Theoretical and Empirical Investigations into Adaptation

by

Kevin Wright

Department of Biology
Duke University

Date:_____

Approved:

_____
John H. Willis, Co-Chair

_____
Mark D. Rausher, Co-Chair

_____
A. Jonathan Shaw

_____
Gregory Wray

_____
Thomas Mitchell-Olds

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2010

ABSTRACT

Theoretical and Empirical Investigations into Adaptation

by

Kevin Wright

Department of Biology
Duke University

Date:_____

Approved:

_____
John H. Willis Co-Chair

_____
Mark D. Rausher Co-Chair

_____
A. Jonathan Shaw

_____
Gregory Wray

_____
Thomas Mitchell-Olds

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2010

Abstract

The problem is two fold: how does natural selection operate on systems of interacting genes and how does natural selection operate in natural populations. To address the first problem, I have conducted a theoretical investigation into the evolution of control and the distribution of mutations in a simple system of interacting genes, a linear metabolic pathway. I found that control is distributed unevenly between enzymes, with upstream enzymes possessing the greatest control and accumulating the most beneficial mutations during adaptive evolution. To address the second problem, I investigated the evolution of copper tolerance in the common yellow monkeyflower, *Mimulus guttatus*. I genetically mapped a major locus controlling copper tolerance, *Tol1*. A Dobzhansky-Muller incompatibility was hypothesized to also be controlled by *Tol1*, however, we have demonstrated that it maps to another, tightly linked locus, *Nec1*. Finally, we investigated the parallel evolution of copper tolerance in multiple new discovered mine populations. We found that copper tolerance has evolved in parallel multiple times via at least two distinct physiological mechanisms. In four mine populations, there was a strong signal of selection at markers linked to Tol1, implying that copper tolerance has evolved via the same genetic mechanisms in these populations.

Dedication

This thesis is dedicated to my parents Douglas and Victoria Wright.

Contents

List of Tables

List of Figures

Acknowledgements

This thesis would not be possible without assistance and inspiration from a great many people. First I need to thank my advisors, John Willis and Mark Rausher. I need to thank my committee, Jon Shaw, Gregory Wray, Thomas Mitchell-Olds and Trudy MacKay. I need to thank everyone in the Willis and Rausher lab groups, in particular, David Lowry, Young-Wha Lee, and Jessica Selby for their help on too many things to enumerate. I need to acknowledge my close collaborators on these projects, Deborah Llyod, Mark Macnair and Ryan O'Dell. I also need to acknowledge the many undergraduate students associated with these projects, Margaret Peterson, Michael Yan, Zhirui Zhu, Jeffery Shi, and Daniel Gillon. Also, this thesis would not have been possible without assistance from the great staff in the Duke Biology Department and Sequencing Center: Anne Lacey, Jim Tunney, Beverly Calhon, Jon Mays, Amy O'Tuel, Michael Barnes, Lisa Bukovnik, Todd Smith, and Mel Turner.

1. The Evolution of Control and the Distribution of Adaptive Mutations in a Metabolic Pathway

1.1 Introduction

Theoretical research on the process of adaptation has focused primarily on describing the size and number of genetic changes underlying phenotypic change (Fisher 1930, Kimura 1983, Orr 1998, 2002, 2003). By contrast, comparatively little theoretical attention has been given to the question of whether certain genes or types of genes are preferentially involved in the process of adaptation. Yet the current debate over the relative importance of regulatory vs. structural genes in morphological evolution (Hoekstra and Coyne 2007, Stern and Orogogozo 2008) clearly indicates that this question is of interest to evolutionary biologists.

One situation in which this question is pertinent is the evolution of characters that are influenced by the concentration of end products of metabolic pathways. Often change in end-product concentration can be achieved by substitutions in any one of several genes in the pathway. One example is the intensity of floral pigmentation. To a first approximation, final pigment concentration, and hence color intensity, can be viewed as being determined by the flux rate down the pigment biosynthetic pathway for a fixed time corresponding to the duration of floral development. More generally, any situation in which flux rate determines phenotype is likely to fall in this category. In such situations, metabolic control theory (MCT) (Kacser and Burns 1973) and similar

approaches (Heinrich and Rapoport 1974; Savageau 1976) indicate that changes in flux can be achieved by changing the activity of any enzyme in the pathway. We seek to determine whether, and if so, why, enzymes differ in the probability that they contribute to evolutionary change in pathway flux.

It has been suggested as a general principle that enzymes with the greatest control over flux will be disproportionately involved in such evolutionary change (Hartl *et al*. 1985; Eanes 1999; Watt and Dean 2000). This argument is based on the theoretical expectation that the probability of fixation of an advantageous allele is roughly proportional to its selection coefficient (Hedrick 2000). Since mutations equivalent in terms of enzyme kinetic properties will have greater effects on flux, and hence on fitness, in enzymes with greater metabolic control, mutations in those enzymes will be substituted preferentially.

While this argument is likely sound, it simply pushes back the question of which genes evolve preferentially to the question of which enzymes are expected to have greatest control over flux. Although we are unaware of any theoretical attempts to model the evolution of flux control, many authors have speculated about where in pathways control is expected to be highest.

Kacser and Burns (1973) hypothesized that the magnitudes of flux control exerted by different enzymes may be very similar. This hypothesis was based on the

result from MCT that in linear pathways, overall flux control can be shared by all enzymes. Since metabolic pathways often consist of many enzymes, each would be expected to have only a limited potential to influence flux. Subsequent theoretical analysis of this hypothesis demonstrated that a given flux is consistent with many different flux-control distributions, including, at one extreme, equal flux control by all enzymes and, at the other extreme, major control by one or a few enzymes and little control for all others (Mazat *et al*. 1996). However, the question of which of these possibilities, if any, are likely to be favored by selection has not been addressed.

Another hypothesis, the epistatic or synergistic principle, predicts that control will be vested in a single enzyme at any given time, but will shift over time among enzymes (Dykhuizen *et al.* 1987; Keightley 1989; Bost *et al.* 2001) According to this hypothesis, starting from equal control among enzymes in the pathway, selection to increase (or decrease) flux will cause the activity of one enzyme to increase (decrease). This change results in a decrease (increase) in flux control for the enzyme that changes, and an increase (increase) in control for the other pathway enzymes, causing control to be unequally shared. While this argument seems plausible, there has been no analysis of whether over time all enzymes have equal chance of having elevated control.

Finally, Eanes' (1999) review of enzyme polymorphisms found that control is often centered in enzymes at pathway branch points, which constitute the most

upstream enzymes of their specific branch. Flowers *et al*. (2007) also demonstrated that branching enzymes tend to exhibit more adaptive substitutions than downstream enzymes as would be expected under the principle that evolutionary change will be concentrated in enzymes with the largest control coefficients. In addition, evolutionary changes in these enzymes may be favored because they allow organisms to modify flux allocation to alternate functions and track environmental fluctuations. This suggestion is supported by the "branch point effect", a theoretical demonstration that control coefficients can dramatically shift between enzymes depending on the kinetic rates of the two competing enzymes (LaPorte *et al*. 1984). However, this study does not address the question of how the distribution of control is likely to evolve, but only describes which distributions of control are mathematically possible. Thus, Eanes (1999) concludes his review stating: "all enzymes in [a] contributing pathway may not be equal; determining the rule[s] for these inequalities should be a major goal in studies of enzyme polymorphism."

A Control Coefficient (CC) indicates the degree to which flux through a pathway is altered by a small change in the activity of an enzyme (see Appendix A; this is equivalent to the Sensitivity Coefficient of Kacser and Burns 1973). The "rules" governing the distribution of control coefficients are determined by the biological evolution of metabolic systems. While research demonstrates that there are many

possible distributions of control coefficients, none has examined which of these is most likely to evolve. The optimization of metabolic systems has been explored in detail (Heinrich *et al.* 1991; Heinrich *et al*. 1997; Heinrich and Schuster 1998). In these studies, however, the investigators employ as optimization criteria maximizing flux, maximizing transient times, or minimizing metabolic intermediates, criteria whose biological and evolutionary relevance is unclear.

In an effort to understand how control is expected to be shared among enzymes, and predict which enzymes are most likely to contribute to adaptive genetic changes, we present two models of the evolution of flux control in a simple linear pathway. The first model employs the framework of MCT. Although, there have been many challenges to the MCT framework (Savageau 1976; Cornish-Bowden 1989; Savageau and Sorribas 1989; Savageau 1992), it should be made clear that our aim is not to construct a precisely parameterized model of a particular biological system, but to use this generalized framework to address a single, critically ignored question: what are the rules governing how control will evolve to be distributed among enzymes? The use of the MCT framework to address questions in evolutionary genetics is firmly established, with investigation focused on the molecular basis of dominance (Kacser and Burns 1981; Keightley 1996a, Phandis and Fry 2005, but see Bagheri and Wagner 2004), the relationship between metabolic flux and fitness (Dykhuizen *et al.* 1987, Szathmary 1993),

the amount of additive and non-additive genetic variance in metabolic systems (Keightley 1989), whether this variation can be explained by mutation-selection balance (Clark 1991) and patterns of response of quantitative traits to selection (Keightley 1996b). The second model, saturation kinetics (SK) we examine is based on Michealis-Menten kinetics, and enables us to relax one major assumption of MCT: that enzymes are far from saturation.

Here we limit our analysis to linear pathways as an initial attempt to examine these issues. We find that for such pathways control coefficients will generally evolve to be unequal; that the magnitude of this inequality depends on the *thermodynamic* properties, rather than the kinetic properties, of each reaction step; that upstream enzymes tend to evolve higher control coefficients than downstream enzymes; and that upstream enzymes fix advantageous mutations in greater numbers, and those mutations have larger effects than in downstream enzymes.

1.2 General Modeling Approach

We concentrate here on a simple, linear metabolic pathway containing two, three or ten enzymes. Following Dykhuizen *et al.* (1987), Clark (1991), and Szathmary, (1993) we assume that flux through the pathway is the sole component of fitness and flux is subject to stabilizing selection. Such selection would arise, for example, when there is stabilizing selection on a character (e.g. lactose metabolism, floral pigment intensity,

glucosinolate production following herbivory) whose value is determined by flux. This framework also allows us to examine the effects the effects of directional selection that arises when the optimal level of flux changes, as might occur for floral pigment intensity if there were environmental shift in the optimal intensity for attracting pollinators. We assume, as is usual in an MCT framework, that enzymes are at sufficiently high concentration that they are substantially below saturation (Kacser and Burns 1973). This assumption is subsequently relaxed to investigate the effect of enzyme saturation on the distribution of control coefficients.

To model the evolution of flux control, we assume that flux may be changed either by altering the level of expression of individual enzymes in the pathway (increasing expression level generally increases flux) or by altering the kinetic properties of individual enzymes in the pathway (increasing the specific rate of reaction of an enzyme generally increases flux). We subsume both of these effects by focusing on changes in the overall rate of reaction of an enzyme under standard metabolic conditions. Specifically, we represent the forward rate of reaction i, k_i, to be a product of (1) the amount of the enzyme present, and (2) the standard per-unit-enzyme forward rate constant. For a given k_i, the reverse rate of reaction i, k_{-i}, is a thermodynamic property of the system; it is determined by the difference in free energy between products and reactants (Nelson and Cox 2000), and is not subject to organic evolution.

To determine the control properties of each enzyme at evolutionary equilibrium, we initialize replicate populations described by the enzyme activities (k_1 , k_2 . . . k_i) of a metabolic pathway determining fitness of organisms in the population. Each replicate population is seeded with random k_i values.

Mutations occur in a randomly chosen enzyme, *i*. We do not allow enzymes to differ in their mutation rates. The effect size of each mutation, Δk_i , is drawn from a Gaussian distribution centered at zero. We measure the effect of each mutation on flux and fitness (ΔW) and determine whether the mutation is fixed based on the standard population-genetic fixation probability. Mutations are allowed to occur repeatedly until the flux reaches its optimum. This process produces one evolutionary trajectory, defined as the route across an adaptive landscape from an initial flux toward a new optimal flux. Trajectories are determined for a large number of different initial sets of randomly chosen k_i values all evolving toward the same optimal flux. In essence, this procedure simulates a situation in which the optimal flux has just shifted, and the starting points represent many different previous optima. We do not assert that all of these starting values are biologically likely, but adopt this approach in order to explore the entire space of possible activity values. We perform additional simulations to model a situation in which populations repeatedly shift between two different environmental optima. All simulations are repeated assuming a range of k_{-i} values to determine how

the control evolves under varying thermodynamic environments. Details of the models used are presented in the Appendix A.

This analysis yields two types of information: a frequency distribution of the final degree of flux control for each enzyme at evolutionary equilibrium and the distribution of number and size of mutations in every evolutionary trajectory. By analyzing two models of metabolic control (MCT and SK) and conducting multiple sets of runs with different values of model parameters (e.g. mutational distribution, reversibility of reactions, initial size of substrate pool, intensity of stabilizing selection; Table 1), we are able to evaluate the robustness of our conclusions.

Table 1: Parameters for MCT and SK simulations

k_i	Rate of reaction for enzyme i. (MCT model)
V_i	Maximum rate of reaction for enzyme i. (*SK* model)
C_i	Control coefficient for enzyme i.
α_i	k_{-i} / k_i ; Ratio of reverse to forward rate of reaction.
σ_J	Standard deviation of optimal flux in fitness function, strength of selection.
σ_k	Standard deviation of normal distribution describing mutational effect sizes.
J_{opt}	Optimal flux/ optimal fitness.
N	Population size.
I	Substrate input into pathway.
T	Threshold limit for substrates (*SK* model).

1.3 Results

1.3.1 General behavior of MCT model

In our model, N, α_i , σ_J , σ_k , I, and J_{opt} are fixed parameters (Table 1). We first consider in detail the behavior of the model for a population size of N = 1000, I = 10, an

optimal flux of J_{opt} =1, a strength of selection of σ_J = 0.5, and a variance in mutation size of σ_k = 0.05. Subsequently we discuss how the behavior of the model changes as these parameters are varied. We also assume that there is some maximal enzyme efficiency that is the same for each enzyme, and we scale this maximum to 1 (i.e. we assume that k_i < 1) by choosing the appropriate unit for time. For simplicity, we consider the case where the enzyme reversibilities, α_i, are the same for all i and equal to α, and we let α = 0.001, 0.05, 0.50, or 0.95, representing respectively conditions of very low, low, intermediate, and high reaction reversibility. A total of 200 trials were run with each model for each α value, each trial beginning from a randomly chosen starting point (k_1 , k_2).

Simulations usually reached the optimal flux in 500 – 2000 mutation/fixation cycles, typically having undergone 10-25 substitutions. In most simulations, pathways with two-steps reached the fitness peak more quickly than models with three or ten-steps. The reversibility parameter α also influenced the rate of adaptation: when α= 0.001 or 0.05, pathways reached equilibrium more rapidly than when α= 0.50 or 0.95.

1.3.2 Evolution of control under the two-step MCT model:

The two, three, and ten-step pathways exhibit similar evolutionary patterns, and because these patterns are easier to visualize for two-step pathways, we begin by examining the evolution of control in two-step pathways.

For largely irreversible reactions ($\alpha = 0.001$ and 0.05), populations starting at random (k_1, k_2) values evolve highly inequitable control coefficients, with the control coefficient (CC) for the first enzyme being on average about three to twenty times greater than that of the second enzyme (Table 2).

Table 2: Control coefficient, Ci , for enzyme i, for two and three-step MCT simulations.

Pathway Length	α	C_1	C_2	C_3	C_1 v C_2 v C_3
2	0.001	0.946 (0.192)	0.054 (0.192)	-	$C_1 > C_2$
	0.05	0.780 (0.316)	0.220 (0.316)	-	$C_1 > C_2$
	0.5	0.580 (0.323)	0.420 (0.323)	-	$C_1 > C_2$
	0.95	0.502 (0.297)	0.498 (0.297)	-	*NS*
3	0.001	0.909 (0.255)	0.036 (0.145)	0.055 (0.219)	$C_1 > C_2 = C_3$
	0.05	0.769 (0.329)	0.155 (0.285)	0.076 (0.213)	$C_1 > C_2 = C_3$
	0.5	0.500 (0.312)	0.298 (0.287)	0.202 (0.262)	$C_1 > C_2 > C_3$
	0.95	0.355 (0.226)	0.332 (0.225)	0.312 (0.220)	*NS*

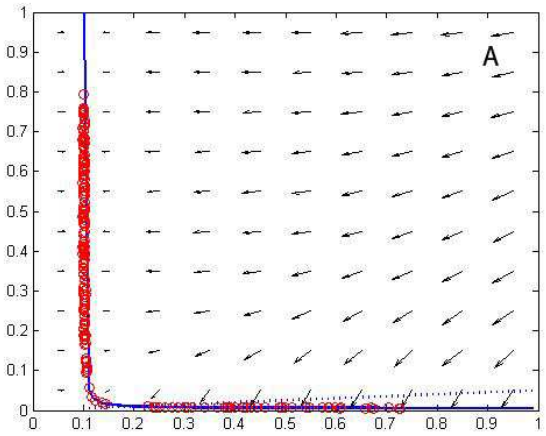
As expected (Figure 1a), populations converge to an equilibrium line corresponding to the hyperbola

$$k_2 = \alpha_1 / [(I/J_{opt}) - (1/k_1)] \qquad (1) ,$$

which is obtained from the two-enzyme equation for flux analogous to (A1). Although different runs converged to different points on this line, the majority of these points lie in the region corresponding to $C_1 > C_2$ (Figure 1a) above the (dashed) $C_1 = C_2$ line:

$$k_2 = \alpha_1 k_1 \qquad (2)$$

derived from Equation A3.



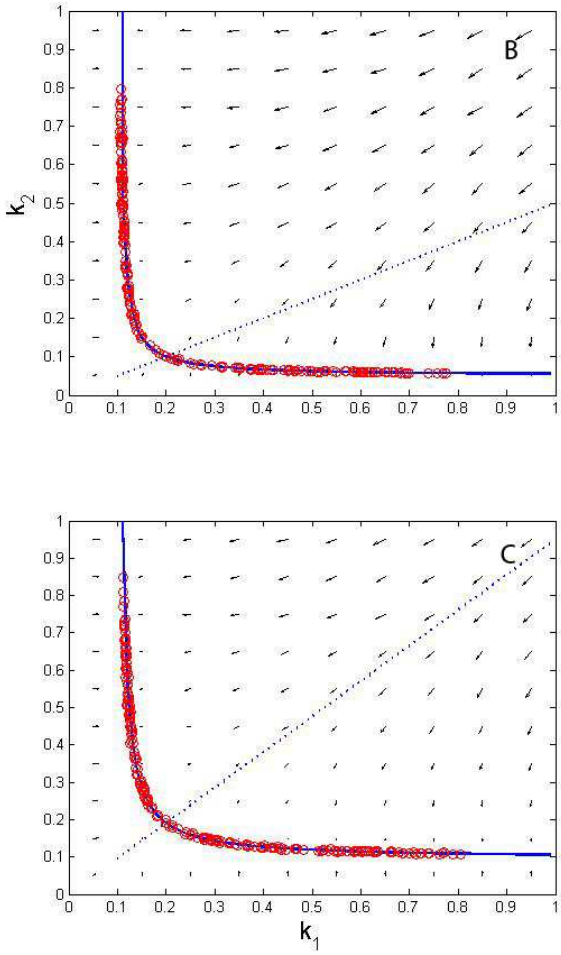


Figure 1: Equilibrium distribution of k1 and k2 for two-step MCT simulations.

Equilibrium reaction rates (red circles) from 200 replicate simulations of the two-enzyme model. Paramerers were: N = 1000, σ^J = 0.5, σ^k = 0.05, I = 10, and Jopt = 1.0. Solid blue lines is equilibrium predicted for MCT model (Equation 1). Dashed lines denotes combinations of k1 and k2 for which C1 = C2 (Equation 2). For all points above this line, C1 > C2 and for all points below, C1 < C2 . Arrows indicate direction of average instantaneous evolutionary trajectories for given starting values. Length of arrows scaled to be visible. A. MCT Simulations with α_i = 0.05. B. MCT Simulations with α_i = 0.5. C. MCT Simulations with α_i = 0.95.

Two factors contribute to this bias in evolutionary outcomes. First, for a randomly chosen starting (k_1 , k_2) there is an initial bias for the CC of the first enzyme to be greater than that of the second enzyme. This bias arises because the space above the $C_1 = C_2$ line (corresponding to $C_1 > C_2$) is much larger than the space below the line (corresponding to $C_1 < C_2$) (Figure 1a). Second, except for starting points very near the $C_1 = C_2$ line, the average instantaneous trajectories are virtually horizontal, *i.e.* the average change in k_1 is large while the average change in k_2 is small (Figure 1a). This means that trajectories from most starting points will tend to move populations toward a point on the equilibrium line lying in the $C_1 > C_2$ region. In fact, this second property is a direct consequence of the first. When $C_1 > C_2$, equivalent mutations affecting k will have larger effects on flux if they occur in enzyme 1. The selection coefficient favoring a beneficial mutation will thus on average be greater for mutations in enzyme 1. Since probability of fixation of an advantageous mutation is approximately proportional to its selection coefficient, mutations in enzyme 1 will have a greater chance of fixation than mutations in enzyme 2.

For reactions with intermediate reversibility ($\alpha_i = 0.5$), the evolutionary outcome is still biased toward larger CC in the first enzyme, but this bias is not as large as in the previous case (Table 2). And for largely reversible reactions ($\alpha_i = 0.95$), there is little bias (Table 2). Paralleling this decrease in bias with increasing reversibility is both a decrease

in the expected proportion of starting points with $C_1 > C_2$ and a decrease in the proportion of instantaneous trajectories for which k_1 is substantially larger in absolute value than k_2 (Figure 1b, c).

To assess the generality of these results, we performed analogous simulations for different combinations of the values of the parameters N, α_i , σ_J , σ_k , I, and J_{opt}. Simulations were run in a factorial design varying all six parameters by at least two orders of magnitude, yielding 972 different comparisons, each with 20 replicates. We used an ANOVA (JMP, SAS Institute Inc., 2005) to analyze the effect of altering these parameters on the equilibrium control coefficient for the two enzymes. As expected, varying α significantly effected mean control coefficients, but changes in the other five of the parameters (N, σ_J , σ_k , I, and J_{opt}) did not significantly alter the mean control coefficient (Table 3).

Table 3: Effects of varying parameters in two-step MCT model.

ANOVA testing effects of varying parameters on the evolution of C1 in two enzyme model. Note that the results for C2 are completely analogous because of the summation property. Parameter levels used in simulations are: αi – 0.001, 0.05, 0.50, 0.95; N - 10, 1000, 100000; σJ - 0.01, 0.10, 1.0; σk- 0.01, 0.10, 1.0; I –0.1, 1.0, 10.0; Jopt - 0.01, 0.10, 1.0.

Variable	df	SS I	*F*	p
α	3	350.66	12.24	0.000
σ_J	2	0.23	0.11	0.894
σ_K	2	52.39	2.68	0.108
J_{opt}	2	20.18	0.97	0.407
N_e	2	1.07	0.35	0.714
I	2	12.24	0.9	0.433
$\alpha * \sigma_J$	6	0.77	1.27	0.265
$\alpha * \sigma_K$	6	10.66	17.67	<.0001
$\alpha * J_{opt}$	6	26.47	43.86	<.0001
$\alpha * N_e$	6	1.01	1.67	0.123
$\alpha *$ ***I***	6	20.79	34.45	<.0001
$\sigma_J * \sigma_K$	4	2.04	5.06	0.000
$\sigma_J * J_{opt}$	4	0.87	2.15	0.072
$\sigma_J * N_e$	4	1.04	2.57	0.036
$\sigma_J *$ ***I***	4	1.23	3.06	0.016
$\sigma_K * J_{opt}$	4	18.73	46.56	<.0001
$\sigma_K * N_e$	4	2.79	6.94	<.0001
$\sigma_K *$ ***I***	4	9.97	24.78	<.0001
$J_{opt} * N_e$	4	2.66	6.62	<.0001
$J_{opt} *$ ***I***	4	3.29	8.18	<.0001
$N_e *$ ***I***	4	0.55	1.36	0.243

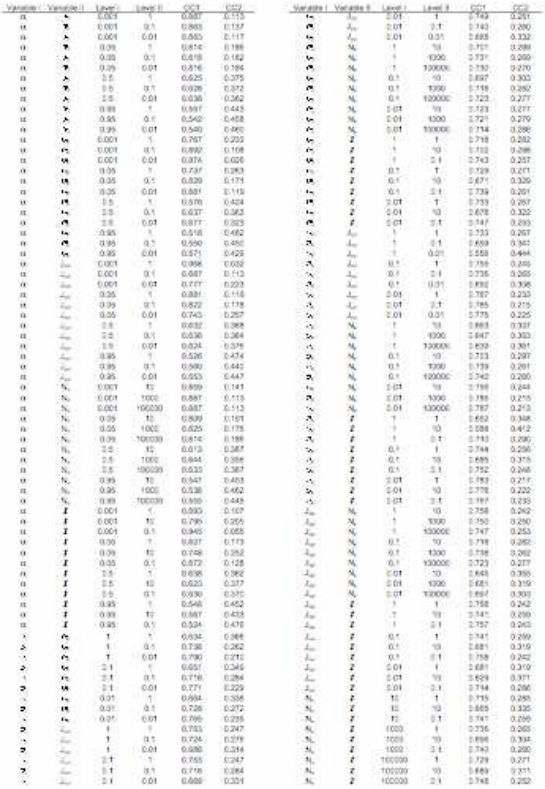
Many of the interaction terms were statistically significant (Table 3), but these effects are subtle and do not change the fundamental evolutionary trend: the mean control coefficient for the first enzyme is larger then that of the second for every comparison (Table 4).

Table 4: Least Square Mean control coefficients for a. each variable level and b. two way interaction between variables for the two-step MCT model.

A.

Variable	Level	CC_1	CC_2
α	0.001	0.87763759	0.122
α	0.05	0.81577334	0.184
α	0.5	0.63023165	0.370
α	0.95	0.54651633	0.453
σ_J	1	0.72066554	0.279
σ_J	0.1	0.71274675	0.287
σ_J	0.01	0.71920689	0.281
σ_K	1	0.6496859	0.350
σ_K	0.1	0.72718991	0.273
σ_K	0.01	0.77574337	0.224
J_{tot}	1	0.75179215	0.248
J_{tot}	0.1	0.72643792	0.274
J_{tot}	0.01	0.67438912	0.326
N_e	10	0.70707531	0.293
N_e	1000	0.72343627	0.277
N_e	100000	0.7221075	0.278
I	1	0.72635841	0.274
I	10	0.68335918	0.317
I	0.1	0.7429016	0.257

B.



The previous analyses demonstrate that a shift in the flux optimum will, *when averaged over random starting points*, tend to yield a new equilibrium at which flux control is biased toward the upstream enzyme. However, it is not known whether all possible starting combinations of (k_1, k_2) are equally likely. Moreover, the analysis does not necessarily imply that equilibria at which the majority of control is vested in the *second*

enzyme are not evolutionarily stable in a shifting environment. It is thus conceivable that the above analysis fails to completely capture the dynamics of CC evolution in an environment in which the optimal flux varies.

To examine this issue, we simulated evolutionary trajectories in which the optimal flux fluctuated between two values, with shifts in the optimum occurring every 30,000 mutation-fixation cycles. Simulations were performed with six pairs of optimal flux (1.0, 0.5; 1.0, 3.0; 1, 5.0; 3.0, 0.5; 3.0, 1.0; 3.0, 5.0). Simulations were initiated with all replicates at optimal fitness, equally spaced along the high fitness hyperbola, with half of the starting points at $C_1 > C_2$ and half $C_2 > C_1$. The parameters α and σ_k were varied over at least an order of magnitude across replicate simulations to determine the effect on the equilibrium distribution of control. To determine the long-term distribution of genotypes along the equilibrium curve we used our simulations to estimate transition probabilities from one class of (k_1, k_2) values to all other classes. Using this information, we constructed a transition matrix, M. Each element, m_{ij} of this matrix represents the proportion of replicates starting at the midpoint of segment *j* that ended within segment *i*. Under the approximation that this transition probability is the same for any starting point within segment *j*, the long-term, equilibrium probability that a population will be in segment *k* is just the k^{th} element of the leading eigenvector of M.

In general, modeling evolution in a fluctuating environment produced results similar to those already obtained: many more trials yielded greater control by the upstream enzyme ($C_1 > C_2$) than the reverse (Table 5). The few exceptions to this trend occur only when mutations are small and reactions are of intermediate or high reversibility, a condition that appears uncommon in biochemical reactions (see DISCUSSION).

Table 5: Proportion of pathways that evolve C1 > C2 in shifting environmental optimum.

Populations were initialized with equal number of pathways with C1 > C2 and C1 < C2. Pathways evolved under six regimes of shifting environmental optima: 1.0 – 0.5; 1.0 – 3.0; 1.0 – 5.0; 3.0 – 0.5; 3.0 – 1.0; 3.0 – 5.0. Only 3 combinations of parameters (bolded type) resulted in a majority of pathways evolved C2 > C1.

J_{opt} Shift	α	σ_k 0.01	0.05	0.1
1.0 - 0.5	0.001	1.000	1.000	1.000
	0.05	0.993	1.000	0.999
	0.5	0.999	1.000	0.992
	0.95	0.806	0.948	0.913
1.0 - 3.0	0.001	0.831	0.850	0.726
	0.05	0.870	0.823	0.799
	0.5	0.778	0.861	0.810
	0.95	**0.430**	0.778	0.795
1.0 - 5.0	0.001	0.688	0.748	0.835
	0.05	0.643	0.652	0.715
	0.5	**0.405**	0.659	0.688
	0.95	**0.213**	0.560	0.653
3.0 - 0.5	0.001	1.000	1.000	0.972
	0.05	1.000	0.967	0.951
	0.5	1.000	0.953	0.852
	0.95	0.809	0.941	0.898
3.0 - 1.0	0.001	0.972	0.959	0.960
	0.05	1.000	0.862	0.871
	0.5	0.996	0.903	0.909
	0.95	0.813	0.937	0.919
3.0 - 5.0	0.001	1.000	1.000	1.000
	0.05	0.998	1.000	1.000
	0.5	0.994	0.999	0.989
	0.95	0.528	0.843	0.929

1.3.3 Evolution of control under the three-step MCT model

As with the two-step MCT model, a total of 200 trials were run with this model for each α value, each trial beginning from a randomly chosen starting point (k_1 , k_2 , k_3). In these simulations, the evolution of CC in a three-step pathway exhibits similar patterns to those described above for a two-step pathway. All trials converged to an equilibrium hyperboloid represented by Equation (1), with $J = J_{opt} = 1$ (Figure 2).

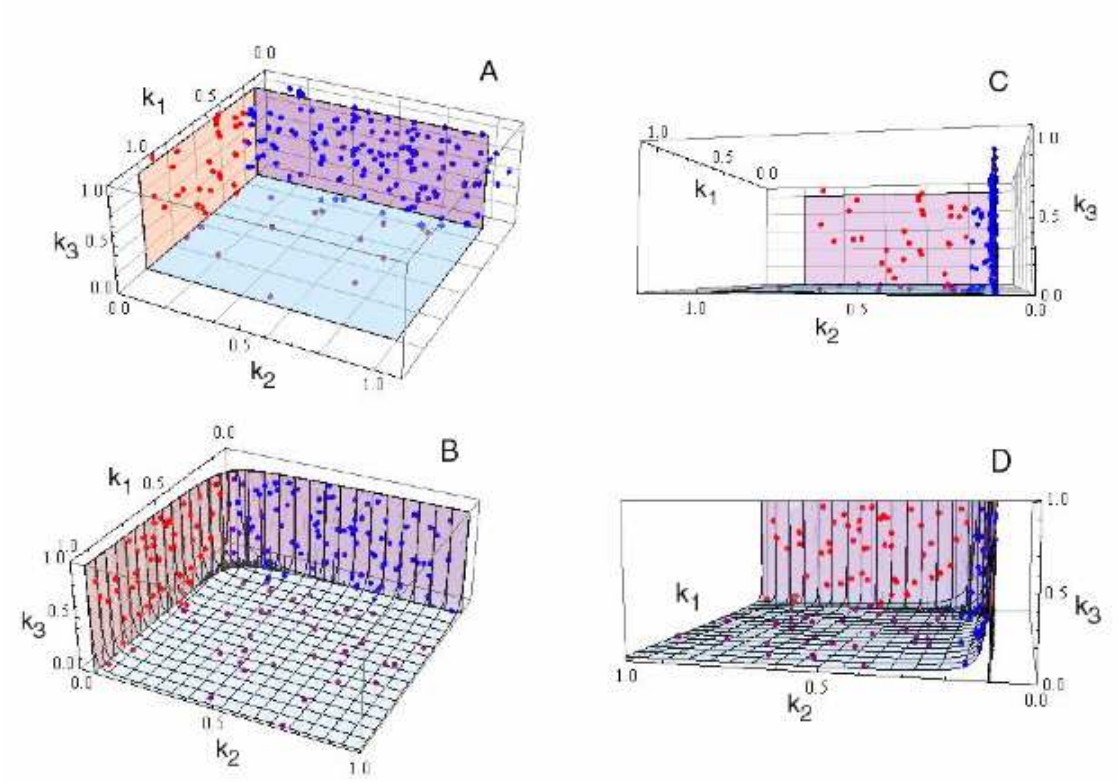


Figure 2: Equilibrium distribution of k1 , k2 , k3 for three-step MCT simulations.

Optimal flux surface and equilibrium reaction rates from 200 replicate simulations of the three-enzyme MCT model. Parameters as described for Figure 1.

Blue points: equilibria for which C1 > C2, C3. Red points: equilibria for which C2 > C1, C3. Purple points: equilibria for which C3 > C1, C2 A. αi = 0.05. Two alternate views of the optimal flux surface (grid lines) is divided into three regions, differently colored: purple corresponds to C1 > C2, C3. Orange corresponds to C2 > C1, C3. Light blue corresponds to C3 > C1, C2 . B. As A, but with αi = 0.5.

As for the two-step pathway, when $\alpha = 0.001$ and $\alpha_i = 0.05$, there is a strong bias for trajectories to converge to points having high CC for the first enzyme: the mean C_1 is twenty times greater than the mean C_2 and C_3 for $\alpha = 0.001$ and more than three times greater for $\alpha = 0.05$ (Table 2; Figure 2a). This bias is reduced for $\alpha_i = 0.5$ and essentially absent when $\alpha_i = 0.95$ (Table 2; Figure 2b, Figure 3).

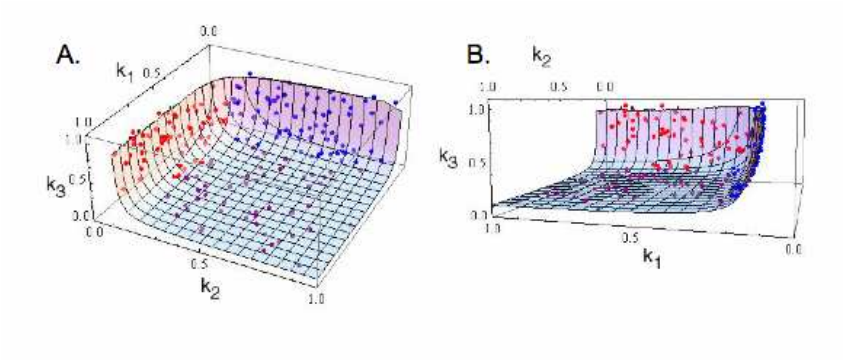


Figure 3: Equilibrium distribution of k1 , k2 , k3 for three-step MCT simulations when αi = 0.95.

Optimal flux surface and equilibrium reaction rates from 200 replicate simulations of the three-enzyme MCT model with αi = 0.95. Parameters as in Figure 2. A. and B. are same graph from different viewpoints.

This pattern is not dependent on the values of the fixed parameters in our model. As in the two-step model, we generated simulations for all combinations of parameters values (N, α_i , σ_J , σ_k , I, and J_{opt}), each of which was varied over two orders of magnitude. Once

again, we found that effect of α_i on the mean values of the control coefficients for each enzyme were significant, while variation in the other five parameters (N, σ_J, σ_k , I, and J_{opt}) had little effect on CC for any of the enzymes (Tables 6, 7).

Table 6: Effects of varying parameters in three-step MCT model.

ANOVA testing effects of varying parameters on the evolution of control in the three-step pathway; a. C1 b. C2 c. C3 . Parameter levels used in simulations are: αi – 0.001, 0.05, 0.50, 0.95; N - 10, 1000, 100000; σJ - 0.01, 0.10, 1.0; σk- 0.01, 0.10, 1.0; I –0.1, 1.0, 10.0; Jopt - 0.01, 0.10, 1.0.

A.

Variable	df	SS I	F	p
α	3	749.43	14.69	<.0001
σ	2	0.07	0.1	0.907
σ	2	1.87	0.11	0.898
J	2	1.75	0.13	0.881
N	2	8.46	1.37	0.309
I	2	1.44	0.1	0.908
α * σ	6	1.10	2.23	0.037
α * σ	6	30.16	61.41	<.0001
α * J	6	29.20	59.45	<.0001
α * N	6	6.04	12.3	<.0001
α * I	6	37.48	76.32	<.0001
σ * σ	4	0.89	2.7	0.029
σ * J	4	0.38	1.15	0.330
σ * N	4	0.40	1.23	0.297
σ * I	4	0.34	1.03	0.390
σ * J	4	4.98	15.2	<.0001
σ * N	4	8.19	25.01	<.0001
σ * I	4	1.53	4.66	0.001
J * N	4	0.37	1.13	0.338
J * I	4	3.34	10.2	<.0001
N * I	4	0.64	1.95	0.099

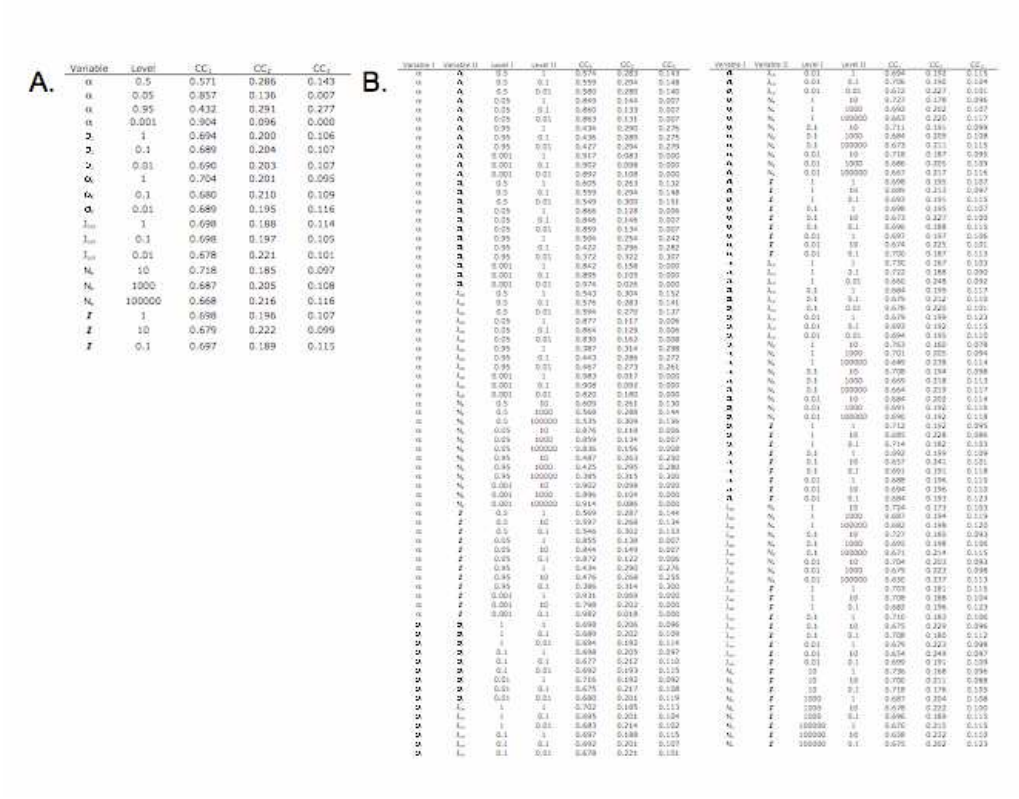
B.

Variable	df	SS I	F	p
α	3	147.96	4.14	0.021
σ	2	0.05	0.09	0.916
σ	2	0.75	0.07	0.937
J	2	3.72	0.34	0.719
N	2	3.23	1.06	0.396
I	2	3.91	0.33	0.731
α * σ	6	0.82	2.78	0.011
α * σ	6	18.92	64.25	<.0001
α * J	6	22.00	74.73	<.0001
α * N	6	2.25	7.65	<.0001
α * I	6	28.69	97.44	<.0001
σ * σ	4	0.57	2.89	0.021
σ * J	4	0.35	1.77	0.133
σ * N	4	0.28	1.41	0.229
σ * I	4	0.21	1.06	0.374
σ * J	4	4.40	22.44	<.0001
σ * N	4	4.41	22.47	<.0001
σ * I	4	1.78	9.08	<.0001
J * N	4	0.11	0.58	0.679
J * I	4	3.15	16.07	<.0001
N * I	4	0.59	2.98	0.018

C.

Variable	df	SS I	F	p
α	3	250.95	90.22	<.0001
σ	2	0.00	0.31	0.854
σ	2	1.49	1.38	0.299
J	2	0.56	1.61	0.259
N	2	1.24	1.62	0.247
I	2	0.76	2.11	0.209
α * σ	6	0.06	1.18	0.315
α * σ	6	2.33	44.85	<.0001
α * J	6	0.88	16.91	<.0001
α * N	6	1.38	26.66	<.0001
α * I	6	1.12	21.52	<.0001
σ * σ	4	0.07	1.88	0.111
σ * J	4	0.01	0.32	0.867
σ * N	4	0.02	0.63	0.642
σ * I	4	0.02	0.58	0.676
σ * J	4	0.06	1.86	0.114
σ * N	4	0.61	17.5	<.0001
σ * I	4	0.01	0.37	0.830
J * N	4	0.11	3.05	0.016
J * I	4	0.07	2.07	0.082
N * I	4	0.01	0.33	0.859

Table 7: Least Square Mean of control coefficients for a. each variable level and b. two way interaction between variables in three enzyme MCT model.



1.3.4 Evolution of control under the ten-step MCT model

As a final examination of the generality of our results, additional simulations using a model of a longer metabolic pathway containing ten enzymes. Two hundred replicates pathways were evolved for 20,000 mutation/fixation cycles with $\alpha = 0.001$ and 0.05, CC decrease rapidly with position in the pathway and on average the degree of control for the first enzyme is approximately eight and three times greater, respectively, than that of the second enzyme (Table 8). With $\alpha = 0.5$, this bias is less, but the first two

enzymes share two-thirds of the control, and the first three enzymes account for more than 75% of the control. Finally, as with the shorter pathways, when $\alpha = 0.95$, control is much more equitably distributed on average, with the CC of the first and last enzymes differing by only 0.032 (Table 8).

Table 8: Mean (standard deviation) of equilibrium control coefficients for ten-step MCT simulations.

α	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
0.001	0.849 (0.317)	0.096 (0.254)	0.054 (0.216)	0	0	0	0	0	0	0
0.05	0.665 (0.368)	0.197 (0.310)	0.041 (0.153)	0.031 (0.151)	0.044 (0.202)	0.022 (0.080)	0	0	0	0
0.5	0.38 (0.262)	0.247 (0.236)	0.121 (0.159)	0.088 (0.154)	0.059 (0.118)	0.028 (0.083)	0.022 (0.078)	0.024 (0.089)	0.02 (0.107)	0.01 (0.076)
0.95	0.117 (0.021)	0.115 (0.022)	0.109 (0.018)	0.106 (0.019)	0.103 (0.020)	0.099 (0.019)	0.093 (0.017)	0.088 (0.017)	0.085 (0.018)	0.085 (0.019)

1.3.5 Properties of fixed mutations under the MCT model

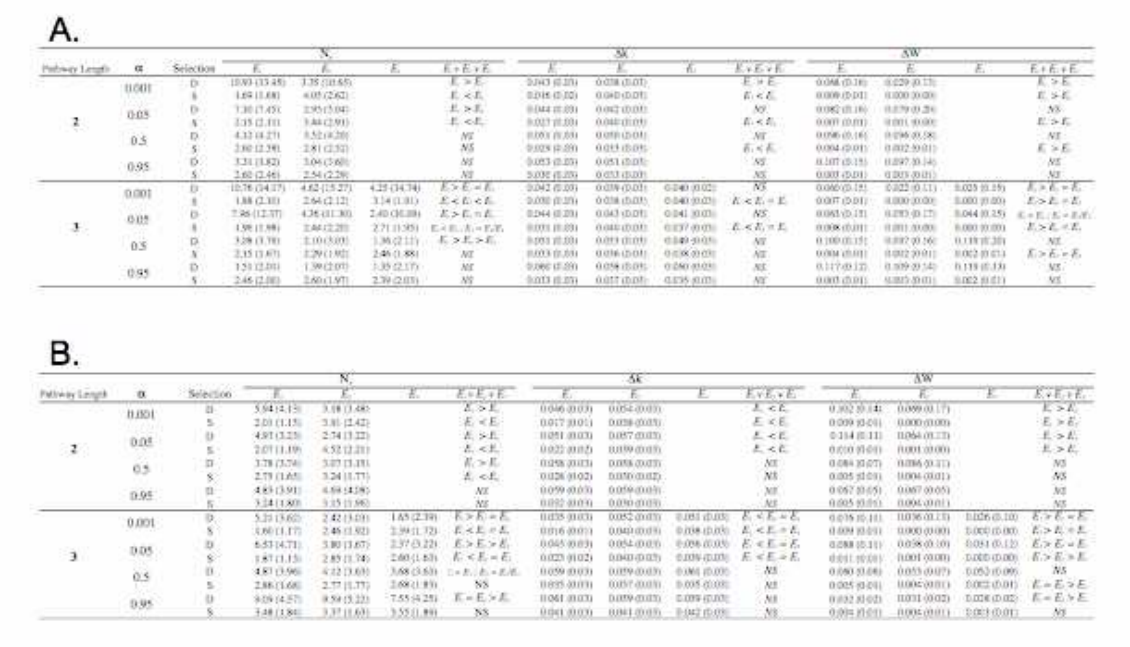
We examined three properties of fixed mutations: the number of substitutions (N_μ) in a trajectory, the change in enzyme activity caused by each mutation (Δk_i), and the change in fitness caused by each mutation (ΔW). We compared the value of these properties for different enzyme steps, during different phases of an evolutionary trajectory. Adaptive walks were divided into two phases: the first phase when the population is far from the optimum ($W(\boldsymbol{J}) < 0.95$) and evolution is dominated by directional selection, and a second phase when the population is near the optimum ($W(\boldsymbol{J}) > 0.95$) and evolution is dominated by stabilizing selection.

For pathways with low reversibility ($\alpha_i = 0.001$ and 0.05), during the first phase of the adaptive walk, 2-3 times as many substitutions occurred in the first enzyme as in the second or third enzymes (Table 9a). Although the average effect of these substitutions

on *k* were similar for the different enzymes, the average effect on fitness was larger for the first enzyme (Table 9a). Moreover, many more substitutions improved fitness in the first enzyme, compared to those downstream (α_i = 0.001: 2E pathway- 44.0% versus 10.3%; 3E pathway- 38.2% versus 7.3% & 2.7%). In the second phase, the same pattern emerges in the distribution of beneficial mutations between up and downstream enzymes (α_i = 0.001: 2E pathway- 91.6% versus 26.1%; 3E pathway- 93.7% versus 35.1% & 0.5%). There were many more substitutions that had no effect on fitness during the first phase then during the second phase. This unexpected result is caused by mutations that were fixed primarily when populations were far from the fitness optimum and the landscape was essentially flat. The average effect on *k* during the second phase was significantly smaller for the first enzyme than for the more downstream enzymes, presumably reflecting the fact that for enzymes with greater flux control, only mutations with very small effects will either not overshoot the optimum or will have small enough fitness effects to be fixed by drift (Table 9a). Nevertheless, the effects on fitness were largest for the first enzyme, though ΔW is approximately an order of magnitude lower than in the first phase (Table 9a). These patterns indicate that during the directional selection phase, adaptive substitutions are fixed primarily in the first enzyme, and that neutral mutations occur in both phases of selection, but are fixed most frequently in downstream enzymes.

Table 9: The mean values of Nμ , Δk, ΔW for two and three step MCT and SK models.

The mean values (standard deviation in parentheses) of Nμ , Δki, ΔW. All values were calculated from 200 replicate simulations of two and three-step pathways initialized with random values of ki. Mutations were determined to occur under one of two selective regimes: directional (D) W < 0.95 or stabilizing (S) W ≥ 0.95. Comparison of means between pathway steps conducted using post-hoc T-test or Tukey test, significant threshold: α = 0.05. A. MCT simulations. B. SK simulations.



Pathways with intermediate to high reversibility (α_i = 0.5 or 0.95) yield a very different pattern. In particular, enzymes tend not to differ in either phase for number of substitutions, or their effects on enzyme activity or fitness (Table 9). These results demonstrate that a fairly high degree of irreversibility is needed to cause upstream and downstream enzymes to differ in the properties of their associated fixed mutations.

The results for three properties of fixed mutations investigated are robust to the defined threshold between the two phases of adaptive evolution ($W > 0.95$ or $W > 0.99$) and the distribution of mutational effects (results not shown). We analyzed three Gaussian distributions with mutations skewed toward decreasing enzyme function ($\mu = 0.0$, $\mu = -\sigma_k$, $\mu = -2\sigma_k$) and found no effect on final distributions of N_μ , ΔW, or Δk_1.

1.3.6 Evolution of control under the SK model:

The saturation kinetics model contains one more parameter than the MCT model. Under saturation kinetics, when downstream enzymes have slower rates of reaction, there is the potential for the buildup of large pools of pathway intermediates. Because high concentrations of intermediates are likely to be deleterious, we follow Clark (1991) in introducing a fitness penalty for high substrate pools, represented by the parameter **T** (see Appendix A). We first examine an exemplar set of simulations for the two- and three-enzyme models in which $I = T = 10$, $N = 1000$, $J_{opt} = 0.5$, $\sigma_J = 0.05$, and $\sigma_k = 0.05$, then consider how the behavior of the model changes as these parameters are varied. For simplicity, we again consider the case where α_i are the same for all i and equal to α. Two-hundred trials were run with for the SK model for each α value, each trial beginning from a randomly chosen starting point (V_1 , V_2).

The evolution of control in these exemplar simulations of two- and three-enzyme pathways yields patterns that are very similar to those exhibited by the MCT model (Supplementary Table 6).

Table 10: Equilibrium control in two and three-step SK simulations.

Mean (and standard deviations) control coefficients for SK model. Average enzyme control generated from 200 replicate simulations. Variation in enzyme control analyzed using post-hoc Tukey test (a = 0.05).

Pathway Length	α	C_1	C_2	C_3	C_1 v C_2 v C_3
2	0.001	0.904 (0.006)	0.005 (0.007)	-	$C_1 > C_2$
	0.05	0.806 (0.085)	0.080 (0.080)	-	$C_1 > C_2$
	0.5	0.401 (0.196)	0.355 (0.211)	-	$C_1 > C_2$
	0.95	0.330 (0.131)	0.315 (0.140)	-	*NS*
3	0.001	0.829 (0.245)	0.033 (0.152)	0.030 (0.142)	$C_1 > C_2 = C_3$
	0.05	0.806 (0.076)	0.061 (0.064)	0.017 (0.021)	$C_1 > C_2 > C_3$
	0.5	0.237 (0.090)	0.228 (0.103)	0.185 (0.125)	$C_1 = C_2 > C_3$
	0.95	0.161 (0.042)	0.152 (0.038)	0.143 (0.046)	*NS*

In this model, Equation A12 with $v = J_{opt}$ defines a line (2-enzyme pathway) or surface (3-enzyme pathway) of optimal flux corresponding to the set of enzyme maximal velocities, V_1 , V_2, V_3, that make the reaction velocity equal to the optimum (blue curves in Figure 4 a,b, surfaces in Figure 5 a,b). This line or suface is divided into distinct regions corresponding to control by the different enzymes (dashed line in Figure 4 a,b, blue line in Fig. 5a).

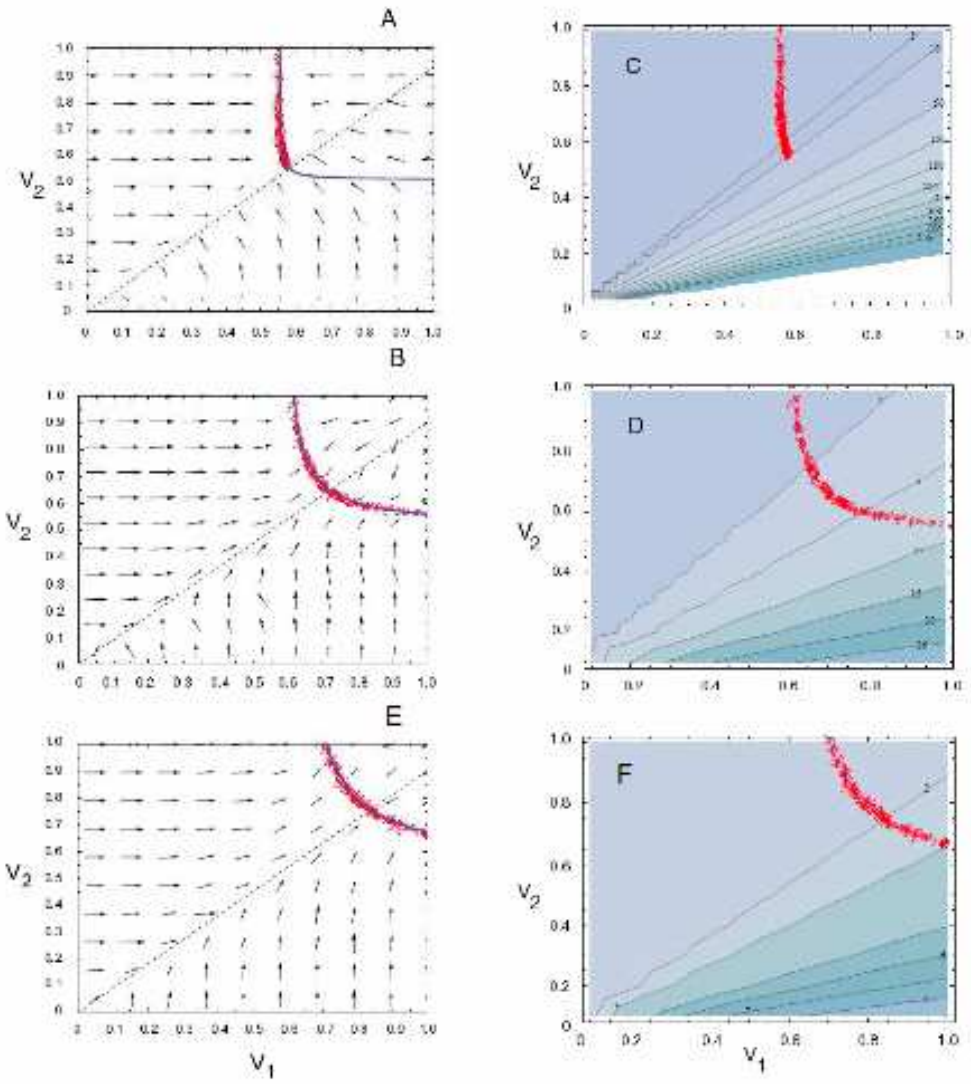


Figure 4 Equilibrium distribution of V1 and V2 for two-step SK simulations.

Equilibrium maximum reaction velocities (red circles) from 200 replicate simulations of the two-enzyme SK model. Parameters were: N = 1000, σJ = 0.05, σk = 0.05, I = T = 10, and Jopt = 0.5. Vectors as in Figure 1. Dashed line represents combinations of V1 and V2 for which C1 = C2 derived from Equation A13. A. αi = 0.05. B. αi = 0.5. E. αi = 0.05. Contours of pool size of intermediate B derived from

Equation A14. C. countour of pool size $\alpha i = 0.05$. D. countour of pool size $\alpha i = 0.50$. F. countour of pool size $\alpha i = 0.95$.

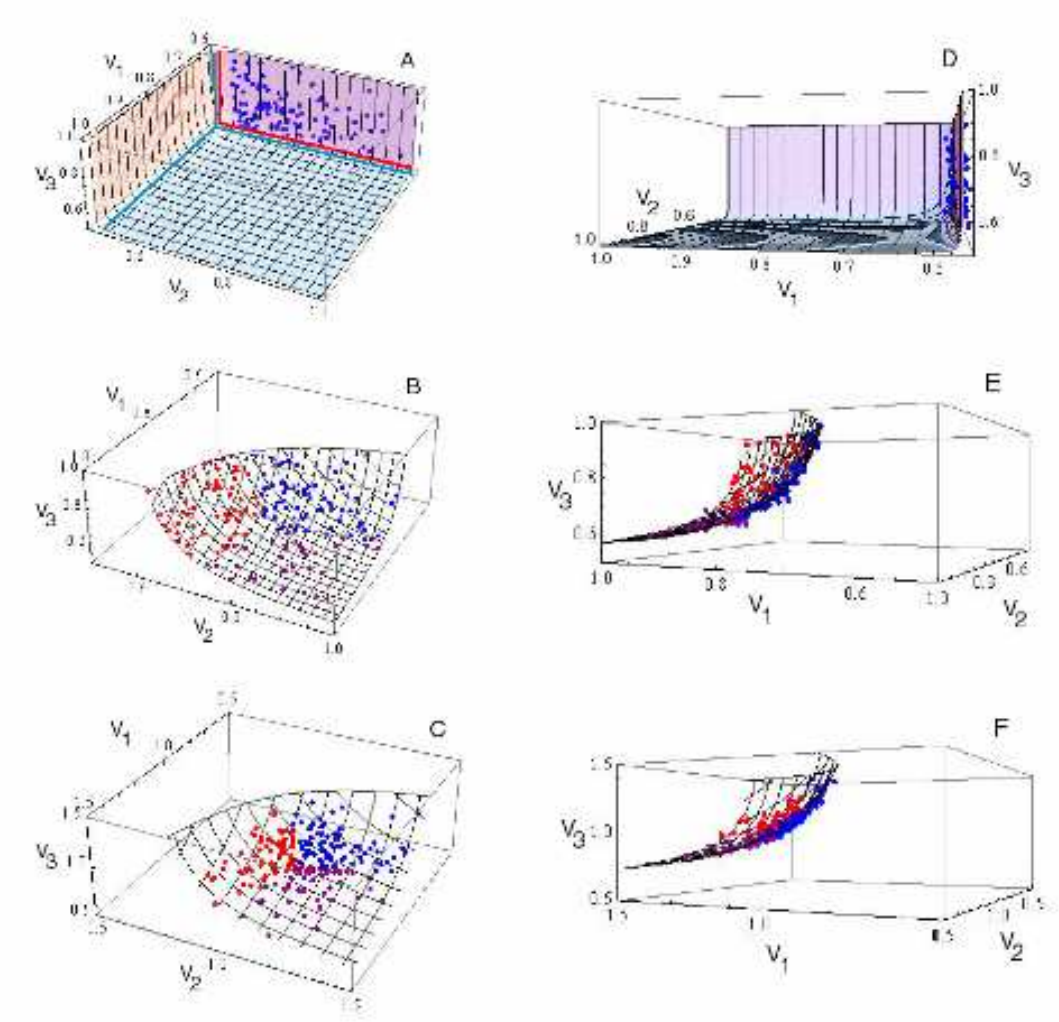


Figure 5: Equilibrium distribution of V1 , V2 , V3 for three-step SK simulations.

Optimal flux surface and equilibrium maximum reaction velocities from 200 replicate simulations of the three-enzyme SK model. Parameters as described for Figure 4. Blue points: equilibria for which C1 > C2, C3. Red points: equilibria for which C2 > C1, C3. Purple points: equilibria for which C3 > C1, C2 A. αi = 0.05.

Alternate views of optimal flux surface. Blue line divides regions based on which enzyme exerts greatest control: purple corresponds to C1 > C2, C3. Orange corresponds to C2 > C1, C3. Light blue corresponds to C3 > C1, C2 . For positions on the optimal flux surface above and to the right of the red line, the size of the intermediate pool is lower than the threshold T. B. As A, but for αi = 0.5. All points on the optimal flux surface have pool sizes < T.

For both two- and three-enzyme pathways, our simulations yielded final values of the V_i that lie on or very close to the line/surface of optimal flux (Figures 4 and 5). Slight deviations from the surface largely reflect nearly neutral substitutions fixed by drift after a population has evolved very close to the optimum. However, the portion of the optimal line/surface occupied depends on the reversibility of the reaction. This point is clearly illustrated for the 2-enzyme model with low reversibilty: all points lie in the $C_1>C_2$ region because of the threshold on the size of substrate pools (Figures 4c), even though the optimal flux hyperbola extends into the $C_2>C_1$ region (blue line, Figure 4a). Enzyme activity values on the optimum line/surface will produce very high intermediate substrate levels when V_2 (or also possibly V_3 in the 3-enzyme pathway) is very close to v, making $V_2 - v$ (or $V_3 - v$) very small. Because the concentration of the intermediate, B, is inversely proportional to $V_2 - v$ (and, in the 3-enzyme pathway, intermediate C concentration is proportional to $V_3 - v$) (Appendix A), the intermediate pool becomes very large and a strong fitness penalty is imposed. Thus, for largely irreversible reactions, fitness penalties associated with large pools of the intermediate constrain the equilibrium to be in the region in which $C_1>C_2$. If the penalty threshold T

were greater, some trials would evolve to equilibria with enzyme 2 (or 3) exhibiting greatest control. However, unless T is unreasonably large, a majority of trials will yield equilibria with enzyme 1 having greatest control.

With moderately to highly reversible reactions ($\alpha = 0.5, 0.95$), this constraint is absent (Figure 4d, f) and consequently the system evolves to equilbria that lie along the entire line/surface of optimal flux (Figure 4b,e, and Figure 5b,c). In the two-enzyme pathway, approximately half the equilibria correspond to $C_1 > C_2$, and half to $C_2 > C_1$; in the three-enzyme pathway, approximately one-third of the equilibria correspond to greatest control by each of the enzymes. Our interpretation of this result is as follows: in a reversible reaction, the net forward flux is countered by a backward flux, even if the intermediate pool is small. Consequently, in order to achieve a given optimal flux, the V_i must be higher than when reactions are less reversible. This effect displaces the line/surface of optimal flux toward higher values of V_1, V_2 (Figure 4 a,b:, blue lines shifted toward higher V_1 and V_2 with increasing reversibility) or V_1, V_2, V_3 (Fig 5a,b: optimum surface shifted toward higher V_i). Because of this displacement, at all points along the line/surface V_2 (and V3) is substantially larger than the actual flux v (0.5 in this example). Consequently, $V_2 - v$ (and $V_3 - v$) remains relatively large and the intermediate pool remains relatively small, avoiding the fitness penalty associated with high intermediate pools. In these examples, pools corresponding to the line/surface of

optimal flux never exceed the threshold value of T (Figure 4d,f and Figure 5b,c). There is thus no constraint preventing equilibria outside the region in which enzyme 1 exerts greatest control.

In summary, a bias favoring dominant control by enzyme 1 can arise in pathways with low reaction reversibility if there is a fitness penalty associated with large pools of intermediates. This penalty does not cause the V_i to deviate from the surface of optimal flux, but rather restricts the portion of that surface that corresponds to a high-fitness equilibrium. By contrast, this restriction greatly decreases in pathways with moderately or highly reversible reactions because reversibility prevents large intermediate pools from building up. Consequently, all enzymes have a roughly equal chance of exerting predominant control over flux.

To assess the generality of these results, we performed simulations for different combinations of the values of the parameters α, J_{opt}, σ_k, σ_J, I, and T. Simulations were run in a factorial design varying all six parameters by at least two orders of magnitude, each with 25 replicates. We employed a restricted factorial design to avoid simulations run with unrealistic parameter values and decrease the computer running time of these simulations (for details see Table 11). The data was analyzed with an ANOVA testing effects of single parameters and two-way interactions on the distribution of control.

Table 11 - Effects of varying parameters in two-step SK model.

ANOVA testing effects of varying parameters on the distribution of control in the two-step SK model. Note that the results for C1 and C2 are completely analogous, because they are analyzed as percentage of total control distributed among each enzyme. Parameter levels used in simulations are: αi – 0.001, 0.05, 0.50, 0.95; Jopt - 0.01, 0.10, 1.0; σk- 0.01, 0.10, 1.0; σJ - 0.001, 0.01, 0.1; I – 1.0, 10.0, 100; T – 1.0, 10.0, 100.0. Not all pairwise combinations were investigated, the restrictions employed were: 1) Jopt ≥ σJ 2) Jopt ≥ σk 3) I > Jopt 4) T > Jopt 5) T ≥ I . The restrictions on specific parameter combinations, resulted in a loss of degrees of freedom for many statistical comparisons.

Variable	df	SS I	*F*	p	
α	3	8.05	20.47	<.0001	
J_{opt}	1	0.24	1.81	0.18	LostDFs
σ_K	0	0.00	.	.	LostDFs
σ_J	1	3.16	24.08	<.0001	LostDFs
I	2	2.86	10.92	<.0001	
T	0	0.00	.	.	LostDFs
α * J_{opt}	6	6.38	8.11	<.0001	
α * σ_K	6	4.37	5.56	<.0001	
α * σ_J	6	0.75	0.96	0.45	
α * ***I***	6	13.48	17.14	<.0001	
α * **T**	6	9.94	12.64	<.0001	
J_{opt} * σ_K	1	0.90	6.83	0.009	LostDFs
J_{opt} * σ_J	3	6.71	17.06	<.0001	LostDFs
J_{opt} * ***I***	3	3.05	7.76	<.0001	LostDFs
J_{opt} * **T**	3	4.72	12.00	<.0001	LostDFs
σ_K * σ_J	4	2.23	4.25	0.002	
σ_K * ***I***	3	0.55	1.40	0.24	LostDFs
σ_K * **T**	3	0.43	1.10	0.35	LostDFs
σ_J * ***I***	4	1.02	1.95	0.10	
σ_J * **T**	4	1.62	3.09	0.01	
I * **T**	1	1.94	14.80	0.0001	LostDFs

Varying two parameters, σ_J, and *I*, had a significant effect on the distribution of control (Table 11 and Table 12), but these effects are subtle and do not change the fundamental pattern: control coefficients evolve to be larger for upstream enzymes (Table 13 and Table 14). However, there were two instances in the three-enzyme case in which the second enzyme had the greatest control coefficient in the pathway: α =0.95, σ_J = 0.01 ; σ_J = 0.01, *I* = 100 (Table 14). Additionally, there is a third instance (*I* = 100, *T* = 100) in which control is quite evenly shared between all three enzymes. These three cases

demonstrate that, for a few parameter combinations, control can evolve to be centered in a downstream enzyme. Regardless, the fundamental evolutionary pattern for this model, and every other model investigated, is that control evolves to be centered in the most upstream enzyme.

Table 12: Effects of varying parameters in three-step SK model.

ANOVA testing effects of varying parameters on the evolution of CC in three enzyme SK model. a. C1 b. C2 c. C3 Parameter levels used in simulations are: αi – 0.001, 0.05, 0.50, 0.95; Jopt - 0.01, 0.10, 1.0; σk- 0.01, 0.10, 1.0; σJ - 0.001, 0.01, 0.1; I – 1.0, 10.0, 100; T – 1.0, 10.0, 100.0.

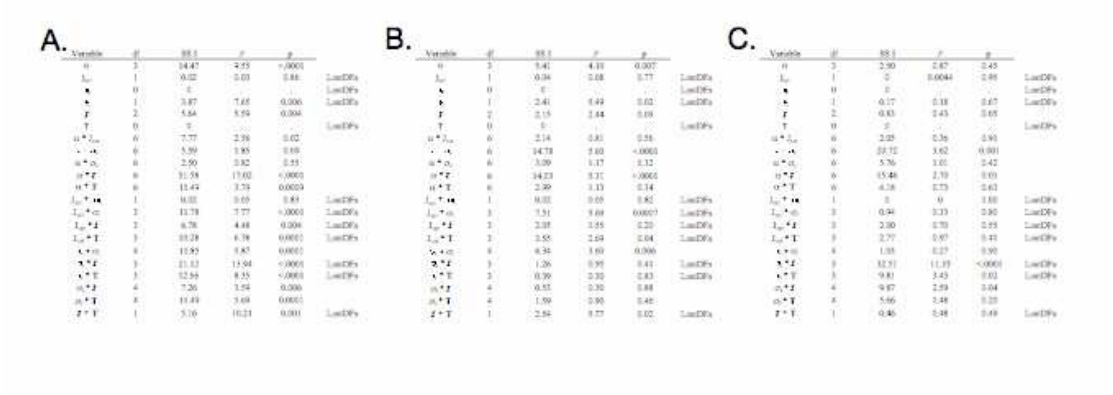


Table 13: Least Square Mean control coefficients for a. each variable level and b. two-way interaction between variables for two step SK model.

A.

Variable	Level	% C_1	% C_2
α	0.001	1.046	-0.046
α	0.05	1.055	-0.055
α	0.5	0.915	0.085
α	0.95	0.725	0.275
λ_{int}	0.01	1.046	-0.046
λ_{int}	0.1	0.983	0.017
λ_{int}	1	-	-
λ_{ex}	0.01	1.046	-0.046
λ_{ex}	0.1	-	-
λ_{ex}	1	-	-
κ	0.001	1.046	-0.046
κ	0.01	0.852	0.148
κ	0.1	-	-
J	1	1.046	-0.046
J	10	0.971	0.029
J	100	0.828	0.172
T	1	1.046	-0.046
T	10	-	-
T	100	-	-

B.

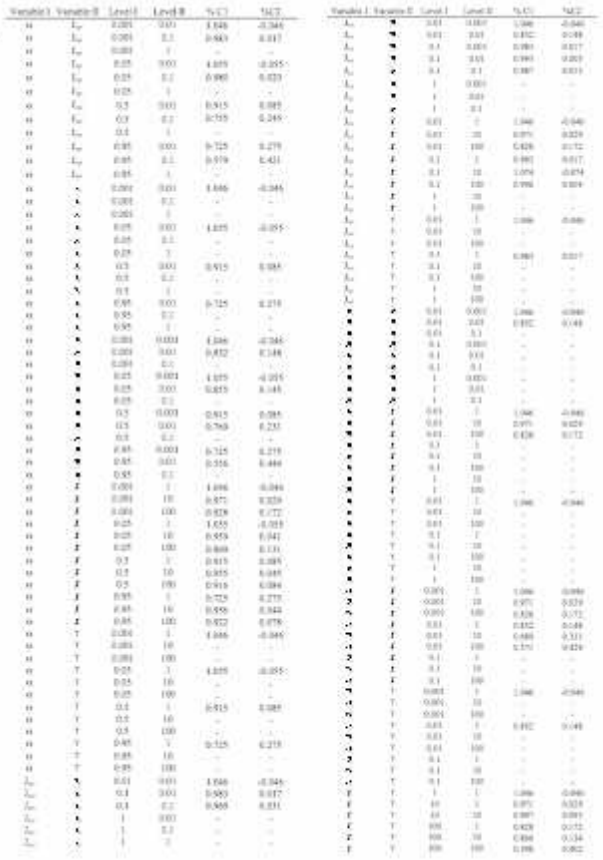
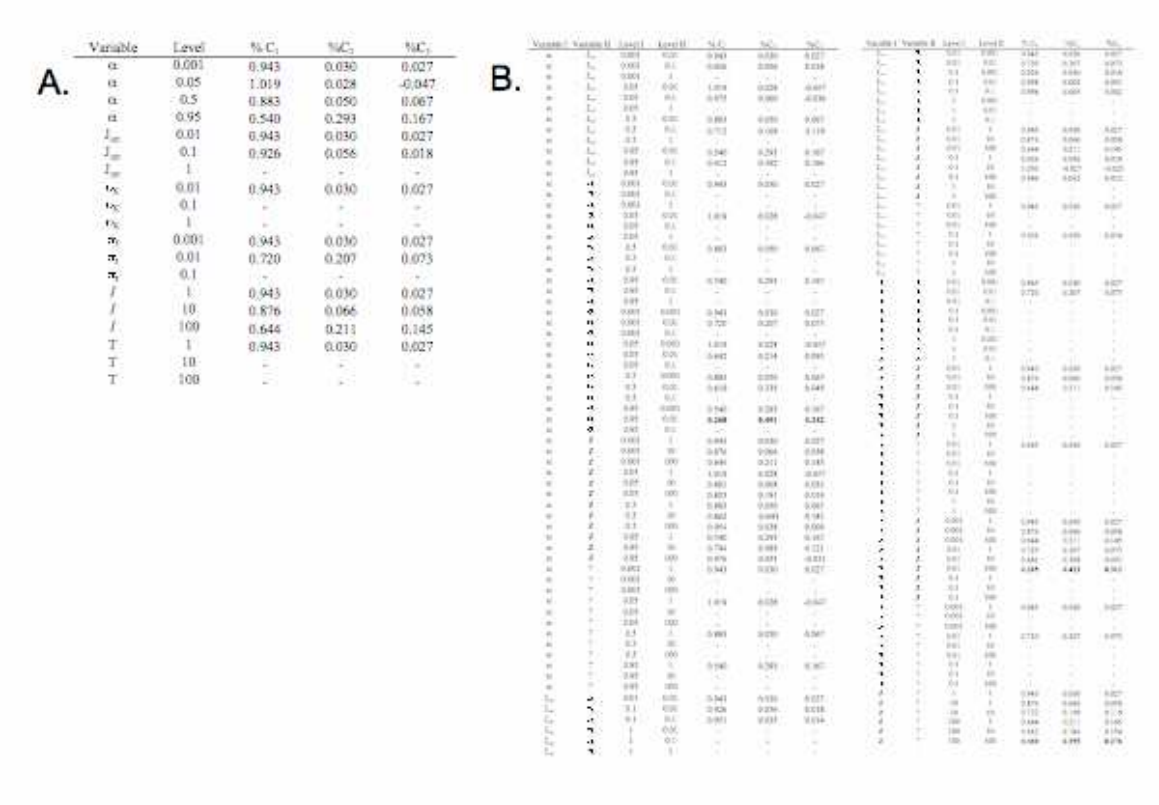


Table 14: Least Square Mean control coefficients for a. each variable level and b. two-way interaction between variables for three-step SK model.



A.

Variable	Level	$\%C_1$	$\%C_2$	$\%C_3$
α	0.001	0.943	0.030	0.027
α	0.05	1.019	0.028	-0.047
α	0.5	0.883	0.050	0.067
α	0.95	0.540	0.293	0.167
J_{x0}	0.01	0.943	0.030	0.027
J_{x0}	0.1	0.926	0.056	0.018
J_{x0}	1	-	-	-
v_x	0.01	0.943	0.030	0.027
v_x	0.1	-	-	-
v_x	1	-	-	-
π_0	0.001	0.943	0.030	0.027
π_0	0.01	0.720	0.207	0.073
π_0	0.1	-	-	-
J	1	0.943	0.030	0.027
J	10	0.876	0.066	0.058
J	100	0.644	0.211	0.145
T	1	0.943	0.030	0.027
T	10	-	-	-
T	100	-	-	-

B.

The three properties of mutations examined in this analysis, N_μ , Δk_i , and ΔW, were similar between the SK and MCT (Table 9). For reactions with low reversibility, during the directional selection phase, most substitutions occurred in the first enzyme and the fitness effects of these substitutions was larger for the first enzyme, despite the fact that substitutions in downstream enzymes have a larger effect on enzyme activity (Table 4b). During the stabilizing selection phase, the last enzyme in the pathway accumulated the greatest number of mutations, and a substantial portion of these tended

to have neutral or detrimental fitness effects ($\alpha_i = 0.001$: 2E pathway- 10.6% versus 55.0%; 3E pathway- 11.2% versus 49.6% & 60.8%) and were thus fixed by drift. These differences among enzymes were greatly reduced when reactions were largely reversible (Table 9b).

1.4 Discussion

1.4.1 Evolution of control coefficients

The evolution of metabolic control has been the subject of much speculation. A number of different arguments have been put forward: (1) metabolic control is likely to be shared roughly evenly across enzymes in a pathway (Kacser and Burns 1973); (2) control will shift between enzymes in a pathway and no one enzyme is expected to be more likely to have high control (Dykhuizen et al. 1987; Keightley 1989; Bost et al. 2001); and (3) control will be unequally shared and is likely to be highest for enzymes just below branch points because of fluctuating selection for allocation of flux along different branches (Eanes 1999).

Our simulations lend little support to any of these arguments. For example, except when reactions are largely reversible, control is very unevenly shared among enzymes. The unequal distribution is a result of selection for control to be centered in a single enzyme and is not simply a consequence of the summation property, as hypothesized by Bost *et al*. (2001). Our simulations find no evidence for constant shifts in

control between enzymes (Dykhuizen et al. 1987; Keightley 1989). This result hinges on the assumption of directional versus stabilizing selection; under directional selection to always increase flux, control will shift between enzymes, while we assumed stabilizing selection because a pathway's flux will eventually be inhibited at some rate. Under stabilizing selection, the strong tendency for the most upstream enzyme to gain predominant control tends to prevent frequent shifts in control among enzymes and means that most enzymes have a greatly reduced chance of exercising major control. Finally, we are catious about extending our conclusions concerning linear pathways, to separate branches of a branching pathway. We believe additional research focused on branching pathways needs to be conducted to fully address this question. Nevertheless, to the extent that this is possible, our results suggest that fluctuating selection on allocation between branches is not necessary for the evolution of major control by the most upstream enzyme in a pathway branch.

Our analysis does reveal, however, two properties of control that are expected to evolve, at least in linear pathways, when stabilizing selection acts on pathway flux and reactions have low reversibility: (1) metabolic control will be concentrated in one or a few enzymes, and (2) control tends to be located in upstream enzymes. These are general properties of the system which evolve under a large range of parameter values and regardless of whether enzymes are assumed to be saturated or unsaturated.

In the MCT model, these patterns result from two interacting properties. The first is that a very large portion of the (k_1, k_2, . . .) phenotypic space corresponds to greater control for upstream enzymes (e.g. Figure 1a). The second factor arises when comparing mutations of equivalent magnitude on enzyme kinetic properties in enzymes with small and large control coefficients. A beneficial mutation will more likely fix in the enzyme with greatest control, because it will have a greater effect on flux and hence on fitness. Thus, the initial biased distribution of control in upstream enzymes is maintained and intensified by this biased distribution of beneficial mutations.

These same patterns arise for different reasons under the SK model. In our examples, the regions of the line/surface of optimal flux in which the different enzymes exert dominant control are roughly equal in size. Random starting points thus have a roughly equal probability of corresponding to dominant control of a specific enzyme. Instead, a fitness penalty associated with large intermediate pools differentially penalizes downstream enzymes and prevents them from evolving to exert dominant control.

The concentration of intermediate metabolites may also contribute to favoring major control in upstream enzymes in the MCT model. Although this was not formally included in our simulations, in the Supplementary Material it is shown that the concentration of a pathway's intermediate is proportional to the k_i's of the reactions

preceding the intermediate and inversely proportional to the k_i's of the reactions following it. If there is a cost associated with maintaining pools of intermediates, and if that cost increases with the concentration of the intermediate, then there will no longer be an equilibrium *surface* given by Equation 1. Instead, there will be an equilibrium point corresponding to a low value of k_1 and high values of the remaining k_i , *i.e.* to a high CC for the most upstream reaction and a low CC for all others. Thus, the inequality of flux control due to selection against large intermediate pools in the SK model is likely to be seen in the MCT model as well.

1.4.2 Expected and observed bias in control

Our simulations indicate that reaction reversibility, α, has the greatest influence on the degree of bias among enzymes in flux control. As shown in Appendix A, the value of α for a particular reaction is directly related to the equilibrium constant for the reaction, *i.e.* $\alpha = 1/K_{eq}$. Because K_{eq} is a *thermodynamic* property of the reaction (it is determined by the difference in free energy between products and reactants (Nelson and Cox 2000)), α is not influenced by the *kinetic* properties of the enzyme associated with that reaction. Because, unlike kinetic properties, thermodynamic properties cannot evolve, the extent of bias in control is set *a priori* by intrinsic chemical and physical properties of the reactions.

Whether in general one would expect to see substantial bias in control thus depends on whether metabolic reactions tend to be reversible or irreversible. Equilibrium constants have been measured for many enzymatic reactions and they overwhelmingly tend to be large, indicating irreversibility. For example, the tryptophan biosynthetic pathway has equilibrium constants ranging from $7.6 \bullet 10^{8}$ to $9.2 \bullet 10^{12}$, corresponding to α values much less than the $\alpha = 0.001$ used in our simulations (Kishore *et al*. 1998). Two additional studies (Tewari *et al.* 2002a, b) of metabolic pathways found that the majority of K_{eq} measured are greater than 10^{8}, and only two reactions had smaller equilibrium constants (*e.g.* $K_{eq} = 1.7$ and 4.6, corresponding to $\alpha = 0.59$ and 0.22 respectively). Although these examples do not constitute an exhaustive survey, they indicate that for many biochemical pathways, if not most, reactions are largely irreversible, and therefore the bias toward high CC in upstream enzymes predicted by our model should often be found. Given that most estimated values of α are several orders of magnitude smaller than the smallest value (0.001) used in our simulations, we would expect this bias to be substantially stronger than we found (*e.g.* Table 2).

Evaluating agreement between the predictions of our model and patterns of flux control in real pathways is complicated by the fact that most pathways are not strictly linear. Instead, they are often branched or cyclical, and it is not clear whether and how predictions from our model carry over to such pathways. Nevertheless, a literature

survey of linear and nearly linear metabolic pathways provides strong support for our first conclusion that control is distributed unequally in pathways and marginal support for the second claim that control should be vested in upstream enzymes. In thirteen control analysis studies, all but one (Wisniewski *et al.* 1995), demonstrated significant evidence for non-uniform distributions of control (Groen *et al.* 1986; Dykhuizen *et al.* 1987; Albe and Wright 1992; Hill *et al.* 1993; Kashiwaya *et al.* 1994; van der vlag *et al.* 1995; Roussel *et al.* 1998; Thomas and Fell 1998; Bost *et al.* 2001; Cronwright *et al.* 2002; Pritchard and Kell 2002; Wu *et al.* 2004). These studies are sampled across a broad spectrum of organisms and metabolic pathways. Half of them focused on the glycolysis pathway, but other pathways included the tricarboxylic acid cycle, oxidative phosophorylation pathway, gluconeogenesis, lactose catabolism, and the succinate pathway.

Our second prediction, that upstream enzymes tend to evolve larger CC than downstream enzymes, is less well supported. Five studies provided strong support for this prediction: results indicated that the first enzyme had the highest control coefficient (Groen *et al.* 1986; Dykhuizen *et al.* 1987; Roussel *et al.* 1998; Cronwright *et al.* 2002; Pritchard and Kell 2002). Pritchard and Kell's (2002) analysis of the entire glycolysis pathway indicates that the two most upstream enzymes exhibit the majority of control in this pathway. By contrast, five investigations found that either an enzyme in the center

of the metabolic pathway (Albe and Wright 1992; Kashiwaya *et al.* 1994; van der vlag *et al.* 1995; Thomas and Fell 1998) or at the end of the pathway (Wisniewski *et al.* 1995) exerted the majority of control.

It should be noted that our analysis does not predict that downstream enzymes will never exhibit the greatest flux control. Even with low reaction reversibility (α = 0.001 or 0.05), approximately 5%-25% of our simulations evolved to equilibria at which the highest CC corresponded to an enzyme downstream of the first enzyme in the pathway. Moreover, our simulations of shifting optima revealed that pathways with control centered in the most upstream enzyme can evolve majority control in another enzyme, and that there is a low, but not insignificant, probability that at any given time, a downstream enzyme will exert majority control. It is thus not surprising that examples exist in which control is vested in downstream enzymes. The real question is whether this situation is as common as the reverse. Although the slight bias in these studies toward control being exerted in upstream enzymes is consistent with our expectation, this small sample clearly does not provide sufficient evidence to either support or refute the prediction that control should most commonly be vested in upstream enzymes.

1.4.3 Which genes participate in bouts of adaptive evolution?

The ultimate goal of our analysis was to determine whether predictions can be made about which genes in a pathway are likely to be involved in adaptive change. In doing so, we aim to extend the general theory of adaptation (Fisher 1930, Kimura 1983, Orr 2005) beyond its current focus on predicting the number and size distribution of mutations involved in bouts of adaptation. Initially, we suspected that a simple principle would govern the choice of genes involved in an adaptive walk: mutations in genes for which flux control is highest would be used preferentially because such mutations are likely to have greater effects on flux and hence on fitness (Hartl *et al.* 1985; Eanes 1999; Watt and Dean 2000).

Our analyses confirm the operation of this principle. In both the MCT and SK models, when reactions are largely irreversible and the distribution of control is the most inequitable between enzymes, most substitutions, as well as substitutions with the largest effect on fitness occur in upstream enzymes during adaptive walks toward an optimal flux. However, this pattern reverses, such that the majority of substitutions occur in downstream enzymes, as the population nears the optimum. Moreover, many substitutions fixed in this phase appear to be nearly neutral (Ohta 1973, Kimura 1983). This shift occurs because near the optimum most mutations are deleterious and because substitutions in downstream enzymes have smaller effects on fitness because of their

reduced control. Consequently, a mutation with a given effect on enzyme activity will have a smaller detrimental effect, and thus a greater probability of being fixed by drift, when it occurs in a downstream enzyme. Some adaptive substitutions continue to be fixed, however, as compensation for slightly deleterious substitutions, and these occur preferentially in the upstream enzymes. To summarize, the difference in control causes upstream enzymes to be subject to strong directional selection during the first phase of an adaptive walk and purifying selection during second.

The distribution of substitutions among genes within metabolic pathways has been investigated in three systems of which we are aware. The anthocyanin pathway is a linear pathway comprised of six enzymes that produce pigments involved in flower coloration across angiosperms. Three studies of molecular evolution in the anthocyanin pathway, looking at broad and narrow phylogenetic scales (between three angiosperm families and within the genus *Ipomea*), found that upstream enzymes had the lowest rates of substitution and the downstream enzymes had the greatest rates of substitution (Rausher *et al.* 1999, Lu and Rausher 2003; Rausher *et al.* 2008). These studies found no evidence of positive selection, suggesting that most substitutions were of neutral or slightly disadvantageous mutations. To the extent that these substitutions were disadvantageous, this pattern is consistent with expectations under our model: because of reduced flux control, mutations that cause deviation from the flux optimum are more

likely to have smaller effects on fitness in downstream enzymes, and are thus more likely to be fixed by genetic drift (Hedrick 2000). An independent study of four terpenoid biosynthesis pathways obtained similar results, a strong correlation between elevated rates of substitution and downstream enzyme position (Ramsay *et al.* 2009). This study finds that the elevated rates of substitution in downstream enzymes are, at least partly, caused by relaxed selection in the downstream enzymes, a pattern consistent with the expectations of our model. Tests for positive selection at a broad phylogentic scale (between angiosperm families) found significant effects for multiple genes distributed between up- and downstream positions of the pathways. Ramsay *et al.* (2009) conclude that decreased pleiotropy in downstream genes resulted in relaxed selection on these enzymes; however, our model provides another explanation: the control of downstream enzymes is greatly reduced compared to upstream enzymes and they accumulate many neutral or nearly-neutral mutations.

An additional study by Flowers *et al.* (2007) examined rates of molecular evolution in *Drosophila* for 17 metabolic enzymes in five pathways (glycolytic, gluconeogenic, glycogenic, trehalose, and pentose shunt) that intersect at glucose-6-phosphate enzyme. This study found a strong signature of adaptive evolution on the *D. melanogaster* and *D. simulans* lineages. The three enzymes that showed statistically significant elevated rates of adaptive evolution are all upstream enzymes occurring at

the branch points between pathways. Flowers *et al.* (2007) interpret this pattern as arising from fluctuating selection on the relative magnitudes of flux on different branches. However, this pattern may also be consistent with our models' prediction that adaptive substitutions will be concentrated in the most upstream enzyme of a terminal linear pathway, if the evolution of a pathway branch behaves similarly to the evolution of the linear pathways described here. This issue will only be resolved, however, by examining the evolution of control in branched pathways.

1.4.4 Conclusion

We present population genetic models of the evolution of metabolic pathway flux in order to investigate whether predictions can be made about whether different genes in a pathway will be differentially involved in the process of adaptation. Although the models pertain only to linear pathways, they indicate that there are likely to be strong differences among enzymes in the numbers and types of substitutions they accumulate. While the extent to which the specific predictions of our model may be extended to pathways with more complex topology is unclear, the fact that our model generates strong patterns of bias suggests that similarly strong, though perhaps different, patterns may be expected in more complicated pathways. We suggest that the approaches presented here will be useful for examining these more complex situations.

2 The Genetic Control of Copper Tolerance in *M. guttatus*

2.1 Introduction

The relationship between adaptation and speciation has been a thoroughly contested topic in evolutionary biology. There is wide support for the hypothesis that adaptation to novel environments drives reproductive isolation through the development of prezygotic barriers such has mating preference, location and timing (Mayr 1942; Simpson 1953; Shemske and Bradshaw 1999; Schluter 2000; Nosil et al. 2002). The model of intrinsic post-zygotic isolation put forth independently by Bateson (1909), Dobzhansky (1937) and Muller (1942) is agnostic toward the evolutionary forces, drift or natural selection, that generate reproductive isolation. This model states that Dobzhansky-Muller incompatibilities (DMIs) may arise when unique mutations from two different gene pools are brought together and interact epistatically to reduce fitness. The beauty of this model is that new substitutions can be neutral or beneficial in their native background, and only when they are exposed to a foreign genome, are they deleterious. Populations do not have to cross a fitness valley during the process of speciation. Dobzhansky (1951) hypothesized that intrinsic post-zygotic barriers would likely be a by-product of natural selection because the genotype is an integrated set of co-adapted genes shaped by a specific ecological niche and hybridization would 'produce discordant gene patterns' or intrinsic post-zygotic barriers. Initial research on

the genetic basis of intrinsic post-zygotic isolating barriers supported this hypothesis; three of the four "speciation genes" first identified had strong signals of positive selection (Coyne and Orr 2004). However, as more cases of intrinsic post-zygotic isolation have been genetically dissected, greater evidence has accumulated that alternative evolutionary forces such as: mutation pressure, drift, and intragenomic conflict, may drive the fixation of DMI (Barbash et al. 2003; Brideau et al. 2006; Harrison and Burton 2006; Masley et al. 2006; Mihola et al. 2008; Bikard et al. 2009; Ferree and Barbash 2009). A possible counter example to this thesis is the adaptation to a copper mine by *Mimulus guttatus* driving the fixation of a post-zygotic incompatibility locus (Coyne and Orr 2004; Futuyma 2005). In this chapter we aim to further investigate this hypothesis, by investigating the genetic basis of copper tolerance and hybrid lethality and the evolutionary forces that shaped these two phenotypes.

Adaptation to heavy metal enriched environments by plants is a classic example of microevolution, because, much like antibiotics and pesticides, heavy metal enriched soils impose strong selection for tolerance and prevent most species from colonizing these environments (Futuyma 2005). *M. guttatus* is a common wildflower inhabiting many different environments in western North America, including serpentine soils enriched for heavy metals such as zinc, and copper mines. Copper tolerant *M. guttatus* were first described inhabiting copper mine tailings of four mine sites within Calaverous

County, CA, the largest of these populations was found at the Keystone mine at Copperopolis (Allan and Sheppard 1971). Macnair (1981; 1983) has investigated the genetic basis of copper tolerance with multiple crossing experiments and established that this phenotype has a simple genetic basis, controlled by a single dominant locus. Repeated backcrosses of tolerant to nontolerant plants found that tolerance, scored as ability to grow roots above a threshold of copper, segregated in a 50:50 ratio over successive generations (Macnair 1983; Strange and Macnair 1991). However, these experiments cannot exclude the possibility that tolerance is controlled by multiple, tightly linked loci or that there is an inversion segregating between the parental lines. When tolerance is measured as a quantitative character, as opposed to a threshold character, the tolerance locus does not explain all of the genetic variation in tolerance (Macnair et al. 1987; Smith and Macnair 1998). We propose to measure the amount of quantitative genetic variation explained by this major tolerance locus.

A by-product of this investigation into the genetic basis of the tolerance locus experiments is that crosses with one population, Cerig-y-drudion (Cerig), produced hybrid lethal offspring, while crosses to another population, Stinson Beach, were completely viable (Macnair and Christie 1983). When hybrid lethality was observed, it always cosegregated with copper tolerance, leading Macnair and Christie (1983) to hypothesize that the tolerance and hybrid lethality were controlled by the same locus.

However, these experiments cannot exclude the possibility that tolerance and hybrid lethality are controlled by multiple, tightly linked loci or that there is an inversion segregating between the parental lines. An investigation into the distribution of the incompatibility alleles found that the DMI cosegregating with tolerance is fixed in the Copperopolis population (Macnair and Christie 1983; Christie and Macnair 1987). The fixation of the Copperopolis DMI allele in this population could be due to drift or to selection on this locus in the Copperopolis population, because the tolerance and DMI are controlled by the same locus, or there are two tightly linked loci and the DMI was swept to fixation, or there are two loci, and each one is beneficial in the mine environment. Hybrid lethality in the Cerig population was variable; the results from crossing experiments indicated that at least two polymorphic DMI loci are segregating within this population (Macnair and Christie 1983; Christie and Macnair 1987). The segregating polymorphisms in the Cerig DMI alleles indicate that they may be evolving neutrally or subject to balancing selection. An understanding of the physiological and genetic basis of the DMI loci will enable us to distinguish between these possibilities.

Hybrid lethality in crosses between Copperopolis and Cerig manifests as F1 hybrid necrosis, seedling leaves become yellow and die very early in development (Macnair and Christie 1983). Hybrid necrosis has been noted by breeders for a hundred years, but has only recently been appreciated as a common postzygotic isolating barrier

within plant species (Bomblies and Weigel 2007). The physiological phenotype of tissue necrosis is similar across many angiosperm clades, implying they all may share a common genetic basis (Bomblies 2009). For systems in which the genetic basis of this trait has been investigated, all studies have implicated the disease resistance genes, R genes, and their interacting proteins (Kruger et al. 2002; Bomblies et al. 2007; Alcazar et al. 2009; Jeuken et al. 2009). It is hypothesized that tissue necrosis is a result of the plant hypersensitive response; the R genes from one background are negatively interacting with proteins from the other genome and triggering this response. The tissue necrosis phenotype is very similar to the plant hypersensitive response, in which cell are locally aborted to prevent the spread of obligate biotrophic pathogens. These results suggest that tolerance and hybrid incompatibility are unlikely to be controlled by the same locus.

In this study, we propose to investigate the genetic basis of tolerance to determine: 1) is copper tolerance controlled by a single locus?, 2) how much phenotypic variation does this locus explain?, 3) are tolerance and hybrid incompatibility controlled by the same locus?, and 4) do R genes underlie the DMI loci in crosses between Copperopolis and Cerig plants?

2.2 Methods

M. guttatus is a common wild flower distributed throughout western North American. *M. guttatus* is the most common member of a species complex composed of

many interfertile subspecies varying in habitat preference, mating system, many life history traits (Vickery 1959). Adaptation to extreme edaphic environments is very common in this group, multiple populations and subspecies are locally adapted to serpentine soils and copper mine tailings (Macnair and Gardner 1998). This group has been subject to intense ecological research for over 50 years and has recently developed into a model system for ecological and evolutionary functional genomics (Wu et al. 2008).

2.2.1 Mapping Copper Tolerance Locus

To map the major copper tolerance locus we will use an introgression line developed by Macnair with the tolerance allele backcrossed into a nontolerant genome (Macnair 1983; Strange and Macnair 1991). In keeping with precedence (Macnair 1983), we refer to this as a major effect locus and acknowledge that the effect of this locus on quantitative variation in tolerance has not been measured. Macnair created this line by crossing plants from three different populations (Figure 6). The first cross was between two outbred plants, a tolerant plant (COP10) from the central smelter in Copperopolis, CA and the nontolerant plant (CER34) from Cerig, Wales, UK (Allan and Sheppard 1971). *M. guttatus* is native to North America, but has been found as a garden escapee in the UK for over a hundred years (Allan and Sheppard 1971). A single tolerant F1 plant was backcrossed to a second outbred, nontolerant plant from Stinson Beach (SB), Marin

County, CA (Macnair 1983; Strange and Macnair 1991). Macnair repeated backcrossing a single tolerant line to different outbred plants from SB for six generations (Figure 6). Each backcross generation segregates 50:50 plants that are heterozygous for the tolerance allele (BC_{7T}) and homozygous for the nontolerant allele (BC_{7NT}). We generated our mapping population by intercrossing a line BC_{7T} and BC_{7NT} line to produce 5014 $F1_{BC7}$ plants.

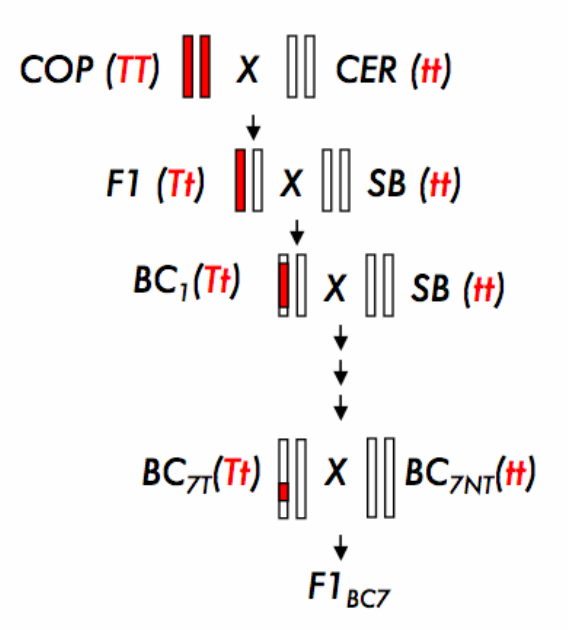


Figure 6: Generation of Major Tolerance Locus Mapping Lines

We scored copper tolerance in our mapping population as a threshold character, root growth in solution with elevated copper concentration (Macnair 1983). We planted $F1_{BC7}$ seeds into 12" x 18" trays filled with potting soil and stratified them for one week

at 4C. They were then transferred to greenhouse at University of Exeter, United Kingdom, and grown in under 8 hour light, 16 hour dark. Upon germination plants were thinned to 32 plants per tray. Plants were maintained in short day conditions to encourage vegetative growth and prevent flowering. When natural day length exceeded 8 hours, black out screens were used to prevent additional light. After eight weeks of growth, cuttings were taken, stripped of roots, and placed into solution of 0.5 g ml^{-1} $CaNO_3$ and 0.5 µg ml^{-1} $CuSO_4$. Cuttings were grown in solution for five days and scored as tolerant if they initiated new root growth or nontolerant if they were unable to initiate new root growth. Plants were grown in seven successive blocks of 768 plants from 2005 - 2007. We repeated tolerance assays on all plants with ambiguous root growth results.

Our approach in mapping the tolerance locus was to first identify the heterozygous genomic regions in the tolerant introgression line BC_{7T}, and second to demonstrate that a heterozygous region associates with the tolerance phenotype in the $F1_{BC7}$ mapping population. To genotype these lines we collected leaf and bud tissue into 96 well plates and shipped the tissue on dry ice to Duke University, Durham, NC, USA. We extracted genomic DNA from bud and leaf tissue using a CTAB/chloroform protocol (Doyle and Doyle 1990) modified for use in 96-well format (Fishman et al. 2005). We used a previously developed set of *MgSTS* markers that amplify intron length polymorphisms (described in Wu et al. 2008). We used these markers to identify

heterozygous genotypes in the BC_{7T} line and homozygous genotypes in BC_{7NT}. We screened 541 markers in BC_{7T} and BC_{7NT} lines and confirmed any putatively linked markers by screening them in 72 $F1_{BC7}$ plants and testing for a correlation with tolerance and marker segregation. For fine mapping, we designed additional intron length polymorphism markers using the M. guttatus 7X genome assembly produced by the Department of Energy's Joint Genome Institute (www.phytozome.net). Marker size was analyzed by capillary electrophoresis and fragment analysis on an ABI 3730x1 DNA Analyzer. Markers were scored manually in GENEMARKER (SoftGenetics, 2005, State College, PA).

2.2.2 Quantitative Copper Tolerance Analysis

To estimate the effect of the major tolerance locus, we needed to measure copper tolerance as a quantitative trait. Many quantitative phenotyping methods have been proposed, such as: measuring root length or biomass after growing in solution at a single heavy metal concentration (Smith and Macnair 1998), measuring of accumulation of heavy metals in root/shoot tissue (Salt et al. 2008), and the sequential method which measures the concentration of heavy metals that inhibits new root growth (Schat et al. 1996). In this method, plants are grown in hydroponic solution and exposed to a series of sequentially increasing heavy metal concentrations. Prior to each treatment, roots are stained black with activated charcoal and after an interval of a few days, new root

growth is scored by the presence of white root tips. Tolerance is scored as the heavy metal concentration at which new root growth ceases (Schat et al. 1996). This method is very laborious, but it is preferred to measuring root length or biomass, because of the possible confounding effects of the rate of plant growth. Measuring the accumulation of heavy metal ions is very precise, but it is very expensive and not feasible for a large number of samples. Thus, we sought to modify the sequential method for measuring tolerance in a large QTL mapping population.

For our quantitative mapping experiment, we tested a large mapping population using the sequential method. Instead of scoring new growth by staining roots black with activated charcoal, we grew plants in straws suspended in hydroponic solution and monitored root growth with rubber bands placed around the straws. To implement this new phenotyping method, we constructed 21 watertight boxes (6" X 12" X 15") made of PVC foam (Piedmont Plastics, Morrisville, NC) with wells drilled through the tops of the boxes. We used fishing line to attach clear drinking straws, length: 7", diameter: ¾" (Dispozo Plastics, Columbia, SC) to the wells in the lid, so the straws were extending down into the boxes. We punched between 20-30 small holes in each straw to facilitate the movement of solution. This design allowed plants to grow roots from the top of the well and into the straws suspended in hydroponic solution. The straws enabled us to track the root growth of each individual plant without them becoming intertwined.

Boxes were filled with ¼ strength Hoglands solution (Hewitt 1966), set at pH 5.7; this was our baseline hydroponic solution for all experiments. Copper was added to this baseline in the form of $CuSO_4$. The solution was constantly mixed and supplemented with oxygen using aquarium air pumps.

To phenotype copper tolerance, plants were grown in the hydroponic boxes and monitored for root growth in solution of increasing copper concentration. Plants were grown from seed on an inert hydroponic medium, Rockwool. We filled 0.5 mL microcenterfuge tubes with Rockwool and planted the seeds into the substrate. After germination, seedlings were thinned to one per tube, the bottom of the microcenterfuge tube was cut off, and placed into a well in one of the boxes. Seedlings were grown in the baseline solution for a minimum of two weeks before initiating copper treatment. Prior to initiation of the first treatment, we marked the root length of all plants. To mark root length, we placed rubber bands on the outside of each straw and positioned them to be level with the bottom of the plants longest root. Rubber bands were stiff enough to prevent accidental shifting in between census dates, but were easily rolled up and down the straw to track growing roots on census dates. At the initiation of the first treatment, we replaced the solution in all boxes with the baseline, supplemented with 1 $\mu g\ ml^{-1}$ Cu. Plants were censused for increased root growth after three days; we recorded which plants had extended their roots, moved the rubber bands to mark the new maximum

root length, and replaced the baseline solution with the next treatment level. If root length exceeded the length of the straws, we used rulers to measure root length at each census date. We tested the plants at seven treatment levels: 1, 2, 3, 4, 6, 8, 10 μg ml^{-1} Cu. Our measurement of tolerance was the concentration of copper at which root growth ceased.

We used this quantitative phenotyping method to measure copper tolerance in a large F2 mapping population. We used two inbred lines, from mine and off-mine habitats respectively, to generate our mapping population. The tolerant line, COP52, was collected from Copperopolis, CA, at the same site as tolerant line in the first mapping experiment, COP10. The nontolerant plant, MED84, was collected at a site free of copper contamination near Moccasin, Tuolumne County, CA, located 20 km from Copperopolis. We selfed parental lines for 4 generations and reciprocally crossed them to generate two F1 lines, differing only in their maternal backgrounds. We selfed each F1 plant to develop a mapping population segregating autosomal and organelle genomes. In total we measured copper tolerance of 1387 plants: 32, COP52; 98, MED84; 95, F1; 1162, F2. We stratified seeds for 1 week at 4C and transferred them to a growth chamber, set at 20C, 8 hr light and 16 hr dark. Plants were completely randomized among 21 blocks containing 72 or 144 wells. Due to a large environmental variability in this experiment we subsequently repeated this experiment.

In order to reduce the environmental variability and increase the accuracy of our measurements, we repeated this experiment with cuttings from a subset of plants from the first experiment. The second experiment was initiated six weeks after the completion of the first experiment, in which the plants were allowed to grow in the baseline solution containing no excess copper. We tested 356 plants in the second experiment: 29 MED84, 31 COP52, 44 F1s, and 252 F2s from 96 unique genotypes. All roots we stripped from cuttings and they were immediately placed in first copper treatment. In this experiment, we increased the level of copper plants were tested against, the treatment levels: 2, 4, 6, 9, 12, 15, 18 μg ml^{-1} Cu. As in the first experiment, plants were censused every three days for new root growth and then transferred to the next treatment level. Our measurement of tolerance was the concentration of copper at which root growth ceased.

We analyzed the phenotypic variation in copper tolerance, and estimated the heritability of tolerance and the effect size of the tolerance locus. We calculated the mean and variance of tolerance for each genotypic class (COP parent, MED parent, F1, F2) and tested the phenotypic distributions for normality (Shapiro-Wilks *W*-test). To statistically control for block effects in each experiment we fitted a single factor ANOVA. We used the residuals of this model to calculate the mean tolerance for each parental, F1 and unique F2 genotype. We calculated $2a$ as the difference between the mean for each parental class. Dominance, d, was calculated as the difference between the F1 mean

phenotype and the midpoint between the two parents (Lynch and Walsh 1998). We estimated the environmental variance, V_E, using phenotypic variance from the three genetically homogenous classes: the F1 hybrids and the two parental lines. We calculated V_E as a weighted average of the parental and F1 phenotypic variances, V_E = [2 Var(F1) + Var(COP) + Var(MED)] / 4 (Lynch and Walsh 1998). We calculated the genotypic variance, V_G, by subtracting the F2 phenotypic variance from V_E and then estimated broadsense heritability for tolerance as $H^2 = V_G$ /Var(F2) (Lynch and Walsh 1998). To estimate the effect size of the tolerance locus, we first genotyped each F2 line for a marker linked to the tolerance locus and then compared the difference between the two homozygous genotypic classes, $2a^{Marker}$, to the difference between the parental classes. We present these differences in phenotypic variation scaled by the environmental standard deviation (ESD), which is the square of the environmental variance V_E. All statistically analyses were conducted in JMP V. 7.0 (SAS Institute, Cary, NC).

2.2.3 Fine Mapping Copperopolis Incompatibility Locus

In order to determine if the copper tolerance and hybrid incompatibility phenotypes are controlled by the same locus, we crossed an incompatible Cerig line, CER10, to 19 $F1_{BC7}$ plants (10 nontolerant and 9 tolerant) identified as having a recombination breakpoint near the tolerance locus and scored their progeny for necrosis.

As a control, we also crossed CER10 to 8 nonrecombinant *F1*$_{BC7}$ plants (6 nontolerant and 2 tolerant). We measured necrosis as the percentage of progeny with yellow or dead leaf tissue. Plants were scored three weeks after germination, any ambiguous seedlings were not included in further analyses. We scored necrosis in 2-8 replicate grow-outs/genotype, two blocks/grow-out, 30 seedlings/block. All progeny were grown in the glass house at the University of Exeter. Plants were grown under ambient light and temperature conditions. Temperature was measured every day in the greenhouse. Seedlings were monitored every day for yellow and necrotic tissue. Recombinant and control lines were genotyped for multiple markers developed using the *M. guttatus* 7X genome (www.phytozome.net). Most of the markers amplified intron length polymorphisms and were scored as described above. Three markers, *Sc84.180kb*, *Sc84.234kb*, and *Sc84.297kb* were sequence-based markers used to identify SNPs. We direct sequenced genomic DNA from each line and ran these markers and identified SNPs as base pairs that consistently had two peaks in tolerant control and a single peak in the nontolerant control plants in Sequencer.

2.2.4 Mapping Cerig Incompatibility Locus(i)

To map the DMI(s) in the Cerig population interacting with the Copperopolis locus we created an F2 population segregating for the Cerig DMI(s), phenotyped them for hybrid lethality and genotyped them for markers near our candidate necrosis genes.

The two parents in our mapping population were chosen based on a preliminary set of crosses between 10 Cerig lines and a Copperopolis and scoring the hybrid progeny for necrosis. We found that CER10 had the highest offspring lethality, between 80%-100%, and CER35 had the lowest, between 0%-10% offspring lethality (Macnair and Christie 1983; Christie and Macnair 1987). We intercrossed these lines, and selfed a single F1 plant to generate an F2 mapping population of 168 plants. To measure hybrid lethality for each F2, we conducted two crosses to two distinct, outbred Copperopolis genotypes. A total a seven Copperopolis lines were used in the entire experiment, so not all F2s were crossed to the same Copperopolis genotype. This is unlikely to introduce much error into this experiment because it has previously established that the Copperopolis DMI is fixed in the mine population (Macnair and Christie 1983; Christie and Macnair 1987). We scored the progeny for necrosis as described previously. We scored necrosis in four successive growouts from May-August 2009.

Previous research has demonstrated that R genes are the root cause of many cases of hybrid necrosis (Bomblies and Weigel 2007; Alcazar et al. 2009). R genes are disease resistance loci that trigger plant defense responses after they identify signals of pathogen attack. These genes are very numerous, and they are often clustered together, in different genomic regions (Meyers et al. 2003). We identified the number and distribution of R genes in the *M. guttatus* genome. To identify R genes in the *M. guttatus*

genome, we conducted BLAST searches against the 7X genome assembly (phytozome.net). Our query amino acid sequences were three genes from each of the two major *A. thaliana* R gene sub-families, CC-NBS-LRR and TIR-NBS-LRR (Meyers et al. 2003). We identified all genomic scaffolds containing four or more R genes. We then identified the *MgSTS* markers located in those genomic scaffolds, or designed new markers for scaffolds lacking markers. Clearly, this approach will not identify all of the R genes in the M. guttatus genome, however it will provide a coarse estimate of the number of genes and their location. To map QTLs contributing to hybrid lethality we screen markers linked to our candidate genes in the F2 population.

To map the DMI in the Cerig population, we conducted a bulk-segregant QTL mapping experiment. We screened 73 markers linked to R gene clusters in 12 plants taken from the each end of the F2 phenotypic distribution. All markers with a skewed allele distribution between the two F2 pools were then screened in the total set of F2s. We statistically controlled for variation among growout date, block and Copperopolis line by fitting a three-factor ANOVA. We used the residuals of this model to calculate the mean and standard deviation for the marker genotypes. We assessed significant differences between marker genotypes with a single factor ANOVA.

2.3 Results

2.3.1 Tolerance Locus Mapping

In order to map the tolerance locus, we first needed to identify heterozygous genomic regions in the tolerant introgression line, BC_{7T}, and homozygous regions in the BC_{7NT} lines. To confirm the heterozygous region affects tolerance we will then test for an association with tolerance in the mapping population, $F1_{BC7}$. We screened 541 BC_{7T} and BC_{7NT} lines and identified 52 markers that were heterozygous in the BC_{7T} line. To confirm which markers are linked to the tolerance locus, we screened them in two panels of $F1_{BC7}$ lines and tested for a correlation between phenotype and allele segregation. We screened 52 markers in an initial panel of 8 $F1_{BC7}$ lines and found a correlation between with six markers. We then screened these markers in 72 $F1_{BC7}$ and winnowed the list of linked markers to four: *MgSTS217*, *MgSTS242*, *MgSTS536*, *MgSTS745*. Previous mapping experiments showed all of these are within 20 cM of each other on linkage group (LG) 9 (mimulusevolution.org). The tightest linked marker, *MgSTS242*, had no recombinants between tolerance phenotype, and all other markers are arrayed on the same side of the tolerance locus as e242. We name the copper tolerance locus *Tol1*.

We then proceeded to fine map the tolerance locus in a large growout of $F1_{BC7}$ lines. We scored 2229 $F1_{BC7}$ plants as tolerant, 2609 plants as nontolerant, 176 lines were removed from the analysis because we were unable to accurately score their phenotype.

Of the plants that were unambiguously phenotyped, we genotyped 4340 of these lines with marker *MgSTS242*. We mapped marker *MgSTS242* to be 1.57 cM from *Tol1*, there were 68 recombinants between this marker and the tolerance locus (Figure 7). Marker *MgSTS242* is located on the end scaffold 103, which is 972kb in length (Figure 7). We soon discovered that this region of the M. guttatus genome is filled with repetitive DNA and the assembly was quite poorly. There are no large scaffolds that span sizeable portion of the linkage group, LG 9. In order to proceed with fine mapping *Tol1*, we needed to identify additional genomic scaffolds in this region. We identified one other marker that mapped to this region of LG 9, *MgSTS481*. This marker is located in the middle of genomic scaffold 84, which is 1.01 MB in length. We hypothesized that this scaffold would contain the tolerance locus because previous estimates of the ratio of physical to genomic distances in M. guttatus range from 50-250 kb/cM (D.B. Lowry, unpublished data). We designed markers *Sc84.1788* and *Sc84.Atg5* located at opposite ends of scaffold 84 to determine if this scaffold contains the tolerance locus. We genotyped these new markers in 68 $F1_{BC7}$ recombinant lines and discovered that scaffold 84 did not include *Tol1* (Figure 7). This genomic region, (at least in this cross) has a greatly reduced recombination rate, the physical distance between *Sc84.1788* and *Sc84.Atg5* is 693 kb and the genetic distance is 0.51 cM, which corresponds to 1.36Mb/

cM. We measured the recombination rate between *Sc84.Atg5* and *Tol1* to be 0.74 cM, thus we estimate that the physical distance to the *Tol1* to be 1 MB (Figure 7).

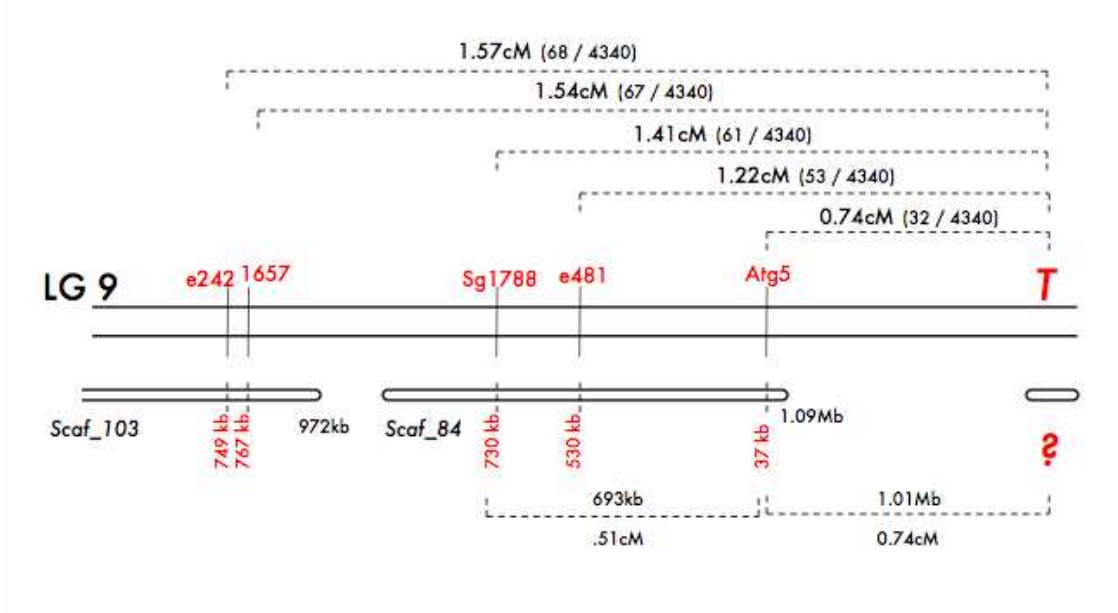


Figure 7: Fine Mapping of Major Copper Tolerance Locus

2.3.2 Quantitative Measurements of Copper Tolerance

To estimate the effect size of *Tol1* we conducted a sequential tolerance assay to quantitatively measure copper tolerance in a F2 population segregating copper tolerance. We measured the presence or absence of new root growth in 1293 plants at seven levels of copper, from 1 - 10 µg ml^{-1} Cu. The parental lines had overlapping distributions of copper tolerance and the distribution of tolerance in the F1 and F2 populations was significantly skewed toward higher levels tolerance (Figure 8). The

mean difference between the two parental lines, *2a*, is 3.25, the ratio of *2a*/*ESD* is 1.09, and the broad sense heritability is 0.14 (Table 15). Due to the overlap between the parental lines and the skewed distribution in the F1 and F2 lines, we repeated this experiment testing lines at higher levels of copper.

Table 15- Mean copper tolerance from 1st quantitative root growth assay

Source	n	Mean	SD	2a /ESD	H^2
CCC52	30	0.22	3.38	1.09	0.14
MED84	97	-3.03	3.36		
F1	96	1.61	2.55		
F2	1070	0.068	3.22		

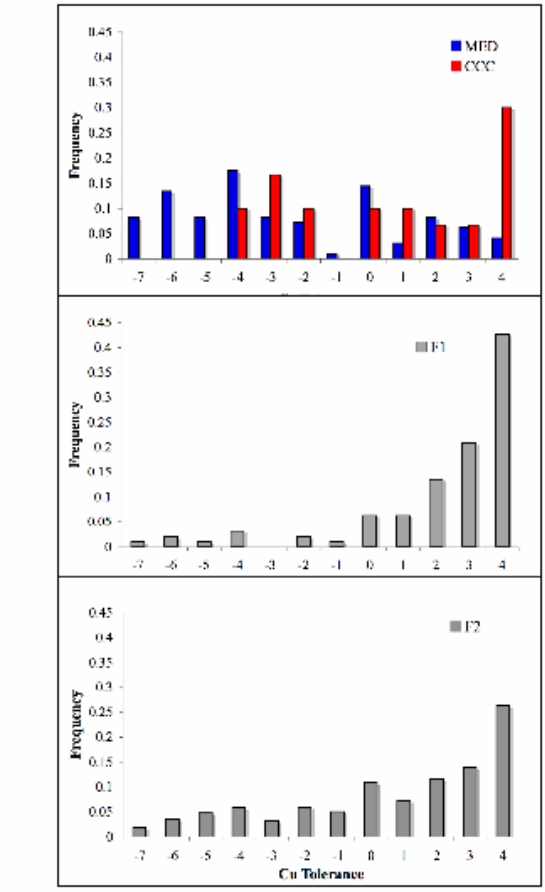


Figure 8: Distribution of copper tolerance from 1st root growth assay for parental (top), F1 (middle), and F2 (bottom) classes.

We conducted a second quantitative tolerance assay. In order to reduce the amount of environmental variability we took multiple cuttings from plants in the first experiment and to decrease the overlap between parental lines we tested lines at higher concentrations of copper. We took multiple cuttings from each F2 line (average 2.6 replicates / F2 line). Plants were tested for new root growth at seven treatment levels,

from 2 -18 μg ml^{-1} Cu. The overlap between parental lines was greatly reduced in the second experiment (Figure 9). Tolerance was normally distributed in the F1 (W = 0.954; p = 0.082) and F2 (W = 0.98; p=0.15) classes (Figure 9). The difference between the two parental lines was larger then previously measured, $2a$ = 7.55, as was the ratio of parental divergence to environmental error was also larger, $2a$/ESD = 2.15 (Table 16). We measured tolerance to be partially dominant, d is 2.59 and d/a is 0.69. We were unable to reduce the amount of environmental variation in this experiment, ESD = 2.98 in the first experiment and ESD = 3.51 in the second experiment. The broad sense heritability remained approximately the same, H^2 = 0.13.

Table 16: Mean copper tolerance from 2nd root growth assay.

Source	n	Mean	SD	2a /ESD	H^2
CCC52	29	3.11	3.07	2.15	0.13
MED84	27	-4.44	3.74		
F1	43	1.93	3.59		
F2	96	-0.088	3.76		

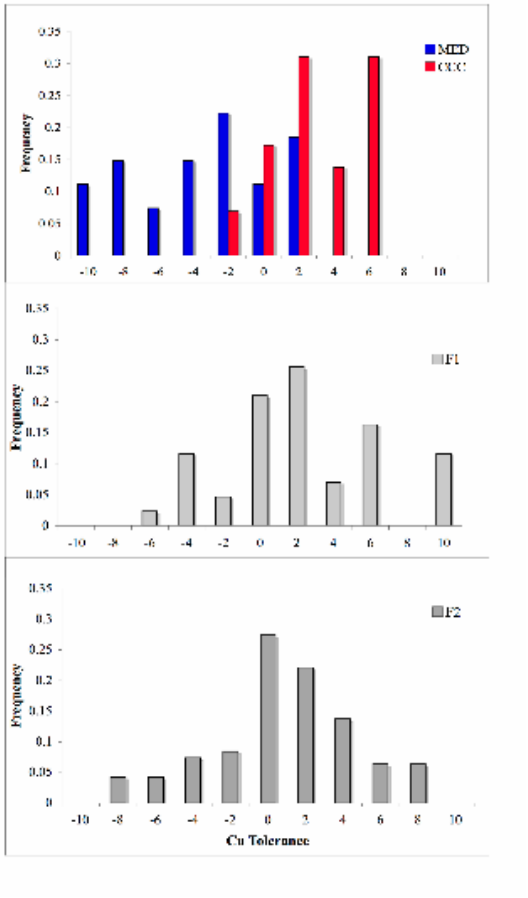


Figure 9: Distribution of copper tolerance from 2nd root growth assay for parental (top), F1 (middle), and F2 (bottom) classes.

To test for an effect of the tolerance locus on the quantitative tolerance phenotype, we screened one marker, *Sc103.1657*, in 96 F2s from the second experiment. This marker is 1.54cM from the *Tol1*, adjacent to *MgSTS242* at the end of scaffold 103 (Figure 7). We conducted a single-marker ANOVA and found a significant effect of *Sc103.1657* on tolerance (Table 17). The difference between the alternative homozygous alleles at this marker is $2a^{1657} = 2.34$ (Table 18). Comparing the ratio of the genotypic class

divergence to the environmental standard deviation, we find $2a^{1657}$/ESD = 0.67 , which accounts for 31.1% the parental divergence. The Copperopolis allele at marker 1657 is completely recessive, d = -1.18 and d/a = -1.01, this contrasts with our estimate of partial dominance of tolerance in F1 class.

Table 17: Single marker ANOVA testing for effect of genotype on copper tolerance.

	Sc103.1657					
Source	df	SS	MS	F	p	R^2
Model	2	94.4	47.2	3.64	0.030	0.077
Error	87	1126.7	13.0			
Total	89	1221.1	13.7			

Table 18: Copper tolerance least square mean (LSM) and standard error (SE) for alternate genotypes, Copperopolis (C) and Moccasin (M).

	Sc103.1657			
Genotype	n	LSM	SE	2a/ESD
CC	23	1.66	0.75	0.67
CM	42	-0.69	0.56	
MM	26	-0.68	0.72	

2.3.3 Fine Mapping Copperopolis Incompatibility Locus

In order to determine if the incompatibility and copper tolerance phenotypes are caused by the same locus, we crossed 19 recombinant $F1_{BC7}$ plants (9T, 10NT) to CER10, a line that yields high offspring lethality when paired with the tolerance locus. We found two lines, 25E01 and 25E11 where the tolerance and lethality phenotypes were disassociated (Figure 10). This indicates that the tolerance and lethality phenotypes are

controlled by different loci and these lines must have had a recombination event in between these two loci. We name the Copperopolis DMI locus, *Nec1*. To identify the location of the incompatibility locus we genotyped 19 lines for eight markers along scaffold 84. We found that 16 of 19 lines have recombination breakpoints that map the incompatibility locus to a region in-between markers at *Sc84.234kb* and *Sc84.297kb* on scaffold 84 (Figure 11). There are no annotated genes located within this interval, our markers are located in each of the two nearest genes. *Sc84.234kb* amplifies the 4th exon, 4th intron, and 5th exon and has high similarity to a Jumanji (JMJ5) transcription factor. The *Sc84.297kb* marker amplifies ~ 500bp of the 5th exon of a protein with high similarity to a glycosyltransferase protein. We found that the Copperopolis allele at this marker has a large indel in the 5′ end of the 5th exon.

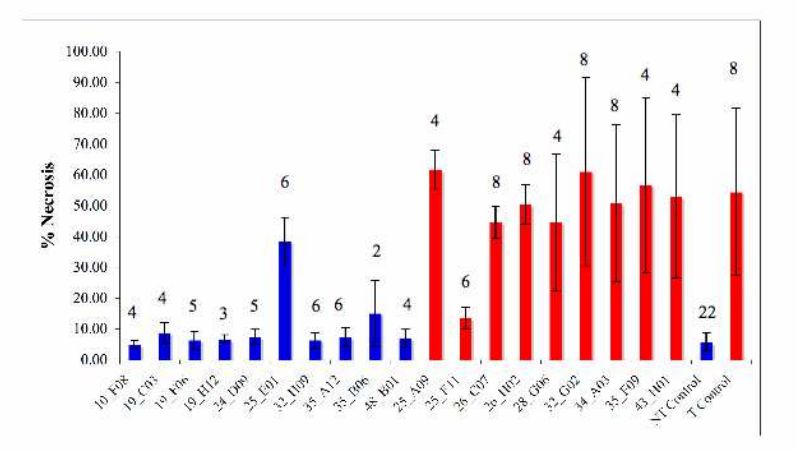


Figure 10: Mean percentage of necrotic offspring for recombinant F1$_{BC7}$ lines crossed to Cerig10 genotype. Red bars are tolerance and blue lines are nontolerant. Numbers above bars are number of replicate lethality assays.

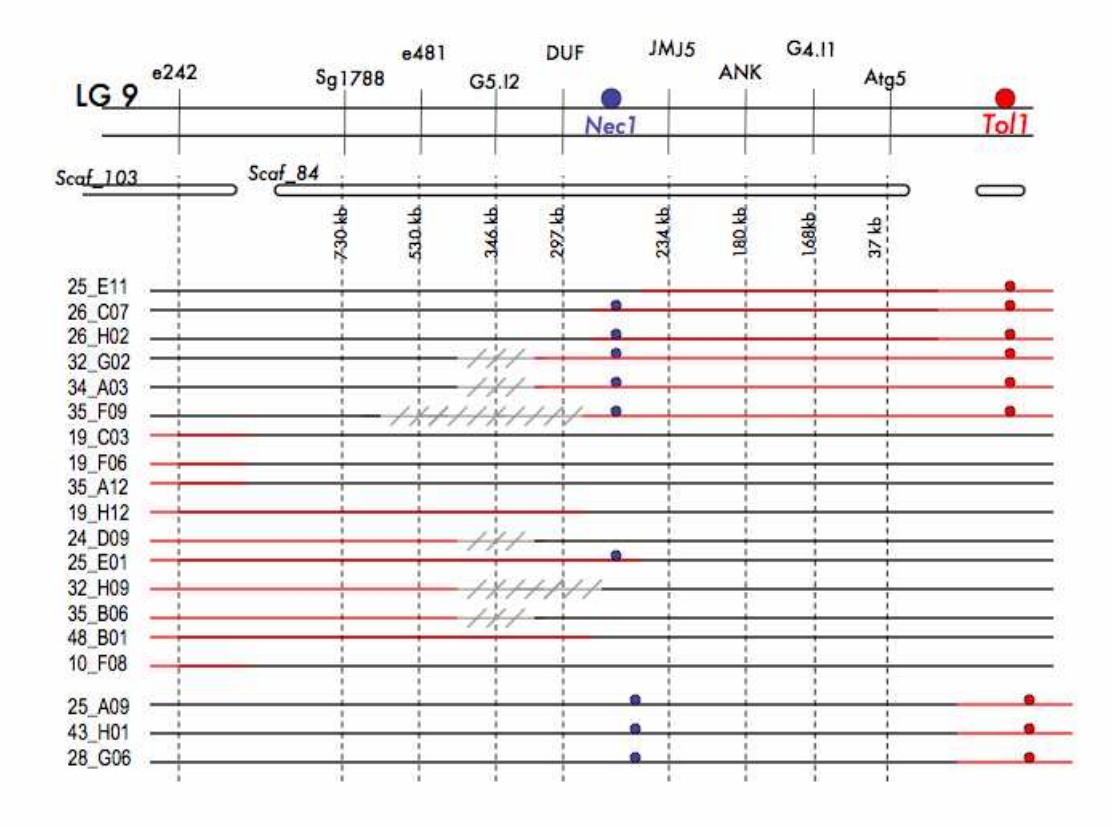


Figure 11: Fine mapping hybrid incompatibility locus, Nec1. Red lines denote COP allele, black lines denote SB (nontolerant) allele, and grey hatched lines are equivocal. Circles represent phenotypic measurements: tolerant genotypes have red circles and high offspring lethality genotypes have blue circles.

2.3.4 Mapping Cerig Incompatibility Locus(i)

To map QTLs underlying the Cerig DMI(s), we generated an F2 mapping populations, from two lines, CER10 and CER35, which produce different levels of lethal offspring when crossed to Copperopolis. We measured the level of offspring lethality by scoring the percentage of lethal offspring from crosses between 168 F2 plants and Copperopolis lines. We measured offspring lethality of 131 F2s when crossed against

two different Copperopolis genotypes, we measured 33 F2s for hybrid lethality when crossed to only a single Copperopolis genotype, and four F2 lines perished during the experiment before any crosses could be completed. We measured the parental mean hybrid offspring lethality as CER35 = 6.7% and CER10= 79.7%. The distribution of offspring lethality is not normal ($W = 0.94$, $p<0.001$), it is skewed toward lower levels of necrosis (Figure 12). The F2 distribution encompasses the parental mean on the low end, but only one plant approaches the parental mean at the high end (Figure 12).

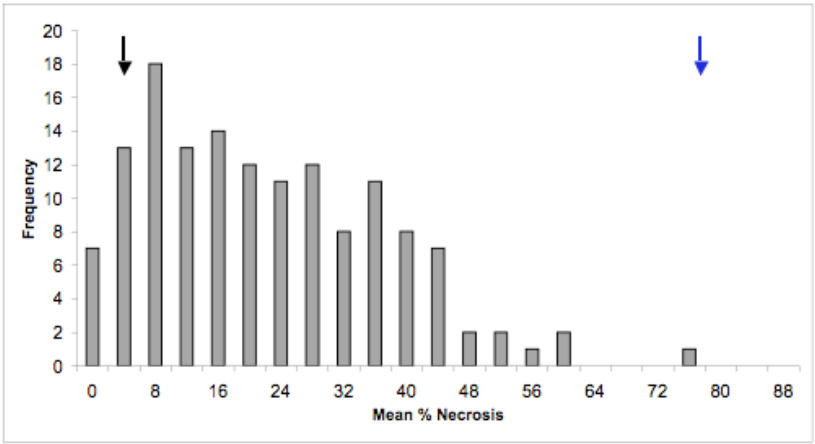


Figure 12: Distribution of necrosis in F2 population derived from CER10 and CER35. Black arrow is mean for CER35 parent, blue arrow is mean for CER10 parent.

To map the genetic basis of offspring lethality we first needed to identify the location of the R genes in the *M. guttatus* genome. We used BLASTp to query the *M. guttatus* genome for amino acid matches to *A. thaliana* R genes with e-value scores $<e^{-10}$. We identified 476 unique R gene hits distributed among 133 scaffolds. The distribution

of hits was highly skewed, the four scaffolds with the greatest number of hits contained 28% of the total number of R genes identified and 72 scaffolds had a single R gene hit. We identified 24 scaffolds with four or more unique R gene clusters.

To determine if any of these R gene clusters affect the level of hybrid necrosis, we genotyped the Cerig F2s for markers linked to the R gene clusters. We found 73 markers in the 24 scaffolds with four or more R genes. To identify allelic variation between the parental lines we screened them in the CER10 and CER35 parents, 25 of 73 markers were polymorphic. We screened the polymorphic markers in 24 F2s taken from the extremes of the distribution, 12 high plants, mean necrosis 49.5%, and 12 low plants, mean necrosis 0.3%. We found that 3 of the 25 markers screened showed skewed allelic distribution between the two F2 pools. We screened these three markers in 162 F2s and found that two markers were significantly associated with hybrid lethality (Table 5). Coincidently, these two markers, *MgSTS217* and *MgSTS416* are located on LG 9 on scaffold 68 near a cluster of seven R genes approximately 21 cM from the primary incompatibility locus (mimulusevolution.org). We name this locus *Nec2*. The two markers comprising *Nec2*, *MgSTS217* and *MgSTS416*, had R^2 values of 0.32 and 0.35 (Table 19). Both of these markers exhibit segregation distortion, with an excess of heterozygotes compared to either homozygote class (Table 20).

Table 19: Single marker ANOVA testing for association between markers and necrosis in Cerig F2s.

	MgSTS217						MgSTS416						MgSTS473					
Source	df	SS	MS	F	p	R^2	df	SS	MS	F	p	R^2	df	SS	MS	F	p	R^2
Model	2	0.86	0.43	37.2	<.0001	0.32	2	0.961	0.48	43.8	<.0001	0.35	2	0.04	0.02	1.23	0.295	0.016
Error	160	1.86	0.01				160	1.754	0.01				149	2.48	0.02			
Total	162	2.72					162	2.715					151	2.52				

Table 20: Hybrid necrosis least square mean (LSM) and standard error (SE) for alternate genotypes, Cerig10 (C10) and Cerig35 (C35).

	MgSTS 217			MgSTS416		
Genotype	n	LSM	SE	n	LSM	SE
C10C10	13	0.029	0.030	16	0.038	0.026
C10C35	92	0.055	0.011	89	0.058	0.011
C35C35	60	-0.099	0.014	60	-0.104	0.014

2.4 Discussion

In this study, we demonstrate that the copper tolerance locus, *Tol1*, maps to an unassembled region of the *M. guttatus* genome. We further demonstrate that *Tol1* explains a large amount of the quantitative variation in copper tolerance. We next show that the Copperopolis DMI locus does not map to *Tol1*, but is controlled by a separate locus, *Nec1*. Lastly, we demonstrate that *Nec1* interacts with at least one DMI in the Cerig genome, *Nec2*, to cause F1 hybrid lethality. *Nec2* maps near a cluster of seven R genes.

2.4.1 Genetic Basis of Copper Tolerance

Macnair (1983) originally proposed that tolerance was controlled by a single dominant locus of large effect. However, this experiment could not distinguish between a single major locus, multiple tightly linked loci, or an inversion containing a large

genomic region. We mapped a single locus, *Tol1*, to an unassembled region of the *M. guttatus* genome. There is no inversion in this genomic region, but we cannot eliminate the possibility that tolerance is controlled by multiple tightly linked loci. We will continue fine mapping this region as the assembly of the *M. guttatus* 7X genome improves. Experiments are currently underway to genetically map all of the unassembled genomic scaffolds by re-sequencing 64 recombinant inbred lines at low coverage to identify SNPs in every scaffold and map the scaffolds onto the existing linkage map. This procedure could place many more scaffolds near scaffold 84, and we can continue fine mapping *Tol1*. We would also like to know the phenotypic effect size of this locus to qualitatively estimate the strength of selection on this locus.

Macnair (1983) hypothesized that the copper tolerance locus has large phenotypic effects. However, this conclusion was based on the results from experiments that measured tolerance as a threshold character. When tolerance was measured as a quantitative trait, the tolerance locus did not explain all of the genetic variation in tolerance (Macnair et al. 1993; Smith and Macnair 1998). In these studies, tolerance was selected in two different directions, high and low, while holding the Copperopolis alleles at *Tol1* constant. There was a response to selection in both directions and they concluded that additional loci contribute to tolerance. This experimental design did not allow for clean estimation of the phenotypic effects of the tolerance locus. We conducted

an experiment to quantitatively measure the effect of *Tol1* in an F2 mapping populations, and we found that this locus accounts for 31% of the divergence between parental lines. Although there has been much debate over what constitutes a QTL of large effect (Fishman et al. 2002), this data indicates that this locus does explain a substantial amount of the quantitative variation in tolerance. Therefore, this indicates that this locus is under strong selection in the mine environment.

We had hoped to map additional tolerance QTLs in our quantitative mapping experiment, but we decided against conducting a genome-wide QTL analysis because of the high environmental variation in our quantitative root growth assays. We may be able to decrease environmental variation and increase our estimate of heritability from 0.14 by repeating our quantitative root assay, and measure the rate of root growth at each copper treatment. Macnair and colleagues conducted a selection experiment and estimated the realized heritability of copper tolerance to be 0.39 when tolerance was phenotyped as the rate of root growth in a single copper treatment level (Macnair et al. 1993).

2.4.2 Hybrid incompatibility, a pleiotropic consequence of selection for tolerance?

One of the few examples of habitat-mediated selection causing the fixation of an intrinsic post-zygotic isolation factor is the association of a DMI factor and copper tolerance locus in the Copperopolis population of *M. guttatus* (Coyne and Orr 2004,

Futuyma 2005). Post-zygotic isolation in this system manifests as hybrid necrosis, which has been linked to incompatibility between interacting plant pathogen response genes (Kruger et al. 2002; Jeuken et al. 2009; Bomblies 2009). Is hybrid lethality a pleiotropic effect of copper tolerance or is it caused by plant pathogen response genes?

Macnair and Christie (1983) proposed that strong selection in the mine environment resulted in the fixation of a copper tolerant allele that also caused post-zygotic reproductive isolation. They performed numerous crosses between mine and off-mine genotypes and found that copper tolerance and hybrid lethality always segregated together. Macnair and Christie did not have any genetic markers in this experiment, so they could not rule out the possibility that the phenotypes are controlled by two loci that are tightly linked or reside in an inversion. We found that copper tolerance and hybrid necrosis phenotype are controlled by two tightly linked loci: *Tol1* and *Nec1*. We reject the hypothesis that hybrid necrosis is a direct pleiotropic byproduct of an allele conferring copper tolerance. We find no evidence of inversions in this genomic region, although we will need to continue fine mapping *Tol1* to confirm this result. If tolerance and hybrid lethality are controlled by two separate loci, how did they both become fixed in the Copperopolis populations?

There are two hypotheses describing the fixation of both tolerance and hybrid offspring lethality phenotypes in the mine population: it was swept to fixation along

with the tolerance locus, or *Nec1* is beneficial in the mine environment. We conclude that it is unlikely there was selection on *Nec1* in the mine environment because the two loci nearest our mapped interval, a Jumanji transcription factor and a glycosyltransferase, seem unlikely candidate genes for conferring fitness in the mine environment. We have presented evidence that *Tol1* has a large effect on copper tolerance, so we expect that selection is very strong on this locus in the mine environment. Therefore, we conclude that the tight linkage between *Nec1* and *Tol1* and strong selection on *Tol1* caused *Nec1* to be swept to fixation in the Copperopolis population. We present additional data supporting this conclusion in the next chapter. This data seems to support a mechanism that DMI can become swept to fixation because of tight genetic linkage to another locus under direct selection.

2.4.3 Genetic basis of hybrid incompatibility

The final goal of this project was to test the hypothesis that hybrid lethality is controlled by genes involved in the plant pathogen response pathway. The offspring from crosses between Copperopolis and Cerig genotypes develop yellow leaves and necrotic tissue. Similar tissue necrosis traits have been observed in hybrid offspring from many plant systems, this indicates that necrosis has a shared underlying physiological mechanism (Bomblies 2009). Tissue necrosis is a result of induced cell death, this phenotype is characteristic of the hypersensitive response of plants used to limit the

spread of pathogens that require living tissue to proliferate. To test the hypothesis that hybrid necrosis is controlled by plant pathogen response genes, we genetically mapped DMI loci controlling hybrid inviability between Copperopolis and Cerig genotypes.

We mapped the Copperopolis DMI, *Nec1*, to an interval of 63kb, between markers *Sc84.234kb* and *Sc84.297kb* on scaffold 84 of the *M. guttatus* 7X genome assembly. Three of 19 recombinant plants do not support this locus mapping to this region (Figure 6). We believe that these genotypes must have been contaminated while being maintained as perennial plants for two-three years in the University of Exeter greenhouse. There are no annotated genes within the mapped interval and the nearest genes are the two markers that define the interval, neither of which have a perfect association with the phenotype. These two genes are in opposite orientation along the chromosome, such that the 5′ region of each gene extends into the mapped interval. Our markers are located in the 3′ end of each gene. We hypothesize that the functional change underlying *Nec1* is in an exon or a regulatory region at the 5′ end of one of these genes. Marker *Sc84.234kb* is a Jumanji transcription factor, these are involved in methylation and histone modification (Hong et al. 2009). Marker *Sc84.297kb* is a glycosyltransferase, these genes are involved in the widespread modification of plant secondary compounds and the regulation of hormones, including salicylic acid (Gachon et al. 2005). This gene is an interesting candidate because of the central role salicylic acid

plays in pathogen response. To determine the functional mutation underlying *Nec1*, we plan to continue fine mapping this interval and compare the expression of these two genes in green living tissue and yellow necrotic tissue in hybrid offspring. The ultimate confirmation that the glycosyltranferase underlies *Nec1* would be to transform the Copperopolis allele into a Cerig compatible genotype and cross the transformed line to Cerig see if we recover the lethality phenotype. To test our hypothesis that hybrid necrosis is caused by plant pathogen response genes, we also investigated the genetic basis of the DMI locus in the Cerig genome.

We hypothesized that the Cerig DMI locus, *Nec2*, would be an R gene. To map the Cerig DMI locus we identified all of the genomic scaffolds with four or more R genes and screened markers within those scaffolds in an F2 population segregating for the Cerig DMI. We identified a single locus effecting hybrid lethality, *Nec2*, which is defined by two markers tightly linked to a cluster of 7 R genes on scaffold 68. *Nec2* accounts for a large portion of the variation in necrosis: the single marker ANOVA for *MgSTS217* and *MgSTS416* have R^2 values of 0.32 and 0.35. However, these results also imply that other loci also affect hybrid lethality in the Cerig background. Our mapping approach was not comprehensive and additional R genes or other defense response genes may contribute to hybrid lethality. There is strong segregation distortion for both *Nec2* associated markers, the ratio of genotypic classes is 1:6:4, with the homozygote genotype for the

high allele, Cerig10, being underrepresented. This distortion could not be caused by hybrid lethality because these genotypes are of the Cerig F2 population and the lethality only manifests in Cerig x Copperopolis crosses. We are unsure of the cause of this distortion, but this phenomenon is often found in crosses between M. guttatus populations (Fishman et al. 2002; Hall and Willis 2005). Our finding that an R gene contributes to the Cerig DMI locus is congruent with previous research that the incompatibility factor is polymorphic in the Cerig population (Macnair and Christie 1983; Christie and Macnair 1987). R genes are highly polymorphic within populations because pathogen pressures produce strong balancing selection on these loci (Bakker et al. 2006). Conversely, we find that the Copperopolis DMI, *Nec1*, does not map to a known R gene, and the incompatibility allele is fixed within this population (Macnair and Christie 1983; Christie and Macnair 1987). Our results are consistent with previous findings that R genes are the genetic basis for hybrid necrosis.

Recent investigations into the genetic basis of hybrid necrosis indicate that plant pathogen response genes are the genetic basis for this trait (Kruger et al. 2002; Bomblies et al 2007; Alcazar et al. 2009; Jeuken et al. 2009). Tissue necrosis in tomato was identified to be associated with resistance to the mold fungus, *Cladosporium fulvum* (Langford 1948). Recent molecular characterization of wild and domesticated tomato lines determined that *Cf-2* is a NB-LRR gene that monitors conformation changes in a

cystenine protease gene, *Rcr3*, that is often targeted during pathogen attack (Kruger et al. 2002). Tissue necrosis arises when *Cf-2* from wild tomato interacts with *Rcr3* from the domestic tomato, but this reaction is prevented when *Cf-2* and *Rcr3* are both from the wild tomato. In *A. thaliana*, the two hybrid necrosis systems that have been genetically characterized found R genes involved in hybrid necrosis (Bomblies et al. 2007; Alcazar et al. 2009). Hybrid necrosis in interspecific crosses between cultivated and wild lettuce species, *Lactuca sativa* and *L. saligna* cosegregates with resistance to powdery mildew (Jeuken et al. 2009). Molecular characterization of this phenotype indicates that this locus is homologous to the *A. thaliana* gene, *Rin4*, which is a plasma membrane associated protein targeted by at least three pathogen effectors and is protected by two R genes (Jeuken et al. 2009). To summarize, these studies have found two genes, *Rin4* and *Rcr3*, that act as DMI loci when paired with R genes from another genomic background. Under normal conditions, these proteins are "guarded" by R genes and any change in their conformation triggers plant defense response (Bomblies 2009). The proposed model of hybrid necrosis in these cases is that R genes and the guardee proteins are co-evolving in each lineage and when different alleles come together through hybridization, they do not recognize each other and the hyper-sensitive response is triggered, killing the plant. Our candidate gene at the Copperopolis DMI locus, a glycosyltransferase protein, does not conform to this model because it directly interacts with salicylic acid, not R genes. If

the glycosyltransferase protein proves to be the functional gene underlying *Nec1*, then the physiological model of how the hypersensitive response is triggered in necrotic plants will have to be modified. How do these conclusions influence current paradigms about the relationship between adaptation and speciation?

2.4.4 Adaptation driving the fixation of DMI loci

Many researchers have argued about the role of adaptation in driving speciation of plant lineages. There are many examples of habitat-mediated selection driving prezygotic isolation through changes in mating behavior, flowering time, pollinator visitation, and habitat preferences (Mayr 1942; Simpson 1953; Shemske and Bradshaw 1999; Schluter 2000; Nosil et al. 2002). There is less consensus on the effect of adaptation on the evolution of intrinsic post-zygotic isolating barriers. Recent discoveries have linked the neutral processes, such as genomic rearrangements, and genomic conflict to driving intrinsic post-zygotic reproductive isolation (Presgraves 2010). Genome rearrangements have been shown to cause intrinsic post-zygotic isolation in *Drosophila* and *A. thaliana* because essential genes are moved into different parts of the genome and hybridization can produce F2s progeny that maybe missing essential loci (Masley et al. 2006; Bikard et al. 2009). Genomic conflict has been demonstrated to cause intrinsic post-zygotic isolation because of the break up of tightly co-evolved selfish elements and the repressors of those elements in each genomic background (Barbash et al. 2003; Brideau

et al. 2006; Mihola et al. 2008; Ferree and Barbash 2009). As discussed here, intrinsic post-zygotic incompatibilities have also been linked to co-evolution of interacting plant defense response genes due to pathogen pressures (Bomblies 2009).

We present data supporting two evolutionary forces contributing to the fixation of DMI loci within a single system. First, out evidence suggests that an R gene contributes to hybrid incompatibility in the Cerig background. R genes may be under strong selection because of their critical role in detecting plant pathogens. Previous studies have indicated that some R genes are under strong balancing selection and maintain ancient polymorphisms (Bakker et al. 2006). The DMI locus in the Copperopolis background may act in the plant pathogen response system through regulation of salicylic acid, but it is unlikely that the R gene and glycosyltransferase interact directly. Previous models of plant pathogen response genes causing hybrid necrosis state that incompatibilities arise when hybridization brings together divergent proteins that interact directly (Bomblies 2009). Our results do not fit this model, so it is not clear how pathogen-mediated selection pressures may drive the incompatibility of *Nec1* and *Nec2* in this system.

The second evolutionary force driving the fixation of a DMI locus in this system is mediated, indirectly, through adaptation to a novel environment. Our evidence suggests that *Nec1* was swept to fixation in the Copperopolis population due to tight

linkage between *Tol1* and *Nec1*. We found that *Tol1* has a large effect on copper tolerance and is likely under strong selection in this environment. We will continue to explore this strength of selection on *Tol1* in the Copperopolis population in the following chapter. For now, we offer a tentative conclusion that a DMI locus can become fixed between species or populations because of habitat mediated selection on another locus in tight linkage.

3 Parallel Evolution of Copper Tolerance in *M. guttatus*

3.1 Introduction

The parallel evolution of ecologically important phenotypes is a common observation in the natural world. Parallel evolution describes the evolution of the same phenotype in genetically independent populations inhabiting similar environments. This process has been observed in multiple systems, including: the reduction of body armor and pelvic spines in independent freshwater three-spine stickleback (*Gasteroteus aculeatus*) populations (Bell and Foster 1994), the repeated evolution of various *Anolis* ecomorphs on numerous Caribbean islands (Losos et al. 1998), the divergence between dwarf and normal morphs in multiple lake whitefish *Coregonus clupeaformis* (Pigeon et al. 1997), and the colonization of serpentine soils and other edaphically stressful habitats by multiple species of *Lashenia* (Rajakaruna et al. 2003). Evolutionary biologists have long studied parallel evolution because this is strong evidence for natural selection driving phenotypic change in natural populations (Simpson 1953, Endler 1986). A question arising from these studies is whether replicate phenotypic changes arise via similar physiological or genetic changes?

To examine this question researchers have investigated the physiological and genetic basis of traits undergoing parallel phenotypic evolution. Recently a great deal of

progress has been made in identifying the genetic basis of ecologically important phenotypes and determining if independent populations have substitutions in the same locus or, possibly, the same nucleotide. One of the best examples of this involves the evolution of reduced body armor plating in freshwater populations of threespine sticklebacks. Researchers demonstrated that independent freshwater populations are more closely related to adjacent marine populations, yet all but one populations has evolved reduced body armor by the same genetic mechanism, a mutation in the *Eda* gene (Colisomo et al. 2005). This mutation segregates at low frequency in the marine population, indicating that ancestral variation was important in the evolution of reduced body armor. However, the researchers did find a single freshwater population that evolved reduced body armor via changes in a different gene, supporting the notion that there is more than one way to evolve a given phenotype. Additional research on the genetic basis of traits evolving in parallel will enable researchers to determine the relative contribution of shared ancestral polymorphism or independent novel mutations in driving phenotypic change.

The colonization of extreme edaphic environments by plant species is an ideal system to study parallel evolution. Extreme edaphic environments, such as serpentine (ultramafic) soils, high salinity soils, and mine tailings all create strong selective environment for organisms living on them (Antonivics et al 1971; Flowers et al 1986;

Macnair and Gardner 1998; Brady et al. 2003). Metal ore mine tailings often present multiple stress factors, including low water holding capacity, cation exchange capacity, pH (acidic), and macronutrient concentration, as well as high concentrations of bioavailable heavy metals (Tordoff et al. 2000). These environments are distributed as terrestrial islands, they are often restricted on a local scale and replicated on a regional scale, enabling comparative investigation of physiological mechanisms of tolerance. In response, organisms often display specific morphological or physiological adaptations to cope with edaphic stress. The physiological and genetic basis of heavy metal tolerance has been studied in many plants inhabiting heavy metal-rich environments (Baker 1987; Shaw 1989). Physiological heavy metal tolerance mechanisms may include exclusion, sequestration/detoxification, and toxicity tolerance, all of which are typically coupled with translocation regulation at the plant organ level (Baker 1987; Shaw 1989; Hall 2002; Marschner 2002). Different physiological mechanisms for tolerance suggest parallel evolutionary trajectories may not always proceed via the same physiological changes. The molecular mechanisms underlying these physiological process has been well studied in *Arabidopsis* (reviewed in Rauser 1999; Cobbett 2000; Hall and Williams 2003; Burkhead et al. 2009). Natural variation between mine and off mine populations of multiple species has also been investigated (Macnair 1983; Schat et al. 1996; Turner et al.

2010). We aim to study the evolution of copper tolerance in the common monkey flower, *Mimulus guttatus*.

M. guttatus has become a model plant species in the study of ecological divergence (Wu et al. 2008). *M. guttatus*, and its many sibling species comprise the morphologically and physiologically diverse *M. guttatus* species complex. *M. guttatus* is an herbaceous, rosetted annual (seasonally dry environments) or rhizomatous perennial (perennially moist environments). *M. guttatus* inhabits a wide diversity of habitats throughout western North America including coastal, riparian, alpine, thermal hot springs, serpentine, and mine tailings (Wu et al. 2008). The colonization of copper mine tailings by *M. guttatus* has been investigated for many years.

Investigations into the evolution of heavy metal tolerance in M. guttatus have focused on populations in California's Sierra Nevada foothills copper belt, primarily a population inhabiting the large North Union-Keystone Union mine complex at Copperopolis (Allan and Shepard 1971, Macnair 1983, Macnair and Christie 1983; Macnair et al. 1993; Tilstone and Macnair 1997). Common garden experiments demonstrate a genetic basis for copper tolerance in mine populations (Allan and Shepard 1971; Macnair 1981; Macnair et al. 1993). An investigation into the distribution of copper tolerant plants found that tolerance is fixed in six mine populations located in the foothills copper belt near Copperopolis, and is segregating < 5% in off mine

populations, despite these populations growing in close proximity (Christie and Macnair 1987). To investigate the genetic basis of tolerance in the Copperopolis population, Macnair (1983) analyzed the segregation of tolerance in a series of crosses and found that tolerance was primarily controlled by a single, dominant locus of large effect. The physiological effect of this locus was investigated using an introgression line with the Copperopolis tolerance allele backcrossed into a nontolerant genomic background and measuring the concentration of copper in root and shoot tissue and cell membrane integrity after exposure to copper ions (Strange and Macnair 1991; Tilstone and Macnair 1997). These experiments found that the Copperopolis tolerance allele introgression line had lower concentration of copper in its root tissue and greater cell membrane integrity then nontolerant off mine plants (Strange and Macnair 1991; Tilstone and Macnair 1997). In collaboration with Macnair we have fine mapped this large effect tolerance locus, *Tol1*, to a small region of linkage group 9 (Chapter 2). In this chapter, we aim to expand our investigation into the evolution of tolerance by continuing to examine the physiological basis of copper tolerance and patterns of selection on *Tol1* at Copperopolis and compare these finding to other mine populations of *M. guttatus*.

There are three major deposits of copper in California. The aforementioned Sierra Nevada foothills belt is the largest, extending from Madera county in the south to

Nevada county in the north, the Plumas copper belt is restricted to Plumas county on the eastern side of the Sierra Nevada mountains, and the Shasta belt occupies portions of Shasta, Trinity, and Siskiyou counties at the northern end of the state (Jenkins 1948). *M. guttatus* has been found on eight copper mines within the foothills belt: Copperopolis, Little Quail Hill, McNulty, Napoleon, Nassau, Newton, Penn, and Star-Excelsior (Allan and Shepard 1971, Macnair et al. 1987). We have recently discovered new populations inhabiting the Iron Mountain mine/Keswick smelter and Bully Hill mines in the Shasta belt and the Walker mine in the Plumas belt. The mines and smelters of the Sierra Nevada foothills and Shasta Cu belt began operations around 1860, while mines of the Plumas Cu belt did not begin operations until around 1900 (Jenkins 1948). Mining and smelting activity at most of the mines ceased around 1945. Non-metallicious populations are quite abundant throughout all three of these regions. The distribution of habitat islands and variation in tolerance begs the question: did copper tolerant alleles evolve once and then migrate to other mines or did tolerance evolve independently at each of these mines? If tolerance did evolve independently, did they use the same physiological and genetic mechanisms to achieve tolerance?

We will investigate the parallel evolution of copper tolerance in *M. guttatus* by comparing the physiological mechanisms of tolerance and selection on the major tolerance locus between Copperopolis and three other mine populations from the

foothills belt, and a single mine population from the Shasta and Plumas belts. We will address a number of specific questions: 1) do other mines have soil copper concentrations similar to Copperopolis?, 2) do other mine populations have reduced concentration of copper in their root tissue similar to Copperopolis?, 3) are mine populations more share more neutral genetic variation with each other or with adjacent off mine neighbors?, and finally, 4) we will determine if there is evidence of selection on the tolerance locus by measuring the extent of genetic differentiation, F_{st}, between the Copperopolis and adjacent off-mine populations at markers linked and unlinked to *Tol1.* We will compare the estimates of Fst at Copperopolis to other paired mine and off-mine populations.

3.2 Methods

3.2.1 Soil Collection and Analyses

In order to compare the abiotic edaphic habitat between mine and off-mine populations, we measured the pH and the concentration of four heavy metals, Cu, Zn, Fe, and Mn at six paired mine and off mine sites, two mine populations were sampled at the McNulty population (Table 21). We did not collect soil at two unpaired off mine sites, CVR_{OM1} and CVR_{OM2}, but these populations are located in different water drainage systems more than 4 km from the nearest mine. Soil samples were collected at 0 – 15 cm depth from where each population was growing, this depth represents the rooting zone

of *M. guttatus*. Soil samples were air dried and sieved to < 2mm. For most soil samples, pH (saturated paste extract; Rhoades and Miyamoto 1990) and bioavailability of heavy metals (DTPA extract; Lindsay and Norvell 1978) was analyzed by A&L Agricultural Labs, Inc. (Modesto, California, USA; analysis suite S3C). Soil from two populations (MCN_{M2} and MCN_{OM}; Table 21) was analyzed at the Texas A&M soil analysis lab using the same extraction method.

Table 21: Description of population habitat, location, pH and concentration of four heavy metals.

							DTPA extractable (Bioavailable)			
Population Location	Population Abbreviation	Habitat	Latitude	Longitude	Elevation (m)	pH	Cu^{2+} (ug g^{-1})	Zn^{2+} (ug g^{-1})	Fe^{2+} (ug g^{-1})	Mn^{2+} (ug g^{-1})
Keswick	KES_M	Mine	40.631627	-122.463645	231	5.9	116	48	8	2
	KES_{OM}	Off	40.617606	-122.482405	288	5.8	4	3	69	27
Walker	WKR_M	Mine	39.965041	-120.665661	1869	6.1	295	1	2	5
	WKR_{OM}	Off	39.964167	-120.660556	1903	6	4	28	66	10
Nassau	NAS_M	Mine	38.044588	-120.648879	506	4.7	56	49	29	5
	NAS_{OM}	Off	38.046111	-120.646667	493	7.2	5	4	52	15
Penn	PEN_M	Mine	38.232421	-120.877191	86	4.6	156	34	1	9
	PEN_{OM}	Off	38.235278	-120.871389	119	5.9	15	18	66	23
Copperopolis	COP_M	Mine	37.977145	-120.637969	293	6	199	2	3	3
	COP_{OM}	Off	37.962067	-120.614329	238	6.4	11	2	84	44
McNulty	MCN_M	Mine	37.912222	-120.723889	252	4.4	252	50	37	34
	MCN_{M2}	Mine	37.934575	-120.732822	260	5.9	64	39	41	42
	MCN_{OM}	Off	37.840102	-120.626463	164	7.3	2	4	18	26
Calaverous Cty	CVR_{OM1}	Off	37.929255	-120.639902	260	-	-	-	-	-
	CVR_{OM2}	Off	37.959805	-120.689355	282	-	-	-	-	-

3.2.2 Copper Concentration in Root and Shoot Tissue

In order to compare the physiological mechanisms of tolerance between Copperopolis and the other mine and off-mine populations we grew plants in three different copper treatments and analyzed the concentration of copper ([Cu]) in root and shoot tissue. We investigated paired mine and off-mine populations from four foothill

mines: Copperopolis, McNulty, Nassau and Penn, as well as paired populations from the Plumas belt, Walker mine, and Shasta belt, Keswick mine (Figure 13, Table 21). For this analysis we did not include a paired off mine population for the McNulty mine. Seed was collected by pooling seed pods from 50 maternal families. We germinated all seeds on potting soil (UC mix). Seven days after germination, prior to the emergence of true leaves, seedlings were randomly selected and transplanted into plastic conetainers (164 ml; SC10, Stuewe and Sons, Inc., Corvallis, Oregon, USA) containing sand culture. Plants were watered twice weekly with full strength Hoaglands solution containing 0.5, 2.0, or 4.0 μg ml^{-1} $CuSO_4$. Macnair (1981) previously established 2.0 μg ml^{-1} Cu as the toxicity threshold for off mine M. guttatus populations. A total of 10 replicates per population per [Cu] treatment were established. Plants were housed outdoors in Hollister, California, USA (ambient climate). Plants were harvested 45 days post-transplant when all individuals were at flowering stage. At harvest, roots were separated from shoots, cleaned, and oven dried at 60°C. Root and shoot samples were then dry ashed in a muffle furnace at 500°C, dissolved in 1N HCl, and analyzed by flame atomic absorption spectrophotometry (Atomic Absorption Spectrometer AAnalyst 800, PerkinElmer Instruments, Waltham, Massachusetts, USA). We analyzed the measurements of [Cu] in root and shoot tissue with an ANOVA or t-test in JMP v.7.01

(SAS Institute Inc. 2007). All statistical comparisons for the McNulty population were made to the nearest off mine population at Copperopolis, 10 km.

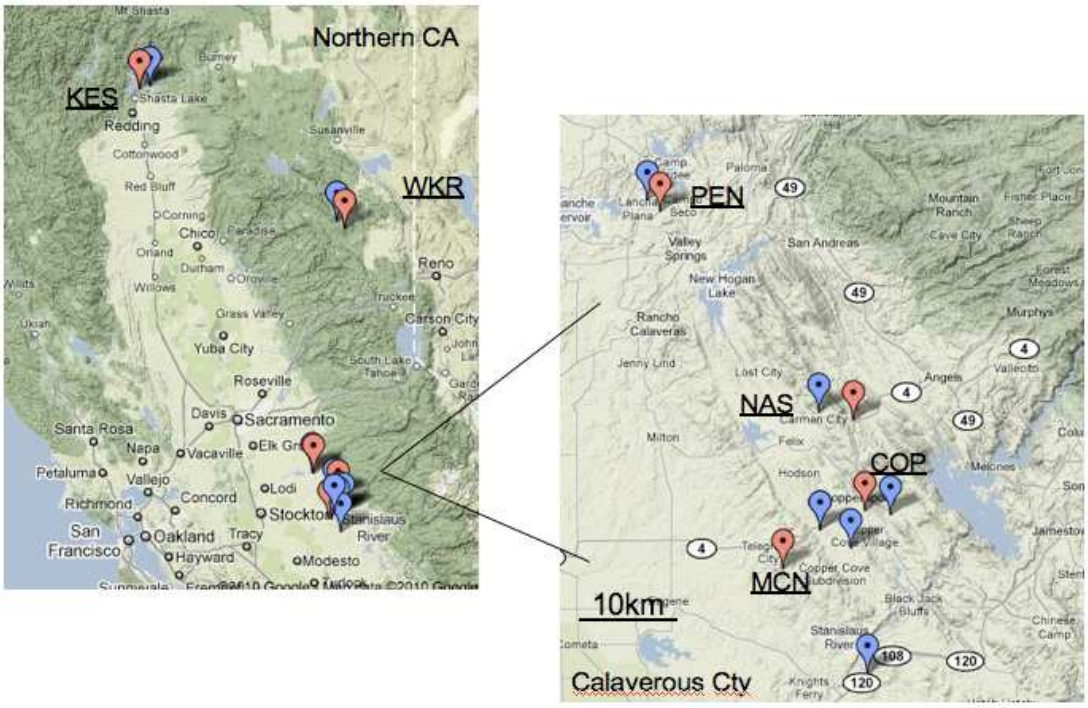


Figure 13: Map of northern California, red markers indicate mines and blue markers are off-mine populations.

3.2.3 Population Structure Plant Material

To measure the genetic differentiation between mine and off mine populations we collected genomic DNA from 15 populations of *M. guttatus*, we sampled all 11 populations included in the physiological testing, an additional population from contaminated soil near the McNulty mine and three additional off mine populations from Calaverous county (Figure 13, Table 21). We sampled 9 - 30 (mean 17) field

collected individuals per population, for a total of 239 individuals, 125 mine and 114 off-mine. Seeds used in population structure analysis were collected as individual maternal families, with the exception of PEN_M and PEN_{OM}. Samples from the Penn mine were collected as 50 pooled maternal families, below we discuss how we accounted for possible sampling of siblings in these populations. Seeds were planted into 2.5in pots with Fafard 4P potting soil and stratified at 4C for seven days. Seeded pots were transferred to a greenhouse (Duke University, Durham, NC), watered daily, and grown under 18 hr day lengths at 20° C. Thirty days after germination, tissue was collected and genomic DNA was extracted from bud and leaf tissue using a CTAB/chloroform protocol (Doyle and Doyle 1990) modified for use in 96-well format (Fishman et al. 2005).

We genotyped a total of 16 molecular markers in our 15 *M. guttatus* populations. Eleven markers are unlinked and dispersed across the genome (mimulusevolution.org), this represents our "neutral" marker dataset, but we acknowledge that we have not formally established that these are evolving neutrally. The five remaining markers are all tightly linked to a QTL for copper tolerance, *Tol1* (Chapter 2). Hereafter, we will refer to the five *Tol1* linked markers as $Tol1_{LINK}$, and the 11 *Tol1* unlinked markers as $Tol1_{UL}$. $Tol1_{LINK}$ are located at the end of linkage group 9, on scaffold 84 and scaffold 103 (M. guttatus genome, 7X build, Department of Energy, Joint Genome Institute,

www.phytozome.net). All markers are on the same side of the tolerance locus, we previously estimated the genetic and physical distances from the tolerance QTL as: *Sc84.Atg5*: 0.74cM/1.01MB, *Sc84.G4*: 0.86cM/1.18MB, *Sc84.G5*: 1.04 cM/1.41MB, *Sc84.G9*: 1.22cM/1.67MB and *Sc103.1657*: 1.54 cM/2.11MB (Chapter 2). These physical distances are an estimate based on the recombination rate between two markers at opposite ends of scaffold 84 (Chapter 2). The markers are all size fragment polymorphisms. All markers, save one, were designed with primers in exons flanking introns containing microsatellites or other indel polymorphisms (as described in Wu et al. 2008). Marker aat217 amplifies a microsatellite located in an intergenic region (Hall and Willis 2005). Marker size was analyzed by capillary electrophoresis and fragment analysis on an ABI 3730x1 DNA Analyzer. Markers were scored manually in GENEMARKER (SoftGenetics, 2005, State Collage, PA).

To determine if plants from the Penn populations could have originated from the same maternal parent, we analyzed the number of shared alleles between all plants within each population. Two plants that were found to share at least one allele at each of the 16 markers were considered to be potential siblings and they were removed from all analyses.

3.2.4 Population Structure Analysis

To determine if copper tolerance evolved in parallel, we first need to establish if mine populations are more related to each other or to their neighboring off mine populations. To address this problem, we first conducted an AMOVA to measure how genetic variation is distributed between habitats, populations, and individuals (Weir and Cockerham 1984). We also conducted a MANTEL test to determine if there was a correlation between the genetic distance matrix and geographic distance matrix (Sokal & Rolf 1994). Genetic distance was calculated as pairwise F_{st}. Pairwise geographic distance was calculated using the Geographic Distance Matrix Generator (http://biodiversityinformatics.amnh.org). The AMOVA and MANTEL tests were implemented in Arlequin v.3.1 (Excoffier et al. 2005).

Our second method for measuring the genetic independence of mine populations was to analyze the extent of shared genetic variation between populations using the program STRUCTURE (Pritchard et al. 2000). Briefly, STRUCTURE uses a model-based clustering algorithm to group individuals into a predefined number of populations, *k*, by maximizing the Hardy-Weinburg equilibrium of marker alleles among individuals within the population. To determine the number of significantly differentiated populations within our data set we ran two independent simulations for values of each k = 2-13, and performed likelihood ratio tests to determine the value of k which best

explains the data. Each simulation was run for 100,000 iterations, with a 50,000 burn-in, assuming a uniform amount of admixture within individuals. We only used the *Tol1*$_{UL}$ dataset in all analyses because we wanted to estimate the amount of shared genetic variation at a neutral set of loci and because STRUCTURE assumes that all genetic markers are unlinked. We acknowledge that we have not formally tested if these markers are evolving neutrally, but we have an expectation that they will be a better proxy for neutral evolution then the *Tol1*$_{LINK}$ dataset.

3.2.5 Selection on Tolerance Locus in Natural Populations

We will investigate if the major tolerance locus identified in the Copperopolis population affects tolerance in other mine populations. We will estimate selection on *Tol1* in the Copperopolis population and compare these findings to the other mine populations. We assume that evidence for selection on *Tol1*in Copperopolis and other mine populations indicates that these mines have a shared genetic basis for copper tolerance. If selection is strong, the loci under selection are expected to exhibit reduced genetic variation and extreme levels of differentiation between divergently selected populations (Beaumont and Nichols 1996; Charlesworth et al. 1998; Beaumont and Balding 2004; Novembre and Di Rienzo 2009). As a general description of allelic diversity for our two marker sets in each population, we calculated the average number of alleles (N_A), the observed heterozygosity (H_o), expected heterozygosity (H_e) for the

Tol1$_{UL}$ and *Tol1*$_{LINK}$ markers in all populations using the program Arlequin v.3.1 (Excoffier et al. 2005). We used the same data sets to estimate the level of genetic differentiation between pair mine and off-mine populations.

To estimate the level of genetic differentiation we measured F_{st} for every marker between paired mine and off-mine populations. If selection is acting on the tolerance locus, we predict that the *Tol1*$_{LINK}$ markers will show greater F_{st} then *Tol1*$_{UL}$ markers. We compared the estimates of F_{st} at Copperopolis to the other paired mine and off-mine populations. We predict that if Copperopolis and the five other mines share a genetic basis for tolerance, then they will also share patterns of selection at *Tol1*. We measured F_{st} and H_e between paired mine and off mine populations for each marker using *Fdist* (Beaumont and Nichols 1996). Note, this measure of H_e includes all individuals from the paired populations, it is important to measure H_e because the observed level of genetic differentiation between populations will depend on the number of alleles segregating at a particular marker. Paired mine and off-mine populations were analyzed separately to control for the effect of population structure. Each population was used in only a single comparison to avoid multiple testing. For this analysis, we combined the two McNulty mine populations, MCN_{M1} and MCN_{M2}. Included in these data are two off mine populations (CVR_{OM1}, CVR_{OM2}) that are located in between the Copperopolis and McNulty mines and not paired with either mine population (Figure 1) We used

information from our STRUCTURE analysis to determine whether to pair each population with either the Copperopolis or McNulty mines.

In contrast to many genome scan studies that attempt to identify Fst outliers from a large set of markers, we are testing the specific hypothesis that the Fst values for the *Tol1*$_{LINK}$ are significantly elevated compared to the *Tol1*$_{UL}$ dataset. This was accomplished by comparing observed levels of F_{st} for the *Tol1*$_{LINK}$ to expected distribution of F_{st} under neutral evolution. To generate the expected distribution of F_{st} under neutral evolution, we used the *Tol1*$_{UL}$ dataset to generate 90,000 coalescence simulations in LOSITAN, a GUI for *Fdist* (Antao et al. 2008). *Tol1*$_{LINK}$ were not included in coalescence simulations because these simulations assume markers are unlinked (Beaumont and Nichols 1996). We activated two recommended features of the program, 'neutral marker simulations' and 'force mean Fst'.

3.3 Results

3.3.1 Soil Charcteristics

The soil at Copperopolis was very high in copper, it has 199 μg g^{-1} Cu and was comparable to all other mine sites, which range from 56 – 295 μg g^{-1} Cu (Table 21). The Cu concentrations at all off-mine sites were less than 15 μg g^{-1} (Table 21). Soil pH and Zn, Fe, and Mn concentrations were variable between mine and off-mine sites with no clear trends.

3.3.2 Copper Concentration in Root and Shoot Tissue

We measured the concentration of copper in root and shoot tissue for six mine and five off-mine populations under three experimental treatments: 0.5, 2.0 and 4.0 μg g^{1} CuSO4. Root tissue [Cu] for four mine populations COP_M, MCN_M, PEN_M, and WKR_M was significantly lower than their paired off mine population plants at all three Cu treatment levels (Figure 14). Shoot tissue [Cu] were only significantly lower than off mine populations at the 4.0 μg g^{-1} Cu treatment (Figure 14). Hereafter, we describe the physiological mechanism of copper tolerance in these four populations as copper "avoidance". Two mine populations, NAS_M and KES_M, had root and shoot [Cu] equal to the paired off mine populations (Figure 14). These two populations will be referred to as "non-avoidance" physiological mechanism. There was no heterogeneity in the Cu concentration among off-mine populations (Figure 14).

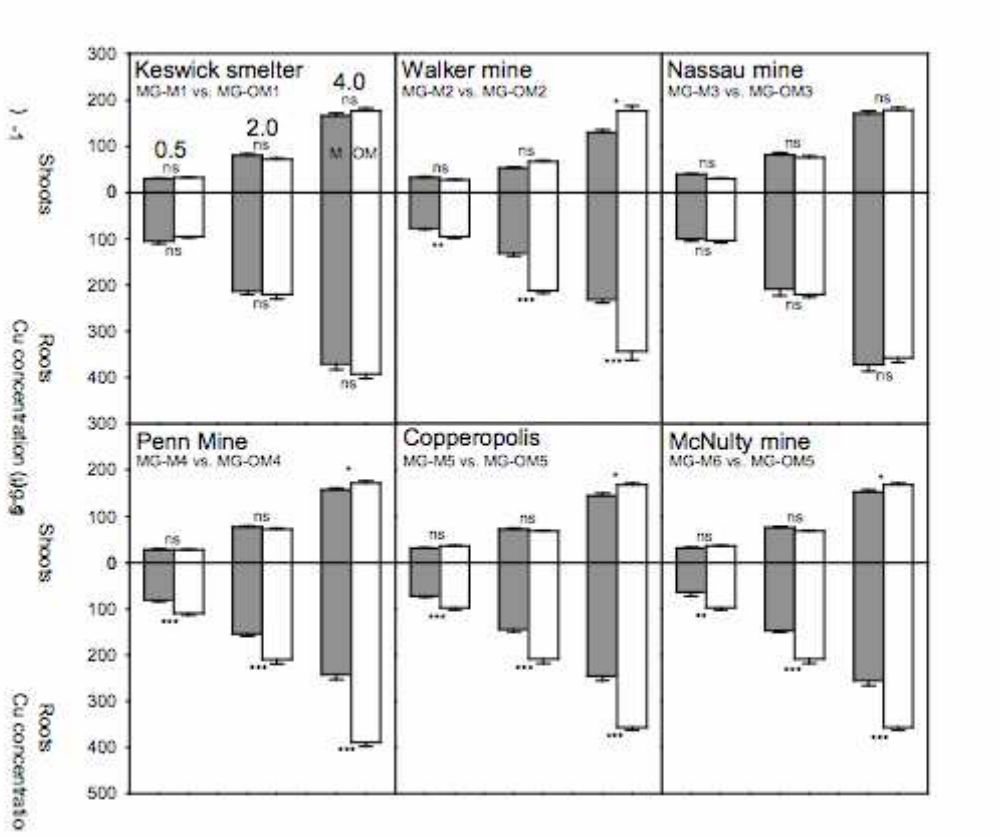


Figure 14: Root and shoot concentration of copper ions. Plants collected from paired mine (grey bars) and offmine (white bars) populations. * p<0.05. **p<0.001. ***p<0.0001.

3.3.3 Population Structure Analysis

To test for genetic isolation between mine and off-mine populations, we conducted an analysis of molecular variance (AMOVA) using the *Tol1*$_{UL}$ data set of neutral markers. We analyzed 239 individuals, 6 genotypes were removed because >30% of the markers did not successfully amplify. We found no significant effect of habitat (mine versus off-mine) on the distribution of genetic variation (Table 22). There was

significant variation distributed within habitat between populations, and within populations, within individual plants (Table 22). We found a significant correlation between geographic and genetic distance ($r=0.641$, $p<0.001$) using a Mantel test. Adjacent populations, on and off-mine, are more closely related at neutral loci then populations from the same habitat located at greater geographic distance.

Source	df	SS	Var	%Var	F	p
Among Habitat	1	20.2	-0.05	-0.01	-0.99	0.82
Among Pop, w/in Habitat	13	358.8	0.74	0.16	16.10	<0.001
Among Individ., w/in Pop	218	1074.1	1.00	0.22	21.70	<0.001
Within Individ.	233	680.5	2.92	0.63	63.19	<0.001
Total	465	2133.6	4.62			

Table 22: AMOVA analysis of distribution of genetic variation. Habitats: mine versus off mine. *** p<0.0001.

To test the consistency of these results, we analyzed the same dataset in the program STRUCTURE. We ran STRUCTURE for k = 2-13, with two independent runs at each value of k. We found that k = 9 yielded the greatest maximum likelihood estimate for the number of populations in our dataset (Table 23). To identify the populations with the greatest genetic divergence we examined how populations were assigned in STRUCTURE as k is increased from 2-9. We found that as k is increased, the Keswick, Walker, and Penn mine complexes form distinct genetic groups, with on and off mine populations at each complex clustering together (Figure 15). The large number of populations in southern Calaverous county presents a more complicated picture. MCN_M

and MCN_{OM} were the first to break from the rest of southern Calaverous county and form their own cluster (k=4, Figure 15). NAS_M and COP_M form their own independent clusters as k increases, but there is more admixture between these mines and between the neighboring off mine populations (Figure 3). Our results also inform the question of which mine population, MCN_M or COP_M, is most closely related to the two Calaverous county off-mine populations, CVR_{OM1} and CVR_{OM2}. These populations are located in between the Copperopolis and McNulty mines and the results from this STRUCTURE analysis indicates that CVR_{OM1} is more closely related to McNulty, and CVR_{OM2} is more closely related to Copperopolis (k=4-7; Figure 15). These grouping will inform the pairing of mine and off-mine populations in the subsequent analysis.

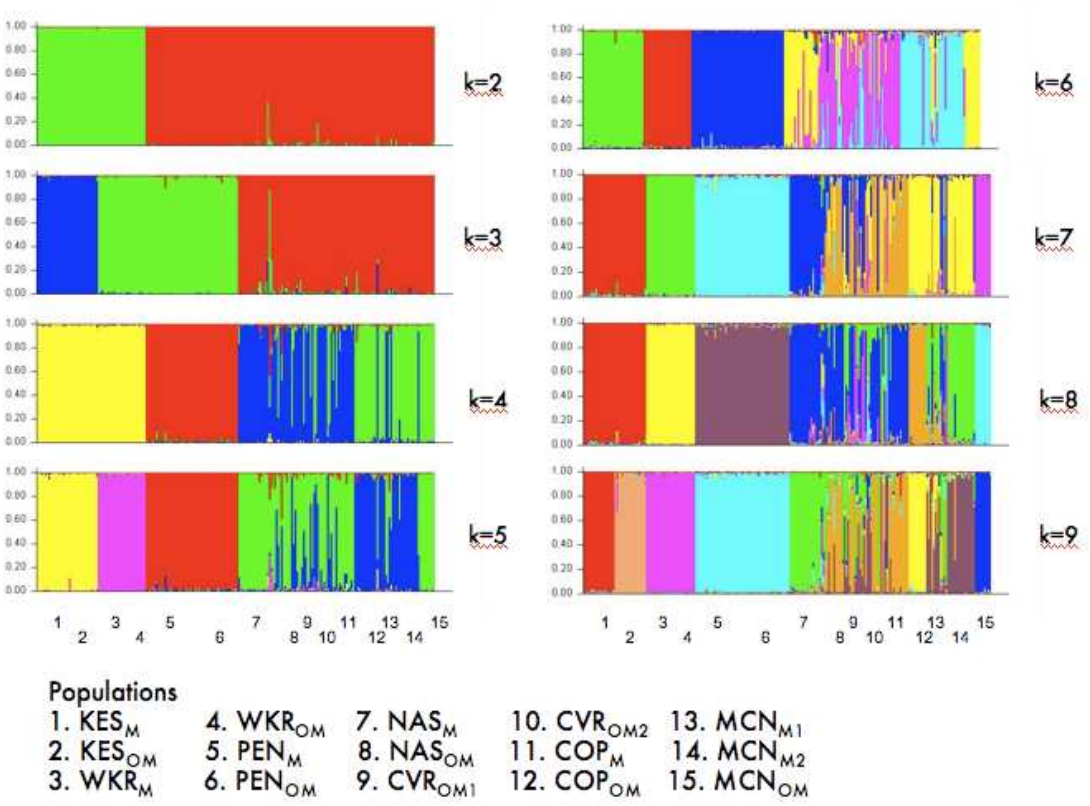


Figure 15: STRUCTURE results for allocation of genetic variation between populations.

Table 23: Maximum likelihood probability of number of populations (K) and admixture (a). Bold k=9 value best explains the data; p<0.0001 from likelihood ratio test.

K	Ln P(D)	Var[LnP(D)]	a
2	-12495.9	369.6	0.0282
2	-12494.2	366.7	0.0282
3	-12102.8	556.3	0.0291
3	-12245.1	656.2	0.0314
4	-11723.6	777.3	0.0288
4	-11714.6	763.1	0.0284
5	-11489.5	949.9	0.0283
5	-11483.3	938.9	0.0284
6	-11460.9	1067.6	0.028
6	-11394.9	1098.4	0.0279
7	-11400.3	1246.9	0.028
7	-11346.3	1211	0.0279
8	-11535.7	1452.4	0.0277
8	-11452.2	1370.2	0.0276
9	**-11319.7**	**1428.6**	**0.0275**
9	-11410.6	1487.6	0.0275
10	-11461	1438.1	0.0273
10	-11408.9	1468.6	0.0272
11	-11482.2	1471.6	0.027
11	-11389.9	1590.6	0.0272
12	-11403.2	1601.4	0.027
12	-11399.9	1593.2	0.0272
13	-11414.8	1609.1	0.027
13	-11482	1575.5	0.027

3.3.4 Evidence for Selection at Tolerance Locus in Natural Populations

We found substantial genetic variation in both marker datasets within and among populations (Table 24). The average number of alleles (N_A) at *Tol1*$_{UL}$ ranges from 4.3 to 9.9 and N_A at the *Tol1*$_{LINK}$ ranges from 1.8 to 7.8 (Table 24). In general there is a reduction in N_A at *Tol1*$_{LINK}$ markers compared to *Tol1*$_{UL}$ markers for all mine populations, this trend was not as strong in the off-mine populations. Average H_o is less then H_e for all populations, implying large deviations from Hardy-Weinberg equilibrium in these populations (Table 24). We found a reduction in H_o and H_e for all mine populations, and there was a greater reduction for the copper avoidance populations compared to the non-avoidance populations (Table 24).

Table 24: N_a, H_e, H_o for *Tol1*$_{UL}$ and *Tol1*$_{LINK}$ markers.

	Copperopolis								McNulty												Nassau							
	COP_M				COP_{OM}				MCN_{M1}				MCN_{M2}				MCN_{OM}				NAS_M				NAS_{OM}			
Loci	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He
332	24	3	0.250	0.236	26	2	0.154	0.271	28	4	0.357	0.323	28	3	0.143	0.140	18	2	0.111	0.111	40	3	0.150	0.145	22	3	0.364	0.325
836	24	11	0.583	0.902	26	12	0.769	0.923	28	12	0.571	0.910	28	3	0.143	0.680	18	5	0.778	0.752	38	10	0.684	0.844	22	10	0.818	0.788
571	24	10	0.667	0.906	26	7	0.692	0.766	28	6	0.714	0.733	28	5	0.500	0.701	18	2	0.333	0.425	38	6	0.684	0.795	22	5	0.727	0.771
617	24	8	0.250	0.848	26	11	0.462	0.815	28	12	0.714	0.857	28	9	0.643	0.733	18	5	0.556	0.601	40	10	0.550	0.881	22	9	0.545	0.887
648	24	8	0.583	0.873	26	9	0.538	0.865	28	9	0.643	0.825	28	12	0.571	0.870	18	3	0.333	0.307	38	9	0.421	0.841	22	6	0.455	0.632
278	24	10	0.500	0.902	26	14	0.615	0.914	28	13	0.500	0.931	28	10	0.571	0.899	18	6	0.667	0.817	40	13	0.800	0.912	22	10	0.727	0.892
423	24	10	0.250	0.888	26	12	0.462	0.926	28	11	0.571	0.905	28	11	0.643	0.892	18	6	0.222	0.771	40	13	0.650	0.905	22	6	0.364	0.758
672	24	11	0.583	0.924	24	10	0.667	0.797	28	10	0.571	0.910	28	5	0.571	0.696	18	3	0.667	0.621	40	9	0.600	0.851	22	11	0.727	0.918
837	22	7	0.636	0.844	26	8	0.154	0.797	26	12	0.538	0.935	26	7	0.077	0.742	18	5	0.556	0.752	40	7	0.450	0.649	20	9	0.500	0.858
641	24	13	0.583	0.938	26	15	0.846	0.945	26	13	0.769	0.935	28	17	0.786	0.950	18	6	0.889	0.824	40	16	0.800	0.913	22	9	0.636	0.844
a217	24	6	0.583	0.746	26	4	0.231	0.662	28	7	0.571	0.765	28	7	0.786	0.825	18	4	0.444	0.673	40	5	0.600	0.695	22	5	0.545	0.528
Mean		8.82	0.497	0.819		9.45	0.508	0.789		9.91	0.593	0.821		8.09	0.494	0.739		4.27	0.505	0.605		9.18	0.581	0.766		7.55	0.583	0.745
s.d.		2.79	0.164	0.200		4.01	0.242	0.192		3.05	0.115	0.179		4.25	0.256	0.220		1.56	0.240	0.231		3.84	0.188	0.224		2.62	0.156	0.183
1657	24	4	0.333	0.627	24	8	0.583	0.826	26	8	0.538	0.825	28	7	0.571	0.765	18	4	0.556	0.647	40	6	0.450	0.745	22	6	0.455	0.788
Atg5	24	2	0.333	0.290	26	7	0.462	0.812	28	2	0.143	0.349	28	3	0.571	0.582	18	4	0.333	0.575	38	3	0.368	0.411	22	7	0.636	0.840
G4.2nd	24	2	0.833	0.507	16	5	0.625	0.842	14	2	0.429	0.363	2	1	0.000	0.000	10	4	0.200	0.733	4	3	0.500	0.833	16	10	0.750	0.942
G5.I1	14	2	0.000	0.264	24	10	0.500	0.859	26	7	0.385	0.523	28	3	0.143	0.265	18	5	0.444	0.641	28	3	0.214	0.204	16	7	0.625	0.833
G9.I5	22	3	0.182	0.515	26	9	0.308	0.705	28	5	0.429	0.717	28	4	0.357	0.513	18	5	0.444	0.405	38	4	0.579	0.622	18	7	0.333	0.837
Mean		2.60	0.336	0.441		7.80	0.496	0.809		4.80	0.385	0.555		3.60	0.329	0.425		4.40	0.396	0.600		3.80	0.422	0.563		7.40	0.560	0.848
s.d.		0.89	0.310	0.157		1.92	0.123	0.061		2.77	0.147	0.212		2.19	0.256	0.298		0.55	0.135	0.123		1.30	0.139	0.256		1.52	0.165	0.057

	Penn								Keswick								Walker								Calaverous Cty Offmine							
	PEN_M				PEN_{OM}				KES_M				KES_{OM}				WKR_M				WKR_{OM}				CVR_{OM1}				CVR_{OM2}			
Loci	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He	N	Na	Ho	He
332	52	3	0.231	0.324	56	3	0.143	0.137	36	4	0.500	0.543	36	5	0.333	0.495	36	3	0.667	0.552	20	3	0.400	0.358	20	2	0.400	0.505	24	2	0.083	0.083
836	52	10	0.462	0.808	56	16	0.750	0.903	36	10	0.722	0.781	36	8	0.722	0.816	36	5	0.389	0.687	20	3	0.100	0.279	20	9	0.600	0.884	24	9	0.333	0.870
571	52	7	0.654	0.777	56	7	0.714	0.805	36	3	0.500	0.624	36	4	0.444	0.410	36	3	0.500	0.641	20	4	0.700	0.647	20	4	0.400	0.563	24	6	0.750	0.670
617	52	8	0.346	0.777	56	11	0.464	0.853	34	6	0.294	0.711	36	5	0.722	0.592	36	7	0.722	0.827	20	5	0.900	0.747	20	3	0.500	0.574	24	9	0.833	0.837
648	52	7	0.769	0.697	54	7	0.481	0.655	36	2	0.167	0.157	36	6	0.778	0.733	36	6	0.667	0.829	20	4	0.900	0.684	20	6	0.600	0.821	24	7	0.667	0.833
278	52	10	0.500	0.792	56	12	0.464	0.823	36	7	0.611	0.729	36	5	0.556	0.778	36	7	0.778	0.778	20	5	0.500	0.721	20	6	0.700	0.789	24	10	0.833	0.920
423	52	9	0.462	0.767	56	11	0.500	0.845	36	6	0.389	0.649	36	8	0.611	0.824	36	6	0.278	0.810	20	5	0.600	0.726	18	7	0.889	0.856	24	7	0.417	0.833
672	52	13	0.692	0.736	56	6	0.464	0.540	36	6	0.556	0.706	36	9	0.889	0.771	36	5	0.389	0.598	20	5	0.700	0.695	20	5	0.700	0.668	24	9	0.667	0.804
837	52	9	0.423	0.747	54	9	0.407	0.843	36	6	0.444	0.825	36	7	0.500	0.775	34	6	0.118	0.784	20	6	0.200	0.832	18	8	0.333	0.889	24	7	0.000	0.797
641	52	11	0.615	0.760	56	12	0.607	0.762	20	9	0.100	0.911	14	3	0.000	0.484	36	6	0.778	0.775	20	7	0.800	0.832	20	6	0.400	0.716	24	8	0.750	0.891
a217	52	7	0.654	0.678	56	7	0.321	0.323	36	6	0.944	0.763	36	8	0.833	0.789	36	4	0.667	0.722	20	2	0.600	0.505	20	4	0.800	0.711	24	7	0.667	0.848
Mean		8.55	0.528	0.715		9.18	0.483	0.681		5.91	0.475	0.673		6.18	0.581	0.679		5.27	0.541	0.728		4.45	0.582	0.639		5.45	0.575	0.725		7.36	0.545	0.763
s.d.		2.62	0.163	0.135		3.63	0.170	0.249		2.34	0.242	0.198		1.94	0.257	0.153		1.42	0.221	0.096		1.44	0.264	0.182		2.11	0.184	0.136		2.16	0.294	0.234
1657	50	4	0.160	0.349	56	5	0.429	0.641	32	7	0.375	0.734	36	7	0.500	0.819	36	3	0.611	0.589	20	5	0.800	0.779	20	6	0.400	0.579	20	6	0.700	0.842
Atg5	52	2	0.077	0.145	56	5	0.250	0.479	36	2	0.500	0.513	36	5	0.667	0.721	36	2	0.056	0.056	20	5	0.700	0.816	18	5	0.778	0.817	24	6	0.417	0.725
G4.2nd	42	1	0	0	54	6	0.370	0.553	28	2	0.357	0.495	30	6	0.267	0.805	8	2	0.250	0.250	14	3	0.143	0.560	20	8	0.900	0.853	24	9	0.583	0.772
G5.I1	34	1	0	0	40	2	0.050	0.450	32	2	0.000	0.121	22	7	0.364	0.840	6	2	0.000	0.533	16	3	0.125	0.642	20	11	0.700	0.889	16	6	0.250	0.783
G9.I5	50	1	0	0	54	5	0.370	0.435	36	5	0.444	0.730	36	10	0.722	0.810	36	3	0.389	0.379	20	5	0.500	0.774	20	6	0.300	0.811	24	7	0.833	0.888
Mean		1.80	0.047	0.099		4.60	0.294	0.512		3.60	0.335	0.518		7.00	0.504	0.799		2.40	0.261	0.361		4.20	0.454	0.714		7.20	0.616	0.790		6.80	0.557	0.802
s.d.		1.30	0.071	0.153		1.52	0.151	0.085		2.30	0.196	0.250		1.87	0.194	0.046		0.55	0.250	0.217		1.10	0.311	0.108		2.39	0.255	0.122		1.30	0.230	0.064

To determine if there is evidence of selection at *Tol1* in natural populations we measured the extent of genomic differentiation markers linked to *Tol1* between paired mine and off mine populations. We used the program *Fdist* (Beaumont and Nichols 1996) to measure F_{st} and H_e for the *Tol1*$_{LINK}$ markers and compared those observed values to an expected distribution of F_{st} generated from our *Tol1*$_{UL}$ dataset. We first examined the distribution of F_{st} values for the two marker datasets between the mine and off-mine populations at Copperopolis. We found that the three markers with tightest linkage to *Tol1*, *Sc84.Atg5*, *Sc84.G4*, and *Sc84.G5*, show significantly elevated levels of F_{st} compared to neutral expectations (Table 25, Figure 16). The mean F_{st} for all *Tol1*$_{LINK}$ markers was five times greater than for the *Tol1*$_{UL}$ markers. These results demonstrate colocalization between a genomic region of significantly elevated Fst between mine and off-mine populations and a previously mapped copper tolerance QTL.

We next examined the distribution of F_{st} and H_e at each marker dataset for the five remaining pairs of mine and off-mine populations. We found that three of the five populations had elevated F_{st} at three or more *Tol1*$_{LINK}$ markers, similar to Copperopolis (Table 25, Figure 16). The McNulty mine showed elevated F_{st} at the four markers most tightly linked to *Tol1*. The Walker mine showed significantly elevated F_{st} at three *Tol1*$_{LINK}$ markers, the other two markers, *Sc84.G5* and *Sc84.G9* had poor amplification in this population. The Penn mine had significantly elevated F_{st} at three of the five markers,

although the marker closest to *Tol1* was not significantly elevated. In these three populations the average F_{st} is 4-5 times greater for *Tol1*$_{LINK}$ markers than the *Tol1*$_{UL}$ markers (Table 25). Two mines, Keswick and Nassau, had significantly elevated F_{st} at a single marker, Sc84.G5 (Table 25, Figure 16). The average F_{st} is 1.5 – 2.5 times greater for *Tol1*$_{LINK}$ markers than the *Tol1*$_{UL}$ markers in these two populations, however this trend is entirely attributable to differentiation at *Sc84.G5* (Table 25).

The four populations that show significantly elevated F_{st} at multiple *Tol1*$_{LINK}$ markers are the same populations with the avoidance mechanism of tolerance. The two populations with a greatly reduced level of differentiation at the *Tol1*$_{LINK}$ markers have the nonavoidance mechanism of tolerance. The avoidance populations also had reduced H_o and H_e compared to non-avoidance mine populations and off-mine populations.

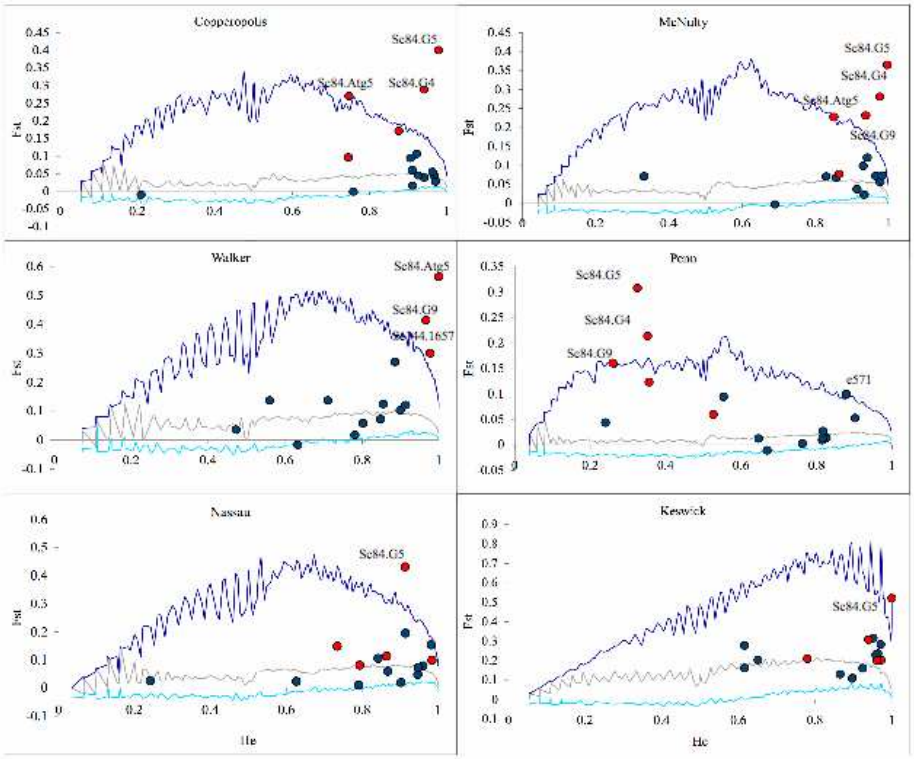


Figure 16: Distribution of H_e by F_{st} for six paired on and off mine populations. Red circles represent observed values of *Tol1*$_{LINK}$ and blue circles represent observed values *Tol1*$_{UL}$. Dark blue line is 95% CI for higher then expected F_{st}, light blue line is 95%CI for lower then expected F_{st}, grey line is mean value of F_{st}.

Table 25: Values of F_{st} and H_e for *Tol1*$_{LINK}$ and *Tol1*$_{UL}$ for six pair mine and offmine population pairs. p values >0.95 indicate markers with significantly elevated amount of shared variation and p<0.05 indicate markers with significantly reduced amount of shared variation.

		COP_M v. COP_{OM} & CVR_{OM2}			MCN_{M1+2} v. MCN_{OM} & CVR_{OM}			PEN_M v. PEN_{OM}			WKR_M v. WKR_{OM}			NAS_M v. NAS_{OM}			KES_M v. KES_{OM}		
Marker	Location	H_e	F_{st}	p	H_e	F_{st}	p	H_e	F_{st}	p	H_e	F_{st}	p	H_e	F_{st}	p	H_e	F_{st}	p
e332	*Tol1*$_{UL}$	0.209	-0.010	ns	0.332	0.071	ns	0.241	0.043	ns	0.472	0.036	ns	0.241	0.025	ns	0.651	0.203	ns
e836	*Tol1*$_{UL}$	0.942	0.039	ns	0.914	0.037	ns	0.902	0.052	ns	0.560	0.137	ns	0.867	0.059	ns	0.897	0.110	ns
e571	*Tol1*$_{UL}$	0.922	0.106	ns	0.690	-0.003	>0.995	0.878	0.100	>0.95	0.633	-0.017	>0.995	0.791	0.010	ns	0.617	0.162	ns
e617	*Tol1*$_{UL}$	0.911	0.060	ns	0.858	0.068	ns	0.826	0.014	ns	0.849	0.072	ns	0.960	0.079	ns	0.952	0.315	ns
e648	*Tol1*$_{UL}$	0.926	0.045	ns	0.931	0.098	ns	0.668	-0.011	ns	0.803	0.058	ns	0.914	0.194	ns	0.616	0.277	ns
e278	*Tol1*$_{UL}$	0.968	0.044	ns	0.978	0.056	ns	0.815	0.009	ns	0.856	0.124	ns	0.947	0.047	ns	0.865	0.129	ns
e423	*Tol1*$_{UL}$	0.963	0.055	ns	0.934	0.023	ns	0.816	0.013	ns	0.782	0.018	ns	0.982	0.153	ns	0.964	0.236	ns
e672	*Tol1*$_{UL}$	0.911	0.016	ns	0.943	0.120	ns	0.646	0.013	ns	0.886	0.270	ns	0.902	0.020	ns	0.958	0.229	ns
e837	*Tol1*$_{UL}$	0.905	0.094	ns	0.965	0.072	ns	0.817	0.027	ns	0.900	0.102	ns	0.843	0.106	ns	0.897	0.108	ns
e641	*Tol1*$_{UL}$	0.971	0.027	ns	0.983	0.073	ns	0.763	0.002	ns	0.914	0.121	ns	0.948	0.073	ns	0.971	0.282	ns
a217	*Tol1*$_{UL}$	0.758	-0.001	ns	0.830	0.071	ns	0.553	0.094	ns	0.711	0.137	ns	0.626	0.023	ns	0.924	0.160	ns
Mean		**0.853**	**0.043**		**0.851**	**0.062**		**0.721**	**0.032**		**0.760**	**0.096**		**0.820**	**0.072**		**0.847**	**0.201**	
sd		**0.22**	**0.04**		**0.19**	**0.03**		**0.19**	**0.04**		**0.15**	**0.08**		**0.22**	**0.06**		**0.14**	**0.07**	
	cM - *Tol1*$_{LINK}$																		
Sc103.1657	1.54	0.876	0.171	ns	0.865	0.077	ns	0.526	0.060	ns	0.978	0.301	<0.05	0.865	0.114	ns	0.972	0.201	ns
Sc84.G9	1.22	0.745	0.097	ns	0.938	0.231	<0.05	0.259	0.160	<0.05	0.967	0.414	<0.01	0.794	0.082	ns	0.962	0.200	ns
Sc84.G5	1.04	0.979	0.401	<0.005	0.997	0.365	<0.005	0.325	0.308	<0.005	-	-	-	0.913	0.432	<0.01	1.000	0.520	<0.05
Sc84.G4	0.86	0.942	0.289	<0.005	0.977	0.281	<0.005	0.352	0.214	<0.05	-	-	-	0.984	0.098	ns	0.939	0.308	ns
Sc84.Atg5	0.74	0.747	0.270	<0.05	0.851	0.228	<0.05	0.356	0.123	ns	1.000	0.564	<0.005	0.734	0.148	ns	0.780	0.209	ns
Mean		**0.858**	**0.245**		**0.926**	**0.237**		**0.364**	**0.173**		**0.981**	**0.426**		**0.858**	**0.175**		**0.931**	**0.288**	
sd		**0.11**	**0.12**		**0.07**	**0.10**		**0.10**	**0.09**		**0.02**	**0.13**		**0.10**	**0.15**		**0.09**	**0.14**	

3.4 Discussion

We investigated parallel evolution of copper tolerance in *M. guttatus*. We found that the soil in mine populations had more copper then off mine populations, although there was no consistent trend with the other three heavy metal investigated. We found a difference in the [Cu] in root and shoot tissue between mine populations, four populations had reduced [Cu] and two had the same level of [Cu] when compared to off-mine populations. We found that adjacent mine and off-mine populations share more genetic variation at a genome-wide marker set than other mine populations, indicating that the mine populations are genetically independent from each other. Finally, we find strong evidence that populations that maintain lower [Cu] in their root tissue have higher genomic divergence at markers linked to a QTL for copper tolerance.

3.4.1 Parallel Evolution of Copper Tolerance

Parallel evolution describes the evolution of the same phenotype in genetically independent populations inhabiting similar environments. We find that all mine populations inhabit soils with 10-25 times the concentration of copper as the average off-mine population. To inhabit these environments, all mine populations must have evolved some level of copper tolerance because high [Cu] is toxic to plants (Baker 1987; Shaw 1989). To determine if the mine populations are genetically independent, and therefore if they display parallel evolution, we analyzed the distribution of neutral

genetic variation with an AMOVA and MANTEL test. We found that genetic variation sorts according to geographic distance and not habitat differences. The STRUCTURE analysis confirmed these patterns, although the three mines in southern Calaverous county (Copperopolis, McNulty and Nassau) are not as differentiated from each other as the other mine complexes. The high differentiation between populations is consistent with other observations of genetic structure in *M. guttatus* (Lowry et al. 2008). Thus, we conclude that copper tolerance has evolved in parallel due to similar selection pressures on genetically independent populations. This conclusion is in agreement with the consensus emerging from many other studies on the evolution of heavy metal tolerance (Bradshaw 1991). However, our results are also consistent with a scenario of tolerance evolving in a single population, followed by migration of tolerance alleles to other mines and hybridization with local populations. We believe this maybe unlikely because most of these mines began operation in the 1860s, providing a window of only ~150 generations for a chance migration event and interbreeding to obscure any evidence of shared ancestry. This scenario only seems likely in the case of the Copperopolis and McNulty mines because they have the same mechanism of tolerance, they are located with 10km of each other, and they share more genetic variation then any other mine populations. These results enable us to ask, do populations experiencing parallel

selection evolve tolerance via the same physiological mechanisms? Does tolerance evolve via the same genetic mechanisms?

3.4.2 Physiological Mechanism of Copper Tolerance

We contend that copper tolerance in *M. guttatus* has evolved in parallel via at least two independent physiological mechanisms. This conclusion is supported by our finding that four mine populations, Copperopolis, McNulty, Penn, and Walker, maintain lower root [Cu] and two mine populations, Keswick and Nassau, have equal root [Cu] compared to off mine populations (Figure 2). Our finding of lower root [Cu] in the Copperopolis population is consistent with previous measurements of root [Cu] in this population (Strange and Macnair 1991; Tilstone and Macnair 1997). However, our finding of reduced shoot [Cu] contrasts with previous results of Copperopolis plants accumulating more copper in their shoots then off-mine plants (Tilstone and Macnair 1997). There are two possible mechanisms of tolerance in populations with lower root [Cu]: plants may reduce the influx or increase the efflux of copper ions into/out of their roots. Molecular characterization of heavy metal transporters in *Arabidopsis thaliana* has found that exclusion of copper can be achieved through the reduced expression or functionality of the cell membrane copper transport proteins: *AtCOPT1,2*, *AtZIP2,4*, and *AtFRO2,3* or through increased expression of *AtSPL7*, a transcription factor repressor of *AtCOPT1,2* (Burkhead et al. 2009). Conversely, a reduction in root [Cu] could be

achieved through the increased expression or functionality of the heavy-metal efflux protein *AtHMA5* (Burkhead et al. 2009). This avoidance mechanism of tolerance has been observed in many plant species (Wu and Antonovics 1975; Baker et al. 1983; Lolkema et al. 1984; de Vos et al. 1992). However, this is not the only mechanism of copper tolerance in *M. guttatus*.

Tolerance in the Keswick and Nassau populations is not achieved through avoidance of copper; [Cu] in root and shoot tissue for these mine populations was indistinguishable from off-mine populations. It is unlikely that these populations are intolerant, because the soil copper levels at these mines is much greater then levels seen in off-mine populations (Table 1). Other studies have found the same concentration of heavy metals in root tissue of tolerant and nontolerant genotypes (Lin and Wu, 1994; Hall et al. 2002). In these cases, tolerance could be achieved through the intracellular sequestration of heavy metals. Investigation into the genetic basis of Zinc tolerance in *A. hallerii* identified three QTLs that co-localize with the heavy metal transporters *AhHMA4*, *AhMTP1-A,B* (Willems et al. 2007). *AhMTP1-A,B* are Zn-specific vacuole transporters (Hall and Williams 2003). *AhHMA4* also transports heavy metals into cellular organelles but this protein contributes to Zn hyperaccumulation rather then tolerance (Roosens et al. 2008, Hanikenne et al. 2008). Other studies of serpentine soil adapted plant species *Sedum anglicum* and *Lasthenia californica* found that the

sequestration of magnesium is an important aspect of tolerance (Tibbetts and Smith 1992; Rajakaruna et al. 2003). Research in *A. thaliana* indicates that intracellular sequestration of copper could be achieved through the increased expression or functionality of vacuole copper transport proteins, *AtCOPT3*, *5* (Burkhead et al. 2009). It will be interesting to test if this protein affects tolerance in the Keswick and Nassau populations, however this is not the only possible non-avoidance mechanism of tolerance.

Tolerance can be achieved through intracellular detoxification. The two most investigated mechanisms of detoxification are: the heavy metal binding peptides, phytochelatins, and the heavy metal chelators, metallothioneins (Macnair 1993; Burkhead et al. 1993; Hall et al. 2002). It is clear that phytochelatins and metallothioneins act as intracellular metal homeostasis molecules/proteins and that they are effective in ameliorating heavy-metal toxicity at low levels, but there has been much debate as to whether they contribute to tolerance at high heavy metal levels (Schultz and Hutchinson 1988, De Knecht et al. 1992; Hall et al 2002; Jack et al. 2007; Hassinen et al. 2009). An investigation into metallothioneins from Copperopolis lines found that only 6% of the available copper was bound to metallothioneins in root tissue, so the authors concluded this was the not mechanism of tolerance (Robinson and Thurman 1986). Blocking the activity phytochelatins in the roots of the same lines from Copperopolis reduced their

copper tolerance, but it remains unknown as to whether this is the causal mechanism of tolerance (Salt et al. 1989). Given our results, it seems that the Copperopolis population may achieve tolerance through avoidance and detoxicification, however the relative contribution of these two mechanisms remains to be determined. Additionally, it would be interesting to repeat these experiments with the Keswick and Nassau populations and determine if this mechanism contributes to tolerance in these lines. Our measurements of genetic differentiation at the major tolerance locus in our six mine populations will inform the differentiation of the avoidance and non-avoidance populations.

3.4.3 Selection on Tol1 in Copper Avoidance Populations

We found that the four populations with the avoidance mechanism have significantly elevated F_{st} at 3-4 markers tightly linked to *Tol1*. We also observe that these populations had fewer alleles and reduced heterozygosity for *Tol1*$_{LINK}$ markers. This contrasts with the two non-avoidance populations, which have significantly elevated F_{st} at only a single *Tol1*$_{LINK}$ marker. The strong genetic differentiation between populations and reduced genetic variation is consistent with models of strong selection, over short time span, on initially rare alleles (Beaumont and Nichols 1996; Charlesworth et al. 1998; Beaumont and Balding 2004; Novembre and Di Rienzo 2009). Selection on alleles at higher frequency or over long periods of time would reduce the signal of genetic

differentiation between populations (Hermission and Pennings 2005). Many researchers have cautioned against concluding that selection drives elevated Fst between populations (Storz 2005, Stinchecombe and Hoekstra 2008, Excoiffer et al. 2009, Nosil et al. 2009). However, we are not conducting a genome-wide scan for outliers, we are testing a specific hypothesis that *Tol1LINK* markers have significantly elevated F_{st} compared to unlinked markers. We find this pattern is repeated for multiple markers in four independent populations. Similar patterns of genetic variation are seen in domestication alleles and drug resistance (Palaisa et al. 2004; Caicedo et al. 2007; Sutter et al. 2007) and in natural, ecologically mediated, selection in sticklebacks (Cano et al 2006; Makinen et al. 2008). Does our finding of elevated Fst in the avoidance populations indicate that the physiological effects of *Tol1* are to reduce the [Cu] in root tissue?

We found a perfect correlation between the mine populations with reduced [Cu] in their root tissue and elevated Fst at the *Tol1LINK* markers. These data suggest that *Tol1* affects the influx or efflux of copper into root tissue. This assertion is supported by a previous experiment that found reduced [Cu] in root tissue of near isogenic line with *Tol1* introgressed into a nontolerant background (Strange and Macnair 1991). Thus, *Tol1* may be an intercellular copper transport proteins or a regulator of these proteins discussed in the previous section. Could the four avoidance populations share the same genetic basis for tolerance?

The co-localization between the elevated region of F_{st} and a previously mapped locus for copper tolerance suggests that the tolerance locus mapped in the Copperopolis population also contributes to tolerance in the McNulty, Walker and Penn populations. Examining how marker F_{st} is effected by distance to the tolerance locus, we found that the tightest linked marker, Sc84.Atg5 (position: 0.74cM/1.01MB), has significantly elevated F_{st} in all avoidance populations except Penn, however even in this population F_{st} at this locus is still an order of magnitude higher then nearly all of the *Tol1*$_{UL}$ loci (Table 25). Conversely, the furthest marker, Sc103.1657 (position: 1.54 cM/2.11MB), is only significantly elevated in a one of the four avoidance populations. These data suggest that conclusion that tolerance has a shared genetic basis in all four avoidance populations. Three previous studies came to similar conclusions after finding colocalization between QTLs and elevated regions of F_{st} in replicated populations (Olafsdottir et al. 2006; Rodgers and Bernatchez 2007; Baxter et al. 2010), however another study of edaphic specialization in *Helianthus* failed to find any such colocalization (Yatabe et al. 2007). Because we have not conducted specific experiments (i.e., QTL mapping) to determine the genetic architecture of copper tolerance in the avoidance populations, save Copperopolis, we cannot definitively state whether they have a shared genetic basis. Furthermore, the region of elevated F_{st} extends over 1MB and contains many genes that may affect tolerance in different populations.

Evidence from two other investigations into parallel evolution to extreme edaphic habitats also finds evidence for shared genetic basis for tolerance. Complementation crosses between two copper tolerant subspecies of *Siliene vulgaris* of indicate that this phenotype is controlled by the same locus (Schat et al 1996). A genome scan for elevated F_{st} between serpentine tolerant and nontolerant *A. lyrata* populations identified many loci that may contribute to tolerance, including Mg and Ca ion transporters and metallithiones; one of these loci was also shown to be differentiated in an independent serpentine *A. lyrata* (Turner et al. 2010). We conclude that our data suggest that copper tolerance has a shared genetic basis in the four avoidance populations, however we need to conduct additional experiments to specifically test this hypothesis. What is the origination of copper tolerance in these populations?

3.4.4 Origination of Copper Tolerance Alleles

We found that there are at least two independent derivations of copper tolerance, each with different physiological mechanisms, in *M. guttatus*. When comparing populations that share the same mechanism for tolerance, we would like to determine if tolerance evolved from independent new mutations or from shared ancestral polymorphisms. Our data demonstrate that the avoidance populations share the same physiological mechanism of tolerance, but the two non-avoidance populations maybe achieve tolerance through either detoxification or intracellular sequestration. We

observed elevated F_{st} in the *Tol1*$_{LINK}$ markers in all four avoidance populations, this indicates that the causal allele could not have been maintained at intermediate to high frequency, but must have been a low frequency segregating polymorphism or a new mutation (Hermission and Pennings 2005). Our finding of selection at the same locus in all four avoidance populations supports the hypothesis that copper tolerance evolved from shared ancestral polymorphism segregating at low frequency. There are multiple copper transport proteins and transcription factors that could be mutated to achieve the same avoidance phenotype (Burkhead et al. 2009), so it is unlikely that independent mutations would all target the same protein. Two caveats of this assertion are that we do not know how many of these possible candidate genes are linked to *Tol1*, nor do we have any knowledge of the mutation rate or pleiotropic constraint for each of these proteins. The recent age of these mines, all are less than 150 years old, also argues against the independent mutation hypothesis. Alleles that reduce the level of copper uptake may be segregating in ancient populations because of high concentration of naturally occurring copper outcrops throughout California (Jenkins 1948). Previous research on heavy metal tolerance indicates that segregating polymorphisms for tolerance are a major determinant of whether populations will be able to colonize metal-enriched environments.

The consensus from studies of heavy metal tolerance is that segregating polymorphisms for tolerance are the primary factor determining whether a population may colonize a metal enriched habitat (Bradshaw 1991). Studies of plant communities on and off copper mines in the UK found that species with a low frequency of tolerant plants in off-mine populations were able to colonize mine environments, while those species lacking tolerant individuals in off-mine populations were not found in the mine environment (Gartside and McNeilly 1974). This result was replicated in a study of plant communities growing under galvanized electricity pylons (Al-Hiyaly et al. 1988, 1990). Rainwater causes zinc to leach from these structures into surrounding soils. The recent origin of these structures argues against opportunities for gene flow between disjunct regions and their ubiquity allows for multiple comparisons of which species are present in the contaminated areas. Al-Hiyaly and colleagues (1990) found that particular species were unable to colonize the contaminated soil at every pylon, even though it may be abundant in the area. Colonization of contaminated soils occurs only when neighboring populations also exhibit low frequency of tolerance (Al-Hiyaly et al. 1993). Additional studies have found that high levels of zinc tolerance has evolved in genetically independent populations of *A. halleri*, although this work is complicated by the fact that this species is a Zn hyperaccumulator and all populations have some level of tolerance (Pauwels et al. 2005). Although these studies indicate that segregating polymorphisms

are important for the evolution of heavy metal tolerance, we will not be able to fully address this question in our system until we fine map the actual causal locus in the Copperopolis population and determine if the same locus contributes to tolerance in the other avoidance populations.

3.4.5 Selection for Copper Tolerance Fixes Hybrid Incompatibility Allele at Copperopolis

These results also address an interesting question about the evolution of a post-zygotic incompatibility factor in the Copperopolis population. Macnair and Christie (1983) proposed that a hybrid incompatibility factor became fixed in the Copperopolis population as a pleiotropic byproduct of selection for copper tolerance. Macnair and Christie never found a recombinant between copper tolerance and post-zygotic incompatibility and hypothesized that the same locus caused both phenotypes. We previously demonstrated that these two phenotypes are controlled by two separate, but very tightly linked loci, *Nec1* and *Tol1* (Chapter 2). In this chapter, we demonstrate that the region of elevated F_{st} in the Copperopolis population includes *Nec1*, implying that selection on copper tolerance caused a hybrid incompatibility factor to become fixed in this population.

We previously demonstrate that *Nec1* maps to a region in between two markers located at position 234kb and 297kb of scaffold 84 (Figure XX). Here we find that the window of elevated Fst in the Copperopolis populations extends to marker Sc84.G5,

which is located at 346kb on scaffold 84 (orientation of markers shown in Figure XX). This provides strong evidence that selection for copper tolerance resulted in *Nec1* being swept to fixation in the Copperopolis populations. The elevated Fst between on and off mine populations could not be caused by the F1 incompatibility itself because 1) the hybrid incompatibility only manifests in crosses between Copperopolis genotypes and four other populations, none of which are located immediately adjacent to the mine (Christie and Macnair 1987) and 2) we find the same window of elevated Fst in three additional mine populations, all of which lack F1 hybrid necrosis. The elevated Fst was caused by selection for copper tolerance, this is the first example of habitat-mediated selection indirectly causing the fixation of a post-zygotic inviability locus.

3.4.6 Conclusion

The objective of this study was to determine if copper tolerance has independently evolved in multiple *M. guttatus* mine populations and, if it has occurred multiple times, to determine if it was through similar or divergent genetic mechanisms. We find that tolerance has evolved independently multiple times, via two distinct physiological mechanisms. All four populations that exclude copper have undergone changes at the same locus, which colocalizes with a previously mapped QTL for copper tolerance. These results indicate that copper tolerance sometimes evolves via the same

physiological and genetic mechanisms in independent populations or can evolve via alternative physiological mechanisms and alternative genetic mechanisms.

Appendix A

Here we demonstrate, for two- and three-step pathways, that the concentration of pathway intermediates is proportional to the k_i for reactions upstream of the intermediate and inversely proportional to k_i for reactions downstream of the intermediate. Proof for longer pathways is analogous.

Consider first a two-reaction pathway with intermediate B. The rate of change of the concentration of B, [B] is given by

$$d[B]/dt = k_1 I - \alpha k_1 [B] - k_2 [B] \ .$$

When flux reaches an equilibrium, this rate of change is zero, which means that the equilibrium concentration of B is

$$[B]_{eq} = k_1 I/(\alpha k_1 + k_2) \ .$$

Increasing k_1 results in an increase in $[B]_{eq}$, as can be seen by taking the derivative with respect to k_1:

$$d[B]_{eq}/dk_1 = k_2 I/ (\alpha k_1 + k_2)^2 \ ,$$

which is always positive because α, I, k_1 and k_2 are positive. Similarly, increasing k_2 results in a decrease in $[B]_{eq}$, since $d[B]_{eq}/dk_2 = - \alpha k_1 I/ (\alpha k_1 + k_2)^2$, is always negative. Thus, an increase in k_1, corresponding to the reaction upstream of B, causes $[B]_{eq}$ to increase, while an increase in k_2, which is downstream of B, causes $[B]_{eq}$ to decrease.

Now consider a three-reaction pathway, with intermediates B and C. The rates of change of the concentrations of these two intermediates are

$$d[B]/dt = k_1 I + \alpha k_2[C] - \alpha k_1[B] - k_2[B]$$

$$d[C]/dt = k_2[B] - \alpha k_2[C] - k_3[C]$$

At flux equilibrium, these to equations will equal 0. Solving for the equilibrium concentrations of [B] and [C] gives

$$[B]_{eq} = k_1(\alpha^2 k_2 + k_3)I/(\alpha^2 k_1 k_2 + \alpha k_1 k_3 + k_2 k_3)$$

$$[C]_{eq} = k_1 k_2 I/(\alpha^2 k_1 k_2 + \alpha k_1 k_3 + k_2 k_3)$$

Taking derivatives,

$$d[B]_{eq}/dk_1 = k_2 k_3(\alpha k_2 + k_3)I/D^2 \quad > 0 \text{ always}$$

$$d[B]_{eq}/dk_2 = -k_1(k_3)^2 I/D^2 \quad < 0 \text{ always}$$

$$d[B]_{eq}/dk_3 = -\alpha k_1(k_2)^2 I/D^2 \quad < 0 \text{ always}$$

$$d[C]_{eq}/dk_1 = k_1(k_2)^2 k_3 I/D^2 \quad > 0 \text{ always}$$

$$d[C]_{eq}/dk_2 = \alpha(k_1)^2 k_3 I/D^2 \quad > 0 \text{ always}$$

$$d[C]_{eq}/dk_3 = -k_1 k_2 I(\alpha k_1 + k_2)/D^2 \quad > 0 \text{ always}$$

where $D = (\alpha^2 k_1 k_2 + \alpha k_1 k_3 + k_2 k_3)$.

Thus, $[B]_{eq}$ is increased by an increase in k_1 (upstream) and is decreased by an increase in either k_2 or k_3 (downstream); and $[C]_{eq}$ is increased by an increase in either k_1 or k_2 (upstream) and decreased by an increase in k_3 (downstream).

References

Al-Hiyaly, S. A., T. McNeilly, and A. D. Bradshaw. 1988. The effects of zinc contamination from electricity pylons-evolution in a replicated situation. *New Phytologist* 110: 571-580.

Al-Hiyaly, S. A. K., T. McNeilly, and A. D. Bradshaw. 1990. The effect of zinc contamination from electricity pylons. Contrasting patterns of evolution in five grass species. *New Phytologist* 114: 183-190.

Albe, K. R., and B. E. Wright, 1992. Systems analysis of the tricarboxylic acid cycle in Dictyostelium discoideum. II. Control analysis. *Journal of Biological Chemistry* 267: 3106-3114.

Alcazar, R., A. V. Garcia, J. E. Parker, and M. Reymond. 2009. Incremental steps toward incompatibility revealed by arabidopsis epistatic interactions modulating salicylic acid pathway activation. *Proceedings of the National Academy of Science* 106: 334-339.

Allen, W. R. and P. M. Sheppard. 1971. Copper tolerance in some Californian populations of the monkey flower, Mimulus guttatus. *Proceedings of the Royal Society London, Series B* 177: 177-196.

Antao, T., A. Lopes, R. Lopes, A. Beja-Pereira, and G. Luikart. 2008. Lositan: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9: 323-333.

Antonovics, J., A.D. Bradshaw, and R.G. Turner. 1971. Heavy metal tolerance in plants. *Advanced Ecological Research* 7: 1-85.

Bagheri, H. C., and G. P. Wagner, 2004. Evolution of dominance in metabolic pathways. *Genetics* 168: 1713-1735.

Baker, A., R. Brooks, A. Pease, and F. Malaisse. 1983. Studies on copper and cobalt tolerance in three closely related taxa within the Genussilene l. (caryophyllaceae) from Zaire. *Plant and Soil* 73: 377-385.

Baker, A. J. M. 1987. Metal tolerance. *New Phytologist* 106: 93-111.

Bakker, E. G., C. Toomajian, M. Kreitman, and J. Bergelson. 2006. A genome-wide survey of r gene polymorphisms in Arabidopsis. *The Plant Cell* 18: 1803-1818.

Barbash, D. A. 2008. Clash of the genomes. *Cell* 135: 1002-1003.

Barbash, D. A., D. F. Siino, A. M. Tarone, and J. Roote. 2003. A rapidly evolving myb-related protein causes species isolation in Drosophila. *Proceedings of the National Academy of Science* 100: 5302-5307.

Bateson, W., ed. 1909. Heredity and variation in modern lights. Edited by A. C. Seward. *Darwin and modern science.* Cambridge: Cambridge University Press.

Baxter, S. W., N. J. Nadeau, L. S. Maroja, P. Wilkinson, B. A. Counterman, et al. 2010. Genomic hotspots for adaptation: The population genetics of mullerian mimicry in the Heliconius melpomene clade. *PLoS Genetics* 6: e1000794.

Beaumont, M. A. 2005. Adaptation and speciation: What can Fst tell us? *Trends in Ecology & Evolution* 20: 435-440.

Beaumont, M. A. and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13: 969-980.

Beaumont, M. A. and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society London, Series B* 263: 1619-1626.

Bell, M.A. and S. A. Foster. 1994. *The evolutionary biology of the threespine stickleback*. New York: Oxford University Press.

Bikard, D., D. Patel, C. Le Mette, V. Giorgi, C. Camilleri, M. J. Bennett, and O. Loudet. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. *Science* 323: 623-626.

Bomblies, K. 2009. Too much of a good thing? Hybrid necrosis as a by-product of plant immune system diversification. *Botany* 87: 314-323.

Bomblies, K., J. Lempe, P. Epple, N. Warthmann, C. Lanz, J. L. Dangl, and D. Weigel. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biology* 5: e236.

Bomblies, K. and D. Weigel. 2007. Hybrid necrosis: Autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* 8: 382-393.

Bost, B., D. de Vienne, F. Hospital, L. Moreau, and C. Dillmann, 2001. Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics* 157: 1773-1787.

Bradshaw, A. D. 1991. The croonian lecture, 1991: Genostasis and the limits to evolution. *Philosophical Transaction Royal Society London, Series B* 333: 289-305.

Brady, K. U., A. R. Kruckeberg, and H. D. Bradshaw Jr. 2005. Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of. Ecology & Systematics* 36: 243-266.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang, and D. A. Barbash. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. *Science* 314: 1292-1295.

Burkhead, J. L., K. A. Gogolin Reynolds, S. E. Abdel-Ghany, C. M. Cohu, and M. Pilon. 2009. Copper homeostasis. *New Phytologist* 182, no. 4: 799-816.

Caicedo, A. L., S. H. Williamson, R. D. Hernandez, A. Boyko, A. Fledel-Alon, et al. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* 3: e163.

Cano, J. M., C. Matsuba, H. Mikinen, and J. Meril. 2006. The utility of qtl-linked markers to detect selective sweeps in natural populations; a case study of the Eda gene and a linked marker in threespine stickleback. *Molecular Ecology* 15: 4613-4621.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research* 70: 155-174.

Christie, P. and M. R. Macnair. 1987. The distribution of postmating reproductive isolating genes in populations of the yellow monkey flower, Mimulus guttatus. *Evolution* 41: 571-578.

Clark, A. G., 1991. Mutation-selection balance and metabolic control theory. *Genetics* 129: 909-923.

Clemens, S. 2001. Molecular mechanisms of plant metal tolerance and homeostasis. *Planta* 212: 475-486.

Cobbett, C. S. 2000. Phytochelatins and their roles in heavy metal detoxification. *Plant Physiology* 123: 825-832.

Colosimo, Pamela F., Kim E. Hosemann, Sarita Balabhadra, Guadalupe Villarreal, Jr., Mark Dickson, et al. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307: 1928-1933.

Cornish-Bowden, A., 1989. Metabolic control theory and biochemical systems theory: different objectives, different assumptions, different results. *Journal of Theoretical Biology* 136: 365-377.

Coyne, J. A. and H. A. Orr. 2004. *Speciation*. Sunderland, Mass: Sinauer Associates.

Cronwright, G. R., J. M. Rohwer, and B. A. Prior, 2002. Metabolic Control Analysis of Glycerol Synthesis in Saccharomyces cerevisiae. *Applied Environmental Microbiol*ogy 68: 4448-4456.

De Vos, C. H. R., M. J. Vonk, R. Vooijs, and H. Schat. 1992. Glutathione depletion due to copper-induced phytochelatin synthesis causes oxidative stress in Silene cucubalus. *Plant Physiology* 98: 853-858.

Dobzhansky, Theodosius. 1937. *Genetics and the origin of species.* New York: Columbia University Press.

_____. 1951. *Genetics and the origin of species, 3rd edition.* New York: Columbia University Press.

Doyle, J. J. and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Dykhuizen, D. E., A. M. Dean, and D. L. Hartl, 1987. Metabolic flux and fitness. *Genetics* 115: 25-31.

Eanes, W. F., 1999. Analysis of selection on enzyme polymorphisms. *Annual Review of. Ecology & Systematics.* 30: 301-326.

Ellison, C. K., O. Niehuis, and J. Gadau. 2008. Hybrid breakdown and mitochondrial dysfunction in hybrids of Nasonia parasitoid wasps. *Journal of Evolutionary Biology* 21: 1844-1851.

Endler, John A. 1986. *Natural selection in the wild*. Princeton, New Jersey: Princeton University Press.

Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103: 285-298.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.

Ferree, P. M. and D. A. Barbash. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in drosophila. *PLoS Biol* 7: e1000233.

Fisher, R. A., 1930. *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.

Fishman, Lila and John H. Willis. 2005. A novel meiotic drive locus almost completely distorts segregation in Mimulus (monkeyflower) hybrids. *Genetics* 169: 347-353.

Flowers, J. M., E. Sezgin, S. Kumagai, D. D. Duvernell, et al., 2007. Adaptive evolution of metabolic pathways in Drosophila. *Molecular Biology and Evolution* 24: 1347-1354.

Flowers, T. J., M. A. Hajibagheri, and N. J. W. Clipson. 1986. Halophytes. *The Quarterly Review of Biology* 61: 313-337.

Futuyma, D. J. 2005. *Evolution*. Sunderland, Mass: Sinauer Associates.

Gachon, C. M. M., M. Langlois-Meurinne, and P. Saindrenan. 2005. Plant secondary metabolism glycosyltransferases: The emerging functional analysis. *Trends in Plant Science* 10: 542-549.

Gartside, D. W. and T. McNeilly. 1974. The potential for evolution of heavy metal tolerance in plants. *Heredity* 32: 335-348.

Groen, A., C. van Roermund, R. Vervoorn, and J. Tager, 1986. Control of gluconeogenesis in rat liver cells. *Biochemistry Journal* 237: 379-389.

Haldane, J. B. S. 1932. *The causes of evolution*. New York: Harper Brothers.

Hall, J. L. 2002. Cellular mechanisms for heavy metal detoxification and tolerance. *Journal of Experimental Botany* 53: 1-11.

Hall, J. L. and Lorraine E. Williams. 2003. Transition metal transporters in plants. *Journal of Experimental Botany* 54: 2601-2613.

Hall, M. C. and J, H. Willis. 2005. Transmission ratio distortion in intraspecific hybrids of Mimulus guttatus: Implications for genomic divergence. *Genetics* 170: 375-386.

Hanikenne, M., I. N. Talke, M. J. Haydon, C. Lanz, A. Nolte, et al. 2008. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of hma4. *Nature* 453: 391-395.

Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985. Limits of adaptation: the evolution of selective neutrality. *Genetics* 111: 655-674.

Hassinen, V. H., M. Tuomainen, S. Peraniemi, H. Schat, S. O. Karenlampi, and A. I. Tervahauta. 2009. Metallothioneins 2 and 3 contribute to the metal-adapted phenotype but are not directly linked to zn accumulation in the metal hyperaccumulator, Thlaspi caerulescens. *Journal of Experimental Botany* 60: 187-196.

Hedrick, P. W., 2000. *Genetics of Populations*. Boston, Mass: Jones and Bartlett.

Heinrich, E. M., F. Montero, and J. C. Nuno, 1999. The structural design of glycolysis: an evolutionary approach. *Biochemistry Society Transactions* 27: 294-298.

Heinrich, R., and S. Schuster, 1998. The modelling of metabolic systems. Structure, control and optimality. *Biosystems* 47: 61-77.

Heinrich, R., and T. A. Rapoport, 1974. A linear steady-state treatment of enzymatic chains. *European Journal of Biochemistry* 42: 89-95.

Heinrich, R., F. Montero, E. Klipp, T. G. Waddell, and E. Melendez-Hevia, 1997. Theoretical approaches to the evolutionary optimization of glycolysis. *European Journal of Biochemistry* 243: 191-201.

Heinrich, R., S. Schuster, and H. Holzhutter, 1991. Mathematical analysis of enzymic reaction systems using optimization principles. *European Journal of Biochemistry* 201: 1-21.

Hermisson, J. and P. S. Pennings. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335-2352.

Hill, S. A., J. H. Bryce, and C. J. Leaver, 1993. Control of succinate oxidation by cucumber (Cucumis sativus L.) cotyledon mitochondria. *Planta* 190: 51 - 57.

Hoekstra, H. E., and J. A. Coyne, 2007. The locus of adaptation: EvoDevo and the genetics of adaptation. *Evolution* 61: 995-1016.

Hong, E., Y. Jeong, J. Ryu, R. Amasino, B. Noh, and Y. Noh. 2009. Temporal and spatial expression patterns of nine Arabidopsis genes encoding jumonji c-domain proteins. *Molecules and Cells* 27: 481-490.

Jack, E., H. W. J. Hakvoort, A. Reumer, J. A. C. Verkleij, H. Schat, and W. H. O. Ernst. 2007. Real-time pcr analysis of metallothionein-2b expression in metallicolous and non-metallicolous populations of Silene vulgaris (moench) garcke. *Environmental and Experimental Botany* 59: 84-91.

Jenkins, O. P. 1948. *Copper in California*. San Francisco, California: Department of Natural Resources, Division of Mines.

Jeuken, M. J. W., N. W. Zhang, L. K. McHale, K. Pelgrom, E. den Boer, et al. 2009. Rin4 causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. *The Plant Cell* 109: 703-734.

Kacser, H., and J. A. Burns. 1973. The control of flux. *Symposium of the Society of Experimental Biology.* 32: 65-104.

Kacser, H., and J. A. Burns, 1981. The molecular basis of dominance. *Genetics* 97: 639-666.

Kashiwaya, Y., K. Sato, N. Tsuchiya, S. Thomas, et al., 1994. Control of glucose utilization in working perfused rat heart. *Journal Biological Chemistry* 269: 25502-25514.

Keightley, P. D. 1989. Models of quantitative variation of flux in metabolic pathways. *Genetics* 121: 869-876.

Keightley, P. D. 1996. Metabolic models of selection response. *Journal of Theoretical Biology* 182: 311-316.

Keightley, P. D. 1996. A metabolic basis for dominance and recessivity. *Genetics* 143: 621-625.

Kimura, M., 1983. *The neutral theory of molecular evolution.* Cambridge: Cambridge University Press.

Kishore, N., Y. B. Tewari, D. L. Akers, R. N. Goldberg, and E. Wilson Miles, 1998. A thermodynamic investigation of reactions catalyzed by tryptophan synthase. *Biophysical Chemistry* 73: 265-280.

Knecht, J. A. de, P. L. M. Koevoets, J. A. C. Verkleij, and W. H. O. Ernst. 1992. Evidence against a role for phytochelatins in naturally selected increased cadmium tolerance in Silene vulgaris (moench) garcke. *New Phytologist* 122: 681-688.

Kruger, J., C. M. Thomas, C. Golstein, M. S. Dixon, M. Smoker, S. Tang, L. Mulder, and J. D. G. Jones. 2002. A tomato cysteine protease required for cf-2-dependent disease resistance and suppression of autonecrosis. *Science* 296: 744-747.

Langford, A. N. 1948. Autogenous necrosis in tomatoes immune from Cladosporium fulvum cooke. *Canadian Journal of Research* 26: 35-64.

LaPorte, D. C., K. Walsh, and D. E. ,. J. Koshland, 1984. The branch point effect. Ultrasensitivity and subsensitivity to metabolic control. *Journal of Biological Chemistry* 259: 14068-14075.

Lee, H., J. Chou, L. Cheong, N. Chang, S. Yang, and J. Leu. 2008. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135: 1065-1073.

Lin, S. and L. Wu. 1994. Effects of copper concentration on mineral nutrient uptake and copper accumulation in protein of copper-tolerant and nontolerant Lotus purshianus. *Ecotoxicology and Environmental Safety* 29: 214-228.

Lindsay, W. L. and W. A. Norvell. 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. *Soil Science Society of America Journal* 42: 421-428.

Lolkema, P. C., M. H. Donker, A. J. Schouten, and W. H. O. Ernst. 1984. The possible role of metallothioneins in copper tolerance of Silene cucubalus. *Planta* 162: 174-179.

Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. guez-Schettino. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279: 2115-2118.

Lowry, D. B., R. C. Rockwood, and J. H. Willis. 2008. Ecological reproductive isolation of coast and inland races of Mimulus guttatus. *Evolution* 62: 2196-2214.

Lu, Y., and M. D. Rausher, 2003. Evolutionary rate variation in anthocyanin pathway genes. *Molecular Biology and Evolution* 20: 1844-1853.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* 4: 981-994.

Lynch, M. and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sunderland, Mass: Sinauer Associates.

Macnair, M. R. 1981. The uptake of copper by plants of Mimulus guttatus differing in genotype primarily at a single major copper tolerance locus. *New Phytologist* 88: 723-730.

_____. 1983. The genetic control of copper tolerance in the yellow monkey flower, Mimulus guttatus. *Heredity* 50: 283-293.

Macnair, M. R. and P. Christie. 1983. Reproductive isolation as a pleiotropic effect of copper tolerance in Mimulus guttatus? *Heredity* 50: 295-302.

Macnair, M. R., Suzanne E. Smith, and Q. J. Cumbes. 1993. Heritability and distribution of variation in degree of copper tolerance in Mimulus guttatus at Copperopolis, California. *Heredity* 71: 445-455.

Marschner, H. 1995. *Mineral nutrition in higher plants.* London: Academic Press.

Masly, J. P., C. D. Jones, M. A. F. Noor, J. Locke, and H. A. Orr. 2006. Gene transposition as a cause of hybrid sterility in Drosophila. *Science* 313: 1448-1450.

Mayr, E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.

Mazat, J., C. Reder, and T. Letellier, 1996. Why are most flux control coefficients so small? *Journal of Theoretical Biology* 182: 253-258.

Meiklejohn, C. D. and Y. Tao. 2009. Genetic conflict and sex chromosome evolution. *Trends in Ecology & Evolution* 25: 215-223.

Meyers, B. C., A. Kozik, A. Griego, H. Kuang, and R. W. Michelmore. 2003. Genome-wide analysis of nbs-lrr-encoding genes in Arabidopsis. *The Plant Cell* 15: 809-834.

Mihola, O., Z. Trachtulec, C. Vlcek, J. C. Schimenti, and J. Forejt. 2009. A mouse speciation gene encodes a meiotic histone h3 methyltransferase. *Science* 323: 373-375.

Mikinen, H. S., J. M. Cano, and J. Meril. 2008. Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback Gasterosteus aculeatus populations. *Molecular Ecology* 17: 3565-3582.

Muller, H. J. 1942. Isolating mechanisms, evolution, and temperature. *Biological Symposium* 6: 12-86.

Nelson, D. L., and M. M. Cox, 2000. *Lehninger principles of biochemistry*. New York: Worth Publishers.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* 18: 375-402.

Novembre, J. and A. Di Rienzo. 2009. Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics* 10: 745-755.

Orr, H. A., 1998. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52: 935-945.

Orr, H. A., 2002. The population genetics of apatation: The adaptation of DNA sequences. *Evolution* 56: 1317-1330.

Orr, H. A., 2003. The distribution of fitness effects among beneficial mutations. *Genetics* 163: 1519-1526.

Orr, H. A., 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* 6: 116-127.

Palaisa, K., M. Morgante, S. Tingey, and A. Rafalski. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize y1 gene are indicative of an asymmetric selective sweep. *Proceedings of the National Academy of Science* 101: 9885-9890.

Pauwels, M., P. Saumitou-Laprade, A. Catherine Holl, Daniel Petit, and Isabelle Bonin. 2005. Multiple origin of metallicolous populations of the pseudometallophyte Arabidopsis halleri (Brassicaceae) in central europe: The cpDNA testimony. *Molecular Ecology* 14: 4403-4414.

Phadnis, N., and J. D. Fry, 2005. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics* 171: 385-392.

Presgraves, D. C. 2010. The molecular evolutionary basis of species formation. *Nature Reviews Genetics* 11: 175-180.

Presgraves, D. C., L. Balagopalan, S. M. Abmayr, and H. A. Orr. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature* 423: 715-719.

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

Pritchard, L., and D. B. Kell, 2002. Schemes of flux control in a model of Saccharomyces cerevisiae glycolysis. *European Journal of Biochemistry* 269: 3894-3904.

Rajakaruna, N., B. G. Baldwin, R. Chan, A. M. Desrochers, B. A. Bohm, and J. Whitton. 2003. Edaphic races and phylogenetic taxa in the Lasthenia californica complex (Asteraceae: Heliantheae): An hypothesis of parallel evolution. *Molecular Ecology* 12: 1675-1679.

Rauser, W. E. 1999. Structure and function of metal chelators produced by plants. *Cell Biochemistry and Biophysics* 31: 19-48.

Rauser, W. E. 2003. Phytochelatins. *Annual Review of Biochemistry* 59: 61-86.

Rausher, M. D., R. E. Miller, and P. Tiffin, 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Molecular Biology and Evolution* 16: 266-274.

Rausher, M. D., Y. Lu, and K. Meyer, 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *Journal Molecular Evolution* 67: 137-144.

Robinson, N. J. and D. A. Thurman. 1986. Involvement of a metallothionein-like copper complex in the mechanism of copper tolerance in Mimulus guttatus. *Proceedings of the Royal Society London, Series B* 227: 493-501.

Rogers, S. M. and L. Bernatchez. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonus sp. Salmonidae) species pairs. *Molecular Biology and Evolution* 24: 1423-1438.

Roosens, N. H. C. J., G. Willems, and P. Saumitou-Laprade. 2008. Using Arabidopsis to explore zinc tolerance and hyperaccumulation. *Trends in Plant Science* 13: 208-215.

Roussel, R., P. G. Carlier, J. Robert, G. Velho, and G. Bloch, 1998. 13 C/ 31 P NMR studies of glucose transport in human skeletal muscle. *Proceedings of the National Academy of Science* 95: 1313-1318.

Salt, D. E., I. Baxter, and B. Lahner. 2008. Ionomics and the study of the plant ionome. *Annual Review of Plant Biology* 59: 709-733.

Salt, D. E., D. A. Thurman, A. B. Tomsett, and A. K. Sewell. 1989. Copper phytochelatins of mimulus guttatus. *Proceedings of the Royal Society of London. Series B* 236: 79-89.

Savageau, M. A., 1976. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Reading, Mass: Addison-Wesley.

Savageau, M. A., 1992. Dominance according to metabolic control analysis: major achievement or house of cards? *Journal of Theoretical Biology* 154: 131-136.

Savageau, M. A., and A. Sorribas, 1989. Constraints among molecular and systemic properties: Implications for physiological genetics. *Journal of Theoretical Biology* 141: 93-115.

Sawamura, K. and M. Yamamoto. 1997. Characterization of a reproductive isolation gene, zygotic hybrid rescue, of Drosophila melanogaster by using minichromosomes. *Heredity* 79: 97-103.

Schat, H., R. Vooijs, and E. Kuiper. 1996. Identical major gene loci for heavy metal tolerances that have independently evolved in different local populations and subspecies of Silene vulgaris. *Evolution* 50: 1888-1895.

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 Science* 96: 11910-11915.

Schluter, D. 2000. *The ecology of adaptive radiation*. Oxford: Oxford University Press.

_____. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737-741.

Schultz, C. L. and T. C. Hutchinson. 1988. Evidence against a key role for metallothionein-like protein in the copper tolerance mechanism of Deschampsia cespitosa. *New Phytologist* 110: 163-171.

Simpson, G. G. 1953. *The major features of evolution*. New York: Columbia University Press.

Smith, S. E. and M. R. MacNair. 1998. Hypostatic modifiers cause variation in degree of copper tolerance in Mimulus guttatus. *Heredity* 80: 760-768.

Sokal, R. R. and F. J. Rolf. 1994. *Biometery*. New York: W. H. Freeman.

Stern, D. L., and V. Orgogozo, 2008. The loci of evolution: How predictable is genetic evolution? *Evolution* 62: 2155-2177.

Stinchcombe, J. R. and H. E. Hoekstra. 2007. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* 100: 158-170.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14: 671-688.

Strange, J. and M. R. Macnair. 1991. Evidence for a role for the cell membrane in copper tolerance of Mimulus guttatus. *New Phytologist* 119: 383-388.

Sutter, Nathan B., Carlos D. Bustamante, Kevin Chase, Melissa M. Gray, Keyan Zhao, et al. 2007. A single igf1 allele is a major determinant of small size in dogs. *Science* 316: 112-115.

Szathmary, E., 1993. Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics* 133: 127-132.

Tao, Y., L. Araripe, S. B. Kingan, Y. Ke, H. Xiao, and D. L. Hartl. 2007. A sex-ratio meiotic drive system in Drosophila simulans. II: An x-linked distorter. *PLoS Biology* 5: e293.

Tao, Y., S. Chen, D. L. Hartl, and C. C. Laurie. 2003. Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. I. Differential accumulation of hybrid male sterility effects on the x and autosomes. *Genetics* 164: 1383-1397.

Tao, Y., Z. Zeng, J. Li, D. L. Hartl, and C. C. Laurie. 2003. Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics* 164: 1399-1418.

Tilstone, G. H. and M. R. Macnair. 1997. The consequence of selection for copper tolerance on the uptake and accumulation of copper in Mimulus guttatus. *Annals of Botany* 80: 747-751.s

Ting, C., S. Tsaur, M. Wu, and C.-I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282: 1501-1504.

Tordoff, G. M., A. J. M. Baker, and A. J. Willis. 2000. Current approaches to the revegetation and reclamation of metalliferous mine wastes. *Chemosphere* 41: 219-228.

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. 2010. Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature Genetics* 42: 260-263.

Tewari, Y. B., P. Y. Jensen, N. Kishore, M. P. Mayhew, et al., 2002. Thermodynamics of reactions catalyzed by PABA synthase. *Biophysical Chemistry* 96: 33-51.

Tewari, Y. B., R. N. Goldberg, A. R. Hawkins, and H. K. Lamb, 2002. A thermodynamic study of the reactions: {2-dehydro-3-deoxy--arabino-heptanoate 7-phosphate(aq)=3-dehydroquinate(aq) + phosphate(aq)} and {3-dehydroquinate(aq)=3-dehydroshikimate(aq) + H2O(l)}. *Journal of Chemistry Thermodynamics* 34: 1671-1691.

Thomas, S., and D. A. Fell, 1998. A control analysis exploration of the role of ATP utilisation in glycolytic-flux control and glycolytic-metabolite-concentration regulation. *European Journal of Biochemistry* 258: 956-967.

van der Vlag, J., R. van't Hof, K. van Dam, and P. W. Postma, 1995. Control of glucose metabolism by the enzymes of the glucose phosphotransferase system in Salmonella typhimurium. *European Journal of Biochemistry*. 230: 170-182.

Vickery, R. K., Jr. 1959. Barriers to gene exchange within Mimulus guttatus (Scrophulariaceae). *Evolution* 13: 300-310.

Watt, W. B., and A. M. Dean, 2000. Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. *Annual Review of Genetics* 34: 593-622.

Weir, B. S. and C. C. Cockerham. 1984. Estimating f-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Willems, G., D. B. Drager, M. Courbot, C. Gode, N. Verbruggen, and P. Saumitou-Laprade. 2007. The genetic basis of zinc tolerance in the metallophyte Arabidopsis halleri ssp. halleri (Brassicaceae): An analysis of quantitative trait loci. *Genetics* 176: 659-674.

Wisniewski, E., F. N. Gellerich, and W. S. Kunz, 1995. Distribution of flux control among the enzymes of mitochondrial oxidative phosphorylation in calcium-activated saponin-skinned rat musculus soleus fibers. *European Journal of Biochemistry* 230: 549-554.

Wittbrodt, J., D. Adam, B. Malitschek, W. Maueler, F. Raulf, A. Telling, S. M. Robertson, and M. Schartl. 1989. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing tu locus in Xiphophorus. *Nature* 341: 415-421.

Wu, L., W. Wang, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen, 2004. A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. *European Journal of Biochemistry* 271: 3348-3359.

Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2007. Mimulus is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100: 220-230.

Wu, L. and J. Antonovics. 1975. Zinc and copper uptake by Agrostis stolonifera, tolerant to both zinc and copper. *New Phytologist* 75: 231-237.

Yatabe, Y., N. C. Kane, C. Scotti-Saintagne, and L. H. Rieseberg. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, Helianthus annuus and H. petiolaris. *Genetics* 175: 1883-1893.

Biography

I was born in Downey, California in 1982 to Douglas James and Victoria Anne Wright. I graduated Warren High School in Downey in the year 2000. I then attended the University of California, San Diego from 2000-2003, where I completed my Bachelor of Science degree in Ecology, Behavior, and Evolution. At UCSD, I worked in the lab of Dr. Lin Chao on experimental evolution and natural variation in the bacteriophage, $\phi 6$. I came to Duke University in 2004, became a candidate for PhD in 2006, and completed my PhD in the Spring of 2010.