary constraint, is maintained by the pleiotropic effects of these loci. In the context of an adaptive landscape, these pleiotropic QTL would limit the possible routes a population could take toward an optimum phenotype.

Constraints were not universal across all loci; there were QTL identified that did not have significant effects on both flowering and spikelet number. For example, a flowering time QTL was located on LG21 and was not associated with reproductive output. Likewise, a QTL for an increased number of spikelets on LG1 and spikelets and tillers on LG6 had no significant effect on flowering time. From these results it is likely that flowering and reproductive output cannot evolve independently from each other, however there appears to be the presence of loci for the individual traits that do not contribute to any covariation among traits, perhaps providing a mechanism to alleviate the evolutionary constraint posed by the negative correlation. However, a caveat to this interpretation is whether or not some of these seemingly "single trait" QTL have undetectable, but true, effects on several traits. For example, the spikelet and tiller number QTL on LG6 has some effect (negative) on flowering time (Appendix Two) but does not reach significance. Likewise, the flowering QTL on LG21 has some effect (positive) on spikelets and tillers, but neither effect is significant (Appendix Two). It is perhaps tempting to infer true pleiotropic effects in these instances (and below) since their additive effects among the traits are in the expected direction, however it is

unknown if these effects are real QTL or simple "background noise". Thus I take a conservative approach and limit the discussion of loci contributing covariance among traits to QTL that are at least significant at the chromosome wide level (see Chapter Seven for further discussion).

It has been suggested that limitations in the accumulation of resources, and genetic variation in the utilization of resources, can lead to tradeoffs among life history traits (Roff, 2002; Mitchell-olds, 1996). In this study, allocation of biomass to vegetative growth, measured as leaf mass, and allocation of biomass to reproductive structures, measured via tiller mass, were presumed to be in a tradeoff, indicated by the strong negative genetic correlation ($r = -0.71$) between the traits (see Chapter Three for details). Two genomic regions, already known to have effects on flowering and fitness, were identified where QTL for each of these mass traits co-localized. On LG10 a decreasing leaf mass QTL was associated with an increasing tiller mass QTL, however the confidence limits of QTL position for each QTL barely overlapped. A second genomic location (LG17) had QTL for leaf and tiller mass but the signs on the additive effects were in the opposite direction as on LG10. Again it appears that while it is possible for one parent to harbour alleles of different signs for the same trait, a negative relationship among traits can be maintained via antagonistic pleiotropic QTL even if the parental alleles are rearranged into transgressive genotypes.

While the QTL on LG10 and LG17 provide a genetic explanation for the tradeoff between growth and reproduction, there were a total of five other QTL for leaf and tiller mass that did not have significant effects on both traits. This demonstrates that a considerable amount of variation in each trait was not associated with the tradeoff

(However, as stated above, there may be some undetectable pleiotropic effects which may contribute trait covariance to some degree). Interestingly, two of the five loci co-localize with the only two QTL identified for total mass. In each case the xeric additive effects are in the same direction, with the QTL on LG3 increasing leaf mass and total mass and the QTL on LG15 decreasing tiller mass and total mass. Because these loci generate positive covariance between total mass and its constituent parts, putatively independent from the tradeoff, it is possible these loci are involved in the acquisition of resources (Houle, 1991; Chapter Two). If this is the case, then the antagonistically pleiotropic loci underlying the tradeoff between leaf and tiller mass described above would certainly represent loci controlling resource allocation (Houle, 1991; Mitchell-Olds, 1996).

The fact that the mass allocation QTL were not detected for total mass by itself illustrates the difficulty in identifying loci underlying some evolutionary constraints. In this case the antagonistic QTL effects on tiller and leaf mass roughly cancel out, leaving no net effect on total mass. This serves to reduce the amount of genetic variation in total mass as any additive variance in total mass gets "swamped" by the negative covariance between tiller and leaf mass. This provides an explanation as to why the heritability of total mass is much less than its component traits (Chapter Three). Therefore, without examining the component traits separately it is likely the mass allocation QTL would have gone undetected.

Interactions among marker loci and the possible role of epistasis in releasing variation in the F6 generation:

In this study I took two approaches to identify interacting loci in order to infer whether epistasis was an important source of variation in the novel greenhouse environment. First, following the method of Cheverud (2000) I tested all pairwise combinations of marker loci with each trait. In contrast to the results of Kim and Rieseberg (2001), I found many more significant interactions than could be explained by chance alone, indicating the presence of true epistatic interactions for most traits (Table 5.3). In particular, growth traits such as dry mass at 60 days, leaf, tiller, and total mass all had double the number of significant tests than expected by chance. This was surprising because there was no net epistatic effect detected for any of these mass traits, as analyzed in Chapter Three. This suggests that there must be plus and minus epistatic loci for these mass traits operating in the parental ecotypes, that essentially cancel out when summed across all loci. If this is the case then the breakup of certain interactions through recombination may allow for a release of variation over and above that possible by rearranging additive alleles, thus providing a means of producing a wider array of transgressive phenotypes. However, the pattern is less clear for the major fitness traits, flowering time and spikelet number, and the fitness axis of variation (PC1), each of which had only a few more interactions than expected by chance.

The second approach was to identify individual interactions among loci that could be declared significant at the genome wide level (i.e. highly unlikely they are false positives), and observe both direction and magnitude of the nonadditive effects. This

proved to be an arduous task, resulting in only three interactions (out of 416) significant at the adjusted 10 % significance level. While infrequent, this number of significant interactions was comparable to both Kim and Rieseberg (2001) and Cheverud (2000) who found nine and three epistatic loci respectively. In each case (this study included), the highly significant interactions usually described the same interacting segments of each genome. For example, I found all three interacting pairs of loci were situated on LG10 near marker number eight and LG14 near marker two; each pair of loci had effects on total mass either directly or through the mass principle component (PC2). Interestingly, while these two locations did not have QTL identified for total mass or PC2, they were regions known to carry QTL for leaf mass (LG10, LG14) and tiller mass (LG10).

The epistatic effects on total mass at the interacting loci were found to be as large as the additive effects at the main effect QTL for total mass, and incidentally, larger than the additive effects at the tiller and leaf mass QTL at this position. This is evidence of the type of interaction effect needed for epistasis to help shape the adaptive landscape in Wright's shifiting balance theory.

Chapter Six:

Genetic Correlations and QTL Analysis for Fitness in a Reciprocal Transplant Experiment

Introduction:

In previous chapters I have presented evidence, at the phenotypic (Chapter Three) and molecular level (Chapter Four), that substantial genetic variation exists between the mesic and xeric ecotypes of *A. barbata* and among their inbred progeny for a suite of quantitative traits, measured in a greenhouse environment. The broadsense heritabilities for growth and reproduction in the F6 generation were generally high (>50%), indicating a substantial release of genetic variation upon crossing the parental ecotypes. Clear evidence of transgressive segregation was also found for the fitness related traits in the F6 lines suggesting the presence of complementary gene action in the parental ecotypes. There were obvious constraints to the range of possible phenotypes however, with many traits being highly genetically correlated in the line cross progeny. In fact, fitness and related traits such as flowering time and mass allocation were found to vary along one major axis of variation, indicating they function as an integrated unit. I made efforts to identify the individual loci underlying fitness related traits and have identified QTL for fitness that co-localize to the same genomic regions as QTL for related traits, such as flowering time and mass, suggesting a common genetic basis for variation in these traits due to pleiotropy or tightly linked loci.

The main caveat to interpreting these results in an evolutionary context is the unknown relationship between performance in the greenhouse and performance in the

two parental environments. The goal of this chapter is to present results of a field experiment where the parental ecotypes and the F6 line population are grown in a common garden, within each native parental environment, in order to address questions about adaptive population divergence of the parental ecotypes. Specifically, the questions of interest are 1) Does fitness correlate across environments? 2) Is there evidence for coadapted gene complexes, and do they affect fitness across environments? 3) Is it possible to detect QTL for fitness in the natural environments? 4) Do QTL for fitness in the parental environments map to the same location as those detected in the greenhouse?

Very few studies have addressed these and similar questions by probing the genetic basis of growth and fitness in a field reciprocal transplant QTL mapping experiment. Instead, most multi-environment QTL mapping studies involve growing domesticated species (e.g. cereals) in environments that are novel in relation to that inhabited by their natural progenitors. For example, Schon *et al.*, (2004) describe a study where inbred maize lines were grown across 19 environments. Likewise, DeKoeyer *et al.* (2004) presented a QTL mapping study conducted on a cultivated oat (*Avena sativa)* population that was grown in 13 field environments and measured for 16 traits. In both cases the authors were able to detect QTL for a suite of yield related traits and, in the case of DeKoeyer *et al.* (2004), also found substantial epistatic and QTL X environment interactions, thus demonstrating the possibility of uncovering a complex genetic architecture even in uncontrolled field environment. The utility of such studies is clear, both from a practical and methodological viewpoint, but since the field environments do not represent the conditions that their ancestral genpotypes adapted to, linking the

observed genetic architecture with that responsible for the origins of the species/ecotype becomes problematic.

Recently however, Hawthorne and Via (2001) conducted a transplant experiment by growing ecotypes of pea aphid (*Acyrthosiphon pisum pisum)* on two native host plants, namely alfalfa and clover. They succeeded in identifiying loci that had antagonistic fitness effects in the two different host environments. Specifically, they found that the loci for increasing acceptance of alfalfa as a host plant, and fecundity on alfalfa, generally co-localized with loci for decreasing acceptance and fecundity on clover, fitting well with the known negative genetic correlation among fitness traits between host plant environments (Via and Hawthorne, 2001). Likewise, Jiang *et al.* (1999) describe identifying loci for fitness that map to the same genomic position but have opposite additive effects in environments native to several locally adapted ecotypes of tropical maize. In contrast, Verhoeven *et al.* (2004) analyzed the performance of an F3 population of wild barley (*Hordeum spontaneum*) in contrasting native environments in Israel, and found that most QTL for fecundity related traits were detectable only in one environment. Interestingly, the QTL that did map to the same genomic location in both environments did not have opposing effects on fitness, suggesting performance in one environment did not preclude performance in the other environment (Verhoeven *et al.*, 2004).

Materials and Methods:

Experimental Design:

In the fall of 2002 a field experiment was conducted at two sites in California. The first site, located at the Hopland Research and Extension Centre in Hopland, was situated in a shallow valley and was designated the "mesic" site (Hutchinson, 1982). Approximate rainfall accumulation at the Hopland site was 967 mm and 964 mm for 2002 and 2003 respectively. Native *A. barbata* at Hopland were entirely hairy stemmed, a trait indicative of the mesic genotype (Clegg and Allard, 1972; Jain and Rai, 1980). The second site, located at the Sierra Foothills Research and Extension Centre near Brown's Valley in the central valley of California, and was designated the "xeric" site. Rainfall was lower at this site with an approximate accumulation of 627 mm in 2002, and 784 mm in 2003. The field plot at Sierra was situated on the crest of a hill in a cattle pasture that contained patches of mainly xeric genotypes (glaborous stemmed) with intermittent patches of mesic genotypes.

At each site three randomized complete blocks were set up using individuals from the parental, F2 and F6 generations of the slender wild oat population previously analyzed in the greenhouse (see Chapter Three; Johansen, 2004; Latta *et al.*, 2004). Specifically, 50 individuals from each parental ecotype, 100 F2 individuals and 889 F6 individuals from 188 lines (three seeds from each of 163 lines, 16 seeds from the remaining 25 lines) were randomly assigned to the three blocks with each generation/line represented at least once in each block. Data from the F2 generation was used to assess early versus late hybrid performance by Johansen (2004) and was not analyzed in this

study. Seeds were prepared for germination by first removing the husk and placing them in petri dishes lined with moistened filter paper. The petri dishes were placed in a refrigerator for three days to allow the seeds to imbibe water, then were placed in the dark at room temperature for two days allowing uniform germination. Prior to being placed in the field plots, the seedlings were sown in individual plastic cones (Stuewe and Sons Inc. Oregon) labeled with family/line number filled with moistened germination soil mix (Sunshine Mix, SunGro Vancouver, Canada). The plastic cones had been previously cut in half such that the bottoms could be removed upon planting, so as to allow the seedling's roots to grow in the native soil during the experiment. The seedlings were transported to the field sites and planted in the blocks, each of which was 1.5 meters by 25 meters in size. The three blocks at each site were completely enclosed by fencing to exclude grazing animals, but otherwise the surrounding vegetation was undisturbed. Once the sites were planted the experiment was left unattended for seven months to allow plants to flower and reach senescence.

In May 2003 each site was harvested, during which maximum height and total spikelets produced were recorded. The above ground portion of each plant was then bagged and dried at 55 C for two weeks. The dried material was then weighed to get an estimate of total above ground biomass accumulation.

Statistical Analysis:

Analysis of the quantitative trait data, collected for the xeric and mesic parental ecotypes and their inbred progeny, were carried out in a manner similar to that used in the

greenhouse experiment described in Chapter Three. Data were coded as missing if there was evidence the plants were broken off during the experiment, or there was uncertainty about whether an individual plant would have matured existing spikelets. This is a more conservative approach than that used by Johansen (2004), and was chosen because the uncertainty of the data can be accounted for in the QTL analysis by permutation testing. All other statistical analysis was carried out using SPSS version 11.5. Briefly, differences between the parental ecotypes for fitness and the fitness related traits at the two field sites were tested using multivariate ANOVAs, with *ecotype* (Fixed), *block* (Random) and *family within ecotype* (Random) as main factors and *ecotype*block* and *family within ecotype*block* as the interaction terms. If the interaction terms were not significant they were removed from the ANOVA model and the analysis was repeated. Variation among families within each parental ecotype for each trait was tested for significance in separate univariate ANOVAs, and quantified as the broadsense heritability (H^2). Heritabilities were estimated as the proportion of variation attributed to among family differences, calculated by the REML estimation function of SPSS.

In order to test for departure from an additive inheritance model, the mean values from the parents and the F6 generation were used in the weighted least squares (WLS) procedure (see Chapter Three; Kearsey and Pooni, 1996) designed to estimate simple quantitative genetic parameters. First, a model including only the trait mean (m) and additive effect ([a]) was fit to the generation means data for each trait. A departure from additivity was detected if the model residual was significantly different from zero. In this instance a new model was fit which included an additive by additive epistatic effect ([aa]).

Differences among the F6 lines were tested using separate univariate ANOVAs for each trait at both field sites, with *line* (Random) and *block* (Random) as main factors and *line*block* as the interaction term. Broadsense heritabilities were calculated as described above for the parental families. Relationships among the traits were quantified by estimating the genetic correlations for all pairwise combinations of traits. Genetic correlations were calculated by partitioning the among line and residual sums of squares and cross product matrices (SSCP) and equating each to their expectations and fitting the values in the standard formula for the bivariate correlation (Lynch and Walsh, 1998). In order to compare genetic correlations across environments the line means for each trait in all environments, including the greenhouse (Chapter Three), were combined into one data set and the product moment correlation was calculated. In addition, the line means for all traits in all environments (in all generations) were used in a principle components analysis (i.e. an extension of the analysis conducted in Chapter Three) using the Factor function of SPSS. Components were retained if they had an eigenvalue greater than one, and a Varimax rotation was employed to generate the factor loading matrix.

Single marker analysis and Hybrid Index:

In order to test whether genome composition correlates with fitness in the field the proportion of xeric alleles, calculated for each line in Chapter Five, was correlated to the mean trait value for each F6 line. Significance of association was calculated using a randomization test with 1000 randomizations using the RESAMPLING program (Howell, 2000).

Single point QTL analysis was conducted for all marker loci across all the field traits using the simple linear regression method described in Chapter Five. Allelic effects are given as the effect of the xeric allele. The single markers tests were only used as a rough guide to QTL identification so no correction for multiple tests was applied.

QTL mapping using field data:

The QTL mapping process employed in Chapter Five was used to identify regions of the oat genome controlling fitness in the field environments. Trait means for each F6 line were used to map QTL for seed number, total mass and maximum height at each of the field sites. Component scores from the principal component analysis were added as separate traits. Composite Interval Mapping (CIM) was conducted for each trait individually using QTL Cartographer (Wang *et al.*, 2004) with the number of markers used as cofactors determined via stepwise regression. Mapping parameters were chosen to scan the linkage map every two centiMorgans with a cofactor window set to 15 centiMorgans. Significance thresholds for declaring the presence of a QTL for each trait were determined using permutation tests with 1000 randomizations (Churchill and Doerge, 1994).

All pairwise combinations of marker loci were tested for significant interactions with each trait using the Epistat computer package (Chase *et al.*, 1997) described in Chapter Five. The analysis was carried out as in Chapter Five, with an initial significance threshold of $p = 0.0015$. All pairs of significantly interacting loci were used in randomization tests with 250 000 permutations to obtain approximate p-values for each

interaction. The modified Bonferroni method described in Chapter Five was used to adjust the 10 % and 5 % significance thresholds to account for multiple testing. Any interactions found to exceed the adjusted significance thresholds were then fit to a fixed effects two way ANOVA model in SPSS 11.5 to allow estimation of epistatic effects.

Results:

Genetic differences within and between xeric and mesic ecotypes:

Analysis of the trait values of height and total mass at the two field sites revealed significant differences between the parental ecotypes (Table 6.1). However, the differences were not detected until the effect of block and all the interaction factors, which were all non significant sources of variation, were removed from the analysis (Tables 6.2a and 6.2b). Overall, the mesic ecotype grew taller and accumulated more biomass than the xeric ecotype but no detectable difference was found between the ecotypes for spikelet production at either field site (Table 6.1). Separate ANOVAs on each trait did not detect any significant variation among any of the families within either parental ecotype. Estimates of the broadsense heritability within ecotypes (H^2) for height, mass and spikelet number at each field site were either negative (shown as zero) or very close to zero with confidence limits including zero (Table 6.1).

Quantitative genetic parameter estimation via weighted least squares:

At Sierra highly significant additive effects were detected for height (Table 6.3, $p < 0.0001$), with the xeric additive effect decreasing height by five centimeters. No significant additive effects were detected for fitness or total mass. However, significant departure from a simple additive model was detected for height and fitness, with both epistatic effects increasing the trait values in the parental ecotypes (Table 6.3). In fact, the non additive effect on height was over double the magnitude of the additive effect on

Table 6.1. Means, ranges and variance components for traits measured in the field environments. Variance components (and coefficients of variation, CV) are given for genetic, block and environment sources. (ht.se = height at Sierra, spike.se = spikelet number at Sierra, mass.se = mass at Sierra, ht.hop = height at Hopland, spike.hop = spikelet number at Hopland, mass.hop = mass at Hopland).

	ht.se	spike.se	mass.se	ht.hop	spike.hop	mass.hop
Means						
Xeric	92.280	23.840	1.090	81.120	11.440	0.400
Mesic	102.230	26.040	1.690	102.230	10.850	0.770
F6 lines	86.500	17.460	0.890	69.480	7.480	0.440
Range						
Xeric	(48-129)	(0-126)	(0.12-5.73)	(10-120)	(4-82)	(0.11-0.78)
Mesic	(0-134)	(0-81)	(0-5.73)	(55-128)	(0-37)	(0.31-2.17)
F6 lines	(0-188)	(0-167)	(0-9.04)	(0-145)	(0-39)	(0.07-1.71)
Xeric families	Variance components					
Var gen.	0.000	0.000	0.006	0.000	0.000	0.002
(CV gen)	0.000	0.000	0.072	0.000	0.000	0.109
Var block	0.000	0.000	0.000	146.998	3.925	0.003
(CV block)	0.000	0.000	0.000	0.149	0.173	0.129
Var env.	797.707	435.850	1.031	1829.642	138.355	0.053
(CV env)	0.306	0.876	0.931	0.527	1.028	0.574
H^2 (Xeric)	0.000	0.000	0.006	0.000	0.000	0.035
95% conf.	-0.105, 0.028	-0.088, 0.161	-0.068, 0.310	-0.090, 0.140	-0.062, 0.344	-0.055, 0.379

Table 6.1 continued.

	ht.se	spike.se	mass.se	ht.hop	spike.hop	mass.hop
Mesic families						
Var gen.	0.000	0.000	0.000	148.702	5.928	0.029
(CV gen)	0.000	0.000	0.000	0.119	0.224	0.221
Var block	1.006	0.000	0.000	0.000	2.001	0.001
(CV block)	0.010	0.000	0.000	0.000	0.130	0.044
Var env.	632.186	325.836	1.423	1546.622	41.141	0.190
(CV env)	0.246	0.693	0.706	0.385	0.591	0.565
H^2 (Mesic)	0.000	0.000	0.000	0.088	0.126	0.133
95% conf.	-0.105, 0.061	-0.120, -0.074	-0.117, -0.042	-0.032, 0.431	-0.007, 0.506	-0.010, 0.498
F6 Lines						
Var gen.	126.317	29.519	0.071	65.152	0.000	0.003
(CV gen)	0.130	0.311	0.300	0.116	0.000	0.127
Var block	9.929	4.431	0.007	0.000	0.000	0.000
(CV block)	0.036	0.121	0.095	0.000	0.000	0.000
Var env.	339.902	164.178	0.496	985.483	19.417	0.042
(CV env)	0.213	0.734	0.792	0.452	0.589	0.467
H^2 (F6)	0.271	0.152	0.126	0.062	0.104	0.069
95% conf.	0.349, 0.201	0.234, 0.094	0.210, 0.074	0.176, 0.000	0.192, 0.016	0.180, 0.001

Table 6.2a. Differences between ecotypes tested using nested multivariate ANOVA. Table shows multivariate between factor F ratio statistics and degrees of freedom. Items shown in bold are significant at $p<0.05$.

Source	d.f.	height.se	seed.se	mass.se	height.hop	seed.hop	mass.hop
Ecotype	1	**7.85**	0.46	**6.90**	**15.74**	0.05	**10.26**
Family within Ecotype	9	0.98	0.50	0.68	1.39	0.45	0.57
Block	2	1.55	0.01	0.01	0.09	0.20	0.55
Ecotype by Block	2	0.07	0.83	0.86	1.01	0.61	0.14
Family within Ecotype by Block	18	0.72	0.60	0.64	0.99	0.31	0.50
Error d.f.		67	67	67	33	33	33

Table 6.2b. Differences between ecotypes tested using a reduced multivariate ANOVA model. Table shows multivariate F ratio statistics and degrees of freedom. Items shown in bold are significant at $p<0.05$.

Source	d.f.	height.se	seed.se	mass.se	height.hop	seed.hop	mass.hop
Ecotype	1	**6.30**	0.18	**6.61**	**20.19**	.024	**16.88**
Error d.f.		98	98	98	63	63	63

Table 6.3. Weighted least squares estimates of generation mean (m), additive ([a]) and epistatic ([aa]) effects in the field environments. Values shown in bold are significant at $p<0.05$. Additive effects are in relation to the xeric genotype. Epistatic effects indicate the effect of having the parental genotypes (i.e. AABB and aabb). SE = one standard error.

Trait	m	SE (m)	a	SE(a)	aa	SE(aa)
height.se	**86.49**	0.87	**-4.97**	1.18	**10.75**	1.47
spike.se	**17.46**	0.70	-1.10	1.16	**7.47**	1.36
mass.se	**0.89**	0.16	-0.30	0.28	0.50	0.32
height.hop	**69.47**	1.17	**-10.55**	1.30	**22.17**	1.75
spike.hop	**7.48**	0.42	0.21	0.99	**3.74**	1.08
mass.hop	**0.44**	0.09	-0.19	0.15	0.15	0.18

height. An epistatic effect causing an increase of 0.5 grams was detected for mass at Sierra, however this was not significantly different from zero at the 5% level ($p = 0.12$).

Similarly, at Hopland a highly significant additive effect was estimated for height, with the xeric effect decreasing height 24 centimeters from the population mean. Again, no additive effects were detected for mass or fitness. Relatively large non additive effects were however detected for height and spikelet production, with both parameters being highly significantly different from zero ($p < 0.0001$). No departure from additivity was detected for mass at Hopland.

Broadsense heritiability and genetic correlations among traits in the F6 generation:

Analysis of the F6 generation shows there is significant variation among the recombinant inbred lines for all three traits including fitness at the Sierra site (Table 6.1). At the Hopland site the F6 lines differed for spikelet production and total mass but not for height. Broadsense heritabilities were found to be low at each site, with the proportion of variation attributed to among line differences ranging from nine to sixteen percent (Table 6.1). Lower 95 % confidence limits for most traits were marginally above zero. The one exception was height at the Sierra site which had a heritability of 27.5 %.

Within each field site the three traits were found to be significantly genetically correlated with each other, when analyzed using the variance components method (Tables 6.4a and 6.4b). At the Sierra field site there were strong positive correlations between larger growth and spikelet production with the largest correlation being between total mass and fitness ($r = 0.925$). A similar pattern of positive genetic correlation was

Table 6.4a. Genetic correlations derived from variance components among traits measured at the Sierra field site. Items in bold are significant at p<0.001

	height.se	seed.se	mass.se
height.se	-	**0.777**	**0.819**
seed.se		-	**0.925**
mass.se			-

Table 6.4b. Genetic correlations derived from variance components among traits measured at the Hopland field site. Items in bold are significant at p<0.001

	height.hop	seed.hop	mass.hop
height.hop	-	**0.797**	**0.498**
seed.hop		-	**1.019**
mass.hop			-

observed for the traits measured at Hopland, with the mass-fitness correlation essentially equating to 1.00.

When the trait correlations were compared across field sites, using the line means method, it was shown that performance at Sierra correlated pooly with performance at Hopland. For the most part the across site correlations were significantly different from zero, however the overlap of variation, measured as the coefficient of determination (R^2), was generally between 10-14 %. Within sites, the family means correlation estimates were found to be lower than the variance component correlation estimates. By comparing the line means of the field traits with the line means of the greenhouse traits (Chapter Three) it was possible to estimate genetic correlations among fitness related traits across environments (Table 6.5). Overall, the traits measured in the field environments including spikelet production did not correlate strongly with any of the traits measured in the greenhouse. There was a weak positive correlation ($r = 0.223$) between number of spikelets at Sierra and spikelets in the greenhouse, coupled with a weak negative correlation between spikelets at Sierra and flowering time in the greenhouse. Vegetative growth, predominantly height at 20 days and maximum height, measured in the greenhouse was also weakly correlated to height measured at Sierra. No pattern of correlations were identified between greenhouse traits and growth and reproduction at Hopland.

Table 6.5. Genetic correlation across field and greenhouse environments estimated from line means. Estimates shown in bold are significant ($p<0.05$) after a Bonferroni correction for multiple tests.

Corr	GM	HT20	DM60	FL	MH	TN	SN	TM	LM	TOTM	HT.SE	SN.SE	M.SE	HT.HP	SN.HP	M.HP
GM	1.00	0.18	-0.01	0.01	-0.09	0.01	0.00	0.05	0.04	0.10	-0.16	-0.14	-0.16	-0.12	0.01	-0.03
HT20		1.00	0.20	0.19	0.18	-0.16	-0.23	-0.10	0.20	0.08	0.18	0.06	0.09	0.01	-0.02	0.10
DM60			1.00	-0.15	0.11	0.09	0.10	0.22	0.14	**0.39**	0.19	0.06	0.09	0.13	-0.01	0.12
FL				1.00	-0.23	**-0.75**	**-0.82**	**-0.81**	**0.64**	**-0.26**	-0.08	-0.15	-0.07	-0.13	-0.08	-0.08
MH					1.00	-0.11	0.21	**0.49**	**-0.39**	0.16	0.20	0.12	0.11	0.12	0.13	0.17
TN						1.00	**0.84**	**0.66**	**-0.45**	**0.29**	0.08	0.18	0.11	0.02	0.12	0.11
SN							1.00	**0.85**	**-0.62**	**0.33**	0.10	0.21	0.09	0.02	0.13	0.12
TM								1.00	**-0.56**	**0.55**	0.16	0.20	0.14	0.07	0.17	0.18
LM									1.00	**0.38**	-0.05	-0.16	-0.07	-0.07	-0.09	-0.05
TOTM										1.00	0.13	0.06	0.08	0.00	0.09	0.15
HT.SE											1.00	**0.66**	**0.66**	0.03	0.17	**0.25**
SN.SE												1.00	**0.91**	0.11	**0.37**	**0.38**
M.SE													1.00	0.11	**0.32**	**0.37**
HT.HP														1.00	**0.34**	**0.33**
SN.HP															1.00	**0.67**

Factor analysis and major axes of variation across environments:

Using factor analysis on the lines means across all environments (greenhouse included), six principal components were identified and together explain 80 % of the variation in all 15 traits (Table 6.6). Four of the components (labeled 1, 2, 3, 4 in this chapter) were identified in Chapter Three and have factor loadings that show an association of these components with the greenhouse traits. The two novel components identified in this chapter (PC5 and PC6) associate with growth and reproduction in only one field environment each, with component five associated with fitness at Sierra; component six at Hopland. These two "field" components did not have significant association with any of the traits measured in the greenhouse. In addition the field components accounted for more trait variation than PC2, PC3, and PC4.

The range of component scores (based on line means) among the F6 lines completely encompassed the ranges of both parental ecotype along PC5 and PC6 (Figure 6.1) indicating transgressive segregation in both parental environments. Generally the F6 lines that had high scores component on PC5 were not the same as those on PC6. The one exception was F6-87 which was ranked in the top five lines at Hopland and Sierra.

Table 6.6. Principal components analysis on trait data collected from the greenhouse and field sites. Varimax rotated component loadings greater than 0.50 are shown in bold. Abbreviations for greenhouse traits are gien in Table 3.1.

Component	1	2	3	4	5	6
Variation explained (%)	27.300	9.100	8.000	7.100	17.800	11.100
GM	0.029	-0.062	-0.074	**0.870**	-0.168	0.021
HT20	-0.272	0.294	0.367	**0.531**	0.143	-0.009
DM60	0.044	**0.726**	0.155	-0.098	0.071	0.029
FL	**-0.904**	-0.045	-0.123	0.091	-0.001	-0.049
MH	0.180	0.064	**0.907**	-0.003	0.084	0.100
TN	**0.877**	0.091	-0.281	0.017	0.080	0.033
SN	**0.953**	0.053	0.042	-0.029	0.071	0.033
TM	**0.864**	0.254	0.342	0.044	0.073	0.085
LM	**-0.670**	**0.555**	-0.410	0.085	-0.038	-0.028
TOTM	0.254	**0.859**	-0.050	0.137	0.042	0.064
ht.se	0.010	0.134	0.148	-0.055	**0.839**	-0.012
spike.se	0.126	-0.030	-0.014	-0.025	**0.918**	0.223
mass.se	0.043	0.027	-0.007	-0.033	**0.932**	0.197
ht.hop	0.000	0.060	0.151	-0.343	-0.099	**0.687**
spike.hop	0.086	-0.049	-0.048	0.126	0.217	**0.836**
mass.hop	0.057	0.099	0.050	0.117	0.280	**0.795**

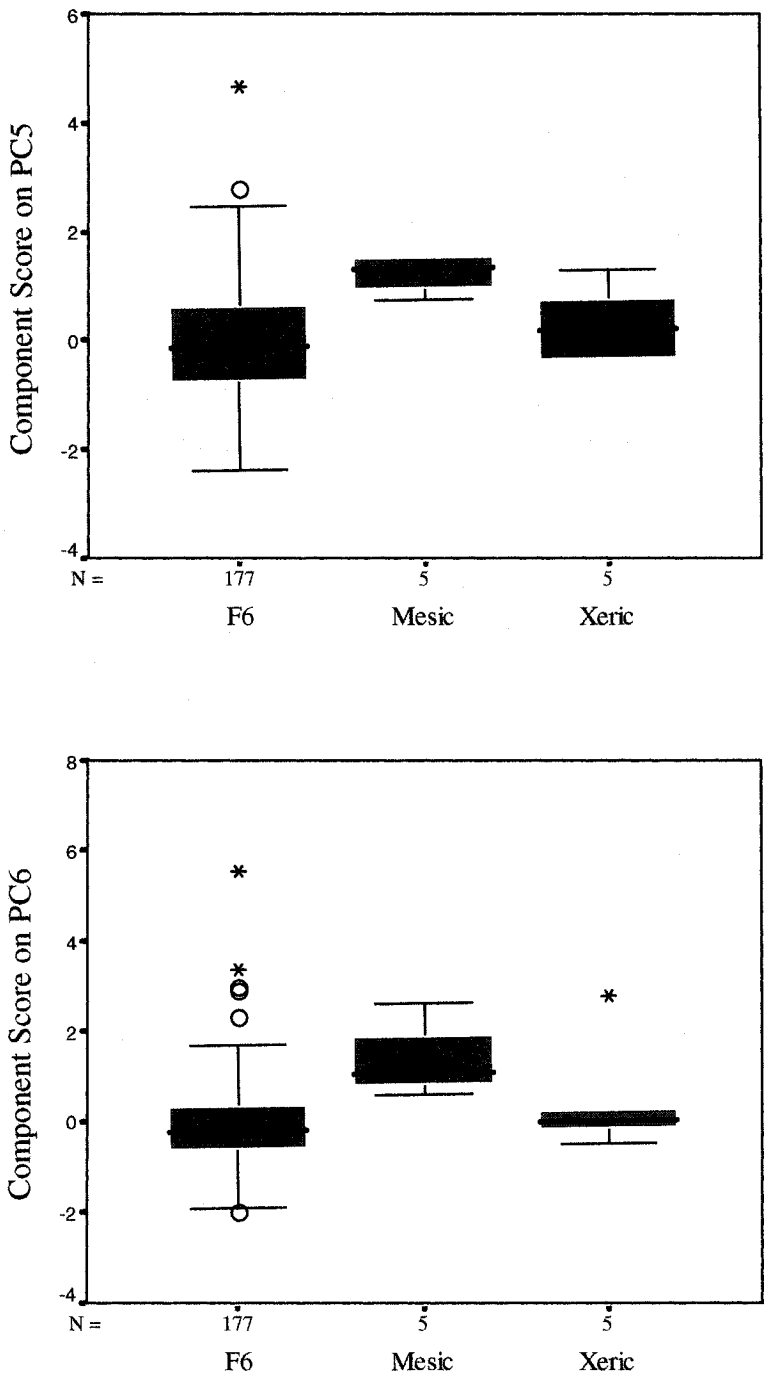


Figure 6.1. Boxplots of component scores along the major axes of variation in the parental environments. a) PC5 (Sierra), b) PC6 (Hopland).

Hybrid index and performance in the field environments:

No significant correlations were detected between the line means for each trait in each environment and the proportion of xeric alleles carried by each line (which was used as a proxy for total of genome composition as in Chapter Four).

Single marker QTL analysis:

Single marker tests were carried out for the six field traits at all loci including the mapped loci and the AFLP markers that were unlinked to the framework map (Appendix One). No significant associations were found among the unlinked marker and the traits at either field site.

Composite Interval Mapping QTL for fitness traits in the field:

Principal component five, growth and fitness at Sierra:

One significant QTL (LR ~> 12; see Appendix Two) and one suggestive QTL (LR > 9.2; see Appendix Two) were detected for spikelet production at Sierra and were located on linkage LG1 and LG7 respectively (Table 6.7; Figure 6.2). The xeric allele of the significant QTL on LG7 was found to decrease spikelet production and could explain approximately 9% of the variation in fitness. The xeric allele of the putative spikelet QTL on LG 1 increased spikelet number and explained 5% of the variation in fitness. A similar pattern was found for mass at Sierra, with one significant QTL mapping to approximately the same location as the Sierra spikelet QTL on LG7 and a suggestive

Table 6.7. QTL detected for fitness related traits measured in two field environments. QTL position is given as the number of the first marker in the interval (Marker), and as cumulative map distance from the first marker of the linkage group (Position). Significant QTL are shown in bold (see Appendix Two for significance values). Additive effects are given as the effect of the xeric allele. %Var is the variation among line means explained by the QTL.

Trait	QTL #	Group	Marker	Position	LR	Additive	% Var
height.se							
seed.se	1	1	9	30.7	11.62	2.02	5.3
	2	**7**	**7**	**50.2**	**15.34**	**-2.59**	**9**
mass.se	**1**	**7**	**7**	**52.2**	**14.46**	**-0.139**	**8.4**
	2	10	10	45.4	10.3	0.116	5
height.hop	1	21	1	1	10.24	-6.51	6
seed.hop	**1**	**18**	**1**	**8**	**11.69**	**0.916**	**5.4**
mass.hop	**1**	**1**	**8**	**25**	**15.43**	**0.044**	**7.4**
	2	7	4	44.5	9.47	-0.034	4.3
PC5	**1**	**1**	**7**	**23.3**	**11.90**	**0.247**	**5.7**
	2	**7**	**1**	**30**	**15.25**	**-0.339**	**10.4**
PC6	1	18	1	6	9.48	0.244	5.5
	2	21	1	2	10.68	-0.251	5.9

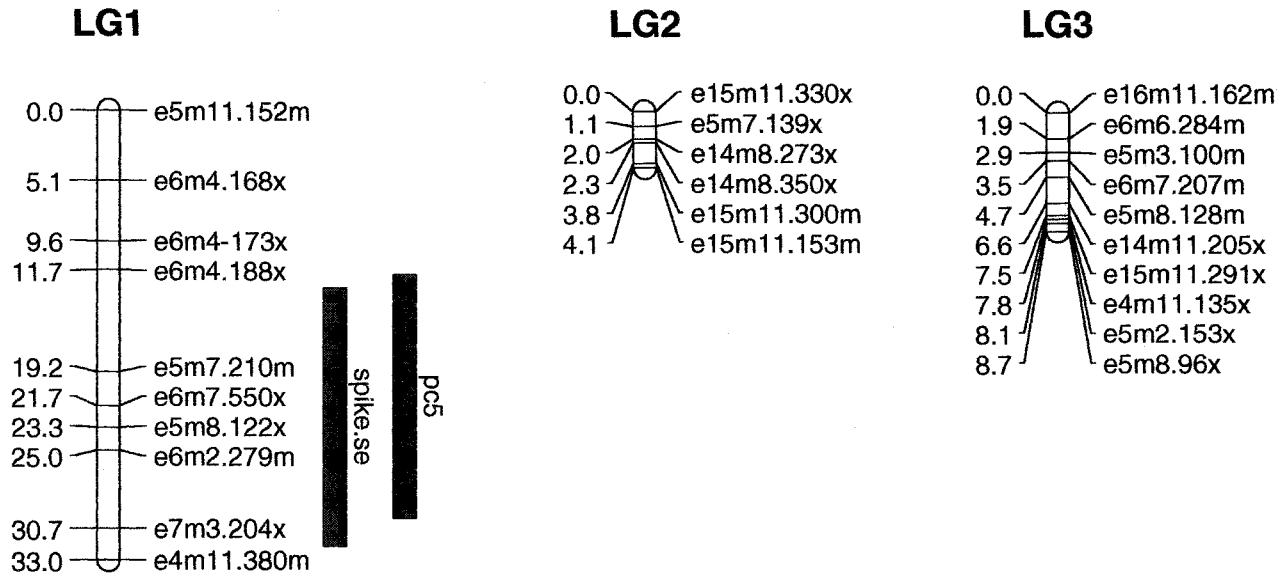


Figure 6.2. Genetic linkage map with positions of QTL for traits measured in two field environments. All QTL that are significant at the chromosome wide threshold are shown (see Appendix Two for further details). One LOD confidence intervals around the QTL position are shown as boxes. Red shading indicates the xeric allele increases the trait value. Blue shading indicates the xeric allele decreases the trait value.

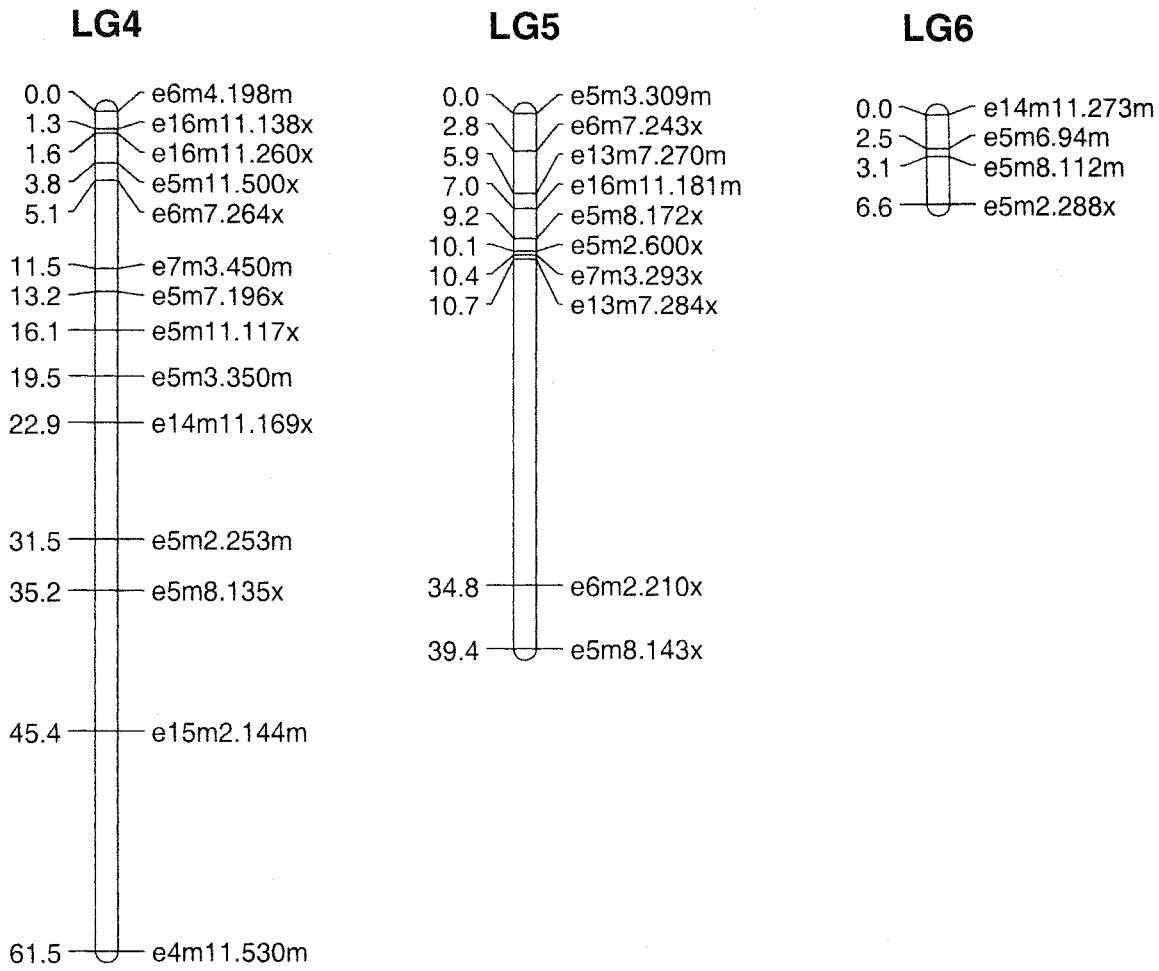


Figure 6.2 continued.

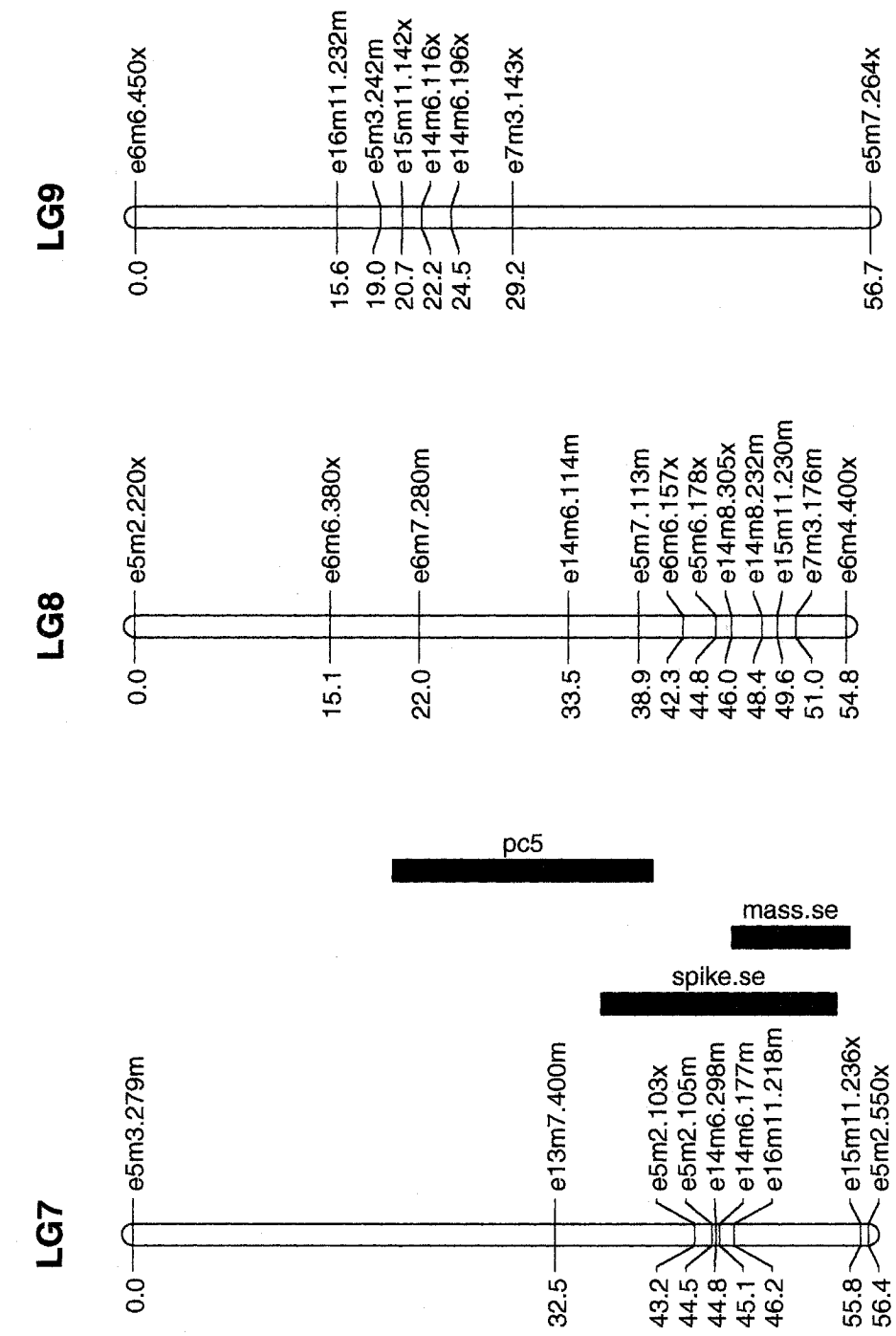


Figure 6.2 continued.

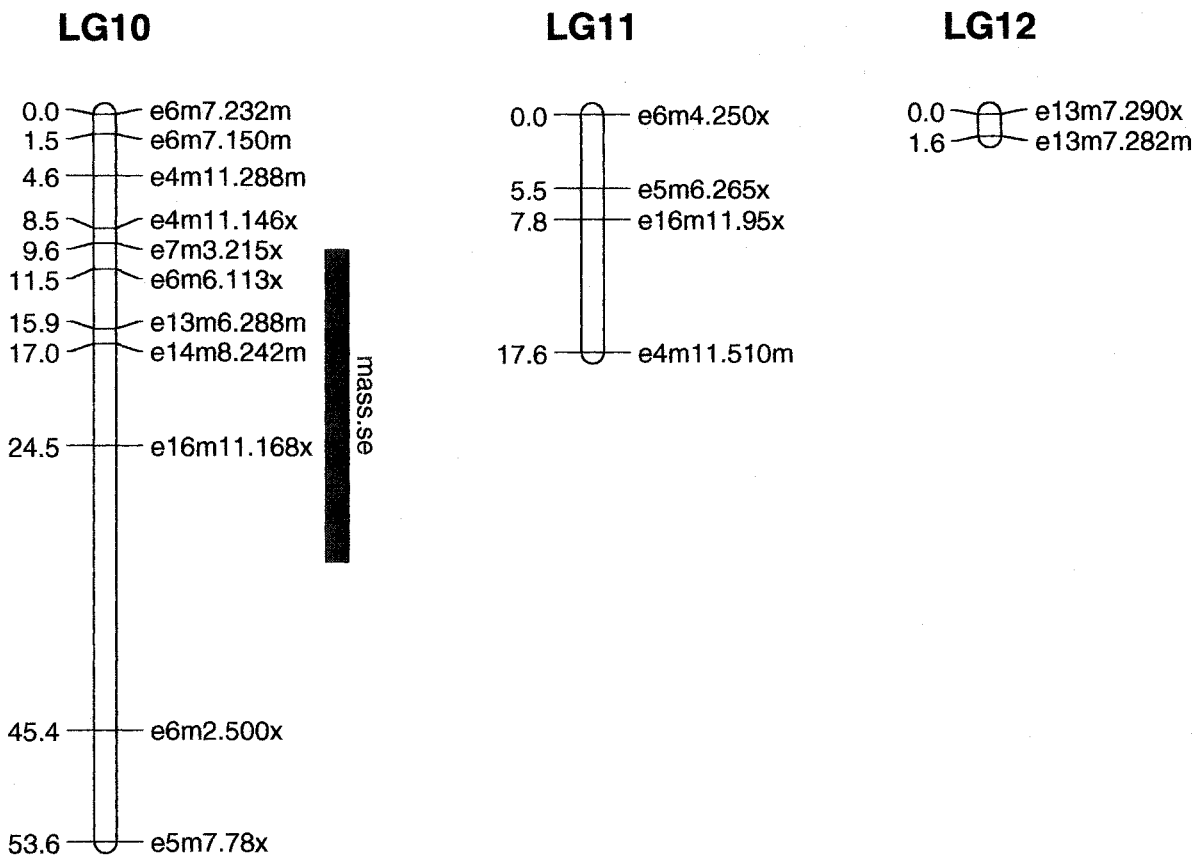


Figure 6.2 continued.

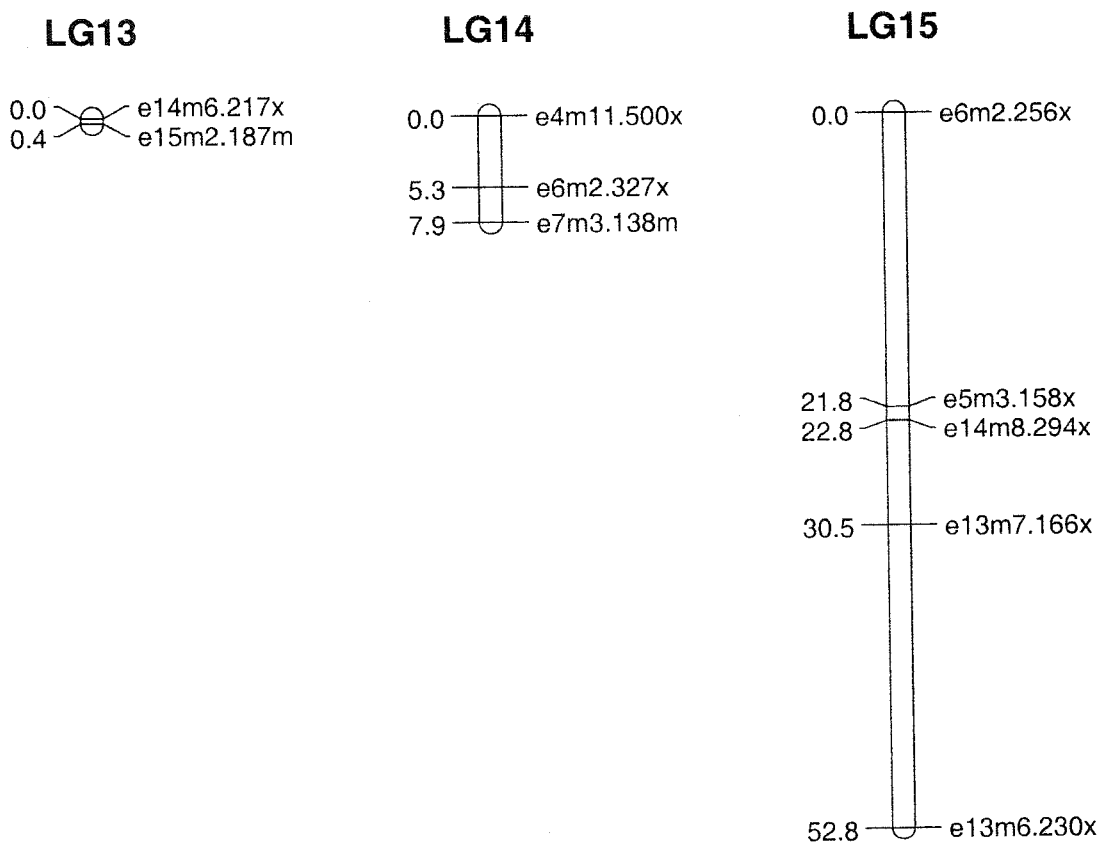


Figure 6.2 continued.

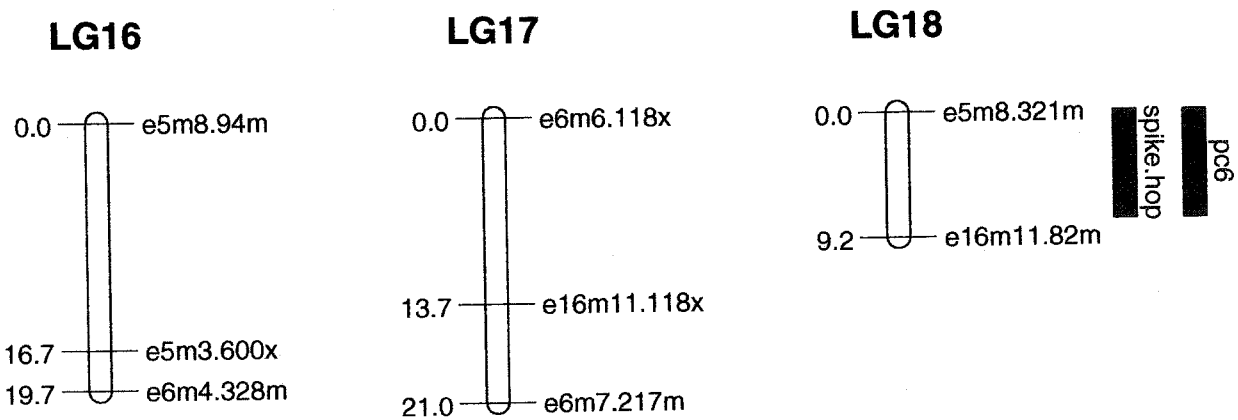


Figure 6.2 continued.

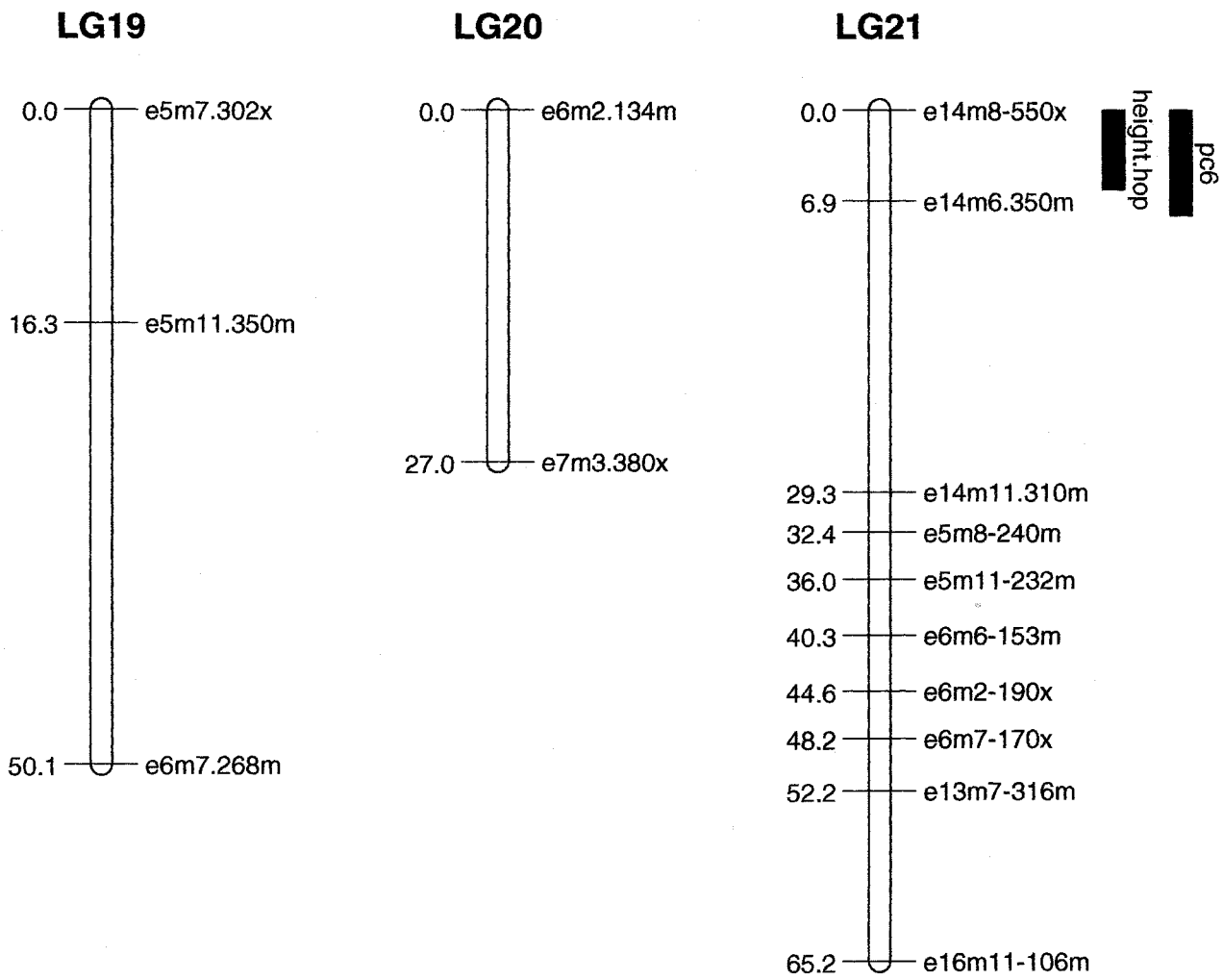


Figure 6.2 continued.

QTL mapping to the bottom of LG10. Again there was a mix of signs on the additive effects of the xeric allele, with the mass QTL on LG7 decreasing mass and the LG10 QTL increasing mass. Strangely, no QTL could be identified for height at Sierra even though there are significant single marker tests indicating the presence marker-QTL linkage on LG1 and LG7 (Appendix One).

Component scores from the factor analysis were used as trait values in order to map QTL underlying correlated complexes of traits. Two significant QTL were identified for PC5, the component known to be associated with growth and fitness at the Sierra site. The QTL on LG1 increased the trait value and mapped to a position that has been shown to have a QTL increasing spikelet number at Sierra. The second QTL co-localized to a portion of LG7 known to carry QTL for spiklet production and total mass a Sierra and mass at Hopland. Overall these two QTL could explain 17 % of the variation in the among line component scores.

Principal Component six, growth and fitness at Hopland:

At Hopland, a QTL was identified for spikelet production and was located on LG18 (Table 6.7; Figure 6.2). The additive effect of the xeric allele at this locus was to increase spikelet number, however the QTL was only found to explain 5 % of the variation in fitness. No other QTL were identified on Group 18 for any greenhouse or field trait. One suggestive QTL was also detected for height at Hopland and was located near the left hand end of LG21. The QTL was estimated to explain 6 % of the variation in height at Hopland, with the xeric additive effect reducing tiller height. This height QTL

was loosely linked to QTL identified for height at 20 days, flowing time, and stem pubescence in the greenhouse, however there was no overlap of confidence intervals on position with these "greenhouse QTL". In contrast, two QTL were identified for mass at Hopland and each was found to map to a position near a QTL of another trait. Specifically, one Hopland mass QTL was detected on LG1 near QTL identified for spikelet number in the greenhouse and spikelet number at Sierra. The additive effect of the xeric allele at all three was to increase the trait. The second Hopland mass QTL was located on LG7 within the position confidence intervals for mass at Sierra, spikelet number at Sierra, mass of the tillers in the greenhouse, maximum height in the greenhouse and germination time. The effect of the xeric allele was to decrease the trait values of the spikelet, mass and height traits, but interestingly increased germination time.

Two QTL were identified for PC6, and each co-localized to a position that had known effects on growth and reproduction at Hopland. One of the PC6 QTL located on LG18 had significant effects on fitness at Hopland with the xeric allele increasing the value of both of these traits. The remaining QTL mapped to LG 21 and decreased the component score by 0.34 while also decreasing height at Hopland. The QTL model containing these two QTL could only explain 8% of the variation in component scores among lines.

Epistatic loci for traits in the field:

A total of 317 significant two-way marker interactions were detected for the three traits measured at the two field sites (Table 6.8). In each case there were more significant interactions than the 13 expected by chance alone, indicating the presence of true epistatic interactions among loci. At Sierra, all three traits had at least 33 significant interactions, while at Hopland each trait had at least 22 significantly interacting marker pairs up to a maximum of 139 significant tests for mass at Hopland.

Only ten of the marker interactions significant at the p = 0.0015 level were still deemed significant at the Bonferroni adjusted 10 % significance level, and of these only seven were significant at the adjusted 5 % level (Table 6.8). All three significant interactions for spikelet number at Sierra describe the same interaction between LG1 and LG7, two groups that are known to have main effect QTL for spikelet production at Sierra. The epistatic effect (aa) of having the parental genotypes at the two loci (AABB or aabb) was to increase spikelet production by nearly three spikelets. Interestingly, the same marker loci, notably e7m3.204m and e16m11.218m, were found to have significant interaction effect on mass at Sierra (Table 6.9), with the epistatic effect increasing total mass by 0.70 grams. In both cases the epistatic effects between the loci on LG1 and LG7 on spikelet number and mass at Sierra were roughly equal to the magnitude of the additive QTL effects. The remaining interacting pair of loci had a significant effect on mass at Hopland. These loci were located on LG15 and LG18, the latter of which was known to have a QTL for spikelet production at Hopland. Surprisingly, the epistatic effect at these loci was to decrease mass at Hopland by 0.05 grams. The magnitude of the

Table 6.8. Number of two-way interactions among marker loci detected in the field environments.

Trait	Number of Sig. Interactions	Number expected by chance at p~0.0015	Sig. at p< 3.8*10-5 (Bonferroni adjusted alpha = 0.1)	Sig. at p< 1.9*10-5 (Bonferroni adjusted alpha = 0.05)
ht.se	32	13	0	0
spike.se	33	13	3	2
mass.se	77	13	6	4
ht.hop	22	13	0	0
spike.hop	139	13	0	0
mass.hop	64	13	1	1

Table 6.9. Epistatically interacting pairs of marker loci significant at the Bonferroni adjusted 10% significance level.

Trait	Locus A (group-marker)	Locus B	LR	p-value	aa
seed.se	e7m3.204x (1-9)	e16m11.218m (7-7)	12.46	0.000004	2.71
seed.se	e7m3.204x (1-9)	e14m6.298m (7-5)	11.93	0.000008	2.58
seed.se	e7m3.204x (1-9)	e14m6.177m (7-6)	11.26	0.000028	2.60
mass.se	e7m3.204x (1-9)	e14m6.298m (7-5)	15.04	0.000001	0.148
mass.se	e7m3.204x (1-9)	e16m11.218m (7-7)	16.42	0.000001	0.160
mass.se	e7m3.204x (1-9)	e14m6.177m (7-6)	14.07	0.000012	0.151
mass.se	e7m3.204x (1-9)	e5m2.105m (7-4)	13.50	0.000016	0.153
mass.se	e5m8.122x (1-7)	e16m11.218m (7-7)	13.29	0.000032	0.140
mass.se	e6m7.550x (1-6)	e16m11.218m (7-7)	12.07	0.000036	0.130
mass.hop	e5m3.158x (15-1)	e5m8.321m (18-1)	14.26	0.000016	-0.05

epistatic effect was larger than either additive effect at the main effect QTL identified for mass at Hopland.

Discussion:

Fitness correlates poorly across environments:

Clearly fitness in the parental environments must involve mechanisms that differ from those controlling fitness in a novel, stress free greenhouse environment. This is demonstrated both by the by the weak genetic correlations among traits across field and greenhouse environments and the tendency for fitness related traits to group together on environment specific axes of variation. If the same genes had been responsible for trait variation in multiple environments then the genetic correlation between the same trait measured in each environment should have been close to one (Falconer and MacKay, 1996). In the case of the traits measured at Hopland and Sierra, the overlap of variation (i.e. R^2) was in the 10-15% range (Table 6.5), implying that growth and fitness at each site were under predominantly independent genetic control. Similarly, no highly significant genetic correlations were detected among the field traits and the greenhouse traits, providing evidence that growth and fitness are controlled by separate loci in these different environments. However, within each field site the fitness traits were found to be highly genetically correlated, with larger and taller plants generally being more fit.

Thus the prediction was to uncover separate QTL underlying fitness in each environment. However, the prospect of finding QTL that could explain this pattern was significantly lower in the field, owing to significantly smaller heritiabilities in the F6 generation in relation to the greenhouse. This was largely due to relatively more environmental variation affecting the traits rather than a reduction in genetic variation (Table 6.1). For example, the coefficients of variation among F6 lines for height in the

greenhouse ($CV_{Among\ F6}$ = .111) and the field ($CV_{Among\ F6}$ = .130 at Sierra, $CV_{Among\ F6}$ = .116 at Hopland) were roughly the same, while the environmental coefficients of variation for height were much higher at the field sites ($CV_{Environment}$ = .213 at Sierra, $CV_{Environment}$ = .452 at Hopland) than the greenhouse ($CV_{Environment}$ = .095).

Variation within and among ecotypes in contrasting environments and coadapted gene complexes for growth and fitness:

Despite the increased influence of the environment, significant genetic variation was detected between the mesic and xeric ecotypes, at both field sites, for height and total mass but not for fitness (Table 6.2a). The same result was found using a reduced ANOVA model, suggesting the lack of a fitness difference among the ecotypes was not simply a function of overfitting the model with too many nonsignificant factors and interactions (Table 6.2b). This was surprising since it has been speculated that each ecotype is locally adapted and therefore should have out performed its rival genotype when grown in its own natural environment (Clegg and Allard, 1972). This also contrasted sharply with the greenhouse environment where the mesic parent outperformed the xeric with a near twofold difference in spikelet production (Chapter Three). It is possible the increased non-genetic variability of the field environments may mask some of the fitness difference among the ecotypes, however given that it is known the parental genotypes carry complementary alleles at several QTL (Chapter Five), it is likely the additive effects at the fitness QTL cancel out (see below), yielding a similar net phenotype among mesics and xerics. As well, no significant differences among families

within either parental ecotype were detected. This is further evidence that the parental ecotypes are each genetically uniform. In addition, it implies that the significant variation among parental families that was found in the greenhouse may be artifactual to that environment and may have no relevance in the natural environments. At present however this is only speculation, as traits such as flowering time (variable among xeric families in the greenhouse) have not yet been measured at the field sites.

In addition to the apparent lack of a difference in fitness between ecotypes, there were three other surprising results of the biometrical analysis. The first was that the xeric additive effect on height was negative in both parental environments, despite the fact it is known that taller (and heavier) plants are more fit at both field sites (Tables 6.4a and 6.4b). If the xeric ecotype really is locally adapted to the conditions observed at Sierra than it is reasonable to assume it should have been taller (and more fit) at that site.

The second result was the large epistatic effects that increased height and fitness in both parental environments, the magnitude of which overshadowed the magnitude of the additive effects on these traits. In fact, the epistatic effects on height at Sierra and Hopland essentially allowed both parents to exceed the mean performance of their F6 progeny. Thus nonadditive genetic effects appear to be a major contributor to growth and fitness in the field, which provides support for the presence of coadapted gene complexes in the parental ecotypes. However, it is unknown whether these epistatic effects are manifest from the suite of allozyme loci originally described by Allard *et al.* (1972) as being coadapted in this species. Future studies can address this question by screening the inbred line population with these allozyme loci. What can be concluded in this study is, given the presence of genic interactions, there must be multiple genes affecting height

and spikelet production in each natural environment. Furthermore, because there are epistatic effects for spikelet production in the absence of net additive effects it is likely that the alleles carried by each parent at some loci must have opposing effects that cancel out to a net additive effect close to zero (Table 6.3).

The third result was the pattern of transgressive segregation at both field sites. The range of component scores along the major axis of variation in the F6 generation was found to be more extreme than the mesic and xeric families, indicating the parental ecotypes harbour complementary alleles for growth and fitness in the field. While this is similar to what was observed in the greenhouse, the production of F6 lines that do more poorly than the least fit parent (i.e. maladaptive) is more pronounced in the field. As with the greenhouse study, most F6 lines did not score highly on more than one axis of variation. The one exception was F6-87 which scored highly on each environment specific component, PC5 and PC6. This suggests it may be possible to produce a recombinant genotype that can perform well in one environment without a loss in fitness in another anvironment.

Colocalizing QTL and transgressive segregation in the RIL population:

Several significant and suggestive QTL were detected for traits, including fitness, in each field environment with the exception of height measured at Sierra. Within the Sierra field environment, one QTL for spikelet production colocalized with a QTL for total mass consistent with the known positive genetic correlation of these traits (Figure 6.1). A similar pattern was not observed at Hopland, with the QTL for mass and fitness

mapping to separate genomic regions. Across field sites however, QTL for mass at Hopland mapped to the same locations as QTL for spikelet production at Sierra on LG1 and LG7, and mass at Sierra on LG7. This provides direct evidence that there are some similar genetic mechanisms controlling variation in the natural environments and provide an explanation of the weak genetic correlation of traits across the two field sites.

In addition, QTL with opposing signs on the additive effects were detected for spikelet number and mass at Sierra, and mass at Hopland thus providing evidence of a mechanism for transgressive segregation in the natural environment, very similar to what was observed in the greenhouse.

When comparing the field results to what was found in the greenhouse the prediction, that there are environment specific loci, seemed to hold in part as the QTL that were identified for spikelet production in the field environments mapped to positions (LG7 and LG18) that did not have QTL for fitness in the greenhouse (Figure 6.1). However, the suggestive QTL for fitness at Sierra on LG1 does slightly overlap the confidence limits on position for the suggestive QTL for spikelet production in the greenhouse. This perhaps provides the basis of the weak (nonsignificant) genetic correlation observed between these traits (Table 6.5).

A further surprising result of the QTL analysis was identifying loci associated with total mass in the field environments, on LG7 and LG10, that colocalized with QTL for tiller and leaf mass in the greenhouse (Figure 6.1). This suggests the presence of some loci controlling growth that contribute to trait variation in multiple environments. In this case mass at Sierra seems to have a positive relationship with tiller mass in the greenhouse and a negative relationship with leaf mass in the greenhouse. Similarly, mass

at Hopland is positively related to tiller mass in the greenhouse. However, given the poor correlation among mass in the greenhouse and mass in the field it is likely there are many more mass QTL that only have main effects in one environment thus reducing the influence of the trait covariance generated by these colocalizing QTL.

Interactions among marker loci and the magnitude of epistatic effects in the field environments:

Across all pairwise combinations of loci there was evidence of some true digenic epistasis, as each trait had more marker interactions than expected by chance (Table 6.8). This was not unexpected for traits such as fitness and height, both of which have significant net epistatic effects (estimated biometrically), and therefore provides evidence at the molecular level that some of the interacting loci reside within the mapped portion of the genome (Table 6.3). It was unexpected however for mass at Sierra and Hopland because there was no net epistatic effects detected in either of these cases. In order to reconcile the biometric and molecular data in these latter two instances it is necessary to look at the individual pairs of loci that are significant at the genome wide level (i.e. interactions unlikely to be false positives).

While only ten pairs of markers were significant at the 10 % genome wide error level, they do provide insight into the arrangement of epistatic effects across the genome. Two regions of the molecular map, namely LG1 and LG7, were found to have significant epistatic effects on fitness and mass at Sierra. In both cases the direction and magnitude

of the interaction effect (i.e. the effect of having the parental genotypes at the two marker loci), was to increase the trait by an amount comparable to the additive main effects of QTL identified for these traits (Table 6.9, Table 6.7). This is both the direction and magnitude of effects expected if the parental ecotypes really have coadapted gene complexes (Allard *et al.*, 1972) and these complexes are important contributors to fitness in the natural environment (Wright, 1931; Kim and Rieseberg, 2001). However, this still does not explain why there is no net epistasis for mass at Sierra (and Hopland).

A possible explanation comes from looking at the one interaction between LG15 and LG18 that has an effect on mass at Hopland. Interestingly, the interaction effect *decreases* mass at Hopland contrary to the expected performance of a coadapted genotype. While unexpected, this is good evidence that there are some negatively epistatic loci underlying trait variation in the field environments. Given this evidence, it is possible the lack of net epistasis for the mass traits is due to the canceling of increasing and decreasing interaction effects.

Chapter Seven:

Summary and Conclusions

In Chapter Two I conducted a review of the genetics literature in order to find studies that examined both genetic correlations among quantitative traits and their underlying QTL. I was interested in testing whether the genetic architecture underlying correlated versus uncorrelated traits was different with regard to both the numbers of pleiotropic (co-localizing) QTL and their relative additive effect sizes. I found that on average significantly genetically correlated traits shared more QTL (33%) than a pair of uncorrelated traits (11%). Surprisingly, the actual number of QTL shared between the average pair of correlated traits was only one more than the average uncorrelated trait pair. This suggests that a relatively small number of shared loci can account for a considerable amount of the covariation among particular pairs of traits.

I then defined a QTL based genetic correlation estimator called r_Q, incorporating additive QTL effects as measures of variation and covariation among pairs of traits. The QTL based estimator yielded somewhat accurate estimates of the biometric genetic correlation (reffered to as r_G), with variation in r_Q explaining about 50 % of the variation in r_G. Error in estimation was predominantly in magnitude of the correlation rather than correlation sign.

The third goal of the review was to characterize traits that had identifiable antagonistically pleiotropic QTL to determine whether or not they lead to negative genetic correlations, and potentially tradeoffs among traits. Approximately 25 % of trait pairs in the data set had at least one QTL with antagonistic effects on more than one trait. The majority of these trait pairs (177 out of 276) were found to be negatively genetically

correlated as estimated by standard biometric techniques. Nevertheless, 54 trait pairs were found to have net positive genetic correlations, providing direct evidence for Fry's (1993) conjecture that mechanisms for tradeoffs between traits can be "hidden" within positive genetic correlations. Surprisingly, these net positively correlated traits had the same average number of antagonistically pleiotropic QTL as the negatively correlated traits. However, the positively correlated traits did have more total pleiotropic QTL than negatively correlated traits and it was these additional loci that generated the positive covariance between traits that masked the effect of the antagonistic loci.

In Chapter Three I conducted a common garden experiment in the greenhouse, where the recombinant inbred lines were grown along with their parental ecotypes, in order to characterize the genetic differences that have led to the differentiation of the *A. barbata* ecotypes. The mesic and xeric ecotypes were found to differ for a suite of growth and fitness traits such as flowering time, allocation to tillers versus leaves and spikelet number, the latter of which was used as a fitness measure. Within ecotypes however there was no detectable genetic variation, suggesting the mesic and xeric ecotype each represent naturally occurring inbred lines. Overall, the mesic ecotype was the more fit genotype, flowering a month earlier than the average xeric, allocating more biomass to reproductive structures (tillers) and producing near double the number of spikelets than the xeric genotype. The xeric genotype in contrast was larger overall, predominantly through allocating more biomass to vegetative growth. I speculate these differences in life histories between ecotypes are a result of ancestral population differentiation to the two natural moisture regimes originally used to describe this study system (Allard *et al.* 1972; Clegg and Allard 1972; Hamrick and Allard 1972; Perez de la Vega *et al.* 1991).

The genetic differences among the parental ecotypes translated into a substantial release of genetic variation in their progeny, with most traits having broadsense heritabilities above 50%. This is good evidence for the possible role of rare outcrossing events providing the fuel for selection, thus initiating adaptive divergence in this otherwise selfing species. As well, significant transgressive segregation was observed for most traits, including fitness, indicating that within this pool of variation certain genotypes could out perform the parental ecotypes, or thrive in a novel habitat.

The relationships among quantitative traits revealed strong constraints to the types of possible genotypes in the progeny. Reproductive output in both tillers and spikelets was highly negatively genetically correlated to flowering time indicating that reproductive timing was the major determinant of fitness in the stress free greenhouse environment. In fact, these traits predominantly varied along one major axis of variation in the greenhouse suggesting they are functionally integrated. In contrast, the traits describing early growth were not highly correlated to fitness or the fitness axis of variation and were relatively unimportant to overall performance in the greenhouse, likely due to the absence of competition for seedling establishment. The only traits that exhibited a clear genetic tradeoff were allocation to tiller mass and allocation to leaf mass, with the negative correlation between them exceeding -0.70. Interestingly, each trait was positively correlated to total mass, indicating the presence of loci that are independent of the tradeoff.

Using Amplified Fragment Length Polymorphism (AFLP, Vos *et al.*, 1995) molecular markers I screened the parental ecotypes for genetic variation and genotyped the F6 generation in order to construct a genetic linkage map *A. barbata* genome

(Chapter Four). In total, I screened over 80 AFLP selective primer combinations in order to find markers that were polymorphic between a subset of the available mesic and xeric families. In only one instance (out of >200 loci) was there a locus that was polymorphic within one of the parental ecotypes. This is conclusive evidence that the parental ecotypes are genetically uniform within ecotypes, and supports the assertion that the ecotypes are naturally occurring inbred lines.

A total of 133 reliable AFLP loci were genotyped across 180 recombinant inbred lines in order to construct the genetic map (Figure 4.1). The markers grouped into 21 linkage groups that spanned approximately 640 cM, or 40 % of the estimated 1600 cM wild oat genome. The level of genome coverage was less than expected, and was most likely due to the standard AFLP protocol, using the methylation sensitive restriction endonuclease *Eco*R I, only targeting certain portions (i.e. unmethylated) of the wild oat genome. Future efforts to expand and refine the genetic map should ideally incorporate different types of molecular markers (e.g. microsatellites, RFLP, RAPD) to avoid problems of markers clustering in certain portions of the genome.

The overall level of segregation distortion (16.2%) among AFLP loci was comparable to hexaploid (9.8%) and diploid (16%) oat mapping experiments. One linkage group in particular contained six loci that were all distorted toward an overrepresentation of the xeric parent's alleles, regardless of which parent contributed the "band present " AFLP allele. This linkage group most likely contains a segregation distorter factor or pollen viability locus that was preferentially inherited early on in the inbred line propagation.

By using the combination of phenotypic data collected in Chapter Three and the molecular map data from Chapter Four I was able to identify some of the QTL underlying the divergence of the mesic and xeric ecotypes (Chapter Five). At least one suggestive (i.e. significant at chromosome wide threshold) or significant (i.e. significant at genome wide threshold) QTL was identified for all traits except dry mass at 60 days. In all instances there was at least one QTL with additive effects in the opposite direction than expected from the difference in parental means, thus providing an explanation for the transgressive segregation observed among the F6 lines. Clearly some lines received mostly beneficial, or mostly deleterious, alleles from each parent resulting in the lines exhibiting more extreme phenotypes than either parent.

Several clusters of QTL were identified that explain the correlated relationship among fitness traits and therefore describe constraints to the independent evolution of these traits. In particular, regions of LG10, LG15 and LG17 had QTL for flowering time that also had antagonistic effects on measures of reproductive output in spikelets and tillers (Figure 5.1). It is likely these QTL contain pleiotropic, or tightly linked, genes and therefore pose a formidable constraint between reproductive timing and overall fitness in this species. Moreover, while it is apparently possible to recombine the complementary QTL alleles carried by the parents at these loci, the negative relationship among flowering and fitness remains, thus significantly limiting the types of possible phenotypes in the progeny.

Likewise, the same regions of LG10 and LG17 were also identified to have antagonistic effects on tiller mass and leaf mass, therefore accounting for a portion of the tradeoff between these traits. Interestingly, there were QTL that were not involved in the

tradeoff between traits. Rather these other QTL reveal the predominantly positive relationship between total mass and its constituent traits separate from the tradeoff.

A number of epistatically interacting pairs of markers were detected for traits in the greenhouse, over and above the number expected by chance (Table 5.2). This indicated that at least some of the significant pairs of loci exhibited true epistasis for traits such as total mass and tiller number. Other traits, such as flowering time and germination time had as many interactions expected by chance alone. Identifying the individual pairs of epistatic loci that were truly interacting was difficult, due to the extreme significance threshold (e.g. $p<1 \times 10^{-5}$) needed to ensure a reasonable false positive rate across the entire molecular map. Nevertheless, three interacting pairs of markers were significant at the genome wide level and all had effects on total mass (directly or indirectly through PC2; Table 5.3). All three pairs indicate the same regions of the map, namely LG10 and LG14, are interacting to increase total mass in the parental genotypes (i.e. AABB and aabb). Interestingly, the magnitude of the epistatic effect was as large as the additive effects of the main effect QTL for total mass.

A field reciprocal transplant experiment was carried out where the F6 inbred line population was grown in each parental environment in California, alongside the mesic and xeric ecotypes, in order determine how fitness correlate across environments (Chapter Six). Both the biometrical and QTL analysis indicate there are different mechanisms underlying variation in fitness and growth traits in the field in comparison to the greenhouse. First, while the traits measured in the greenhouse could be largely explained by additivity, there were large epistatic effects on height and spikelet production in the field, allowing both mesic and xerics to exceed the mean performance

of their F6 progeny with respect to fitness and height in both parental environments. Obviously, the interactions among genes within xeric and mesic genotypes (coadapted gene complexes) contribute to growth and fitness to a greater degree in the more variable field environments. Second, growth and fitness are highly genetically correlated within both parental environments, but do not correlate strongly across field and greenhouse environments. Moreover, each suite of "field traits" were found to group together on environment specific axes of variation, with little or no relation to any of the greenhouse traits. This is strong evidence that similar sets of loci control variation among traits within environments, but that separate groups of loci control trait variation among environments. This result is significant because it implies that fitness in one environment does not necessarily come at a cost of lower fitness in other environments. This contrasts the study of Via and Hawthorne (2002) who showed that there is a negative correlation between aphid fecundity on two different host plants and thus a tradeoff in fitness among environemts.

The results of the QTL analysis largely support the results of the genetic correlation and principal component analyses; QTL identified for fitness in the parental environments did not map to the positions of the loci for fitness in the greenhouse. However, there was some overlap of QTL positions on LG7 and LG10 for mass and spikelet production at Sierra with mass and height in the greenhouse. This indicates there may be some loci contributing to trait variation in multiple environments but, given the small genetic correlation of growth in the field with growth in the greenhouse, there are most likely other loci that are specific to one environment. If this is the case then

different portions of the genome could become targets for selection in different environments.

There was evidence for epistasis among marker loci for all traits measured in the field environments. This was unexpected for mass at Sierra and Hopland because neither had any net epistatic effects in the biometrical analyses. Narrowing the total number interactions down to individual pairs of interacting marker loci was difficult, and as such only 10 interactions (out of 367) were deemed significant at the genome wide error threshold. Nine of these 10 pairs of loci indicate the same regions of LG1 and LG7 (already known to carry QTL) have epistatic effects on spikelet production and mass at sierra, with the parental genotypes having 2.5 more spikelets and 0.16 grams more mass. These epistatic effects were as large as the additive effects of the main effect QTL, and thus indicate importance of coadapted gene complexes in contributing to fitness in the natural environment. The remaining interaction suggested LG15 and LG18 have epistatic effects on mass at Hopland. Surprisingly, the effect of the interaction was to reduce mass in the parental genotypes, which is contrary to the expected effect of a coadapted pair of loci. The presence of such negative epistasis may cause the canceling of other positive intereaction effects thus explaining explain why there is no net epistasis for mass in the field environments.

Overall, the experiments outlined in this thesis yield insight into the genetics of population divergence with respect to the release of genetic variation and constraints to this release both in a novel environment and under natural conditions. In summarizing my experiments a logical question to ask is "How well does the QTL analyses explain the variation among mesic and xeric ecotypes observed at the phenotypic level?". To answer

this question I compare the genetic parameters estimated biometrically among generation means, expected to represent the average of effects across all loci, to the sum of the individual effects of the identified QTL (Table 7.1). I also compare the amount of genetic covariance (rQ) that can be explained by the QTL in order to see how well it fits with the biometric data.

In the greenhouse and field analysis, the sum of additive effects of the QTL ($\sum a_{qtl}$) were generally the same sign as the net additive effect [a] for most traits, however they could not account for all of the effect among generation means. In contrast, the sum of the QTL additive effects for maximum height, spikelet number and total mass were of opposite sign than expected. In each of these cases it is the undetected loci that must account for the difference, thus providing a means to infer the net effects of the undetected (or undetectable) QTL.

A major concern of the mapping analysis was the amount of genome coverage afforded by the current AFLP map. I have speculated that if the entire genome differed between xeric and mesic genotypes then the 644 cM coverage, presented in Chapter Four, represents approximately 40% of the wild oat genome. It is possible however, only limited segments of the *A. barbata* genome actually differ among ecotypes, due to recent shared ancestry. If this is the case then the mapped portion of the genome might represent the only polymorphic pieces of the oat genome. While there is no direct test of this supposition, the amount of genetic variation explained by the identified QTL could be an indicator of map coverage. If the map is nearly complete one would likely find all the

Table 7.1. Comparison of genetic parameters measured at the phenotypic and genetic level

	WLS estimates				QTL estimates			
Trait	m	[a]	[aa]	H^2	$\sum a_{qtl}$	$\sum aa_{qtl}$	%Vp/100	%Vg/100
GM	5.27	0.03	-0.11	0.16	0.06		0.13	0.35
HT20	12.22	-1.44	0.00	0.39	-0.51		0.24	0.37
DM60	0.39	0.01	0.12	0.09				
FL	135.81	17.79	2.08	0.59	8.91		0.42	0.47
MH	137.60	7.07	4.79	0.58	-4.99		0.21	0.25
TN	11.83	-4.97	-1.53	0.55	-2.56		0.43	0.55
SN	424.70	-	-	0.59	25.68		0.48	0.58
		113.50	38.20					
TM	7.79	-1.49	-0.13	0.57	-1.76		0.42	0.52
LM	7.37	2.69	0.34	0.58	0.67		0.40	0.47
TOTM	15.16	1.08	0.11	0.33	-0.12	0.70	0.23	0.39
height.se	86.49	-4.97	10.75	0.28				
spike.se	17.46	-1.10	7.47	0.16	-0.57	2.71	0.18	0.49
mass.se	0.89	-0.30	0.50	0.14	-0.02	0.16	0.13	0.44
height.hop	69.47	-10.55	22.17	0.09	-6.51		0.04	0.62
spike.hop	7.48	0.21	3.74	0.10	0.92		0.04	0.28
mass.hop	0.44	-0.19	0.15	0.09	0.01	-0.05	0.14	0.95

QTL underlying the difference among parental genotypes (and explain a large proportion of the genetic variation in the process).

On the surface the amount of phenotypic variance explained by the QTL (Vp, Table 7.1) for most traits is approximately what would be expected (~40%) if the map covers less than half of the genome. However, this is the variance among line means which is not all genetic (i.e. contains a portion of the within line variance) and is perhaps misleading. When corrected for within line variation, the amount of genetic variation (V_G) explained by the identified QTL was still only approximately 50% for most traits. It is likely these amounts are biased upward as in most QTL experiments (Beavis, 1998), however the bias would be more pronounced for the traits that have heritabilities less than 0.20 (Beavis, 1998). Therefore without any other information, it would appear there are still portions of the genome to be mapped and presumably contain some additional loci responsible for the unexplained genetic variation among traits.

While most QTL studies do report how much variation is accounted for by identified loci (see Orr, 2001), very few studies report how much covariation among traits is explained by these same loci. Therefore I conclude by applying the QTL based correlation estimator defined in Chapter Two, to the data collected in this thesis. The correlation of the r_G and r_Q was calculated to be 0.679, thus approximately 46 % of the genetic correlation among traits could be explained by the effects of QTL identified for traits in the greenhouse and in the field ($R^2 = 0.461$, Figure 7.1). This was similar to the correlation observed across all studies in Chapter Two (i.e. 0.711). This indicates that a relatively small number of loci can make a substantial contribution to variation and covariation among traits in this species.

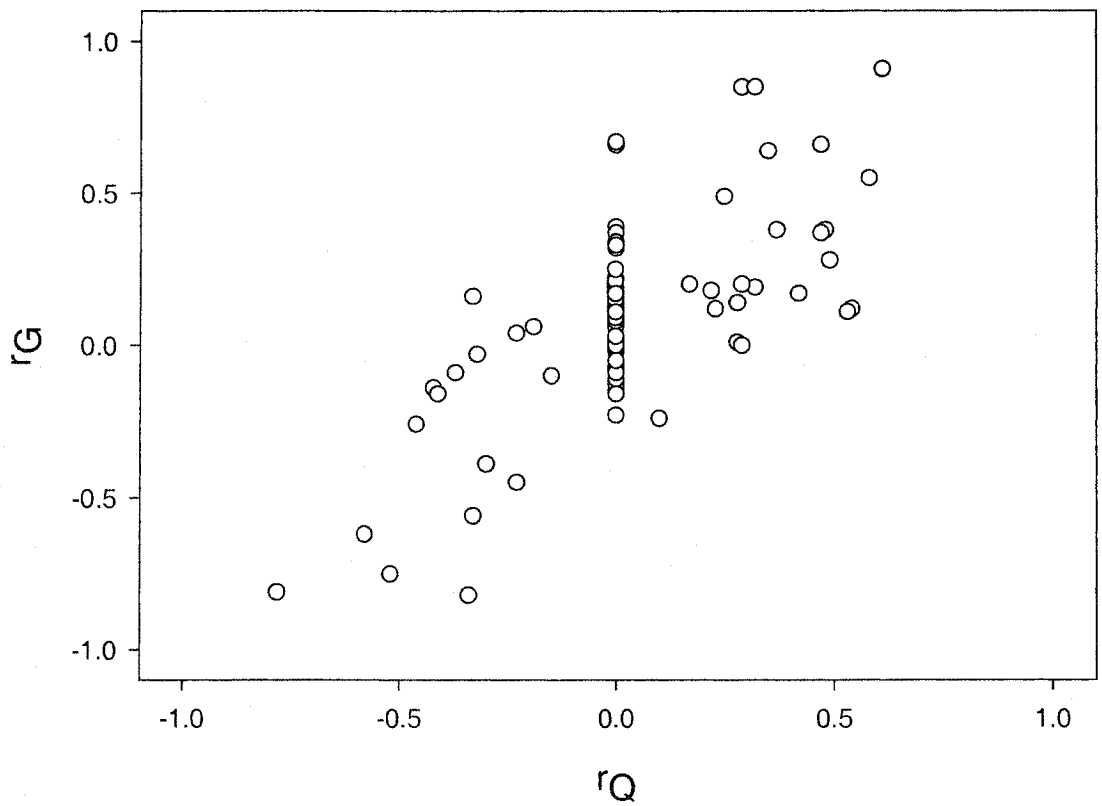


Figure 7.1. Scatter plot of r_G versus r_Q for pairs of traits in *A. barbata*. Each point represents one trait pair.

References

Allard, R.W., Babbel, G.R., Clegg, M.T., Kahler, A.L. 1972. Evidence for Coadaptation in *Avena barbata. Proc. Nat. Acad. Sci.* 69: 3043-3048.

Allard, R.W., Garcia, P., Saenz-de-Miera, L.E., Perez de la Vega, M. 1993. Evolution of Multilocus Genetic Structure in *Avena hirtula* and *Avena barbata. Genetics* 135: 1125-1139.

Arnold, S.J. 1992. Constraints on phenotypic evolution. *Am. Nat.* 140: S85-S107.

Austin, D.F., Lee, M. 1996. Genetic resolution and verification of quantitative trait loci for flowering and plant height with recombinant inbred lines of maize. *Genome* 39:957-968.

Austin, D.F., Lee, M. 1998. Detection of Quantitative trait loci for grain yield and yield components in Maize across generations in stress and non stress environments. *Crop Sci.* 38:1296-1308.

Bai, G.H., Shaner, G., Ohm, H. 1999. Inheritance of resistance to Fusarium graminearum in wheat. *Theor. App. Genet.* 100: 1-8.

Barton, N., Keightley, P.D. 2002. Understanding quantitative genetic variation. *Nat. Rev. Gen.* 3: 11-21.

Beavis, W.D. 1998. QTL analyses: power, precision, and accuracy, IN Molecular Dissection of Complex Traits, Ed. A.H. Patterson. CRC Press, London.

Becker, J., Vos, P., Kuiper, M., Salamini,F., Heun, M. 1995. Combined mapping of AFLP and RFLP markers in barley. *Mol. Gen. Genet.* 249: 249-265.

Ben Chaim, A., Paran, I., Grube, R.C., Jahn, M. van Wijk, R., Peleman, J. 2001. QTL mapping of fruit related traits in pepper (Capsicum annuum). *Theor. Appl. Genet.* 102:1016-1028.

Bohn, M., Khairallah, M.M., Gonzalez-de-Leon, D., Hoisington, D.A., Utz, H.F., Deutsch, J.A., Jewell, D.C., Mihm, J.A., Melchinger, A.E. 1996. Mapping QTL in tropical Maize: Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits. *Crop Sci.* 36:1352-1361.

Botstein, D., White, R.L., Skolnick, R., Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Gen.* 32: 314-331.

Broman, K.W., Speed, T.R. 2002. A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of The Royal Statistical Society Series B-Statistical Methodology 64: 641-656.

Castiglioni, P., Ajmone-Marsan, P., van Wijk, R., Motto, M. 1999. AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor. Appl. Genet. 99: 425-431.

Chase, K., Adler, F.R., Lark, K.G. 1997. Epistat: A computer program for identifying and testing interactions between pairs of quantitative trait loci. *Theor. Appl. Genet.* 94: 724-730.

Cheverud, J.M. 2000. Detecting epistasis among quantitative trait loci. IN Epistasis and the Evolutionary Process. J.B. Wolf, E.D. Brodie III, M.J. Wade Eds.. Oxford University Press, London.

Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.

Clegg, M.T., Allard, R.W. 1972. Patterns of Genetic Differentiation in the Slender Wild Oat Species *Avena barbata*. *Proc. Nat. Acad. Sci.* 69: 1820-1824.

Cluster, P.D., Allard, R.W. 1995. Evolution of Ribosomal DNA (rDNA) Genetic Structure in Colonial Californian Populations of *Avena barbata*. Genetics 139: 941-954.

Conner, J.K. 2002. Genetic mechanisms of floral trait correlations in a natural population. *Nature*. 420:407-410.

Cui, K.H., Peng, S.B., Xing, Y.Z., Xu, C.G., Yu, S.B., Zhang, Q. 2002. Molecular dissection of seedling-vigor and associated physiological traits in rice. *Theor. Appl. Genet.* 105:745-753.

de Jong, G., van Noordwijk, A.J. 1992. Acquisition and allocation of resources: genetic (co)variances, selection, and life histories. *Am. Nat.* 139:749-770.

De Koeyer, D.L., Tinker, N.A., Wight, C.P., Deyl, J., Burrows, V.D., O'Donoughue, L.S., Lybaert, A., Molnar, S.J., Armstrong, K.C., Fedak, G., Wesenberg, D.M., Rossnagel, B.G., McElroy, A.R. 2004. A molecular linkage map with associated QTLs from a hulless x covered spring oat population. *Theor. Appl. Genet.* 108: 1285-1298.

deVicente M.C., Tanksley, S.D. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134: 585-596.

Doerge, R.W., Zeng, Z-B., Weir, B.S. 1997. Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Stat. Sci.* 12:195-219.

Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nature Rev. Genet.* 3:43-51.

Falconer, D.S. 1952. The problem of environment and selection. *Am. Nat.* 86: 293-298.

Falconer, D.S., Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edn. Harlow-Longman, London.

Fishman, L., Kelly, A.J., Morgan, E., Willis, J.H. 2001. A Genetic Map in the Mimulus guttatus Species Complex Reveals Transmission Ratio Distortion due to Heterospecific Interactions. *Genetics* 159: 1701-1716.

Foolad, M.R., Chen, F.Q., Lin, G.Y. 1997. RFLP mapping of QTLs conferring salt tolerance during germination in an interspecific cross of tomato. *Theor. Appl. Genet.* 97: 1133-1144.

Fry, J.D. 1993. The "General Vigor" problem: Can antagonistic pleiotropy be detected when genetic covariances are positive? Evolution 47: 327-333.

Fry, J.D. 1996. The evolution of host specialization: Are tradeoffs overrated? *Am. Nat.* 148: S84-S107.

Garcia, P., Vences, F.J., Perez d la Vega, M., Allard, R.W. 1989. Allelic and Genotypic Composition of Ancestral Spanish and Colonial Californian Gene Pools of *Avena barbata*: Evolutionary Implications. *Genetics* 122: 687-694.

Georgiady, M.S., Whitkus, R.W., Lord, E.M. 2002. Genetic Analysis of Traits Distinguishing Outcrossing and Self-Pollinating Forms of Currant Tomato, Lycopersicon pimpinellifolium (Jusl.) Mill. *Genetics* 161: 333-344.

Groh, S., Kianian, S.F., Phillips, R.L., Rines, H. L., Stuthman, D. D., Wesenberg, D. M., Fulcher, R.G.. 2001. Analysis of factors influencing milling yield and their association to other traits by QTL analysis in two hexaploid oat populations. *Theor. Appl. Genet.* 103:9-18.

Haldane, J.B.S., Waddington, C.H. 1930. Inbreeding and linkage. *Genetics* 16: 357-374.

Hakim-Elahi, A., Allard, R.W. 1983. Distribution of homeoalleles at two loci in a diplodized tetraploid: leucine aminopepitase loci in *Avena barbata*. J. Heredity 74: 379-380.

Hamrick, J.L., Allard, R.W. 1972. Microgeographical Variation in Allozyme Frequencies in *Avena barbata*. *Proc. Nat. Acad. Sci.* 69: 2100-2104.

Hamrick, J.L., Allard, R.W. 1974. Correlations Between Quantitative Characters and Enzyme Genotypes in *Avena barbata. Evolution* 29: 438-442.

Hawthorn, D.J., Via, S. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*. 412:904-907.

Herve,D., Fabre, F., Berrios, E., Leroux, N., Al Chaarani, G., Planchon, C., Sarrafi, A., Gentzbittel, L. 2001. QTL analysis of photosynthesis and water statue traits in sunflower (Helianthus annuus) under greenhouse conditions. *J. Exp. Botany* 52:1857-1864.

Hoffmann, A.A., Sgro, C.M., Lawler, S.H. 1995. Ecological Population Genetics: The interface between genes and the environment. *Annu. Rev. Genet.* 25: 371-399.

Houle, D. 1991. Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. *Evolution* 45:630-648.

Howell, D.C. 2000. RESAMPLING version 1.3. University of Vermont.

Hutchinson, E. S. 1982. *Genetic markers and Ecotypic differentiation of Avena barbata Pott. ex Link*. PhD dissertation, University of California Davis.

Hutchinson, E.S., Hakim-Elahi, A., Miller, R.D., Allard, R.W. 1983. The genetics of diploidized tetraploid *Avena barbata. The Journal of Heredity* 74: 325-330.

Igrejas, G.,Leroy, P., Charmet, G.,Gaborit, T.,Marion, D., Branlard, G. 2002. Mapping QTLs for grain hardness and puroindoline content in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 106:19-27.

Jain, S.K., Rai, K.N. 1980. Population biology of *Avena*. VIII. Colonization experiment as a test of the role of natural selection in populations divergence. *Am. J. Bot.* 67: 1342-1346.

Jenczewski, E., Gherardi, M., Bonnin, I., Prosperi, J.M., Olivieri, I., Huguet, T. 1997. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). *Theor. Appl. Genet.* 94: 682-691.

Jiang, C., Edmeades, G.O., Armstead, I., Lafitte, H.R., Hayward, M.D., Hoisington, D. 1999. Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers. *Theor. Appl. Genet.* 99: 1106-1119.

Jin, H., Domier, L.L., Shen, X.J., Kolb, F.L. 2000. Combined AFLP and RFLP mapping in two hexaploid oat recombinant inbred populations. *Genome* 43: 94-101.

Johansen, A.D. 2004. Fitness consequences of hybridization between ecotypes of *Avena barbata*. PhD. Dissertation. Dalhousie University.

Juenger, T, Purugganan, M., MacKay, T.F.C. 2000.Quantitative Trait Loci for Floral Morphology in Arabidopsis thaliana. *Genetics* 156: 1379-1392.

Kamoshita, A., Wade, L.J.,Ali, M.L., Pathan, M.S., Zhang, J., Sarkarung, S., Nguyen, H.T. 2002. Mapping QTLs for root morphology of a rice population adapted to rainfed lowland conditions. *Theor Appl Genet.* 104:880-893.

Kato, K., Miura, H., Sawada, S. 2000. Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. Theor Appl Genet. 101:1114-1121.

Katsiotis, A., Hagidimitriou, M., Heslop-Harrison, J.S. 1997. The close relationship between the A and B genomes in Avena L (Poaceae) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. *Annals of Botany* 79: 103-109.

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

Kearsey, M.J., Pooni, H.S. 1996. The genetical analysis of quantitative traits. Chapman and Hall, London.

Kearsey, M.J., Pooni, H.S., Syed, N.H. 2003. Genetics of quantitative traits in *Arabidopsis thaliana*. Heredity 91: 456-464.

Kim, S.C., Rieseberg, L.H. 2001. The contribution of epistasis to species differences in annual sunflowers. *Mol. Ecol.* 10: 683-690.

Kremer, C.A., Lee, M., Holland, .JB. 2001. A restriction fragment length polymorphism based linkage map of a diploid *Avena* recombinant inbred line population. *Genome* 44: 192-204.

Lande, R., Arnold, S.J. 1983. The measurement of selection on correlated charaters. Evolution 37: 1210-1226.

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

Lander, E.,Green, P.,Abrahamson, J.,Barlow, A., Daley, M., Lincoln, S., Newburg, L. 1987. MAPMAKER: An Interactive Computer Package for Constructing Primary Genetic Linkage Maps of Experimental and Natural Populations. *Genomics* 1: 174-181.

Latta, R.G., MacKenzie, J.L., Vats, A., Schoen, D.J. 2004. Divergence and variation of quantitative traits between allozyme genotypes of Avena barbata from contrasting habitats. *Journal of Ecology* 92: 57-71.

Lexer, C., Randell, R.A., Rieseberg, L.H. 2003. Experimental hybridization as a tool for studying selection in the wild. *Ecology* 84: 1688-1699.

Liedloff, A. 1999. MANTEL version 2.0: a nonparametric test calculator.

Lincoln, S.E., Lander, E.S. 1992. Systematic Detection of Errors in Genetic Linkage Data. *Genomics* 14: 604-610.

Lubberstedt, T., Melchinger, A.E., Scon, C.C., Utz, H.F., Klein, D. 1997. QTL mapping in testcrosses of European Flint line Maize: comparison of different testers for forage yield traits. *Crop Sci.* 37:921-931.

Lynch, M., Walsh, B. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland.

MacKenzie, A. 1996. A Trade-Off for Host Plant Utilization in the Black Bean Aphid, *Aphis fabae*. *Evolution* 50: 155-162.

Manly, KF, Cudmore, Jr, RH, Meer, JM. 2001.Map Manager QTX, cross-platform software for genetic mapping. *Mammalian Genome* 12: 930-932.

Mather K., Jinks, J.L. 1982. Biometrical Genetics The study of continuous variation. Chapman and Hall, London.

McKay, J.K., Richards, J.H., Mitchell-Olds, T. 2003. Genetics of drought adaptation in Arabidopsis thaliana: I. Pleiotropy contributes to genetic correlations among ecological traits. *Mol. Ecol.* 12: 1137-1151.

Merila, J., Sheldon, B.C. 1999. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity* 83: 103-109.

Mesfin, A., Smith, K.P., MacKay, R., Evans, C.K., Waugh, R., Gustus, C.D., Muehlbauer, G.J. 2003. Quantiative trait loci for Fusarium head blight resistance in barley detected in a two rowed by six rowed population. *Crop Sci.* 43:307-318.

Mitchell-Olds, T. 1995. The molecular basis of quantitative genetic variation in natural populations. *Trend. Ecol. Evol.* 10: 324-328.

Mitchell-Olds, T. 1996. Genetic constraints on life history evolution: quantitative trait loci influencing growth and flowering in *Arabidopsis thaliana*. *Evolution* 50:140-145.

Mitchell-Olds, T., Pedersen, D. 1998. The molecular basis of quantitative genetic variation in central and secondary metabolism in Arabidopsis. Genetics 149:739-747.

Nagy, E.S. 1997. Selection for native characters in hybrids between two locally adapted plant subspecies. *Evolution* 51: 1469-1480.

Nei, M., Li, W-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Nat. Acad. Sci.* 76:5269-5273.

O'Donoughue, L.S., Wang, Z., Roder, M., Kneen, B., Leggett, M., Sorrells, M.E., Tanksley, S.D. 1992. An RFLP-based linkage map of oats based on a cross between 2 diploid taxa (*Avena-atlantica x A-hirtula*). *Genome* 35: 765-771.

O'Donoughue, L.S., Kianian, S.F., Rayapati, P.J., Penner, G.A., Sorrells, M.E., Tanksley, S.D., Phillips, R.L., Rines, H.W., Lee, M., Fedak, G., Molnar, S.J., Hoffman, D., Salas, C.A., Wu, B., Autrique, E.,Van Deynze, A. (1995) A molecular linkage map of cultivated oat. *Genome* 38:368–380

Orr, H.A. 1999. The evolutionary genetics of adaptation: a simulation study. *Genet. Res.* 74: 207-214.

Orr, H.A. 2001. The genetics of species differences. *Trend. Ecol. Evol.* 16: 343-350.

Otto, S.P., Jones, C.D. 2000. Detecting the Undetected: Estimating the Total Number of Loci Underlying a Quantitative Trait. *Genetics* 156: 2093-2107.

Perez de la Vega, M., Garcia, P., Allard, R.W. 1991. Multilocus genetic structure of ancestral Spanish and colonial Californian populations of *Avena barbata*. *Proc. Natl. Acad. Sci.* 88: 1202-1206.

Phillips, P.C., Otto, S.P., Whitlock, M.C. 2000. Beyond the Average: The evolutionary importance of gene interactions and variability of epistatic effects. IN Epistasis and the evolutionary process. J.B. Wolf, Brodie III, E.D., Wade, M.J. Eds. Oxford University Press, London.

Pradhan, A.K., Gupta,V., Mukhopadhyay, A., Arumugam, N., Sodhi, Y.S., Pental, D. 2003. A high-density linkage map in *Brassica juncea* (Indian mustard) using AFLP and RFLP markers. *Theor. Appl. Genet.* 106: 607-614

Rajhathy, T., Thomas,H. 1974. Cytogenetics of oats (*Avena*) L. *Misc. Pub. Genet. Soc. Canada* No. 2.

Rayapati, P.J., Gregory, J.W., Lee, M., Wise, R.P. 1994. A linkage map of diploid *Avena* based on RFLP loci and a locus conferring resistance to 9 isolates of *Puccinia-coronata* var *avenae*. *Theor. Appl. Genet.* 89: 831-837.

Remington, DL., O'Malley, D.M. 2000. Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. Genetics 155: 337-348.

Reznick, D. 1985. Costs of reproduction: an evaluation of the empirical evidence. *Oikos* 44: 257-267.

Rieseberg, L.H., Archer, M.A., Wayne, R.K. 1999. Transgressive segregation, adaptation and speciation. *Heredity* 83: 363-372.

Rieseberg, L.H., Widmer, A., Arntz, A.M., Burke, J.M. 2002. Directional selection is the primary cause of phenotypic diversification. *Proc. Nat. Acad. Sci.* 99:12242-12245.

Rieseberg, L.H., Widmer, A., Arntz, A.M., Burke, J.M. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. *Phil. Trans. R. Soc. Lond. B.* 358: 1141-1147.

Roff, D.A. 1996. The evolution of genetic correlations: an analysis of patterns. *Evolution.* 50:1392-1403.

Roff, D. 2002. Life history evolution. Sinauer, Sunderland.

Saliba-Columbani, V., Causse, M., Langlois, D., Philouze, J., Buret, M. 2001. Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. *Theor Appl Genet* 102:259-272.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: A laboratory manual, Second Edition. Cold Spring Harbour Laboratory, NY.

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

Schon, C.C., Utz, H.F., Groh, S., Truberg, B., Openshaw, S., Melchinger, A.E. 2004. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167: 485-498.

Schwarzbach, A.E., Donovan, L.A., Rieseberg, L.H. (2001) Transgressive character expression in a hybrid sunflower species. *Am. J. Bot.* 88: 270-277.

Sokal, R.R., Rohlf, F.J. 1995. Biometry 3rd edition. W.H. Freeman and Company, New York.

Stirling, G., Roff, D., Fairbairn, D. 1999. Four characters in a trade-off: dissecting their phenotypic and genetic relations. Oecologia 120: 492-498.

Tan, Y.F., Sun, M., Xing, Y.Z., Hua, J.P., Sun, X.L., Zhang, Q.F., Corke, H. 2001. Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. *Theor. Appl. Genet.* 103:1037-1045.

Tani, N., Takahashi, T., Iwata, H., Mukai, Y., Ujino-Ihara, T., Matsumoto, A., Yoshimura, K., Yoshimaru, H., Murai, M., Nagasaka, K., Tsumura, Y. 2000. A consensus linkage map for sugi (Cryptomeria japonica) from two pedigrees, based on microsatellites and expressed sequence tags. Genetics 165: 1551-1568.

Tanksley, S.D. 1993. Mapping Polygenes. *Annu. Rev. Genet.* 27: 205-233.

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

Tinker, N.A., Mather, D.E., Rossnagel, B.G., Kasha, K.J., Kleinhof, A., Hayes, P.M., Falk, D.E., Ferguson, T., Shugar, L.P., Legge, W.G., Irvine, R.B., Choo, T.M., Briggs, K.G., Ullrich, S.E., Franckowiak, J.D., Blake, T.K., Graf, R.J.,, Dofing, S.M., Saghai Maroof, M.A., Scoles, G.J., Hoffman, D., Dahleen, L.S., Kilian, A., Chen, F., Biyashev, Kudrna, R.M., Steffenson, B.J. 1996. Regions of the genome that affect agronomic performance in two rowed barley. *Crop Sci.* 36:1053-1062.

Ungerer,M.C., Halldorsdottir, S.S., Modliszewski, J.L., MacKay, T.F.C., Purugganan, M.D. 2002. Quantitative Trait Loci for Inflorescence Development in Arabidopsis thaliana. *Genetics* 160: 1133-1151.

Veldbloom, L., Lee, M. 1996. Genetic mapping of quantitative trait loci in Maize in stress and nonstress environments I: grain yield and yield components. *Crop Sci.* 36:1310-1319.

Verhoeven, K.J.F., Vanhala, T.K., Biere, A., Nevo, E., Van Damme, J.M.M. 2004. The genetic basis of adaptive population differentiation: A quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. Evolution 58: 270-283.

Veldbloom, L., Lee, M. 1996. Genetic mapping of quantitative trait loci in Maize in stress and nonstress environments II: plant height and flowering. *Crop Sci.* 36:1320-1327.

Via, S., Hawthorne, D.J. 2002. The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *Am. Nat.* 159:S76-S88.

Vieira, C., Pasyukova, E.G., Zeng, Z-B., Hackett, J.B., Lyman, R.F., MacKay, T.F.C. 2000. Genotype-Environment Interaction for Quantitative Trait Loci Affecting Life Span in *Drosophila melanogaster. Genetics.* 154: 213-227.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. 1995. AFLP - a new technique for DNA fingerprinting. *Nuc. Acid. Res.* 23: 4407-4414.

Wang S., C. J. Basten, and Z.-B. Zeng 2001-2004. Windows QTL Cartographer 2.0. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.

Wight, C.P., Tinker, N.A., Kianian, S.F., Sorrells, M.E., O'Donoughue, L.S., Hoffman, D.L., Groh, S., Scoles, G.J., Li, C.D., Webster, F.H., Phillips, R.L., Rines, H.W., Livingston, S.M., Armstrong, K.C., Fedak, G., Molnar, S.J. 2003. A molecular marker map in 'Kanota' x 'Ogle' hexaploid oat (*Avena spp.*) enhanced by additional markers and a robust framework. *Genome* 46: 28-47.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are usedful as genetic markers. *Nuc. Acids Res.* 18: 6531-6535.

Mueller, U.G., Wolfenbarger, L.L. 1999. AFLP genotyping and fingerprinting. *TREE* 14: 389-394.

Yan, J., Zhu, J., He, C., Benmoussa, M., Wu, P. 1999. Molecular marker assisted dissection of genotype X environment interaction for plant type traits in rice (*Oryza sativa*). *Crop Sci.* 39:538-544.

Yu, G.X., Wise, R.P. 2000. An anchored AFLP- and retrotransposon-based map of diploid *Avena*. *Genome* 43: 736-749.

Zeng, Z., Kao, C., Basten, C.J. 1999. Estimating the genetic architecture of quantitative traits. *Genet. Res.* 74: 279-289.

Zeng, Z.-B. 1993. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proc. Nat. Acad. Sci.* 90:10972-10976.

Zeng, Z-B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136:1457-1468.

Zhou, Y., Li, W., Wu, W., Chen, Q., Mao, D., Worland, A.J. 2001. Genetic dissection of heading time and its components in rice. *Theor. Appl. Genet.* 102:1236-1242.

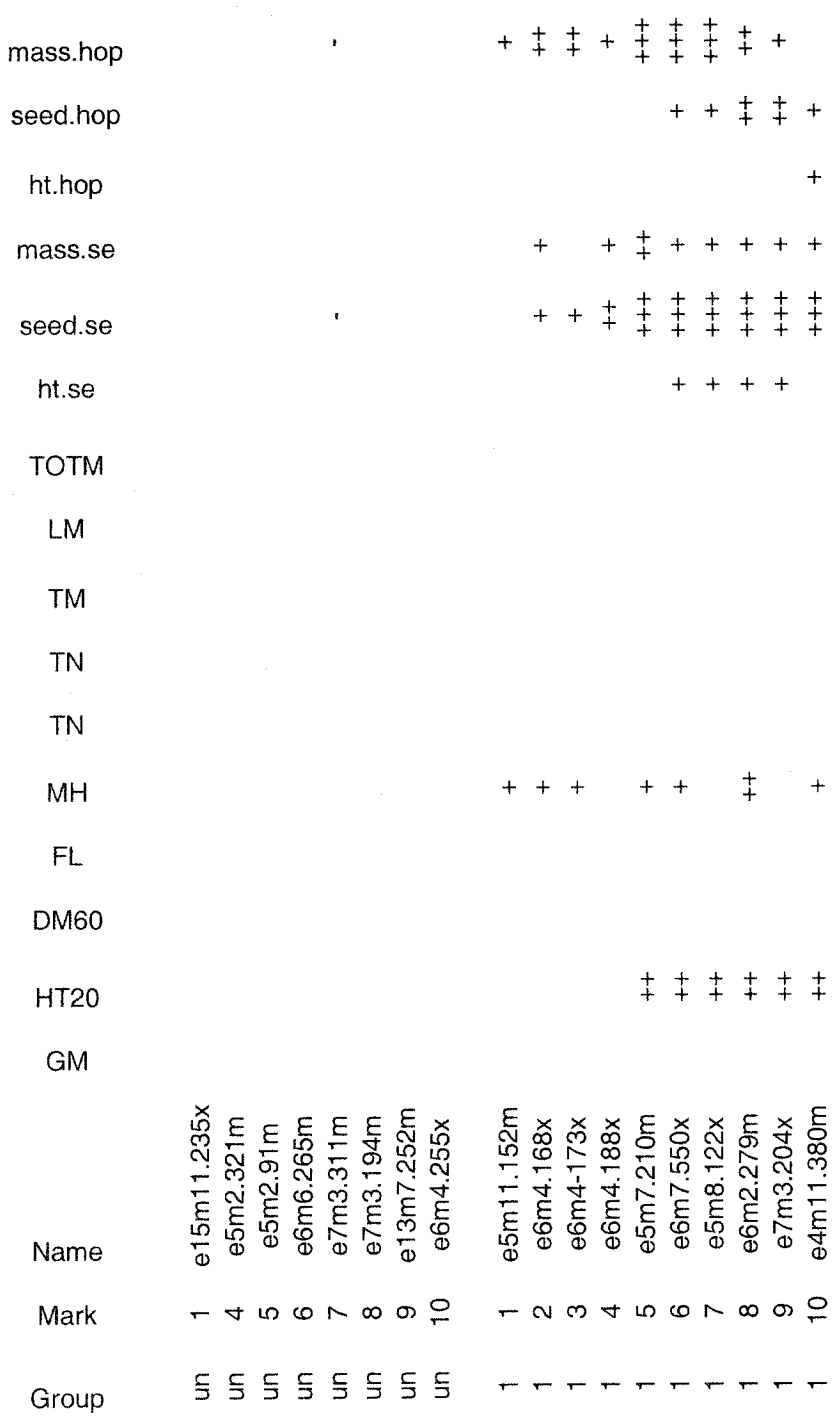
Zhu, S., Kaeppler, H.F. 2003. Identification of quantitative trait loci for resistance to crown rust in oat line MAM17-5. *Crop Sci.* 43:358-366.

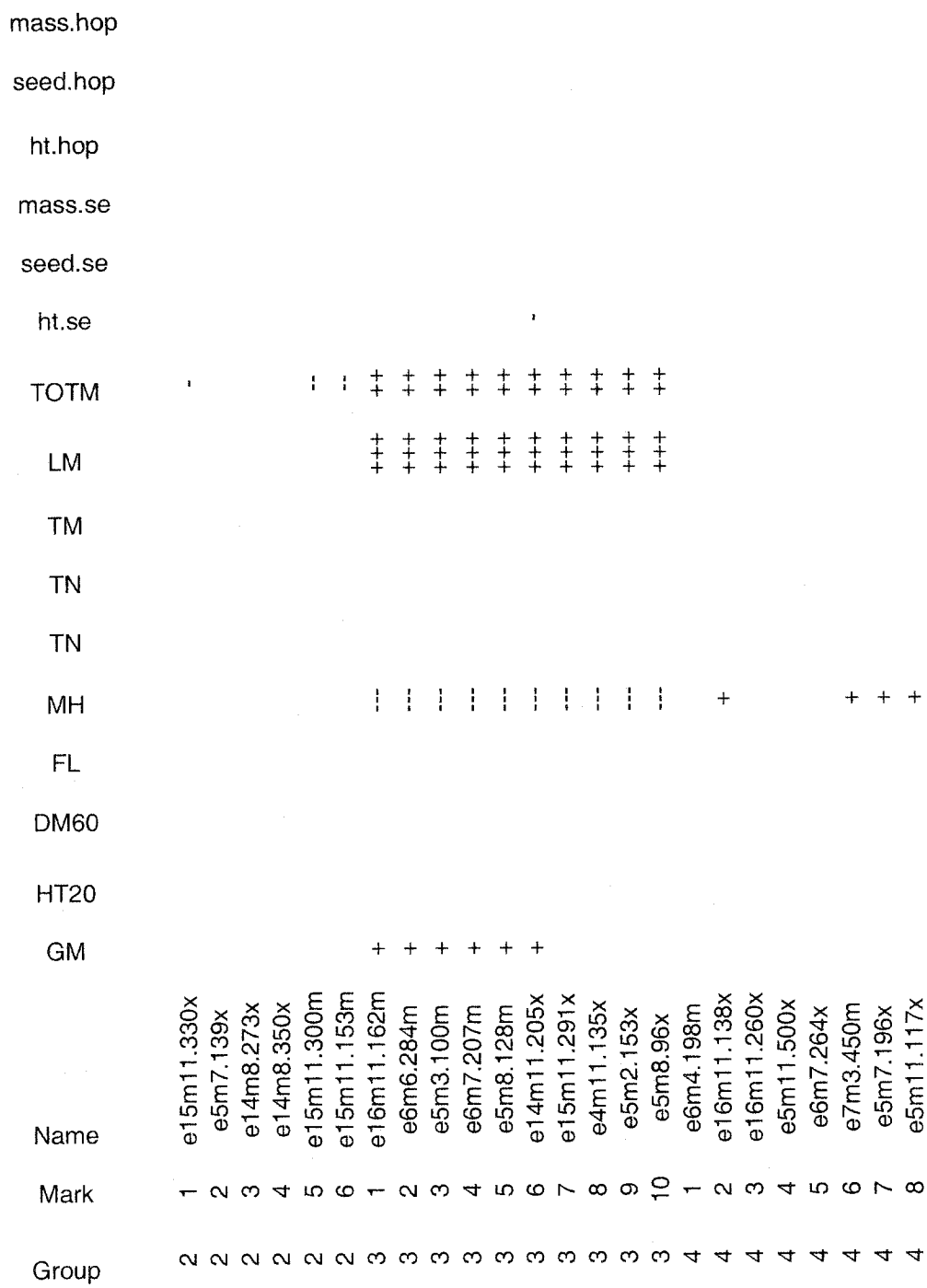
Zhuang, J.Y., Fan, Y.Y., Rao, Z.M., Wu, J.L., Xia, Y.W., Zheng, K.L. 2002. Analysis on additive effects and additive by additive epistatic effects of QTLs for yield traits in a recombinant inbred line population of rice. *Theor. Appl. Genet.* 105:1137-1145.

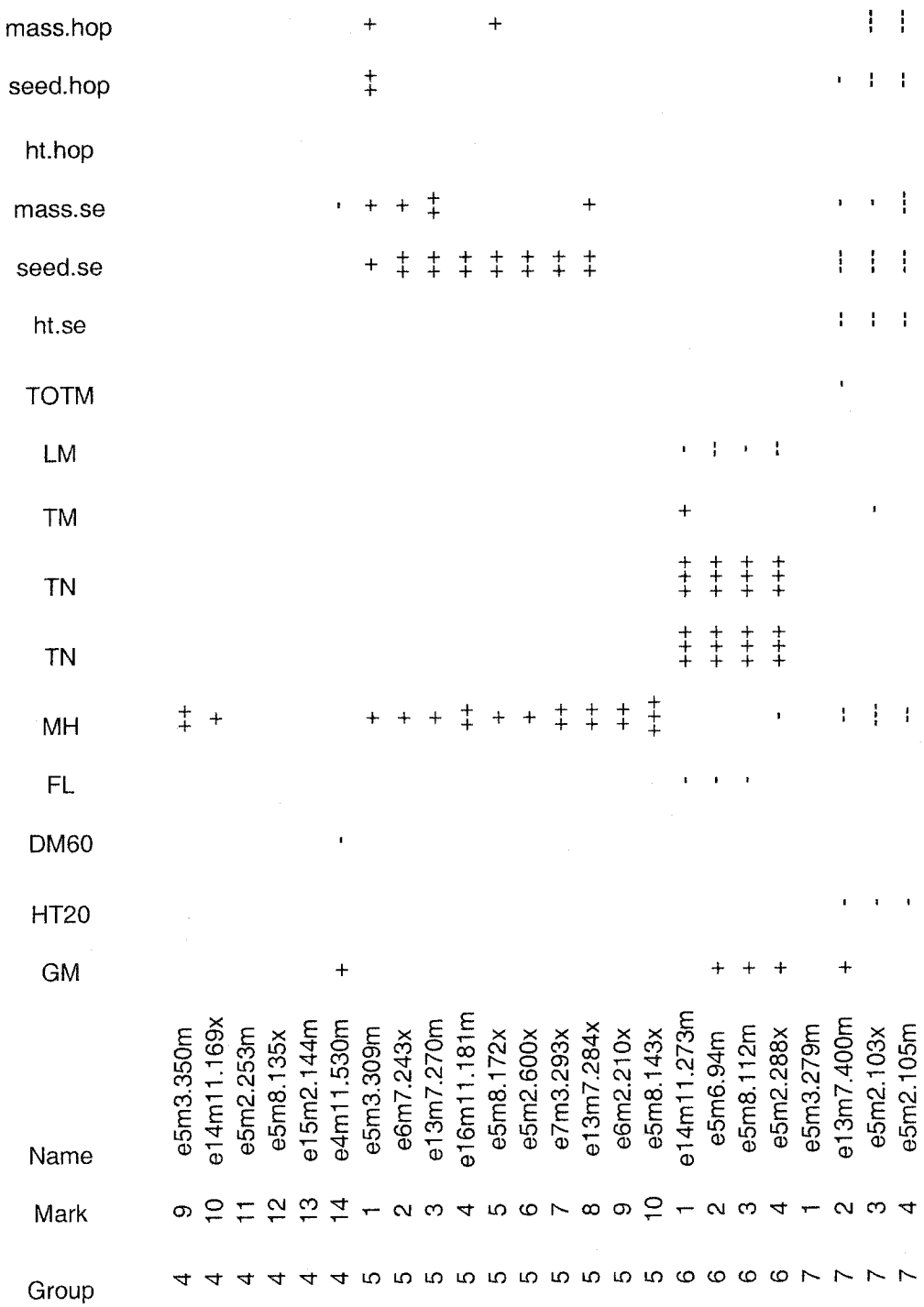
Appendix One

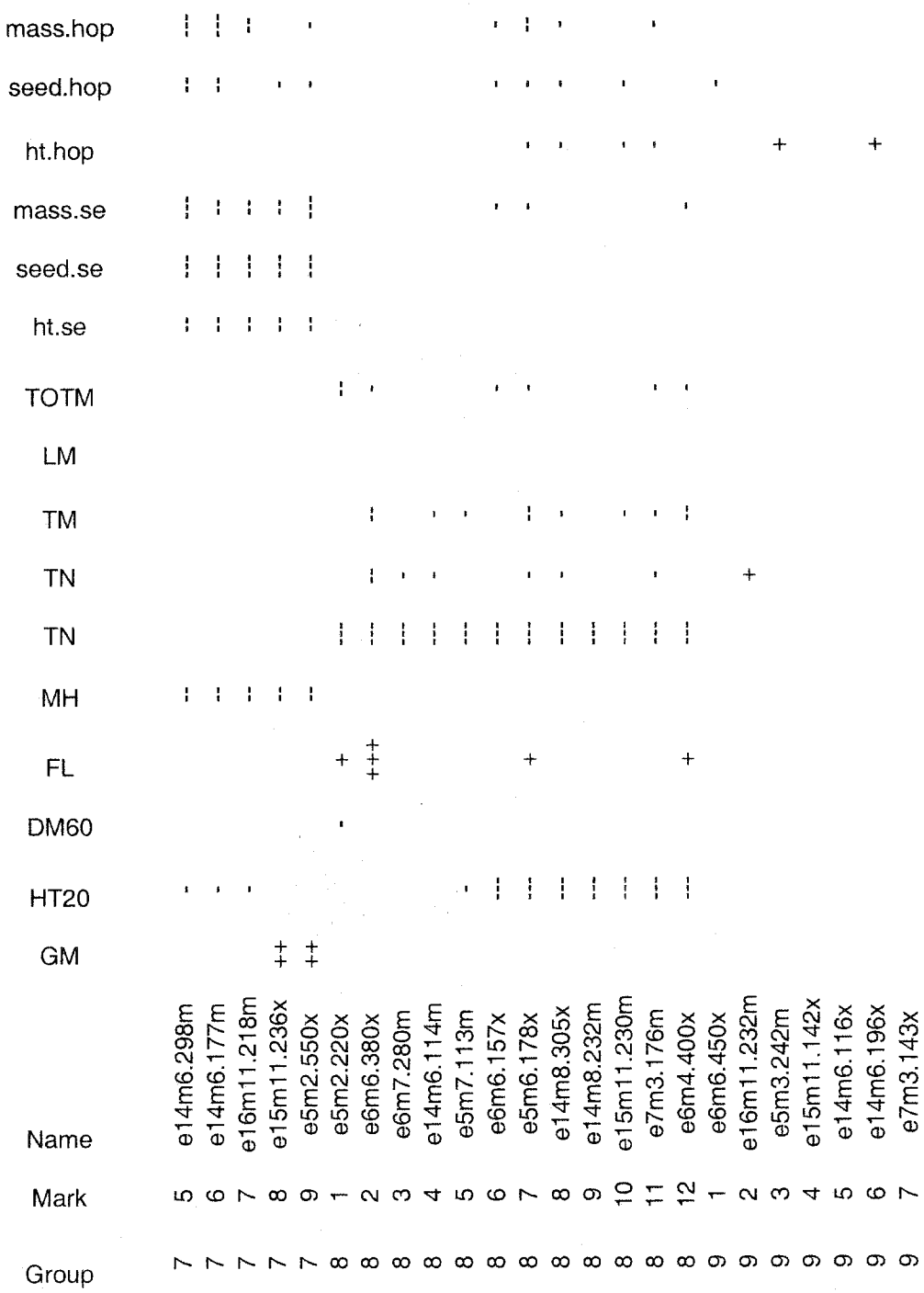
Single Marker Analyses

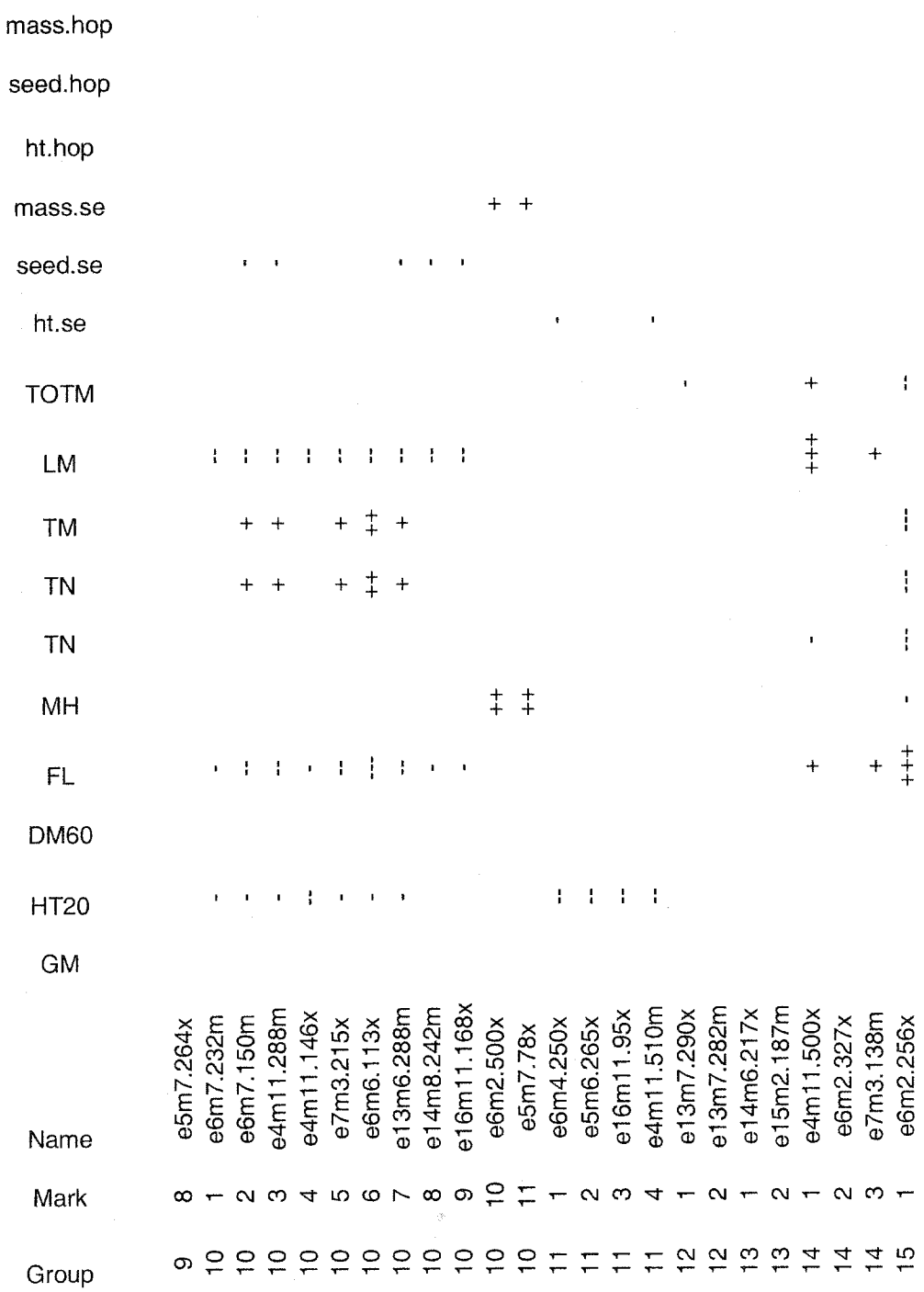
Results of single marker tests are given on the following pages. Marker loci are given in order of their position within linkage groups, preceded by the unlinked markers. Direction of the marker effect is given by "+" if the xeric allele increases the trait value and "-" if the xeric decreases the trait value. Statistical significance is indicated by the number of +'s and –'s for each marker. (e.g. +++ = $p<0.001$, ++ = $p< 0.01$, + = $p<0.05$)

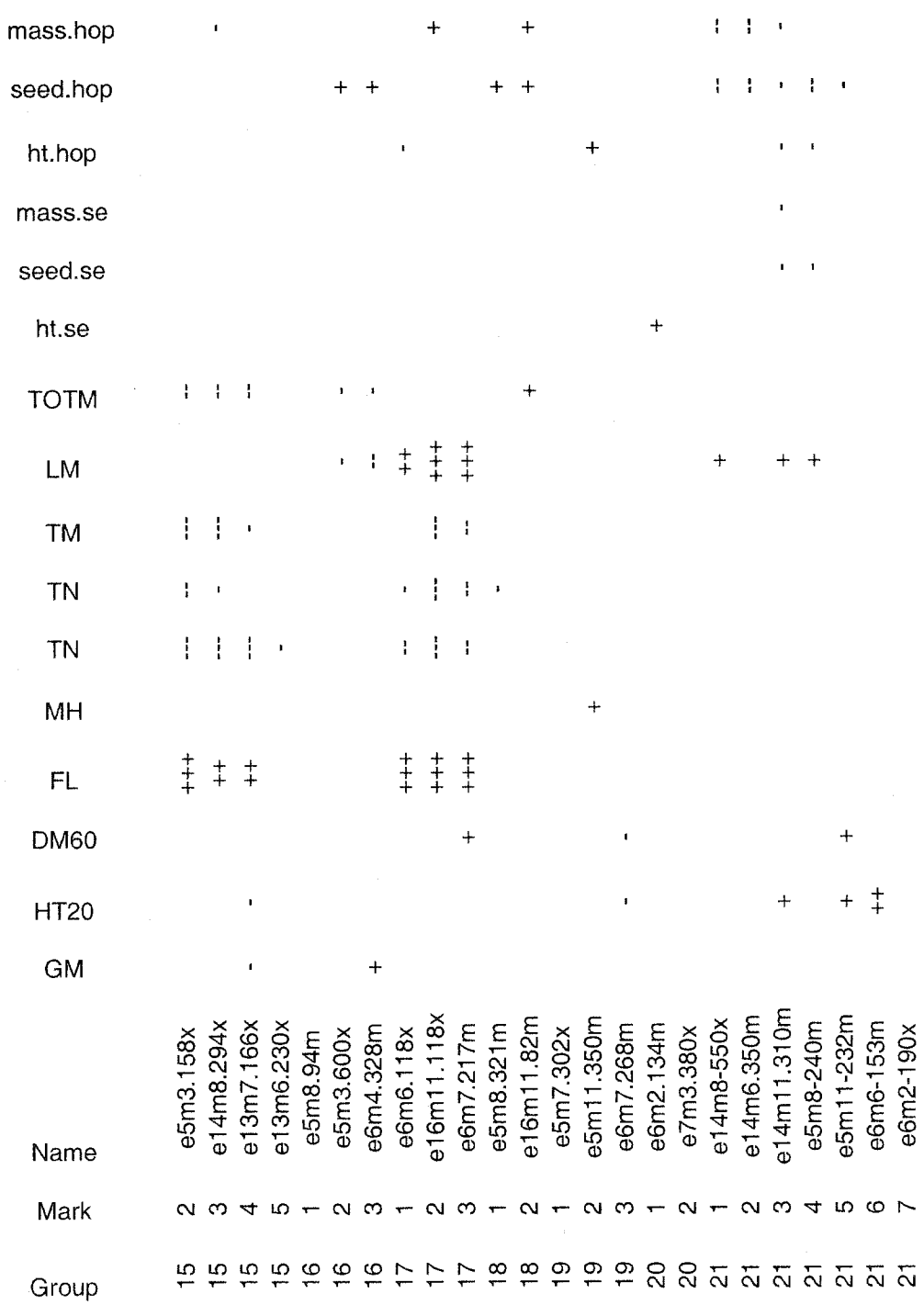

mass.hop
seed.hop
ht.hop
mass.se
seed.se
ht.se
TOTM
LM
TM
TN
TN
MH
FL
DM60
HT20
GM
Name
Mark
Group
e15m11.235x
e5m2.321m
e5m2.91m
e6m6.265m
e7m3.311m
e7m3.194m
e13m7.252m
e6m4.255x
e5m11.152m
e6m4.168x
e6m4-173x
e6m4.188x
e5m7.210m
e6m7.550x
e5m8.122x
e6m2.279m
e7m3.204x
e4m11.380m


mass.hop
seed.hop
ht.hop
mass.se
seed.se
ht.se
TOTM
LM
TM
TN
TN
MH
FL
DM60
HT20
GM
Name
Mark
Group
e15m11.330x
e5m7.139x
e14m8.273x
e14m8.350x
e15m11.300m
e15m11.153m
e16m11.162m
e6m6.284m
e5m3.100m
e6m7.207m
e5m8.128m
e14m11.205x
e15m11.291x
e4m11.135x
e5m2.153x
e5m8.96x
e6m4.198m
e16m11.138x
e16m11.260x
e5m11.500x
e6m7.264x
e7m3.450m
e5m7.196x
e5m11.117x


mass.hop
seed.hop
ht.hop
mass.se
seed.se
ht.se
TOTM
LM
TM
TN
TN
MH
FL
DM60
HT20
GM
Name
Mark
Group
e5m3.350m 9 4
e14m11.169x 10 4
e5m2.253m 11 4
e5m8.135x 12 4
e15m2.144m 13 4
e4m11.530m 14 4
e5m3.309m 1 5
e6m7.243x 2 5
e13m7.270m 3 5
e16m11.181m 4 5
e5m8.172x 5 5
e5m2.600x 6 5
e7m3.293x 7 5
e13m7.284x 8 5
e6m2.210x 9 5
e5m8.143x 10 5
e14m11.273m 1 6
e5m6.94m 2 6
e5m8.112m 3 6
e5m2.288x 4 6
e5m3.279m 1 7
e13m7.400m 2 7
e5m2.103x 3 7
e5m2.105m 4 7


mass.hop
seed.hop
ht.hop
mass.se
seed.se
ht.se
TOTM
LM
TM
TN
TN
MH
FL
DM60
HT20
GM
Name
Mark
Group
e14m6.298m 5 7
e14m6.177m 6 7
e16m11.218m 7 7
e15m11.236x 8 7
e5m2.550x 9 7
e5m2.220x 1 8
e6m6.380x 2 8
e6m7.280m 3 8
e14m6.114m 4 8
e5m7.113m 5 8
e6m6.157x 6 8
e5m6.178x 7 8
e14m8.305x 8 8
e14m8.232m 9 8
e15m11.230m 10 8
e7m3.176m 11 8
e6m4.400x 12 8
e6m6.450x 1 9
e16m11.232m 2 9
e5m3.242m 3 9
e15m11.142x 4 9
e14m6.116x 5 9
e14m6.196x 6 9
e7m3.143x 7 9


mass.hop
seed.hop
ht.hop
mass.se
seed.se
ht.se
TOTM
LM
TM
TN
TN
MH
FL
DM60
HT20
GM
Name
Mark
Group
9 8 e5m7.264x
10 1 e6m7.232m
10 2 e6m7.150m
10 3 e4m11.288m
10 4 e4m11.146x
10 5 e7m3.215x
10 6 e6m6.113x
10 7 e13m6.288m
10 8 e14m8.242m
10 9 e16m11.168x
10 10 e6m2.500x
10 11 e5m7.78x
11 1 e6m4.250x
11 2 e5m6.265x
11 3 e16m11.95x
11 4 e4m11.510m
12 1 e13m7.290x
12 2 e13m7.282m
13 1 e14m6.217x
13 2 e15m2.187m
14 1 e4m11.500x
14 2 e6m2.327x
14 3 e7m3.138m
15 1 e6m2.256x


mass.hop
seed.hop
ht.hop
mass.se
seed.se
ht.se
TOTM
LM
TM
TN
TN
MH
FL
DM60
HT20
GM
Name
Mark
Group
15 2 e5m3.158x
15 3 e14m8.294x
15 4 e13m7.166x
15 5 e13m6.230x
16 1 e5m8.94m
16 2 e5m3.600x
16 3 e6m4.328m
17 1 e6m6.118x
17 2 e16m11.118x
17 3 e6m7.217m
18 1 e5m8.321m
18 2 e16m11.82m
19 1 e5m7.302x
19 2 e5m11.350m
19 3 e6m7.268m
20 1 e6m2.134m
20 2 e7m3.380x
21 1 e14m8-550x
21 2 e14m6.350m
21 3 e14m11.310m
21 4 e5m8-240m
21 5 e5m11-232m
21 6 e6m6-153m
21 7 e6m2-190x

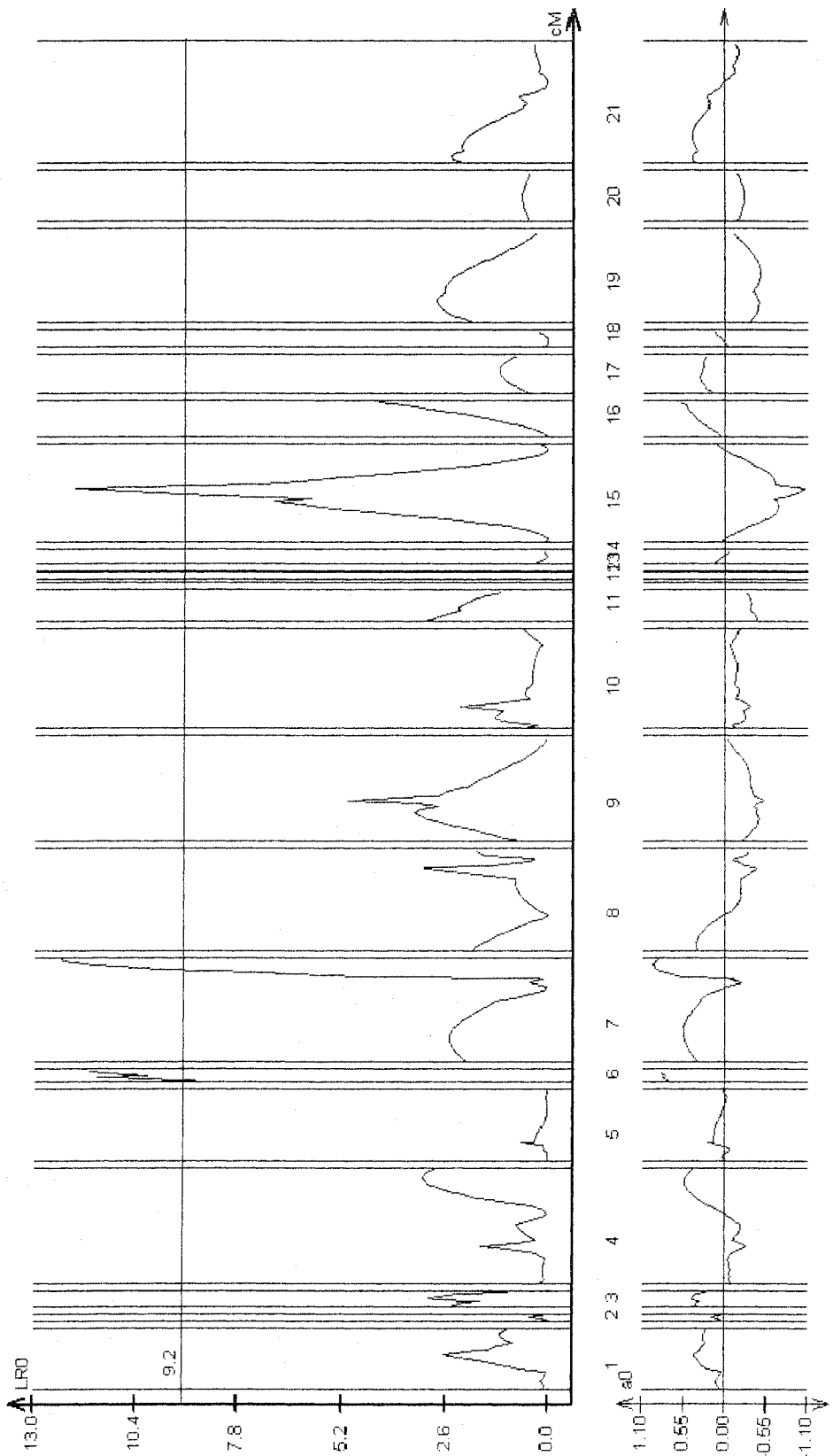
mass.hop			
seed.hop			
ht.hop			
mass.se			
seed.se			
ht.se			
TOTM			
LM			
TM			
TN			+
TN			
MH			
FL			
DM60			
HT20			
GM			
Name	e6m7-170x	e13m7-316m	e16m11-106m
Mark	8	9	10
Group	21	21	21

Appendix Two

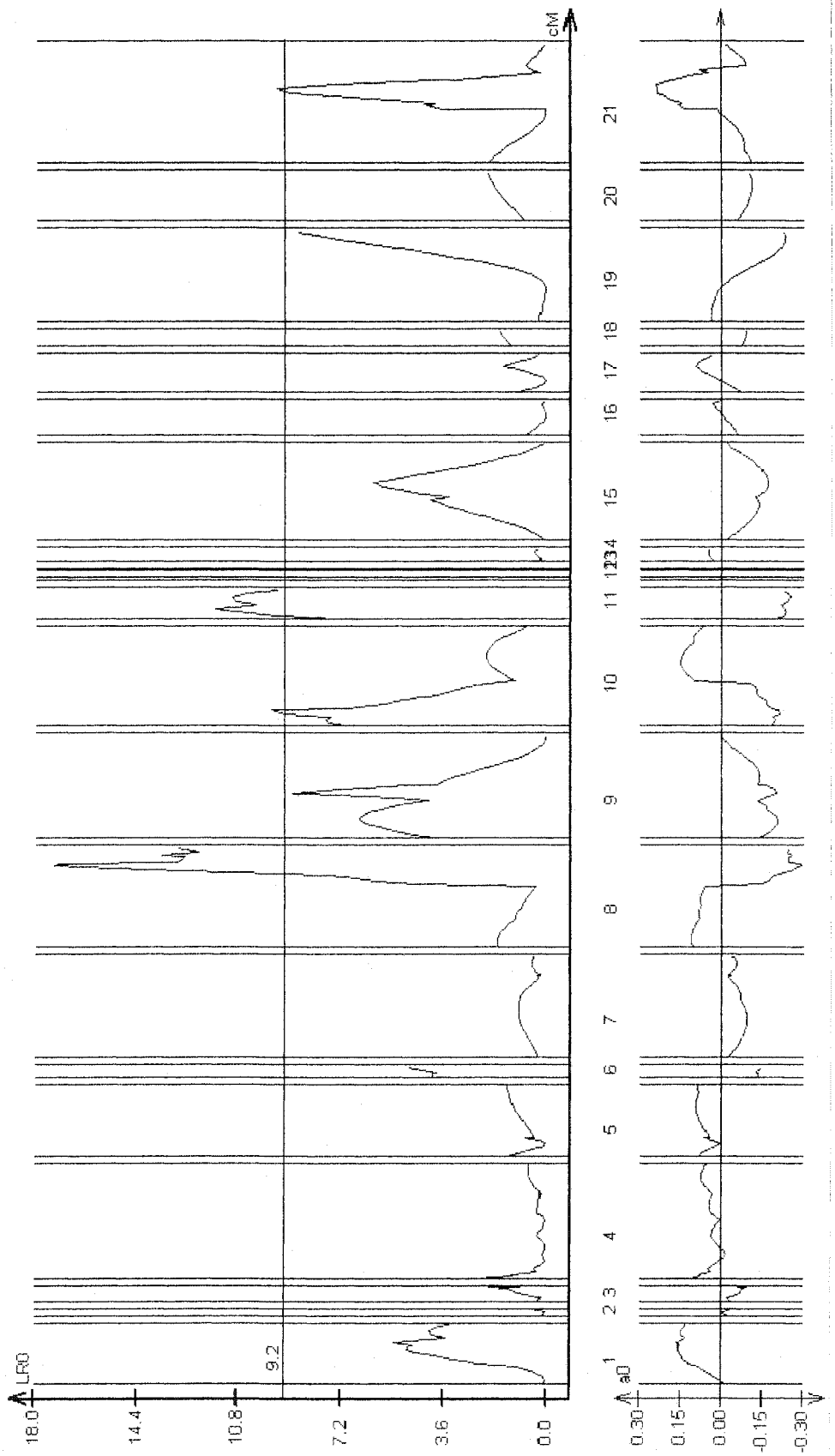
Significance Thresholds and Likelihood Ratio Profiles for all Greenhouse and Field Traits

Likelihood ratio significance thresholds and likelihood ratio profiles for greenhouse and field traits. Thresholds (below) are derived from 1000 permutations of the data following the method of Churchill and Doerge (1994). Likelihood ratio (top panel) and additive effect (lower panel) profiles for each linkage group are given on the following pages. Linkage group number is indicated between panels.

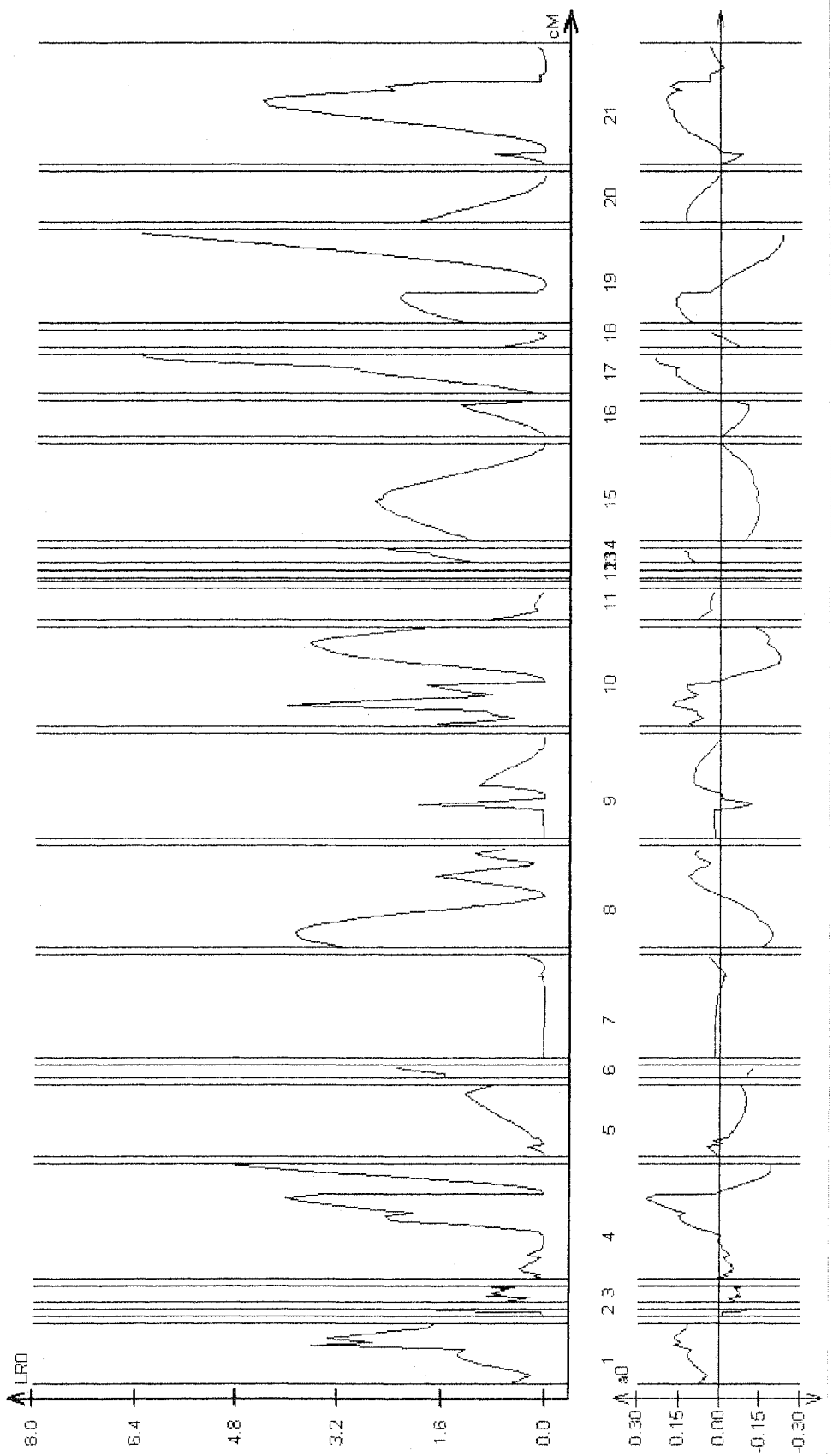
	Highly Significant (alpha = 0.01)	Significant (alpha = 0.05)
GM	19.0	12.4
HT20	19.2	12.3
DM60	17.3	12.0
FL	21.4	12.8
MH	20.2	12.2
TN	18.0	11.9
SN	23.7	12.1
TM	22.6	12.1
LM	18.5	12.0
TOTM	19.0	12.0
ht.se	20.1	11.9
seed.se	17.9	12.1
mass.se	19.0	11.3
ht.hop	17.6	11.8
seed.hop	20.0	10.7
mass.hop	17.4	11.9
PC1	21.4	12.0
PC2	19.0	12.0
PC3	24.3	12.4
PC4	19.0	11.9
PC5	19.3	11.9
PC6	18.7	11.5



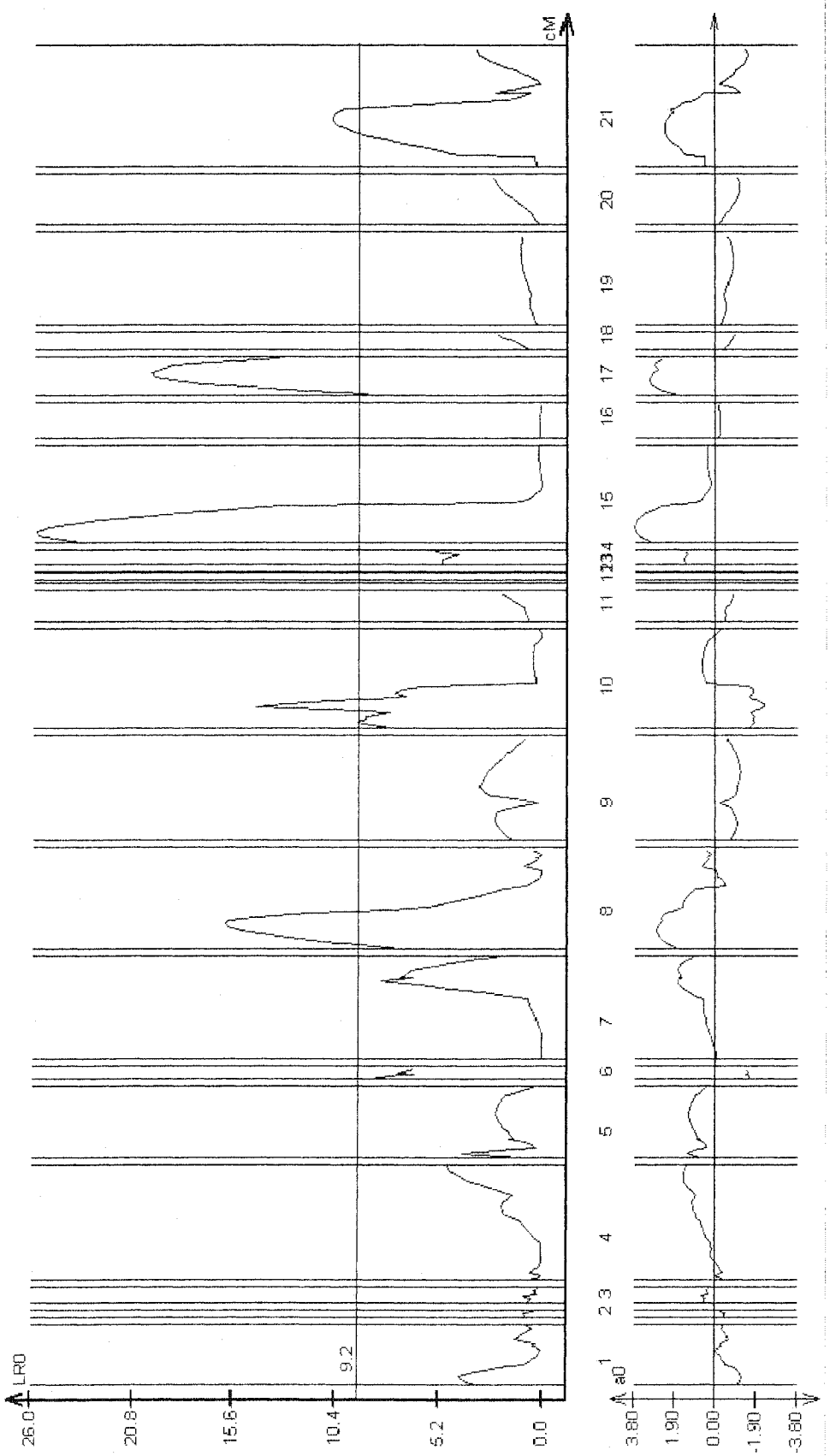
Germination time (GM).



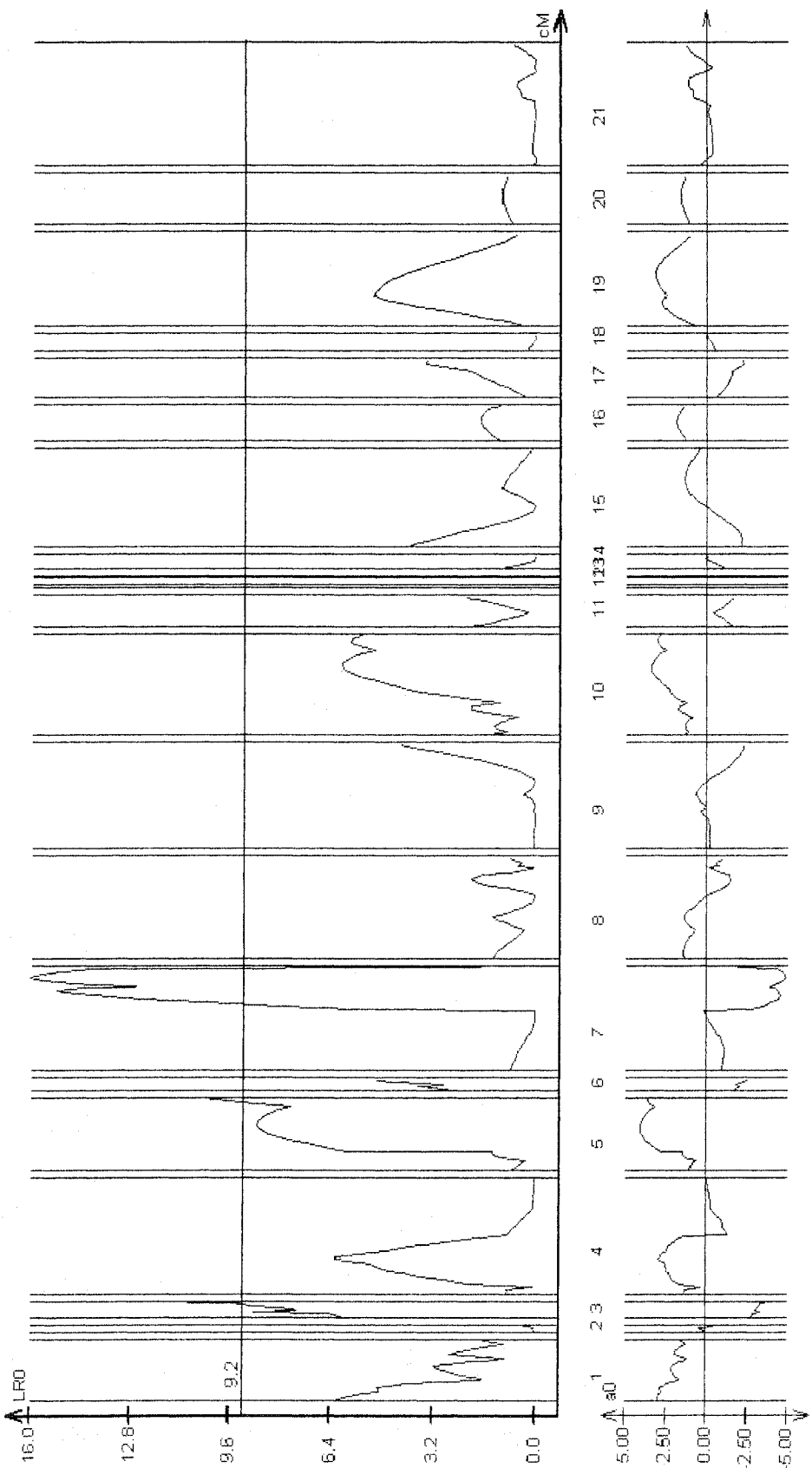
Height at 20 days (HT20).



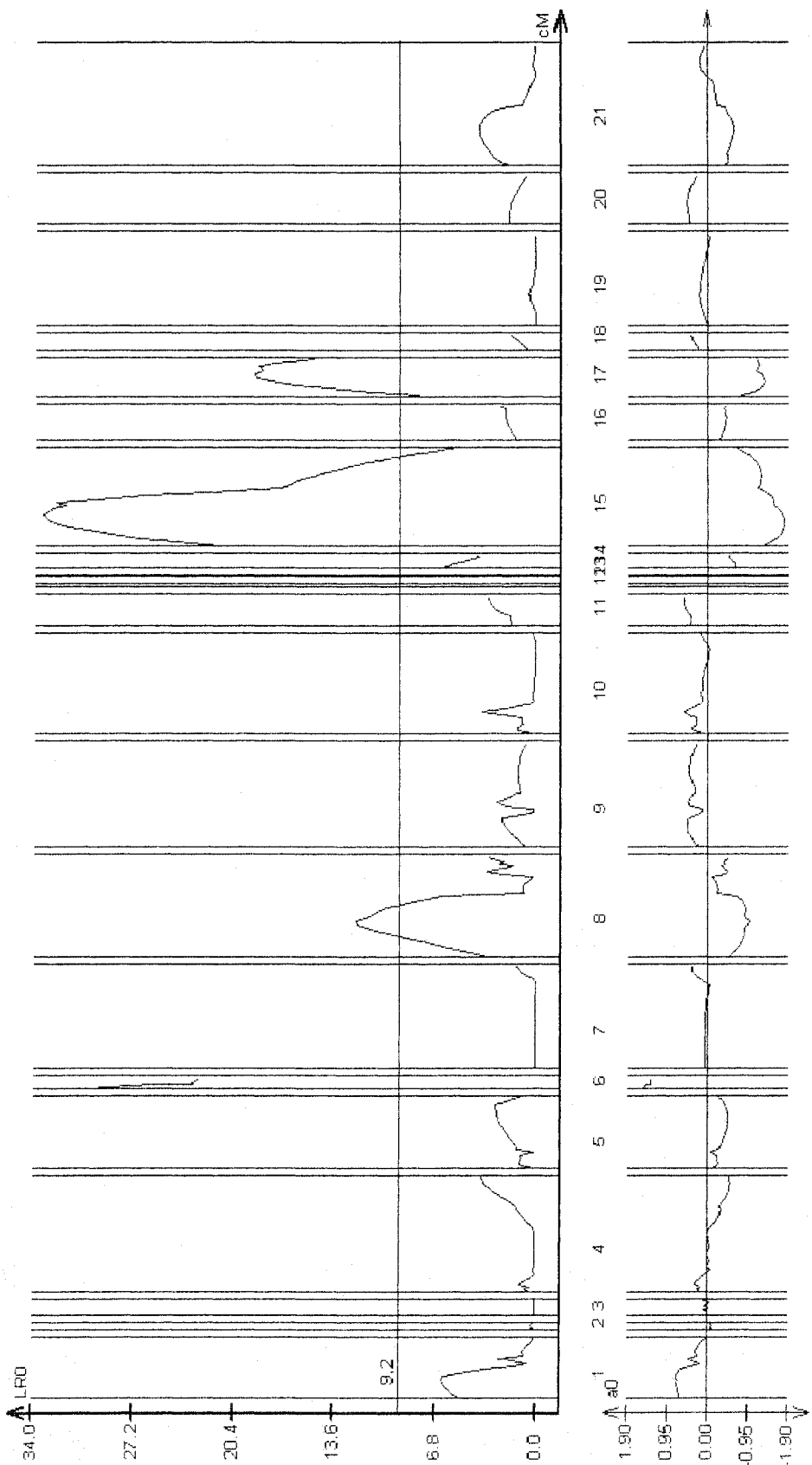
Dry mass at 60 days (DM60).



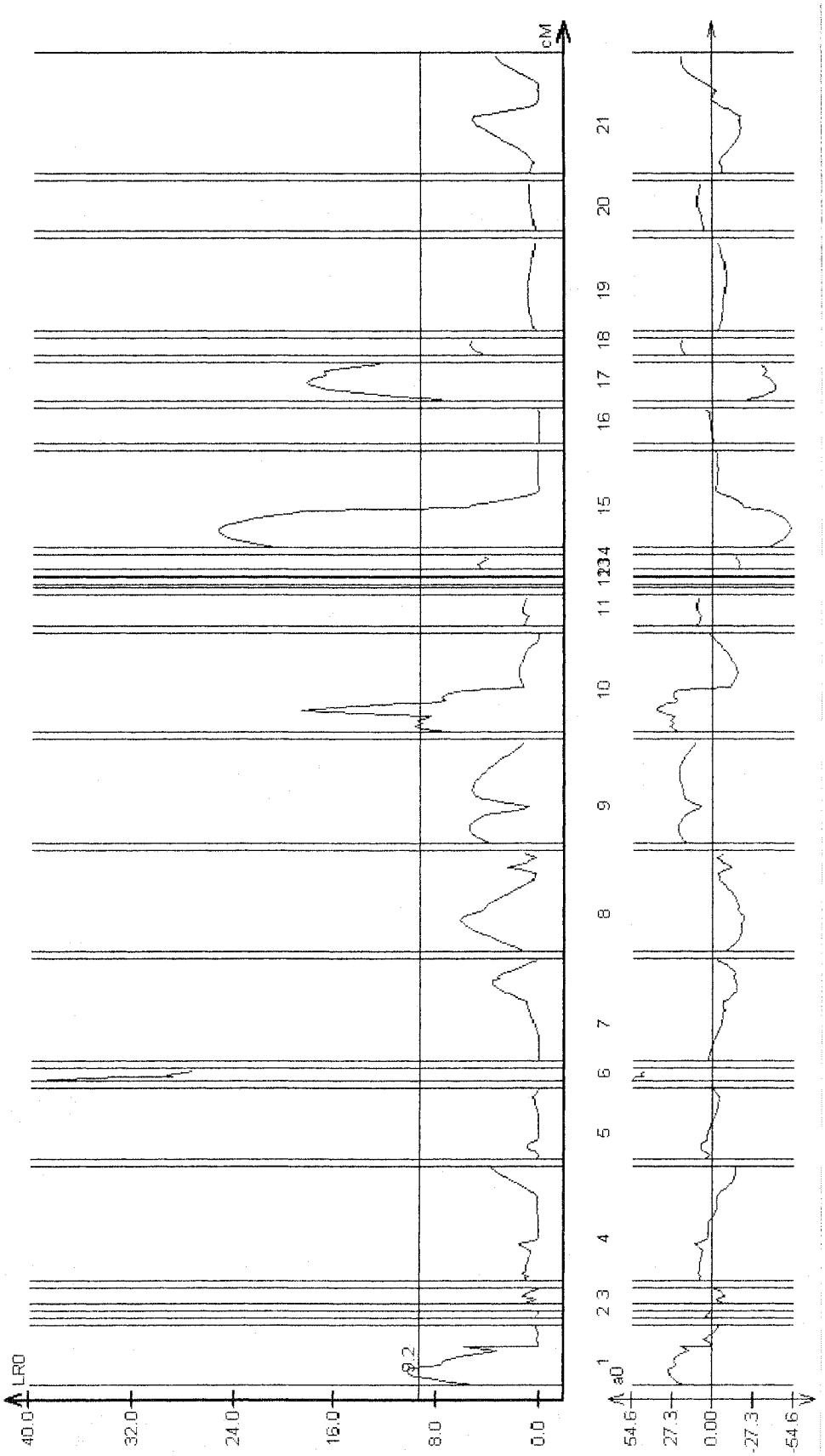
Flowering time (FL).



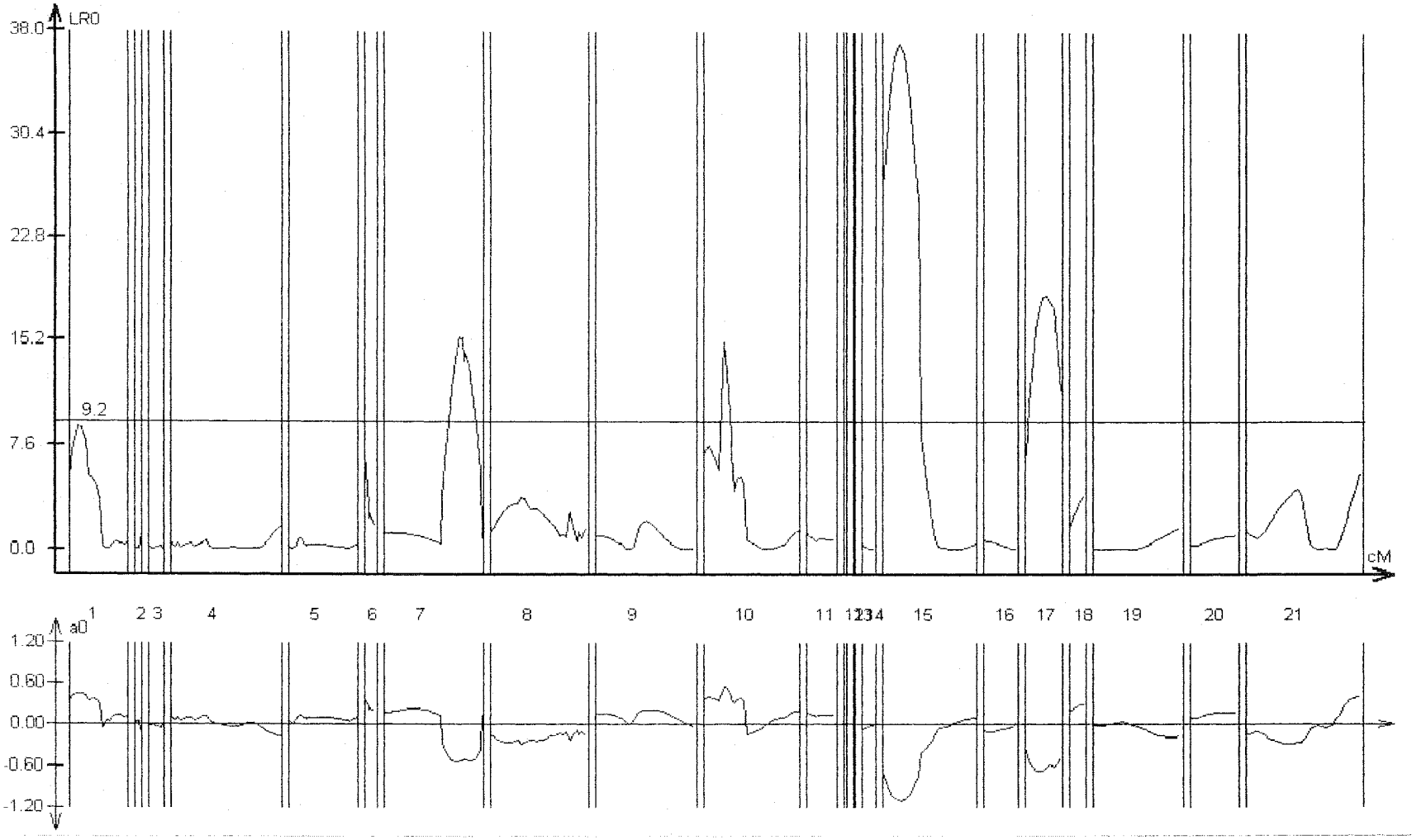
Maiximum height (MH).



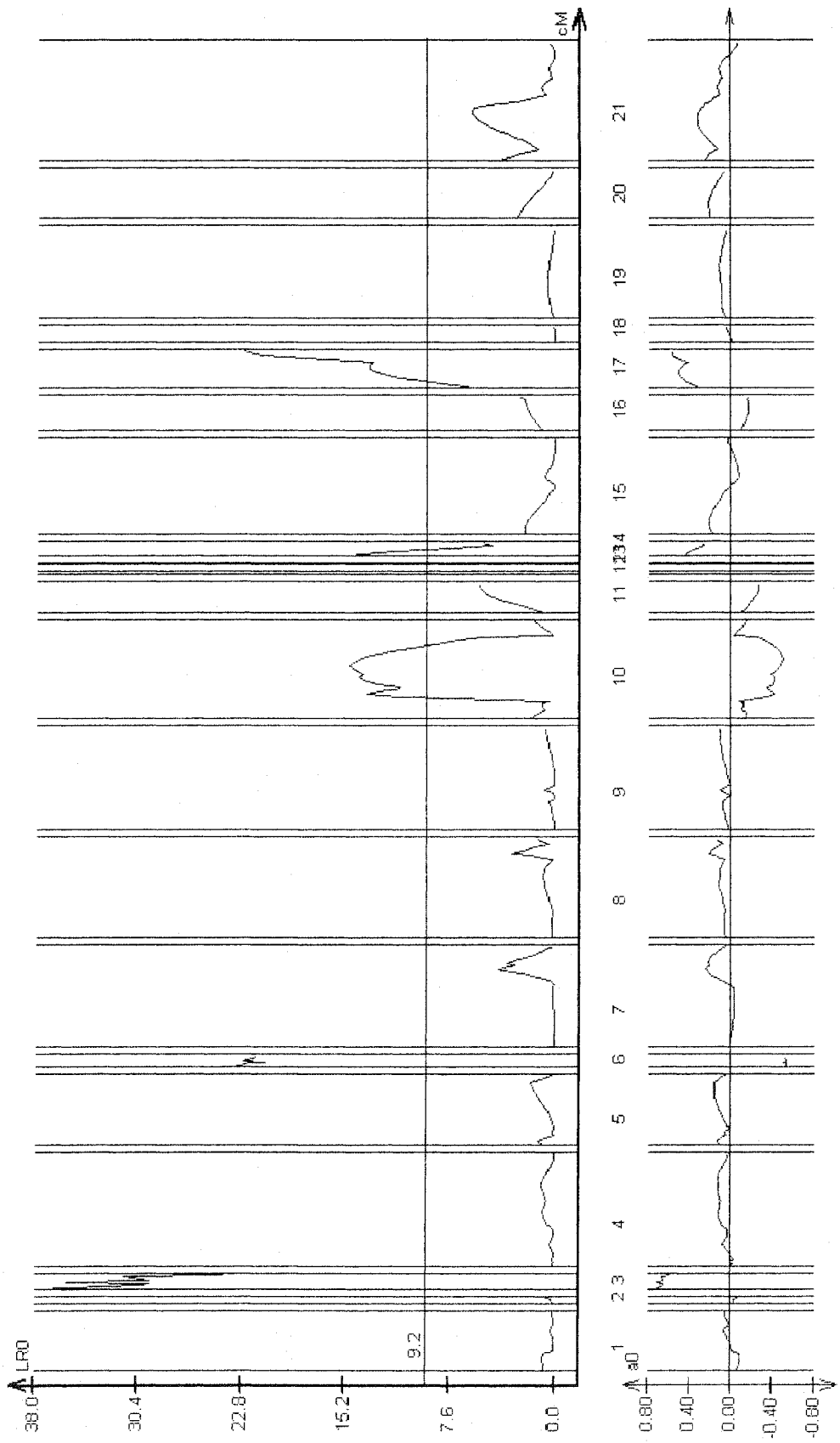
Number of tillers (TN).



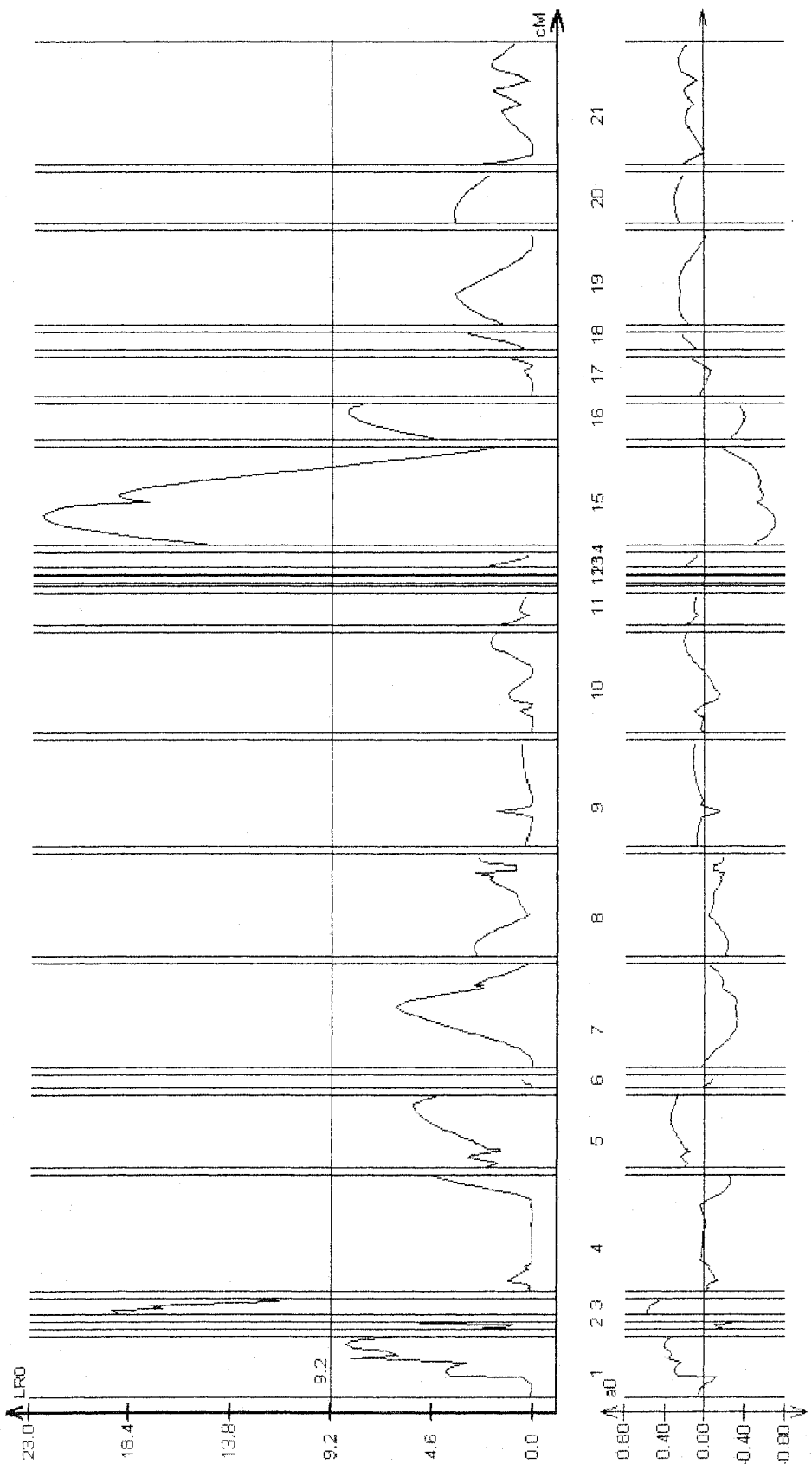
Number of Spikelets. (SN).



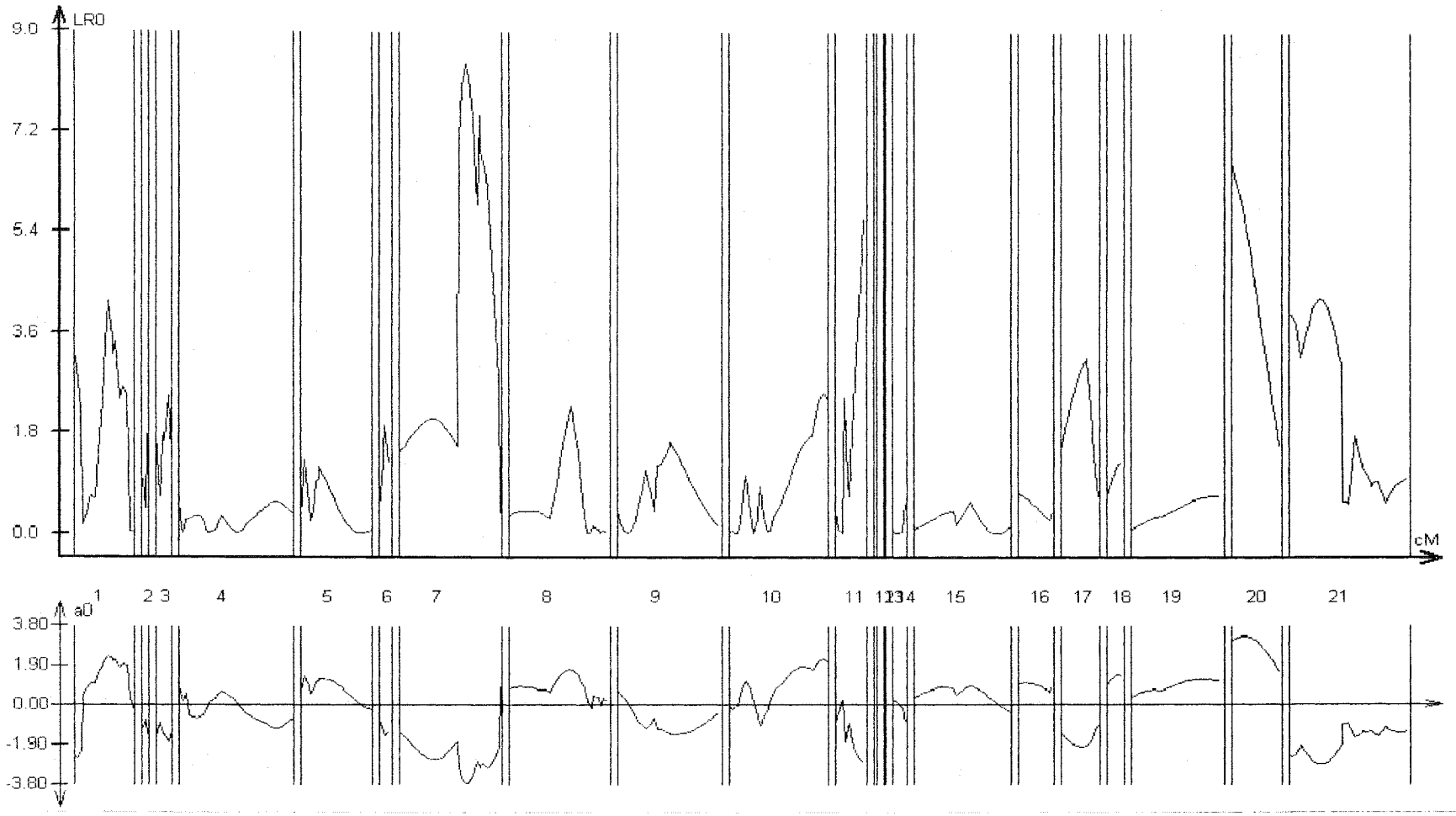
Tiller mass (TM).


LRD
38.0
30.4
22.8
15.2
7.6
0.0
9.2
cM
1
2 3
4
5
6
7
8
9
10
11 12 13 14
15
16
17
18
19
20
21
aD
0.80
0.40
0.00
-0.40
-0.80

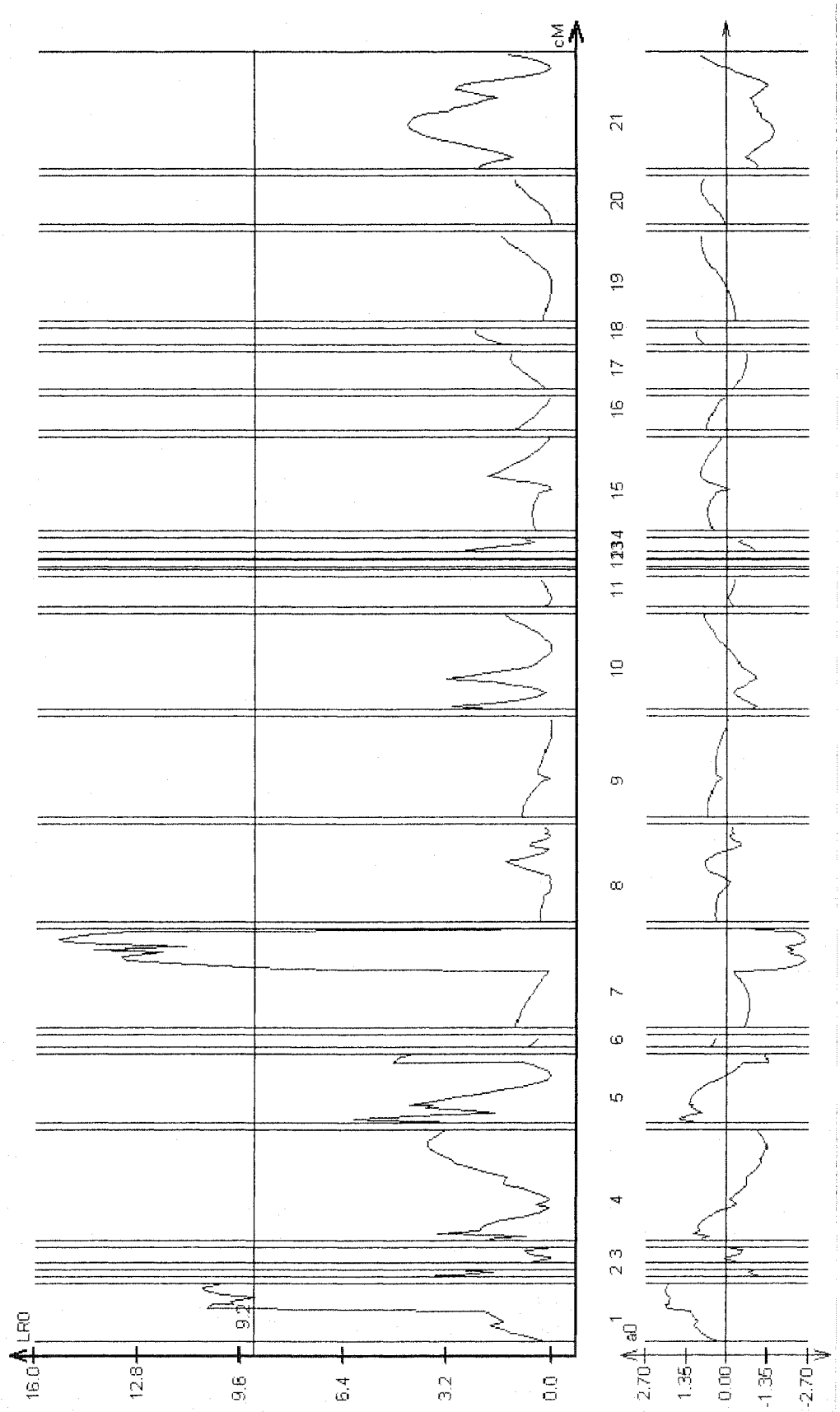
Leaf mass (LM).



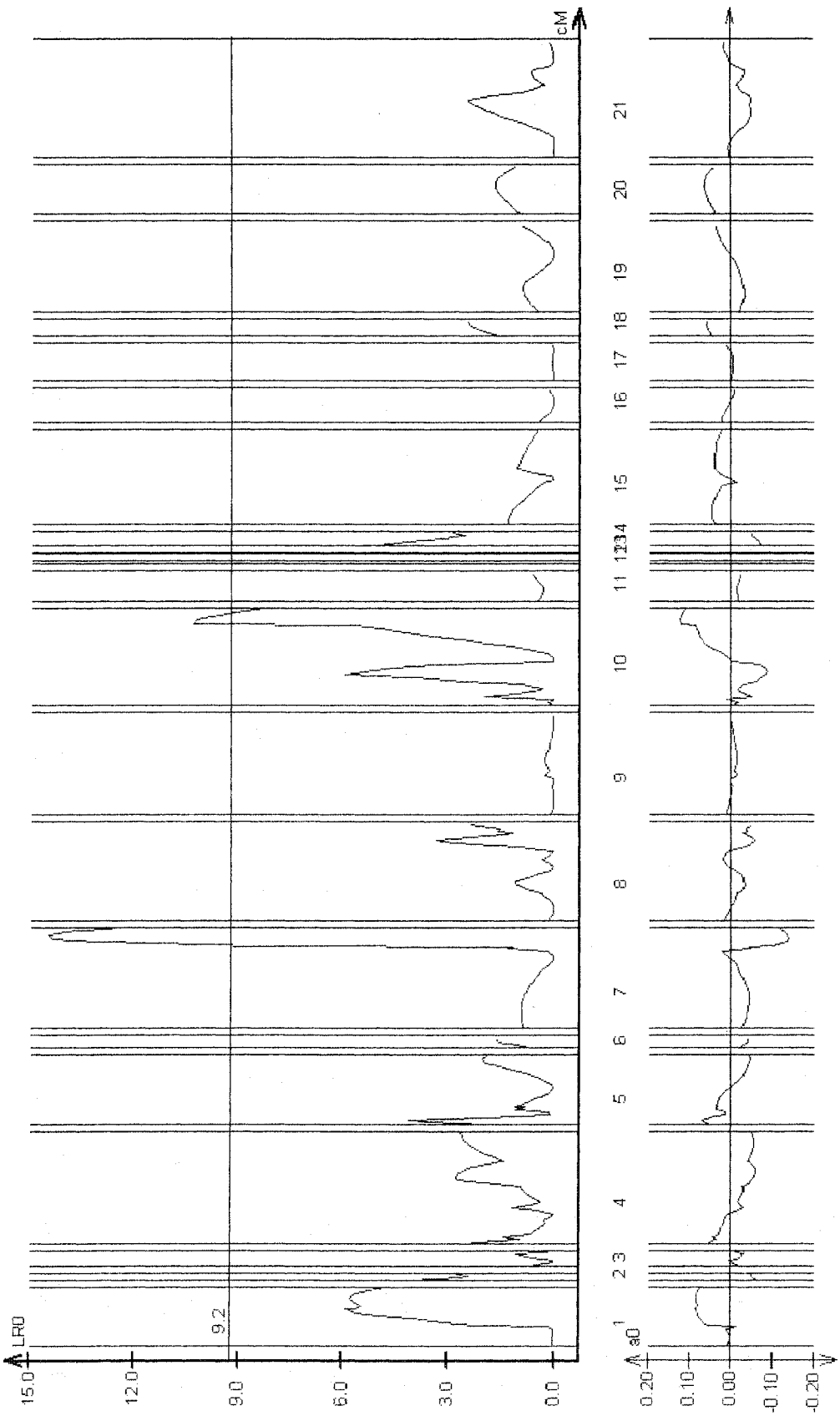
Total mass (TOTM).



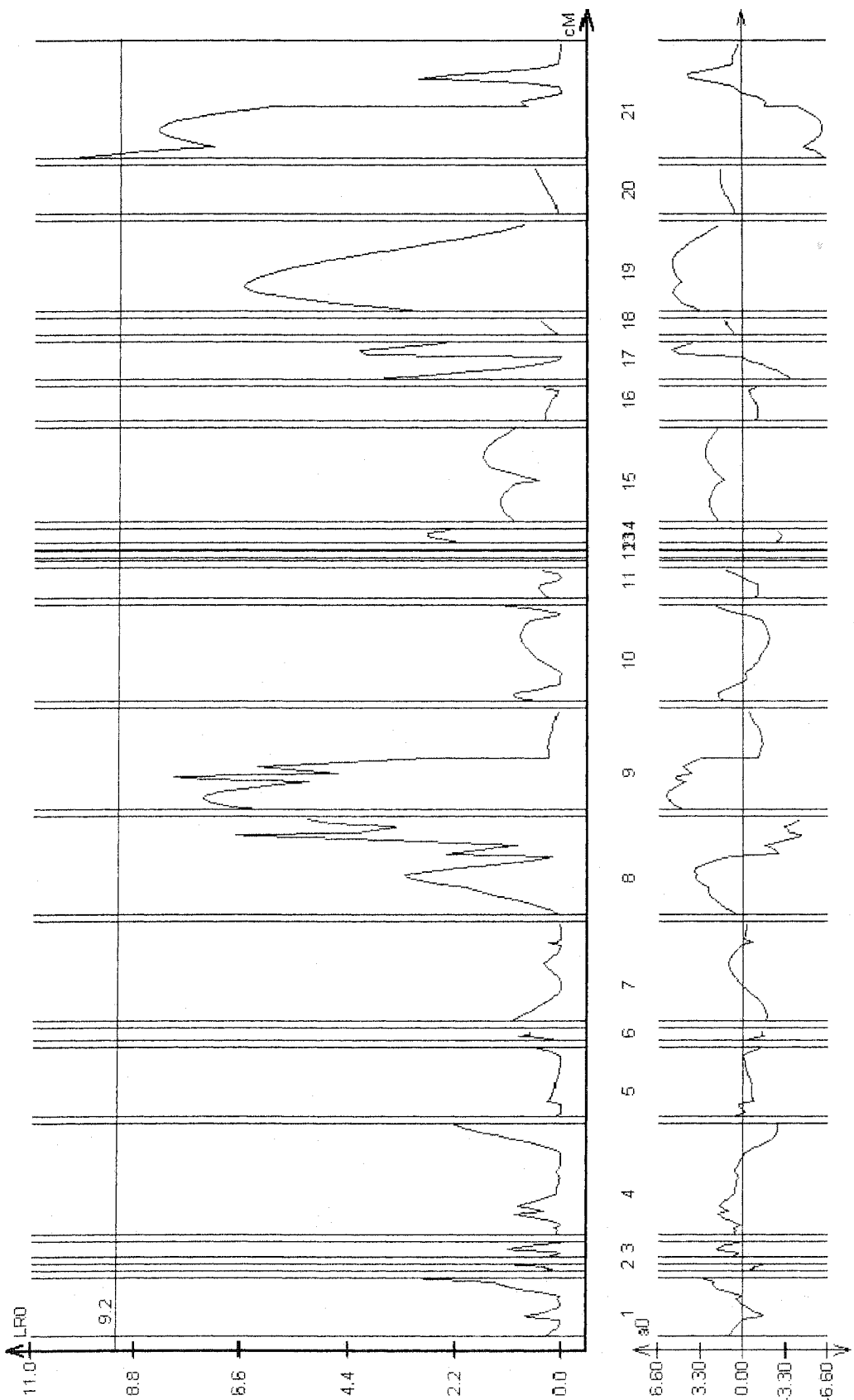
Height at Sierra (ht.se).



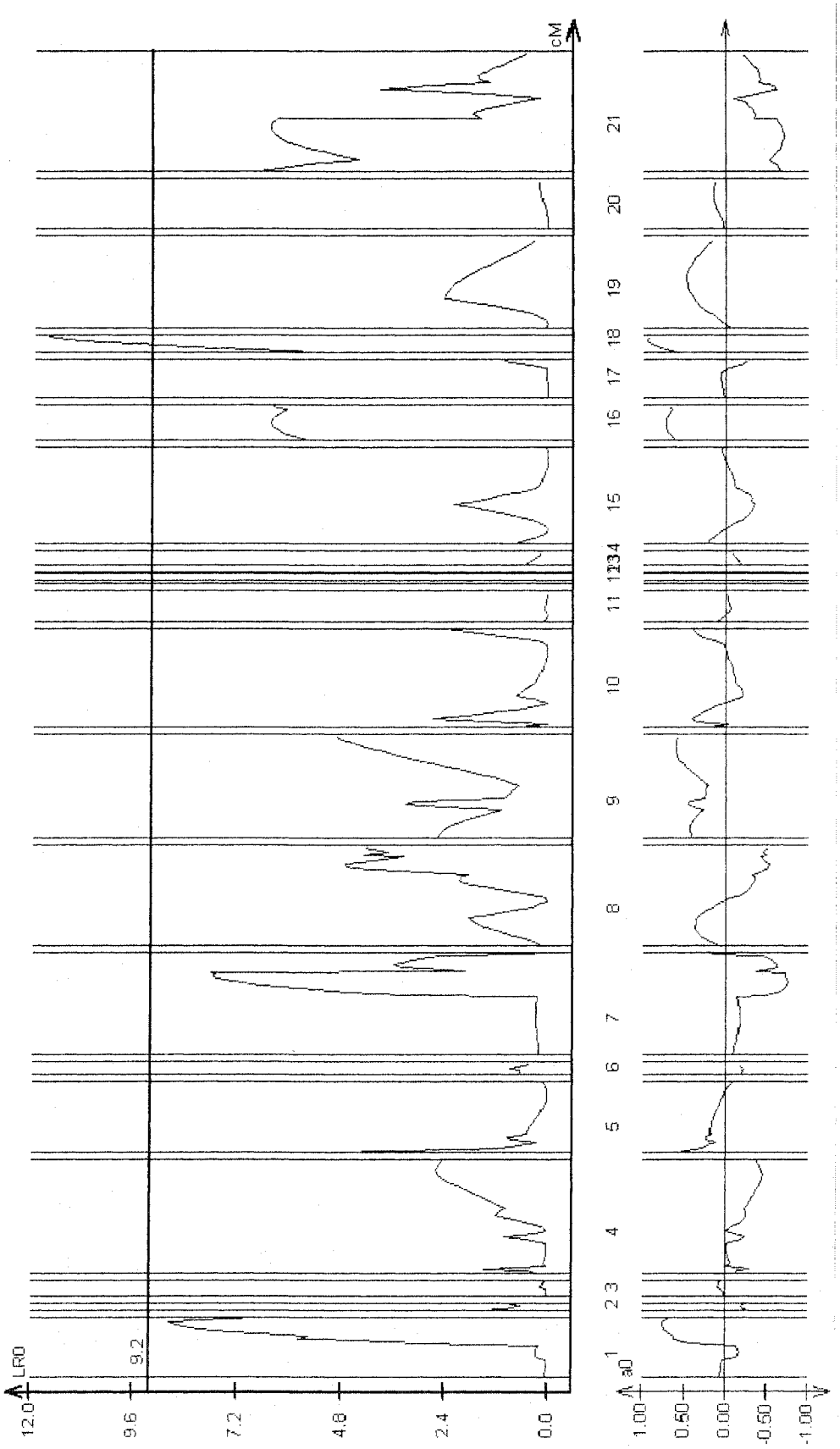
Number of spikelets at Sierra (spike.se).



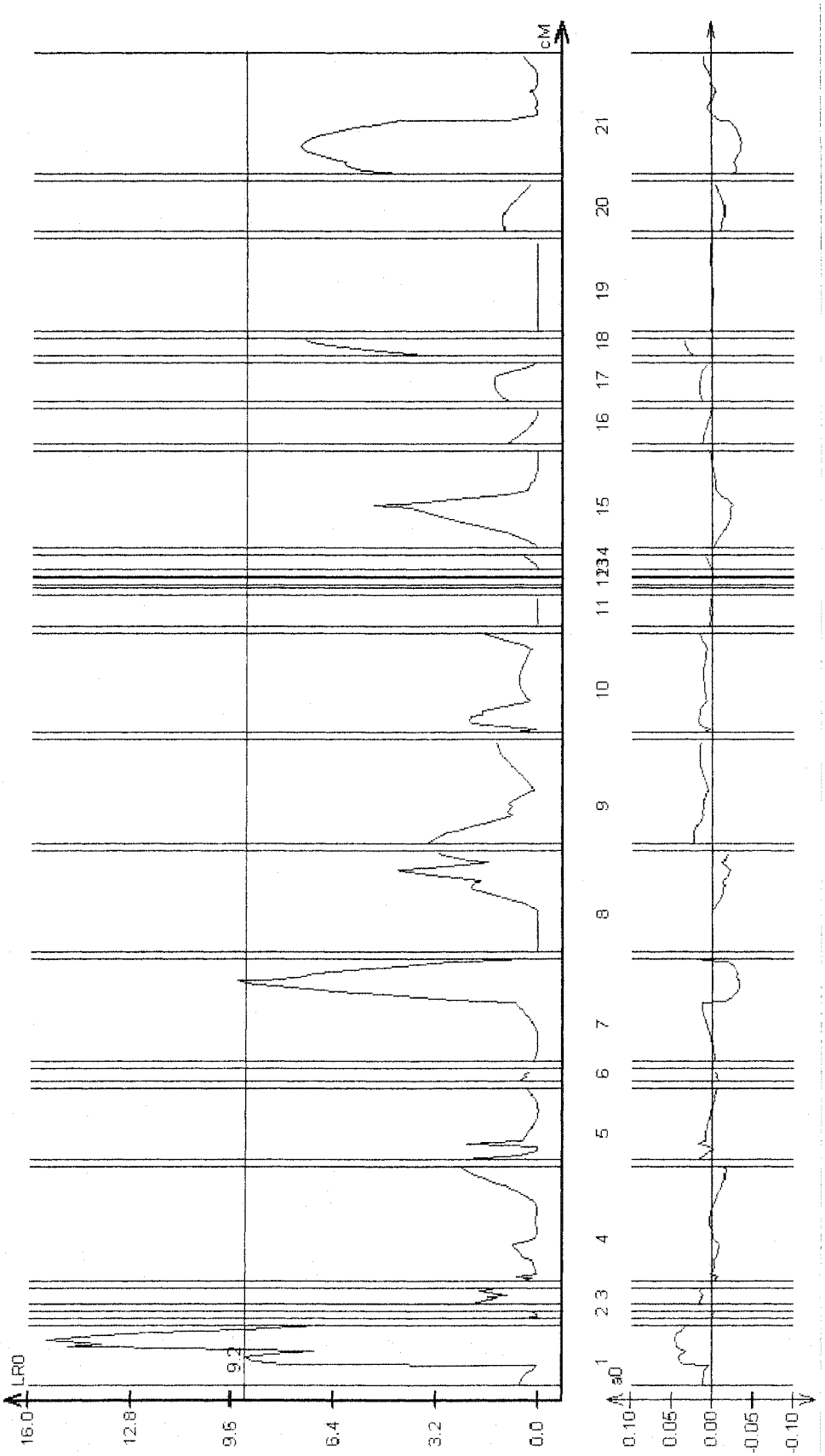
Mass at Sierra (mass.se).



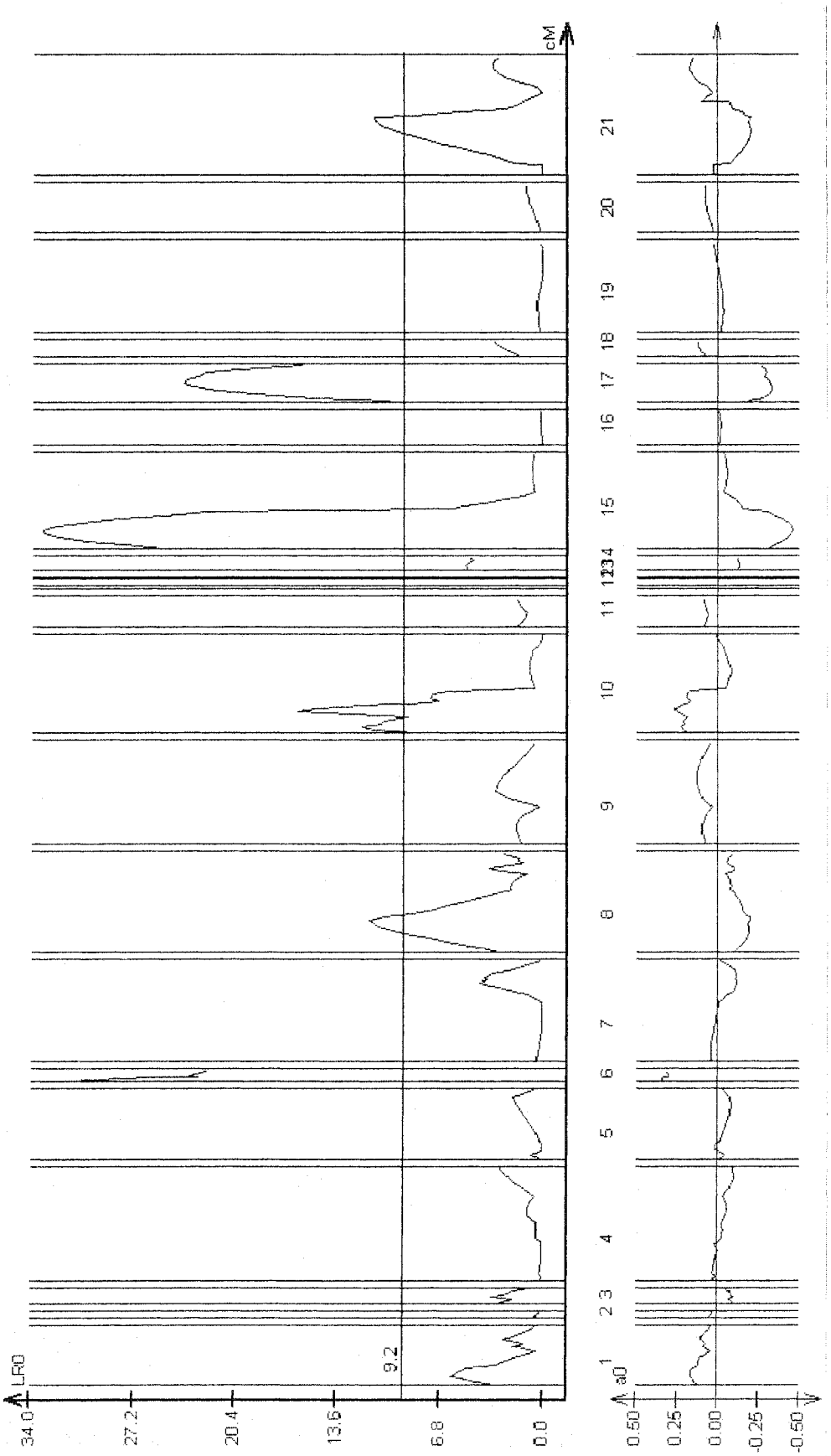
Height at Hopland (ht.hop).



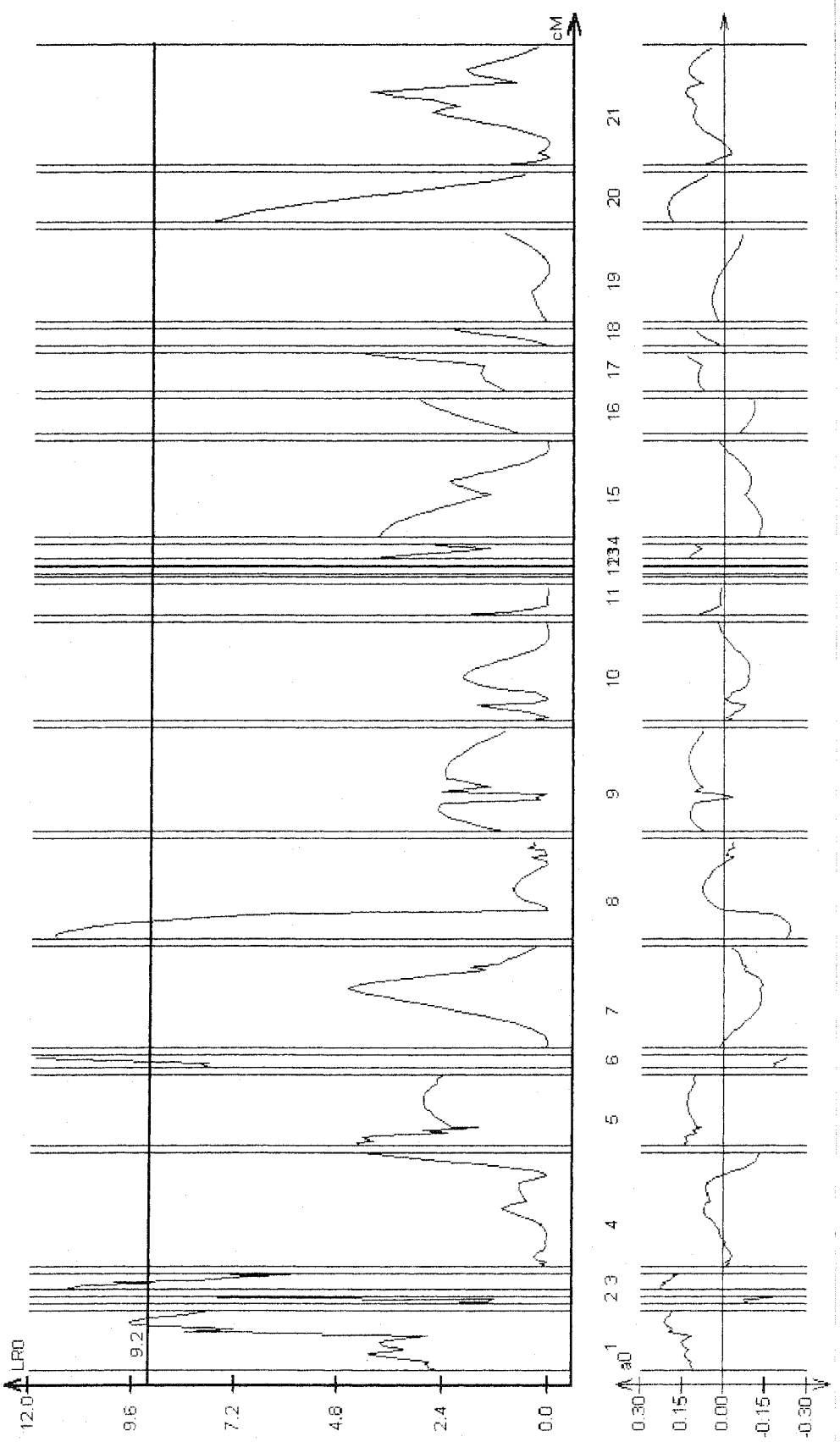
Number of spikelets at Hopland (spike.hop).



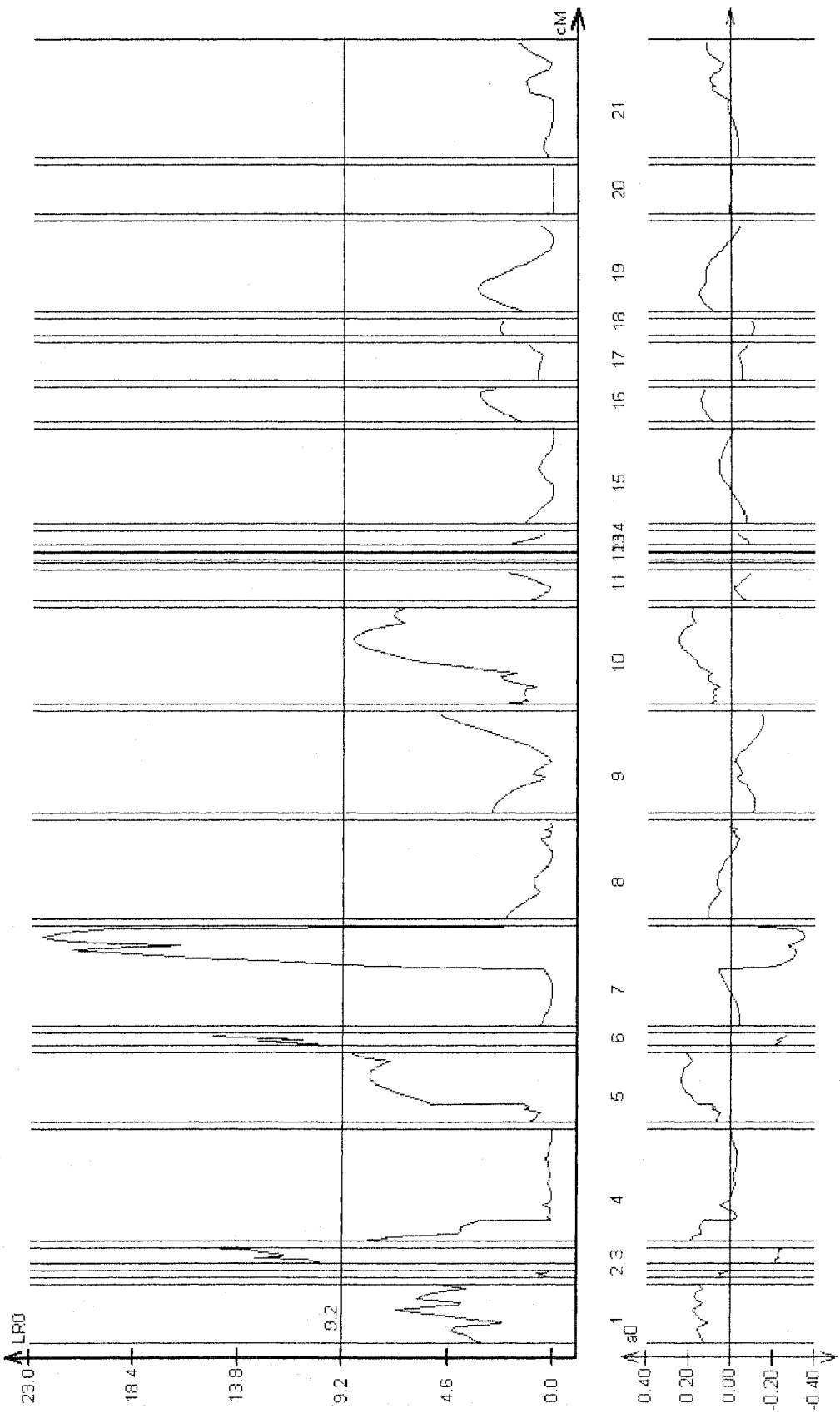
Mass at Hopland (mass.hop).



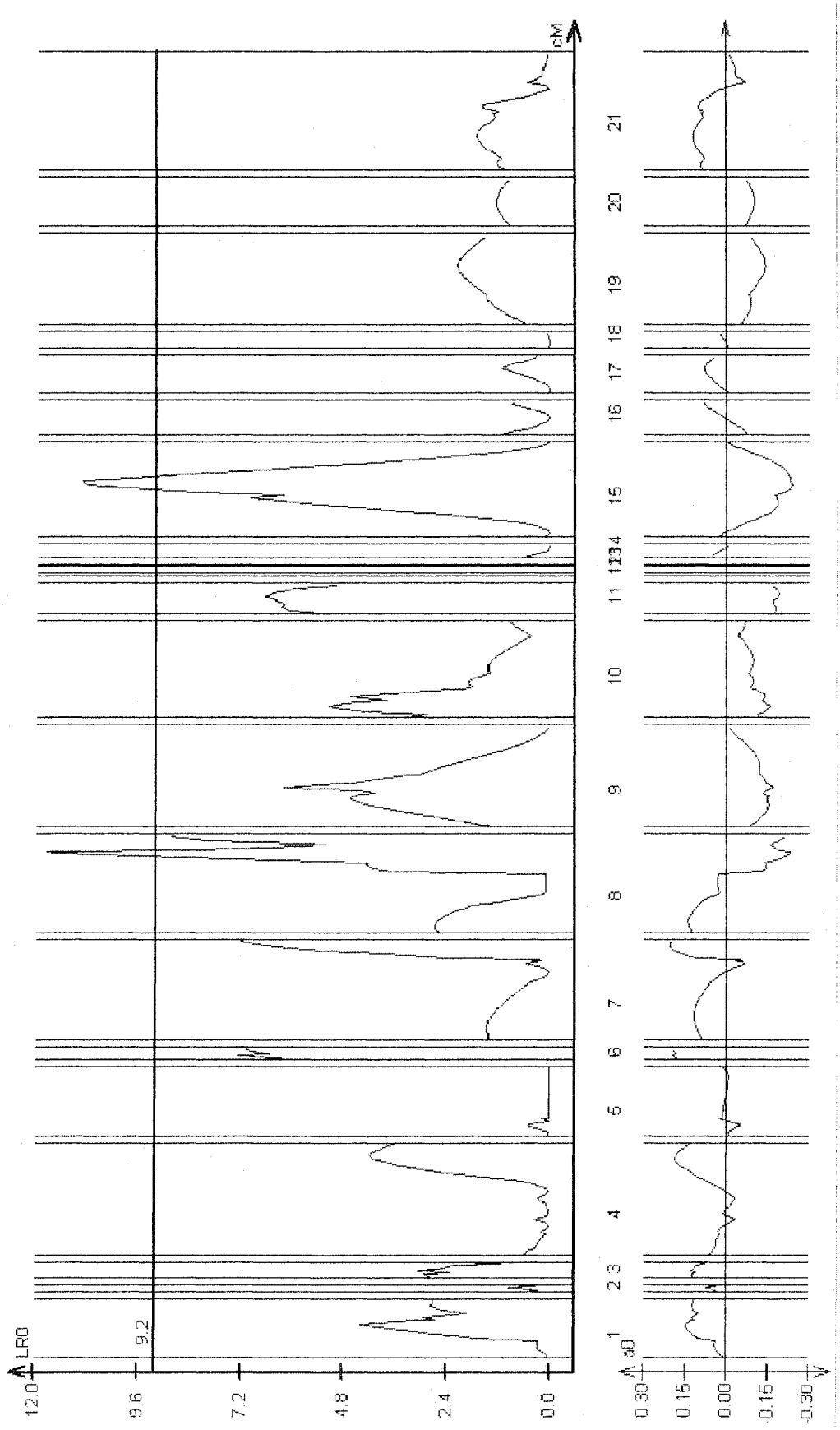
Principal component one (PC1).



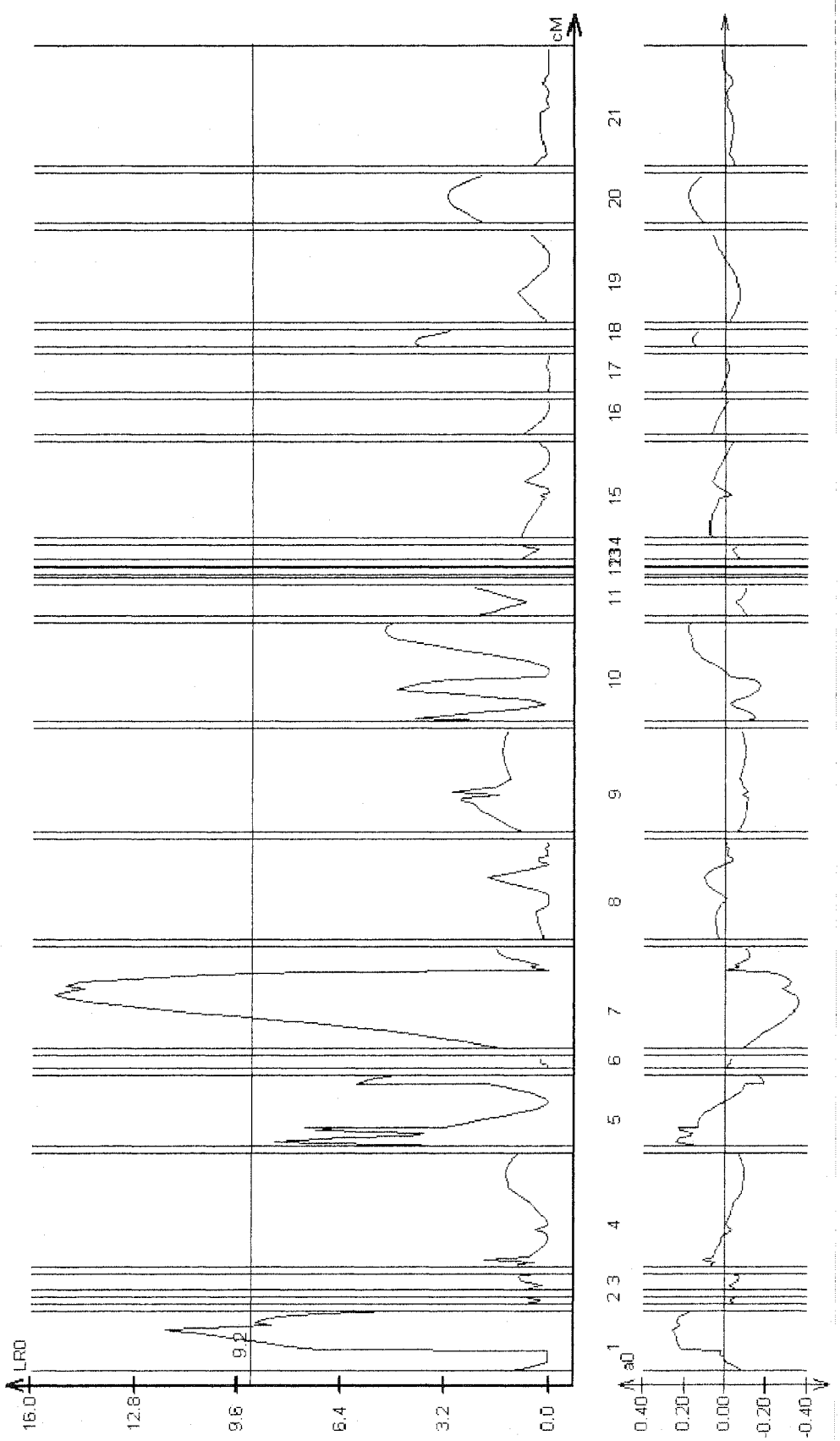
Principal component two (PC2).



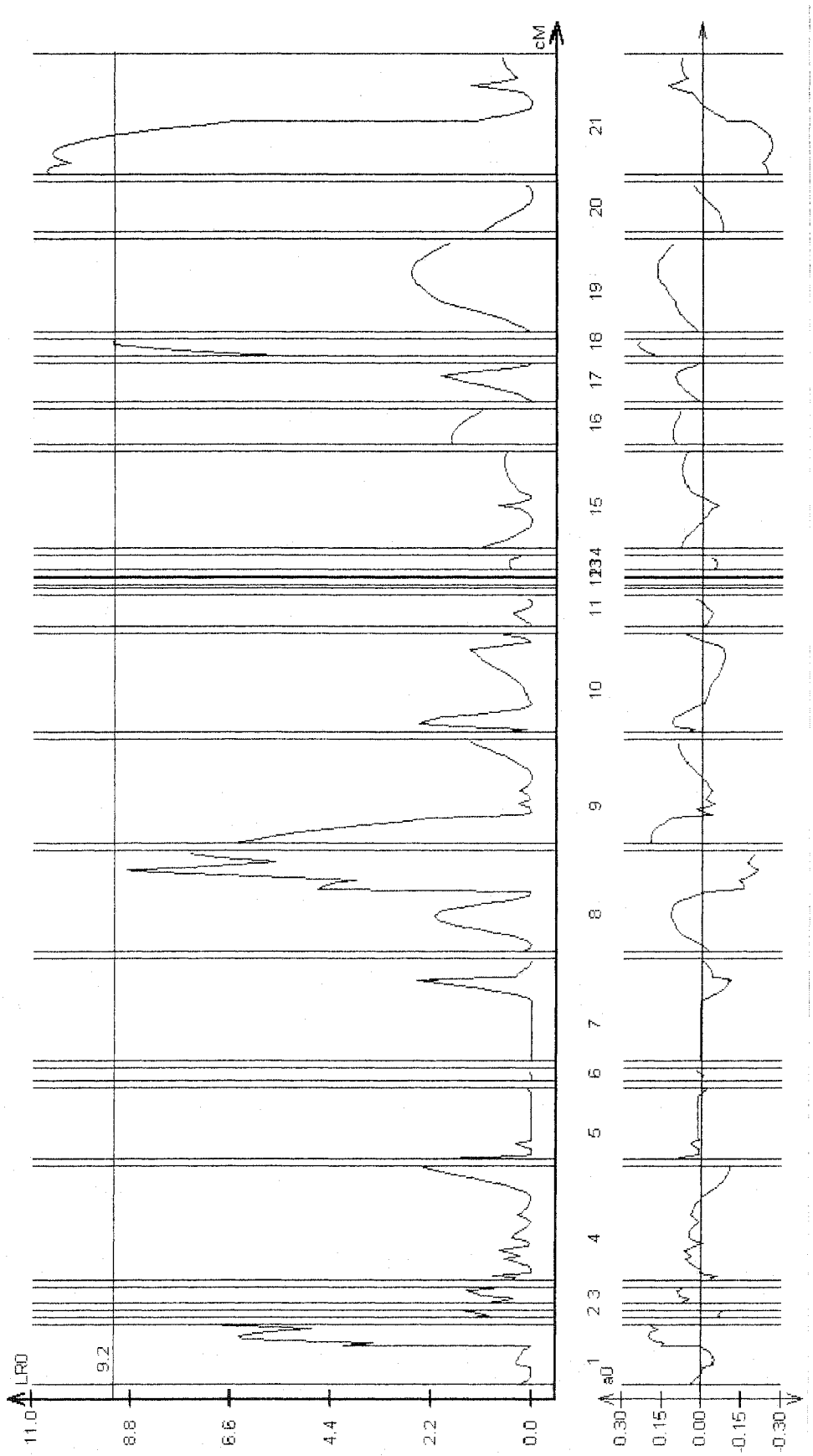
Principal component three (PC3).



Principal component four (PC4).



Principal component five (PC5).



Principal component six (PC6).