Chapter 3

METABOLOMICS AND PLANT QUANTITATIVE TRAIT LOCUS ANALYSIS – THE OPTIMUM GENETICAL GENOMICS PLATFORM?

Daniel J. Kliebenstein
Department of Plant Sciences, Mail Stop 3, University of California, One Shields Ave, Davis, CA 95616, USA

1 INTRODUCTION

Biologists have long strived to understand what causes phenotypic differences between two individuals. This includes differences in morphology, disease susceptibility, and physiology as well as potential metabolic differences underlying these higher-order phenotypes. The diversity between individuals is partitioned into both environmental and genetic variation. Most genetic variation studied to date tends to be qualitative such that there are one or more distinct and non-overlapping phenotypic states. However, most phenotypic differences are quantitative such that there are numerous overlapping phenotypic states (Mackay, 2001; Flint et al., 2001; Lynch and Walsh, 1998; Mauricio, 2001). It has been known for nearly a century that the approximate genetic position of loci controlling these quantitative traits can be identified through associating marker and phenotype variation in a structured population (Sax, 1923). This association is the foundation for Quantitative Trait Locus (QTL) mapping experiments that attempt to identify the number, phenotypic impact and interaction of loci controlling a quantitative trait.

The latest incarnation of the QTL experiment is genetical genomics that phenotypes genetic mapping populations with genomics technology (Jansen and Nap, 2001). The goal is to merge the genomics technologies high-throughput and highly parallel phenotyping capacity, i.e., microarrays, proteomics, and metabolomics, with genetic segregation to test or generate specific hypothesis. The rationale is that a specific genes expression level is easier to quantify than the more complex developmental or physiological traits. Thus, by identifying loci controlling the differential gene expression

B.J. Nikolau and E. Syrkin Wurtele (eds.), Concepts in Plant Metabolomics, 29–44.

patterns for all an organism's genes and comparing this to those loci controlling a specific physiological trait, the researcher could develop a systems biological understanding of more complex traits. Genetical genomics has predominantly utilized microarray analysis of stable mapping populations in a variety of species (Morley et al., 2004; Yvert et al., 2003; Brem and Kruglyak, 2005; Brem et al., 2002; Schadt et al., 2003).

Microarrays present a basic fiscal problem in that it is expensive to phenotype all lines in a mapping population much less replicate the phenotyping. Thus, microarrays are fiscally limited to small, highly defined mapping populations and replication limited to genes with highly heritable expression differences. This is a serious limitation, as most complex physiological traits are moderate-to-low heritability and controlled by numerous loci that require large populations with replicated experimental designs for reliable detection. Therefore, fiscal limitations alone will hinder microarray use in genetical genomics to all but the very largest or well funded of laboratories. Metabolomics platforms may provide a more widespread entry into genetical genomics. Metabolomics is much cheaper per sample than transcriptomics, enabling large populations to be studied with sufficient replication for moderate-to-low heritability traits. Additionally, most metabolomics platforms are higher-throughput than transcriptomics, allowing for rapid analysis (Fiehn, 2001; Fiehn et al., 2000; Hall et al., 2002).

Numerous studies have investigated QTL controlling plant metabolites but none with a metabolomics purview (Kliebenstein et al., 2002a; Kliebenstein et al., 2001a; Kliebenstein et al., 2002b; Monforte et al., 2001; Santos and Simon, 2002; Bushman et al., 2002; Thorup et al., 2000; McMullen et al., 1998; Byrne et al., 1996). This chapter's goal is to help provide guidance in developing, designing, and interpreting metabolomics genetical genomic experiments. I will focus on three questions that are frequently asked by individuals starting a metabolite QTL project: (1) How do I design the experiment? (2) What traits/variables do I measure? (3) What will I find? This will draw on literature both involved with the theory of QTL formation as well as experimental analysis of metabolite QTL detection and interpretation.

2 QTL QUESTIONS AND FINDINGS FOR METABOLOMICS

2.1 How do I design the experiment?

This question is best handled in three interrelated parts: population structure, population size, and replication. All three aspects are intertwined, such that population structure will influence the other two and vice versa, but I will deal with them separately. For more detailed information see the enclosed references (Mackay, 2001; Mauricio, 2001).

2.1.1 Which population do I chose?

For genetical genomics experiments, the optimal population structure is either Recombinant Inbred or Advanced Intercross lines. These populations allow for recombination and transgressive segregation similar to an F_2 population but are taken to homozygosity allowing independent replicated measurements of a given line. Homozygosity also increases the populations' power by forcing each genomic position to only have one of the two opposing haplotypes instead of the three possibilities in F_2 populations. Inbred line populations are not feasible in all systems due to generation time, inbreeding depression, or self-incompatability. In these species, the next best population structures are typically backcross populations as there are only two allelic classes at each locus, heterozygote, and one homozygote. Another factor that should be considered in determining the population is the availability of previously genotyped populations with phenotypic differences of interest. This is valuable as the majority of time and expense in any new population is not phenotyping but instead generating and genetically mapping the population. Thus, previously existing populations are highly desirable even if the structure is not optimal.

2.1.2 What population size do I use?

The next decision to resolve is the population's size. The general rule in determining the optimum population size is the larger the better. Ideally, populations should contain at least 300 individuals or lines. Larger populations provide several benefits. The first is that they have more recombination events increasing precision in measuring a QTLs position. Secondly, larger populations have more power to separate closely linked QTL due to the increased recombination. The increased line numbers also allow for better capacity to detect two- and three-way epistatic interactions because there are more lines in each combinatorial class. Finally, the larger population sizes allow for higher replication in terms of number of lines with Allele X at

position Y. Populations with less than 300 individuals can be utilized but will have limited power for traits with more than a couple QTL or moderate-to-low phenotypic effect QTL. In genetical genomics experiments, most traits may be controlled by numerous QTL with predominantly low-to-moderate phenotypic effect and thus small populations should be avoided (Mackay, 2001; Lynch and Walsh, 1998; Brem and Kruglyak, 2005).

2.1.3 How many replicates should I conduct?

Once the population is chosen, the next question is how many replicate measurements per line should be conducted and how these should be organized. The key to deciding these issues is to measure the experimental sources of variation. This involves designing an experiment whereby several samples are taken per plant with multiple independent plants per parental genotype per replicate. Multiple independent replicates are conducted and all samples independently analyzed *via* metabolomics. Analysis of variance for this experiment will allow the researcher to estimate the variation from spatial differences within a plant, from differences between plants, from differences between replicate experiments, and from differences between genotypes as well as any interactions between the different levels. The optimum result is that most of the variance is genetic with the rest of the error being split between plants within a replicate or between replicates. If this is the case, it is best to take one measurement per plant with each line being represented by two or more plants per replicate.

The analysis of parental variance also allows the researcher to obtain a very rough estimate of each traits heritability be estimating the variance due to genotype difference. There is a common perception that low heritability traits require high-replicate numbers to successfully map QTL. However, calculations show that even for traits with 30% heritability, only six replicates are required to diminish the error in the mean trait estimate to approximately 10% (Denby et al., 2004). Thus, it should be possible to identify QTL for most traits with less than 10 and as few as 6 replicates per line. Metabolomics platforms are probably the best current technology for fiscally achieving this replication in large populations. The analysis of parental variance will allow the researcher to identify the heritability distribution for the metabolites and make an informed decision on replication. Previous metabolite profiling projects have found heritabilities that range from 20% to 90% with most being in the range of 50–70% (Kliebenstein et al., 2002a; McMullen et al., 1998; Byrne et al., 1996; Kliebenstein et al., 2001b).

The ability to measure interactions in the above variance test is a key element of properly designing a QTL experiment. If there is a significant interaction between genotype and replicate, this suggests the presence of genotype × environment interactions. Previous metabolite profiling and

microarray QTL mapping projects have identified significant genotype × environment interactions (Brem and Kruglyak, 2005; Kliebenstein et al., 2002a). One option to minimize this is that each line should be repeated enough times per replicate to allow for QTL analysis within each replicate as there could be different QTL identified depending upon environmental fluctuation. Additionally, the researcher could attempt to better control the environmental variance by controlling the growth conditions between replicates to minimize this difficulty. Alternatively, the researcher may only be interested in QTL that impact the trait in all environments and would thus conduct the analysis in multiple environments.

The identification of a significant interaction of genotype with either within plant variance or between plant variance in the parental analysis suggests that there may be a developmental difference between the parents that is impacting the sampling. The best way to minimize this variance is to ensure that the same tissue at the same developmental stage is being sampled in all cases. A detailed analysis of the sources of variance before conducting a QTL mapping experiment will greatly enhance both the potential for success and the resulting QTL maps interpretability. This is especially important in a metabolomics genetical genomics experiment where thousands of traits will be analyzed simultaneously.

3 WHAT TRAITS/VARIABLES DO I MEASURE?

There are several aspects to this question. This includes what guidelines to use in deciding upon a metabolomics platform. Another important question to contemplate is which variables to use in the QTL mapping. Finally, should the data be altered to conform to the expectation of normality and what potential errors does this introduce? Each of these questions is dealt with below.

3.1 Which metabolomics platform to utilize?

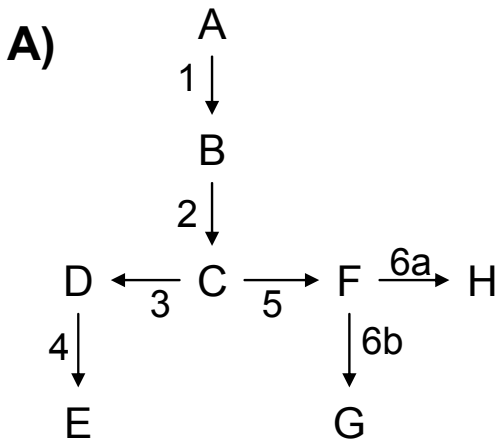
The first decision is which metabolomics platform should be utilized. This involves a compromise between the analytical speed and information content per analysis. The optimum platform should have significant high-throughput capacity to allow for the thousands of samples that are required for a statistically powerful QTL mapping experiment. In addition to high-throughput the best technology would individually quantify specific compounds and provide identification where possible and structural information for all compounds detected. This optimum requirement for individual quantification and identification provides the maximal power in the downstream QTL analysis. A number of high-throughput platforms like IR and NMR platforms are limited in providing specific compound

information and do not measure as many compounds per sample as other platforms. Thus, when QTL are identified the researcher will not be fully sure of the phenotypes identification. Thus, the researcher will be challenged to develop specific hypothesis about the locus's molecular function possibly even after cloning the underling gene. However, with platforms such as GC-MS and the rapidly gaining (LC-MS)/monolithic columns, it is possible to quantify and identify hundreds if not thousands of compounds (Bino et al., 2004; Tolstikov et al., 2003). Thus, when a QTL is found, the researcher will know the exact compound that locus is regulating and will be aided in developing specific hypothesis about the underlying molecular function.

3.2 What variables should I use for mapping?

The second part of this question is what aspects of the output are actually valid variables for QTL analysis. The most obvious variables are the actual amount of each individual compound. Considering that most metabolomics platforms can reliably detect hundreds of compounds, this creates a massive number of traits for QTL mapping. There are suggestions to decrease this dimensionality by using regression analysis to identify metabolite clusters, and then use an individual compound within each cluster to identify QTLs for that cluster. While this will decrease the computational power required, it will also decrease the experiments information content. The base assumption in regression clustering is that if two compounds are 80% correlated, that the other 20% is due to measuring error. However, it is equally likely that this 20% discrepancy is due to differences in the genetic control for the two compounds. Using a single compound per cluster would lose this genetic information. A better solution is to generate QTL software that can analyze 1000s of traits on the same population and present the results in a coherent manner. A challenge that is equally present for genetical genomics experiments using transcriptomics and proteomics.

Numerous variables/traits can be generated for QTL mapping using metabolomics data. The first is the absolute value of each variable (Figure 3-1B). Often times, there are known or predicted metabolic pathways providing relational context to the metabolites (Figure 3-1A). This relational context provides the ability to generate variables interrogating the interrelation between compounds (Figure 3-1C–E) (Weckworth et al., 2004; Steuer et al., 2003). These variables can either be the sum of specific groups of metabolites, the ratio between specific metabolites, or the ratio between different groups (Kliebenstein et al., 2001a). For instance, the equations in Figure 3-1C sequentially ask about the loci controlling the accumulation of the whole pathway (A–H), the accumulation of only those compounds on the right side (F–H), and the accumulation of those on the left side (D, E). These



B) A, B, C …

C) $\sum_{i=A}^{H} i \quad \sum_{i=F}^{H} i \quad \sum_{i=D}^{E} i$

D) $\frac{D}{E} \quad \frac{G}{H}$

E) $\frac{\sum_{i=D}^{E} i}{\sum_{i=C}^{H} i}$

Figure 3-1. Metabolomics Variables for QTL Mapping.

A. A hypothetical biosynthetic pathway is shown. The letters refer to the individual compounds. The numbers refer to the enzymes. Enzyme 6a and 6b are two different alleles of the same enzyme that lead to two different compounds. Arrows represent the direction of the biochemical reaction.
B. The first variable level is the individual compounds.
C. The second variable level is the broad summation meant to represent different branches of the pathway. i = the amount of specific compounds.
D. The third variable level is the ratio of two related compounds that may provide insight into particular enzymatic processes.
E. The final variable level is the ratio of different biosynthetic branches that may provide insight into more global regulation. i = the amount of specific compounds.

will identify a subset of common QTL as well as unique QTL. For example, it is possible to have a locus that has a 5% effect across the entire pathway. This

effect would most likely not be identified as QTL for any of the individual compounds but due to the smoothing impact of summing all of the compounds this effect may be seen at the pathway level. In addition to summations, ratios are other potential variables derived from a metabolomics data set (Figure 3-1D and E). These can allow the investigator to identify loci controlling regulation at specific branch points. For instance, the equation in Figure 3-1E measures the level of D and E with regards to all compounds produced from C. This would test for the presence of loci that impact the decision to go from compound C to either D or F. However, there are some statistical difficulties introduced in the use of ratio statistics that will be discussed later. When guided by known or predicted metabolic linkages, ratios, and summations provide powerful tools at querying the population for loci regulating whole branches or branch points in metabolic networks.

3.3 Should I worry about normality?

The final aspect before proceeding with the QTL analysis is data preparation. There is the underlying assumption that biological variables/traits should show a parametric distribution. This, however, presumes that the true biological distribution is in fact parametric and the skewing was technically introduced *via* the measurement. In metabolomics, this may not be the case especially in secondary metabolism (Kliebenstein et al., 2001a; McMullen et al., 1998; Byrne et al., 1996; Kliebenstein et al., 2001c; Yencho et al., 1998). In Figure 3-2A and B, the parents of a recombinant inbred mapping population differ in their capacity to make specific compounds due to enzymatic polymorphisms. Parent 1 contains a null allele of enzyme 5 but a hidden "a" allele at enzyme 6 and thereby does not accumulate compounds F, G, or H. Parent 2, however, contains a functional enzyme 5 allele but only the "b" allele of enzyme 6, leading to the accumulation of F and G (Figure 3-2A). When these two parents are mated, the recombinant inbred progeny will represent a mixture of parental genotypes and two recombinant genotypes, recombinant 1 will phenotypically look like parent 1 due to the enzyme 5 null allele while recombinant 2 will be a transgressive segregant producing F and H due to the "a" allele at enzyme 6 (Figure 3-2B). When the accumulation of either H or G in the progeny is plotted on a histogram, it will be a bimodal distribution due to the epistatic interaction between variation at enzymes 5 and 6 in controlling (Figure 3-2C). Normalization would destroy the information about both enzymes 5 and 6. The requirement for parametric distributions is a result of the QTL analysis algorithms. Most algorithms can handle skewed parametrics without normalization by using the bootstrapping methodology to empirically determine the significance threshold. Bimodal and true non-parametric distributions should instead be handled using non-parametric QTL

analysis techniques to obtain the maximal information (Diao et al., 2004; Kruglyak and Lander, 1995).

Metabolic pathway variation can generate other non-parametric distributions *via* transgressive segregation. The above epistasis example can generate non-parametric distributions if the enzyme 5 null allele hides functional enzymes such as 6b (Figure 3-2A). Another way for these non-parametric distributions to occur is when the compound is present in levels near the level of detection. QTL segregation can generate lines with undetectable levels while other lines are readily measurable. A common impulse in these situations is to take the undetectable lines and record them as no data/measurement when it is actually valid to assume that they are less than the other lines. By recording these lines as no measurement, the researcher is lowering the QTL detection power by diminishing the number of lines available for QTL analysis. A potential remedy is to give all undetectable lines a value equal to the detection threshold for the compound in question. This allows the researcher to include the fact that these lines are lower than the rest in the QTL analysis. However, if a significant number of lines are below the detection threshold, this may create a skewed parametric or non-parametric distribution. The skewed parametrics can be handled by the bootstrapping methodology as described. There are algorithms to handle the non-parametric distributions but they are not typically included into the common QTL mapping packages (Diao et al., 2004).

4 WHAT WILL I FIND IN THE QTLS?

Upon generating the metabolomics data and variables for QTL mapping there are numerous software options available to map QTL that are discussed elsewhere (Basten et al., 1999). These generally rely on the same composite or multiple interval mapping algorithms (Doerge and Churchill, 1996; Haley and Knott, 1992; Lander and Botstein, 1989; Zeng, 1994). Most programs, however, were not made to handle or present the massive number of traits generated in a standard genetical genomics experiment and thereby need to be modified to handle this data set. Once these hurdles are overcome and a QTL map is in hand for each trait, there are numerous questions to ask of the data. These include the size and number of QTLs for each trait, are the QTLs for different traits co-localized and is this because of a common polymorphism, as well as what is the level of epistasis and transgressive segregation in the population. I will briefly describe below what may be expected for each question.

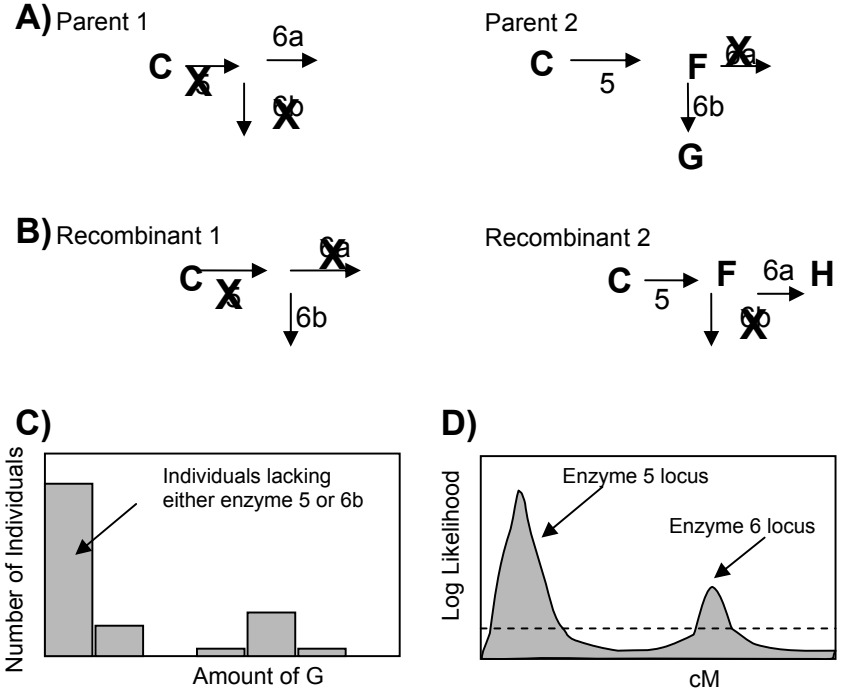


Figure 3-2. The Control of Epistasis and Transgressive Segregation by Enzymatic Variation.

A. The genotype and chemical composition of the parents is shown. The letters refer to the compounds present in each parent. The numbers refer to the enzymes. Enzyme 6a and 6b are two different alleles of the same enzyme that lead to two different compounds. Arrows represent the direction of the biochemical reaction. The X's indicate the presence of non-functional alleles for each enzyme.
B. The genotype and chemical composition of the recombinant individuals is shown. The letters refer to the compounds present in each genotypic class. The numbers refer to the enzymes. The X's indicate the presence of non-functional alleles for each enzyme.
C. The distribution of compound G's accumulation in the RIL population generated from crossing Parent 1 × Parent 2.
D. The QTL map generated for the accumulation of compound G in the RIL population generated from crossing Parent 1 × Parent 2.

4.1 QTL number and phenotypic effect

Recent analysis of a small yeast mapping population with 1 × replication *via* microarray has allowed a glimpse at what may be expected from a metabolomics genetical genomics experiment. This analysis found that most traits required at least 5 QTL's to partially explain the variation (Brem and Kruglyak, 2005). This experiment, however, was limited to a small number of lines with 1 × replication and as such, the analysis was limited to those

genes with at least 69% heritability. Nevertheless, it shows that most variable traits are under highly complex genetic regulation (Brem and Kruglyak, 2005). Metabolomics of higher-throughput and financial scale will allow for these experiments to be conducted with greater power, and therefore, to detect small to medium effect QTL. Thus, one could readily expect that the microarray indication is only the iceberg's tip.

One caveat should be made to the interpretation of both QTL number and phenotypic effect. Both the power to detect a QTL and the estimate of its phenotypic effect are dependent upon the populations' background variation. There could be epistatic interactions with other loci in one population that are not present in another. Alternatively, if the QTL is the only locus impacting the trait in one population, it will have a large phenotypic effect, whereas if the QTL is one of many in another population, it may have a smaller phenotypic effect. Thus, it should not be expected that a QTL or its phenotypic effect would be identical amongst all populations in which it is variable. An excellent example of this is shown in a paper investigating the quantitative inheritance of chlorogenic acid and flavones in three different maize populations. These populations were chosen as they are a pyramid such that all pair wise crosses of three different inbred lines were investigated (Bushman et al., 2002). While numerous QTLs were identified in more than one population, their significance and phenotypic effect were dependent upon the population studied. For example, one QTL was identified in two populations and controlled 34% of the trait variance in one population but only 8% in another (Bushman et al., 2002). Thus, the phenotypic effect and power to detect a QTL is relational and not absolute when comparing populations (Mackay, 2001; Lynch and Walsh, 1998).

4.2 QTL proximity – causality or proximity?

After the QTLs for each trait have been identified and surveyed, the next goal is to identify those QTL that pleiotropically impact different metabolites and as such may have global metabolic impacts. There are several major difficulties in this analysis. The first deals with validating if overlapping QTL for two different traits are caused by the same locus or closely linked loci. There are two possible techniques to try and differentiate between these two distinct possibilities. The first is a statistical approach to testing the possibility that the two QTL positions overlap by chance and hence are probably due to closely linked loci (Lebreton et al., 1998; Varona et al., 2004). It is possible to take the QTL models for each trait, fix the position and effect of the non-overlapping QTL as well as the effect of the overlapping QTL. Then randomly modify the position of the overlapping QTL from the same genetic position to gradually larger unlinked distances. At each step, use every lines genotype and the QTL model to predict all of

lines trait value. Then test the predictive strength of the model at each distance from identical to unlinked position and identify the genetic distance that maximizes the predictive power of both traits model (Lebreton et al., 1998; Varona et al., 2004). This would indicate whether the overlap is due to closely linked loci or a single locus.

Even if there is statistical support for a single locus, it is still possible that the overlap is due to two extremely tightly linked loci. The only way to validate the single pleiotropic QTL hypothesis is to clone the underlying molecular polymorphism and confirm that it impacts the expected traits. This requires fine-scale recombinational mapping in conjunction with some form of transgenic confirmation of the phenotypic effect (Mackay, 2001). Thus, once a pleiotropic region is identified, there still remains significant work to validate the pleiotropic QTL hypothesis. A further potential difficulty with highly pleiotropic QTL is to understand the mechanism by which it works and differentiate between direct and indirect effects. For example, in Arabidopsis, the ERECTA locus impacts leaf morphology, root architecture, floral shape and size, silique shape and size, pathogen resistance and numerous other traits (Qi et al., 2004; Xu et al., 2003; Godiard et al., 2003; Shpak et al., 2003; Douglas et al., 2002). However, even though the gene is a known receptor kinase with global impact, little is known about what the primary and secondary impacts are and how it controls these traits. Thus, highly pleiotropic QTL may not be easily interpreted panaceas of biological information.

4.3 Epistasis and transgressive segregation

Variation in biosynthetic pathways can easily form epistatic interactions measurable in metabolomics QTL mapping projects. One potential epistasis interaction is when variation at a preceding enzymatic step controls the production of another variable enzymes substrate (Figure 3-2A). In the example shown, functional variation at enzyme 5 determines whether the functional variation at enzyme 6a is seen. Thus, the accumulation of compound G in the population will form a bimodal distribution where the low accumulating lines have either a non-functional enzyme 5 or enzyme 6a (Figure 3-2C). Only those lines with functional enzymes 5 and 6a will accumulate compound G. When the level of compound G is used for QTL mapping it will identify at least two locations that epistatically interact in controlling the level of compound G, enzyme 5 and enzyme 6a, (Figure 3-2E). Other ways in which epistatic interactions may occur in biosynthetic pathways is if two proteins physically interact such that the variants from each parent prefer interacting with each other such as might occur in metabolic channeling. This will lead to any recombinant progeny between the loci having lower efficiencies and less compound accumulation. Taken

together, this suggests that epistatic interactions should be expected in metabolomics QTL mapping projects. It will be interesting to compare microarray and metabolomics estimates on epistatic interaction frequencies to see if metabolism is more prone to such interactions than gene expression.

In addition to epistasis, biosynthetic pathways readily generate transgressive segregation in both compound structure and amount. The easiest transgressive segregation to visualize is that impacting structure. In the example shown, the knockout in enzyme 5 hides the presence of the enzyme 6b allele. When recombination shuffles together a functional enzyme 5 and enzyme 6b, compound H is produced (Figure 3-2A and B). Compound H was not produced in either the parental genotype and is thus the product of transgressive segregation. This form of segregation has been identified in Plant Secondary Metabolite QTL projects (Kliebenstein et al., 2002a; Kliebenstein et al., 2001a). In addition to structural transgressive segregation, there is also the likelihood of transgressive segregation in compound amounts. If, for example, one parent has a bottleneck at one step in a biosynthetic pathway while the other is bottlenecked at a different step, segregation will produce lines that have both bottlenecks and compound levels lower than either parent while other lines will have neither bottleneck nor compound levels higher than both parents (Figure 3-3). This also illustrates how two parents can be indistinguishable *via* compound accumulation yet the progeny have highly variable levels. Both forms of transgressive segregation have been readily found in metabolite profiling experiments and should be expected in metabolomics QTL mapping (Kliebenstein et al., 2002a; Kliebenstein et al., 2001a). Further, over half of the highly heritable transcripts in yeast showed evidence of transgressive segregation, again supporting the idea that this will be a common hallmark of genetical genomics experiments (Brem and Kruglyak, 2005).

5 SUMMARY

Combining genomics technologies with segregating populations is becoming an area of increasing interest. The first experiments in this area were conducted with microarrays but it is likely that metabolomics due to cost and throughput advantages will become the “omics” platform of choice for genetical genomics. The hope in these experiments is to combine the massive parallel capacity albeit still reductionist “omics” technologies in a systems biological approach to understand why two organisms are different. In the process, fundamental aspects of biology may be illuminated. However, these same projects and data sets can be used to address base issues of quantitative genetics such as; How many loci control each trait?

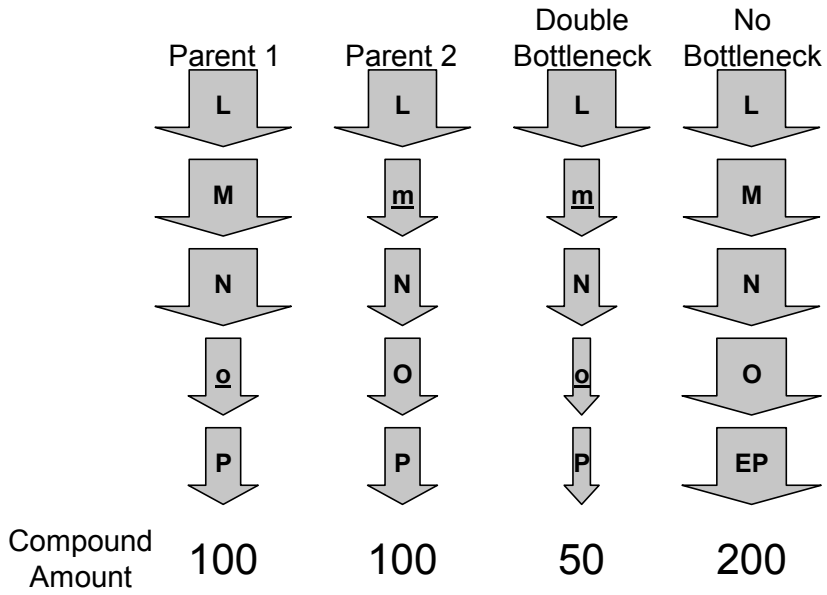


Figure 3-3. Transgressive Segregation Produced by Bottlenecks.
The enzymatic flux for two parents with different bottlenecks is shown. The specific bottlenecks are marked as small underlined letters with diminished arrow size representing the decreased flux. The two potential recombinant inbred line recombinant progeny are shown. The number at the bottom of each biosynthetic pathway shows the relative amount of the final product produced in each line.

What is the basis of the control, either additive or epistatic? Is there directionality in the parents with regards to the QTLs effect? Thus, it may be possible using QTL mapping and metabolomics to both understand key aspects of metabolic relationships as well as fundamental questions of quantitative genetics.

REFERENCES

Basten, C.J., Weir, B.S., and Zeng, Z.-B., 1999, *QTL Cartographer, Version 1.13*. Department of Statistics, North Carolina State University, Raleigh, NC.

Bino, R.J., et al., 2004, Potential of metabolomics as a functional genomics tool, *Trends In Plant Science* **9**(9):418-425.

Brem, R.B., and Kruglyak, L., 2005, The landscape of genetic complexity across 5,700 gene expression traits in yeast, *Proc. Natl. Acad. Sci. U S A*. **www.pnas.org/cgi/doi/10.1073/pnas.0408709102**(on-line).

Brem, R.B., et al., 2002, Genetic dissection of transcriptional regulation in budding yeast, *Science* **296**(5568):752-755.

Bushman, B.S., et al., 2002, Two loci exert major effects on chlorogenic acid synthesis in maize silks, *Crop Science* **42**(5):1669-1678.

Byrne, P.F., et al., 1996, Quantitative trait loci and metabolic pathways: Genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **93**(17):8820-8825.

Denby, K.J., Kumar, P., and Kliebenstein, D.J., 2004, Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana, *Plant Journal* **38**(3):473-486.

Diao, G.Q., Lin, D.Y, and Zou, F., 2004, Mapping quantitative trait loci with censored observations, *Genetics* **168**(3):1689-1698.

Doerge, R.W., and Churchill, G.A., 1996, Permutation tests for multiple loci affecting a quantitative character, *Genetics* **142**(1):285-94.

Douglas, S.J., et al., 2002, KNAT1 and ERECTA regulate inflorescence architecture in Arabidopsis, *Plant Cell* **14**(3): p. 547-558.

Fiehn, O., 2001, Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic networks, *Comparative and Functional Genomics* **2**(3):155-168.

Fiehn, O., et al., 2000, Metabolite profiling for plant functional genomics, *Nature Biotechnology* **18**(11):1157-1161.

Flint, J., and Mott, R., 2001, Finding the molecular basis of quantitative traits: Successes and pitfalls, *Nature Reviews Genetics* **2**(6):437-445.

Godiard, L., et al., 2003, ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt, *Plant Journal* **36**(3):353-365.

Hall, R., et al., 2002, Plant metabolomics: The missing link in functional genomics strategies, *Plant Cell* **14**(7):1437-1440.

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

Jansen, R.C., and Nap, J.P., 2001, Genetical genomics: the added value from segregation, *Trends In Genetics* **17**(7):388-391.

Kliebenstein, D.J., Figuth, A., and Mitchell-Olds, T., 2002a, Genetic architecture of plastic methyl jasmonate responses in Arabidopsis thaliana, *Genetics* **161**(4):1685-1696.

Kliebenstein, D.J., Pedersen, D., and Mitchell-Olds, T., 2002b, Comparative analysis of insect resistance QTL and QTL controlling the myrosinase/glucosinolate system in Arabidopsis thaliana, *Genetics* **161**(1):325-332.

Kliebenstein, D.J., Gershenzon, J. and Mitchell-Olds, T., 2001a, Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in Arabidopsis thaliana leaves and seeds*, Genetics* **159**(1):359-370.

Kliebenstein, D.J., et al., 2001b, Genetic control of natural variation in Arabidopsis thaliana glucosinolate accumulation, *Plant Phys.* **126**(2):811-825.

Kliebenstein, D., 2001c, et al., Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in Arabidopsis thaliana, *Plant Cell* **13**:681-693.

Kruglyak, L., and Lander, E.S, 1995, A Nonparametric Approach For Mapping Quantitative Trait Loci., *Genetics* **139**(3):1421-1428.

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

Lebreton, C.H., et al., 1998, A nonparametric bootstrap method for testing close linkage vs. pleiotrophy of coincident quantitative trait loci, *Genetics* **150**(2):931-943.

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

Mackay, T.F.C., 2001, The genetic architecture of quantitative traits, *Annual Review Of Genetics* **35**:303-339.

Mauricio, R., 2001, Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology, *Nature Reviews Genetics* **2**(5):370-381.

McMullen, M.D., et al., 1998, Quantitative trait loci and metabolic pathways, *Proceedings Of The National Academy Of Sciences Of The United States Of America* **95**(5):1996-2000.

Monforte, A.J., et al., 2001, Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization, *Theoretical And Applied Genetics* **102**(4):572-590.

Morley, M., et al., 2004, Genetic analysis of genome-wide variation in human gene expression, *Nature* **430**(7001):743-747.

Qi, Y.P., et al., 2004, ERECTA is required for protection against heat-stress in the AS1/AS2 pathway to regulate adaxial-abaxial leaf polarity in Arabidopsis, *Planta* **219**(2):270-276.

Santos, C.A.F., and Simon, P.W., 2002, QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots, *Molecular Genetics and Genomics* **268**(1):122-129.

Sax, K., 1923, The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris, *Genetics* **8**:552-60.

Schadt, E.E., et al., 2003, Genetics of gene expression surveyed in maize, mouse and man, *Nature* **422**(6929):297-302.

Shpak, E.D., Lakeman, M.B., and Torii, K.U., 2003, Dominant-negative receptor uncovers redundancy in the Arabidopsis ERECTA leucine-rich repeat receptor-like kinase signaling pathway that regulates organ shape, *Plant Cell* **15**(5):1095-1110.

Steuer, R., et al., 2003, Interpreting correlations in metabolomic networks, *Biochemical Society Transactions* **31**:1476-1478.

Thorup, T.A., et al., 2000, Candidate gene analysis of organ pigmentation loci in the Solanaceae, *Proceedings of the National Academy of Sciences of the United States of America* **97**(21):11192-11197.

Tolstikov, V.V., 2003, et al., Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics, *Analytical Chemistry* **75**(23):6737-6740.

Varona, L., et al., 2004, Derivation of a Bayes factor to distinguish between linked or pleiotropic quantitative trait loci, *Genetics* **166**(2):1025-1035.

Weckwerth, W., et al., 2004, Differential metabolic networks unravel the effects of silent plant phenotypes, *Proceedings of the National Academy of Sciences of the United States of America* **101**(20):7809-7814.

Xu, L., et al., 2003, Novel as1 and as2 defects in leaf adaxial-abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity, *Development* **130**(17):4097-4107.

Yencho, G.C., et al., 1998, QTL mapping of foliar glycoalkaloid aglycones in Solanum tuberosum x S-berthaultii potato progenies: quantitative variation and plant secondary metabolism, *Theoretical And Applied Genetics* **97**(4):563-574.

Yvert, G., et al., 2003, Transacting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors, *Nature Genetics* **35**(1):57-64.

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

.