EVOLVABILITY IN A VARIABLE WORLD: GENETIC ARCHITECTURE IN *ARABIDOPSIS THALIANA* AND ITS IMPLICATIONS FOR ADAPTATION

by

Tarek Wahid Elnaccash

B.S., State University of New York at Stony Brook, 2000

M.S., University of Nebraska-Lincoln, 2004

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences, Department of Biological Sciences,

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Tarek Wahid Elnaccash

It was defended on

September 9, 2011

and approved by

Dr. John K. Kelly, Associate Professor, Ecology & Evolutionary Biology, KU

Dr. Susan Kalisz, Professor, Biological Sciences

Dr. M. Brian Traw, Assistant Professor, Biological Sciences

Dr. Jeffrey Lawrence, Professor, Biological Sciences

Dissertation Advisor: Dr. Stephen J Tonsor, Associate Professor, Biological Sciences

EVOLVABILITY IN A VARIABLE WORLD: GENETIC ARCHITECTURE IN *ARABIDOPSIS THALIANA* AND ITS IMPLICATIONS FOR ADAPTATION

Tarek Wahid Elnaccash, PhD

University of Pittsburgh, 2011

Evolvability, the ability of a population to adapt to its environment, is critically affected by the genetic architecture of key traits which are further affected by the environmental context. Modern approaches to quantifying genetic architecture at several scales of biological organization allow elucidation of constraints and accelerants to evolutionary change. In chapter one, I describe a novel approach to quantifying genetic architecture that combines recombinant inbred lines (RIL) with line cross analysis. By defining genetic effects relative to an F2 population and incorporating RIL (which are available for many model species), the sampling variance of several nonadditive genetic effect estimates is greatly reduced. The RIL population can be simultaneously used for quantitative trait locus (QTL) identification, thus uncovering the effects of specific loci or genomic regions as elements of genetic architecture. In chapter two, I investigate constraints to evolvability in a set of *Arabidopsis thaliana* RIL populations grown under four levels of nitrogen (N) availability ranging from saturating to stressfully low N-supply rates. I show that changes in N-availability can alter genetic covariances, QTL location and effect magnitude, principle component (PC) structure, and the constraint due to the mismatch between the axes of multivariate genetic variation (particularly PC1 or g-max) and the direction of evolutionary change favored by selection. In chapter three, I show that the G-matrix structures of *Arabidopsis* RIL populations in different N-environments possess patterns of trait associations different enough to alter simulated evolutionary trajectories. I discuss the role of genetic covariances and main-effect QTL in determining the different adaptive trajectories. In chapter four, I report on an extensive QTL mapping study using *A. thaliana* RIL to determine the genetic basis of plastic responses to shifts in the N-environments as well as the role of epistasis in quantitative trait architecture. Exhaustive searches for QTL x QTL interactions at 1cM intervals for 78 trait-environment combinations revealed the presence of several epistatic QTL with no main effect and resolved several seemingly pleiotropic QTL in tightly liked interacting loci. The implications of these patterns of genetic architecture are discussed in the concluding chapter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The work presented here would not have been possible without the help and support of numerous people. I thank my committee members John K. Kelly, Jeffrey Lawrence, Susan Kalisz, and Brian Traw for their helpful advice throughout my graduate career. In particular, Brian's enthusiasm and excitement for biology and Sue's clear thinking and warm encouragement have created a stimulating and enjoyable learning environment for me.

The following people have contributed greatly to my development as a scientist: Tia-Lynn Ashman, Rachel Spigler, Maya Groner, Alison Hale, Sasha Rhode, John Paul, April Randle, Tom Pendergast, Josh Auld, and Christopher Haeckel. Tia-Lynn has listened to my research ideas for several years and has always given me insightful and practical advice on experimental design. I admire her ability to see into the complexities of biological theory and pull out the right questions to ask. My ability to design experiments and develop testable hypotheses has improved thanks to her influence. Rachel has often given me thoughtful research advice. I value her equally for her intellectual rigor and her up-beat and light-hearted personality; she's a great person to work with. Maya has been a helpful colleague and a great friend for several years now; her opinions on personal and academic matters mean a lot to me. John and April helped me early on in my graduate career, both through research discussions and by being exemplary graduate students for me to emulate. I also thank other members of Pitt's Ecology and Evolution Program for fun, friendship, and an intellectually stimulating environment: Henry Schumacher, Ji Hao, Nathan Brouwer, Heather Schaffery, Jessica Hua, Will Brogan, and Aaron Stoler.

Anthony Bledsoe and Melanie Popa deserve thanks for being excellent teachers. When I worked as a teaching assistant for Tony, I took almost as many notes on how to be a good teacher as on his actual lectures. His passion for evolutionary biology and ornithology comes across in every class. Melanie is a wonderfully open-minded teacher and a good friend. She makes improvements to her excellent microbiology lab each year and she has always been open to new ideas and discussion on how to improve undergraduate education. I hope to be as good a teacher as these two in the future.

I have gained so much from weekly discussions in the Tonsor lab with its past and present members: Alicia Montesinos, Marnin Wolfe, John Paul, Matt Simon, Timothy Helbig, Ellen York, and Natalie Settles. Alicia often pulled me away from the lab computer, reminding me to appreciate biology

first hand. Ellen patiently helped me learn how to grow *Arabidopsis* for my first experiments. Marnin helped me learn SAS macro-programming and spent many hours with me discussing genetic architecture and evolvability. It has been a pleasure for me to learn from and learn with people who share my excitement for evolutionary biology.

I owe thanks to friends and family. Laura Bilski took over many of my responsibilities while I was writing this dissertation. She kept me well fed, my apartment organized, and generally kept my cats Kuma and Garbanzo from eating drafts of my dissertation. My brother Ihab and my sister Yasmin have both been enormously supportive throughout my years in graduate school. Ihab and Mina always remind me that there is a funny side to everything and not to take anything too seriously. They make me happy every time I see them and I plan to see them both more often in the future.

Most importantly, I thank my advisor, collaborator, and friend Stephen Tonsor. It's impossible to do him justice in these acknowledgments. Steve is one of the sharpest and most generous people I have met. He created an ideal learning environment for me in his lab: challenging, stimulating, critical when necessary, and always supportive. I am excited about continuing our collaboration and our friendship.

Financial support

I received financial support from the University of Pittsburgh first year fellowship (2004), the K. Leroy Irvis Fellowship (2005) and The National Science Foundation Grade K-12 Teaching Fellowship (2006). The work described in Chapters 2-4 was made possible by financial support to Stephen Tonsor from the National Science Foundation.

1.0 INTRODUCTION

'Evolvability' means the propensity to produce high fitness variants; it is the ability for a population or species to adapt to its environment. Until recently, evolvability was quantified as the proportion of the trait variation present in a population that is genetically based, and thus can be passed from parents to their progeny. This description of variation, more commonly known as heritability (abbreviated h^2), has been widely used in animal and plant breeding as well as in evolutionary ecology and evolutionary genetics research. Heritability is a component of one of the most widely used models of short-term evolutionary change known as the breeder's equation. In this context, heritability is used as a weight or a conversion factor to describe how a change in a trait mean within a generation (due to selection) can be converted into a change in trait mean across generations (the evolutionary response). Heritability has been criticized as an insufficient measure of evolvability (Hansen et al 2011). It presents a 'snapshot' of the standing genetic variation in a population. Theoretical studies have suggested that the amount of standing genetic variation in a population can be explained by a balance between the input of variation due to deleterious mutation and the reduction in variation due to natural selection (Barton and Keightley 2002). If this is an accurate explanation for standing variation, then heritability may tell us more about short-term maladaptation than about the potential for evolutionary improvement. Alternatively, evolvability can be better understood with more comprehensive measures of genetic architecture including estimates of how genotype maps to phenotype. Genetic architecture can be measured with statistical abstractions such as genetic variances of traits as well as the covariances between traits, and also includes more mechanistic estimates such as of the number of loci that affect a trait, their effect magnitudes, their multi-trait effects (known as pleiotropy), and interactions within and between loci (dominance and epistasis, respectively). Many of these aspects of genetic architecture have been shown to change across environments, thus measures environmental effects on phenotype (i.e phenotypic plasticity) and genotype x environment interactions are import components of genetic architecture. As all models of adaptation, population divergence and speciation assume a particular genetic architecture of traits under study, yet little is known about the validity of these assumptions, detailed empirical estimates of genetic architecture are invaluable in assessing and improving predictive models of evolutionary change.

My approach to measuring and understanding evolvability is to use quantitative genetics (Lynch and Walsh 1998), the study of heredity at the level of the phenotype. In my view, quantitative genetic approaches can allow researchers to balance mechanistic and higher level analyses of evolvability. On the one hand, mechanistic understandings of genotype-phenotype maps can be attained through approaches like quantitative trait locus (QTL) mapping. On the other, measuring genetic variances and covariances can allow predictive models of evolutionary change when the number of relevant genes, genotypes, or environments are too numerous to analyze in a more reductionist framework. Using complementary analyses at different levels of biological organization (QTL, traits, multi-trait phenotypes, and populations) will present a clearer picture of the potential to evolve than any single approach in isolation.

The empirical work reported here was accomplished using *Arabidopsis thaliana*, a model plant species ideal for investigations of genetic architecture and evolvability (Tonsor et al 2005). A wealth of genetic tools are available for use with *A. thaliana*, the most notable for my research is the availability of recombinant inbred lines (RIL). RIL are genotypes produced by the crossing of two divergent parental lines, followed by inbreeding to produce lineages that are homozygous at nearly every locus but vary between loci in the parent of origin (figure 1.1). Being homozygous at each locus means each genotype breeds true, producing (essentially) genetically identical offspring. This allows replicated studies of genotypes across different environments and greatly facilitates QTL mapping (figure 1.2). Figures 1.1 and 1.2 outline the creation of RIL and their use in QTL mapping. *A. thaliana* RIL are used for all analyses in chapters 2-4 of this thesis.

In chapter one, I describe an approach that incorporates RIL (whether *A. thaliana* or other model species) in the context of line cross analysis. By producing crosses between phenotypically divergent parents and comparing the means of various progeny (e.g. F1, F2, etc.) one can learn about the genetic basis of parental differences. Incorporating RIL in these breeding designs can dramatically increase the power to detect dominance and epistasis with a relatively small increase in research effort. This work was published in PlosOne and coauthored with my advisor Stephen Tonsor (Elnaccash and Tonsor 2010).

In chapter two, I investigate constraints to evolvability and how the environmental context can alter these constraints. I use univariate (e.g. heritabilities), bivariate (e.g. genetic correlations), and multivariate (e.g principle component analyses) quantitative genetic approaches to quantify genetic architecture and how it changes across a nitrogen supply gradient. I then use QTL mapping and multivariate measures of constraint based on the vectors of selection and genetic variation to understand how constraints to evolvability can change with shifts in the environment.

In chapter three, I ask whether the differences in evolvability detected across nitrogen environments translate into differences in evolutionary trajectories in response to simulated selection

pressures. I do this by simulating random natural selection on *A. thaliana* RIL grown in each nitrogen environment and comparing each population's predicted direction of evolutionary response. I further ask whether these trajectory differences are due to the constraining effects of genetic covariances or to the pattern of QTL expression that is specific to each environment.

In chapter four, I report on an extensive QTL mapping study to elucidate the genetic basis of plastic responses to changes in nitrogen environments as well as the importance of epistasis to quantitative trait variation. This QTL mapping study is novel for several reasons. It compares the genetic basis of phenotypic plasticity for a diverse set of traits in their sensitivity to changes at limiting or nearly saturating nitrogen levels. It also tests for epistatic QTL X QTL interactions (at 1 cM resolution) for traits in four nitrogen environments as well as for phenotypic plasticity to changes at limiting and saturating nitrogen levels. The single QTL analysis of plasticity includes analysis of 26 traits. The analysis of epistatic QTL X QTL interactions includes 78 traits. This number of traits is uncommon except in meta-analyses, and meta-analyses rarely can address issues of pleiotrypy, plasticity, and epistasis without serious caveats due to differences in study locations.

I use the pronoun "we" throughout all of my chapters because all my work (and understanding of evolution) is a collaborative effort with my advisor Dr. Stephen Tonsor. Chapter 1 was already published with Dr. Tonsor as a coauthor; all other chapters will be as well. For our first publication, my contribution to this paper (chapter 1) was the analytic results and the writing of the paper. Stephen Tonsor contributed editing and invaluable advising and constructive discussion on this topic as well as all others in this thesis. For the work involved in chapters 2-4, planting and measuring *Arabidopsis* took place in the Tonsor lab before I began as a graduate student. The conceptual framework, data analysis, statistical programming, and writing of these chapters was my own work.

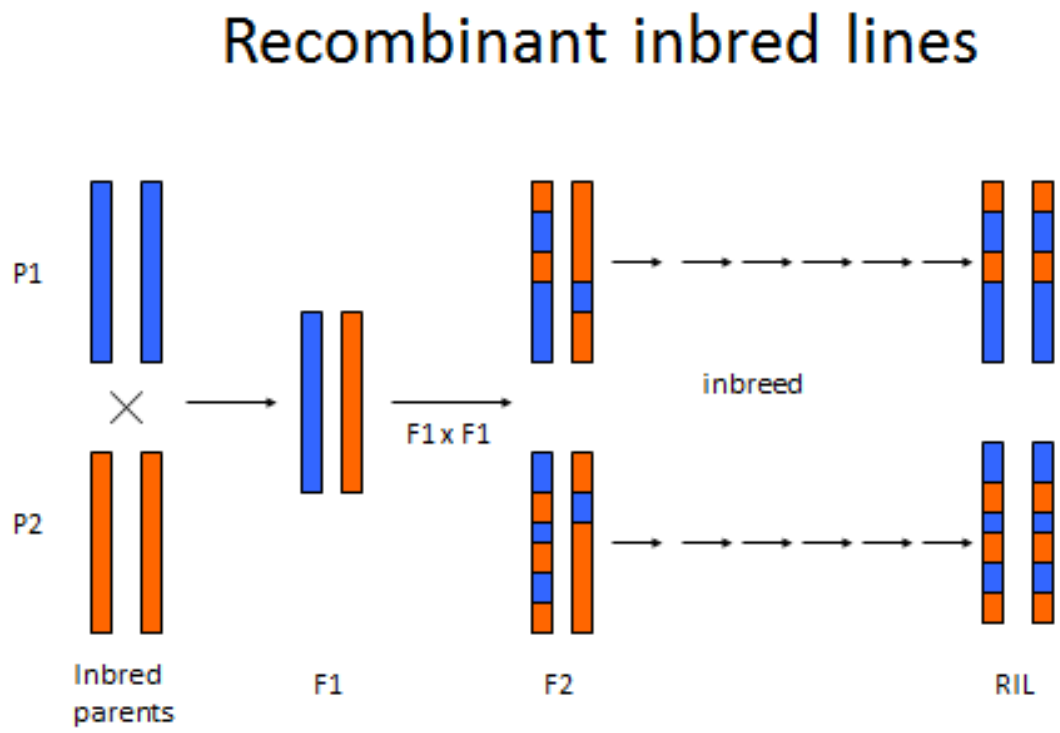


Figure 1.1. Recombinant inbred lines. RIL are created by first crossing two inbred parental lines (P1 and P2) that have diverged in the trait of interest (e.g. biomass). Solid red and blue chromosome diagrams indicate parents are homozygous for each locus. The P1 and P2 cross produces F1 which are heterozygous for each locus. Crosses between F1 produce a genetically variable F2 population which, after several generations of self-fertilization (inbreeding) results in a set of genotypes homozygous for each locus. Shown here are two resulting genotypes which possess varying amounts of parental genomes (red and blue segments). Note that the two RIL differ from each other and are entirely homozygous. Typically this is done to produce hundreds of RIL per parental cross, which can then be used in quantitative trait locus (QTL) mapping.

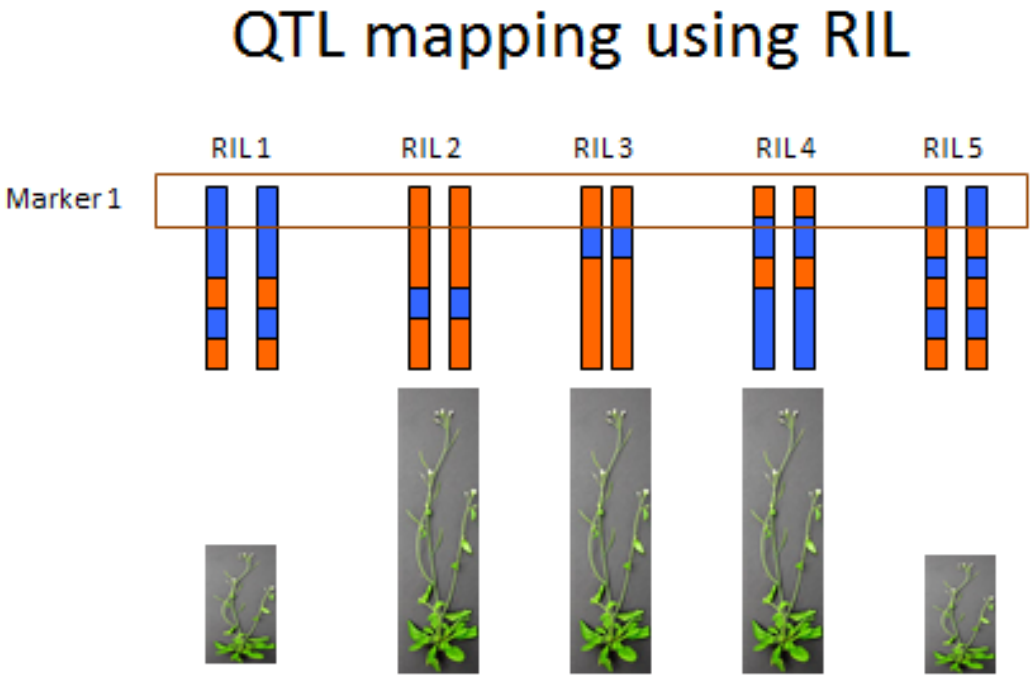


Figure 1.2. QTL mapping using RIL. QTL mapping in its simplest form begins by using molecular markers (e.g. Marker 1) to identify the parental allele present in each RIL at a particular genomic position (orange box). Plant traits (e.g. biomass) are measured and variation in phenotype is compared to variation in the marker genotype. In this example, the orange allele is associated with high biomass and the blue allele is associated with low biomass. From these data, the investigator would conclude that there is a QTL for biomass on this particular chromosome near Marker 1. Such tests are repeated using many markers located throughout the genome.

2.0 SOMETHING OLD AND SOMETHING NEW: WEDDING RECOMBINANT INBRED LINES WITH TRADITIONAL LINE CROSS ANALYSIS INCREASES POWER TO DESCRIBE GENE INTERACTIONS

2.1 ABSTRACT

In this paper we present a novel approach to quantifying genetic architecture that combines recombinant inbred lines (RIL) with line cross analysis (LCA). LCA is a method of quantifying directional genetic effects (i.e. summed effects of all loci) that differentiate two parental lines. Directional genetic effects are thought to be critical components of genetic architecture for the long term response to selection and as a cause of inbreeding depression. LCA typically begins with two inbred parental lines that are crossed to produce several generations such as F1, F2, and backcrosses to each parent. When a RIL population (founded from the same P1 and P2 as was used to found the line cross population) is added to the LCA, the sampling variance of several nonadditive genetic effect estimates is greatly reduced. Specifically, estimates of directional dominance, additive x additive, and dominance x dominance epistatic effects are reduced by 92%, 94%, and 56% respectively. The RIL population can be simultaneously used for QTL identification, thus uncovering the effects of specific loci or genomic regions as elements of genetic architecture. LCA and QTL mapping with RIL provide two qualitatively different measures of genetic architecture with the potential to overcome weaknesses of each approach alone. This approach provides cross-validation of the estimates of additive and additive x additive effects, much smaller confidence intervals on dominance, additive x additive and dominance x dominance estimates, qualitatively different measures of genetic architecture, and the potential when used together to balance the weaknesses of LCA or RIL QTL analyses when used alone.

2.2 INTRODUCTION

Understanding the genetic basis of complex phenotypes, i.e. genetic architecture, is of fundamental importance both for modeling evolutionary change and for genetic manipulation of crop plants. Genetic architecture is a broad term for all factors that influence the determination of phenotype from genotype. It includes all genetic effects on traits: the number of genes, allelic effects, epistasis, pleiotropy, and genotype x environment interactions (Zeng et al. 1999). Knowledge of genetic architecture can inform us about the propensity to evolve (i.e. 'variability' *sensu* Hansen (2006)) on all timescales.

Studies of genetic architecture have revealed that epistasis, i.e. interactions between loci, is a common component of most quantitative traits. For example, biomedical studies have shown an epistatic genetic basis for many human diseases (Moore 2003) including diabetes (Cox et al 1999, Wiltshire et al. 2006), Alzheimer's disease (Combarros et al. 2009), obesity (Ankra-Badu 2009), cardiovascular disease (Lim et al. 2005) and schizophrenia (Qin et al. 2005). Knowledge of the genetic basis of these diseases is important because epistatic traits can evolve in a fundamentally different way than additive traits (Wade 2000, Carlborg et al. 2006, Weinreich et al. 2006, Hallander and Waldmann 2007). Knowledge of gene interactions and genetic architecture is also important for building and evaluating models of evolutionary processes. All models of adaptation, population divergence and speciation assume a particular genetic architecture, but the assumptions vary wildly among models. At two ends of a spectrum, selection analyses used commonly in evolutionary ecology studies implicitly assume an additive genetic architecture (Lande 1980, LAnde and Arnold 1983), while most studies of speciation assume an epistatic genetic architecture (Dobzhansky 1937, Gavrilets 1997, Porter and Johnson 2002). Does trait architecture change from additive to epistatic over some range of genetic distances or geographic distances? While patterns of the genetic architecture of inter-specific differentiation are becoming clear (e.g. Haldane's rule, Dobzhansky-Muller incompatibilities (Coyne and Orr 1997, Presgraves 2002, Kondrashov et al. 2002)), the genetic basis of differences within and among populations is more poorly understood and the genetic architecture in particular instances does not appear to correlate with factors such as genetic, geographic, or even phenotypic differences among populations (Fenster and Galloway 2000, Edmands 2002, Erickson and Fenster 2006, Demuth and Wade 2007a, 2007b). We know very little about the genetic architecture of quantitative traits. This limits theoretical and practical advances in evolutionary genetics and plant breeding.

Line cross analysis (LCA) is a well-established method of quantifying genetic architecture with a long history of use in agriculture. Because of its utility for gene discovery, much recent work has focused on understanding genetic architecture at the level of individual loci or QTL. LCA in contrast measures the

summed, i.e. *directional*, effects of all loci contributing to a trait. Line crosses have become more popular in recent years as interest in quantifying epistasis in quantitative traits has increased; this method offers far greater statistical power than variance component analyses previously used to measure epistasis (Demuth and Wade 2006). Traditionally the nearly exclusive realm of plant and animal breeders, LCA have also been used recently to address more broadly evolutionary issues with genetic architecture (Roff and emerson 2006), and are likely to continue to become more common in evolutionary research for several reasons. Demuth and Wade (2005), refined by Fitzpatrick (2008), have shown how line crosses between populations can be used to study speciation and Haldane's rule. Directional dominance effects are a requirement for inbreeding depression (Lynch and Walsh 1998, p. 257). Hansen and colleagues (Wagner and Altenberg 1996, Carter et al. 2005, Hansen 2006) have shown that the directional epistasis revealed by line cross analysis may be a key to understanding continued response to long term selection, and empirical work in corn and chicken is consistent with this theoretical prediction (Carlborg and Haley 2004, Carlborg et al. 2006, Dudley 2007). The selection responses in corn oil concentration and chicken body weight are also consistent with a large number of loci each with several alleles of small effect (Barton and Keightly 2002) and with a large input of new variation from mutation (Dudley 2008). Clearly we need empirical measures of both locus-specific as well as directional genetic architecture estimates (particularly positive directional epistasis, (Hansen 2006) to determine the relative roles of these hypothesized factors.

In this paper we present a novel approach to quantifying genetic architecture that combines recombinant inbred lines (RIL) with line cross analysis. When RIL are used in line cross analyses, scaling tests can be constructed for non-additive genetic effects with far more precision than traditional methods of estimation. The RIL can be simultaneously used for QTL identification. These two uses of a RIL population yield qualitatively different information about genetic architecture and can be used in a powerful and complementary manner.

2.3 METHODS

Line cross analyses typically begin with two inbred parental lines that are crossed to produce an F1 generation. F2s, backcrosses, and other generations can be produced as well; the directional genetic effects (also called 'composite genetic effects') to be estimated are limited by the number of generation means measured. For example, estimating the mean, additive, dominance, and 3 pairwise epistatic effects requires at least 6 generation means for estimation and 7 for hypothesis testing.

Line cross analyses are primarily carried out using frameworks based on the F2 model of Cockerham (1954) or on the F-infinity model of Hayman and Mather (1955). Here we follow line cross theory based on the F2 model as described by Lynch (1991) and Lynch and Walsh (1998). We refer to it as the F2 model for simplicity. In this model, the F2 is the reference generation relative to which all genetic effects are derived by linear contrasts. Line crosses use linear combinations of generation phenotypic means to estimate composite genetic effects and carry out significance tests. Each generation mean can be written as a function of two coefficients, the source index (θ_S) and the hybridity index (θ_H), multiplied by the additive (A), dominance (D) or epistatic interaction effects (AA, AD, DD, etc.) that potentially differentiate the parental lines (equation 1).

$$\text{Generation mean} = \mu. + \theta_S A + \theta_H D + \theta_S^2 AA + \theta_S \theta_H AD + \theta_H^2 DD... \quad (1)$$

$\mu.$ = the mean of the F2 generation. The source index θ_S determines the coefficients of the additive effects' contribution to each generation's phenotypic mean. The source index is scaled from one to negative one and indicates the proportion of genes in the generation that came from parent one (P1), with +1 indicating 100% and -1 indicating 0%. P1's $\theta_S = +1$ while for F1s, F2s, and RILs $\theta_S = 0$.

The hybridity index determines the contribution of the dominance effects to each generation mean. The hybridity index is also scaled from +1 to -1, with +1 indicating that every locus is heterozygous and -1 indicating that every locus is homozygous. F1s thus have $\theta_H = +1$, while parents have $\theta_H = -1$. Figure 2.1 shows the source and hybridity indices for the P1, P2, F1, F2, B1 (back-cross to P1), B2 (back-cross to P2).

To this traditional set of line cross generation means, the mean of a RIL generation can be added. In this context, 'RILs' or a 'RIL population' is a set of genotypes of highly inbred F2 lines. If these genotypes were replicated, the means of each genotype can be used as individuals for calculating the overall RIL generation mean. RILs asymptotically approach complete homozygosity for all loci as the number of generations of inbreeding approaches infinity. In practice, the convention is to use six to eight generations of inbreeding, resulting in ~99.84 to 99.96% homozygosity respectively. A major advantage of RILs is that the descendents of any one RIL are genetically identical, hence "immortal" (ignoring mutation accumulation), allowing RILs to be marker-genotyped once and phenotyped repeatedly in multiple labs and experiments. In the framework of LCA, RIL can be used to greatly improve power in estimating non-additive genetic effects.

The F2 generation has a value of zero on both the source and hybridity indices. All genetic effects are scaled relative to this F2 generation mean, thus the linear contrasts used to estimate the genetic effects are sometimes called F2 scaling tests. The expected mean of the F2 and RIL generations are

identical and their source indices are both zero (the actual source index for RIL can be approximated from marker data as 2*(number of P1 marker alleles among all lines/total number of alleles)-1, assuming equal spacing of markers throughout the genome. In the absence of segregation distortion, this will be very close to zero). However the F2 hybridity index has zero value, while the RIL hybridity index is in contrast approximately negative one. (Figure 2.1 and Table 2.1).

Products of the source and hybridity indices determine the coefficients for interactions between additive and dominance effects (i.e. epistasis). For example, the product of the additive coefficient (source index) and the dominance coefficient (hybridity index) is the coefficient for the additive x dominance epistatic effect. The coefficients for additive, dominance and pairwise epistatic effects for 7 commonly used generation means are given in table 2.1. The RIL generation mean can used in estimating non-additive effects, since in contrast to the F2 it has a non-zero hybridity coefficient.

Using equation (1) and the first six generations in Table 2.1, Lynch and Walsh (1998: Table 9.3, p. 214) produced equations to estimate the following composite genetic effects:

$$\begin{aligned}
\mu. &= z_{F2} \\
A &= z_{B1} - z_{B2} \\
D &= -z_{P1}/4 - z_{P2}/4 + z_{F1}/2 - 2z_{F2} + z_{B1} + z_{B2} \\
AA &= -4z_{F2} + 2z_{B1} + 2z_{B2} \\
AD &= -z_{P1}/2 + z_{P2}/2 + z_{B1} - z_{B2} \\
DD &= z_{P1}/4 + z_{P2}/4 + z_{F1}/2 + z_{F2} - z_{B1} - z_{B2}
\end{aligned} \qquad (2)$$

z_{Xi} indicates the phenotypic mean of the X_i^{th} generation (eg. X = B, i = 1 for the B1 generation). Note the equations for D and AD in Lynch & Walsh (1998 Table 9.3) estimate –D and –AD so we have included the corrected equations here.

By incorporating the RIL generation's equation (for the contributions of the various genetic effects to the RIL generation mean) in the F2 scaling tests, we can construct tests for non-additive genetic effects with fewer terms than traditional tests, shown by contrasting equations (2) and (3). Incorporating the RIL means equalizes the number of generation means necessary to estimate the additive and dominance effects, and the number of generation means necessary to estimate the AA, and DD epistatic effects. This is important in providing equanimity in the power of tests for both intra- and interlocus additive vs. dominance effects; estimates of A and D both require two generation means while AA and DD both require three generation means. AD is the sole equation which retains four generation means in its estimator because the RIL mean cannot be used to simplify the equation.

$$
\begin{aligned}
\mu. &= z_{F2} \\
A &= z_{B1} - z_{B2} \\
D &= (z_{F1} - z_{RIL})/2 \\
AA &= (z_{P1} + z_{P2})/2 - z_{RIL} \\
AD &= -z_{P1}/2 + z_{P2}/2 + z_{B1} - z_{B2} \\
DD &= (z_{F1} + z_{RIL})/2 - z_{F2}
\end{aligned}
\tag{3}
$$

T-tests can be used to test the null hypothesis that a genetic effect equals zero, assuming that the test statistic is normally distributed under the null hypothesis. The test statistic is simply the estimated genetic effect estimate divided by the standard error of the estimate. For example, the test statistic for the composite dominance effect (using eq. (2)) is

$$
\Delta_D = \frac{\left|\hat{D}\right|}{\sqrt{\mathrm{Var}\,(\hat{D})}} \tag{4}
$$

$$
= \frac{\left| -\frac{z_{P1}}{4} - \frac{z_{P2}}{4} + \frac{z_{F1}}{2} - 2z_{F2} + z_{B1} + z_{B2} \right|}{\sqrt{\frac{\mathrm{Var}[z_{P1}]}{16} + \frac{\mathrm{Var}[z_{P2}]}{16} + \frac{\mathrm{Var}[z_{F1}]}{4} + 4\mathrm{Var}[z_{F2}] + \mathrm{Var}[z_{B1}] + \mathrm{Var}[z_{B2}]}}
$$

Δ_D follows a t distribution with 1 df. Similar test statistics can be constructed for each genetic effect following the same format.

The effect of reducing the number of terms becomes clear when we look at the new RIL-based test statistic for the composite dominance effect:

$$
\Delta_D = \frac{\left|\hat{D}\right|}{\sqrt{\mathrm{Var}(\hat{D})}} = \frac{\left|(z_{F1} - z_{RIL})/2\right|}{\sqrt{\frac{\mathrm{Var}[z_{F1}]}{4} + \frac{\mathrm{Var}[z_{RIL}]}{4}}} \tag{5}
$$

Recall that the variance of a sum equals the sum of the variances multiplied by the square of the coefficients, i.e. Var (cA + dB) = c^2 * Var (A) + d^2 * Var (B), provided that the terms being summed are independent. We can compare the variances associated with the traditional formulae for D, AA, and DD

from equations (2) with the corresponding RIL equations (3). For these comparisons, we assume that all generation means have equal variance (i.e. σ^2 = Var (P1) = Var (P2) = Var (F1) = Var (F2) = Var (B1) = Var (B2) =Var (RIL)).

Based on this assumption, the RIL-based equation for D, AA, and DD have 92%, 94%, and 56% reductions in variance respectively relative to the traditional equations (Table 2.2). The variance reductions occur for two reasons. First, when fewer generation means are summed to estimate a genetic effect, fewer sources of error are summed into this estimate as well. Second, the RIL equations have smaller coefficients for each generation mean than traditional equations. Since these coefficients are squared when summing the variances, lower coefficients can drastically reduce the variances of the genetic effects. Further variance reductions can occur in RIL based estimates due to the sample size of RIL. Since the number of lines composing the RIL generation is typically large since this determines the power of QTL mapping with RIL populations, the genetic effect variance reduction from using RIL equations is even greater than the reductions using equal variances for all generations illustrated in Table 2.2. These reductions in variance produce a substantial increase in power to detect dominance and epistasis and to compare dominance-influenced vs. additive effects.

Frequently, line cross experiments are analyzed using joint scaling tests (e.g. Kelly 2005, Sun et al. 2006, Wegner et al. 2008). The joint scaling test is a weighted least squares regression technique for estimation and significance testing of various models of genetic architecture. A description of this method can be found in Lynch and Walsh (1998, p.215-219, see also: Bradshaw and Holzapfel 2000, Demuth and Wade 2005). Briefly, one starts with a vector of generation means (**Y**), a design matrix (**X**) of coefficients derived from the source and hybridity indices, and a vector of composite genetic effects (**β**) to be estimated. Initially, **β** contains the mean and the composite additive effect and **X** contains two corresponding columns. An estimate of **β** is calculated using $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ (or $(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$, where $\mathbf{V}^{-1}$ is a diagonal matrix of squared standard errors for generation means if sample sizes are unequal). This estimate of **β** is premultiplied by **X** to produce a vector of predicted generation means, $\hat{\mathbf{Y}}$, given an additive genetic architecture. $\hat{\mathbf{Y}}$ is then compared with the observed **Y** using a chi-squared test. If the observed and predicted **Y**'s are significantly different, then the additive model is rejected and an additive and dominance model is tested next. A new **β** vector containing the mean, the composite additive effect, and the composite dominance effect is estimated and multiplied by an **X** matrix with 3 columns to produce a new $\hat{\mathbf{Y}}$. Increasingly complex models of genetic architecture are tested until the predicted and observed vector of generation means is not significantly different.

To illustrate the advantages of using RIL in a joint scaling context, we used seven generation means to estimate a model of additive, dominance and pairwise epistatic effects:

$\mathbf{Y} = \mathbf{X}\,\boldsymbol{\beta}$, where

$$\begin{bmatrix} P1 \\ P2 \\ F1 \\ F2 \\ B1 \\ B2 \\ RIL \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ 1 & -\frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \mu \\ A \\ D \\ AA \\ AD \\ DD \end{bmatrix} \qquad (6)$$

$\mathbf{Y}$ $\qquad\qquad$ $\mathbf{X}$ $\qquad\qquad$ $\boldsymbol{\beta}$

The general formula for solving linear equations is $\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. When we used Mathematica [46] to solve for $\boldsymbol{\beta}$ in terms of the generation means, the solution is:

$$\begin{aligned} \mu. &= (16z_{B1} + 16z_{B2} + 19z_{F2} - 4z_{P1} - 4z_{P2} + 8z_{RIL})/51 \\ A &= z_{B1} - z_{B2} \\ D &= (4z_{B1} + 4z_{B2} + 51z_{F1} - 8z_{F2} - z_{P1} - z_{P2} - 49z_{RIL})/102 \\ AA &= 2(z_{B1} + z_{B2} - 2z_{F2} + 4z_{P1} + 4z_{P2} - 8z_{RIL})/17 \\ AD &= (2z_{B1} - 2z_{B2} - z_{P1} + z_{P2})/2 \\ DD &= (-12z_{B1} - 12z_{B2} + 17z_{F1} - 10z_{F2} + 3z_{P1} + 3z_{P2} + 11z_{RIL})/34 \end{aligned} \qquad (7)$$

As in the individual scaling tests, the variance of the dominance effect and the additive x additive effect in RIL models are reduced by 92% and 94% respectively relative to the traditional equations. When the genetic effects are estimated simultaneously using RILs in the model above, the variance of DD is now reduced by 79% (c.f. 56% in individual scaling tests) and the variance of the estimate of the mean is reduced by 63%.

More precise estimation of non-additive genetic effects will help distinguish whether these non-additive effects are rarely detected within micro-evolutionary studies because they are uncommon or because experimental designs have lacked sufficient statistical power to detect them.

2.4 RESULTS AND DISCUSSION

Phylogenetically broad crosses have gained increasing importance in both plant breeding and evolutionary genetic studies (e.g. Schemske and Bradshaw 1999, Reiseberg et al. 1999, Lopez-Fernendez and Blonick 2007). Directional gene interactions appear to be increasingly important as the genetic distance between lineages increases. However, even when genetic distances between crossed lines are small, the extent of epistatic interaction can be surprisingly large (Kelly 2005, Demuth and Wade 2007a). Line cross analysis is consequently receiving increased attention as a method for detecting directional gene interactive effects.

We show in this paper that the inclusion of a RIL generation in line cross analysis can greatly increase the accuracy with which D, AA, and DD interactions are estimated. The accurate estimation of gene interaction effects can be of substantial value for those interested in describing genetic architecture and its role in a variety of evolutionary processes (Whitlock et al. 1995).

A reviewer has pointed out that one research group has previously incorporated RIL into line cross analysis. Kusterer et al. (2007) crossed *Arabidopsis thaliana* C24 and Col-0 genotypes to produce F7 recombinant inbred lines, then crossed these RIL to both parents and F1 in what is known as a triple test cross (TTC) design. RIL, RIL X C24, RIL X Col-0, and RIL X F1 generations were all used in line cross analysis and their results suggested that pairwise and higher order epistasis are important components of the genetic architecture of heterosis for biomass in C24 X Col-0 *Arabidopsis* lines. While the TTC design allows one to estimate non-additive genetic variance components, these additional crosses are not necessary to reap benefits of using RIL in LCA. We suggest purchasing RIL from stock centers to reduce the time consuming crosses necessary for more complex breeding designs.

The reductions in variance used as an illustration in this paper are predicated on the assumption of equal variances in the estimate of every generation line mean. This is not necessarily a realistic assumption, particularly for the RIL generation. First, RIL populations are perforce large. The best RIL populations in many species contain 200 - 400 RILs and these are often grown and measured in multiple replicates for the purpose of QTL analysis. Line cross generation means are typically calculated with far fewer measures and hence degrees of freedom. Thus we might expect the variance of the mean to be substantially smaller for the RIL mean than for other generations. However, RIL populations very often show transgressive segregation, even when the parents are phenotypically similar. In fact Rieseberg et al. (1999) report that 155 of 171 segregating hybrid populations they examined manifested transgressive segregation. We should therefore expect that the F2 and the RIL generations might show higher phenotypic variance than for example the P1, P2, or F1 generations (all three of which are genetically

identical within generations and thus will have low variance relative to other generations), and this effect will be exaggerated in RIL compared to the F2 because all individuals are homozygous at virtually all loci. There are thus two offsetting effects on the variance associated with the phenotypic mean of the RIL generation: large sample size reducing the variance of the mean and transgressive segregation and homozygosity increasing the mean's variance. The net effect can only be determined empirically.

If a RIL population is used within a line cross analysis, little extra work is required for QTL mapping. The QTL mapping results will give qualitatively different information on genetic architecture, information that compliments the results of the line cross analysis. QTL mapping can potentially find the number of regions with additive effects (QTL) and the magnitude of those effects, as well as additive x additive epistatic regions responsible for the composite effects detected in line cross analysis. Additionally, QTL mapping may detect loci with equal and opposite effect that are invisible to LCA. For example, if the P1 allele at locus A adds 5 units to the phenotype but the P1 allele at locus B reduces the phenotype by 5 units, LCA will not detect this zero net additive difference between parents. Such canceling effects are clearly often present, evidenced by RIL population parents having very similar phenotypes but widely transgressive segregation in the inbred F2 descendants (reviewed in Tonsor et al. 2005).

Comparison of additive and additive x additive effects in LCA and QTL analysis can be used to cross validate each result. One would expect that QTL effects summed across the genome will produce a total equal to the composite directional effect produced in LCA. In practice, this may not be the case. QTL analyses are widely known to produce biased results, with QTL number being underestimated and magnitude being over estimated, especially when the number of RIL is small (Beavis 1998, Xu 2003). Differences between composite A and AA effects from LCA and from the summed effects of all QTL discovered may indicate that such biases are present. Additionally, when QTL effects are directional but are too small to be detected by QTL analysis, their sum may still be detected as a difference between means in the line cross analysis. Finally, line cross analysis complements QTL mapping by detecting genetic architecture invisible to QTL analysis. LCA can detect dominance effects and epistatic effects containing dominance that cannot be detected using RIL based QTL mapping.

In summary line cross analysis is a powerful method based on linear contrasts of generation means. Using recombinant inbred lines as a generation in LCA greatly increases the power to detect non-additive genetic effects. Line crosses can detect additive, dominant and epistatic genetic effects of any kind as long as the number of generation means matches or exceeds the number of genetic effects to be estimated. Line cross analysis may detect small genetic effects missed by QTL mapping when effects are directional. QTL mapping using recombinant inbred lines has the ability to detect effects (QTL) of opposite sign invisible to line cross analysis. It can also detect additive and additive by additive epistatic

QTL. It can be used to find the location of QTL for effects detected in line cross analysis. Recombinant inbred lines can be purchased from stock centers so that the time and work required to produce them is avoided. QTL studies that wish to incorporate additional line crosses will only require small increase in sample size on the order of 20%. On the other hand, line cross studies will require adding a much larger sample size to add a set of RIL lines large enough for QTL mapping. But these additional organisms phenotyped will not require the time-consuming crosses. Adding line crosses to a QTL experiment or a RIL population to a line cross experiment results in a large increase in ability to measure genetic architecture that will more than justify the modest increase in research effort and cost. Increased statistical power, qualitatively different measures, cross-validation of results, and potential to overcome weaknesses of each approach alone makes this a very powerful approach to gaining a fuller understanding of genetic architecture.

Table 2.1. Source and hybridity indices and coefficients of directional genetic effects. Source and hybridity indices and the resulting coefficients for the genetic effects in line cross equations, including all two-way epistatic interactions, after (Lynch and Walsh 1998, chapter 9). Lines are created by crossing inbred parent 1 (P1) with inbred parent 2 (P2) to produce the F1 and F2 generations as well as reciprocal backcrosses to P1 (B1) and P2 (B2). Recombinant inbred lines (RIL) are formed by repeatedly selfing the F2s. The meaning of the columns: S = proportion of genome from P1; H = proportion of heterozygous loci; θ_S = source index, indicating the relative contributions of P1 and P2 to the generation genome; θ_H = hybridity index, indicating expected heterozygosity of the generation's genome on a scale of 1 to -1. μ = the mean phenotype of the F2 generation. The values in the remaining columns indicate expected contribution of the column's genetic effect to the phenotype of the row's generation. The effect types: A = additive; D = dominance; AA = dominance by dominance interaction; AD = additive by dominance interaction; DD = dominance by dominance interaction.

Line	S	H	θ_S	θ_H	μ	A	D	AA	AD	DD
P1	1	0	1	–1	1	1	–1	1	–1	1
P2	0	0	–1	–1	1	–1	–1	1	1	1
F1	1/2	1	0	1	1	0	1	0	0	1
F2	1/2	1/2	0	0	1	0	0	0	0	0
B1	3/4	1/2	1/2	0	1	1/2	0	1/4	0	0
B2	1/4	1/2	–1/2	0	1	–1/2	0	1/4	0	0
RIL	1/2	0	0	–1	1	0	–1	0	0	1

Table 2.2. Comparison of variances of directional genetic effects with and without recombinant inbred lines. Comparison of variances of directional genetic effects with and without recombinant inbred lines. RIL-based traditional line cross equations and variance reduction under the assumption of equal variances in the estimate of the means in all generations. Typically, RIL populations will have a lower variance for the estimate of the mean because of their larger sample size. z_{Xi} = the phenotypic mean of the Xith generation (eg. z_{F1} = mean of the F1 generation).

Effect	RIL equation	Variance of RIL-based estimate	Traditional equation	Variance of Traditional estimate	Variance reduction
D	$(z_{F1} - z_{RIL})/2$	$0.5(\sigma^2)$	$-z_{P1}/4 - z_{P2}/4 + z_{F1}/2 - 2z_{F2} + z_{B1} + z_{B2}$	$6.375(\sigma^2)$	92%
AA	$(z_{P1} + z_{P2})/2 - z_{RIL}$	$1.5(\sigma^2)$	$-4z_{F2} + 2z_{B1} + 2z_{B2}$	$24(\sigma^2)$	94%
DD	$(z_{F1} + z_{RIL})/2 - z_{F2}$	$1.5(\sigma^2)$	$z_{P1}/4 + z_{P2}/4 + z_{F1}/2 + z_{F2} - z_{B1} - z_{B2}$	$3.375(\sigma^2)$	56%

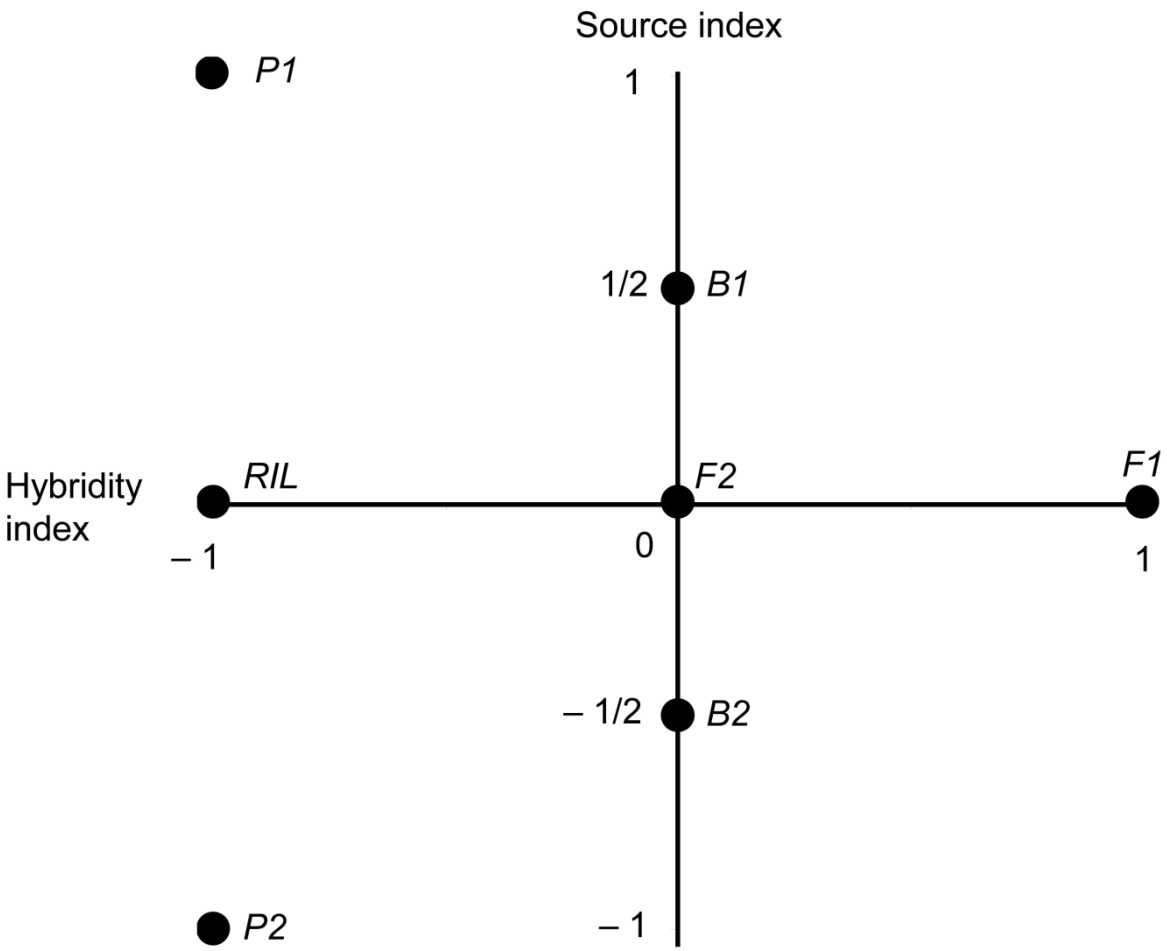


Figure 2.1. Source and Hybridity indices for various line cross derivatives. Source and hybridity indices for the various generations of a line cross population. The vertical axis indicates each generation's source index. A source index value of +1 indicates that all genes originate with P1 while -1 indicates that all genes originate with P2. The horizontal axis indicates a generation's hybridity index such that +1 indicates heterozygosity at every locus, while -1 indicates homozygosity at all loci. The RIL values represent an ideal in which an infinite number of generations of selfing preceded measurement of the RIL population. Real RIL populations asymptotically approach this value as the number of generations of inbreeding increases.

3.0 EVOLVABILITY IN A VARIABLE WORLD- CONSTRAINTS

3.1 ABSTRACT

We investigated potential constraints to evolvability using sets of *Arabidopsis thaliana* recombinant inbred lines grown across a nitrogen (N) supply gradient ranging from saturating to near limiting N- levels. We used several approaches to quantify constraints and genetic architecture of 13 traits in each of four N- environments. Our approaches involved comparisons of genetic correlations within and across environments both individually and through their principle component (PC) axes. Our work emphasized differences in g-max (i.e. PC1), an estimate of the genetic line of least resistance (Schluter 1996). Changes in the direction and magnitude of g-max were compared across N-environments as well as the 'mismatch' between the directions of g-max and a selection gradient imposed by the nitrogen treatments. A mismatch between g-max and selection gradients suggests that a response to multivariate selection would be constrained by a lack of heritable variation. Measures of genetic correlations were complemented with N-environment specific QTL mapping to elucidate the mechanistic basis of changes in trait correlations. Our findings here indicating that patterns of genetic variation are environmentally specific and that under extreme environmental conditions, patterns of genetic constraint can be altered substantially, suggesting evolution may proceed in different directions.

.

3.2 INTRODUCTION

Studies of the genetic architecture of quantitative traits inform us about the potential for traits to evolve by natural and artificial selection. With rapid environmental shifts due to anthropogenic change, understanding of adaptation and constraints to adaptation are of paramount importance (Bell and Gonzalez 2009, Futuyma 2010).

Constraints to adaptation can take several forms (reviewed in Futuyma 2010). A genetically based trait correlation (i.e. due to pleiotropic gene action or linkage disequilibrium between alleles at different loci) can constrain independent evolution and thus optimization of individual characters (McGuigan 2006). This is true whether considering a genetic correlation between two traits or a genetic correlation between a trait and itself measured in two different environments (i.e. cross- environment genetic correlations). Selection on one such trait will produce a correlated response in the other (Lande and Arnold 1983). Negative genetic correlations have been widely documented in many taxa especially for traits that affect fitness; such trade-offs form much of the basis of life history theory (Roff 2002). These genetic correlations (r_g) are not expected to be insurmountable or absolute constraints (Mezey and Houle 2005) if they are not perfect (i.e. if $r_g \neq \pm 1$). Theoretical (Via and Lande 1985) and empirical (Connor 2003) work suggests that r_g different from ± 1 can bias the direction of a response to selection, temporarily resulting in maladaptation, but that character means will eventually reach optima (Roff and Fairbairn 2006). Biases such as these may be especially important when rapid changes in the environment require adaptation to proceed quickly to prevent local extinction (Gomulkiewicz and Houle 2009).

Adaptive response can also be constrained by a lack of genetic variation. Several authors have pointed out that nearly every continuous trait studied displays some level of genetic variation (Lynch and Walsh 1998, Barton and Partridge 2000, Barton and Keightley 2002). While few studies have found particular characters possessing little or no heritable variation (e.g. cotyledon number in wild radish [Connor and Agrawal 2005] and metal tolerance in several plant species [Bradshaw 1991]), the abundance of genetic variation may present a misleading picture of constraints when traits are considered in isolation. Even when individual traits or trait correlations are genetically variable, variation in particular multivariate directions may be lacking and can present absolute constraints as to what can evolve (Blows and Hoffmann 2005, Kirkpatrick 2009). In this situation, the set of genetic covariances among traits, summarized as a G-matrix, will have fewer dimensions than individual traits (Schluter 2000), indicating absolute constraint on the evolution of the corresponding phenotype (Roff and Fairbairn 2006, Kirkpatrick 2009, Gomulkiewicz and Houle 2009) at least until mutation can generate new variation (Jones et al. 2007). Analysis of multivariate genetic variation using principle components (hereafter PC) can identify the directions in trait space in which selective response is most or least likely. The direction of greatest variation, sometimes called the 'genetic line of least resistance' (i.e. 'g-max' or PC1 of the G-matrix, Schluter 1996) can bias the response to selection, and in fact, evidence exists that bias along g-max may shape selection and population divergence over evolutionary timescales (Schluter 1996, Schluter 2000, Begin and Roff 2003, Arnold et al. 2008). Trait combinations lacking genetic variation can cause absolute constraint on the evolution of corresponding phenotypes; such directions are

indicated by the existence of PC axes with eigenvalues equal to zero (Kirkpatrick 2009). When estimates of phenotypic selection are coupled with analyses of patterns of genetic variation, whether adaption will be slowed (biased away from optima) or prevented (absolutely constrained) can be predicted. For example, genetic correlations among leaf traits and reproductive stage in *Chamaecrista fasciculata* opposed the direction of selection caused by simulated global warming conditions in a field experiment, resulting in a greatly reduced response relative to that predicted from univariate trait heritabilities (Etterson and Shaw 2001). Chenowerth et al. (2010) found that the genetic covariance structure strongly biases response to sexual selection on male contact pheromones in *Drosophila serrata*. When selection was nearly orthogonal to g-max, the response was more determined by the direction of g-max than the sexual selection gradient. They suggest that a g-max dominated evolutionary response is likely when genetic variation is much greater in particular directions than others (large differences in eigenvalues of PC axes) and selection acts in a direction of little genetic variation. Coupling selection analysis with identification of the major axes of genetic variation can tell us much about constraints by identifying such 'mismatch' between genetic variation and selection.

Genetic architecture can be altered by environmental factors (Weinig 2002) changing patterns of trait heritabilities (Connor et al. 2003), correlations (Sgro and Hoffman 2004), and functional integration (Tonsor and Scheiner 2007). For example, in a field experiment, wild radish floral traits had lower heritabilities, reduced additive genetic covariances, and changed G-matrix structure relative to individuals from the same source population grown in a greenhouse (Connor et al. 2003). With environmental effects being common, studies of evolvability for organisms that experience multiple environments should be conducted across those environments to determine robustness of constraints. If constraints exist within only one of several commonly experienced environments, we learn little about the potential to evolve from any single environment study. Clearly, laboratory studies of adaptive constraint should include multiple environments that span the relevant ecological natural conations (Sgro and Hoffman 2004).

One particularly relevant set of environmental conditions, especially for plants, is limiting resource supply rates. Because of their sessile nature, plants do not choose their environment actively as animals do. Both course- and fine-grained patchiness of resources are likely experienced by many individuals in natural populations. Conditions of limited resource availability are likely to reveal trade-offs. Trade-offs can result from allocation of shared resources between traits (Sterns 1992) and these trade-offs are expected to be more apparent when resources are limited, since increased allocation to one trait may reduce resources to the other (Tonsor and Scheiner 2007). Although some trade-offs are expected to be more apparent at low resource levels (Gehring and Linhart 1993), trade-offs can be masked by variation in the acquisition of resources (Houle 1991, Sterns 1992, Gardner and Latta 2007, Roff and Fairbairn 2009). In some circumstances, quantitative trait locus (QTL) mapping can reveal

allocational trade-offs, even when positive genetic correlations between traits are produced by variation in resource acquisition (Gardener and Latta 2007).

Measures of genetic correlations and heritabilities can be complimented with QTL mapping (Kelly 2009). Often genetically correlated characters are reported to have shared underlying QTL (Gardner and Latta 2007) and if these QTL are indeed pleiotropic, genetic correlations are more likely to affect long term evolution. If QTL actually represent separate genes within the same chromosomal region, genetic correlations can more easily be altered by recombination. QTL mapping can be a good first approximation to determining whether linkage or pleiotropy is the cause of a correlation. While many QTL mapping approaches do not have the resolution to distinguish pleiotropic gene action from tight linkage, it can be used to rule out the pleiotropy hypothesis when QTL are found that do not occur in the same genomic position. Mackay et al (2003) found pleiotropic effects of loci acting on both flowering time and water use efficiency in *Arabidopsis thaliana*. Gardener and Latta (2007) found in their literature review that QTL with antagonistic effects on two traits (interpreted as a trade-off) failed to produce negative genetic correlations in more than one third of the 276 trait pairs analyzed, due to the masking effects of other QTL with pleiotropic positive effects on the same traits. Thus QTL mapping has the potential to reveal trade-offs in the absence of negative genetic correlations. Additionally, the number and effect size of QTL underlying genetic variation can give information as to the likelihood of the variation becoming depleted over time, as traits determined by a greater number of loci are likely to regenerate variation by mutation more quickly (Houle 1996).

We investigated how the genetic architecture of 13 quantitative traits in *Arabidopsis thaliana* (hereafter Arabidopsis) changes across an ecologically relevant nitrogen (N) gradient from extremely limiting to saturating N supply rates. Our focus was on the constancy of potential adaptive constraints in the form of genetic correlations, available heritable variation, and principle component structure across N-environments. Nitrogen availability is of critical importance to the majority of temperate plant populations, with it being the nutrient most limiting to population growth across a range of species (Vitousek and Howarth 1991, Diaz et al 2006). Nitrogen supply variation is of further interest because soil N levels have been changing on a global scale with increases paralleling the rise of CO_2 and temperature associated with anthropogenic change (Gruber et al 2008).

Arabidopsis recombinant inbred lines (RIL) derived from a cross between Landsberg erecta and Cape Verdi Islands accessions were grown in controlled environment growth chambers for this study. To determine how genetic architecture changes across four N- environments (two limiting and two near saturating N supply rates), we used several approaches. **First**, we compared genetic variation and correlations within and across each environment. **Second**, we compared covariance structure across environments using PC analysis, asking whether g-max changes direction or magnitude, and whether the

effective dimensions of genetic variation are less than the number of traits analyzed within each N-environment. **Third**, we mapped additive QTL within each environment asking if QTL quantity or location changed across N-environments. We then compared QTL locations with cross-environment genetic correlations, asking whether changes in genetic correlations across N-environments corresponded to changes in QTL locations. **Fourth**, we investigated the 'mismatch' between g-max and the selection gradient imposed by N-availability within each growth chamber environment. We present our findings here indicating that patterns of genetic variation are environmentally specific and that under extreme environmental conditions, patterns of genetic constraint can be altered substantially, suggesting evolution may proceed in different directions. In a companion paper (Elnaccash and Tonsor, in prep) we investigate how the pattern of multivariate genetic variation affects simulated evolutionary trajectories.

3.3 METHODS

3.3.1 Plant material

Arabidopsis thaliana is a highly selfing annual plant in the family Brassicaceae. A fully sequenced genome, a wealth of genetic tools, and large number of ecotypes available from stock centers (TAIR) make this plant highly useful for studies of genetics and development. Arabidopsis is becoming a model system of ecological research as well (Tonsor et al 2005, Mitchell-Olds and Schmitt 2006) especially as more data is being collected from relatively undisturbed natural populations (e.g Montesinos et al. 2009, Montesinos-Navarro et al. 2010).

Our research used a set of 158 RIL (CS22000) derived from a cross between Landsberg erecta (Ler-2, CS8581) and Cape Verdi islands (Cvi-1, CS8580) ecotypes (Alonso-Blanco et al.1998) plus the Cvi and Ler parents (i.e. 160 genotypes). All seeds were obtained from the Arabidopsis Biological Resource Center (ABRC) of Ohio State University. Approximately three replicates of each genotype were grown at each of four Nitrogen supply rates in Conviron controlled environment growth chambers.

3.3.2 Plant growth conditions and phenotyping

Planting was staggered to accommodate growth and measurement of the large number of plants and to have replicate growth chamber environments to account for chamber effects. Seeds were cold stratified in dark at 4°C for 5d. Plants were grown in 164 mL SC10 Conetainers plastic pots (www.stuewe.com/products/rayleach.php) filled with washed Turface (www.turface.com/) and stored 24/ rack each rack in a fiberglass bin. Racks and bins were placed in one of 4 PGW 36 Conviron growth chambers, each supplying light at 270 µMoles (ramping from/to 100 µMoles during first/last light hours) photons m^{-2} s^{-1} with 16h/8hr light/dark cycles per 24h period. Temperatures cycled between 15°/22°C during night/day. Water was supplied to bins through an automated ebb and flood system designed by SJT. Water filled to a set height in each bin at simulated dawn and remained full for 45 min before draining. Nitrogen was added at 56 ppm, 51 ppm, 6 ppm, and 1ppm concentrations supplied as NO_3^-. Background NO_3 concentration in the water supply contributed on average an additional 1 ppm. Nutrients were supplied using four Dosatron® D25RE2 (www.dosatronusa.com/) adjustable nutrient apportioners. Supply rate was calibrated weekly using an ion-specific probe to measure NO3 concentration as the water entered the bins. Actual NO_3 concentrations were maintained +/- 1-2 ppm at the two lower concentrations and +/- 2-3 ppm at the two higher concentrations.

These nitrogen levels were originally chosen for an experiment on developmental instability (Tonsor 2011 in prep) to represent saturating and limiting N-supply levels. These N-environments will be referred to as N56, N51, N06, and N01 or occasionally as high (N56 and N51) and low (N06 and N01) N-environments. This range spans then natural range of N availability across several US populations (Diane Beyers, University of Illinois, personal communication to SJT). Three growth chambers were used for four runs each from September 2001 to June 2004.

Thirteen traits were measured on each plant. Traits were chosen because they are likely to be important targets of selection for adapting to various parts of Arabidopsis' natural range, they have been previously shown to have genetic variation, and because they are likely affected by N-supply rates. Morphological traits included number of basal reproductive branches (rb), total length of reproductive branches (totlen), number of rosette leaves at bolting (nlvs), total fruit length (tsl), biomass(totg), age at bolting (boltd), proportion of root/total biomass (proprt), each measured at day 70 past stratification after drying at 65°C. Our fitness proxy, total fruit length, was calculated as the product of total length of reproductive branches, the number of fruit per 10cm branch length, and the average length of 5 fruit. Physiological measures included whole plant instantaneous carbon gain (delco2) and transpiration rate (delh2o), both recorded 29d post stratification were measured at 270 µMoles photons $m^{-2}s^{-1}$ and 500 ppm CO_2 with whole plant gas exchange cuvettes connected to a LiCor 6400 IRGA (see Tonsor & Scheiner

2007, Earley et al. 2009 for details). Fluorescence-based photosynthetic yield (yield) and photosynthetic electron transport rate (dietr) were measured on the same day with a Walz PAM-2000 fluorometer (www.walz.com). Photosynthetic quantum efficiency (fvfm) was calculated from quantum yield and electron transport rate. Percent nitrogen was measured by grinding plants to powder, homogenizing and sub-sampling for measurement of whole-plant %N using a Perkin-Elmer 2400 elemental analyzer (www.perkinelmer.com/).

All plant traits were adjusted for growth chamber runs. For each nitrogen treatment, we calculated the mean trait values in each of the five growth chamber runs to produce a treatment mean. We then adjusted the mean of traits within each growth chamber run to match the treatment mean, effectively removing variation due to the particular run of the chamber. The resulting chamber-adjusted residuals were used for all analyses.

3.3.3 Statistical analyses

All statistical analyses were carried out using SAS/STAT, version 9.2 for Windows (SAS Institute 2008). RIL means were used as the experimental observations for each analysis unless stated otherwise. RIL are genotypic means and we use them to estimate genetic correlations (see below) and genetic covariances (chapter 4). Two analyses were not based on RIL means. Estimates of heritabilities (using genetic variance components) and selection analyses both require analysis of individual plant measurements.

A one-way MANOVA tested the effect of N supply rate on the multivariate plant phenotype (SAS, PROC GLM using MANOVA statement). The thirteen study traits were dependent variables and N-environment was the sole independent variable. Given the significance of the overall test (see Results), the univariate main effects were examined with ANOVA using PROC GLM. Each trait was used in one-way ANOVA as the dependent variable with N-treatment as the independent variable. Means were compared across N-environments using the criterion of non-overlapping 95% confidence intervals. This test criterion produced the same significant differences among means as adjustment using Fisher's Least Significant Difference and Tukey's Honestly Significant Difference, and differed from Bonferroni and Scheffe's multiple comparison procedures in two or fewer out of 78 comparisons (data not shown).

Genetic variances were calculated within each nitrogen environment as the among-RIL variance component from a random effect ANOVA using SAS PROC MIXED in which the univariate phenotype was the response variable and RIL was a random effect. Genetic variances were used to calculate heritability (genetic variance/total phenotypic variance) and coefficients of genetic variation (genetic

variance/ mean, Houle 1992). Differences in heritabilities and coefficients of genetic variation were compared between all pairs of nitrogen treatments. Our heritabilities and coefficients of genetic variation follow an unknown distribution and should not be compared using parametric tests that assume normality. The Wilcoxon sign rank tests, i.e. non-parametric paired t-tests, were used for comparisons. Sequential Bonferroni corrections were used to account for multiple comparisons.

3.3.4 Genetic correlations

Shared genetic basis of traits can be estimated with genetic covariances or genetic correlations. Genetic covariances are useful for evolutionary studies as the genetic variances are directly relevant to models of short term evolutionary change such as the breeder's equation (Lande 1980). Difficulties can arise when using genetic variances and covariances when traits are measured on vastly different scales (e.g. compare units for percent nitrogen, bolting day, and biomass) and differences in variance may arise solely from the units in which traits were measured. Genetic correlations are standardized covariances and they can be advantageous for facilitating comparisons of traits measured on different scales. For this reason, genetic correlations are used for the analyses in this chapter and the use of genetic covariances is deferred to chapter 4 which reports on simulated evolutionary trajectories.

Between-trait (within environment) genetic correlations were estimated as the Pearson product-moment correlations among RIL means. Estimating genetic correlations as the correlations among genotype means is common practice (e.g. Brock and Weinig 2007; for a review of methods to estimate genetic correlations see Astles et al. 2006). This approach does not produce strict additive genetic correlations since dominance, epistasis, and maternal effects can contribute to the estimates (although dominance is expected to be negligible since RIL are homozygous at nearly all loci). Additionally, a fraction of the within RIL variances can contribute to the correlation estimate (e.g. see Via 1984). With these caveats in mind, the correlations should be interpreted as broad sense genetic correlations. These genetic correlations were estimated using all available data, thus samples sizes vary depending on the specific traits or environments.

Each genetic correlation was tested for significance using PROC CORR in SAS. 95% parametric confidence intervals were calculated around each genetic correlation to test the null hypotheses of $r_g = 0$ and to compare correlations across N-environments. Correlation coefficients were were considered significantly different across N-environments if the 95% CIs did not overlap. Note that this is a more conservative test than asking whether one correlation falls within the 95% confidence interval of another; this will help compensate for the large number of statistical comparisons.

Cross-environment genetic correlations were estimated as the correlation among RIL means for a single trait measured in each pair of nitrogen environments. These correlations have the same caveats as the between-trait genetic correlations and should also be interpreted in the broad sense.

3.3.5 Principle component analysis (PCA)

We used a SAS macro program (randomPCA created by SJT, www.pitt.edu/~tonsor/downloads/DOWNLOADS.html) to determine the number of significant PC axes. Briefly, the RandPCA macro randomly permuted the traits relative to one another in the original data, recalculated the PCA, and repeated this 10 000 times. This produced a null distribution of the eigenvalues for each PC axis given random associations among traits. When the actual eigenvalues had a greater magnitude than 95% of the randomly generated eigenvalues, the corresponding PC axes were considered significant.

A second SAS macro, BootPCA (created by SJT, www.pitt.edu/~tonsor/downloads/DOWNLOADS.html), was used to create bootstrapped 95% confidence intervals on eigenvalues and loadings for each PC axis based on 10 000 bootstrapped samples. Eigenvalues of the genetic correlation matrix indicate the maximum amount of the variance in the traits that can be accounted for with a linear model by a single underlying variable (Friedman and Weisberg 1981). No corrections for reflection or reordering of bootstrapped PCA were made (Peres-Neto et al. 2003). This has the effect of reducing power to detect differences in loadings and eigenvalues, i.e. confidence intervals will be wide. Any significant differences between loadings or eigenvalues would only become more significant with corrections. Eigenvalues and PC axes loadings were compared across nitrogen environments using bootstrapped 95% confidence intervals.

Both BootPCA and RandPCA are written to be easily adapted to new datasets with an arbitrary number of user specified traits and treatments/populations and can be downloaded from SJT's website (www.pitt.edu/~tonsor/downloads/DOWNLOADS.html).

We then compared the first principal component (PC1 or g-max) of the r_g matrix across N supply treatments. We treated each PC1 as a vector in multivariate trait space described with the coefficients used to transform the r_g matrix into orthogonal axes. We calculated the angle between these vectors across N supplies as a measure of similarity in potential biases to evolution, with a greater angle indicating greater differences in the direction of maximum evolvability. Vector correlations describe the difference in the directions of these vectors and they equal the cosine of the angle between them. Vector correlations are calculated as:

$$\frac{V_1^T * V_2}{\sqrt{(V_1^T * V_1) * (V_2^T * V_2)}} \quad (1)$$

,with V1 and V2 being vectors (such as PC1 axis loadings), T indicating transpose, and * indicating vector multiplication. The resulting vector correlation will be between (-1, 1), with numbers near zero indicating weak or no correlation (i.e. vectors are nearly orthogonal) and numbers near one indicating vectors point in the same direction.

We used Kirkpatrick's (2009) estimate of the effective number of dimensions (n_D) for traits. Briefly, this is a descriptive measure of the number of independent directions in which selection can proceed, with a maximal value equal to the number of traits analyzed. It is calculated as the sum of the eigenvalues of a mean standardized G-matrix divided by the largest eigenvalue. Kirkpatrick states that this is not the only standardization possible and we chose to use the standardization inherent in our correlation matrices for these analyses. The effective number of dimensions is the reciprocal of the fraction of the total genetic variance explained by PC1 (e.g. if PC1 explains 25% of the total variance then $n_D = 4$). Since this is based on the correlation matrix, with each trait standardized to unit variance, n_D = 12/(eigenvalue of PC1).

3.3.6 Quantitative trait locus mapping

In each N-supply treatment, 160 line means (158 RIL +2 parents) were used for mapping of all thirteen study traits with R/QTL (Broman et al. 2003). RIL means were used for phenotypic input data. Marker genotypes for 163 markers with unique map positions were downloaded from The Arabidopsis Information Resource (ftp://ftp.arabidopsis.org/home/tair/Maps/Ler_Cvi_RIdata/).

The presence of transgressive segregation (TS) was determined by calculating the percentile of the distribution of RIL means in which the parental genotypes fell. Any rank beyond the first or beneath the 99th percentile for either parent indicates TS.

We mapped using interval mapping employing a multiple imputation method designed by Sen and Churchill (2001). Pseudomarkers (simulated marker genotypes) were imputed at 1 cM intervals to give complete and even marker coverage along each chromosome and 100 imputations were used for each study trait /N-environment combination. We report findings of additive QTL for each of 52 traits (13 traits x 4 N-environments); more detailed QTL mapping of plasticity and epistasis (QTL x QTL and QTL x QTL x E interactions) will be reported in a later paper. Our use of multiple imputation interval

mapping was chosen for consistency with our larger 2D survey; multiple imputation methods can currently be implemented for both 1D and 2D genome scans unlike more commonly used approaches such as composite interval mapping. Maternal cytoplasm was used as an "additive covariate" in QTL mapping, i.e. it affected significance tests but not the location of QTL. The LOD threshold for QTL detection was determined by 1000 permutations (Doerge and Churchill 1996). Support intervals around QTL positions were determined by dropping 1.5 LOD from a significant QTL peak and finding the corresponding two locations to the nearest cM. We did not use the loadint() command in R/QTL to find support intervals. Loadint() takes all locations across the chromosome within 1.5 LOD of the peak and merges them into one contiguous support interval; it tended to produce much larger support intervals which often spanned whole chromosomes when multiple peaks were present. When two QTL for the same trait/N-environment combination had overlapping 1.5 LOD support intervals, only the QTL with the highest LOD was reported and retained for analyses.

3.3.7 Measures of selection

We carried out a "laboratory natural selection" experiment (Fuller et al. 2005) and measured selection imposed by the growth chamber/nitrogen environment using fitness regression (Lande and Arnold 1983). Total fruit length was used as a fitness proxy (hereafter fitness) and was standardized to a mean of 1. Fitness was regressed onto the twelve remaining traits (each standardized to a mean of zero and SD = 1) using SAS PROC REG. Selection gradients (i.e. the set of partial regression coefficients) were estimated within each N-environment. The regression used individuals (not RIL means) as the experimental unit and the model was TSL = TOTG PROPRT NPCT NLVS BOLTD DELCO2 DELH2O FVFM DIETR YIELD TOTLEN RB, with each trait a fixed effect.

Significance testing of the selection analyses was carried out by transforming traits to normalize and remove heteroscedasticity from the distribution of residuals. ANCOVA was used to test whether the pattern of selection differed across environments. The ANCOVA model used TSL as the dependent variable and each trait, N-environment, and each trait*N-environment interaction as fixed independent variables. Selection analysis was used to generate point estimates of selection gradients for comparison with g-max as an instantaneous measure of constraint.

3.3.8 Constraints to selection

Directions of multivariate selection were compared to the genetic line of least resistance (i.e. PC1) within each environment via vector correlations. High vector correlations indicate that selection acts in the direction of g-max, the axis of greatest genetic variation, while low vector correlations indicate potential constraint as g-max can bias the short term response to selection in a different direction than optimal for increasing population fitness (Chenoweth et al 2010). Statistical comparisons of PC axes and selection gradients were used to interpret differences in constraint angles; no statistical tests were used for the vector correlations themselves.

3.4 RESULTS

MANOVA revealed a significant main effect of nitrogen treatment (Wilks' $\lambda = 0.21$, F (39, 1280) = 23.03, $p < 0.0001$) on multivariate plant phenotype. Surprisingly, MANOVA was also significant when the two low nitrogen environments (Wilks' $\lambda = 0.41$, F (13, 152) = 16.66, $p < 0.0001$) or the two high nitrogen environments (Wilks' $\lambda = 0.80$, F (13, 268) = 5.07, $p < 0.0001$) were analyzed alone, indicating the phenotype differed across all pairs of N-environments.

While sample sizes tended to decrease with decreasing nitrogen supply rate, we concluded that this was not the cause of mean differences between treatments. The traits with the greatest number of missing values were fluorometry traits (yield, fvfm, dietr), since many plants were too small for their measurement in the lowest N-environment. We coded a categorical variable MISS as '1' if plants were missing data on any of the 3 fluorometry traits within N01 and '0' otherwise. MANOVA testing for multivariate differences in the 10 remaining traits between plants that were missing and not missing fluorometry data (in N01 only) proved to be non-significant (Wilks' $\lambda = 0.77$, F (10, 62) = 1.90, $p = 0.0621$). While the p-value is low, it is large compared with the multivariate differences between the N-environments, each significantly different as indicated by $p < 10^{-4}$. While there may be real differences in N01 between the plants included and excluded from multivariate analyses due to missing data, these differences appear to be orders of magnitude too small to account for the differences between N01 and other nitrogen environments.

Univariate ANOVA showed significant differences among N-environments for all traits (data not shown, all traits $p < 0.003$, except boltd, $p = 0.052$). For many traits, comparisons of 95% confidence intervals showed no significant differences between the two high nitrogen environments (table 3.1). With the exception of transpiration rate, carbon assimilation rate and basal reproductive branches, trait means did not differ between N56 and N51. Plants in N06 tended to differ from N56 and N51 across most traits, with N06 plants having significantly less biomass, greater root investment, lower carbon assimilation and transpiration rates, lower quantum efficiency and yield, and lower total branch length, basal reproductive branches, and total fruit length. Nearly all traits measured in N01 had significantly reduced means relative to the two high nitrogen environments. Additionally, plant traits in N01were often reduced relative to traits measured within N06. For example, N01 plants had significantly less biomass, lower carbon assimilation and transpiration rates, lower quantum efficiency and quantum yield, fewer basal reproductive branches and less total fruit length than plants measured in N06. Only bolting day had confidence intervals overlap across all N-environments. In summary, phenotypes were similar in the two high nitrogen environments, with a steep decline in N06, and a further steep drop-off in N01. Significant reductions in fitness in low relative to high N-environments relative as well as in N01 relative to N06 indicate an increase in stress with decreasing N-supply rate.

Broad sense heritabilities ranged from a high of 0.59 for total fruit length in N51 to near zero for several traits in the low nitrogen environments (table 3.1). Wilcoxon signed rank test showed no differences in heritabilities between any pair of environments (data not shown, $0.22 < p < 0.69$ for each comparison). Wilcoxon signed rank test showed a decrease in coefficients of genetic variation between high and low nitrogen environments (i.e. N56 vs. N06, N56 vs. N01, N51 vs. N06, N51 vs. N01), and these decreases remained significant or marginally significant after sequential Bonferroni correction (all $p < 0.05$, table 3.2). No significant differences in coefficients of genetic variation exist within the two higher or the two lower nitrogen environments.

Between-trait genetic correlations were similar among the three highest nitrogen environments, all of which tended to differ from N01 (table 3.3). For example, the genetic correlation between carbon assimilation rate and each of total biomass, percent nitrogen, number of rosette leaves, and bolting day ranges from $r_g = (0.30, 0.59)$ in N56, N51, and N6, but is statistically indistinguishable from zero in N01. Overall there were 5, 16, and 10 genetic correlations that were significantly different between N01 and N06, N01 and N51, and N01 and N56 respectively (table 3.4). Only 2 genetic correlations differed statistically between any pair of the three highest N-environments. These changes represent an overall decrease in magnitudes of correlations from the higher N-environment to the lower N-environment. The decreases ranged from a 12% to a 40% and never involved a change of sign (whenever a change of sign occurred at least one r_g had 95% CI overlap zero).

Cross-environment genetic correlations tended to be significant and highest magnitude between the two high nitrogen environments (N56 and N51), and smaller and non-significant between N01 any other N-environment (table 3.4). Fluorometry traits (quantum efficiency, quantum yield, and electron transport rate) all tended to have non-significant cross-environment genetic correlations.

3.4.1 Principle component analysis

Randomization tests determined that N56, N51, N06 and N01 had 2, 1, 2, and 3 significant PC axes respectively (figure 3.1, table 3.5). These were lower than n_D, the effective number of dimensions (n_D = 2.78, 2.67, 3.03, 4.18 for N56, N51, N06, and N01 respectively, table 3.5) but had the same rank order. Bootstrapped confidence intervals showed that PC1 eigenvalues were significantly larger for N56 and N51 than N01, indicating that traits were more integrated in the high nitrogen environments (figure 3.2). Additionally, the three highest nitrogen environments all had significantly larger eigenvalues for PC1 than PC2, as was expected. This was not the case for N01. In N01, the eigenvalues were 2.871, 2.492, and 1.625 for the first three axes, with consecutive axes having overlapping confidence intervals. This means PC1 explained only slightly but not significantly more trait variation than PC2, and PC2 only explained slightly but not significantly more trait variation than PC3. A three dimensional plot of PC1, PC2, and PC3 would look more spherical in N01 compared with a similar plot for N56, N51, and N06 which would appear elongated in the direction of PC1. This spherical arrangement of genotypic means suggests a de-structuring of genetic co-variation in N01 compared with the other nitrogen environments.

PC axis loadings (table 3.5) changed across N-environments. In particular, PC1 had substantially different loadings in N01 relative to other environments, often with coefficients reversing sign (figure 3.3). Bootstrapped confidence intervals on PC loadings often overlapped 0 in N01 (figure.3.3), and generally tend to be large across diverse datasets (Tonsor, unpublished).

Vector correlations between the first two PC axes were calculated between each pair of environments. The direction of g-max (PC1) was essentially the same in N56, N51, and N06, with correlations over 0.98 in each of these cases (table 3.6). This indicates that genetic variation is spread out in the same direction in multivariate trait space across these three environments. The lowest N-environment is strikingly different from the other three. Vector correlations between PC1 of N01 and N56, N51, and N06, are 0.19, 0.18, and 0.23 respectively (table 3.6). While not orthogonal, the direction of greatest genetic variation in N01 does not align with genetic variation in other environments. Vector correlations also revealed that the first PC axis in N01 is more similar to PC2 than PC1 in the other three N-environments (with correlations being more than twice the magnitude). A similar pattern for PC2 in

N01 was seen-it has a greater correlation with PC1 than PC2 in N56, N51, and N06. This suggests that the PC axes have changed order in N01 relative to the other N-environments.

3.4.2 QTL mapping

Transgressive segregation was apparent for all traits in all environments (table 3.7). TS was strong for proportion root investment, day of bolting, PS electron transport rate, and PS quantum yield-the parent's range spanned on average less than 25% of the range of the RIL distributions for these traits.

A total of 66 QTL were found for 13 traits measured in 4 nitrogen environments (figure 3.4). Of the 66 QTL, 50 had a positive effect size (Cvi-Ler)/2) indicating that the Cvi allele increased phenotype in these cases (table 3.8). For the most part, allelic effects changed phenotype in the direction of parental divergence. But in nine QTL, an allele of the parent with the greater genotypic mean resulted in a lower value for phenotype. For example, the Cvi parent has a greater total fruit length than the Ler parent in N56 (81^{st} vs. 15^{th} percentiles of distribution or RIL means respectively, table 3.7), but the Cvi allele at chromosome 1 position 10 cM has an effect of reducing TSL by 140 mm (table 3.8).

Six of the 13 traits had detectable QTL within all environments and two fluorometry traits (PS quantum yield and PS electron transport rate) did not have detectable QTL within any environment. This is consistent with the low heritability and coefficients of genetic variation of fluorometry traits (table 3.1) - a lack of trait variation can result in low probability of QTL detection. Many QTL were specific to particular environments. Some traits had QTL specific to low nitrogen (N01 or N06) or high nitrogen (N51 or N56) environments (figure 3.5). About half the traits had QTL that were in essentially the same position regardless of nitrogen environment. When viewed together it becomes apparent that QTL are not evenly distributed across the genome. The majority of QTL discovered were on the first half of chromosome 5. A second QTL hotspot was on the last third of chromosome 2, and a third hotspot was located at the beginning of chromosome 1. In total, ten distinct regions collectively account for nearly every QTL found.

3.4.3 Comparisons of QTL and genetic correlations

The magnitudes of cross-environment genetic correlations (table 3.4) were generally consistent with the pattern of QTL locations (figure 3.4). Traits such as biomass, % nitrogen, number of rosette leaves, bolting day, total branch length, and total fruit length each showed detectable QTL within all four environments. These QTL were located at one of three QTL hotspots (on chromosomes 5, 2, and 1).

Each of these traits had a significant cross-environment genetic correlation for all pairs of environments. Carbon assimilation rate had significant cross-environment genetic correlation for pairs of the three higher nitrogen environments, but not between N01 and any other environment. QTL for carbon assimilation colocate to chromosome 5 in N56, N51, N6 but no QTL were detectable in N01. Basal reproductive branches had detectable QTL in each environment except N01, and no significant cross-environment correlation was found between N01 and either high N environment.

3.4.4 Selection analyses

Phenotypic selection favored greater biomass and branch length across nearly all N-environments (table 3.9). Within N56, selection also favored more basal branches and reduced nitrogen use efficiency (the inverse of percent Nitrogen), reduced transpiration rate and PS quantum yield, although the latter two were only marginally significant. Selection in N51 and N06 favored increases in PS quantum efficiency and later age at first reproduction respectively. In N01, selection favored faster carbon assimilation and reduced PS quantum efficiency ($0.05 < p < 0.1$). Not surprisingly, the pattern of selection changed across nitrogen environments (ANCOVA results, table 3.9). Changes in the pattern of selection on total biomass, bolting day, transpiration rate and total branch length varied with nitrogen environment.

3.4.5 Constraints

Selection imposed by the nitrogen environment tended to act in a similar direction to that of g-max in all environments but N01. In N56, N51, and N06, the vector correlations between the selection gradient and the first PC axis are 0.54, 0.64, and 0.54 respectively (table 3.10). These are neither orthogonal (i.e. vector correlation near 0) nor in the exact same direction (vector correlation near 1) but somewhere in between. In N01, the selection gradient and PC1 axis are nearly orthogonal in trait space with a vector correlation of 0.025.

3.5 DISCUSSION

Understanding the environmentally specific patterns of genetic architecture informs us about the potential for selection to be facilitated or constrained in varying or changing environments. By examining

the same set of recombinant inbred lines grown across an ecologically relevant nitrogen supply gradient, we found evidence for changes in genetic correlation structure, particularly when resources are strongly limited. Furthermore, cross-environment genetic correlations tended to be significant between the three higher nitrogen environments, but not between N01 and the other environments. Overall, we found that genetic architecture changes with resource supply rate and at the most limiting nitrogen level, patterns of genetic architecture change enough to alter evolvability. This was seen across the analyses of genetic correlations, principle components, QTL mapping, and measures of the mismatch between the directions of phenotypic selection and the axes of greatest genetic variation.

3.5.1 Nitrogen environment effects

The multivariate phenotype changed across all four of the nitrogen environments. Trait means were strongly affected by reduced Nitrogen availability, particularly as the resource supply rate changed from saturating to limiting. The changes across all environments revealed by MANOVA were less apparent when analyzing individual traits. In general, a decrease of 5ppm NO3 at high nitrogen levels (between N56 and N51) had little or no measurable effect on individual trait means. Highly significant MANOVA between N56 and N51 as well as a change in significant number of principle component axes together highlight the importance of working with multivariate data in determining environmental effects and evolvability. While the two high N environments differed somewhat in their effect on multivariate plant traits, larger differences occurred between high and low N-environments and between N06 and N01. For example, our fitness proxy, total fruit length, decreased by 50 percent from either high N-environment to N06, then decreased a further 65 percent from N06 to N01. Clearly, the low N-environments resulted in moderate (N06) and strong (N01) nitrogen stress. While some authors have hypothesized that increased stress will result in increased genetic variation (Rutherford and Lindquist 1998), we found that coefficients of genetic variation were substantially reduced in the nitrogen stressed environments. The observed reduction in genetic variation in stressful environments may be the result of a general pattern in the effect of environmental variation on evolvability (reviewed in Charmantier et al. 2005 and McGuigan and Sgro 2009). This reduction in variation in the stressful N-environments was also indicated by reduction in eigenvalues in PC analysis and fewer significant between-trait genetic correlations (see below).

3.5.2 Genetic correlations

Not surprisingly, many between-trait genetic correlations were high. One would *a priori* expect morphological traits to be highly integrated (Wagner et al. 2007). Edwards and Weinig (2011) found high genetic correlations as well as co-localizing QTL for floral and vegetative morphology in *Brassica rapa*. Ashmann and Majetic (2005) reviewed over 900 genetic correlations in vegetative and reproductive traits in flowering plants and found stronger genetic correlations within floral traits and within vegetative traits than between the two categories. One would also expect aspects of plant physiology such as carbon assimilation rate and transpiration rate to be highly correlated within each environment, and previous work has shown these traits to have a genetic correlation indistinguishable from 1 in *Plantago lanceolata* (Tonsor and Goodnight 1997). We estimated this correlation to be between 0.66 and 0.88 in the Ler x Cvi RIL, depending on the environment. Little to no genetic correlation was found between the fluorometry traits quantum efficiency, yield, and electron transport rate. This is likely due to low amounts of genetic variation within each of these traits (table 3.1), which lowers the potential for correlations among fluorometry traits and with other measures.

Between-trait genetic correlations were relatively consistent across three of the four nitrogen environments. The three highest nitrogen environments had a very similar between-trait genetic correlation structure, with only two correlations changing between any pair of the three higher nitrogen environments. There are 468 potential comparisons of between-trait genetic correlations across environments. Thirty-seven of these changed between any pair of environments. Only six changes occurred between any of the three highest nitrogen environments. In contrast, there were 31 changes of between-trait genetic correlations between one of the three high nitrogen environments and N01. In most cases, this represented a significant correlation in the higher nitrogen environment which became statistically indistinguishable from zero in N01. This is a significant result given that the same set of 160 genotypes was measured within each environment.

Tonsor and Scheiner (2007) suggested that trade-offs can be environmentally mediated and that trade-offs driven by competition among processes may drive negative genetic correlations more strongly in low resource conditions. This could occur when two traits are limited by the same resource and the resource supply rates are below saturating levels. Such trade-offs should be masked when resource levels are saturating for the biosynthetic pathways that determine at least one of the traits, as this would remove the dependence between traits. We did not see an increase in negative genetic correlations in the two low nitrogen environments relative to two the high nitrogen environments.

Approximately half of the traits (total biomass, percent nitrogen, number of rosette leaves, bolting day, total branch length, and total fruit length) showed significant cross-environment genetic

correlations between all pairs of nitrogen environments. For several traits, including proportion root mass, carbon assimilation rate, transpiration rate, and basal reproductive branch number, significant cross-environment genetic correlations were limited to between N56 and N51, or to among the three highest nitrogen environments. Traits in N01 tended to be decoupled from those in high N-environments and sometimes from N06, suggesting traits can evolve independently in the nitrogen stressed vs. high nitrogen conditions. Over periods of fluctuating environmental conditions or in a patchy environment with variable nitrogen availability, we would therefore expect plants to be able to adapt to severe resource limitation as well as to more benign conditions simultaneously.

Cross-environment genetic correlations can be altered by additional environmental variables. For example, in a study of *Imaptiens capensis*, Stinchcombe et al. (2010) found negative across-density genetic correlations for shoot elongation for plants grown in high light levels. These trade-offs in growth rate across densities disappeared when the same parameters were estimated in shade conditions; the negative genetic correlations became positive. This suggests the potential for more complex patterns of cross-environment correlations to emerge from our work should we alter additional environmental variables simultaneously with Nitrogen availability. Tonsor and Scheiner (2007) suggested that the resource supply rates of multiple resources (e.g. C, N, H2O) relative to one another can alter trait integration.

3.5.3 Principle components

Principle component analysis showed that the pattern of genetic variation was reoriented in N01 relative to the three higher nitrogen environments. N01 had a significantly smaller eigenvalue for the first principal component of genetic variation than any of the three higher nitrogen environments, indicating a reduction in trait integration. N01 also had more significant PC axes than the other three environments, and the eigenvalues for these three PC axes were more similar in magnitude within N01 than in the other environments. The number of significant PC axes as determined by randomization agrees well with Kirkpatrick's (2009) effective number of dimensions, n_D, indicating that among the 12 traits, there are several directions with little to no genetic variation. This is not surprising considering small coefficients of genetic variation were found for fluorometry traits in most environments. Each trait without genetic variation should reduce the effective number of dimensions by one. The PC results indicate that the pattern of covariation becomes more "spread out" when plants are exposed to severe nitrogen stress. Large differences in the loadings of PC1 in N01 relative to other environments suggest that the axis of greatest genetic variation has been substantially altered by low N-supply rate. This change in direction of

the "genetic line of least resistance" (i.e. g-max, Schluter 1996) was summarized with the vector correlations between PC1 across the nitrogen environments. Vector correlations were 0.98 or higher between pairs of higher nitrogen environments, but dropped to approximately 0.2 between N01 and any other environment. Clearly, g-max points in a different direction in trait space in N01. Taken together, severe nitrogen stress reduced the variation available to selection as well as reoriented it in a different direction relative to other environments. While adaptation may proceed more slowly, it would be less constrained to follow a particular direction in N01.

Interestingly, the first two PC axes may have switched in N01 relative to the three higher nitrogen environments. The loadings of PC1 in N01 are highly correlated with the loadings of PC2 in the other environments and PC2 loadings in N01 are highly correlated with PC1 in the other environments. It appears that in N01, genetic variation is both reduced (e.g. lower CVg, table 3.1, and lower PC1 eigenvalues, figure 3.1) and redirected (changed loadings, figure 3.2, table 3.5, and lower PC1 vector correlations table 3.5b) relative to the other nitrogen environments. On the one hand, g-max has changed substantially, indicating that potential biases in the direction of evolutionary change do not remain constant over long timescales but can change substantially across an environmental gradient. On the other hand, the new direction of g-max in N01 is similar to PC2 in the three higher nitrogen environments (table 3.6), indicating direction of principle components (but not magnitude) is conserved.

Changes in N availability can explain differences in PC score loadings for the recombinant inbred line populations. In N01, plants have the greatest positive loading of proportion root mass and percent nitrogen and they have lower carbon assimilation and transpiration rates, as well as fewer basal reproductive branches than in other nitrogen environments. When nitrogen is scarce, plants may invest in greater root architecture to scavenge more nitrogen from the nutrient solution. Nitrogen is important for photosynthesis; approximately 30% is used in RuBisCO, the photosynthetic molecule that catalyzes the first step of carbon fixation. Nitrogen limitation can result in decreased photosynthesis, slowing both carbon assimilation and transpiration rates simultaneously as well as a change in investment in leaf and root tissues.

Our results differ from those of Mallitt et al. (2010) who studied another member of the Brassicaceae, the Pepper-grass *Lepidium bonariense.* They found the principle components of 12 morphological traits had plastic responses to light and water levels, but no overall change in covariance structure across treatments. The discrepancy between our results and theirs are likely due to the three causes. Morphological traits might be tightly integrated regardless of environmental factors. If so, one would not expect to see changes in covariance structure in Mallitt et al. (2010), but could expect it in our work considering the more diverse set of traits we measured. We measured morphology, fluorometry, physiology, and development time, none of which had correlation structures indicating separate modules

(Wagner et al 2007) yet were functionally different enough to be differentially affected by changes in Nitrogen level. If we analyze only our plant morphology data, we see similarities across all environments. Another possibility is that we saw changes in genetic covariance structure because we looked across environments which included both benign and stressful conditions. A third explanation is

that the environmental treatment itself (nitrogen supply rate vs. water and light availability) may be more likely to alter genetic covariances.

3.5.4 QTL mapping

QTL were not scattered randomly across the genome but were located in several 'hotspots'. Most traits had detectable QTL in the beginning of chromosome 5. The end of chromosomes 2 and the beginning of chromosome 1 have hotspots of decreasing importance. In total, all QTL mapped fall in one of ten different locations, with the three largest hotspots on chromosomes 1, 2, and 5 corresponding to hotspots discovered in Fu et al. (2009). Fu and colleagues mapped 40,580 molecular and 139 phenotypic traits complied from studies using the Ler X Cvi RIL population. Our largest 3 hotspots map to the same locations as the hotspots for 'metabolite abundance traits' and 'phenotypic traits' (Fu et al. 2009), with the greatest match to their phenotypic traits hotspot. The significance of these hotspots is unclear. They could represent highly pleiotropic genomic regions that have diverged because of the divergent selection pressures that Ler and Cvi have been exposed to in their natural habitats. Alternatively, these hotspots might represent deleterious mutations of large effect induced by the X-ray mutagenesis of the original Landsberg accession that produced the Ler genotype. If this were the case, we would expect that a set of RIL created by crossing Landsberg erecta with the original Landsberg line would reveal the same QTL hotspots; unfortunately the exact identity of the original Landsberg ecotype used to create Ler is unclear (see http://arabidopsis.info/CollectionInfo?id=94 for details).

The latter hypothesis, that QTL hotspots are the result of deleterious (potentially X-ray induced) mutations is consistent with our data. In nearly all QTL at chromosome 1, 3, and 5 hotspot locations, possessing the Ler allele is detrimental to plant performance. For example, Ler alleles at the chromosome 5 hotspot decrease biomass, number of rosette leaves, carbon assimilation and transpiration rates, total branch length, and total fruit length (our fitness proxy). At the chromosome 2 hotspot, the Ler allele reduces biomass, branch length and fitness in several N-environments, and also increases percent nitrogen across all environments. If we consider percent nitrogen to be the inverse of nitrogen use efficiency (the amount of nitrogen needed to produce a unit of plant material, NUE), then all of these

effects represent poor performance of plants possessing Ler alleles relative to Cvi. A similar but weaker pattern exists at the chromosome 1 hotspot. Ler alleles result in a later age at first reproduction in all N-environments. This will likely result in lower lifetime fitness as the plants will have a shorter reproductive season. Ler alleles cannot all reduce plant performance as transgressive segregation indicates that some Ler alleles (or combinations of Ler and Cvi alleles) increase fitness over that of both parents in each N-environment.

Furthermore, if QTL hotspots were due to differential adaptation at these loci, we would expect the Ler allele to be beneficial in some environments and worse in others (assuming that there were average differences in N-availability between the parental ranges within Landsberg, Germany and the Cape Verdi Islands). For, example, if one genotype was adapted to a N- rich habitat and the other a N-poor environment, we might expect QTL for percent nitrogen (i.e. Nitrogen use *inefficiency*) to have pleiotropic effects on fitness. Such differential adaptation could result in high percent N (low NUE) coupled with high fitness in high N-environments and low percent N (high NUE) coupled with high fitness in low N-environments. A lack of detectable change in the genetic correlation between percent N and total fruit length across nitrogen environments as well as the lack of co-locating QTL refute differential adaptation. If many of the QTL of large effect are due to deleterious mutations in the Ler background, this could mask trade-offs due to differential adaptation of the parental genotypes. Perhaps adaptive differences between Ler and Cvi are small, or are the result of many loci with small effects which are often missed in QTL mapping studies (Beavis 1998). Additionally, if adaptive divergence occurred in QTL near large effect 'hotspots', our method of excluding lower LOD score peaks when support intervals overlapped may have limited our ability to detect them.

For several traits, QTL locations were constant across environments. These traits without QTL X E interactions also had high cross-environment genetic correlations. For other traits, QTL co-localized in the three higher nitrogen environments, but not in N01, and cross-environment genetic correlations were significant only between pairs of the three higher nitrogen environments. This suggests pleiotropic (or tightly linked) genes affecting traits in N56, N51, and N06, with a different set of genes becoming active in extreme nitrogen stress. Loudet et al (2003) also mapped traits across nitrogen environments using *Arabidopsis thaliana* (Bayreuth X Shadara) RIL. They found significant cross-nitrogen environment phenotypic correlations for percent nitrogen, with QTL co-localizing in the two environments. In both nitrogen treatments, they found percent nitrogen QTL on chromosome 2 at approximately 35 cM, a close position to our percent N QTL at position 49 across all four environments. They found additional QTL that mapped to the same location in both environments on chromosomes 3 and 4 and QTL detectable in only one environment on each of the five chromosomes. The large differences in percent nitrogen QTL

number and locations between our studies may be due to greater divergence of parental alleles in Bay X Sha than in Ler X Cvi.

3.5.5 Correspondence between selection gradients and g-max

The pattern of selection imposed by the nitrogen treatments differed primarily in its effect on morphological measures; selection on total branch length and on total biomass differed across our treatments. Additionally, ANCOVA revealed differential selection on bolting day and transpiration rates across the N-environments but this was only marginally significant ($p < 0.1$). More extensive analysis of the relationship between traits and fitness is being conducted by SJ Tonsor and SM Scheiner using structural equation modeling (SEM). SEM reveals a more complex relationship than is possible to see using this multiple regression approach to measuring selection (Tonsor, unpublished data). Nevertheless, fitness regression tells us that differences in the nitrogen environment are strong enough to impose differential selection upon the RIL populations.

In any natural population, more complex patterns of selection will exist than those imposed by changes in nitrogen availability alone. Our measures of selection are not meant to mimic natural conditions. They do allow us to ask whether or not selection tends to act in the same direction as g-max and whether or not this changes with nitrogen availability. We found that the direction of the selection gradients and of g-max both changed across N- environments. Additionally, the correspondence between the direction of greatest increase in fitness and the most abundant genetic variation changed as N supply rate changed. Specifically, the mismatch between g-max and selection gradient was greatest in the N01 environment. Without replication of stressful environments, it is impossible to conclude that the mismatch between selection gradients and g-max tends to be greater in stressful conditions. But if we consider N01 as a single potential environment that Arabidopsis may experience, constraint due to the mismatch between selection and available genetic variation can change easily with a shift in the environment.

Since the same Arabidopsis genotypes were used in all four nitrogen environments, one could have strong expectations of a constant pattern of genetic covariation. We have shown that potential constraints to adaptation in the form of between-trait and cross-environment genetic correlations, heritabilities and coefficients of genetic variation, the direction of PC axes of genetic variation and the magnitude of their eigenvalues, and the mismatch between the direction of selection and the genetic lines of least resistance can all change across ecologically relevant resource supply gradients, and that this change primarily occurs when resources become strongly limited. Based on our data, we caution

researchers on making long term inferences about constraint imposed by the genetic lines of least resistance, g-max. Several researchers have found that g-max appears to constrain divergence over evolutionary time scales (e.g. Begin and Roff 2003). Here we've shown that g-max as well as several other measures of constraints can change over the smallest of timescales. This requires an alternative explanation for the divergence of populations and species along the first principle component of genetic variation. In Schluter's 1996 paper *Adaptive radiation along genetic lines of least resistance*, he hypothesized that g-max would constrain selection and bias the direction of population divergence and gave several examples consistent with this prediction. He also presented an alternative hypothesis, that g-max does not direct population divergence but that both are affected by some third causal factor such as natural selection. Recent simulation and modeling work suggests that if QTL exist that alter the correlation between two quantitative traits, selection can produce patterns of variation oriented in the direction of phenotypic selection, thus increasing evolvability (Jones et al. 2007, Pavlicev et al 2011). Both of these simulations suggest that selection on two traits (directional or correlational) will favor increased genetic correlation which can result in eigenvectors of G improving alignment with selection gradients. These models are complex and make many assumptions about the genetic architecture of bivariate trait correlations; it will be some time before enough empirical data accumulates to allow assessment of the plausibility of these models, particularly when dealing with more than two traits.

In summary, our results support the notion that g-max does not provide a long term constraint to adaptation. By observing changes in g-max, QTL architecture, and all our measures of constraint across different N supply rates, we find that seemingly strong constraints can be environmentally dependent and may change or disappear when resources supply rates change.

Table 3.1. Trait means and heritabilities. Trait means (and standard errors) were calculated as the average (SE) of 160 genotypic means of the Cvi x Ler recombinant inbred lines. Letters indicate significant differences in trait means across N-environments as determined by non-overlapping 95% confidence intervals. Broad sense heritability (h_B^2) was calculated as the among RIL variance component from random effects ANOVA divided by the variance among plants (phenotypic variance). Coefficients of genetic variation (CVg) were calculated as the among-RIL variance component divided by the trait mean.

	N56					N51					N06					N01				
Variable	N	Mean	SE	h_B^2	CVg	N	Mean	SE	h_B^2	CVg	N	Mean	SE	h_B^2	CVg	N	Mean	SE	h_B^2	CVg
totg	159	0.909 [a]	0.047	0.507	0.000	153	0.840 [a]	0.046	0.516	0.000	157	0.261 [b]	0.017	0.272	0.000	158	0.066 [c]	0.006	0.209	0.000
proprt	159	0.072 [a]	0.003	0.056	0.000	153	0.081 [a,b]	0.004	0.034	0.000	156	0.092 [b,c]	0.006	0.010	0.000	158	0.100 [c]	0.004	0.000	0.000
Npct	160	1.833 [a]	0.030	0.498	0.000	153	1.815 [a]	0.031	0.232	0.000	160	1.677 [b]	0.026	0.227	0.000	158	1.668 [b]	0.028	0.284	0.000
nlvs	157	8.159 [a]	0.295	0.548	0.000	152	7.681 [a]	0.266	0.777	0.000	160	6.708 [b]	0.200	0.821	0.000	159	6.455 [b]	0.129	0.441	0.000
boltd	157	24.638 [a]	0.519	0.653	0.000	152	23.525 [a]	0.526	0.516	0.000	159	23.412 [a]	0.516	0.752	0.000	159	24.186 [a]	0.617	0.677	0.000
delco2	160	16.872 [a]	0.800	0.196	0.000	154	21.807 [b]	1.087	0.069	0.000	134	4.628 [c]	0.375	0.208	0.000	75	1.083 [d]	0.175	0.369	0.000
delh2o	160	5.752 [a]	0.201	0.116	0.000	154	6.685 [b]	0.202	0.190	0.000	134	2.017 [c]	0.081	0.255	0.000	75	1.226 [d]	0.041	0.263	0.000
fvfm	151	0.801 [a]	0.002	0.102	0.000	150	0.802 [a]	0.001	0.137	0.000	137	0.786 [b]	0.002	0.000	0.000	57	0.752 [c]	0.004	0.000	0.000
dietr	155	14.166 [a]	0.684	0.085	0.000	150	13.860 [a]	0.490	0.087	0.000	137	14.366 [a]	0.492	0.004	0.000	58	18.511 [b]	1.688	0.000	0.000
yield	154	0.737 [a]	0.003	0.000	0.000	150	0.731 [a]	0.003	0.123	0.000	136	0.700 [b]	0.008	0.296	0.000	55	0.640 [c]	0.015	0.341	0.000
totlen	160	777.059 [a]	39.094	0.512	0.000	156	758.843 [a]	40.272	0.531	0.000	160	212.757 [b]	12.074	0.458	0.000	160	67.075 [c]	4.865	0.299	0.000
rb	160	9.381 [a]	0.298	0.235	0.000	157	8.146 [b]	0.297	0.325	0.000	160	4.860 [c]	0.145	0.102	0.000	160	2.989 [d]	0.088	0.033	0.000
tsl	160	816.153 [a]	42.048	0.466	0.000	153	830.396 [a]	56.020	0.594	0.000	157	200.565 [b]	10.642	0.232	0.000	145	63.094 [c]	4.142	0.191	0.000

Table 3.2. Wilcoxon signed rank test for coefficients of genetic variation. Non-parametric paired t-test comparing CVg across pairs of nitrogen environments. P-values are listed as well as threshold alpha values using sequential Bonferroni corrections.

Coefficcient of Genetic Variation	p-value	threshold α
N56 vs N51	0.787	0.0500
N56 vs N06	0.008	0.0083
N56 vs N01	0.011	0.0100
N51 vs N06	0.011	0.0125
N51 vs N01	0.040	0.0167
N06 vs N01	0.204	0.0250

Table 3.3. Genetic correlations within each N-environment. Calculated as the correlations among RIL means within each environment. P-values correspond to a test of the null hypothesis of rg = 0. (b) N51 below diagonal\N56 above diagonal and (b) N01 below diagonal\N06 above diagonal.

Pearson Correlation Coefficients N51\N56 Prob > |r| under H0: Rho=0 Number of Observations													
	totg	**proprt**	**Npct**	**nlvs**	**boltd**	**delco2**	**delh2o**	**fvfm**	**dietr**	**yield**	**totlen**	**rb**	**tsl**
totg		-0.3657	-0.4374	0.6152	0.5634	0.5937	0.4215	0.2157	0.1005	0.0583	0.8807	0.4959	0.7999
		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0080	0.2150	0.4739	<.0001	<.0001	<.0001
		159	159	156	156	159	159	150	154	153	159	159	159
proprt	-0.3873		0.2337	-0.1250	-0.1123	-0.2102	-0.1414	0.0319	-0.0673	-0.1073	-0.2860	-0.1613	-0.2856
	<.0001		0.0030	0.1202	0.1628	0.0078	0.0755	0.6985	0.4069	0.1868	0.0003	0.0422	0.0003
	153		159	156	156	159	159	150	154	153	159	159	159
Npct	-0.3673	0.2902		0.0476	0.0542	-0.0216	-0.0117	0.0041	-0.0870	0.0275	-0.4937	-0.2366	-0.1713
	<.0001	0.0003		0.5539	0.5005	0.7860	0.8837	0.9606	0.2817	0.7347	<.0001	0.0026	0.0303
	150	150		157	157	160	160	151	155	154	160	160	160
nlvs	0.4728	-0.2417	-0.0275		0.8199	0.4924	0.2940	0.2488	0.1005	0.1229	0.3962	0.3050	0.5896
	<.0001	0.0032	0.7414		<.0001	<.0001	0.0002	0.0023	0.2179	0.1326	<.0001	0.0001	<.0001
	147	147	147		157	157	157	148	152	151	157	157	157
boltd	0.3974	-0.1668	0.1512	0.7960		0.4678	0.2778	0.2671	0.1267	0.1307	0.3620	0.2674	0.5596
	<.0001	0.0434	0.0665	<.0001		<.0001	0.0004	0.0010	0.1198	0.1098	<.0001	0.0007	<.0001
	147	147	148	151		157	157	148	152	151	157	157	157
delco2	0.5248	-0.2858	-0.1837	0.4444	0.3897		0.8779	0.2335	-0.0332	0.0831	0.5216	0.3257	0.5982
	<.0001	0.0004	0.0244	<.0001	<.0001		<.0001	0.0039	0.6821	0.3058	<.0001	<.0001	<.0001
	151	151	150	147	147		160	151	155	154	160	160	160
delh2o	0.5458	-0.2661	-0.2347	0.3720	0.2848	0.7704		0.1485	-0.1177	0.0783	0.3877	0.2707	0.4501
	<.0001	0.0010	0.0038	<.0001	0.0005	<.0001		0.0688	0.1446	0.3345	<.0001	0.0005	<.0001
	151	151	150	147	147	154		151	155	154	160	160	160
fvfm	0.2718	-0.1787	0.0572	0.3037	0.3619	0.2328	0.1720		0.0131	0.2176	0.2439	0.1721	0.2203
	0.0009	0.0303	0.4932	0.0002	<.0001	0.0043	0.0360		0.8735	0.0075	0.0025	0.0346	0.0066
	147	147	146	143	143	149	149		151	150	151	151	151
dietr	-0.0028	-0.0396	-0.0695	0.1469	0.0614	0.0261	0.0708	-0.1225		-0.1643	0.0293	-0.1145	0.0951
	0.9734	0.6336	0.4049	0.0801	0.4665	0.7522	0.3909	0.1353		0.0417	0.7175	0.1560	0.2391
	147	147	146	143	143	149	149	150		154	155	155	155
yield	0.0310	0.0261	0.0167	0.0263	0.1082	0.0883	-0.0570	0.1361	-0.2475		0.1149	0.1157	0.0719
	0.7094	0.7539	0.8411	0.7557	0.1983	0.2845	0.4897	0.0969	0.0023		0.1559	0.1530	0.3759
	147	147	146	143	143	149	149	150	150		154	154	154
totlen	0.8654	-0.3666	-0.4516	0.4755	0.3463	0.4799	0.5019	0.3052	-0.0369	0.1435		0.5753	0.7084
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0002	0.6555	0.0808		<.0001	<.0001
	153	153	153	150	150	153	153	149	149	149		160	160
rb	0.5157	-0.3612	-0.1840	0.4511	0.3227	0.4327	0.4265	0.1855	0.0024	0.1180	0.5896		0.4699
	<.0001	<.0001	0.0228	<.0001	<.0001	<.0001	<.0001	0.0231	0.9772	0.1505	<.0001		<.0001
	153	153	153	150	150	154	154	150	150	150	156		160
tsl	0.6409	-0.3360	-0.2483	0.6874	0.5757	0.4747	0.4499	0.3120	0.0208	0.1165	0.7338	0.5521	
	<.0001	<.0001	0.0021	<.0001	<.0001	<.0001	<.0001	0.0001	0.8027	0.1601	<.0001	<.0001	
	150	150	151	147	147	150	150	147	147	147	153	153	

Table 3.3 continued.

Pearson Correlation Coefficients N01\N06 Prob > |r| under H0: Rho=0 Number of Observations													
	totg	**proprt**	**Npct**	**nlvs**	**boltd**	**delco2**	**delh2o**	**fvfm**	**dietr**	**yield**	**totlen**	**rb**	**tsl**
totg		-0.3369 <.0001 156	-0.3489 <.0001 157	0.4705 <.0001 157	0.4201 <.0001 156	0.4924 <.0001 133	0.4190 <.0001 133	0.1880 0.0290 135	0.0816 0.3469 135	0.1925 0.0259 134	0.6144 <.0001 157	0.4181 <.0001 157	0.5778 <.0001 154
proprt	-0.1024 0.2006 158		0.3256 <.0001 156	-0.0387 0.6312 156	-0.0581 0.4724 155	-0.2312 0.0077 132	-0.2786 0.0012 132	0.1184 0.1729 134	0.1266 0.1450 134	-0.0437 0.6173 133	-0.3734 <.0001 156	-0.3412 <.0001 156	-0.3007 0.0002 153
Npct	-0.3045 0.0001 157	0.1692 0.0342 157		-0.0289 0.7167 160	-0.0165 0.8362 159	-0.1306 0.1326 134	-0.1470 0.0900 134	0.0741 0.3896 137	0.0096 0.9116 137	-0.0332 0.7013 136	-0.5522 <.0001 160	-0.2858 0.0002 160	-0.3440 <.0001 157
nlvs	0.4249 <.0001 157	0.1341 0.0941 157	-0.0747 0.3527 157		0.8078 <.0001 159	0.5635 <.0001 134	0.3118 0.0002 134	0.1960 0.0217 137	0.1669 0.0513 137	0.1323 0.1247 136	0.2665 0.0007 160	0.2120 0.0071 160	0.4154 <.0001 157
boltd	0.3297 <.0001 157	0.1042 0.1942 157	-0.0006 0.9942 157	0.7351 <.0001 158		0.3350 <.0001 133	0.1813 0.0368 133	0.1686 0.0497 136	0.2495 0.0034 136	0.1231 0.1548 135	0.2944 0.0002 159	0.1161 0.1449 159	0.4938 <.0001 156
delco2	0.0566 0.6297 75	-0.0063 0.9573 75	-0.1692 0.1496 74	-0.0373 0.7510 75	-0.1179 0.3139 75		0.8297 <.0001 134	0.1088 0.2410 118	0.0199 0.8310 118	0.1549 0.0954 117	0.3677 <.0001 134	0.3206 0.0002 134	0.4294 <.0001 132
delh2o	0.1397 0.2319 75	-0.1213 0.2998 75	-0.3850 0.0007 74	-0.0678 0.5633 75	-0.1861 0.1100 75	0.6615 <.0001 75		0.0828 0.3726 118	0.0574 0.5370 118	0.1084 0.2445 117	0.3699 <.0001 134	0.4061 <.0001 134	0.4460 <.0001 132
fvfm	0.0422 0.7555 57	0.2178 0.1037 57	0.1456 0.2800 57	0.3538 0.0069 57	0.2606 0.0503 57	-0.1241 0.3577 57	-0.2630 0.0481 57		0.0402 0.6406 137	0.0267 0.7578 136	0.2018 0.0181 137	0.2418 0.0044 137	0.2555 0.0028 135
dietr	0.0285 0.8316 58	0.2308 0.0814 58	0.0480 0.7204 58	0.0242 0.8571 58	0.0081 0.9522 58	0.2062 0.1205 58	0.0965 0.4713 58	-0.0008 0.9953 57		0.0127 0.8830 136	0.0223 0.7957 137	-0.0209 0.8085 137	0.0562 0.5176 135
yield	0.2044 0.1343 55	0.0752 0.5852 55	0.0057 0.9673 55	0.2201 0.1063 55	0.2551 0.0602 55	-0.0472 0.7324 55	-0.1708 0.2125 55	0.2715 0.0470 54	0.1818 0.1842 55		0.1849 0.0312 136	0.1456 0.0907 136	0.1768 0.0410 134
totlen	0.8477 <.0001 158	-0.1778 0.0254 158	-0.4914 <.0001 158	0.2734 0.0005 159	0.1901 0.0164 159	0.1363 0.2437 75	0.2167 0.0619 75	-0.1233 0.3607 57	0.0163 0.9036 58	0.1536 0.2630 55		0.5863 <.0001 160	0.7639 <.0001 157
rb	0.2693 0.0006 158	-0.1630 0.0407 158	-0.0767 0.3381 158	-0.0561 0.4821 159	-0.0322 0.6871 159	-0.0010 0.9931 75	0.2483 0.0317 75	-0.0027 0.9843 57	-0.0133 0.9212 58	-0.1861 0.1737 55	0.4207 <.0001 160		0.4693 <.0001 157
tsl	0.7967 <.0001 144	-0.0677 0.4204 144	-0.3551 <.0001 143	0.4440 <.0001 145	0.3801 <.0001 144	0.1625 0.1667 74	0.1430 0.2243 74	-0.0541 0.6922 56	0.1241 0.3578 57	0.3413 0.0116 54	0.8409 <.0001 145	0.2714 0.0010 145	

Table 3.4. 95% parametric confidence intervals on genetic correlations. See methods for details of CI estimation and trait abbreviations. Table continued on next two pages.

		N56			N51			N06			N01		
Var	With Var	r_g	95% CL		r_g	95% CL		r_g	95% CL		r_g	95% CL	
totg	proprt	-0.37	-0.49	-0.22	-0.39	-0.51	-0.24	-0.34	-0.47	-0.19	-0.10	-0.25	0.05
totg	Npct	-0.44	-0.55	-0.30	-0.37	-0.50	-0.22	-0.35	-0.48	-0.20	-0.30	-0.44	-0.15
totg	nlvs	0.62	0.51	0.70	0.47	0.34	0.59	0.47	0.34	0.58	0.42	0.29	0.54
totg	boltd	0.56	0.44	0.66	0.40	0.25	0.52	0.42	0.28	0.54	0.33	0.18	0.46
totg	delco2	0.59	0.48	0.68	0.52	0.40	0.63	0.49	0.35	0.61	0.06	-0.17	0.28
totg	delh2o	0.42	0.28	0.54	0.55	0.42	0.65	0.42	0.27	0.55	0.14	-0.09	0.35
totg	fvfm	0.22	0.06	0.36	0.27	0.11	0.41	0.19	0.02	0.35	0.04	-0.22	0.30
totg	dietr	0.10	-0.06	0.25	0.00	-0.16	0.16	0.08	-0.09	0.25	0.03	-0.23	0.28
totg	yield	0.06	-0.10	0.21	0.03	-0.13	0.19	0.19	0.02	0.35	0.20	-0.07	0.44
totg	totlen	0.88	0.84	0.91	0.87	0.82	0.90	0.61	0.50	0.70	0.85	0.80	0.89
totg	rb	0.50	0.37	0.60	0.52	0.39	0.62	0.42	0.28	0.54	0.27	0.12	0.41
totg	tsl	0.80	0.73	0.85	0.64	0.53	0.73	0.58	0.46	0.67	0.80	0.73	0.85
proprt	Npct	0.23	0.08	0.38	0.29	0.14	0.43	0.33	0.18	0.46	0.17	0.01	0.32
proprt	nlvs	-0.12	-0.28	0.03	-0.24	-0.39	-0.08	-0.04	-0.19	0.12	0.13	-0.02	0.28
proprt	boltd	-0.11	-0.26	0.05	-0.17	-0.32	0.00	-0.06	-0.21	0.10	0.10	-0.05	0.26
proprt	delco2	-0.21	-0.35	-0.06	-0.29	-0.43	-0.13	-0.23	-0.39	-0.06	-0.01	-0.23	0.22
proprt	delh2o	-0.14	-0.29	0.02	-0.27	-0.41	-0.11	-0.28	-0.43	-0.11	-0.12	-0.34	0.11
proprt	fvfm	0.03	-0.13	0.19	-0.18	-0.33	-0.02	0.12	-0.05	0.28	0.22	-0.05	0.45
proprt	dietr	-0.07	-0.22	0.09	-0.04	-0.20	0.12	0.13	-0.04	0.29	0.23	-0.03	0.46
proprt	yield	-0.11	-0.26	0.05	0.03	-0.14	0.19	-0.04	-0.21	0.13	0.08	-0.19	0.33
proprt	totlen	-0.29	-0.42	-0.14	-0.37	-0.50	-0.22	-0.37	-0.50	-0.23	-0.18	-0.32	-0.02
proprt	rb	-0.16	-0.31	-0.01	-0.36	-0.49	-0.21	-0.34	-0.47	-0.19	-0.16	-0.31	-0.01
proprt	tsl	-0.29	-0.42	-0.14	-0.34	-0.47	-0.18	-0.30	-0.44	-0.15	-0.07	-0.23	0.10
Npct	nlvs	0.05	-0.11	0.20	-0.03	-0.19	0.14	-0.03	-0.18	0.13	-0.07	-0.23	0.08
Npct	boltd	0.05	-0.10	0.21	0.15	-0.01	0.30	-0.02	-0.17	0.14	0.00	-0.16	0.16
Npct	delco2	-0.02	-0.18	0.13	-0.18	-0.33	-0.02	-0.13	-0.29	0.04	-0.17	-0.38	0.06
Npct	delh2o	-0.01	-0.17	0.14	-0.23	-0.38	-0.08	-0.15	-0.31	0.02	-0.38	-0.56	-0.17
Npct	fvfm	0.00	-0.16	0.16	0.06	-0.11	0.22	0.07	-0.10	0.24	0.15	-0.12	0.39
Npct	dietr	-0.09	-0.24	0.07	-0.07	-0.23	0.09	0.01	-0.16	0.18	0.05	-0.21	0.30
Npct	yield	0.03	-0.13	0.18	0.02	-0.15	0.18	-0.03	-0.20	0.14	0.01	-0.26	0.27
Npct	totlen	-0.49	-0.60	-0.37	-0.45	-0.57	-0.31	-0.55	-0.65	-0.43	-0.49	-0.60	-0.36
Npct	rb	-0.24	-0.38	-0.08	-0.18	-0.33	-0.03	-0.29	-0.42	-0.14	-0.08	-0.23	0.08
Npct	tsl	-0.17	-0.32	-0.02	-0.25	-0.39	-0.09	-0.34	-0.47	-0.20	-0.36	-0.49	-0.20
nlvs	boltd	0.82	0.76	0.86	0.80	0.73	0.85	0.81	0.74	0.86	0.74	0.65	0.80
nlvs	delco2	0.49	0.36	0.60	0.44	0.30	0.56	0.56	0.43	0.67	-0.04	-0.26	0.19
nlvs	delh2o	0.29	0.14	0.43	0.37	0.22	0.50	0.31	0.15	0.46	-0.07	-0.29	0.16
nlvs	fvfm	0.25	0.09	0.39	0.30	0.15	0.44	0.20	0.03	0.35	0.35	0.10	0.56
nlvs	dietr	0.10	-0.06	0.26	0.15	-0.02	0.30	0.17	0.00	0.32	0.02	-0.24	0.28
nlvs	yield	0.12	-0.04	0.28	0.03	-0.14	0.19	0.13	-0.04	0.29	0.22	-0.05	0.46
nlvs	totlen	0.40	0.25	0.52	0.48	0.34	0.59	0.27	0.12	0.40	0.27	0.12	0.41
nlvs	rb	0.31	0.15	0.44	0.45	0.31	0.57	0.21	0.06	0.35	-0.06	-0.21	0.10
nlvs	tsl	0.59	0.48	0.68	0.69	0.59	0.76	0.42	0.28	0.54	0.44	0.30	0.57

Table 3.4 continued.

		N56			N51			N06			N01		
Var	With Var	r_g	95% CL		r_g	95% CL		r_g	95% CL		r_g	95% CL	
boltd	delco2	0.47	0.33	0.58	0.39	0.24	0.52	0.33	0.17	0.48	-0.12	-0.34	0.11
boltd	delh2o	0.28	0.13	0.42	0.28	0.13	0.43	0.18	0.01	0.34	-0.19	-0.40	0.04
boltd	fvfm	0.27	0.11	0.41	0.36	0.21	0.50	0.17	0.00	0.33	0.26	0.00	0.49
boltd	dietr	0.13	-0.03	0.28	0.06	-0.10	0.22	0.25	0.08	0.40	0.01	-0.25	0.27
boltd	yield	0.13	-0.03	0.28	0.11	-0.06	0.27	0.12	-0.05	0.29	0.26	-0.01	0.49
boltd	totlen	0.36	0.22	0.49	0.35	0.20	0.48	0.29	0.14	0.43	0.19	0.03	0.34
boltd	rb	0.27	0.11	0.41	0.32	0.17	0.46	0.12	-0.04	0.27	-0.03	-0.19	0.12
boltd	tsl	0.56	0.44	0.66	0.58	0.45	0.67	0.49	0.36	0.60	0.38	0.23	0.51
delco2	delh2o	0.88	0.84	0.91	0.77	0.70	0.83	0.83	0.77	0.88	0.66	0.51	0.77
delco2	fvfm	0.23	0.08	0.38	0.23	0.07	0.38	0.11	-0.07	0.28	-0.12	-0.37	0.14
delco2	dietr	-0.03	-0.19	0.13	0.03	-0.14	0.19	0.02	-0.16	0.20	0.21	-0.06	0.44
delco2	yield	0.08	-0.08	0.24	0.09	-0.07	0.25	0.15	-0.03	0.33	-0.05	-0.31	0.22
delco2	totlen	0.52	0.40	0.63	0.48	0.35	0.59	0.37	0.21	0.50	0.14	-0.09	0.35
delco2	rb	0.33	0.18	0.46	0.43	0.29	0.55	0.32	0.16	0.46	0.00	-0.23	0.23
delco2	tsl	0.60	0.49	0.69	0.47	0.34	0.59	0.43	0.28	0.56	0.16	-0.07	0.38
delh2o	fvfm	0.15	-0.01	0.30	0.17	0.01	0.32	0.08	-0.10	0.26	-0.26	-0.49	0.00
delh2o	dietr	-0.12	-0.27	0.04	0.07	-0.09	0.23	0.06	-0.12	0.24	0.10	-0.17	0.35
delh2o	yield	0.08	-0.08	0.23	-0.06	-0.22	0.10	0.11	-0.08	0.28	-0.17	-0.42	0.10
delh2o	totlen	0.39	0.25	0.51	0.50	0.37	0.61	0.37	0.21	0.51	0.22	-0.01	0.42
delh2o	rb	0.27	0.12	0.41	0.43	0.29	0.55	0.41	0.25	0.54	0.25	0.02	0.45
delh2o	tsl	0.45	0.32	0.56	0.45	0.31	0.57	0.45	0.30	0.57	0.14	-0.09	0.36
fvfm	dietr	0.01	-0.15	0.17	-0.12	-0.28	0.04	0.04	-0.13	0.21	0.00	-0.26	0.26
fvfm	yield	0.22	0.06	0.36	0.14	-0.03	0.29	0.03	-0.14	0.19	0.27	0.00	0.50
fvfm	totlen	0.24	0.09	0.39	0.31	0.15	0.44	0.20	0.03	0.36	-0.12	-0.37	0.14
fvfm	rb	0.17	0.01	0.32	0.19	0.03	0.34	0.24	0.08	0.39	0.00	-0.26	0.26
fvfm	tsl	0.22	0.06	0.37	0.31	0.16	0.45	0.26	0.09	0.41	-0.05	-0.31	0.21
dietr	yield	-0.16	-0.31	-0.01	-0.25	-0.39	-0.09	0.01	-0.16	0.18	0.18	-0.09	0.43
dietr	totlen	0.03	-0.13	0.19	-0.04	-0.20	0.12	0.02	-0.15	0.19	0.02	-0.24	0.27
dietr	rb	-0.11	-0.27	0.04	0.00	-0.16	0.16	-0.02	-0.19	0.15	-0.01	-0.27	0.25
dietr	tsl	0.10	-0.06	0.25	0.02	-0.14	0.18	0.06	-0.11	0.22	0.12	-0.14	0.37
yield	totlen	0.11	-0.04	0.27	0.14	-0.02	0.30	0.18	0.02	0.34	0.15	-0.12	0.40
yield	rb	0.12	-0.04	0.27	0.12	-0.04	0.27	0.15	-0.02	0.31	-0.19	-0.43	0.08
yield	tsl	0.07	-0.09	0.23	0.12	-0.05	0.27	0.18	0.01	0.34	0.34	0.08	0.56
totlen	rb	0.58	0.46	0.67	0.59	0.48	0.68	0.59	0.47	0.68	0.42	0.28	0.54
totlen	tsl	0.71	0.62	0.78	0.73	0.65	0.80	0.76	0.69	0.82	0.84	0.78	0.88
rb	tsl	0.47	0.34	0.58	0.55	0.43	0.65	0.47	0.34	0.58	0.27	0.11	0.42
boltd	delco2	0.47	0.33	0.58	0.39	0.24	0.52	0.33	0.17	0.48	-0.12	-0.34	0.11
boltd	delh2o	0.28	0.13	0.42	0.28	0.13	0.43	0.18	0.01	0.34	-0.19	-0.40	0.04
boltd	fvfm	0.27	0.11	0.41	0.36	0.21	0.50	0.17	0.00	0.33	0.26	0.00	0.49
boltd	dietr	0.13	-0.03	0.28	0.06	-0.10	0.22	0.25	0.08	0.40	0.01	-0.25	0.27
boltd	yield	0.13	-0.03	0.28	0.11	-0.06	0.27	0.12	-0.05	0.29	0.26	-0.01	0.49
boltd	totlen	0.36	0.22	0.49	0.35	0.20	0.48	0.29	0.14	0.43	0.19	0.03	0.34

Table 3.4 continued.

		N56			N51			N06			N01		
Var	With Var	r_g	95% CL		r_g	95% CL		r_g	95% CL		r_g	95% CL	
boltd	rb	0.27	0.11	0.41	0.32	0.17	0.46	0.12	-0.04	0.27	-0.03	-0.19	0.12
boltd	tsl	0.56	0.44	0.66	0.58	0.45	0.67	0.49	0.36	0.60	0.38	0.23	0.51
delco2	delh2o	0.88	0.84	0.91	0.77	0.70	0.83	0.83	0.77	0.88	0.66	0.51	0.77
delco2	fvfm	0.23	0.08	0.38	0.23	0.07	0.38	0.11	-0.07	0.28	-0.12	-0.37	0.14
delco2	dietr	-0.03	-0.19	0.13	0.03	-0.14	0.19	0.02	-0.16	0.20	0.21	-0.06	0.44
delco2	yield	0.08	-0.08	0.24	0.09	-0.07	0.25	0.15	-0.03	0.33	-0.05	-0.31	0.22
delco2	totlen	0.52	0.40	0.63	0.48	0.35	0.59	0.37	0.21	0.50	0.14	-0.09	0.35
delco2	rb	0.33	0.18	0.46	0.43	0.29	0.55	0.32	0.16	0.46	0.00	-0.23	0.23
delco2	tsl	0.60	0.49	0.69	0.47	0.34	0.59	0.43	0.28	0.56	0.16	-0.07	0.38
delh2o	fvfm	0.15	-0.01	0.30	0.17	0.01	0.32	0.08	-0.10	0.26	-0.26	-0.49	0.00
delh2o	dietr	-0.12	-0.27	0.04	0.07	-0.09	0.23	0.06	-0.12	0.24	0.10	-0.17	0.35
delh2o	yield	0.08	-0.08	0.23	-0.06	-0.22	0.10	0.11	-0.08	0.28	-0.17	-0.42	0.10
delh2o	totlen	0.39	0.25	0.51	0.50	0.37	0.61	0.37	0.21	0.51	0.22	-0.01	0.42
delh2o	rb	0.27	0.12	0.41	0.43	0.29	0.55	0.41	0.25	0.54	0.25	0.02	0.45
delh2o	tsl	0.45	0.32	0.56	0.45	0.31	0.57	0.45	0.30	0.57	0.14	-0.09	0.36
fvfm	dietr	0.01	-0.15	0.17	-0.12	-0.28	0.04	0.04	-0.13	0.21	0.00	-0.26	0.26
fvfm	yield	0.22	0.06	0.36	0.14	-0.03	0.29	0.03	-0.14	0.19	0.27	0.00	0.50
fvfm	totlen	0.24	0.09	0.39	0.31	0.15	0.44	0.20	0.03	0.36	-0.12	-0.37	0.14
fvfm	rb	0.17	0.01	0.32	0.19	0.03	0.34	0.24	0.08	0.39	0.00	-0.26	0.26
fvfm	tsl	0.22	0.06	0.37	0.31	0.16	0.45	0.26	0.09	0.41	-0.05	-0.31	0.21
dietr	yield	-0.16	-0.31	-0.01	-0.25	-0.39	-0.09	0.01	-0.16	0.18	0.18	-0.09	0.43
dietr	totlen	0.03	-0.13	0.19	-0.04	-0.20	0.12	0.02	-0.15	0.19	0.02	-0.24	0.27
dietr	rb	-0.11	-0.27	0.04	0.00	-0.16	0.16	-0.02	-0.19	0.15	-0.01	-0.27	0.25
dietr	tsl	0.10	-0.06	0.25	0.02	-0.14	0.18	0.06	-0.11	0.22	0.12	-0.14	0.37
yield	totlen	0.11	-0.04	0.27	0.14	-0.02	0.30	0.18	0.02	0.34	0.15	-0.12	0.40
yield	rb	0.12	-0.04	0.27	0.12	-0.04	0.27	0.15	-0.02	0.31	-0.19	-0.43	0.08
yield	tsl	0.07	-0.09	0.23	0.12	-0.05	0.27	0.18	0.01	0.34	0.34	0.08	0.56
totlen	rb	0.58	0.46	0.67	0.59	0.48	0.68	0.59	0.47	0.68	0.42	0.28	0.54
totlen	tsl	0.71	0.62	0.78	0.73	0.65	0.80	0.76	0.69	0.82	0.84	0.78	0.88
rb	tsl	0.47	0.34	0.58	0.55	0.43	0.65	0.47	0.34	0.58	0.27	0.11	0.42

Table 3.5. Cross-environment genetic correlations. Correlations of RIL means for a single trait between pairs of nitrogen environments were calculated. Asterisks indicate significance level of cross- environment genetic correlations: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Trait	N56 vs N51	N56 vs. N6	N56 vs. N1	N51 Vs. N6	N51 vs. N1	N6 vs. N1
Total Biomass	0.82***	0.52***	0.52***	0.53***	0.52***	0.42***
% Root Mass	0.22**	0.13	0.02	0.27***	0.02	0.08
% Nitrogen	0.56***	0.48***	0.50***	0.41***	0.39***	0.57***
# R. Leaves	0.74***	0.55***	0.69***	0.81***	0.72***	0.75***
Bolting Day	0.71***	0.76***	0.69***	0.77***	0.76***	0.75***
C-assimilation	0.46***	0.35***	0.07	0.39***	-0.15	0.04
Transpiration	0.44	0.27**	-0.07	0.40***	0.03	0.25*
PS Q. Efficiency	0.22**	0.11	0.20	0.15	0.14	0.21
PS e- transport	0.13	0.17	-0.05	0.11	0.19	0.08
PS Q. Yield	0.17*	-0.06	-0.05	0.08	0.09	-0.01
Total Branch Length	0.86***	0.73***	0.61***	0.74***	0.61***	0.69***
Basal Branches	0.42***	0.26***	0.11	0.27***	0.10	0.38***
Tot. Fruit Leng	0.58***	0.53***	0.35***	0.52***	0.51***	0.47***

Table 3.6. Principle components. Eigenvalues followed by loadings of the first three PC axes within each N-environment are shown in columns. Above the PC axes information are the effective number of dimensions (n_D, Kirkpatrick 2009) and the number of significant PC axes (n_R) determined by 10 000 randomizations, with significance evaluated at $\alpha = 0.01$.

	N56			N51			N06			N01		
nD	2.78			2.67			3.03			4.18		
nR	2			1			2			3		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
Eigenvalue	4.31	1.64	1.35	4.49	1.46	1.33	3.96	1.85	1.19	2.87	2.49	1.63
totg	0.44	-0.13	0.14	0.39	-0.19	-0.06	0.45	-0.01	-0.20	0.20	0.49	-0.10
proprt	-0.23	0.36	-0.04	-0.23	0.18	0.14	-0.18	0.39	-0.03	0.37	-0.17	0.35
Npct	-0.18	0.56	-0.09	-0.16	0.53	0.30	-0.18	0.41	0.43	0.14	-0.41	0.06
nlvs	0.33	0.36	0.29	0.33	0.44	0.03	0.35	0.41	-0.05	0.48	0.20	-0.02
boltd	0.31	0.41	0.30	0.32	0.42	0.16	0.32	0.47	-0.14	0.52	0.10	-0.03
delco2	0.37	0.16	-0.28	0.36	0.07	0.01	0.38	-0.03	0.51	-0.08	0.20	0.59
delh2o	0.30	0.09	-0.43	0.36	0.03	-0.12	0.33	-0.12	0.58	-0.22	0.34	0.41
fvfm	0.16	0.28	-0.06	0.21	0.18	0.27	0.10	0.24	-0.01	0.30	-0.07	-0.20
dietr	0.02	0.00	0.64	0.04	0.30	-0.61	0.07	0.34	-0.19	0.07	0.01	0.44
yield	0.07	0.19	-0.33	0.05	-0.20	0.63	0.15	-0.02	-0.11	0.26	0.13	-0.06
totlen	0.41	-0.24	0.00	0.38	-0.30	0.02	0.39	-0.23	-0.33	0.01	0.54	-0.13
rb	0.30	-0.18	-0.10	0.32	-0.13	0.03	0.27	-0.21	0.02	-0.30	0.18	-0.31

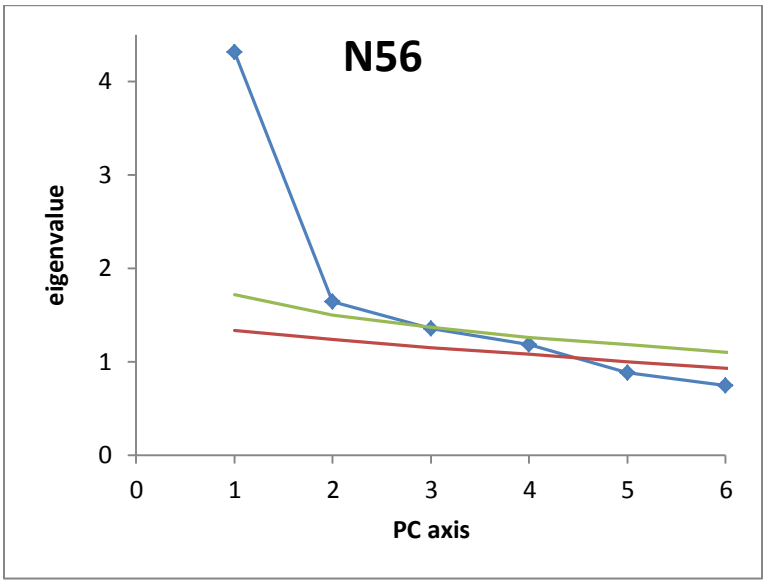
Table 3.7. Vector correlations between the first two PC axes across N-environments.

	PC1_N56	PC1_N51	PC1_N06	PC1_N01	PC2_N56	PC2_N51	PC2_N06	PC2_N01	PC3_N01
PC1_N56	1.00	1.00	0.99	0.19	0.00	-0.05	-0.04	0.88	0.02
PC1_N51		1.00	0.98	0.18	0.04	0.00	-0.01	0.85	0.04
PC1_N06			1.00	0.23	0.03	-0.03	0.00	0.89	0.09
PC1_N01				1.00	0.72	0.58	0.82	0.00	0.00
PC2_N56					1.00	0.84	0.86	-0.34	0.31
PC2_N51						1.00	0.91	-0.41	0.34
PC2_N06							1.00	-0.33	0.27
PC2_N01								1.00	0.00

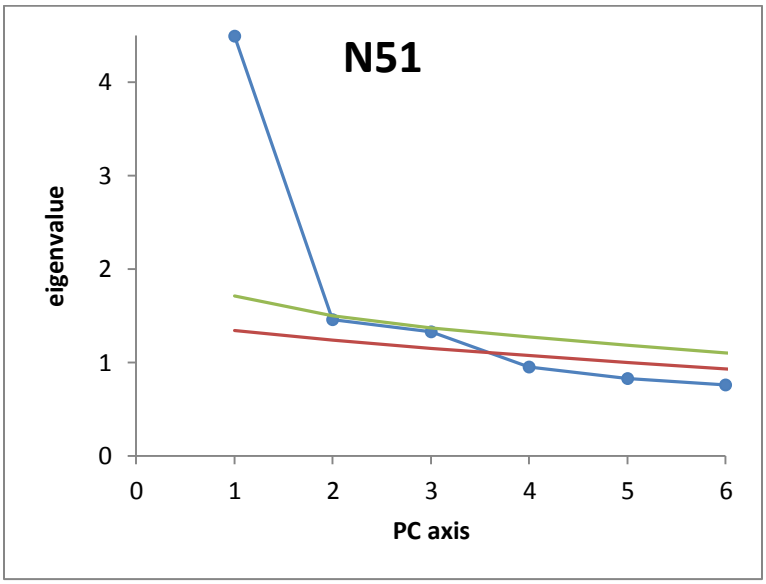
Table 3.8. Transgressive segregation. The table below shows which percentile the Cvi and Ler parents fall into within the distribution of genotypic means for each trait within each N-environment. Values lower than 99 or greater than 1 indicate transgressive segregation, i.e. the range of RIL means is beyond the range of parental means. Transgressive segregation was apparent for all traits in all environments for which we had data. N/A indicates missing data for the parental lines prohibited calculating percentile rank.

	Totg	Proprt	Npct	Nlvs	Boltd	Delco2	Delh2o	Fvfm	Dietr	Yield	Totlen	Rb	Tsl
CVI - N56	0.78	0.27	0.36	0.84	0.89	0.82	0.81	0.46	0.47	0.71	0.77	0.70	0.81
LER - N56	0.13	0.57	0.92	0.18	0.99	0.10	0.06	0.29	0.66	0.65	0.06	0.03	0.15
CVI - N51	0.68	0.38	0.32	0.84	0.83	0.80	0.87	0.97	0.19	0.61	0.63	0.72	0.82
LER - N51	0.55	0.18	0.49	0.46	0.83	0.48	0.38	0.49	0.32	0.58	0.43	0.37	0.41
CVI - N06	0.49	0.81	0.45	0.84	0.67	0.74	0.84	0.89	0.79	0.56	0.56	0.34	0.46
LER - N06	0.21	0.79	0.91	0.16	0.78	0.50	0.54	0.19	0.40	0.39	0.17	0.14	0.42
CVI - N01	0.77	0.48	0.33	0.86	0.80	0.25	N/A	N/A	N/A	N/A	0.70	0.43	0.65
LER - N01	0.26	0.35	0.71	0.63	0.50	0.25	N/A	N/A	N/A	N/A	0.19	0.62	0.08

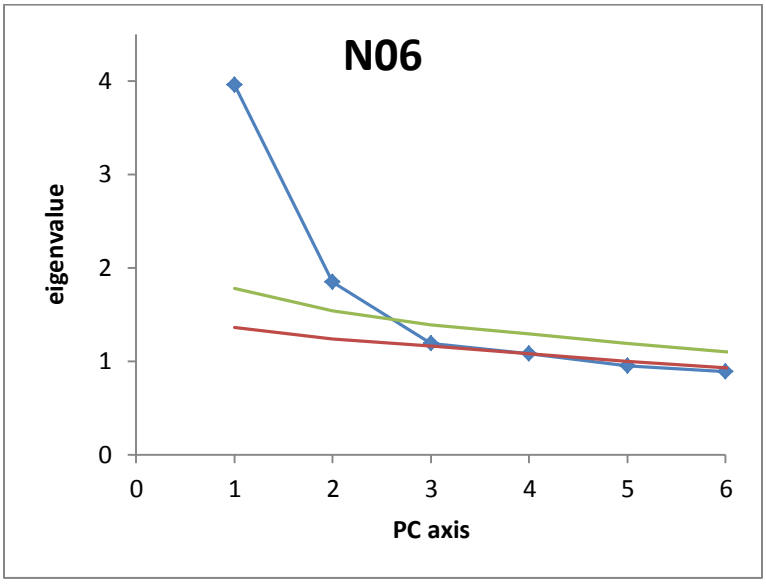
a)


N56
eigenvalue
PC axis
0
1
2
3
4
5
6

b)


N51
eigenvalue
PC axis
0
1
2
3
4
5
6

c)



d)

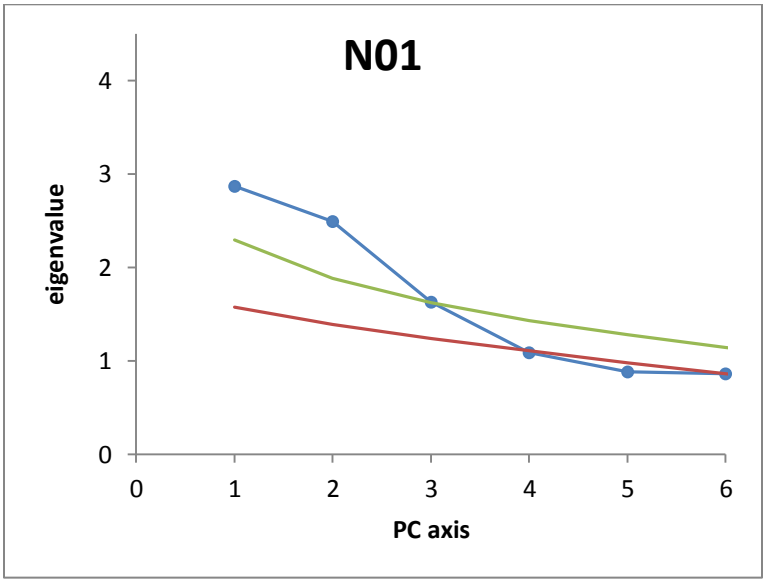


Figure 3.1. Randomization to determine number of significant principle component axes. A SAS macro-program (RandPC) was used to generate 10 000 resampled populations per environment with trait values shuffled relative to each other. These produced a null distribution of eigenvalues for each principle component axis. The 99.5th and 0.5th percentiles of this null distribution are shown in green and red respectively. When eigenvalues of the original data (shown in blue) were greater than 99% of the null distribution (i.e. above the 99.5th%tile), we concluded that the corresponding PC axis was significant at the 99% level. (a) N56 (b) N51 (c) N06 (d) N01.

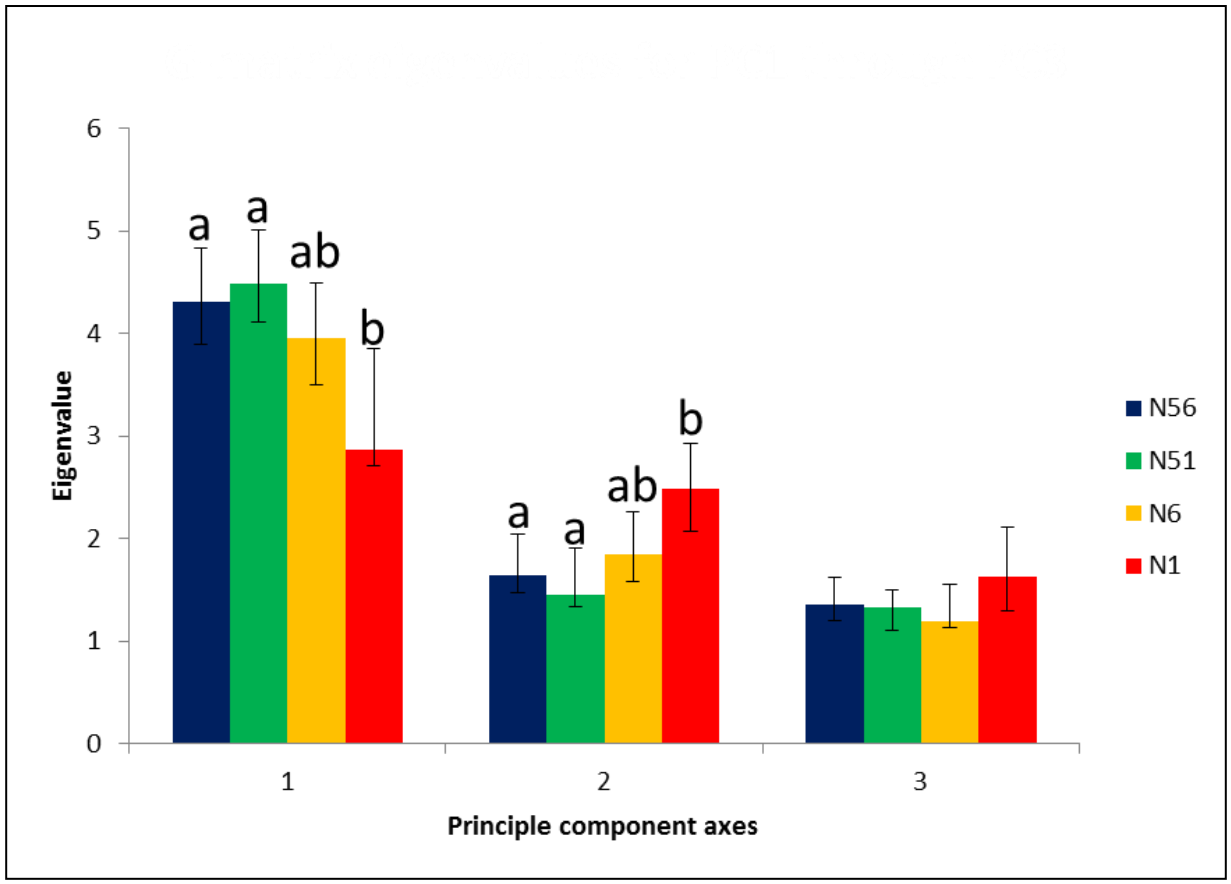


Figure 3.2. Eigenvalues for PC1-PC3 axes within each N-environment. Bootstrapped 95% confidence intervals were estimated using BootPCA, a SAS macro created by SJT. Letters showing significant differences among eigenvalues apply only across N-environments within the same numbered PC axis. No letters were added to PC axis 3 because it only explained a significant amount of trait variation in N01.

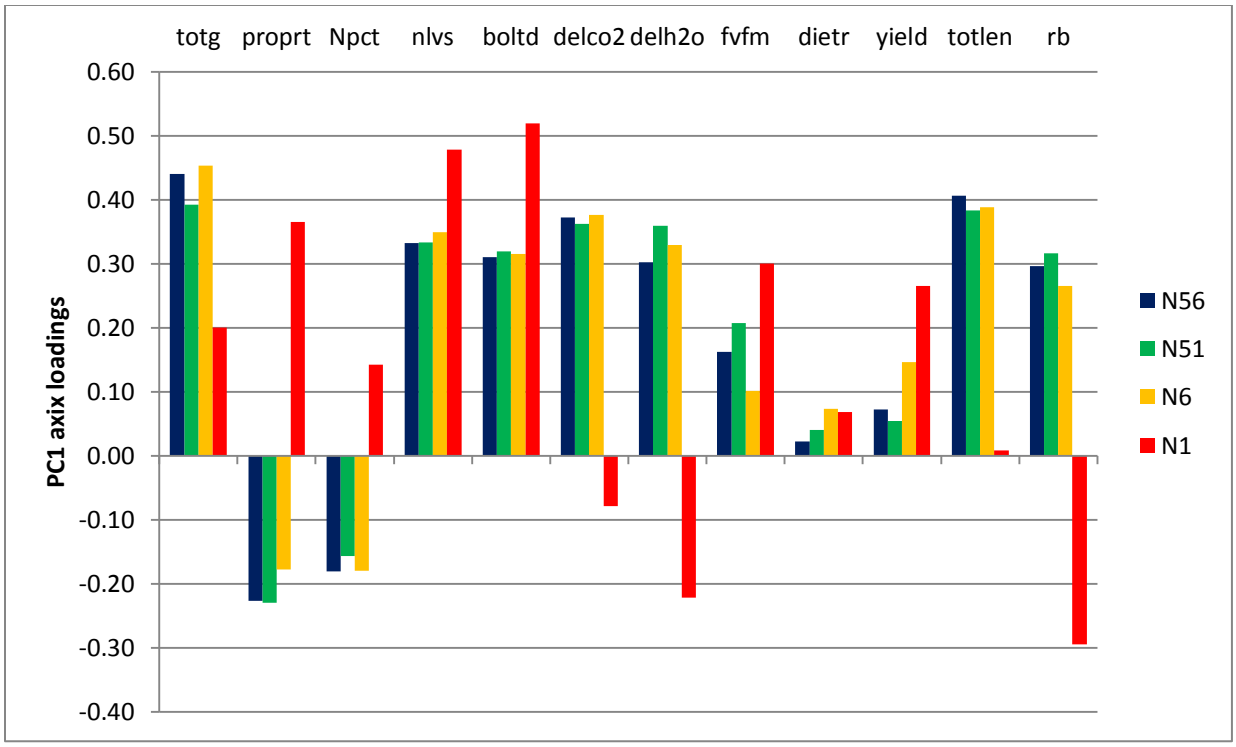


Figure 3.3. Principle component 1 axis loadings across N-environments. See methods for explanation of trait names.

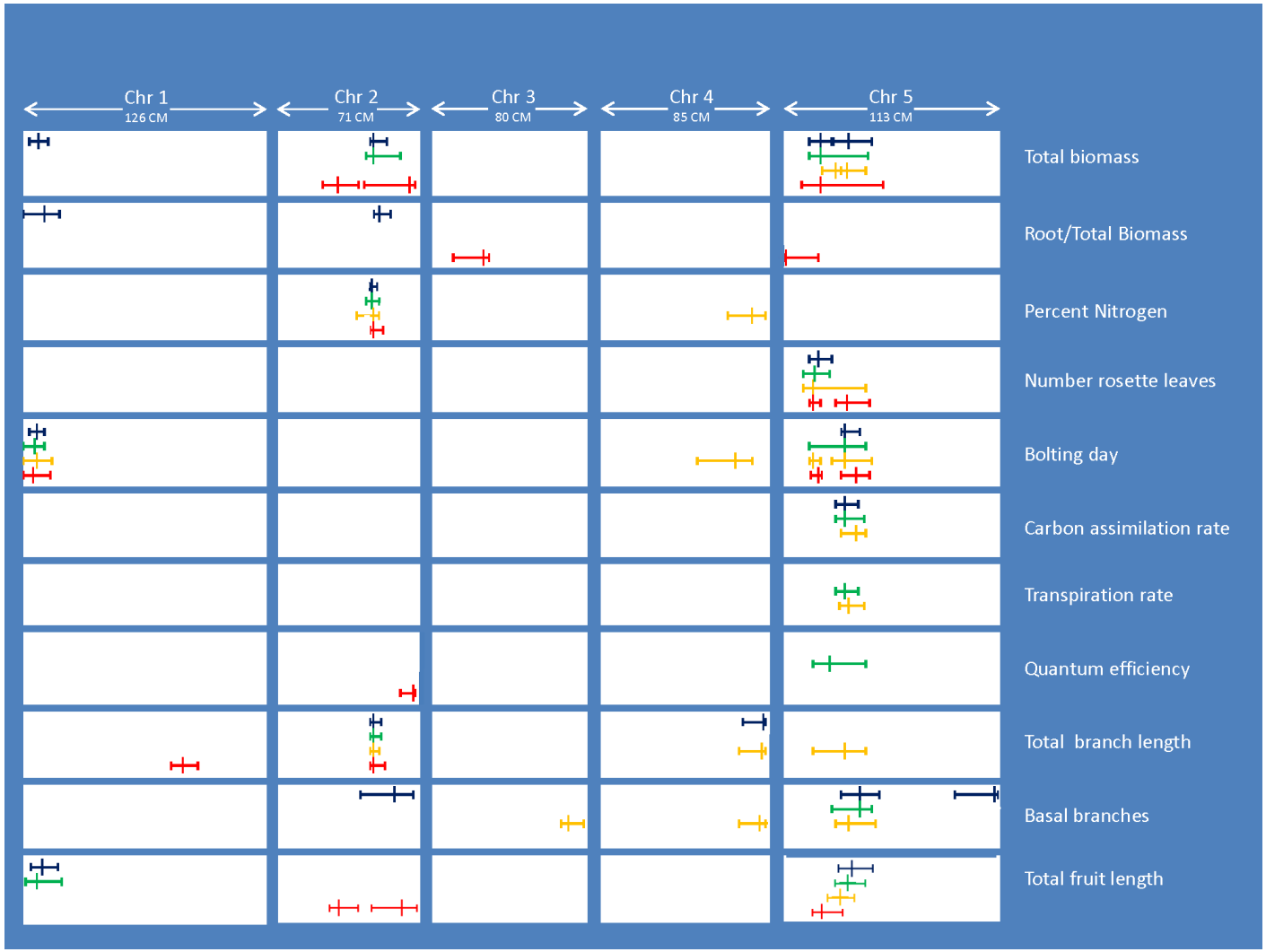


Figure 3.4. Additive QTL locations. QTL for each trait are color coded based on the N-environment in which they were detected: N56 (dark blue), N51 (green), N06 (orange), N01 (red). Error bars around QTL locations indicate 1.5 LOD support intervals. See Methods for details on Multiple Impuation procedure used for QTL detection.

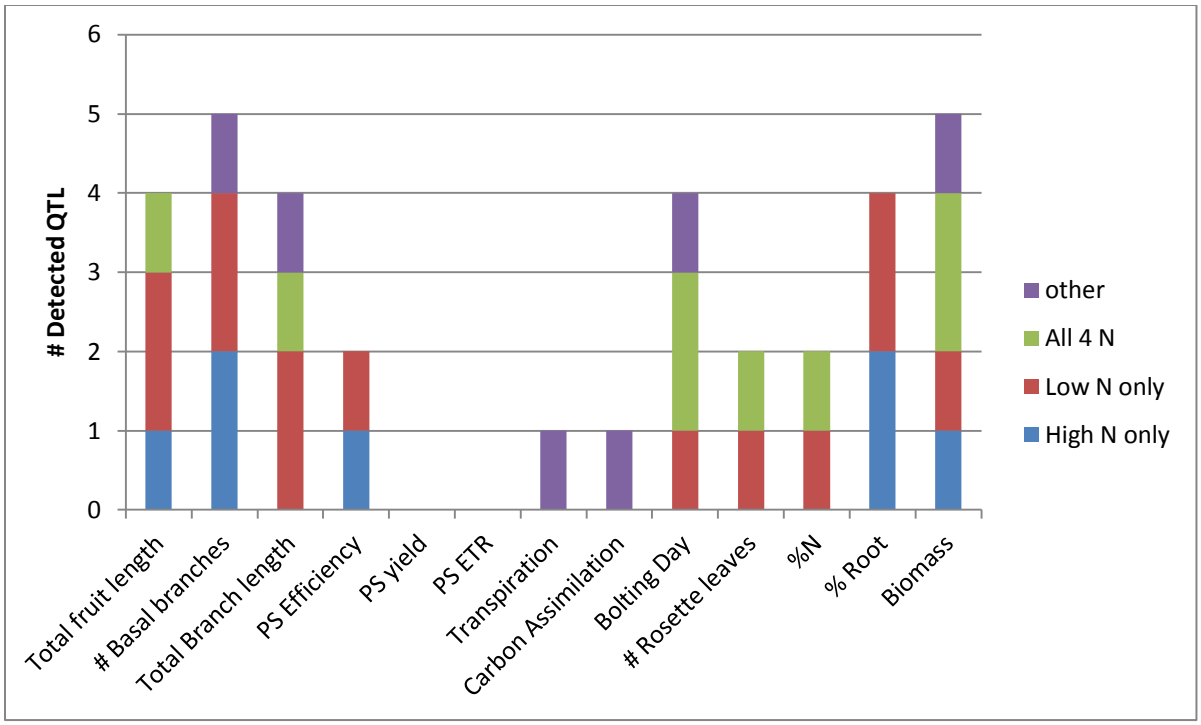


Figure 3.5. Environmentally specific QTL. Some QTL were detected in all four environments (green), while others were only present in high (blue) or low (red) N-environments. QTL found in some combination of high and low N-environments are indicated in purple.

4.0 EVOLVABILITY IN A VARIABLE WORLD- EVOLUTIONARY TRAJECTORIES

4.1 ABSTRACT

We investigated the effect of differences in G-matrix structure for a set of Arabidopsis thaliana genotypes grown across a nitrogen (N) supply gradient from extremely limiting to saturating N-supply rates. Our previous work with these populations identified differences in several measures of potential constraints to evolvability. Here, we continue our analyses by determining whether these populations have different predicted evolutionary responses to simulated selection pressures. This was accomplished through a modified version of Cheverud's (1983, 1996) random skewers test in which two G-matrices are post-multiplied by randomly generated selection gradients and the resulting response vectors are compared through vector correlations. Results consistent with our earlier work suggest that when N-levels become stressfully low, G-matrix x N-supply rate interactions are substantial enough to alter evolutionary trajectories. Differences in evolutionary trajectories were decomposed into effects due to covariance structure and effects due to main effect QTL detected in our earlier study. These goals were accomplished by removing covariances or QTL effects from G-matrix elements and repeating random skewers tests. These analyses indicate that the unique evolutionary paths taken by populations experiencing N-stress are due to both changes in variances and covariances and that variation at main effect QTL, while potentially altering response magnitude, do not affect direction of evolutionary trajectories.

4.2 INTRODUCTION

Understanding the G-matrix, a set of genetic variances and covariances among continuously distributed traits, can tell us about the evolutionary potential of populations and species. The G-matrix can give us indication of both the constraints to adaptation and the potential directions that most easily respond to selection (Schluter 1996, 2000). The G-matrix has been used to make projections of adaptive

trajectories in response to specific selection pressures and has been used retrospectively to reconstruct ancestral phenotypes (e.g. Meagher 1999) or selection pressures (e.g. Cheverud 1997) based on current distributions of genetic covariances. Both of these uses require assumptions of either a constant G-matrix or one that evolves in a predictable manner over projected time intervals (Steppan et al 2003).

While it has become generally accepted that genetic covariances themselves can evolve (Steppan et al 2003), the rate and degree of change remains unclear. Several studies have shown evidence that G-matrices remain constant for long time periods, suggesting that they are useful for extrapolation to perhaps hundreds or thousands of generations in the future (or past). Other studies have shown that G-matrices can change rapidly over short periods of time (Doroszuk et al 2008) or can become reoriented with shifts in in environment potentially biasing or constraining the multivariate response to selection (chapter 2). Recently more attention has been given to changes in G across environments. Understanding plastic changes in genetic covariance structure and trait integration is being used to help aid predictions of response to climate change (Etterson 2001) and understanding the success of invasive species (e.g. purple loostrife, Coulautti and Barrett 2011). There is still much to be learned about how G-matrices evolve.

Many methods exist to study G-matrix differentiation across populations or related groups of various taxonomic rank including T^2 tests (Roff et al.1999), maximum likelihood approaches (Shaw et al. 1995), MANOVA (Roff 2002), common principle components (Phillips and Arnold 1999), and random skewers measures (Cheverud and Marroig 2007). Random skewers (Cheverud 1983, Cheverud 1996, Marroig and Cheverud 2001, Revell 2007) and related approaches (Calsbeek and Goodnight 2009) are particularly interesting because they compare G-matrices based on their differences in response to simulated selection pressures. With a primary motivation in studying the G-matrix being its influence on adaptation, it makes sense to use evolutionary trajectory differences due to G as a measure of matrix differentiation. These random skewers approaches are based on the multivariate breeder's equation, that the change in a vector of trait means ($\Delta\bar{\mathbf{z}}$) is equal to the product of the G-matrix and the selection gradient, β , a vector summarizing the effect of trait values on relative fitness (usually estimated as partial regression coefficients of relative fitness regressed onto standardized trait values, Lande 1979, Lande and Arnold 1983): or a given selection gradient (Lande 1979),

$$\Delta\bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta} \qquad \text{(Eqn. 1).}$$

The random skewers method compares two G-matrices by measuring the correlation between responses to a simulated selection gradient in which the multivariate direction of selection is a random

draw. This is repeated many (1000+) times and the mean (vector) correlation of responses is a measure of similarity between G-matrices. Cheverud (1996) using this approach found G-matrices to be similar between cotton-top and saddle-back tamarins; by assuming stability of G since divergence, he was able to infer past selection differences based on current mean differences in skull morphology. Stinchcombe et al. (2010) recently used random skewers approaches to compare across-density genetic covariance matrices for *Impatiens capensis* between sun and shade treatments, finding only marginally significant support for similarity of G-matrices across environments.

While previous studies have used random skewers tests to compare across populations and species, we use this approach to continue our investigations into changes in the genetic architecture of a set of Arabidopsis thaliana genotypes across a nitrogen (N) supply gradient from extremely limiting to saturating N-supply rates. We previously found differences in genetic correlation structure, directions of eigenvectors and their magnitudes, and quantitative trait locus (QTL) expression across the N-environments for this set of Arabidopsis recombinant inbred lines (RIL). We now ask if these differences in genetic architecture across N-environments are large enough to alter the adaptive responses to simulated selection pressures. We have several goals for this study. First, we make element-by-element comparisons of genetic variances and covariances across N-environments. Second, we use a modified version of Cheverud's random skewers test to determine if G-matrix X N-supply rate interactions are substantial enough to alter evolutionary trajectories. Third, following an approach used by Agrawal and Stinchcombe (2009) we ask whether differences in evolutionary trajectory are due to differences in variance or covariances by comparing G-matrices with covariances set to zero. Finally, we ask whether differences in evolutionary responses are due to environment-specific QTL detected in our pervious study of this population (chapter 2). We address this question by removing the effects of detectable QTL and asking whether the simulated evolutionary trajectories differ by the same amount. We previously discovered several highly pleiotropic (or perhaps, sets of tightly linked) QTL, often with effects on several functionally unrelated traits. In this context, we use pleiotropy to refer to the effect of the genomic region encompassed by a 1.5 LOD support interval around QTL, and fully acknowledge that this may contain many linked genes with no true pleiotropic effects. With this distinction in mind, we use the term pleiotropy for ease of communication. If evolutionary trajectories converge after removal of highly pleiotropic loci, it will suggest that a small number of loci can largely explain adaptation, a matter of ongoing controversy. On the other hand, if removing QTL effects from G-matrices does not alter evolutionary trajectories, it will suggest that many undetectable (and thus of small individual effect) QTL in aggregate can contribute substantially to the direction of response to selection. In such an outcome, the genetic basis of adaptation as revealed by QTL studies not only misses small effect QTL (as is widely

understood), but the aggregate effect of these QTL may be more important than the detected large effect QTL.

4.3 METHODS

Our research used a set of 158 RIL derived from a cross between Landsberg erecta and Cape Verdi islands ecotypes (Alonso-Blanco et al. 1998) plus the Cvi and Ler parents (i.e. 160 genotypes). Approximately three replicates of each genotype were grown at each of four N-supply rates in Conviron controlled environment growth chambers. These N-environments will be referred to as N56, N51, N06, and N01, or as high (N56 and N51) and low (N06 and N01) N-environments. Details of growth chamber conditions and experimental set up can be found in chapter 2.

Traits measured included total biomass, proportion root/total biomass, percent nitrogen, age at first reproduction, number of rosette leaves at bolting, instantaneous carbon assimilation rate, instantaneous transpiration rate, photosynthetic quantum efficiency, photosynthetic quantum yield, and photosynthetic electron transport rate, total branch length, number of basal reproductive branches, and total fruit length. Measurements of these traits were adjusted for growth chamber runs to remove chamber effects. Chamber-adjusted residuals were used for all analyses.

4.3.1 Statistical analyses

All statistical analyses were carried out using SAS/STAT, version 9.3 for Windows (SAS Institute 2010). Broad sense genetic variances and covariances were estimated as the covariance among genotypic (i.e. RIL) means. These covariances are not purely additive; this approach confounds additive genetic variances with additive x additive epistatic variance. Dominance and dominance-based epistatic effects are assumed absent due to homozygosity of RIL at nearly all loci. Additionally, Via (1984) pointed out that a fraction of the within-RIL variance can contribute to the covariances among RIL. These statistical effects will likely affect G-matrix estimation identically within each N-environment, thus we expect that they will not alter the outcomes of any between-environment comparisons which are the focus of this study.

Prior to calculating the covariance among RIL means, each trait was mean standardized (i.e. mean-standardized RIL mean = RIL mean/grand mean of RIL means). This produces a G-matrix with several desirable properties. Mean-standardization converts traits to unit-less proportions facilitating

comparisons across traits and environments. Variance of a mean-standardized trait is equal to the square of the coefficient of genetic variation (CVg = genetic SD/mean), a statistic which Houle (1992) named "evolvability". The G-matrix derived from mean-standardized traits can be applied in the context of the multivariate breeder's equation as follows:

$$\Delta\tilde{\mathbf{z}} = \tilde{\mathbf{G}}\tilde{\boldsymbol{\beta}} = (\mathbf{ERE})\tilde{\boldsymbol{\beta}}$$ Eqn. 2 (Kirkpatrick 2009).

In this formulation of the breeder's equation, $\tilde{\mathbf{G}}$ is the mean-standardized (i.e. normalized) genetic covariance matrix, $\tilde{\boldsymbol{\beta}}$ is the selection gradient indicating the increase in relative fitness per *proportional* change in trait values, and $\Delta\tilde{\mathbf{z}}$ is the *proportional* change in trait means after one generation of selection. The matrix **E** is a diagonal matrix of CVg's (Houle's "evolvability") and **R** is a matrix of genetic correlations (Kirkpatrick 2009)**.** Overall, normalizing traits to estimate $\tilde{\mathbf{G}}$ facilitates comparisons among traits without sacrificing relevance to multivariate selection. All analyses of genetic covariance reported here are based on $\tilde{\mathbf{G}}$, which will be referred to as the G-matrix for convenience. Specific G-matrix estimates for plants grown in the four N-environments N56, N51, N06, and N01 will be referred to as G56, G51, G06 and G01 respectively.

We estimated 95% confidence intervals on elements of G-matrices in each of the four N-environments. Confidence intervals were estimated as the 2.5^{th} and 97.5^{th} percentiles of the distribution of each matrix element from 10 000 bootstrapped samples (Manly 1991). Bootstrapping was carried out using a SAS macro program created by TWE (Appendix A). Genetic covariances were compared across N- environments and were considered significantly different if the 95% CIs did not overlap. While these analyses may be similar to those of genetic correlations (chapter 2), these comparisons are among mean-standardized variances and covariances using a distribution free approach (cf. parametric CI and SD standardized covariances, i.e. correlations). Furthermore, comparisons of covariances allows us to analyze diagonal elements of G-matrices which are always set to 1 when dealing with correlations.

4.3.2 Random skewers G-matrix comparisons

We compared G-matrices consisting of the variances and covariances among twelve traits. The trait *total fruit length* was left out of our random skewers analyses. This allowed our G-matrix estimates

and evolutionary trajectories to be comparable to our earlier principle component analysis based on the same twelve traits (chapter 2).

The random skewers test begins comparisons of two G-matrices by multiplying each by a simulated selection gradient β. For our analyses, β is a 12 x 1 vector with each element randomly determined to be in the interval (-1, 1). This "random skewer" was standardized to unit length by dividing each vector element by the vector magnitude. The product of each Gβ is a 12 x 1 response vector. The vector correlation between the response vectors of the two matrices is calculated as:

$$\frac{V_1^T * V_2}{\sqrt{(V_1^T * V_1) * (V_2^T * V_2)}} \qquad \text{Eqn. 3}$$

, with V_1 and V_2 being the response vectors, T indicating transpose, and * indicating vector multiplication. A vector correlation is bounded between (-1, 1), with ±1 indicating vectors are parallel and zero indicating response vectors are orthogonal. We used the mean vector correlation between response vectors generated by 10 000 random skewers as a measure of G-matrix similarity. A SAS macro program (created by TWE) to generate random skewers tests for various G-matrices and can be found in Appendix B. The mean vector correlation was then compared to a null distribution to test the hypothesis that the mean vector correlation = 1. We tested all pairwise comparisons of G56, G51, G06, and G01 using random skewers.

4.3.3 Null distributions for random skewers tests

As Calsbeek and Goodnight (2009) pointed out, the appropriate null hypothesis for random skewers (and related tests) comparing closely related groups is that of identical G-matrices. With our experiment using the same set of RIL grown in 4 N-environments, one could have strong expectations of identical evolutionary trajectories. In order to create a null distribution for the comparison of two G-matrices, we carried out a random skewers test on bootstrapped replicates of a single G-matrix. For example, to estimate a null distribution based on G56, we compared bootstrapped G56 matrices using random skewers and recorded the vector correlation between 10 000 pairs of responses vectors. These null distributions tell us about the repeatability of the original G-matrix estimate and only vary due to sampling error. A SAS macro program (written by TWE) to create the null distributions of mean-standardized G-matrices can be found in Appendix C.

We chose *a priori* to use the null distribution generated from bootstrapped samples of G56 for comparisons between G56 and other G-matrices. Our justification is that G56 represents the variation expressed by the Ler x Cvi RIL under benign conditions with other environments representing various amounts of nitrogen stress (see chapter 1 for demonstration of increasing stress with decreasing N-supply rate). When comparisons were made between pairs of G51, G06, and G01, we chose to use the higher (less stressful) N-environment to generate the null distribution for the test. We have included data on the null distribution for N01 for completeness.

4.3.4 Evolutionary trajectories: variances or covariances?

Having found differences in evolutionary trajectories among G (see results), we asked whether differences in evolutionary trajectories were due to genetic covariances constraining responses to selection in different directions. To address this, we carried out all pairwise random skewers comparisons among G-matrices, each with their off-diagonal elements set to zero. Null distributions were recalculated as before but using diagonal (i.e. no covariance) matrices.

4.3.5 Evolutionary trajectories: pleiotropic large effect QTL or many undetected loci?

We further used random skewers tests on G-matrices adjusted for the effects of additive QTL discovered through Multiple Imputation Interval Mapping (Sen and Churchill 2001) using R/qtl (Broman et al. 2003). Details of our QTL mapping procedure can be found in chapter 2. Kelly (2009) showed how elements of a G-matrix can be decomposed into the effects of individual QTL. As our analyses used mean-standardized G-matrices, this method was not applicable. In order to remove QTL effects from G-matrices, we adjusted RIL means for all measured additive QTL effects. Specifically, for each QTL in each N-environment, we added twice the additive effect of the Cvi allele (2a) to all genotypes homozygous for Ler alleles at that locus. In effect, we adjusted all plants to the phenotype corresponding to the Cvi genotype at 21 loci nearest to the 60 QTL detected across N-environments. Once QTL effects were removed for all trait-environment combinations, traits were mean standardized and the G-matrices were reestimated. We then repeated random skewers analyses, making all pairwise comparisons of G-matrices lacking QTL effects as well as comparisons of each G-matrix to itself with and without QTL effects included. Null distributions were recalculated for G-matrices after removal of QTL effects.

4.4 RESULTS

4.4.1 Genetic covariances

Element-by-element comparison of G-matrices across N-environments revealed G01 differed substantially from G56, G51, and G06 (tables 4.1 and 4.2). Out of 364 comparisons of variances and covariances across N- environments, there were 48 differences between any pair of nitrogen environments (table 4.2). Seven of these differences were between G56, G51, and G6. Thirty-three of the differences in covariances were between G01 and one of the other G-matrices. Within the three higher N-environments, the differences in G were due to differences in diagonal elements (variances) rather than covariances in 6 of the 7 cases. Of the 33 differences in G-matrix elements between G01 and other environments, 10 were due to differences in genetic variances and the remaining 23 were due to changes in covariance structure.

4.4.2 Random skewers null distributions

The three highest N null distributions of vector correlations were very similar. The null distribution based on G01 (hereafter Null01) had a larger range of response vector correlations than other environments. Within Null56, Null51, and Null06, vector correlations had maxima above 0.999 and minima of 0.33, 0.37, and 0.33, respectively (figure 4.1). Null01 had a similar maximum, but had a minimum vector correlation of 0.006, indicating a nearly orthogonal evolutionary trajectory. While likely an outlier, this indicated more variation (lower repeatability of G) in G01 than other G-matrices (differences in sampling error in G01 relative to other G were addressed using MANOVA in Chapter 2).

4.4.3 QTL effects on G-matrices

The 60 QTL discovered among 12 traits in 4 N-environments were attributable to only 21 different marker positions (table 4.3). This indicates that each position used in adjusting for QTL effects affected an average of 3 traits across environments. In some cases this was the same trait in multiple environments. In others, particularly in QTL 'hotspots' on chromosomes 1, 3, and 5 (Chapter 2), loci affect multiple traits in multiple environments. In the latter two hotspots, loci on chromosomes 2 and 5 affected 10 and 13 trait-environment combinations respectively.

4.4.4 Random skewers G-matrix comparisons

Random skewers analyses among mean-standardized G-matrices showed G01 was significantly different from all other G-matrices in its simulated evolutionary trajectory ($p < 0.03$ all comparisons, table 4.4). In all comparisons of G56, G51, and G06, we were unable to reject the null hypothesis of identical evolutionary trajectories (i.e. H_0: response vector correlation =1).

When covariances were removed from G-matrices, response vector correlations remained significantly different between G01 and all three of the higher N-environments (table 4.4), indicating that both variances and covariances differentiate G01 from other G-matrices. P-values decreased in the G51 vs. G06 comparison when covariances were removed ($p = 0.11$ with full comparison, $p < 0.03$ without covariances, table 4.4), indicating that similar evolutionary trajectories are likely caused by shared covariance structure in these two environments. In each comparison, the mean vector correlations for the no covariance comparisons were greater than the corresponding comparisons with both variances and covariances included. This indicates that removing genetic covariances reduces the range of the null distributions as well as altering trajectories between them.

Removal of QTL effects increased the divergence of response vectors between high and low N-environment G-matrices. We rejected the null hypothesis of identical trajectories for all pairs of QTL adjusted G-matrices except the G56 and G51 pair. All other comparisons revealed divergent responses, with G01 evolving most divergently from all other G in this scenario. The mean vector correlation between G01 and any other environment was lower than all 10 000 bootstrapped null distribution vector correlations indicating H_0 was rejected at $p < 0.0001$.

When G-matrices were tested against themselves with and without QTL effects, evolutionary trajectories were highly correlated and we failed to reject the hypothesis of identical evolutionary trajectories in all cases.

4.5 DISCUSSION

Previous work in this system (chapter 2) has suggested that genetic architecture changes with nitrogen stress in a manner that has the potential to alter constraints to and biases in the responses to selection. Here we elaborate on our previous work to show that changes in genetic architecture across a nitrogen supply gradient alters G-matrices enough to cause populations to diverge substantially in their simulated responses to selection. As before, the differences in covariance structure were largely between

plants in N01 and the other three N-environments. Severe nitrogen stress changes the G-matrix enough to result in plant populations in N01 taking a unique evolutionary trajectory. While some changes in genetic architecture were detected across N56, N51, and N06 through PCA and QTL mapping (chapter 2), these differences failed to alter G enough for populations in these N-environments to respond differently to simulated selection pressures.

G01 differed from other G-matrices in both variances and covariances, whereas the three other G primarily differed in their variances. Additionally, when covariances were removed, G01 still followed a different evolutionary trajectory from other G, indicating that the pattern of trait variances differed substantially in G01 from other G-matrices. Among the three higher N-environments, the majority of differences in G-matrix elements were variances but these differences were relatively small.

Removing covariances increased significance of trajectory divergence relative to unaltered G-matrix comparisons, but the actual mean vector correlation increased in the no covariance tests over the unaltered G-matrix comparisons. This suggests that removing covariances produces a greater reduction in the multi-trait variation within an environment (and thus narrowing the null distribution of vector correlations) than between environments. Having covariances gives populations more "wiggle-room" in multivariate trait space; without them populations in different N-environments diverged less as shown by higher mean response vector correlations. This is almost the opposite of the constraining effects covariances are commonly thought to have.

Our QTL adjusted skewers test results were surprising in that responses did not become more correlated across N-environments after accounting for the effects of 60 QTL at 21 loci. Removing the QTL effects will reduce the variation within that environment (perhaps also the transgressive segregation detected previously), but it actually increased differences between the low and high N-environment G-matrix trajectories. It is possible that by removing variation within environments, covariances may have been altered in a way that constrained response vectors to point in different directions. Regardless of the mechanism of the greater divergence, this result indicates that more genetic architecture differences that impact responses to selection exist than were revealed by QTL mapping.

When G-matrices were compared with and without QTL effects, no differences in the direction of evolutionary trajectories were detected-evolutionary trajectories did not differ from perfectly correlated. Does this suggest that QTL variation has no effect on the direction of multivariate response to selection? There are several hypotheses to explain the lack of differences. Taken at face value, this result could mean that many undetected small-effect QTL collectively make greater contributions to the pattern of multi-trait variation than the QTL we discovered. These QTL may even have moderate sized effects and still go undetected because of our criteria for QTL inclusion. When two QTL had overlapping 1.5 LOD support intervals, we only retained the one with the higher peak for our analyses (chapter 2). Finally, it is

possible that the effects of the QTL we detected affect the multivariate phenotype in the same manner as the undetected QTL. If so, removing the QTL effects would not change the predicted direction of evolution (although it likely affects the magnitude of variation and thus the rate of change). This would be consistent with Orr's (1999, 2006) model of adaptation-that adaptation proceeds by substitutions of initially large effect, with the effect sizes becoming smaller as populations approach fitness peaks. Under this model it is possible that large and small effect QTL both affect trait means (and potentially trait variation) in the same manner.

It is important to keep in mind that these G-matrices being compared are from the same population of genotypes grown under different environmental conditions. This means that regardless of the rate at which G-matrices evolve, the patterns of constraints and accelerants to evolution inherent in G-matrix structure are environmentally specific and can vary enough across environments (particularly extreme environments), to drastically enough to alter the short-term response to selection.

Table 4.1. Genetic covariance matrices. Cells show covariances, upper and lower 95% CI (determined through 10 000 bootstrapped samples), and degrees of freedom. (a) G56 (b) G51 (c) G06 (d) G01.

Mean Standardized Variances and Covariances N56 Covariance / LCI / UCI / DF													
	totg	proprt	Npct	nlvs	boltd	delco2	delh2o	fvfm	dietr	yield	totlen	rb	tsl
totg	0.420	-0.142	-0.060	0.179	0.096	0.231	0.121	0.004	0.039	0.002	0.364	0.129	0.338
	0.341	-0.200	-0.078	0.119	0.064	0.168	0.077	0.001	-0.014	-0.004	0.294	0.087	0.260
	0.497	-0.085	-0.040	0.242	0.126	0.296	0.168	0.006	0.096	0.008	0.435	0.169	0.416
	158	158	158	155	155	158	158	149	153	152	158	158	158
proprt		0.360	0.029	-0.034	-0.018	-0.076	-0.038	0.000	-0.024	-0.004	-0.110	-0.039	-0.112
		0.135	0.016	-0.068	-0.038	-0.115	-0.061	-0.001	-0.074	-0.013	-0.141	-0.070	-0.151
		0.673	0.041	-0.002	0.004	-0.033	-0.011	0.002	0.021	0.003	-0.070	-0.005	-0.065
		158	158	155	155	158	158	149	153	152	158	158	158
Npct			0.044	0.004	0.003	-0.003	-0.001	0.000	-0.011	0.000	-0.066	-0.020	-0.023
			0.036	-0.015	-0.006	-0.023	-0.017	-0.001	-0.026	-0.001	-0.083	-0.030	-0.044
			0.052	0.026	0.013	0.018	0.016	0.001	0.005	0.002	-0.047	-0.008	-0.002
			159	156	156	159	159	150	154	153	159	159	159
nlvs				0.205	0.098	0.133	0.058	0.003	0.028	0.003	0.115	0.056	0.175
				0.129	0.068	0.089	0.026	0.002	-0.013	-0.001	0.074	0.028	0.122
				0.280	0.126	0.176	0.091	0.004	0.076	0.008	0.157	0.085	0.230
				156	156	156	156	147	151	150	156	156	156
boltd					0.070	0.074	0.032	0.002	0.020	0.002	0.061	0.029	0.097
					0.053	0.045	0.012	0.001	-0.001	0.000	0.035	0.012	0.069
					0.084	0.101	0.052	0.003	0.045	0.004	0.086	0.045	0.123
					156	156	156	147	151	150	156	156	156
delco2						0.360	0.233	0.004	-0.012	0.003	0.199	0.078	0.234
						0.284	0.180	0.002	-0.078	-0.002	0.132	0.042	0.170
						0.442	0.291	0.006	0.045	0.008	0.271	0.115	0.301
						159	159	150	154	153	159	159	159
delh2o							0.196	0.002	-0.032	0.002	0.109	0.048	0.130
							0.152	0.001	-0.075	-0.002	0.062	0.024	0.087
							0.242	0.003	0.006	0.006	0.161	0.072	0.173
							159	150	154	153	159	159	159
fvfm								0.001	0.000	0.000	0.004	0.002	0.004
								0.000	-0.002	0.000	0.002	0.001	0.002
								0.001	0.002	0.001	0.007	0.003	0.006
								150	150	149	150	150	150
dietr									0.362	-0.006	0.011	-0.027	0.037
									0.268	-0.014	-0.036	-0.078	-0.016
									0.476	0.002	0.059	0.016	0.092
									154	153	154	154	154
yield										0.003	0.004	0.003	0.003
										0.002	-0.001	-0.002	-0.003
										0.004	0.009	0.007	0.009
										153	153	153	153
totlen											0.405	0.147	0.294
											0.323	0.101	0.220
											0.487	0.193	0.368
											159	159	159
rb												0.161	0.123
												0.122	0.079
												0.202	0.168
												159	159
tsl													0.425
													0.327
													0.524
													159

Table 4.1 continued.

Mean Standardized Variances and Covariances N51 Covariance / LCI / UCI / DF													
	totg	proprt	Npct	nlvs	boltd	delco2	delh2o	fvfm	dietr	yield	totlen	rb	tsl
totg	0.452	-0.144	-0.052	0.115	0.070	0.215	0.135	0.004	-0.001	0.001	0.374	0.152	0.339
	0.351	-0.204	-0.075	0.076	0.044	0.144	0.090	0.002	-0.047	-0.012	0.296	0.107	0.242
	0.562	-0.087	-0.029	0.156	0.098	0.287	0.180	0.006	0.048	0.010	0.453	0.199	0.445
	152	152	149	146	146	150	150	146	146	146	152	152	149
proprt		0.304	0.034	-0.049	-0.024	-0.098	-0.055	-0.002	-0.010	0.001	-0.130	-0.087	-0.146
		0.174	0.011	-0.077	-0.044	-0.152	-0.099	-0.005	-0.045	-0.004	-0.187	-0.125	-0.208
		0.439	0.059	-0.023	-0.006	-0.050	-0.021	0.000	0.030	0.006	-0.072	-0.054	-0.091
		152	149	146	146	150	150	146	146	146	152	152	149
Npct			0.045	-0.002	0.009	-0.023	-0.018	0.000	-0.006	0.000	-0.063	-0.018	-0.044
			0.035	-0.017	-0.004	-0.044	-0.031	0.000	-0.018	-0.002	-0.084	-0.031	-0.068
			0.055	0.015	0.023	-0.003	-0.005	0.001	0.006	0.002	-0.043	-0.006	-0.021
			152	146	147	149	149	145	145	145	152	152	150
nlvs				0.182	0.088	0.118	0.056	0.003	0.028	0.001	0.134	0.089	0.247
				0.098	0.049	0.081	0.037	0.002	-0.010	-0.004	0.074	0.044	0.128
				0.282	0.131	0.159	0.077	0.004	0.072	0.005	0.212	0.153	0.379
				151	150	146	146	142	142	142	149	149	146
boltd					0.076	0.065	0.027	0.002	0.007	0.002	0.063	0.041	0.134
					0.052	0.042	0.011	0.001	-0.017	-0.001	0.029	0.017	0.082
					0.102	0.090	0.043	0.003	0.036	0.004	0.099	0.069	0.187
					151	146	146	142	142	142	149	149	146
delco2						0.383	0.179	0.003	0.007	0.003	0.196	0.122	0.242
						0.296	0.132	0.001	-0.037	-0.004	0.135	0.064	0.170
						0.469	0.226	0.005	0.055	0.010	0.258	0.190	0.316
						153	153	148	148	148	152	153	149
delh2o							0.141	0.001	0.012	-0.001	0.123	0.073	0.138
							0.108	0.000	-0.009	-0.006	0.081	0.050	0.092
							0.176	0.003	0.034	0.003	0.165	0.095	0.184
							153	148	148	148	152	153	149
fvfm								0.000	-0.001	0.000	0.004	0.002	0.006
								0.000	-0.003	0.000	0.002	0.000	0.003
								0.001	0.001	0.000	0.006	0.003	0.008
								149	149	149	148	149	146
dietr									0.188	-0.006	-0.011	0.000	0.008
									0.133	-0.012	-0.050	-0.031	-0.064
									0.257	-0.002	0.033	0.030	0.075
									149	149	148	149	146
yield										0.003	0.006	0.003	0.006
										0.002	-0.001	-0.001	-0.007
										0.005	0.012	0.007	0.018
										149	148	149	146
totlen											0.439	0.179	0.409
											0.353	0.128	0.313
											0.525	0.235	0.506
											155	155	152
rb												0.209	0.211
												0.151	0.137
												0.271	0.293
												156	152
tsl													0.696
													0.441
													0.984
													152

Table 4.1 continued.

Mean Standardized Variances and Covariances N06 Covariance / LCI / UCI / DF													
	totg	proprt	Npct	nlvs	boltd	delco2	delh2o	fvfm	dietr	yield	totlen	rb	tsl
totg	0.693 0.451 0.950 156	-0.201 -0.276 -0.122 155	-0.055 -0.075 -0.036 156	0.147 0.084 0.212 156	0.097 0.054 0.145 155	0.395 0.218 0.583 132	0.167 0.100 0.238 132	0.004 0.001 0.007 134	0.025 -0.014 0.067 134	0.020 0.006 0.039 133	0.369 0.236 0.503 156	0.130 0.082 0.189 156	0.320 0.231 0.410 153
proprt		0.597 0.440 0.754 155	0.048 0.024 0.074 155	-0.011 -0.049 0.036 155	-0.013 -0.035 0.012 154	-0.171 -0.281 -0.054 131	-0.102 -0.153 -0.049 131	0.002 0.000 0.005 133	0.032 -0.047 0.113 133	-0.004 -0.024 0.010 132	-0.208 -0.277 -0.137 155	-0.099 -0.145 -0.055 155	-0.152 -0.227 -0.075 152
Npct			0.037 0.030 0.046 159	-0.002 -0.011 0.008 159	-0.001 -0.008 0.007 158	-0.023 -0.048 0.001 133	-0.013 -0.026 0.000 133	0.000 -0.001 0.002 136	0.001 -0.012 0.012 136	-0.001 -0.007 0.004 135	-0.077 -0.098 -0.055 159	-0.021 -0.031 -0.010 159	-0.044 -0.063 -0.026 156
nlvs				0.143 0.058 0.242 159	0.070 0.032 0.122 158	0.209 0.074 0.357 133	0.057 0.023 0.097 133	0.002 0.001 0.004 136	0.026 0.006 0.048 136	0.007 0.002 0.014 135	0.072 0.039 0.107 159	0.030 0.016 0.046 159	0.105 0.065 0.147 156
boltd					0.077 0.047 0.118 158	0.085 0.030 0.145 132	0.024 0.001 0.047 132	0.001 0.000 0.003 135	0.029 0.009 0.049 135	0.005 0.001 0.010 134	0.059 0.031 0.086 158	0.012 0.000 0.026 158	0.092 0.060 0.122 155
delco2						0.878 0.586 1.215 133	0.361 0.260 0.469 133	0.003 -0.001 0.007 117	0.008 -0.056 0.074 117	0.021 0.006 0.040 116	0.228 0.104 0.364 133	0.108 0.065 0.154 133	0.268 0.143 0.396 131
delh2o							0.216 0.159 0.276 133	0.001 -0.001 0.003 117	0.011 -0.020 0.045 117	0.007 -0.001 0.018 116	0.114 0.063 0.168 133	0.068 0.043 0.094 133	0.137 0.077 0.202 131
fvfm								0.001 0.000 0.001 136	0.000 -0.001 0.002 136	0.000 0.000 0.000 135	0.004 0.001 0.008 136	0.003 0.001 0.005 136	0.005 0.002 0.008 134
dietr									0.160 0.104 0.227 136	0.001 -0.008 0.013 135	0.006 -0.029 0.045 136	-0.003 -0.020 0.015 136	0.015 -0.031 0.063 134
yield										0.019 0.005 0.042 135	0.018 0.003 0.038 135	0.008 -0.001 0.018 135	0.016 0.004 0.033 133
totlen											0.515 0.389 0.648 159	0.159 0.106 0.228 159	0.360 0.273 0.453 156
rb												0.143 0.110 0.182 159	0.117 0.078 0.162 156
tsl													0.442 0.339 0.551 156

Table 4.1 continued.

Mean Standardized Variances and Covariances N01 Covariance / LCI / UCI / DF													
	totg	proprt	Npct	nlvs	boltd	delco2	delh2o	fvfm	dietr	yield	totlen	rb	tsl
totg	1.389 0.679 2.235 157	-0.068 -0.139 0.004 157	-0.077 -0.110 -0.038 156	0.126 0.063 0.196 156	0.126 0.067 0.181 156	0.058 -0.102 0.266 74	0.030 -0.011 0.074 74	0.001 -0.005 0.008 56	0.014 -0.081 0.114 57	0.027 0.000 0.056 54	0.914 0.406 1.523 157	0.115 0.045 0.184 157	0.762 0.338 1.257 143
proprt		0.314 0.164 0.531 157	0.020 0.007 0.035 156	0.019 0.002 0.037 156	0.019 -0.010 0.047 156	-0.003 -0.143 0.118 74	-0.014 -0.042 0.015 74	0.004 0.000 0.008 56	0.063 0.004 0.137 57	0.005 -0.010 0.021 54	-0.091 -0.148 -0.035 157	-0.033 -0.062 -0.004 157	-0.029 -0.088 0.027 143
Npct			0.045 0.035 0.056 157	-0.004 -0.011 0.004 156	0.000 -0.010 0.011 156	-0.040 -0.109 0.013 73	-0.019 -0.032 -0.007 73	0.001 -0.001 0.003 56	0.006 -0.029 0.044 57	0.000 -0.010 0.009 54	-0.096 -0.123 -0.066 157	-0.006 -0.018 0.006 157	-0.058 -0.086 -0.029 142
nlvs				0.063 0.035 0.096 158	0.054 0.031 0.077 157	-0.014 -0.082 0.064 74	-0.005 -0.018 0.007 74	0.004 0.001 0.008 56	0.005 -0.024 0.038 57	0.011 0.001 0.023 54	0.063 0.030 0.093 158	-0.005 -0.021 0.009 158	0.088 0.048 0.129 144
boltd					0.103 0.068 0.142 158	-0.048 -0.136 0.037 74	-0.016 -0.034 0.002 74	0.004 0.001 0.007 56	0.002 -0.036 0.045 57	0.014 0.003 0.026 54	0.056 0.020 0.093 158	-0.004 -0.028 0.027 158	0.090 0.052 0.127 143
delco2						1.955 1.132 3.414 74	0.270 0.121 0.457 74	-0.008 -0.029 0.008 56	0.198 0.018 0.418 57	-0.012 -0.067 0.043 54	0.094 -0.056 0.277 74	0.000 -0.118 0.126 74	0.098 -0.039 0.262 73
delh2o							0.085 0.058 0.113 74	-0.003 -0.005 -0.001 56	0.017 -0.013 0.050 57	-0.008 -0.019 0.004 54	0.031 -0.003 0.067 74	0.025 0.004 0.046 74	0.018 -0.010 0.045 73
fvfm								0.002 0.001 0.003 56	0.000 -0.006 0.005 56	0.002 0.001 0.004 53	-0.003 -0.009 0.003 56	0.000 -0.004 0.004 56	-0.001 -0.008 0.005 55
dietr									0.482 0.221 0.783 57	0.022 -0.006 0.053 54	0.005 -0.051 0.065 57	-0.003 -0.047 0.045 57	0.039 -0.016 0.102 56
yield										0.032 0.018 0.049 54	0.013 -0.006 0.033 54	-0.012 -0.024 0.001 54	0.027 0.004 0.053 53
totlen											0.842 0.477 1.238 159	0.143 0.088 0.200 159	0.618 0.312 0.961 144
rb												0.137 0.109 0.168 159	0.075 0.023 0.129 144
tsl													0.625 0.370 0.877 144

Table 4.2. Number of significantly different genetic covariances across N-environments.

56 vs 51	56 vs 06	56 vs 01	51 vs 06	51 vs 01	06 vs 01
1	3	14	3	19	8

Table 4.3. Adjusted quantitative trait loci. The loci nearest to each QTL was adjusted to remove QTL effects. The marker, chromosome, and position in cM are listed, followed by the number of traits affected by the locus in each environment.

Marker	QTL	N56	N51	N06	N01	Total
AXR-1	Chr 1 Pos 07	2	1	1	1	5
GB.206L/211C-Col	Chr 1 Pos 11	1	0	0	0	1
EC.88C	Chr 1 Pos 84	0	0	0	1	1
CH.65C	Chr 2 Pos 30	0	0	0	1	1
CH.145L-Col/150C	Chr 2 Pos 49	3	3	2	2	10
BH.195L-Col	Chr 2 Pos 53	1	0	0	0	1
BH.120L-Col	Chr 2 Pos 59	1	0	0	0	1
EC.235L-Col/247C	Chr 2 Pos 71	0	0	0	2	2
HH.158L	Chr 3 Pos 26	0	0	0	1	1
HH.171C-Col/173L	Chr 3 Pos 72	0	0	1	0	1
GB.490C	Chr 4 Pos 69	0	0	1	0	1
GB.750C	Chr 4 Pos 78	0	0	1	0	1
BH.342C/347L-Col	Chr 4 Pos 85	1	0	2	0	3
FD.207L	Chr 5 Pos 00	0	0	0	1	1
BH.325L	Chr 5 Pos 15	0	1	2	1	4
BH.107L-Col	Chr 5 Pos 19	3	1	0	2	6
AD.114C-Col	Chr 5 Pos 24	0	1	0	0	1
DF.231C	Chr 5 Pos 25	0	0	1	0	1
GH.473C	Chr 5 Pos 34	3	3	6	1	13
GH.121L-Col	Chr 5 Pos 40	1	1	1	1	4
EG.205L	Chr 5 Pos 113	1	0	0	0	1

Table 4.4. Percentiles of Null distributions. Vector correlation null distributions generated from G56, G51, G06, and G01 through 10 000 bootstrapped datasets. P 5, P 1, etc. indicate the 5th, 1st , etc. Percentiles of the VC=1 null distribution. If a mean vector correlation between two G-matrices is less than P 5, the G-matrices are significantly different at $p < 0.05$.

	Percentiles						
	P 5	P 1	P 0.1	P 0.01	P 0.001	P 0.0001	P 0.00001
Null 56	0.866	0.761	0.634	0.411	0.332	0.332	0.332
Null 51	0.873	0.762	0.590	0.369	0.368	0.368	0.368
Null 06	0.867	0.742	0.491	0.347	0.335	0.335	0.335
Null 01	0.839	0.722	0.549	0.090	0.006	0.006	0.006
Null 56 nocov	0.902	0.840	0.773	0.664	0.647	0.647	0.647
Null 51 nocov	0.961	0.938	0.899	0.832	0.826	0.826	0.826
Null 06 nocov	0.956	0.934	0.897	0.875	0.871	0.871	0.871
Null 01 nocov	0.886	0.815	0.716	0.626	0.596	0.596	0.596
Null 56 noqtl	0.844	0.765	0.674	0.617	0.603	0.603	0.603
Null 51 noqtl	0.946	0.916	0.865	0.813	0.811	0.811	0.811
Null 06 noqtl	0.956	0.930	0.898	0.844	0.841	0.841	0.841
Null 01 noqtl	0.899	0.826	0.738	0.623	0.606	0.606	0.606

Table 4.5. Mean vector correlations (VC) and associated p-values. The test column indicates whether the comparison was between G-matrices, G-matrices with no covariances, or G-matrices with QTL effects removed. Mean vector correlation if for 10 000 random skewers tests. P-values are based on a null hypothesis distribution generated from the higher N-environment of the pair. For example, 'No Cov G51 vs G06' is tested against a null distribution based on a diagonal (i.e. no covariance) G51 matrix.

Test	Environments	Mean VC	p-value	Null
G vs G	56 vs 51	0.958	0.299	G56
G vs G	56 vs 06	0.896	0.084	G56
G vs G	56 vs 01	0.725	**0.005**	G56
G vs G	51 vs 06	0.920	0.110	G51
G vs G	51 vs 01	0.762	**0.010**	G51
G vs G	06 vs 01	0.817	**0.025**	G06
No Cov	56 vs 51	0.976	0.347	G56 nocov
No Cov	56 vs 06	0.924	0.079	G56 nocov
No Cov	56 vs 01	0.838	**0.009**	G56 nocov
No Cov	51 vs 06	0.954	**0.029**	G51 nocov
No Cov	51 vs 01	0.875	**0.000**	G51 nocov
No Cov	06 vs 01	0.926	**0.006**	G06 nocov
No QTL	56 vs 51	0.906	0.136	G56 noqtl
No QTL	56 vs 06	0.857	**0.062**	G56 noqtl
No QTL	56 vs 01	0.546	**0.000**	G56 noqtl
No QTL	51 vs 06	0.888	**0.002**	G51 noqtl
No QTL	51 vs 01	0.640	**0.000**	G51 noqtl
No QTL	06 vs 01	0.698	**0.000**	G06 noqtl
G vs No QTL	56 vs 56	0.891	0.076	G56
G vs No QTL	51 vs 51	0.941	0.168	G51
G vs No QTL	06 vs 06	0.900	**0.083**	G06
G vs No QTL	01 vs 01	0.889	0.106	G01

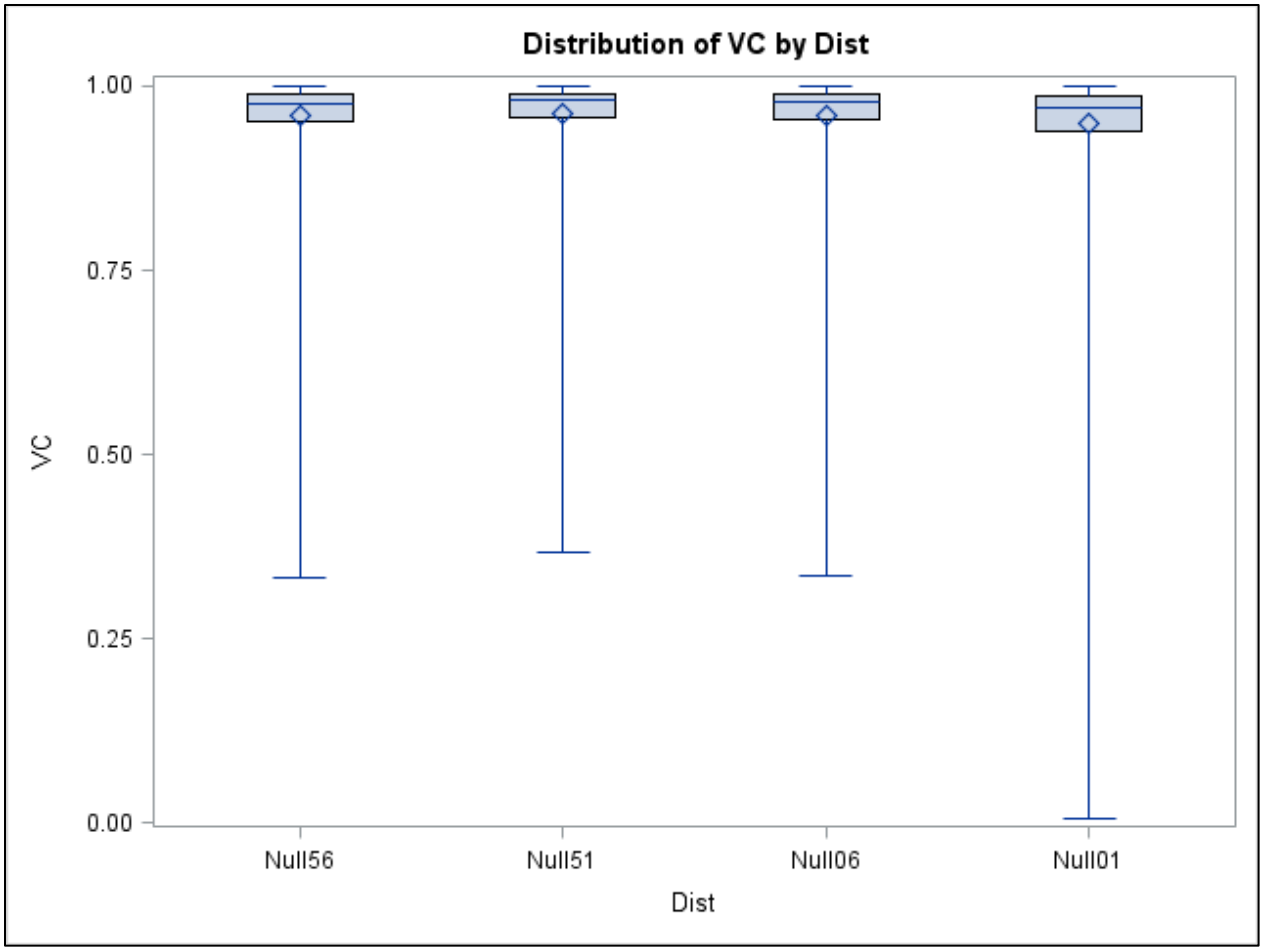


Figure 4.1. Null distributions of random skewers. Box plots of the null distributions of vector correlations for each G-matrix.

5.0 EPISTASIS AND PLASTICITY QTL MAPPING

5.1 ABSTRACT

Phenotypic plasticity and epistasis are two components of genetic architecture which can have substantial effects on the evolvability of populations. Plastic responses to environmental changes can allow maintenance of near-optimal phenotype when exposed to variable conditions or when the environment changes more rapidly than the rate at which the population can adapt. Epistasis may constrain or accelerate responses to evolutionary change depending on the particular effects of the detected gene-interactions. Here we report on an extensive quantitative trait locus (QTL) mapping study using Arabidopsis thaliana recombinant inbred lines to elucidate the genetic basis of plastic responses to changes in nitrogen (N) environments as well as the importance of epistasis to quantitative trait variation. First, we mapped main effect QTL for the slope of the local reaction norm (a measure of phenotypic plasticity) at low (N06-N01) and high (N56-N51) N-supply rates. We then tested for epistatic QTL X QTL interactions (at 1 cM resolution) for traits in four N- environments as well as for phenotypic plasticity to changes at limiting and saturating N-levels. Single QTL analysis of plasticity includes analysis of 26 traits and analyses of epistatic QTL include 78 traits. This number of traits is uncommon outside of meta-analyses. Growing plants in the same growth chambers at the University of Pittsburgh allows us to describe broad patterns of pleiotropy/linkage, phenotypic plasticity, and epistasis without the confounding effects of differences in study sites and growth conditions faced by meta-analyses. Our results indicate that loci responsible for plastic responses to N-supply rate almost exclusively co-locate with trait QTL, giving little evidence for the presence of plasticity QTL with no main effect on traits. Epistasis was common and environmentally specific, with QTL X QTL interactions present between loci both with and without detectable main effects in 1D genome scans. The implications of these patterns of genetic architecture are discussed.

5.2 INTRODUCTION

Epistasis, interactions between alleles at two loci, has the potential to affect all aspects of evolution including adaptation (Naciri-Graven and Goudet 2003), population divergence (Wade 2000, Fenster and Galloway 2000, Demuth and Wade 2007a), and speciation (Coyne and Orr 1997). Its role in evolution has been controversial (Wolf et al. 2000, and references therein), for example with some authors asserting it has little relevance to adaptation (Crow 2010) and others suggesting it can explain continued responses to long term directional selection seen in artificial selection studies (Carlborg et al. 2006, Dudley 2007). Quantifying epistasis has been notoriously difficult (Whitlock et al. 1995), often requiring large sample sizes and complex breeding designs. Methodological and technological advances (e.g. line cross methodology, Demuth and Wade 2005, 2006; the availability of recombinant inbred lines, Broman 2005, Elnaccash and Tonsor 2010, use of MCMC approaches, Walsh 2009), have offered improved ability to detect gene-interactions, particularly at the level of quantitative trait loci (QTL) mapping.

We investigated the genetic architecture of plastic responses of thirteen *Arabidopsis thaliana* traits to limiting and saturating N-supply rates. QTL and QTL x QTL epistatic interactions were mapped for traits in each of 4 nitrogen environments (1, 6, 51, and 56 ppm N as NO3) as well as for the slope of the local reaction norm at low (N06-N01) and high (N56-N51) N-supply rates. We used our data to address the following questions: Are there QTL that affect the plastic response to changes in N-supply rate? Do plasticity QTL and trait QTL colocate? How common are epistatic QTL x QTL interactions? How common are epistatic QTL interactions for which there are no corresponding main-effect QTLs? Do QTL x QTL interactions respond plastically to N-supply rates?

5.3 METHODS

Our research used a set of 158 RIL (CS22000) derived from a cross between Landsberg erecta (Ler-2, CS8581) and Cape Verdi Islands (Cvi-1, CS8580) ecotypes (Alonso-Blanco et al.1998) plus the Cvi and Ler parents (i.e. 160 genotypes). All seeds were obtained from the Arabidopsis Biological Resource Center (ABRC) of Ohio State University. Approximately three replicates of each genotype were grown at each of four Nitrogen supply rates in Conviron controlled environment growth chambers. See chapter 2 for details on growing conditions and experimental design. Briefly, plants were grown in one of four nitrogen (N) environments, chosen to represent saturating and limiting N-supply levels. These

N-environments will be referred to as N56, N51, N06, and N01 or occasionally as high (N56 and N51) and low (N06 and N01) N-environments.

Thirteen traits were measured on each plant. Traits were chosen because they have been previously shown to have genetic variation, and because they are likely affected by N-supply rates. Traits measured included total biomass (totg), proportion root/total biomass (proprt), percent nitrogen (Npct), age at first reproduction (boltd), number of rosette leaves at bolting (nlvs), instantaneous carbon assimilation rate (delco2), instantaneous transpiration rate (delh2o), photosynthetic quantum efficiency (fvfm), photosynthetic quantum yield (yield), and photosynthetic electron transport rate (dietr), total branch length (totleng), number of basal reproductive branches (rb), and total silique length (tsl). Additionally, a set of "plasticities" was calculated for each trait in high and low N-environments. Plasticity to changes at limiting N-levels was calculated as the difference between a trait-value measured in the N06 environment minus the trait value measured in N01. This will be referred to as Low-N Plasticity and abbreviated as PL(trait name). Differences between traits measured in N56 and N51 will be referred to as High-N Plasticity and abbreviated PH(trait name).

5.3.1 1D and 2D genome scans

All of our 1D and 2D genome scans for QTL were based on 160 RIL means within each of 4 N-environments. All QTL mapping analyses used R/QTL (Broman et al. 2003). RIL means were used for phenotypic input data. Marker genotypes for 163 markers with unique map positions were downloaded from The Arabidopsis Information Resource (ftp://ftp.arabidopsis.org/home/tair/Maps/Ler_Cvi_RIdata/).

Our QTL mapping used a multiple imputation algorithm created by Sen and Churchill (2001). One hundred sets of simulated marker genotypes ("Pseudomarkers") per recombinant inbred line were generated at 1 cM intervals and used as the basis of both 1D and 2D genome scans. We have previously reported on mapping of additive QTL for 52 trait-environment combinations (13 traits x 4 N-environments, chapter 2). Here we extend our analysis of 1D mapping to trait plasticities for sensitivity to changes at high and low N (13 traits x 2 plasticities). Additionally we use multiple-imputation mapping to test for QTL x QTL interactions for 52 trait-environment combinations and 26 plasticities.

Maternal cytoplasm was used as an "additive covariate" in QTL mapping to improve the power of QTL detection. The LOD threshold for QTL detection was determined by 1000 permutations (Doerge and Churchill 1996) in both 1D and 2D scans. For main effect QTL, we used 1.5 LOD support intervals to bound different QTL; if two main-effect QTL had overlapping support intervals, the QTL with the lower LOD score was removed from our analyses.

Two dimensional genome scans are far more complex than scanning for main-effect QTL. Mapping each of the 76 traits and applying permutation tests required between 220 and 340 hours of computing time using various computers with Pentium 4 processors (total >18,000 hrs). For this reason, we have made available our empirically determined 5% and 10% significance LOD thresholds for each of our trait-environment combinations in a supplemental table (available upon request). It may allow other researchers working with *Arabidopsis thaliana* RILs to estimate thresholds from our data while avoiding time consuming permutation tests.

Significance tests for QTL x QTL interactions involved calculations of the likelihoods of several QTL models at all pairwise locations at 1cM intervals. For any pair of loci, the 'null' and 'one' QTL models correspond to models of no QTL or one QTL at either position. The 'additive' QTL model allows for the independent effects of QTL at one or both test positions. The 'full' model allows for independent effects of QTL at one or both positions allowing for QTL x QTL interactions. Various LOD scores are derived from the Log10 likelihoods of these models:

LOD FULL = L (full model) – L(null model)
LOD ADDITIVE = L (additive model) – L (null model)
LOD INTERACTION = L (full model) – L (additive model)
LOD ONE = L (one model) – L (null model)

In addition to these Log10 likelihood ratios, further LOD scores comparing two- QTL models with the single-QTL model (LOD FULL VS. ONE and LOD ADDITIVE VS. ONE) give evidence for the presence of a second QTL at a pair of test positions, either allowing for or excluding the possibility of interactions (Broman and Sen 2009). Using two-QTL models (e.g. testing for QTL at two positions using LOD FULL VS. ONE and LOD ADDITIVE VS. ONE scores) can improve the power to detect small effect QTL potentially hidden by QTL of larger effect. Additionally, two-QTL models can help determine if multiple LOD peaks typically found on a single chromosome are linked QTL or artifacts of using single-QTL models to detect linked loci.

We used conservative detection criteria for identifying QTL x QTL interactions. Each pair of chromosomes was compared (including a chromosome with itself) at 1 cM intervals and the two positions that maximized LOD INTERACTION (= LOD FULL – LOD ADDITIVE) were retained. This pair of positions was then required to have both a significant LOD FULL and LOD INTERACTION or a significant LOD FULL and LOD FULL VS ONE (as determined by 1000 permutations) for these test positions.

5.4 RESULTS

5.4.1 Main-effect plasticity QTL

We found 13 additive plasticity QTL for 8 traits (table 5.1). These QTL affect the slope of the reaction norm across two nitrogen environments, either high-N plasticity (N56-N51) or low-N plasticity (N06-N01). One high-N plasticity QTL was detected for proportion root/total biomass (figure 5.1). This QTL was at the end of chromosome one, with its support interval excluding both of the proprt56QTL1 and proprt56QTL2 main effect QTL discovered previously. The remaining twelve plasticity QTL affected eight traits by altering sensitivity to limiting N-supply rates. In each of these cases, Low-N plasticity QTL had 1.5 LOD support intervals that overlapped with those of main effect trait QTL detected in either N01 or N06 (figure 5.2).

5.4.2 QTL X QTL interactions

Seventeen pairs of epistatically interacting loci were detected (table 5.2). Each of these interactions was plotted using the function Plot.Scantwo in R/QTL. Interactions plots of various LOD scores can seem difficult to interpret at first. The upper left triangles in these figures plot the scores calculated for all pairs of chromosomes at 1 cM intervals. The lower right triangles plot the LOD FULL VS ONE scores, which indicate evidence for 2 QTL either with or without interactions over and above the evidence for a single QTL at either of those positions. Peaks in the upper triangle indicate epistasis. When there is no corresponding peak in the lower triangle, this indicates an epistatic interaction without a corresponding main effect for either of the QTL involved. Plots of LOD INTERACTION with LOD ADDITIVE VS ONE can help facilitate visual detection of QTL X QTL with no main effects but were omitted for this paper and are available upon request.

2D mapping of basal branch number in N01 revealed epistasis between QTL located on chromosomes 4 and 5 (figure 5.3a). Neither of these two interacting loci were detected in 1D scans (figure 5.2). Evidence for 2 tightly linked loci that interact epistatically to affect total silique length was discovered in N01 (figure 5.3b) and for total branch length in N06 (figure 5.4). These QTL are closer than the length of the smallest 1.5 LOD support intervals estimated in 1D scans (figure 5.2).

In N51, QTL x QTL interactions between loci on chromosome 5 affected transpiration rate, carbon assimilation rate, and age at first reproduction (figure 5.5). In these cases only one of the interacting QTL was detected in our 1D mapping. In the case of age at first reproduction, the 1D support

interval was wide, suggesting the possibility of a second QTL, further suggested by the two separate QTL peaks that showed up in N01 and N06. Epistasis between tightly linked loci on chromosome 2 affected total biomass in N51. Interactions between tightly linked QTL (such as these for total biomass in N06 and N51, and total fruit length in N01), can potentially be an artifact of permutation testing procedures, particularly if there is no marker located between the two test positions (Broman and Sen 2009, p. 224). To address this, we used ANOVA to test for interactions between nearest flanking markers to these interacting QTL. In these three cases, ANOVA results showed significant interaction effects ($p = 0.0006$ or less), confirming the presence of epistasis between the tightly linked loci (data not shown).

In N56, we detected more epistatic interactions between QTL located on chromosome 5,with epistatic QTL existing for carbon assimilation rate, age at first reproduction, and total biomass (figure 5.6). Percent nitrogen was affected by epistatic QTL on chromosomes 2 and 5; the QTL on chromosome 5 had no significant main effect on percent nitrogen and was not detected in our 1D scan. Age at first reproduction also had interactions between QTL on chromosomes 1 and 5, both of which have independent effects.

Low-N plasticity QTLs were discovered for carbon assimilation rate, quantum efficiency, and total branch length (figure 5.7). In all cases, interacting QTL also had significant main effects as revealed by 1D genome scans. There was weak evidence for QTL on chromosome 1 and on chromosome 5 to interact for total branch length but this was non-significant.

5.5 DISCUSSION

We found evidence for a genetic basis for plastic responses to nitrogen supply rate. QTL for plasticity, that is for the slope of the reaction norm at either limiting or saturating N-supply rates, were primarily discovered affecting sensitivity to changes between 1 and 6 ppm NO3. Only one plasticity QTL was detected, for changes in proportion root/total biomass, across N56 and N51. This may be a 'true' plasticity QTL in the sense that it was the only one that did not colocate with the corresponding main effect QTL for proportion root/total biomass in either N56 or N51. All other plasticity QTL were located at the same genomic position as their corresponding trait QTL. The QTL that affect both the trait and trait plasticity may represent regulatory elements that alter gene expression. In the case of the one high-N plasticity QTL, the mechanism of its effects on sensitivity to environmental variation at near saturating levels is unclear. This locus may be important in canalization of phenotype against environmental variation.

Two dimensional QTL scans revealed much information missed by 1D QTL mapping approaches. Seventeen pairs of epistatically interacting loci were detected. In three cases epistasis was between loci within 2 cM of each other, suggesting extremely tight linkage. This could potentially be an artifact of permutation testing procedures, particularly if there is no marker located between the two test positions (Broman and Sen 2009, p. 224). A finer scale analysis of genes on Chromosome 2 positions 40-42 and chromosome 3 positions 46-48 that affect branch length and silique length respectively could help validate our findings. If we take our findings at face value, then 2D mapping has revealed tightly linked loci that would nearly always be detected as a single locus in 1D genome scans. Additionally, the two-QTL models we used revealed the presence of QTL with no main effect that interacted with additive QTL. This type of epistatic QTL was rare; epistatic QTL with no main effect occurred in only 3 of the 17 pairs of QTL x QTL interactions. For the most part, interacting QTL had detectable main effects.

The epistatic interactions varied considerably across N-environments. With the exception of epistatic carbon assimilation rate QTL in N56 and N51, each QTL X QTL interaction was specific to a single environment. One consistent pattern in both 1D and 2D scans is that loci on chromosome 5 affect nearly every trait and participate in many interactions, both for traits and trait plasticities. Further investigations into the genes near position 15 and position 40 on chromosome 5 could facilitate understanding of the significance of these regions.

Our study revealed complex genetic architecture for a large set of diverse traits in a set of Arabidopsis thaliana recombinant inbred lines. We found plastic changes to epistatic interactions and an epistatic basis for phenotypic plasticity. With our analyses being limited to alleles that diverged between the Landsberg erecta and Cape Verdi Islands parental ecotypes, studies involving broader sampling of Arabidopsis lines will likely find more similar or more complexities in trait architecture.

.

Table 5.1. Plasticity QTL and related trait QTL. PH indicates plasticity QTLs for sensitivity to changes at saturating N levels (N56-N51) and PL indicated plasticity QTLs for sensitivity to limiting N-levels (N06-N01). Below plasticity QTL, the corresponding QTL if detected in relevant environments.

Trait-env	LCI-marker	LCI-pos	QTL-mark	QTLchr	POS	UCI-mark	UCI-pos
PHproprt	EC.88C	84	c1.loc100	1	100	FD.90L-Col	118
proprt56QTL1	PVV4	0	GB.206L/211C-Col	1	11	c1.loc19	19
proprt56QTL2	CH.145L-Col/150C	49	c2.loc52	2	52	c2.loc58	58
PLdelco2	c5.loc32	32	c5.loc41	5	41	c5.loc45	45
PLdelh2o	c5.loc28	28	c5.loc44	5	44	c5.loc50	50
PLfvfm	c2.loc62	62	c2.loc69	2	69	EC.235L-Col/247C	71
PLnlvs	c5.loc1	1	BH.144L	5	8	AD.114C-Col	24
PLrb	c3.loc65	65	DF.65L-Col	3	70	GB.97L-Col/99C	79
PLrb	AD.114C-Col	24	GH.473C	5	34	c5.loc45	45
PLtotg	BF.269C	20	c5.loc26	5	26	BF.164C-Col	30
PLtotg	BF.164C-Col	30	c5.loc33	5	33	c5.loc49	49
PLtotlen	FD.150C	45	CH.145L-Col/150C	2	49	BH.195L-Col	53
PLtotlen	c4.loc72	72	c4.loc82	4	82	BH.342C/347L-Col	85
PLtotlen	BH.325L	15	c5.loc28	5	28	GH.121L-Col	40
PLtsl	c5.loc21	21	c5.loc27	5	27	GH.117C	35
delco206QTL1	BF.164C-Col	30	c5.loc38	5	38	AD.129L-Col	43
delh2o06QTL1	c5.loc29	29	GH.473C	5	34	c5.loc42	42
fvfm01QTL1	DF.140C	63	c2.loc70	2	70	EC.235L-Col/247C	71
nlvs01QTL1	c5.loc27	27	c5.loc33	5	33	c5.loc45	45
nlvs01QTL2	EC.198L-Col	13	BH.325L	5	15	BH.107L-Col	19
nlvs06QTL1	c5.loc10	10	BH.325L	5	15	AD.129L-Col	43
rb06QTL1	c3.loc67	67	c3.loc71	3	71	GB.97L-Col/99C	79
rb06QTL2	c4.loc71	71	c4.loc82	4	82	BH.342C/347L-Col	85
rb06QTL3	c5.loc27	27	GH.473C	5	34	c5.loc48	48
totg01QTL1	CH.284C	22	CH.65C	2	30	c2.loc41	41
totg01QTL2	c2.loc44	44	c2.loc68	2	68	EC.235L-Col/247C	71
totg01QTL3	c5.loc9	9	BH.107L-Col	5	19	BH.96L-Col	52
totlen01QTL1	CH.200C	78	EC.88C	1	84	c1.loc92	92
totlen01QTL2	GD.460L-Col	47	CH.145L-Col/150C	2	49	c2.loc55	55
totlen06QTL1	GD.460L-Col	47	CH.145L-Col/150C	2	49	c2.loc52	52
totlen06QTL2	c4.loc71	71	c4.loc83	4	83	BH.342C/347L-Col	85
totlen06QTL3	BH.325L	15	c5.loc32	5	32	AD.129L-Col	43
tsl01QTL1	GD.460L-Col	47	DF.140C	2	63	EC.235L-Col/247C	71
tsl01QTL2	c2.loc25	25	CH.65C	2	30	FD.85C	40
tsl01QTL3	EC.198L-Col	13	c5.loc18	5	18	c5.loc29	29
tsl06QTL1	c5.loc21	21	c5.loc28	5	28	GH.117C	35

Table 5.2. Epistatic QTL, LOD scores, and significance thresholds. Table shows trait-environment combinations experiencing epistatic QTL x QTL interactions. Columns are trait-environment combinations, chromosome number, corresponding chromosome position in cM, LOD scores for the 'full model', 'full vs 1', 'interaction', 'additive', and 'additive vs 1' models, followed by the 5% significance threshold for each LOD score as determined by 1000 permutations. See Methods for trait abbreviations.

		pos1	pos2	lod.full	lod.fv1	lod.int	lod.add	lod.av1	full	fv1	int	add	av1	one
rb01	c4:c5	68	84	6.1	4.04	4.26	1.84	-0.22	5.59	4.29	3.49	4.29	2.7	2.59
tsl01	c3:c3	46	48	31.1	29.6	28	3.11	1.55	26.7	25.8	19.7	12	11.1	2.37
totlen06	c2:c2	41	42	10.2	3.85	5.92	4.26	-2.06	8.66	7.75	5.68	5.24	4.05	2.5
boltd51	c5:c5	14	41	19.2	10.8	7.22	12	3.56	11.3	10.3	7.23	6.33	5.25	2.51
delco251	c5:c5	15	44	10.8	6.4	7.58	3.22	-1.18	6.54	5.24	3.84	4.72	2.9	2.71
delh2o51	c1:c5	126	46	5.63	0.063	3.71	1.92	-3.64	5.48	4.07	3.36	4.37	2.68	2.55
	c5:c5	15	53	6.62	1.052	4.94	1.68	-3.88	5.48	4.07	3.36	4.37	2.68	2.55
totlen51	c2:c2	40	42	8.73	1.01	4.89	3.84	-3.88	5.97	4.65	3.66	4.41	2.64	2.63
totg56	c5:c5	17	40	8.36	4.59	3.71	4.65	0.888	5.92	4.61	3.52	4.41	2.73	2.6
Npct56	c2:c5	59	97	11.7	-6.46	3.66	8.06	-10.1	5.69	4.1	3.3	4.41	2.49	2.59
boltd56	c1:c5	126	47	15.1	2.81	5.08	9.98	-2.27	6.77	5.42	4.03	4.5	2.91	2.57
	c5:c5	17	39	26.5	14.26	9.67	16.84	4.59	6.77	5.42	4.03	4.5	2.91	2.57
delco256	c5:c5	20	40	8.74	3.56	4.62	4.12	-1.06	6.29	5.19	3.92	4.61	3.15	2.51
totlen56	c2:c4	51	85	17.6	7.3	4.93	12.7	2.37	5.92	4.57	3.53	4.52	2.74	2.66
PLdelco2	c5:c5	32	59	7.25	2.41	4.29	2.96	-1.87	6.27	4.93	3.67	5.01	3.45	2.66
PLfvfm	c2:c5	67	77	7.05	4.17	4.24	2.81	-0.0676	6.79	5.44	4.05	5.18	3.62	2.59
PLtotlen	c2:c4	53	82	10.3	6.29	4.21	6.05	2.07	7.35	6.2	4.37	4.76	3.45	2.58

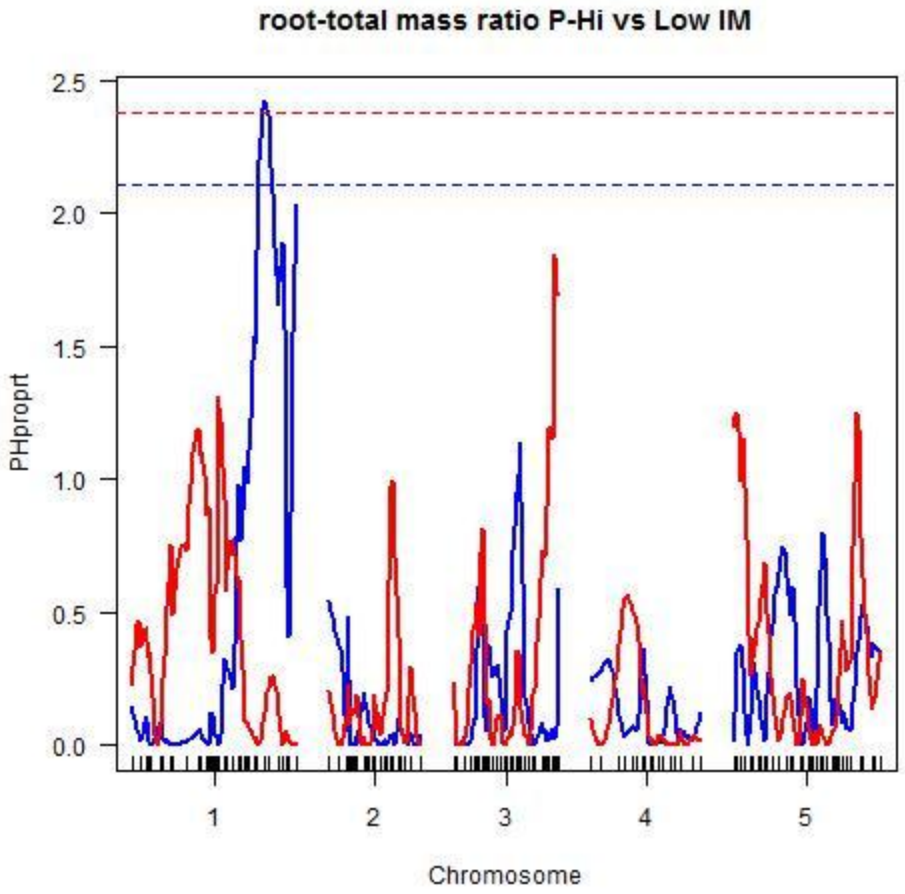


Figure 5.1. LOD profile for low- and high-plasticity (red and blue, respectively) for proportion root/total biomass investment. Significance thresholds were determined through 1000 permutations.

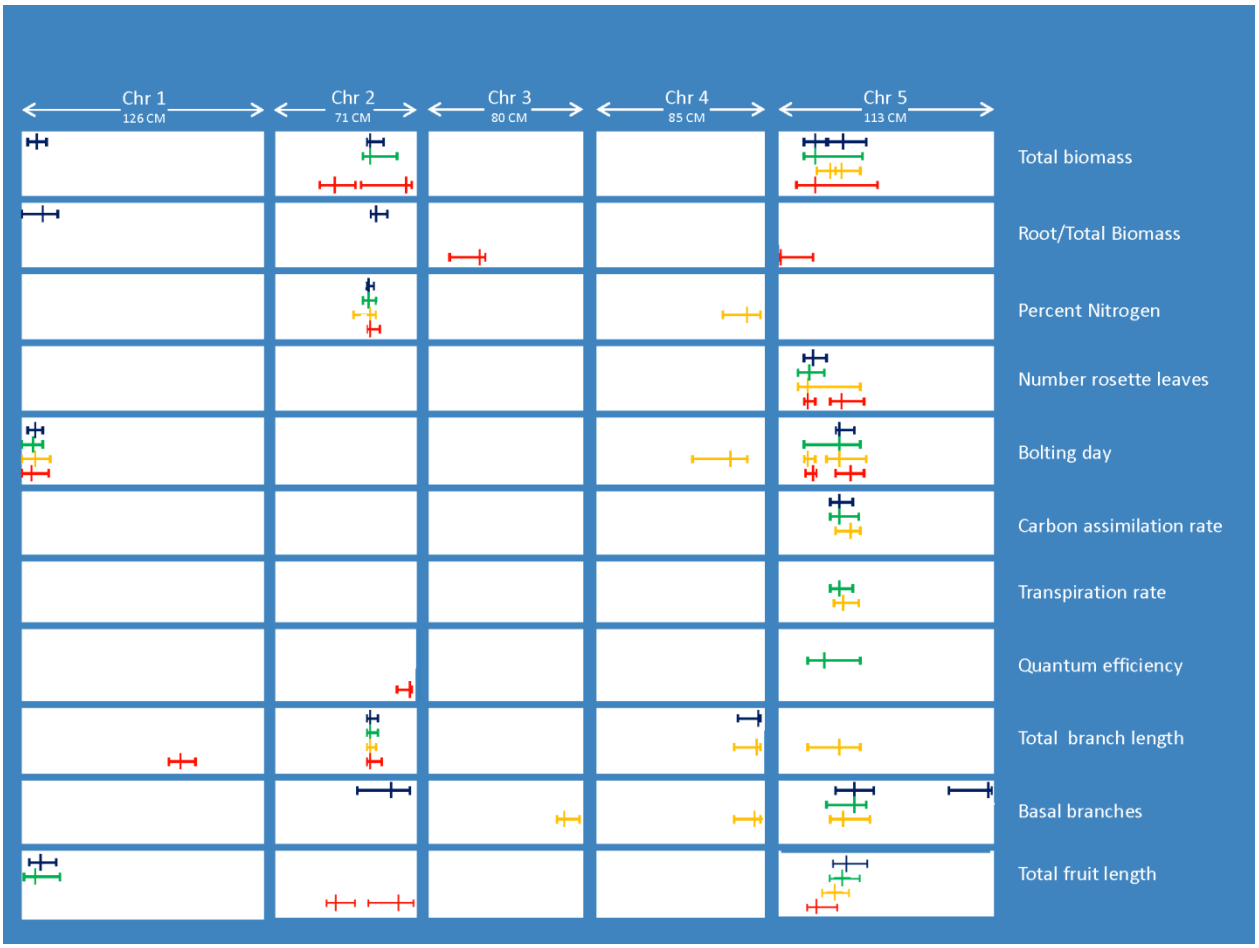
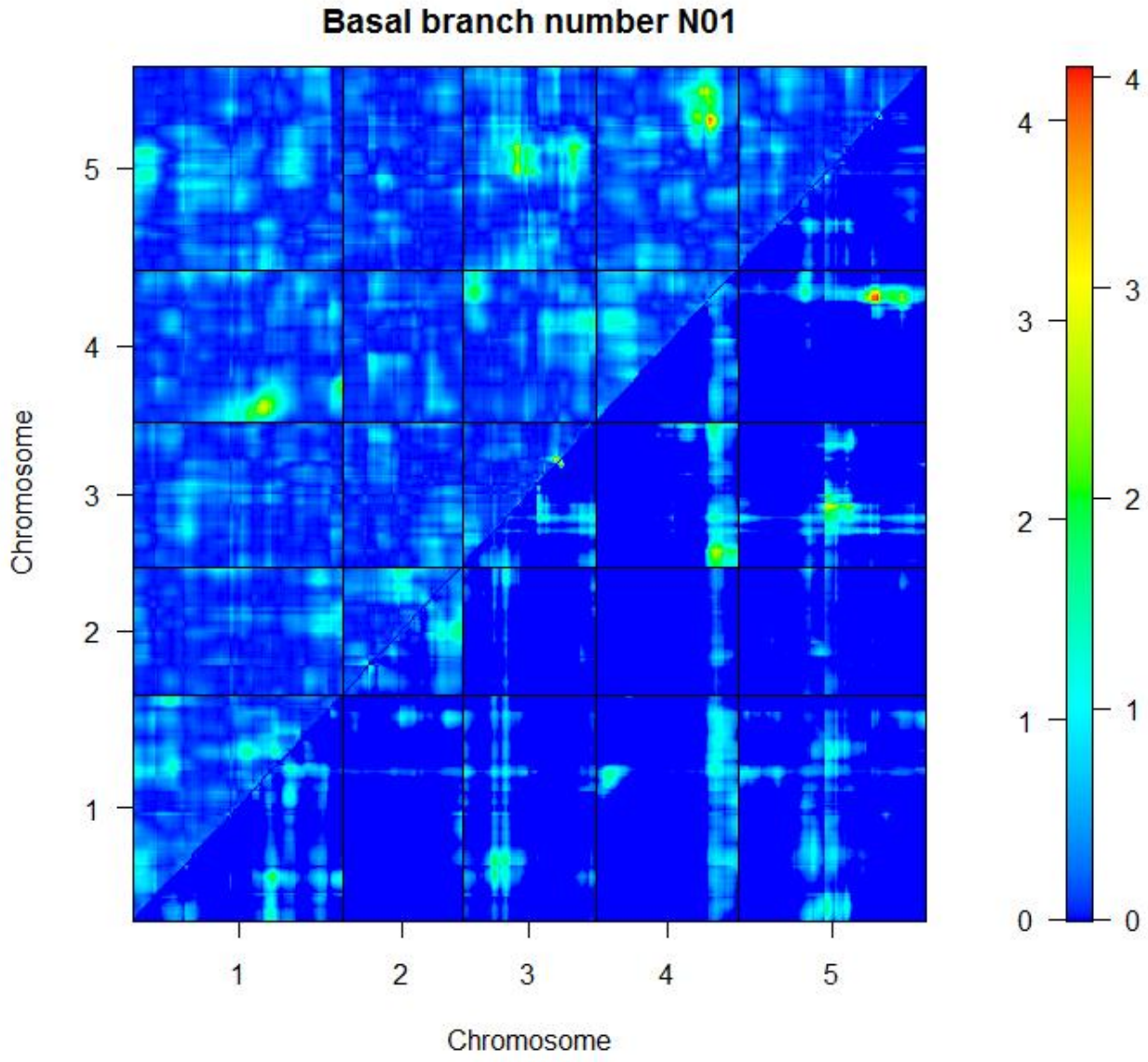


Figure 5.2. 1D QTL summary reprinted from chapter 3 for comparison.

a)


Basal branch number N01
Chromosome
Chromosome
1
2
3
4
5
0
1
2
3
4

b)

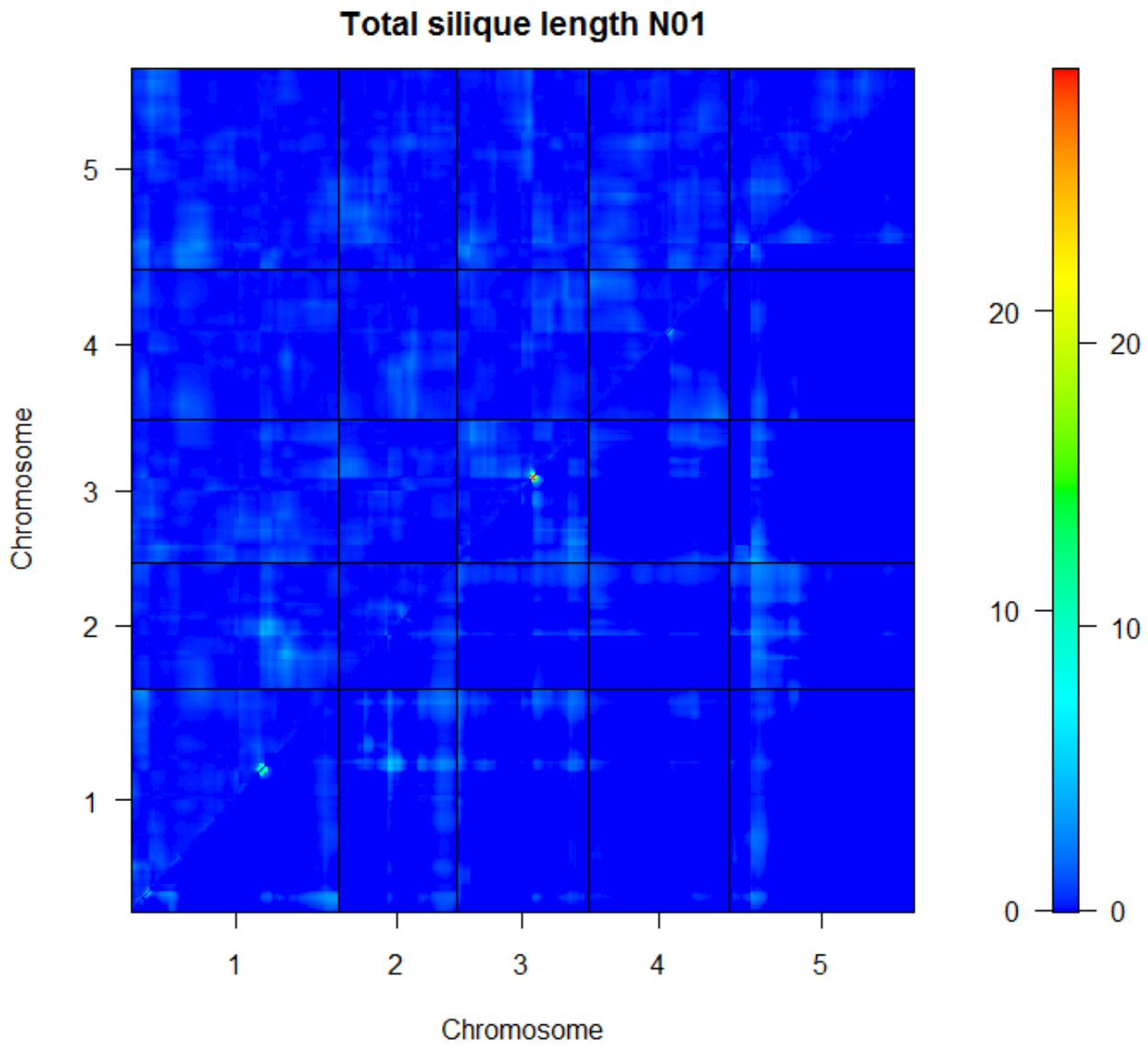


Figure 5.3. QTL X QTL interactions in N01. The upper left triangle plots LOD interaction scores and the lower right triangle plots LOD full vs one scores. Left-hand scale corresponds to LOD interaction and right hand scale corresponds to LOD full vs. one. (a) Basal branch number (b) total silique (i.e. fruit) length.

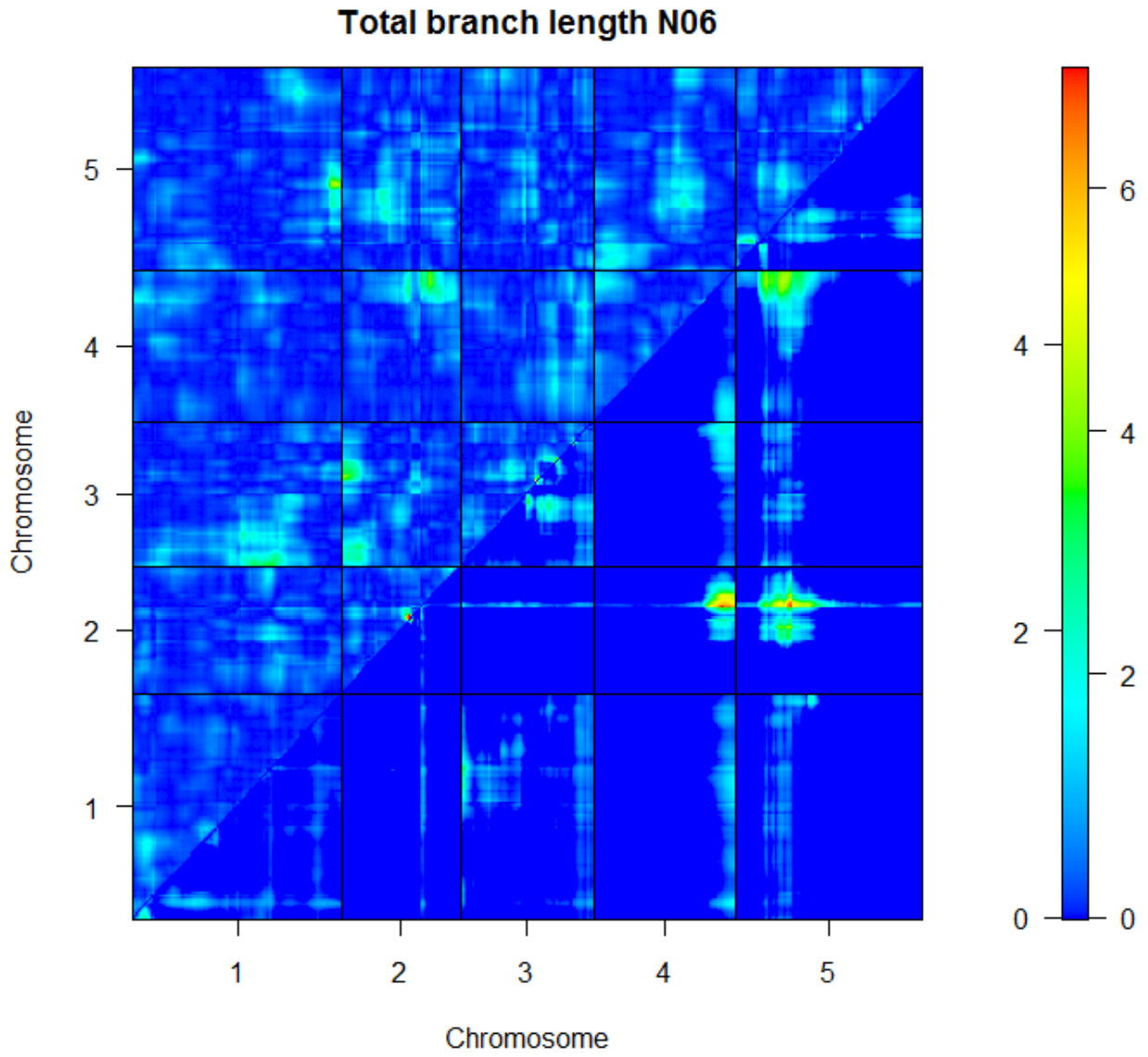
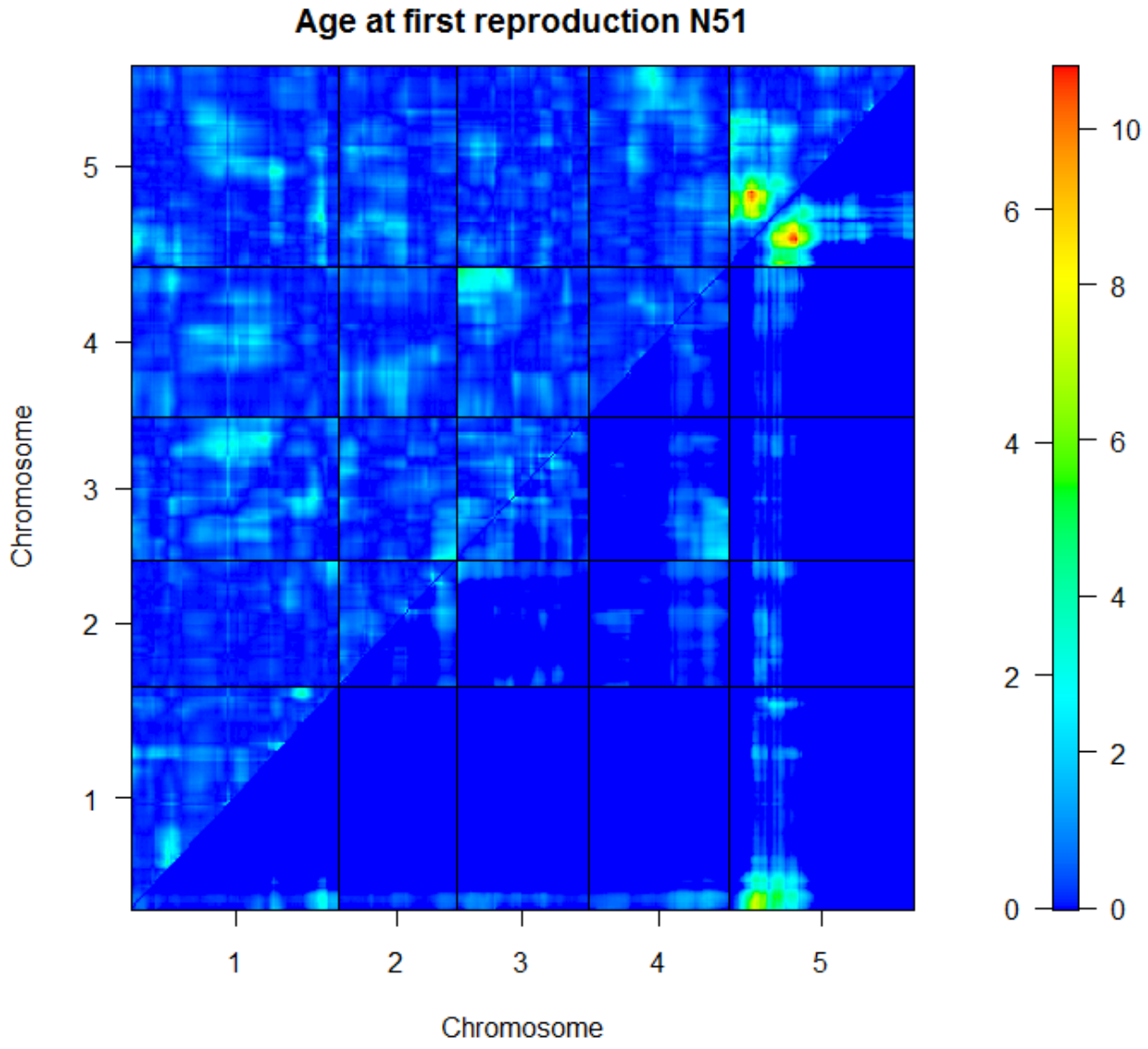
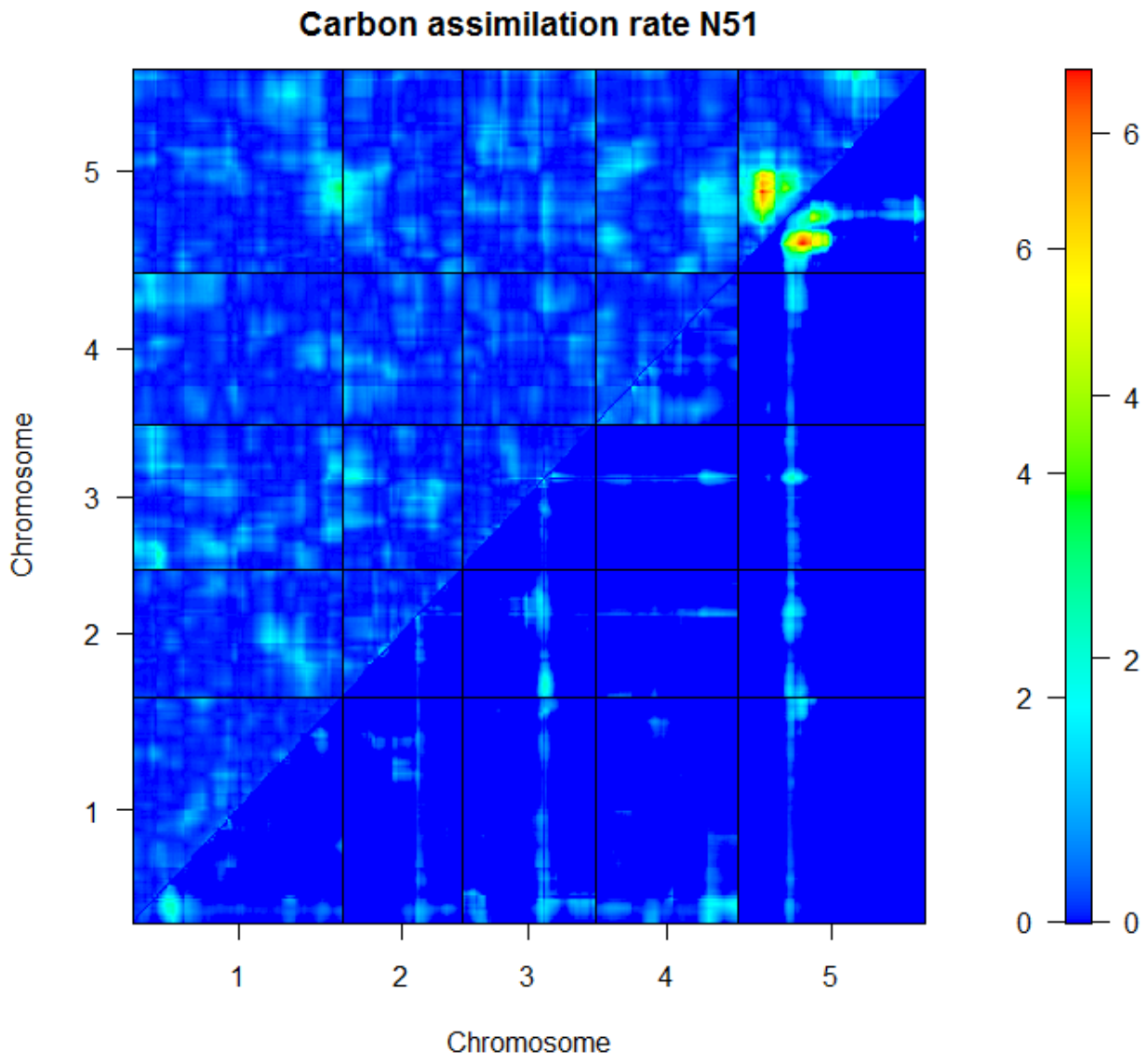


Figure 5.4. QTL X QTL interactions in N06. The upper left triangle plots LOD interaction scores and the lower right triangle plots LOD full vs one scores for Total Branch Length. Left-hand scale corresponds to LOD interaction and right hand scale corresponds to LOD full vs. one.

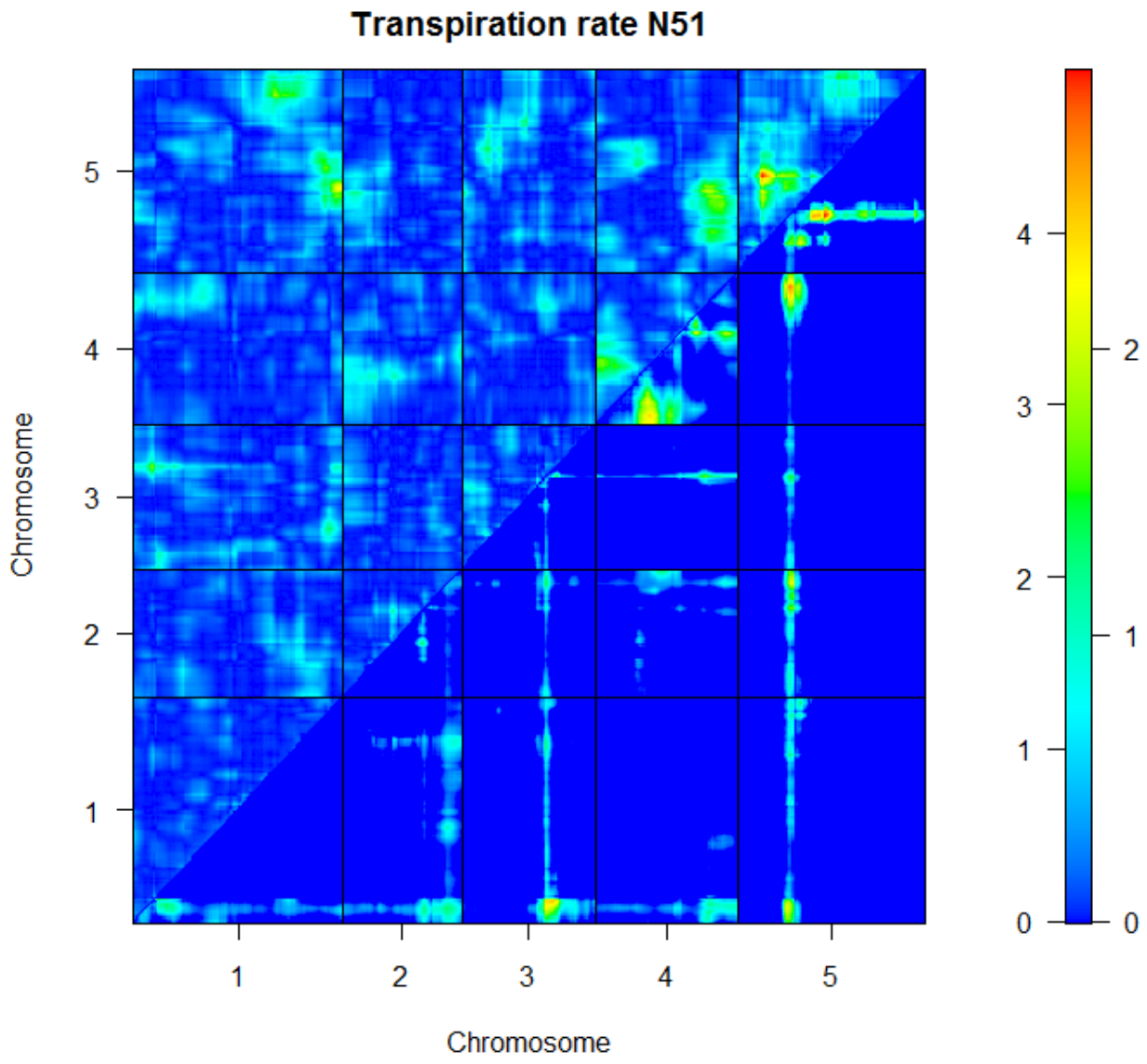
a)


Age at first reproduction N51
Chromosome
Chromosome
1
2
3
4
5
0
2
4
6
8
10

b)


Carbon assimilation rate N51
Chromosome
Chromosome
1
2
3
4
5
0
2
4
6

c)


Transpiration rate N51
Chromosome
Chromosome
1
2
3
4
5
0
1
2
3
4

d)

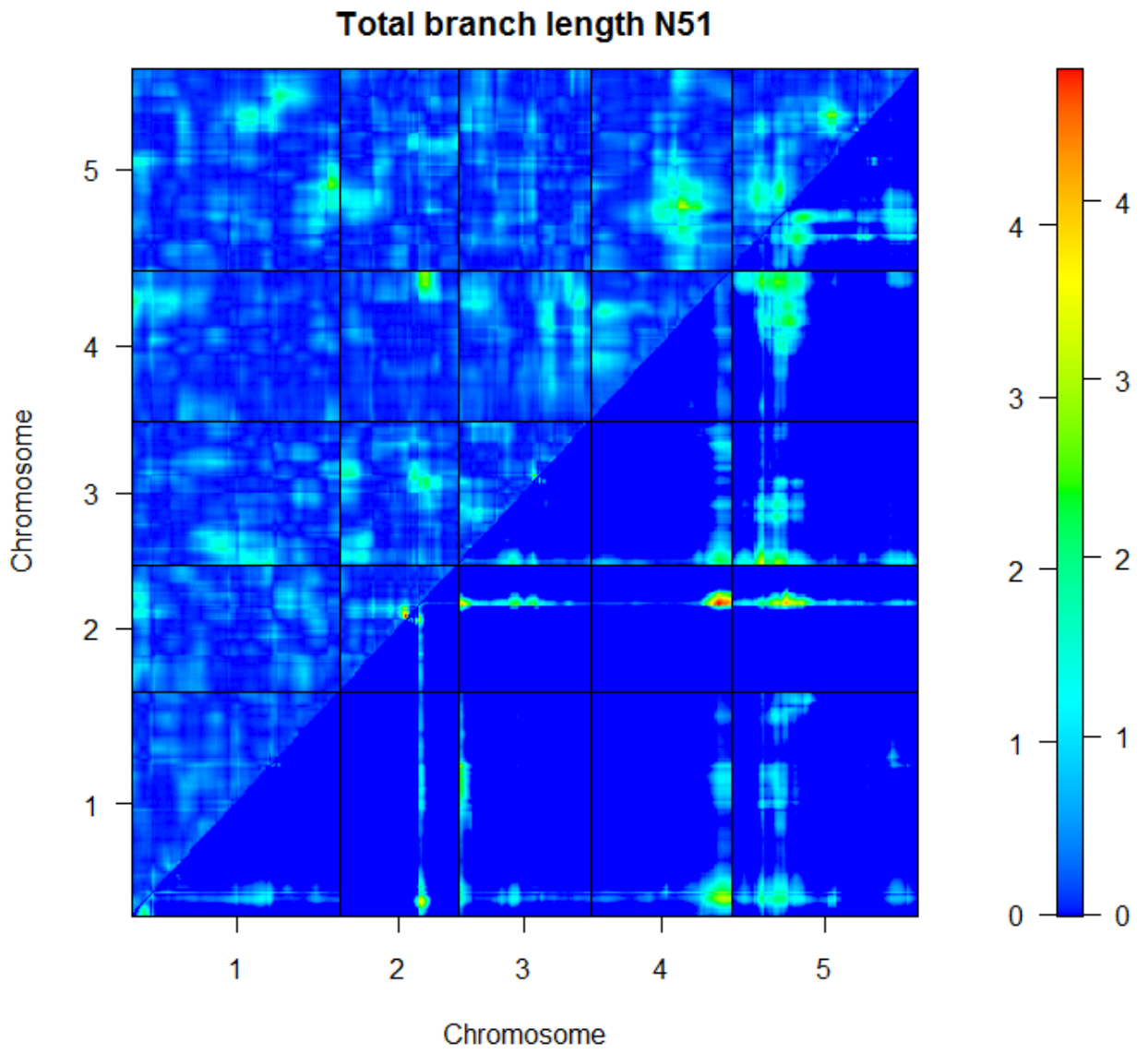
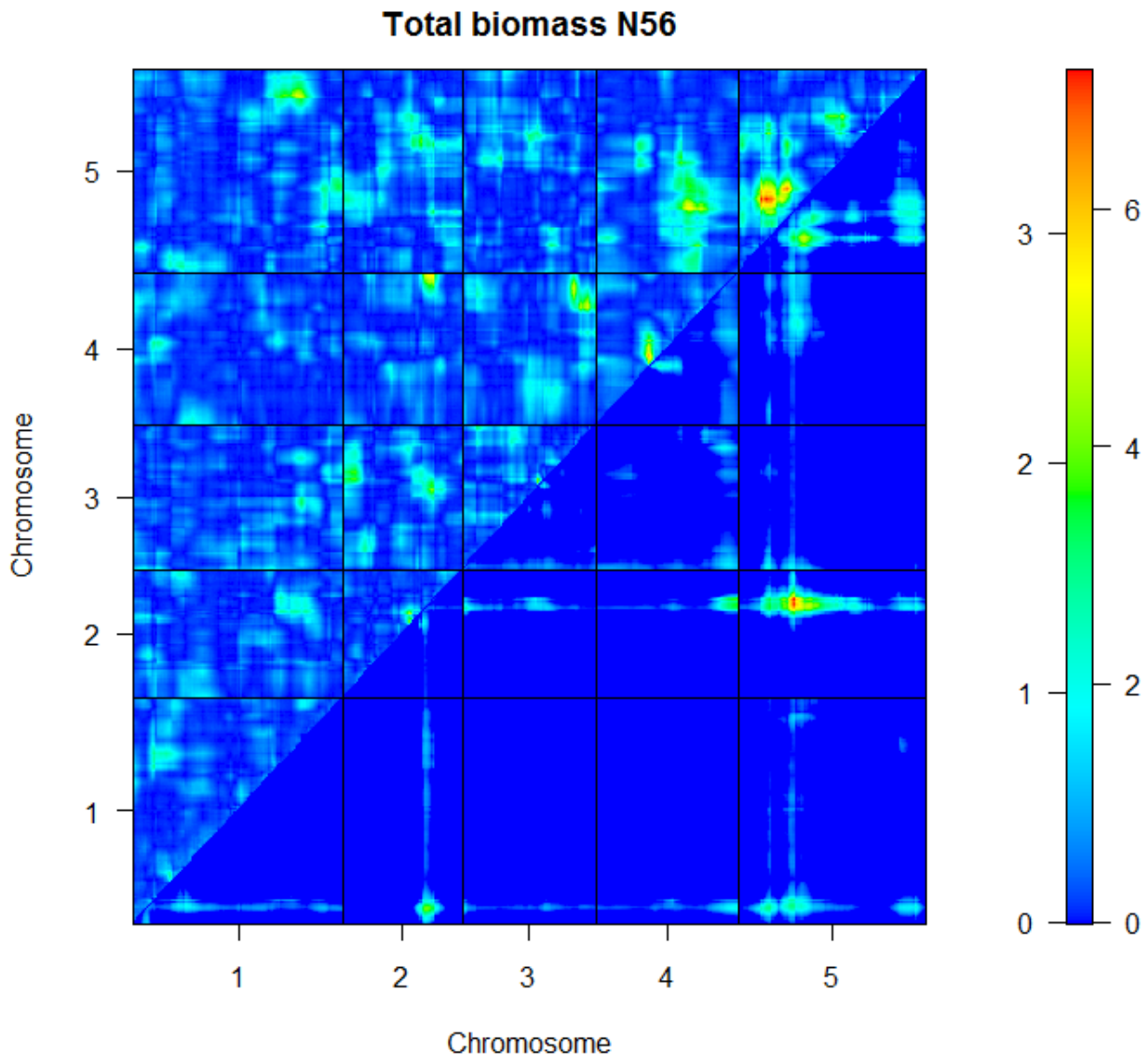
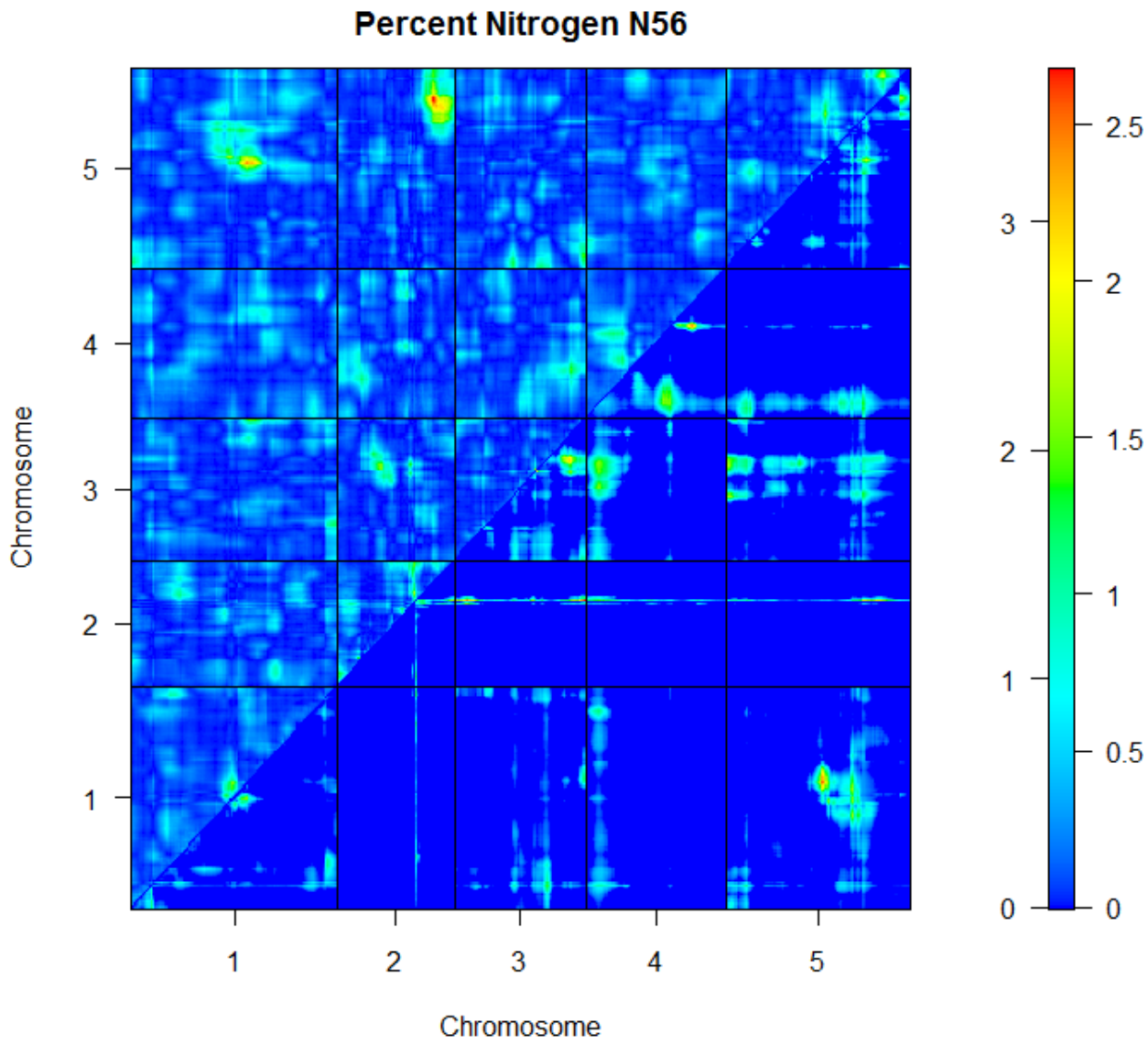


Figure 5.5. QTL X QTL interactions in N51. The upper left triangle plots LOD interaction scores and the lower right triangle plots LOD full vs one scores. Left-hand scale corresponds to LOD interaction and right hand scale corresponds to LOD full vs. one. (a) Age at first reproduction (b) carbon assimilation rate (c) transpiration rate (d) total branch length.

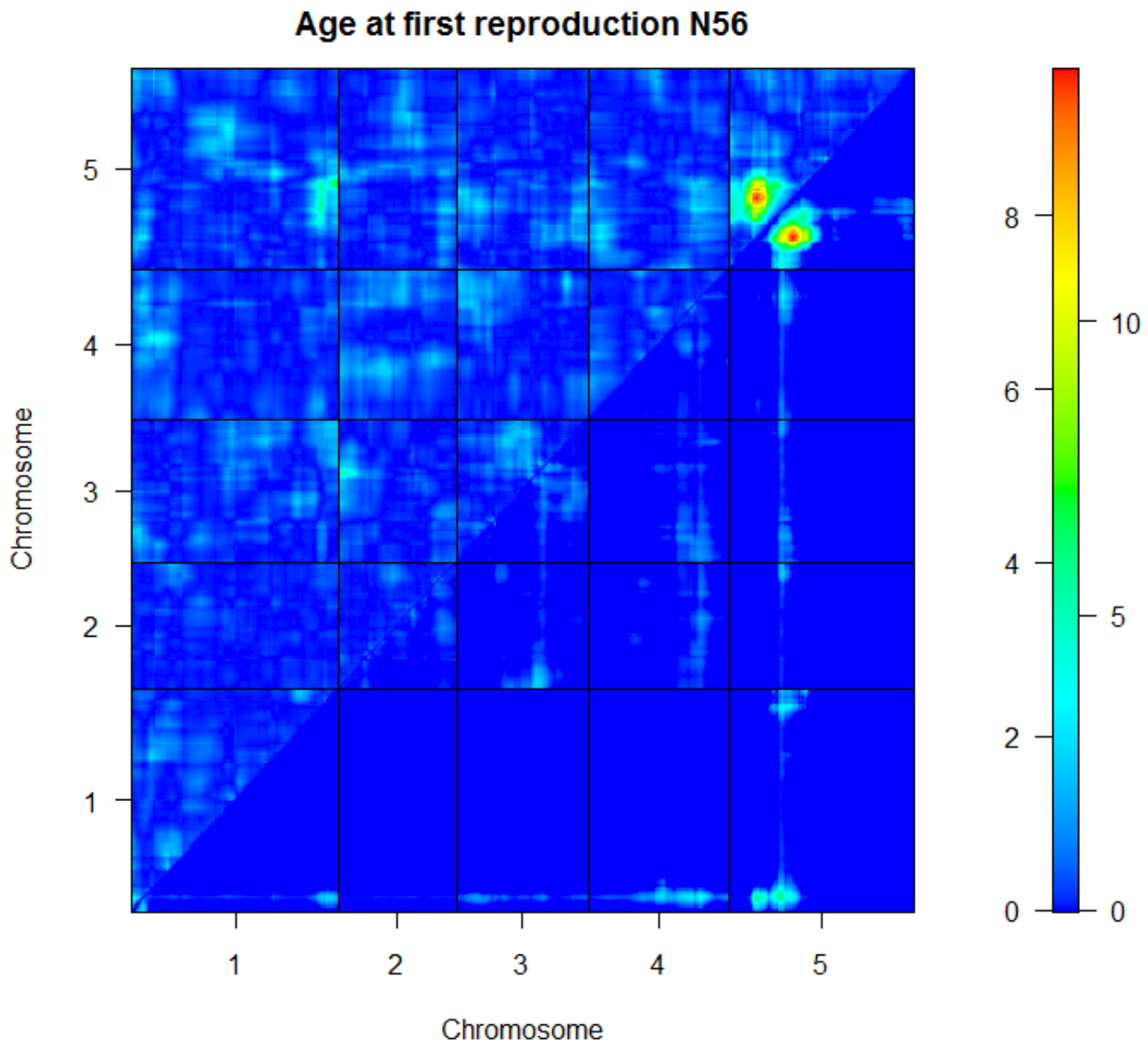
a)



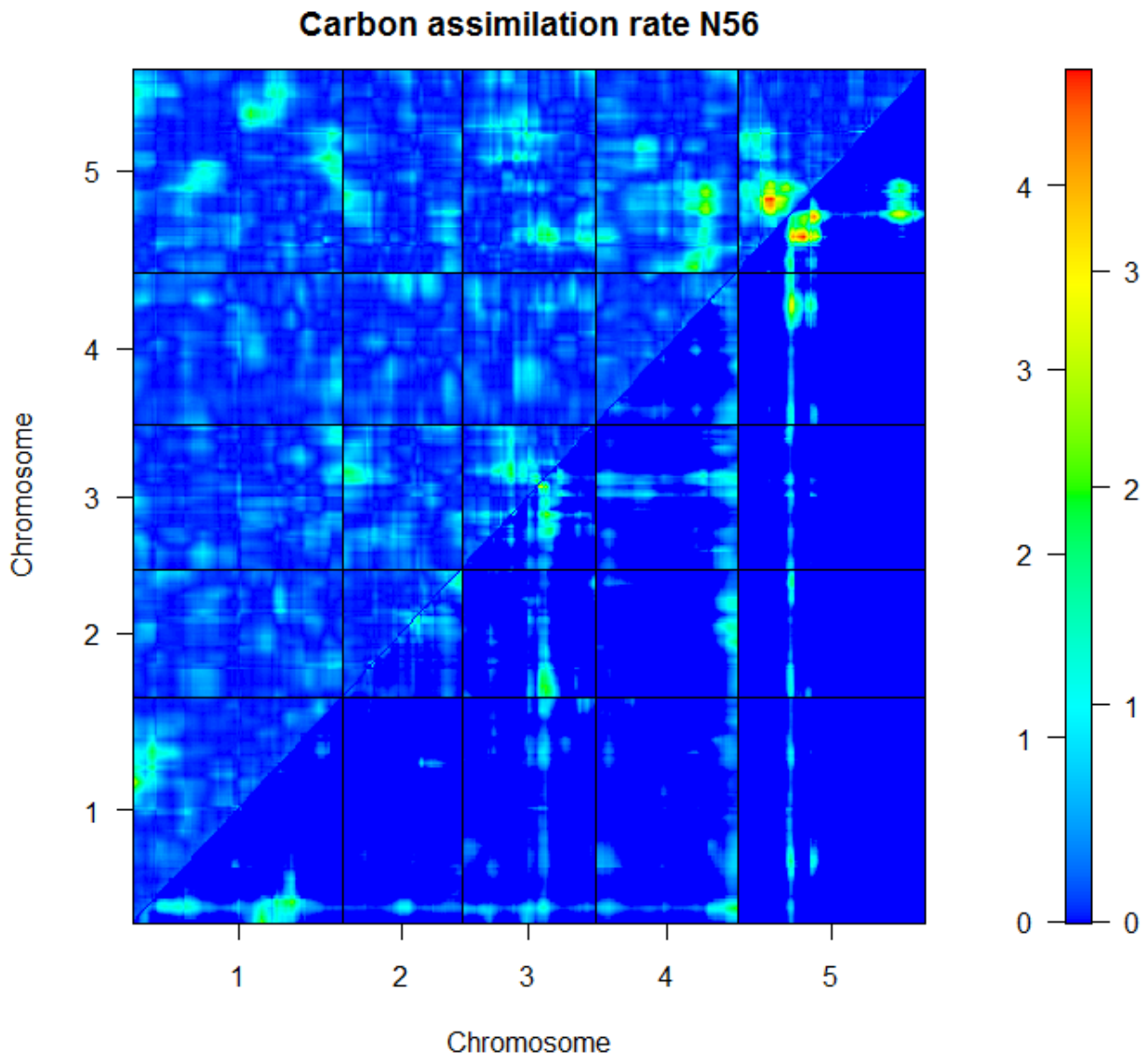
b)


Percent Nitrogen N56
Chromosome
Chromosome
1
2
3
4
5
0
0.5
1
1.5
2
2.5

c)


Age at first reproduction N56
Chromosome
Chromosome
1
2
3
4
5
0
2
4
6
8
10

d)


Carbon assimilation rate N56
Chromosome
Chromosome
1
2
3
4
5
0
1
2
3
4

e)

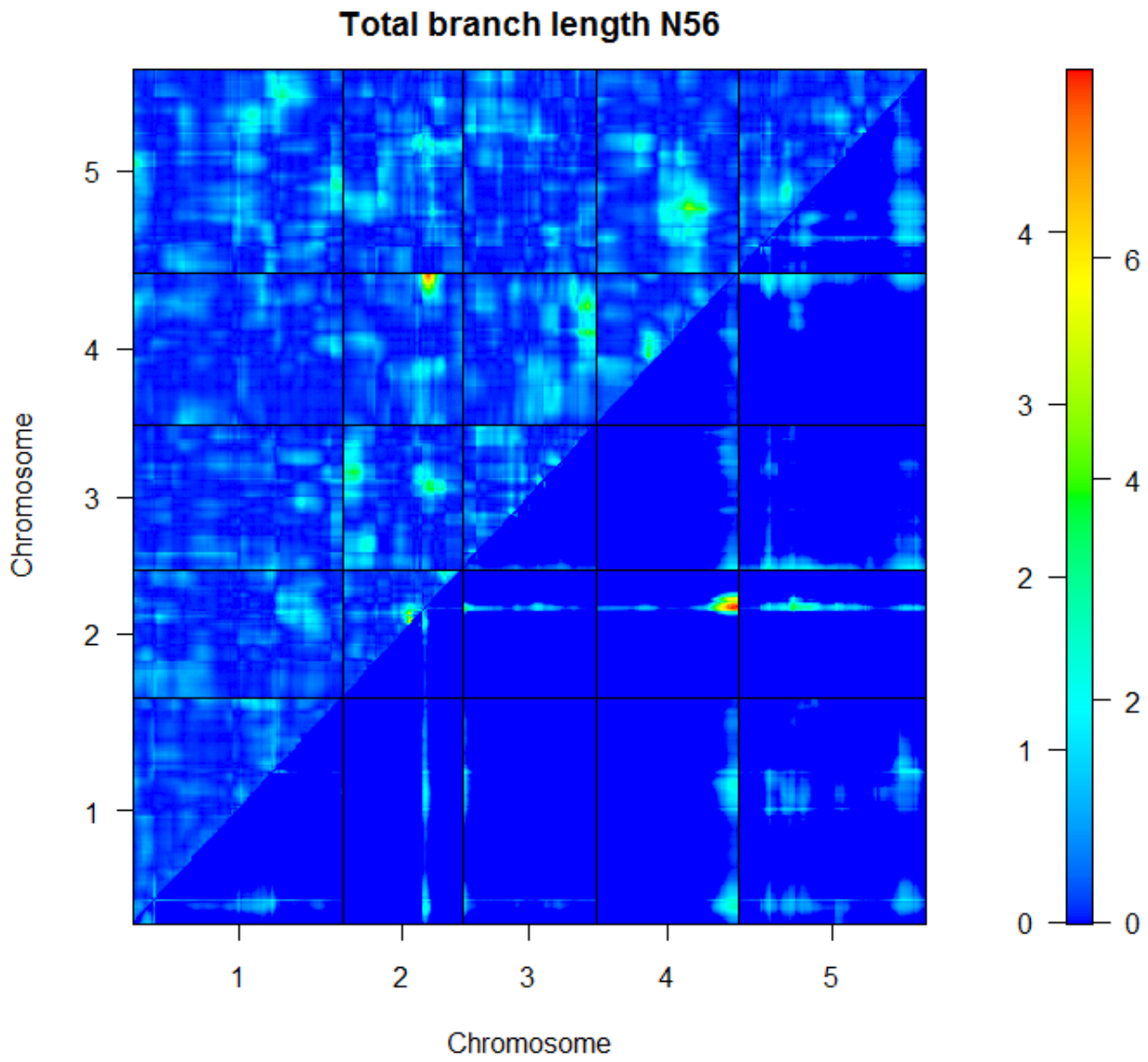
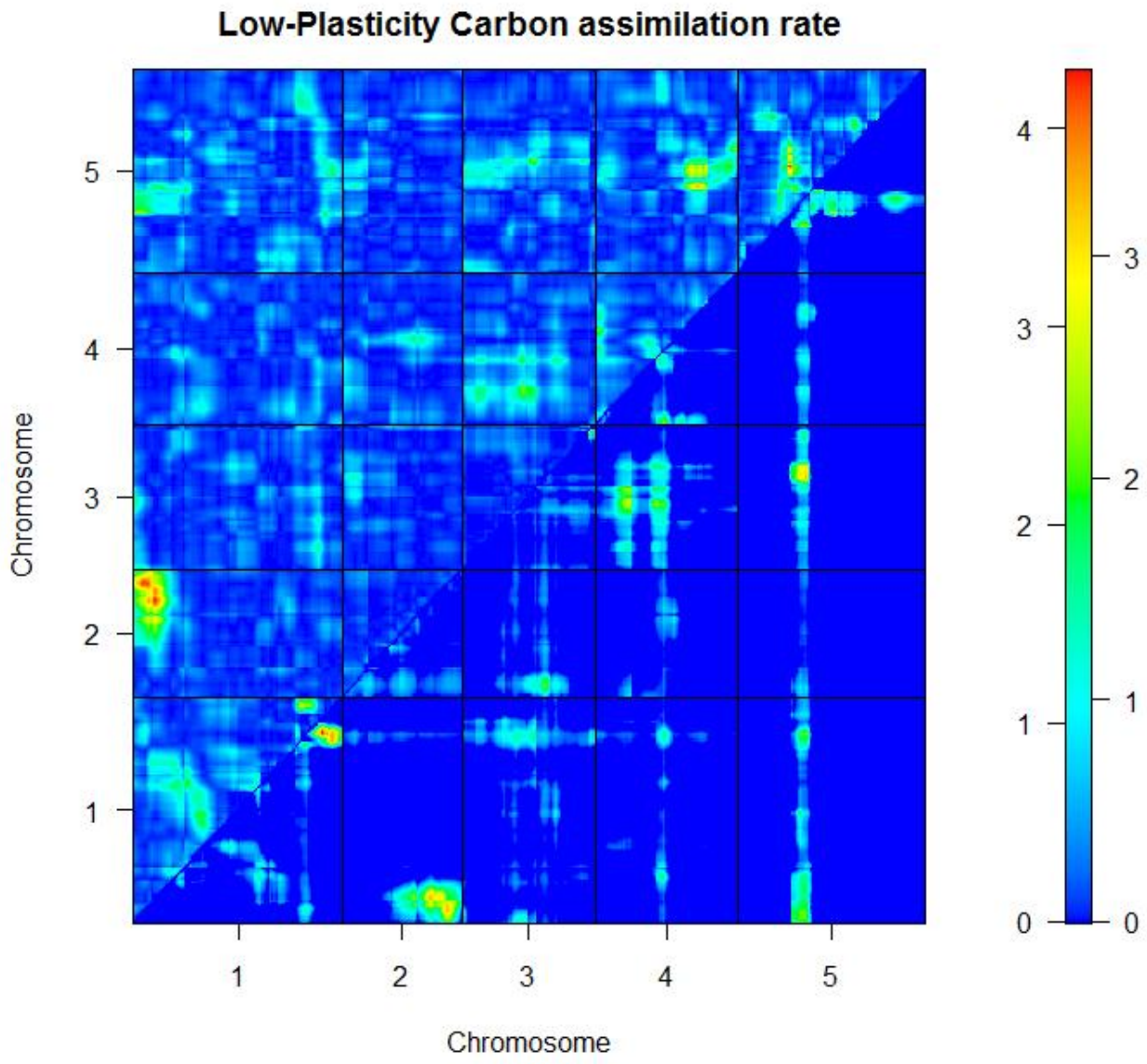
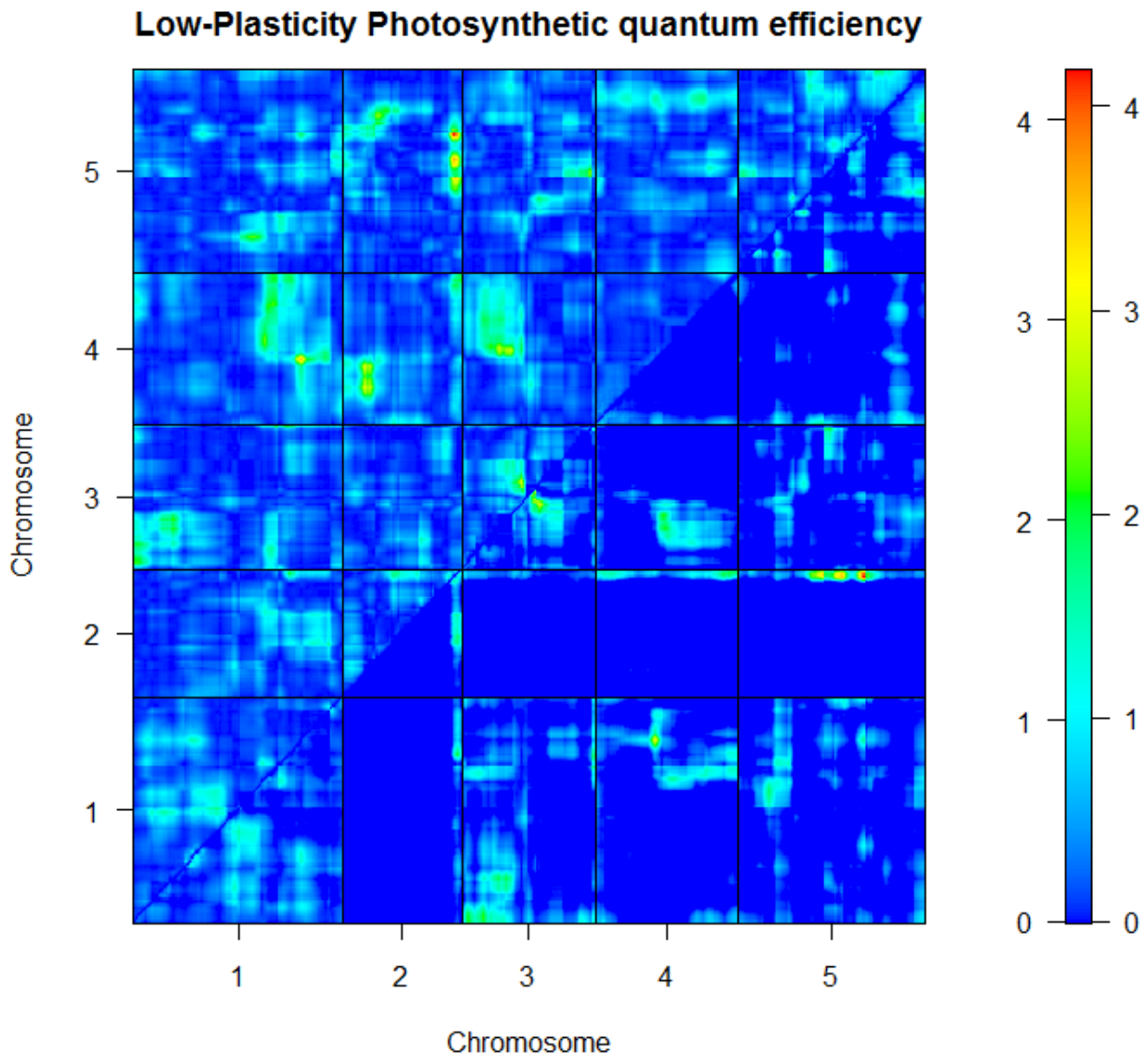


Figure 5.6. QTL X QTL interactions in N56. The upper left triangle plots LOD interaction scores and the lower right triangle plots LOD full vs one scores. Left-hand scale corresponds to LOD interaction and right hand scale corresponds to LOD full vs. one. (a) Total biomass (b) percent nitrogen (c) age at first reproduction (d) carbon assimilation rate (e) total branch length.

a)


Low-Plasticity Carbon assimilation rate
Chromosome
Chromosome
1
2
3
4
5
0
1
2
3
4

b)


Low-Plasticity Photosynthetic quantum efficiency
Chromosome
Chromosome
1
2
3
4
5
0
1
2
3
4

c)

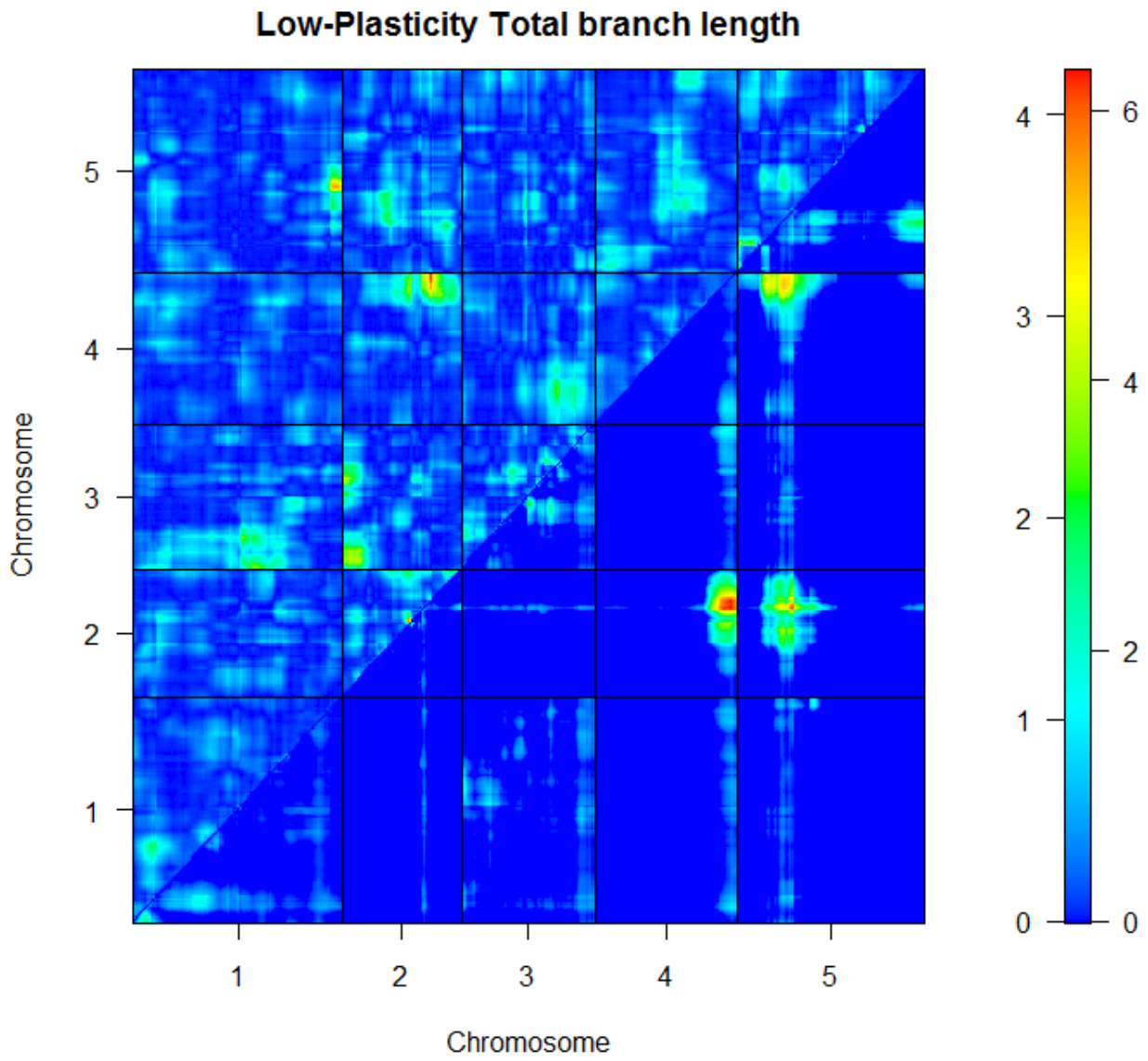


Figure 5.7. QTL X QTL interactions for Low-Plasticity. The upper left triangle plots LOD interaction scores and the lower right triangle plots LOD full vs one scores. Left-hand scale corresponds to LOD interaction and right hand scale corresponds to LOD full vs. one. (a) Carbon assimilation rate (b) photosynthetic quantum efficiency (c) total branch length

6.0 CONCLUSIONS

Improving our understanding of evolvability will be greatly facilitated by analyses at multiple levels of biological organization across several ecologically relevant environments. In my research, descriptions of genetic architecture produced different pictures of evolvability depending on the level of biological organization analyzed, and these pictures changed within different nitrogen (N) environments. At the level of the quantitative trait locus (QTL), genetic architecture of traits is often complex and environmentally specific, with QTL frequently changing location and magnitude across N- environments (chapter 4). Epistatic QTL were common and did not always have main effects detectable in 1D genome scans. Additionally, QTL affected plastic responses to changes at limiting N- levels. With a genetic basis to plasticity, reaction norms can evolve and potentially offer short term solutions to tracking environmental fluctuations. Two-dimensional exhaustive searches for epistatic QTL are computationally demanding and time consuming but worth the effort as they reveal aspects of genetic architecture that are not otherwise detectable. One approach taken by researchers is to test for QTL X QTL interactions only between loci that have displayed main effects. As I have shown, this would result in an underestimate of the importance of epistasis in quantitative genetic architecture. An alternative approach to measuring epistasis is to use line cross analysis. When these include recombinant inbred lines, power to detect epistasis (and dominance) is greatly improved (chapter 1).

Analyses at the level of the individual traits revealed heritable variation in most characters, suggesting each has the potential to evolve by natural selection. Had my research stopped with these univariate analyses, several features of genetic architecture and evolvability that suggested potential constraints and trade-offs would have been missed.

Analyses at the level of covariances and correlations between traits suggested the presence of constraints and trade-offs. Bivariate trait correlations showed that at most N-levels, between-trait and cross-environment correlations were moderate to large, potentially offering constraints (or accelerants depending on the direction of selection) to adaptation (chapter 2). The lack of cross-environment genetic correlation between N-stressed (N01) and higher N-environments (N06, N51, N56) suggest the constraining effects of genetic covariances are altered when plants are N-stressed and that adaptation can occur independently in high and low N-environments.

When sets of traits were jointly analyzed through principle component analyses, additional features of evolvability were detected (chapter 2). A decrease in the magnitude and a shift in direction of the genetic lines of least resistance, g-max, suggested that plants in N-stressed populations will respond more slowly to selection but are not constrained to follow the same evolutionary direction as in the three higher N-environment populations. Additionally, the number of significant G-matrix dimensions, as determined either by Kirkpatrick's estimator, $n_{D,}$ or randomization procedures, suggests that the abundant genetic variation present in each trait is concentrated in only one to three axes in multivariate trait space. This means that there are many possible trait combinations for which there is no genetic variation. These phenotypic combinations cannot be achieved easily through selection and could be described as unattainable because of absolute constraints.

Measures of constraint as determined by the angle between g-max and phenotypic selection gradient further suggested that constraints to adaption can change rapidly with shifts in the environment These analyses call into question the utility of g-max as a measure of biases in responses to selection. For genetic covariance structures with only one significant eigenvector, g-max is likely an adequate descriptor of evolutionary bias. For covariances structures with three significant eigenvectors, as in N01, it is not clear that g-max will constrain adaptation, particularly because all significant eigenvectors explained a similar amount of genetic variation.

These constraints due to changes in in N-availability map to population differences in evolutionary trajectories as determined by random skewers approaches (chapter 3). Random skewers approaches integrate both detected and undetected constraints to evolutionary trajectories (as they are based on the entire G-matrix structure) and allowed me to ask whether populations with the different covariance structures are likely to evolve in different directions. Consistent with my analyses of constraints, the N-stressed population (N01) took a unique evolutionary trajectory. The novel approaches I used allowed me to decompose evolutionary trajectories into effects of covariances and main effect QTL (chapter 3). Random skewers analyses adjusted for genetic covariances suggested that both the genetic variances of traits as well as the covariances between traits are responsible for divergent responses to selection in the N01 population. Random skewers analyses adjusted for main-effect QTL suggest that QTL do not alter the direction of evolutionary trajectory (but likely affect the rate of evolutionary change). The role of epistatic QTL in evolutionary trajectory differences remains unknown.

Taken as a whole, my analyses emphasize the importance of looking at multiple levels of biological organization to understand genetic architecture and evolvability. Looking at individual traits in isolation, or in a single nitrogen environment, would have produced an impression of evolvability that would be severely lacking. Analyses limited to this level would suggest that the set of traits can evolve easily in all environments, although more slowly in N01 due to reduced levels of genetic variation. With

my multi-trait, mult-environment analyses, it is clear that N-stressed plants may evolve more slowly but have fundamentally different constraints to adaptation and have the potential to evolve in different directions. Additionally, assuming an additive genetic architecture and mapping only main effect QTL would miss a large number of loci with no main effect or miss the resolution of single QTL into interactions between tightly linked loci. Further analysis of the data produced in this dissertation will allow me to connect QTL information to evolutionary trajectories in mechanistic models of evolution. This will improve our understanding of the evolutionary process

APPENDIX A

A SAS MACRO PROGRAM TO CREATE 10 000 BOOTSTRAPPED MEAN STANDARDIZED G-MATRICES WITHIN EACH N-ENVIRONMENT.

```
/* I have 6 permanent SAS data sets: N1, N6, N51, N56. They have the RIL
means for each nitrogen level; Mgmatrix are the RIL means for all 4 N-levels
in the same data set; Modcvlclean has individual plants, not RIL means (will
need this for selection analysis). Access permanent data by "DATA
LerCviRI.n1"
      proc print data=LerCviRI.Mgmatrix; run;
      proc print data=LerCviRI.n56; run;
      proc print data=LerCviRI.n51; run;
      proc print data=LerCviRI.n6; run;
      proc print data=LerCviRI.n1; run;
      */
/*This program creates &ITERATION bootstrapped datasets from N56, calculates
the mean standardized COV matrices and outputs the results to a file
"gmatall". The data is then stored in a permanent SAS data set in the
lercviri library called "lercviri.gmatall_iter&sysdate&set".   Change the
variable "set" to a different letter if running this multiple times per day*/

      options nosource nonotes;
      /*options source notes;*/
      /*options mprint mlogic ;*/

      DATA nitro56; /*renames my permanent dataset to nitro56 */
            SET  LerCviRI.n56;
            pot=_n_;
```

```
      RUN; /*PROC PRINT; RUN;*/
DATA nitro51;
      SET  LerCviRI.n51;
      pot=_n_;
      RUN; /*PROC PRINT; RUN;*/
DATA nitro06;
      SET  LerCviRI.n6;
      pot=_n_;
      RUN; /*PROC PRINT; RUN;*/
DATA nitro01;
      SET  LerCviRI.n1;
      pot=_n_;
      RUN; /*PROC PRINT; RUN;*/

%GLOBAL nlvl;
%LET nlvl=56; /*This is the dataset to resample,change to 56 51 06 01*/

%GLOBAL iteration;
%LET iteration=10000; /*number of bootstrapped datasets*/

/*PROC PRINTTO LOG="C:\LerCvi\N&nlvl._&iteration.iter_LOG.txt"; RUN; */
/*This redirects log ouput to prevent overflows*/

%MACRO bootN;
%DO j=1 %TO &iteration;

TITLE "N&nlvl iteration_&j";
      DATA bootset&j;
            DO i=1 to 160;
                  pot=ROUNDE(159*ranuni(0)+1);
/*makes "bootset1" with 160 random numbers between 1 and 160.  This repeats
(bootset2, etc.) niteration times*/
            OUTPUT;
            END;
            RUN;
      /*PROC PRINT DATA = bootset&j; RUN;*/

      PROC SORT DATA= bootset&j;
```

```
                BY pot;
                RUN;
                /*PROC PRINT DATA = bootset&j; RUN;*/
          DATA boot&nlvl._&j;
/*makes "boot1", a merger of random (resampled) pot #s in bootset1 and the
real data from nitro56 */
                MERGE  bootset&j nitro&nlvl;
                BY pot;
                RUN;
                /*PROC PRINT DATA = boot&nlvl._&j; RUN;*/
          DATA boot&nlvl._&j;
/*trims the bootstrapped dataset to the resampled genotypes only */
                SET boot&nlvl._&j;
                IF i^='.';
                RUN;
          /*PROC PRINT DATA = boot&nlvl._&j; RUN;*/
     PROC MEANS DATA=boot&nlvl._&j NOPRINT;
     /*calculates means so that traits can be mean standardized */
          VAR  RDtotg  RDproprt  RDNpct  RDnlvs  RDboltd  RDdelco2  RDdelh2o
          RDfvfm RDdietr RDyield RDtotlen RDrb RDtsl;
          OUTPUT  OUT=  M_boot&nlvl._&j  MEAN  (RDtotg  RDproprt  RDNpct  RDnlvs
          RDboltd  RDdelco2  RDdelh2o  RDfvfm  RDdietr  RDyield  RDtotlen  RDrb
          RDtsl)=M_totg  M_proprt  M_Npct  M_nlvs  M_boltd  M_delco2  M_delh2o
          M_fvfm M_dietr M_yield M_totlen M_rb M_tsl;
          RUN;
     /*PROC PRINT DATA=M_boot&nlvl._&j; RUN;*/

     DATA M_boot&nlvl._&j;   /*adds variable "iter" to means and data set so
                             that they can be merged */
          SET M_boot&nlvl._&j;
          iter=&j;
          RUN; *PROC PRINT; *RUN;

     DATA boot&nlvl._&j;
          SET boot&nlvl._&j;
          iter=&j;
          RUN; *PROC PRINT; *RUN;
```

```
      DATA MS&nlvl._&j;       /*creates 'MS56_1' (mean standardized) which
                              has mean standardized traits (eg MSboltd) */
            MERGE boot&nlvl._&j M_boot&nlvl._&j;
            BY iter;
                  MStotg= Rdtotg/M_totg;
                  MSproprt= Rdproprt/M_proprt;
                  MSNpct= RDNpct/M_Npct;
                  MSnlvs= RDnlvs/M_nlvs;
                  MSboltd= Rdboltd/M_boltd;
                  MSdelco2= RDdelco2/M_delco2;
                  MSdelh2o= RDdelh2o/M_delh2o;
                  MSfvfm= RDfvfm/M_fvfm;
                  MSdietr= Rddietr/M_dietr;
                  MSyield= Rdyield/M_yield;
                  MStotlen= Rdtotlen/M_totlen;
                  MSrb= RDrb/M_rb;
                  MStsl= RDtsl/M_tsl;
            RUN; /*PROC PRINT; RUN;*/

*use the PROC MEANS below to check that all means =1;

/*PROC MEANS DATA= MS&nlvl._&j N MEAN VAR STD;
      VAR MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2 MSdelh2o MSfvfm
      MSdietr MSyield MStotlen MSrb MStsl; RUN;*/


      title "COV matrix, mean standardized MS&nlvl._&j";
      /*Gmatrix output to datasets gmat1, gmat2, etc. */
      PROC CORR DATA =MS&nlvl._&j COV NOCORR NOPRINT NOSIMPLE OUT =
      gmat&nlvl._&j;
            VAR MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2 MSdelh2o
            MSfvfm MSdietr MSyield MStotlen MSrb MStsl; RUN;
      %END;
      %MEND bootN;
      %bootN;
```

```
/*adds a variable (iter) indicating which bootstrapped iteration this data
came from*/

      TITLE;

/* The following initializes the gall dataset so it has the same variables as
the gmat output data sets*/
       DATA gmat&nlvl.all;
            SET gmat&nlvl._1;
            RUN; /*PROC PRINT DATA = gmat&nlvl.all; RUN;*/

/* The following adds results from gmat2, gmat3...etc to gmatall*/

      %MACRO mergeG;
            %DO k=2 %TO &iteration;
                  PROC APPEND BASE=gmat&nlvl.all DATA=gmat&nlvl._&k;
                  *PROC PRINT DATA = gmat&nlvl.all;
            %END;
      %MEND MergeG;
      %MergeG;

      DATA gmat&nlvl.all;
            SET gmat&nlvl.all;
            IF _TYPE_='COV';
      /*PROC PRINT DATA=gmat&nlvl.all; RUN;*/

      PROC SORT DATA=gmat&nlvl.all;
            BY _NAME_; RUN;

      PROC UNIVARIATE DATA=gmat&nlvl.all NOPRINT;
            BY _NAME_;
            VAR MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2 MSdelh2o
            MSfvfm MSdietr MSyield MStotlen MSrb MStsl;
            OUTPUT OUT=all&nlvl
                  PCTLPTS = 2.5 97.5
                  PCTLPRE = MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2
                  MSdelh2o MSfvfm MSdietr MSyield MStotlen MSrb MStsl
                  PCTLNAME = LO HI; RUN;
```

```
      PROC PRINT DATA=all&nlvl;
      TITLE "Mean Standardized N&nlvl Cov matrix 95% CI &iteration
      iterations"; RUN;

/* If running this multiple times per day and want to save several datasets
of output statistics from bootsampled data, change the set=A to B, C, etc.
each time. This way, sets with different numbers of bootstraps can be saved
*/
/*
      %LET set=A;
      %MACRO saveGmatALL;
            DATA  LerCviRI.Gmat56all&iteration.Iter&SYSDATE&SET;
                  SET gmat56all;
                  RUN;
      %MEND  saveGmatALL;
      %saveGmatALL */
```

APPENDIX B

A SAS MACRO PROGRAM TO CALCULATE THE RANDOM SKEWERS TEST FOR PAIRS OF G-MATRICES USING 10 000 RANDOM SKEWERS.

```
DATA MSCOV56;
      SET LerCviRi.MS_Cov56;
      RUN;
DATA MSCOV51;
      SET LerCviRi.MS_Cov51;
      RUN;
DATA MSCOV06;
      SET LerCviRi.MS_Cov06;
      RUN;
DATA MSCOV01;
      SET LerCviRi.MS_Cov01;
      RUN;

%GLOBAL iteration;
%GLOBAL matrixA;
%GLOBAL matrixB;

%LET iteration=10001;
%LET matrixA=06;
%LET matrixB=01;

PROC IML;
      USE MSCOV&matrixA;        /* makes G-matrices GM56 AND GM1*/
```

```
            READ ALL VAR {MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2
            MSdelh2o MSfvfm MSdietr MSyield MStotlen MSrb}
                  INTO gm&matrixA;
            CLOSE MSCOV&matrixA;
            traits={MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2 MSdelh2o
            MSfvfm MSdietr MSyield MStotlen MSrb};
            PRINT gm&matrixA [ROWNAME=traits COLNAME=traits];

            USE MSCOV&matrixB;
            READ ALL VAR {MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2
            MSdelh2o MSfvfm MSdietr MSyield MStotlen MSrb}
                  INTO gm&matrixB;
            CLOSE MSCOV&matrixB;
            PRINT gm&matrixB [ROWNAME=traits COLNAME=traits];

/*Use this module to remove covariances from G-matrices before random skewers
comparisons*/
      /*
      vec&matrixA=VECDIAG(gm&matrixA); PRINT vec&matrixA;
      vec&matrixB=VECDIAG(gm&matrixB); PRINT vec&matrixB;

      diag&matrixA=DIAG(vec&matrixA);   PRINT   diag&matrixA   [ROWNAME=traits
      COLNAME=traits];
      diag&matrixB=DIAG(vec&matrixB);   PRINT   diag&matrixB   [ROWNAME=traits
      COLNAME=traits];

      gm&matrixA=    diag&matrixA;    PRINT    gm&matrixA    [ROWNAME=traits
      COLNAME=traits];
      gm&matrixB=    diag&matrixB;    PRINT    gm&matrixB    [ROWNAME=traits
      COLNAME=traits];
      */

      %MACRO makeresponsevectors;   *calculates vector-correlation(GM56, GM1;
            %DO y=1 %TO (&iteration-1);
                  a=2*RANUNI(0)-1;        *makes a random skewer, each
                                          element between -1 and 1;
                  b=2*RANUNI(0)-1;
                  c=2*RANUNI(0)-1;
```

```
                    d=2*RANUNI(0)-1;
                    e=2*RANUNI(0)-1;
                    f=2*RANUNI(0)-1;
                    g=2*RANUNI(0)-1;
                    h=2*RANUNI(0)-1;
                    i=2*RANUNI(0)-1;
                    j=2*RANUNI(0)-1;
                    k=2*RANUNI(0)-1;
                    l=2*RANUNI(0)-1;
                    beta=j(12,1);
                    beta[1,1]=a;      *assigns  random  numbers  to  the
                                      vector elements;
                    beta[2,1]=b;
                    beta[3,1]=c;
                    beta[4,1]=d;
                    beta[5,1]=e;
                    beta[6,1]=f;
                    beta[7,1]=g;
                    beta[8,1]=h;
                    beta[9,1]=i;
                    beta[10,1]=j;
                    beta[11,1]=k;
                    beta[12,1]=l;
                    mag=
SQRT(a#a+b#b+c#c+d#d+e#e+f#f+g#g+h#h+i#i+j#j+k#k+l#l);  *vector magnitude;
          betaUNIT=beta/mag;      *scales vector to unit length;
          RespVect&matrixA._&y=gm&matrixA*betaUNIT;
          RespVect&matrixB._&y=gm&matrixB*betaUNIT;
     VC&matrixA.vs&matrixB._&y=(RespVect&matrixA._&y`*
     RespVect&matrixB._&y)/
     (SQRT((RespVect&matrixA._&y`*RespVect&matrixA._&y)*(RespVect&matrixB._&
     y`*RespVect&matrixB._&y)));   *calculates vector correlation;

     IF     &y=1     THEN     CREATE     VecCor&matrixA.vs&matrixB     FROM
     VC&matrixA.vs&matrixB._&y;
                    ELSE EDIT VecCor&matrixA.vs&matrixB;
                    APPEND FROM VC&matrixA.vs&matrixB._&y ;
               %END;
```

```
      %MEND makeresponsevectors ;
      %makeresponsevectors;

QUIT;

PROC UNIVARIATE DATA=VecCor&matrixA.vs&matrixB;
      TITLE "VecCorNOCOV&matrixA.vs&matrixB";
      RUN;

*****MAKE PERMENANT RECORDS****;
/*
DATA LerCviRi.MS_VC_NOCOV&matrixA.vs&matrixB;
      SET VecCor&matrixA.vs&matrixB;
      RUN;
      PROC PRINT DATA=LerCviRi.MS_VC_NOCOV&matrixA.vs&matrixB; RUN; */
```

APPENDIX C

A SAS MACRO PROGRAM USED TO CREATE NULL DISTRIBUTIONS OF VECTOR CORRELATIONS=1 FOR RANDOM SKEWERS TESTS.

```
DATA nitro56;           /*renames permanent dataset to nitro56*/
      SET  LerCviRI.n56;
      pot=_n_;    /*pot #s used to merge real data & resampled #s*/
      RUN;
DATA nitro51;
      SET  LerCviRI.n51;
      pot=_n_;
      RUN;
DATA nitro06;
      SET  LerCviRI.n6;
      pot=_n_;
      RUN;
DATA nitro01;
      SET  LerCviRI.n1;
      pot=_n_;
      RUN;


%GLOBAL nlvl;
%LET nlvl=56;           /*Dataset to resample, change to 56 51 06 01*/
%GLOBAL iteration;
%LET iteration=10001;         /*use  1  more  than  the  number  of
                              bootstrapped datasets needed*/
```

```
/* Macro BootN creates (iteration) number of bootstrapped genetic covariance
matrices named gmat56_1, gmat56_2, etc., saving them for analysis in a
separate step using proc IML*/

      options nosource nonotes;

      %MACRO bootN;
      %DO j=1 %TO &iteration;

            DATA bootset&j;
                  DO i=1 to 160;
                        pot=ROUNDE(159*ranuni(0)+1);
/*makes (niteration) datasets "bootset1" with 160 random numbers between 1
and 160*/
                  OUTPUT;
                  END;
                  RUN;

            PROC SORT DATA= bootset&j;
                  BY pot;
                  RUN;

            DATA boot&nlvl._&j;     /*makes datasets merging real data with
                                    resampled numbers, i.e. makes a
                                    bootstrapped dataset*/
                  MERGE  bootset&j nitro&nlvl;
                  BY pot;
                  RUN;

            DATA boot&nlvl._&j;     /* removes the genotypes that weren't
                                    resampled in this iteration of
                                    bootstrapping*/
                  SET boot&nlvl._&j;
                  IF i^='.';
                  RUN;
```

```
PROC MEANS DATA=boot&nlvl._&j NOPRINT;
/*calculates means of traits in boot56_1 and outputs to M_boot56_1 */
      VAR RDtotg RDproprt RDNpct RDnlvs RDboltd RDdelco2 RDdelh2o
      RDfvfm RDdietr RDyield RDtotlen RDrb RDtsl;
      OUTPUT OUT= M_boot&nlvl._&j MEAN (RDtotg RDproprt RDNpct RDnlvs
      RDboltd RDdelco2 RDdelh2o RDfvfm RDdietr RDyield RDtotlen RDrb
      RDtsl)= M_totg M_proprt M_Npct M_nlvs M_boltd M_delco2 M_delh2o
      M_fvfm M_dietr M_yield M_totlen M_rb M_tsl;
      RUN;

DATA M_boot&nlvl._&j;   /*adds variable "iter" to means and original
                        data set so that they can be merged */
      SET M_boot&nlvl._&j;
      iter=&j;
      RUN;

DATA boot&nlvl._&j;
      SET boot&nlvl._&j;
      iter=&j;
      RUN;

DATA MS&nlvl._&j;       /*creates dataset MS56_1 which has mean
                        standardized traits (eg MSboltd) */
      MERGE boot&nlvl._&j M_boot&nlvl._&j;
      BY iter;
            MStotg= Rdtotg/M_totg;
            MSproprt= Rdproprt/M_proprt;
            MSNpct= RDNpct/M_Npct;
            MSnlvs= RDnlvs/M_nlvs;
            MSboltd= Rdboltd/M_boltd;
            MSdelco2= RDdelco2/M_delco2;
            MSdelh2o= RDdelh2o/M_delh2o;
            MSfvfm= RDfvfm/M_fvfm;
            MSdietr= Rddietr/M_dietr;
            MSyield= Rdyield/M_yield;
            MStotlen= Rdtotlen/M_totlen;
            MSrb= RDrb/M_rb;
            MStsl= RDtsl/M_tsl;
```

```
            RUN;

/*Cov among mean standardized RIL means output to datasets gmat56_1,
gmat56_2, etc. with TWELVE TRAITS, NO TSL*/
      PROC CORR DATA =MS&nlvl._&j COV NOCORR NOPRINT NOSIMPLE
      OUT=gmat&nlvl._&j;
            VAR MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2 MSdelh2o
            MSfvfm MSdietr MSyield MStotlen MSrb;
            RUN;

      DATA  gmat&nlvl._&j;    /*removes lines for MEAN STD N from datasets
                              gmat56_1, gmat56_2, etc, leaving covariances
                              only */
            SET gmat&nlvl._&j;
            IF _TYPE_="COV";
            RUN;

      PROC IML;               /* NOTE: matrices must be saved to access them
                              again in later PROC IML*/
            USE gmat&nlvl._&j;      /*makes G- matrices named boot56_1,
                                    boot56_2...*/
            READ ALL VAR {MStotg MSproprt MSNpct MSnlvs MSboltd MSdelco2
            MSdelh2o MSfvfm MSdietr MSyield MStotlen MSrb} INTO
            boot&nlvl._&j;
            CLOSE gmat&nlvl._&j;
            RESET STORAGE = LerCviRI.G_matrix;
            STORE boot&nlvl._&j;
            %END;
            %MEND bootN;
            %bootN;
      QUIT;

      ************************************************************;

      PROC IML;
            RESET STORAGE = LERCVIRI.G_matrix;
            traits={totg proprt Npct nlvs boltd delco2 delh2o fvfm dietr
            yield totlen rb}; *PRINT traits;
```

```
      STORE traits;

%MACRO bringbackcovmatrices;
            %DO j=1 %TO &iteration;
            LOAD boot&nlvl._&j;
            *PRINT boot&nlvl._&j [ROWNAME=traits COLNAME=traits];
            %END;
            %MEND  bringbackcovmatrices;
      %bringbackcovmatrices;



%MACRO makeresponsevectors;          *will calculate vector corr between
                                     Resp1 and Resp2, Resp2 and Resp3,
                                     Resp3 and resp4, etc.;
            %DO y=1 %TO (&iteration-1);
            %Let z=%EVAL(&y+1);
                  a=2*RANUNI(0)-1;        *makes a random skewer;
                  b=2*RANUNI(0)-1;
                  c=2*RANUNI(0)-1;
                  d=2*RANUNI(0)-1;
                  e=2*RANUNI(0)-1;
                  f=2*RANUNI(0)-1;
                  g=2*RANUNI(0)-1;
                  h=2*RANUNI(0)-1;
                  i=2*RANUNI(0)-1;
                  j=2*RANUNI(0)-1;
                  k=2*RANUNI(0)-1;
                  l=2*RANUNI(0)-1;
                  beta=j(12,1);
                  beta[1,1]=a;      *assigns the random numbers to the
                                     vector elements;
                  beta[2,1]=b;
                  beta[3,1]=c;
                  beta[4,1]=d;
                  beta[5,1]=e;
                  beta[6,1]=f;
                  beta[7,1]=g;
```

```
                        beta[8,1]=h;
                        beta[9,1]=i;
                        beta[10,1]=j;
                        beta[11,1]=k;
                        beta[12,1]=l;
                        mag=
SQRT(a#a+b#b+c#c+d#d+e#e+f#f+g#g+h#h+i#i+j#j+k#k+l#l); *vector magnitude;
      betaUNIT=beta/mag;            *scales vector to unit length;
      RespVect&y=boot&nlvl._&y*betaUNIT;
      RespVect&z=boot&nlvl._&z*betaUNIT;
      VC&nlvl._&y=(RespVect&y`*RespVect&z)/
      (SQRT((RespVect&y`*RespVect&y)*(RespVect&z`*RespVect&z)));
                  *calculates vector correlation;
                  *STORE VC&nlvl._&y;
                  IF &y=1 THEN CREATE Vc&nlvl.Null FROM VC&nlvl._&y;
                                    ELSE EDIT Vc&nlvl.Null;
                        APPEND FROM VC&nlvl._&y ;
                  %END;
            %MEND makeresponsevectors ;
            %makeresponsevectors;
      QUIT;
      *PROC PRINT DATA =Vc&nlvl.Null ; *RUN;

      PROC UNIVARIATE DATA =Vc&nlvl.Null;
      OUTPUT OUT=RS_Null&iteration._N&nlvl
                  PCTLPTS = 5 1 0.1 0.01 0.001 0.0001 0.00001
                  PCTLPRE = VC
                  PCTLNAME = P5 P1 P01 P001 P0001 P00001 P000001;
      RUN;
      PROC PRINT DATA=RS_Null&iteration._N&nlvl; RUN;
```

BIBLIOGRAPHY

Agrawal, A. F., and J. R. Stinchcombe. 2009. How much do genetic covariances alter the rate of adaptation? Proceedings of the Royal Society B 276:1183-1191

Alonso-Blanco, C., A. J. M. Peeters, M. Koornneef, C. Lister, C. Dean, N. Van Den Bosch, J. Pot, and M. T. R. Kuiper. 1998. Development of an AFLP based linkage map of Ler, Col and Cvi Arabidopsis thaliana ecotypes and construction of a Ler/Cvi recombinant inbred line population. The Plant Journal 14:259-271.

Ankra-Badu, G. A., D. Pomp, D. Shriner, D. B. Allison, and N. Yi. 2009. Genetic influences on growth and body composition in mice: multilocus interactions. International Journal of Obesity 33:89-95.

Arnold, S. J., R. Burger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the G-matrix. Evolution 62.

Ashman, T. L., and C. J. Majetic. 2006. Genetic constraints on floral evolution: a review and evaluation of patterns. Heredity 96:343-352.

Astles, P. A., A. J. Moore, and R. F. Preziosi. 2006. A comparison of methods to estimate cross-environment genetic correlations. Journal of Evolutionary Biology 19:114-122.

Barton, N., and L. Partridge. 2000. Limits to Natural Selection. BioEssays 22:1075-1084.

Barton, N. H., and P. D. Keightley. 2002. Understanding quantitative genetic variation. Nat. Rev. Genet. 3:11-21.

Beavis, W. D. 1998. QTL Analyses: Power, Precision, and Accuracy *in* A. H. Patterson, ed. Molecular dissection of complex traits. CRC Press, Boca Raton.

Begin, M., and D. A. Roff. 2003. The constancy of the G matrix through species divergence and the effects of quantitative genetic constraints on phenotypic evolution: A case study in crickets. Evolution 57:1107-1120.

Bell, G., and A. Gonzalez. 2009. Evolutionary rescue can prevent extinction following environmental change. Ecology Letters 12:942-948.

Blows, M. W., and A. A. Hoffmann. 2005. A Reassessment of Genetic Limits to Evolutionary Change. Ecology 86:1371-1384.

Bradshaw, A. D. 1991. The Croonian Lecture, 1991: Genostasis and the Limits to Evolution. Philosophical Transactions: Biological Sciences 333:289-305.

Bradshaw, W. E., and C. M. Holzapfel. 2000. The evolution of genetic architectures and the divergence of natural populations *in* J. B. Wolf, E. D. Brodie, III, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford University Press, New York, NY.

Brock, M. T., and C. Weinig. 2007. Plasticity And Environment-Specific Covariances: An Investigation Of Floral-Vegetative And Within Flower Correlations. Evolution 61:2913-2924.

Broman, K. W. 2005. The Genomes of Recombinant Inbred Lines. Genetics 169:1133-1146.

Broman, K. W., and S. Sen. 2009. A guide to QTL mapping with R/qtl. Springer, New York.

Broman, K. W., H. Wu, S. Sen, and G. A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:899-890.

Calsbeek, B., and C. J. Goodnight. 2009. Empirical Comparison of G Matrix Test Statistics: Finding Biologically Relevant Change. Evolution 63:2627-2635.

Carlborg, O., and C. S. Haley. 2004. Epistasis: too often neglected in complex trait studies? 5:618-625.

Carlborg, O., L. Jacobsson, P. Ahgren, P. Siegel, and L. Andersson. 2006. Epistasis and the release of genetic variation during long-term selection. 38:418-420.

Carter, A. J. R., J. Hermisson, and T. F. Hansen. 2005. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. Theoretical Population Biology 68:179-196.

Charmantier, A., and D. Garant. 2005. Environmental quality and evolutionary potential: lessons from wild populations. Proceedings of the Royal Society: Biological Science 272:1415-1425.

Chenoweth, S. F., H. D. Rundle, and M. W. Blows. 2010. The Contribution of Selection and Genetic Constraints to Phenotypic Divergence. The American Naturalist 175:186-196.

Cheverud, J. M. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (Saguinus oedipus) and saddle-back (S. fuscicollis) tamarins. Journal of Evolutionary Biology 9:5-42.

Cheverud, J. M., J. J. Rutledge, and W. R. Atchley. 1983. Quantitative Genetics of Development: Genetic Correlations Among Age-Specific Trait Values and the Evolution of Ontogeny. Evolution 37:895-905.

Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859-882.

Combarros, O., M. Cortina-Borja, A. D. Smith, and D. J. Lehmann. 2009. Epistasis in sporadic Alzheimer's disease. Neurobiology of Aging 30:1333-1349.

Conner, J. K. 2003. Artificial Selection: A Powerful Tool for Ecologists. Ecology 84:1650-1660.

Conner, J. K., and A. A. Agrawal. 2005. Mechanisms of constraints: the contributions of selection and genetic variance to the mainenance of cotyledon number in wild radish. Journal of Evolutionary Biology 18:238-242.

Conner, J. K., R. Franks, and C. Stewart. 2003. Expression of Additive Genetic Variances and Covariances for Wild Radish Floral Traits: Comparison between Field and Greenhouse Environments. Evolution 57:487-495.

Colautti, R. I., and S. C. H. Barrett. 2011. Population Divergence Along Lines Of Genetic Variance And Covariance In The Invasive Plant Lythrum Salicaria In Eastern North America. Evolution 65:2514-2529.

Cox, N. J., M. Frigge, D. L. Nicolae, P. Concannon, C. L. Hanis, G. I. Bell, and A. Kong. 1999. Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nature Genetics 21:213-215.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in Drosophila" revisited. Evolution 51:295-303.

Crow, J. F. 2010. On epistasis: why it is unimportant in polygenic directional selection. Philosophical Transactions: Biological Sciences 365:1241-1244.

Demuth, J. P., and M. J. Wade. 2005. On the theoretical and empirical framework for studying genetic interactions within and among species. American Naturalist 165:524-536.

Demuth, J. P., and M. J. Wade. 2006. Experimental Methods for Measuring Gene Interactions. Annual Review of Ecology Evolution and Systematics 37:289-316.

Demuth, J. P., and M. J. Wade. 2007a. Population differentiation in the beetle Tribolium castaneum. I. Genetic architecture. Evolution 61:494-509.

Demuth, J. P., and M. J. Wade. 2007b. Population differentiation in the beetle Tribolium castaneum. II. Haldane's rule and incipient speciation. Evolution 61:694-699.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia university press, New York, NY.

Dudley, J. W. 2007. From Means to QTL: The Illinois Long-Term Selection Experiment as a Case Study in Quantitative Genetics. Crop Sci 47:S-20-31.

Dudley, J. W. 2008. Epistatic Interactions in Crosses of Illinois High Oil x Illinois Low Oil and of Illinois High Protein x Illinois Low Protein Corn Strain. Crop Sci 48:59-68.

Diaz, C., V. Saliba-Colombani, O. Loudet, P. Belluomo, L. Moreau, F. Daniel-Vedele, J.-F. Morot-Gaudry, and C. Masclaux-Daubresse. 2006. Leaf Yellowing and Anthocyanin Accumulation are Two Genetically Independent Strategies in Response to Nitrogen Limitation in Arabidopsis thaliana. Plant and Cell Physiology 47:74-83.

Doerge, R. W., and G. A. Churchill. 1996. Permutation Tests for Multiple Loci Affecting a Quantitative Character. Genetics 142:285-294.

Doroszuk, A., MarcinÂ W. Wojewodzic, G. Gort, and JanÂ E. Kammenga. 2008. Rapid Divergence of Genetic Varianceâ€• Covariance Matrix within a Natural Population. The American Naturalist 171:291-304.

Edmands, S. 2002. Does parental divergence predict reproductive compatibility? Trends in Ecology & Evolution 17:520-527.

Edwards, C., and C. Weinig. 2011. The quantitative-genetic and QTL architecture of trait integration and modularity in *Brassica rapa* across simulated seasonal settings. Heredity 106:661-677.

Elnaccash, T. W., and S. J. Tonsor. 2010. Something Old and Something New: Wedding Recombinant Inbred Lines with Traditional Line Cross Analysis Increases Power to Describe Gene Interactions. PLoS ONE 5:e10200.

Erickson, D., and C. B. Fenster. 2006. Intraspecific hybridization and the recovery of fitness in the native legume Chamaecrista fasciculata. Evolution 60:225-233.

Etterson, J. R., and R. G. Shaw. 2001. Constraint to Adaptive Evolution in Response to Global Warming. Science 294:151-154.

Fenster, C. B., and L. F. Galloway. 2000. Population differentiation in an annual legume: genetic architecture. Evolution 54:1157-1172.

Fitzpatrick, B. M. 2008. Hybrid dysfunction: Population genetic and quantitative genetic perspectives. American Naturalist 171:491-498.

Friedman, S., and H. F. Weisberg. 1981. Interpreting the first eigenvalue of a correlation matrix. Educational and Psychological Measurement 41:11-21.

Fu, J., J. J. B. Keurentjes, H. Bouwmeester, T. America, F. W. A. Verstappen, J. L. Ward, M. H. Beal, R. C. H. de Vos, M. Dijkstra, R. A. Scheltema, F. Johannes, M. Koornneef, D. Vreugdenhil, R. Breitling, and R. C. Jansen. 2009. System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nature Genetics 41:166-167.

Fuller, R. C., C. F. Baer, and J. Travis. 2005. How and When Selection Experiments Might Actually be Useful. Integr. Comp. Biol. 45:391-404.

Futuyma, D. J. 2010. Evolutionary Constraints and Ecological Consequences. Evolution 64:1865-1884.

Gardner, K. M., and R. G. Latta. 2007. Shared quantitative trait loci underlying the genetic correlation between continuous traits. Molecular Ecology 16:4195-4209.

Gavrilets, S. 1997. Evolution and speciation on holey adaptive landscapes. Trends in ecology and evolution 12:307-312.

Gehring, J. L., and Y. B. Linhart. 1993. Sexual Dimorphisms and Response to Low Resources in the Dioecious Plant Silene latifolia (Caryophyllaceae). International Journal of Plant Sciences 154:152-162.

Gomulkiewicz, R., and D. Houle. 2009. Demographic and Genetic Constraints on Evolution. The American Naturalist 174:218-229.

Gruber, N., and J. N. Galloway. 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451:293-296.

Hallander, J., and P. Waldmann. 2007. The effect of non-additive genetic interactions on selection in multi-locus genetic models. 98:349-359.

Hansen, T. F. 2006. The evolution of genetic architecture. Annual review of ecology and systematics 37:123-157.

Hansen, T. F., C. Pelabon, and D. Houle. 2011. Heritability is not evolvability. Evolutionary Biology 38:258-277.

Hayman, B. I., and K. Mather. 1955. The description of genetic interactions in continuous variation. Biometrics 11:69-82.

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

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. 130:195-204.

Houle, D., B. Morikawa, and M. Lynch. 1996. Comparing Mutational Variabilities. Genetics 143:1467-1483.

Jones, A. G., S. J. Arnold, and R. Burger. 2007. The mutation matrix and the evolution of evolvability. 61:727-745.

Keightley, P. D. 2010. Mutational Variation and Long-Term Selection Response. Pp. 227-247. Plant Breeding Reviews. John Wiley & Sons, Inc.

Kelly, J. C. 2009. Connecting QTLs to the G-matrix of evolutionary quantitative genetics. Evolution 63: 813-825.

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. Genetica 136:271-284.

Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. Proceedings of the National Academy of Sciences of the United States of America 99:14878-14883.

Kusterer, B., J. Muminovic, H. F. Utz, H. P. Piepho, S. Barth, M. Heckenberger, R. C. Meyer, T. Altmann, and A. E. Melchinger. 2007. Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in Arabidopsis. Genetics 175:2009-2017.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain:body size allometry. Evolution 33:402-416.

Lande, R. 1980. sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34:292-305.

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

Lim, U., K. Peng, B. Shane, P. J. Stover, A. A. Litonjua, S. T. Weiss, J. M. Gaziano, R. L. Strawderman, F. Raiszadeh, J. Selhub, K. L. Tucker, and P. A. Cassano. 2005. Polymorphisms in cytoplasmic serine hydroxymethyltransferase and methylenetetrahydrofolate reductase affect the risk of cardiovascular disease in men. Journal of Nutrition 135:1989-1994.

Lopez-Fernandez, H., and D. I. Bolnick. 2007. What Causes Partial F1 Hybrid Viability? Incomplete Penetrance versus Genetic Variation. PLoS ONE 2.

Loudet, O., S. Chaillou, P. Merigout, J. Talbotec, and F. Daniel-Vedele. 2003. Quantitative Trait Loci Analysis of Nitrogen Use Efficiency in Arabidopsis. Plant Physiol. 131:345-358.

Lynch, M. 1991. The genetic interpretation of inbreeding depression and outibreeding depresssion. Evolution 45:622-639.

Lynch, M., and B. Walsh. 1998. Genetics and the analysis of quantitative traits. Sinaur Associates, Sunderland, MA.

Mallitt, K. L., S. P. Bonser, and J. Hunt. 2010. The plasticity of phenotypic integration in response to light and water availability in the pepper grass, Lepidium bonariense. Evolutionary Ecology 24:1321-1337.

Manly, B. F. J. 2001. Randomization, bootstrap and monte carlo methods in biology. Chapman & Hall/CRC, Boca Raton, Florida.

McGuigan, K. 2006. Studying phenotypic evolution using multivariate quantitative genetics. Molecular Ecology 15:883-896.

McGuigan, K., and C. M. Sgrò. 2009. Evolutionary consequences of cryptic genetic variation. Trends in ecology & evolution (Personal edition) 24:305-311.

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

Meagher, T. R. 1999. Quantitative genetics of sexual dimorphism *in* M. A. Geber, T. E. Dawson, and L. F. Delph, eds. .Gender and Sexual Dimorphism in Flowering Plants. Springer, New York.

Mezey, J. G., and D. Houle. 2005. The dimensionality of genetic variation for wing shape in drosophila melanogaster. Evolution 59:1027-1038.

Mitchell-Olds, T., and J. Schmitt. 2006. Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. Nature 441:947-952.

Montesinos-Navarro, A., J. Wig, F. Xavier Pico, and S. J. Tonsor. 2011. Arabidopsis thaliana populations show clinal variation in a climatic gradient associated with altitude. New Phytologist 189:282-294.

Montesinos, A., S. J. Tonsor, C. Alonso-Blanco, and F. Xavier Pico. 2009. Demographic and genetic patterns of variation among populations of Arabidopsis thaliana from contrasting native environments. PLosOne;4(9):e7213.

Moore, J. H. 2003. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Human Heredity 56:73-82.

Naciri-Graven, Y., and J. Goudet. 2003. The Additive Genetic Variance After Bottlenecks Is Affected By The Number Of Loci Involved In Epistatic Interactions. Evolution 57:706-716.

Pavlicev, M., J. M. Cheverud, and G. P. Wagner. 2011. Evolution of adaptive phenotypic variation patterns by direct selection for evolvability. Proceedings of the Royal Society B: Biological Sciences 278:1903-1912.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers. 2003. Giving Meaningful Interpretation To Ordination Axes: Assessing Loading Significance In Principal Component Analysis. Ecology 84:2347-2363.

Phillips, P. C., and S. J. Arnold. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. Evolution 53:1506-1515.

Porter, A. H., and N. A. Johnson. 2002. Speciation despite gene flow when developmental pathways evolve. Evolution 56:2103-2111.

Presgraves, D. C. 2002. Patterns of post-zygotic isolation in Lepidoptera. Evolution 56:1168-1183.

Qin, S. Y., X. Zhao, Y. X. Pan, J. H. Liu, G. Y. Feng, J. C. Fu, J. Y. Bao, Z. Z. Zhang, and L. He. 2005. An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray. European Journal of Human Genetics 13:807-814.

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

Revell, L. J. 2007. The G matrix under fluctuating correlational mutation and selection. Evolution 61:1857-1872.

Roff, D. A. 2002. Life History Evolution. Sinauer Associates, Inc., Sunderland, MA.

Roff, D. A., and K. Emerson. 2006. Epistasis And Dominance: Evidence For Differential Effects In Life-History Versus Morphological Traits. Evolution 60:1981-1990.

Roff, D. A., and D. J. Fairbairn. 2006. The evolution of trade-offs: where are we? Journal of Evolutionary Biology 20:433-447.

Rutherford, S. L., and S. Lindquist. 1998. Hsp90 as a capacitor for morphological evolution. 396:336-342.

Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). Proceedings of the National Academy of Sciences of the United States of America 96:11910-11915.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766-1774.

Schluter, D. 2000. The ecology of adapive radiations. Oxford University Press, Oxford.

Sen, S., and G. A. Churchill. 2001. A statistical framework for quantitative trait mapping. Genetics 159:371-387.

Sgro, C., and A. A. Hoffman. 2004. Genetic correlations, tradeoffs, and environmental variation. Heredity 93:241-248.

Shaw, F. H., R. G. Shaw, G. S. Wilkinson, and M. Turelli. 1995. Changes in genetic variances and covariances: G whiz! Evolution 49:1260-1267.

Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, New York, NY.

Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends in Ecology & Evolution 17:320-327.

Stinchcombe, J. R., R. Izem, M. S. Heschel, B. V. McGoey, and J. Schmitt. 2010. Across-Environment Genetic Correlations and the Frequency of Selective Environments Shape the Evolutionary Dynamics of Growth Rate in Impatiens capensis. Evolution 64: 2887-2903.

Sun, Z., R. L. Lower, and J. E. Staub. 2006. Analysis of generation means and components of variance for parthenocarpy in cucumber (Cucumis sativus L.) Plant Breeding 125:277-280.

Tonsor, S. J., C. Alonso-Blanco, and M. Koornneef. 2005. Gene function beyond the single trait:natural variation, gene effects, and evolutionary ecology in Arabidopsis thaliana. Plant Cell and Environment 28:2-20.

Tonsor, S. J., and C. J. Goodnight. 1997. Evolutionary predictability in natural populations: Do mating system and nonadditive genetic variance interact to affect heritabilities in Plantago lanceolata? Evolution 51:1773-1784.

Tonsor, S. J., and S. M. Scheiner. 2007. Plastic trait integration across a CO2 gradient in Arabidopsis thaliana. American Naturalist 169:119-140.

Via, S. 1984. The Quantitative Genetics of Polyphagy in an Insect Herbivore. II. Genetic Correlations in Larval Performance Within and Among Host Plants. Evolution 38:896-905.

Via, S., and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39:505-522.

Vitousek, P. M., and R. W. Howarth. 1991. Nitrogen limitation on land and in the sea: How can it occur? Biogeochemistry 13:87-115.

Wade, M. J. 2000. Epistasis as a genetic constraint within populations and an accelerant of adaptive divergence among them *in* J. B. Wolf, E. D. Brodie, III, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford University Press, New York, NY.

Wagner, G. P., and L. Altenberg. 1996. Complex adaptations and the evolution of evolvability. Evolution 50:967-976.

Wagner, G. P., M. Pavlicev, and J. M. Cheverud. 2007. The road to modularity. Genetics 8:921-931.

Walsh, B. 2009. Quantitative genetics, version 3.0: where have we gone since 1987 and where are we headed? Genetica 136:213-223.

Wegner, K. M., C. Berenos, and P. Schmid-Hempel. 2008. Nonadditive genetic components in resistance of the red flour beetle Tribolium castanaeum against parasite infection. Evolution 62:2381-2392.

Weinig, C., M. C. Ungerer, L. A. Dorn, N. C. Kane, Y. Toyonaga, S. S. Halldorsdottir, T. F. C. Mackay, M. D. Purugganan, and J. Schmitt. 2002. Novel Loci Control Variation in Reproductive Timing in Arabidopsis thaliana in Natural Environments. Genetics 162:1875-1884.

Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111-114.

Whitlock, M. C., P. C. Phillips, F. B. G. Moore, and S. J. Tonsor. 1995. Multiple fitness peaks and epistasis. Annual review of ecology and systematics 26:601-629.

Wiltshire, S., J. T. Bell, C. J. Groves, C. Dina, A. T. Hattersley, T. M. Frayling, M. Walker, G. A. Hitman, M. Vaxillaire, M. Farrall, P. Froguel, and M. I. McCarthy. 2006. Epistasis between type 2 diabetes susceptibility loci on chromosomes 1q21-25 and 10q23-26 in northern Europeans. Annals of Human Genetics 70:726-737.

Wolf, J. B., E. D. Brodie, III, and M. J. Wade. 2000. Epistasis and the evolutionary process. Oxford University Press, New York, NY.

Wolfram Research, Inc 2007. Mathematica, Version 6.0, Champaigne, IL.

Xu, S. 2003. Theoretical Basis of the Beavis Effect. Genetics 165:2259-2268.

Zeng, Z.-B., C.-H. Kao, and C. J. Basten. 1999. Estimating the genetic architecture of quantitative traits. Genetical Research 74:279-289.