THE GENETIC BASIS OF VARIATION IN A POLYGENIC CHARACTER:
BRISTLE NUMBER IN DROSOPHILA MELANOGASTER

By

ANTHONY DOUGLAS LONG, B.A.&Sc., M.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

(c) Copyright by Anthony Douglas Long, September 1994
THE GENETICS OF BRISTLE VARIATION IN DROSOPHILA
DOCTOR OF PHILOSOPHY (1994)  McMaster University
(Biology)  Hamilton, Ontario

TITLE: The Genetic Basis Of Variation in a Polygenic Character: Bristle Number in *Drosophila melanogaster*

AUTHOR: Anthony Douglas Long,  B.A.&Sc.  (McMaster University)
         M.Sc.  (McMaster University)

SUPERVISOR: Professor R. S. Singh

NUMBER OF PAGES: xii, 153
Abstract

Quantitative traits are important in medicine, evolution, and plant and animal breeding, yet very little is known about the nature of standing variation at the loci that affect quantitative traits. This thesis describes experiments designed to determine the nature of standing genetic variation for a model trait; bristle number in *Drosophila melanogaster*. The first experiment mapped factors responsible for response to short term selection on abdominal bristle number. Eight factors of large effect which accounted for much of the difference between the parental lines were mapped to the X and third chromosomes. Sex specific and epistatic effects of the same order of magnitude as the allelic effects were associated with the mapped factors. In addition, factors mapped to the approximate positions of likely candidate loci (*ASC, bb, emc, h, mab, Dl*, and *E(spl)*), previously characterized by mutations with large effect on bristle number. In a second experiment restriction variation among 47 naturally occurring second chromosomes was associated with variation in abdominal and sternopleural bristle number. Unlike previous studies, the molecular variants associated with phenotypic variation were not large insertion and deletions, but intermediate frequency restriction site variants. A new permutation test based statistical method was developed in this work in order to assess the significance of the observed associations between molecular and phenotypic variation. The experiments described in this thesis do not support the infinitesimal model of quantitative variation, and instead suggest that a great deal of quantitative genetic variation in nature may be due to alleles of large effect at candidate loci that are at intermediate frequency.
Preface

A central tenet of the modern synthesis of Darwinism is that the forces responsible for differences between species or taxa are the same as those acting within species to create change. That is, the same forces of mutation, gene flow, selection, and random genetic drift act on genetic variation to create differences between both populations and taxa. Implicit in this theory is the idea that genetic variation within populations is converted over time to genetic differences between population or species. This idea is very appealing, as biologists generally agree that the most important differences between species are quantitative in nature, and that there appears to be ample genetic variation for almost any measurable quantitative trait (i.e., almost all traits will respond to artificial selection). The largest problem with this concept is that there is no empirical evidence that standing genetic variation and genetic differences between species are one and the same.

The experiments described in chapters 2 and 3 of this thesis are part of a set of experiments designed to answer the question: What is the nature of genetic variation affecting quantitative traits? The work takes two complementary approaches. The first involves mapping the genetic differences responsible for response to short term artificial selection. It is believed that short term selection experiments, unlike long term selection experiments, are likely to fix segregating variants as opposed to newly arising mutants. The second approach involves associating molecular variation at candidate loci with phenotypic differences between chromosomes isolated from the wild. These experiments are necessarily of a large scale and involve a number of people and laboratories (T.F.C. Mackay's laboratory at N.C.S.U. and C.H. Langley's laboratory at U.C.D.). The experiments designed to answer the above question involve technically difficult crossing schemes and use molecular techniques not readily available even a decade ago. I have been
involved in many aspects of this project from near its beginning, below I wish to describe my contribution to parts of the project described in this thesis.

Prior to my arrival on the Quantitative Trait Locus (QTL) mapping project, the selection experiments had been completed and some of the Recombinant Isogenic chromosome (RI) stock construction started with the intent to map the QTLs responsible for the divergence. My role in Davis was to carry out much of the cytogenetic analysis (via in situ hybridization) of the RI lines, scoring of the roo insertion patterns in the RI lines (with the assistance of S. Mullaney (an undergraduate) for the RI X lines), and organization of the data. Simultaneously the RI lines were scored for bristle number variation at N.C.S.U. Once both the phenotypic and genotypic data had been collected I was responsible for mapping the factors causing the differences between the selected lines. In order to efficiently accomplish this I had to modify the algorithms of S. Zeng to the nested design of our study. Previously available methods for QTL mapping were inappropriate for our data set (which had very few lines but replicate genotypes). This may represent one of the first successful applications of this new algorithm to real data and will represent one of the most powerful QTL mapping experiments in animals to date. Finally, I was largely responsible for the interpretation of this data and writing of the manuscript. In my interpretation of the data I made some observations which were not previously noted in QTL mapping experiments. First, the positions of mapped factors coincided with the positions of candidate loci identified by laboratory mutants of large effect. Although mapped factors coincided with candidate loci in the past, the significance of this observation was not generally appreciated. Secondly, this study noticed sex specific allelic effects and epistatic effects of the same order of magnitude as the allelic effects themselves. These observations have not been made in the past, as earlier approaches were biased against the detection of epistasis and confounded sex differences.
Chapter three of this thesis revolves around the significance of observed associations between polymorphic restriction sites at *scabrous* (a candidate gene) and bristle number variation among a series of second chromosomes extracted from the wild. This is the first such study in a series (i.e., *emc*, *h*, *DI*, *E(spl)*, *da*, etc. to follow) in which I am involved. Studies which associate molecular variation at candidate loci with phenotypic variation in a quantitative trait are important as they implicate the candidate locus as harbouring allelic variants affecting the trait and they provide a rough estimate of the frequency of such variants in the population being surveyed. QTL mapping studies accomplish neither of these objectives. C. Lai is the principle author of the *scabrous* manuscript, and as such collected, and analyzed much of the data. An unforeseen problem with the data though, is that it involved a large number of F tests on sites that were largely independent of one another. Thus, unlike previous studies a single test on haplotypes could not be carried out, and although a number of the molecular marker sites were associated with bristle differences at a marginal significance level, a correction for multiple tests (i.e., the Bonferroni test) resulted in no sites being globally significant. My role was to develop a statistical method which would allow analysis of the data and assess its significance. To this end a new permutation test was created and a computer program written by me which would implement this test and assess the significance of associations between restrictions sites and phenotypic data (see Appendix A). This program has a number of features which will be used in future studies of this sort. Without the permutation test employed in this program, the results of this study would not have reached statistical significance and we would not have been able to conclude that *scabrous* harbours alleles of large effect at intermediate frequency affecting bristle number variation. Thus, although not a first author on this work I feel I had a very definable, critical, and substantial role in completing this work.
I would like to thank the many individuals to whom I am indebted. Without their support and insight this work would not exist. Jennifer MacDonald was patient in teaching me how to manipulate DNA in an eppindorf tube instead of in a computer. Chaoqiang Lai was also able to teach me a great deal, and has served as a valuable colleague, and friend throughout the course of this work. Lai and I have had many a long discussion which ended in us knowing the answer to something -- and then never really being able to define who arrived at it. This is what science is about to me. This work has also benefited from the Langley laboratory environment, both the permanent fixtures and the visitors have created endless opportunity for discussion and intellectual advancement. I must also thank my friends in the Genetics Graduate Group at U. C. Davis, especially Chris Waters, Craig Martens, and Scott Erdman, who adopted me as one of the GGG and helped to broaden my scientific interests.

Many call the completion of their dissertation a transition, mine has been more of a transportation. 'Commuting' between Davis and McMaster during my dissertation research has been trying, but rewarding. Nancy Carter and Walter Scott have provided logistical support and been my close friends during my stays in Canada. Pat Hayward has been invaluable in helping me to schedule meetings and fill out the right forms. This is a daunting enough task when I am in Canada, somehow Pat has helped me carry it out from another country. Dick Morton and Rama Singh have always been my friends and intellectual companions while in Canada, I hope they have benefited as much from our 'coffee talks' as I have. Rama was instrumental in allowing me to pursue my research interests unhindered by institutional boundaries, this selfless gesture should not go unnoticed. Finally, I am indebted to my family, who provided unquestioning support, in my strange choice of professions.
Shao-Bang Zeng has always been an exciting colleague, whenever I have the fortune to meet with him. His discussions with me and the sharing of his ideas on QTL mapping have greatly improved this thesis. Trudy Mackay has been an invaluable colleague in her role as my third supervisor. I have learned a great deal from her through our electronic interactions. I am especially indebted to Chuck Langley. He has consistently managed to provide research environment that is both exciting and non-restrictive. More importantly Chuck has served as a constant reminder that Quantitative Genetics is not spelled Quantitative Genetics, something often forgotten in our field. This lesson will not be easily forgotten.
<table>
<thead>
<tr>
<th>Chapter 1 (Introduction)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1 The problem</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 The origins of quantitative genetics</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Problems with the infinitesimal model</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 The infinitesimal model</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Continuous F2 distributions are consistent with a few loci of large effect</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Realized versus estimated values for heritability</td>
<td>6</td>
</tr>
<tr>
<td>1.2.4 Decomposition of variance and underlying gene action</td>
<td>7</td>
</tr>
<tr>
<td>1.2.6 Results from artificial selection experiments</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Theoretical extensions of the infinitesimal model</td>
<td>10</td>
</tr>
<tr>
<td>1.3.1 Maintenance of genetic variation</td>
<td>10</td>
</tr>
<tr>
<td>1.3.2 Response to artificial selection</td>
<td>12</td>
</tr>
<tr>
<td>1.3.3 Long-term response to selection</td>
<td>13</td>
</tr>
<tr>
<td>1.4 How to test the assumptions</td>
<td>13</td>
</tr>
<tr>
<td>1.4.1 What is the architecture of quantitative traits</td>
<td>13</td>
</tr>
<tr>
<td>1.4.2 Testing the assumptions</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 High Resolution mapping of genetic factors affecting abdominal bristle number in <em>Drosophila melanogaster</em> (title page)</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Abstract</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Introduction</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Materials and methods</td>
<td>19</td>
</tr>
<tr>
<td>2.5 Results</td>
<td>22</td>
</tr>
<tr>
<td>2.6 Discussion</td>
<td>30</td>
</tr>
<tr>
<td>2.7 References</td>
<td>48</td>
</tr>
<tr>
<td>2.8 Appendix A</td>
<td>56</td>
</tr>
<tr>
<td>2.9 Appendix B</td>
<td>63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Naturally occurring variation in bristle number associated with DNA sequence polymorphisms at the <em>scabrous</em> locus of <em>Drosophila melanogaster</em> (title page)</td>
<td>65</td>
</tr>
<tr>
<td>3.2 References and notes</td>
<td>79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4 (Discussion)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 QTL mapping</td>
<td>85</td>
</tr>
<tr>
<td>4.1.1 Methodology</td>
<td>85</td>
</tr>
<tr>
<td>4.1.2 Estimates of effects associated with QTLs</td>
<td>86</td>
</tr>
<tr>
<td>4.1.3 Inferences about the base population</td>
<td>87</td>
</tr>
<tr>
<td>4.1.4 Positional cloning of QTLs</td>
<td>88</td>
</tr>
<tr>
<td>4.2 Candidate genes</td>
<td>88</td>
</tr>
<tr>
<td>4.2.1 Systems without well characterized mutants</td>
<td>89</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Quantitative complementation testing</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Gene transformation</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Associating molecular variation with phenotypic variation</td>
</tr>
<tr>
<td>4.3</td>
<td>Quantitative trait nucleotides</td>
</tr>
<tr>
<td>4.4</td>
<td>Progress in identifying candidate genes</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Candidate genes in non-Drosophilids</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Candidate genes in Drosophila</td>
</tr>
<tr>
<td>4.4.2.1</td>
<td>Candidate genes for bristle number</td>
</tr>
<tr>
<td>4.4.2.2</td>
<td>Defining equivalence groups</td>
</tr>
<tr>
<td>4.4.2.3</td>
<td>Lateral inhibition</td>
</tr>
<tr>
<td>4.4.2.4</td>
<td>The daughters of BMC</td>
</tr>
<tr>
<td>4.4.2.5</td>
<td>The nature of candidate loci and the nature of quantitative variation</td>
</tr>
</tbody>
</table>

Chapter 5 (Conclusions)  

Appendix A  

A1   Listing of scab.c  
A2   A sample run of scab.c  

Bibliography  

140
List of Tables

Chapter 1

1.1 Agreement between realized and estimated heritabilities 6
1.2 Hypothetical genic values for two loci with complete dominance and epistasis 8

Chapter 2

2.1 Phenotypic values of the parental lines used to generate the recombinant isogenic lines 33
2.2 Generation means analysis of the abdominal (AB) and sternopleural (ST) bristle number of isogenic abdominal bristle selection sublines 34
2.3 Estimates of the effective number of loci 36
2.4 Estimates of effects associated with mapped factors 42
2.5 Estimates of effects associated with tests of epistasis 45

Chapter 3

3.1 Mean bristle numbers of homozygous second chromosome lines and their molecular variation in the scabrous region. 68
3.2 Polymorphic sites significantly associated with bristle number variation and estimates of their contribution to additive genetic variation in a natural population 71
List of Illustrations

Chapter 2

2.1 Response to 25 generations of divergent selection for abdominal bristle number 31
2.2 The probability that there is no genetic factor in a given interval as a function of the X chromosome map position of the interval midpoint 38
2.3 The probability that there is no genetic factor in a given interval as a function of the map position of the interval midpoint for the chromosome three RI lines 39
2.4 The probability that there is no genetic factor affecting viability relative to TM6B, Tb balancer chromosome in a given interval as a function of the map position of the interval midpoint for the chromosome three RI lines 47

Chapter 3

3.1 Restriction map variation in the scabrous locus region of D. melanogaster among 47 second chromosomes 69
3.2 The difference between line means of chromosome lines containing (+) or lacking (-) a polymorphic site 73
3.3 Association of haplotypes with bristle number variation 76

Chapter 4

4.1 A model for lateral inhibition 96
1.1 INTRODUCTION

1.1.1 The problem: The last 20 years have witnessed an explosion in our understanding of the molecular basis of differences between alleles controlling monogenic traits. In model systems such as yeast, *C. elegans*, Drosophila, mice, and humans literally thousands of genes controlling monogenic traits have been cloned and characterized. These results stand in stark contrast to progress in understanding the molecular genetic basis of polygenic traits. Understanding the genetic basis of variation in continuous characters is pertinent to evolution, animal and plant breeding, and medical genetics where such traits are generally recognized as being of central importance. Knowledge concerning the molecular genetic differences responsible for variation in continuous characters is required in order to assess the applicability of numerous quantitative genetic models to explaining standing variation within populations and the nature of differences between populations or species. In particular, we must understand the architecture of continuous characters at the molecular level in order to determine if standing variation is the raw material for evolution, as the modern synthesis predicts, or if the relationship between standing variation and the differences between species can be uncoupled.

Recent technical advances have resulted in a plethora of neutral, highly polymorphic markers which should be useful in mapping such polygenes as a first step towards their
characterization. In a number of experiments factors have now been mapped which are responsible for quantitative differences between lines, varietals or species (Tanksley 1993). But in moving beyond differences between lines, they are only a few examples of traits for which the genes contributing to standing variation in a continuous morphological or physiological character have been identified. Pristine examples include: GPDH allotypes related to characteristics of flight metabolism in Drosophila (Barnes and Laurie Ahlberg 1986), apoE alleles in humans associated with heart disease risk (Sing et al. 1988), and naturally occurring transposable element insertions in the achaete-scute complex of Drosophila correlated with reductions in bristle number (Mackay and Langley 1990). Advances in unraveling the molecular genetic basis of the development will provide a number of candidate genes which may be important in future work of this sort. But, even with these advances, and the candidate gene approach, I know of no cases in which the molecular difference between alleles causing an observed phenotypic effect in a continuous character is known. This introduction will provide justification for understanding continuous characters at the level of the genetic loci controlling the character by pointing out some limitations of the infinitesimal model of quantitative genetic variation. It will further demonstrate that advances in molecular and developmental biology may open doors which will allow the exploration of continuous variation at the level of the individual genes controlling these traits. There will be particular emphasis on the determination of the basis of variation in bristle number in Drosophila melanogaster. This is the system I have used in my attempts to understand the molecular nature of variation in a quantitative character.

1.1.2 The origins of quantitative genetics: In June of 1916 Fisher's milestone paper, The correlation between relatives on the supposition of Mendelian inheritance, was rejected by the Royal Society of London only to be finally accepted for publication by the Royal Society of Edinburgh two years later (Fisher 1918; Provine
1971). Fisher's paper, although preceded by others with similar ideas of relating Mendelism and Darwinism (Yule 1907; Nilsson-Ehle 1909; East 1910), brought together a number of important ideas. Fisher showed that if continuous traits are composed of a number of Mendelian factors of small, approximately equal, and additive effects, Mendelism was entirely consistent with continuous variation, selection response, and the correlation between relatives commonly observed for such traits. Prior to the publication of Fisher's paper the Biometricians believed that Darwinian evolution acted on continuous characters which were fundamentally different from Mendelian traits - in particular the mode of inheritance of continuous characters was blending. The Mendelians were equally incorrect in thinking that evolution acted only on the 'sports' or large variants commonly analyzed in a Mendelian manner which were likely mutationally limited - with continuous variation being largely inconsequential. Fisher's paper was a synthesis of the ideas of the Biometricians and Mendelians, its appearance marked the beginning of the field of quantitative genetics and the end of the feud between the Biometricians and Mendelians (Provine 1971).

For illustrative purposes the true impact of Fisher's work is overstated above. A great deal of other work certainly contributed to the synthesis of Mendelism and Darwinism early in this century. Nonetheless, it does reinforce the commonly held view that the infinitesimal model united the emerging field of Mendelian genetics and the established Biometrician school. I will advocate an opposing thesis throughout this work. The basis of the infinitesimal model was that a large number of Mendelian factors each of small effect contributed to quantitative variation. This theory was consistent with Mendelian genetics, but it also excluded continuous characters from empirical analyses of the sort which interested the Mendelian geneticists. Since Fisher's paper of 1918 the field of quantitative genetics has almost exclusively attempted to explain phenotypic variation and evolution in
terms of statistical models as opposed to in terms of individually identifiable Mendelian variants (Falconer 1989) - this was exactly the goal of the pre-synthesis Biometricians (Provine 1971)! Thus, as much as the infinitesimal model marked the unification of the Biometricians and the Mendelians, it also marked the beginning of a long divergence between the two groups.

1.2 PROBLEMS WITH THE INFINITESIMAL MODEL

1.2.1 The infinitesimal model: The foundation of the model is that the genetic variance attributable to a single locus can be partitioned into different components, specifically additive and dominance components. As these two components are statistically independent of one another the total variance is simply the sum of the two components, that is \( V_G = V_A + V_D \) where \( V_A = 2pq(a + d(q - p))^2 \) and \( V_D = (2pqd)^2 \), with \( p \) and \( q \) being the frequencies of alternate alleles at the locus under consideration, \( a \) is half the difference between homozygous means, and \( d \) is the dominance deviation of the heterozygote from the average of the two homozygotes (Falconer 1989). This decomposition is statistical, not genetical, in the sense that even for completely recessive gene action the additive component will account for a great deal of the total variation and it is also dependent on allele frequencies in the population. That is, the additive and dominance variance for a trait will be different in two populations with different allele frequencies, even if the gene action is the same in the two populations. The extension to many loci is derived by summing \( V_G \) over all loci contributing to the trait. The central assumption of the infinitesimal model is that quantitative variation is due to many such \( V_G \)'s each of small and approximately equal effect such that the resemblance between relatives (\( X \) and \( Y \)) can be expressed as:

\[
\text{cov}_{X,Y} = rV_A + uV_D + r^2V_{AA} + ruV_{AD} + u^2V_{DD} + ...
\]

(1)

where \( r \) is the probability of individuals \( X \) and \( Y \) having an allele identical by descent, \( u \) is the probability of individuals \( X \) and \( Y \) having a genotype identical by descent, \( V_{AA}, V_{AD}, \)
and $V_{DD}$ are variance components due to two way epistasis between loci, all variance
components are summed over the respective loci contributing to their variance, and (...) represents terms involving higher order epistasis (Crow and Kimura 1970; Falconer 1989). It is important to note that the equation for the additive and dominance variation is a definition so is guaranteed to be true, whereas the extension to the covariance between relative requires the assumptions of the model to be correct - this is an empirical question. If the assumption that a large number of loci (or $V_G$'s) contribute to the total genotypic variance is correct, then as many terms can be estimated in equation (1) as there are covariance between relatives terms estimable in a given breeding design. This is accomplished by equating the observed variance components in a breeding design to 'causal variance components' defined by the expected covariances between relatives (Falconer 1989). In practice all causal components can not be estimated so the higher order terms in (1) are assumed to be negligible. It is important to realize that estimates of the causal components are statistical and they really only tell us about the sums of effects over loci. Such estimates can not tell us how many loci contribute to a trait, the relative effects of the loci contributing to the observed variation, or the nature of gene action at the loci contributing to the trait.

1.2.2 Continuous F2 distributions are consistent with a few loci of large effect: The infinitesimal model is just that - a model. The infinitesimal model appears to have worked well in explaining the major features of response to selection and resemblance between relatives, yet the biological basis of the model is largely untested. Although progress in animal and plant artificial selection experiments is often considered as evidence for the infinitesimal model, these results provide equal support for other models of the genetic basis of quantitative traits. As pointed out by Barton and Turelli (1989), "despite Fisher's reconciliation of Mendelism with the resemblance among relatives (Fisher
1918), the standard equations for the response of phenotypic means to selection are essentially pre-Mendelian". The most commonly cited evidence for the assumption of many alleles of small effect is that the distribution of phenotypes in the F2 of a cross between lines is normally distributed. With heritabilities of 100 percent, just three additive loci contributing to a trait, and samples upwards of 500 individuals, F2 phenotypic distributions can be indistinguishable from a normal distribution (Thoday and Thompson 1976). The paper often cited as the basis for the infinitesimal model (Nilsson-Ehle 1909), observed close to continuous distributions with two to four known Mendelian factors segregating in a cross of wheat, and only further postulated that situations with ten factors segregating may exist (Provine 1971).

1.2.3 Realized versus estimated values for heritability: If the infinitesimal model is correct, then the response to artificial selection should be predictable. In particular, the realized heritability of a trait estimated from the response to artificial selection would be expected to equal the heritability of the same trait estimated in the base population from the resemblance between relatives. This concordance has not generally been observed, as pointed out by Sheridan (1988), whose results are partially reproduced in Table 1. Sheridan showed that the level of agreement between realized and estimated heritabilities (i.e., $100 \times \frac{|\text{Estimated}-\text{Realized}|}{\text{Realized}}$) were actually quite low. In addition, when standard errors were available for both realized and estimated heritabilities,

<table>
<thead>
<tr>
<th>Species</th>
<th>Level of agreement (%)</th>
<th>Total number of comparisons</th>
<th>Tested comparisons p&lt;0.05</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosophila</td>
<td>29</td>
<td>66</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>Other species</td>
<td>29</td>
<td>132</td>
<td>24</td>
<td>66</td>
</tr>
</tbody>
</table>
approximately 25% of the time the estimates significantly differed. Hill and Caballero (1992) point out that although the conclusions that can be drawn from Sheridan's summary are not concrete, the results can not be dismissed. One source of potential discrepancy between realized and estimated heritability is linkage disequilibria between loci which under the infinitesimal model will cause a reduction in realized heritability relative to estimated heritability under intense artificial selection ('Bulmer effect'; Bulmer 1971; Bulmer 1980). Yet even when models incorporating the Bulmer effect are fitted to the data in order to estimate realized heritability there is still a discrepancy between realized and estimated heritability (Beniwal et al. 1992). If selection response is primarily due to a small number of loci of relatively large effect then changes in allele frequency at these loci can significantly alter quantitative genetic parameters in the selected population (Barton and Turelli 1989; Falconer 1989; Hill and Caballero 1992).

1.2.4 Decomposition of variance and underlying gene action:
Quantitative genetic parameters are commonly estimated by equating observed variance components to expected variance components under the infinitesimal model. The expected variance components are derived by decomposing the total variance in a population into variance components attributable to additive genetic, dominant genetic, epistatic genetic, and environmental variation. The additive genetic component of variation most often accounts for much of the genetic variance in any particular experiment. This is taken as evidence that most of the loci effecting quantitative traits are additive both within and between loci (i.e., dominance and epistasis are inconsequential in determining the architecture of quantitative traits) (Falconer 1989). Although the decomposition of variance in this manner is useful in the construction of breeding values (which ultimately determines the response to individual selection), they are very uninformative concerning the genetic details of the loci contributing to the quantitative trait. One reason for this is that in practice
Table 1.2: Hypothetical genic values for two loci with complete dominance and epistasis.

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>Bb</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>bb</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

...all causal components can not be estimated, and as a result they can inflate the estimates of other components. A common example of this is half-sib designs where any additive by additive epistasis increases the estimate of additive variation (Crow and Kimura 1970). A second reason is that the method of variance decomposition which ensures estimates of variance components are independent also results in some variance components not actually estimating the genetic effect they refer to. The dominance variance, for example, does not estimate the variance assuming dominant gene action but the deviation of the heterozygote from the mean of the two homozygotes (i.e., the excess variance after a completely additive model has been fitted). A result is that even in situations in which dominance and epistasis is complete at the genetic level, estimates of additive genetic variation still make up much of the total genetic variation. An example taken from Crow and Kimura (1970; pp. 126) and reproduced in Table 2 will illustrate this point. Table 2 gives the genotypic values for two loci at which all variation is genetic and $p_A = q_B = 0.450$ (such that the frequency of the two phenotypes are equal). From Table 2 it is clear that both loci are completely recessive and epistasis is complete (i.e., aabb = -1 not -2), yet the estimates of the variance components are: $V_A = 58\%$, $V_D = 25\%$, $V_I = 17\%$. Clearly, estimates of variance components with little non-additive variation do not necessarily imply the underlying genetics of the trait is additive in nature.
1.2.5 Estimates of effective number of factors: Estimates of the effective number of factors contributing to a trait have also been used to support the argument that a large number of loci are responsible for quantitative variation for a given trait. Estimates are generally based on the Wright-Castle index, \( N_e = (\text{range})^2 / (8 \times V_A) \), or related variants (Wright 1977), where \( N_e \) is the effective number of factors causing the difference between lines, range is the difference between the parental lines, and \( V_A \) is the additive genetic variance in the base population. Nevertheless, the conclusion after considerable theoretical work and many applications of the index to real data is that "even in the best of circumstances, information from [the index] is very limited and can be misleading" (Zeng et al. 1990; Hill and Caballero 1992). \( N_e \) tends to underestimate \( N \) if allelic effects are unequal, and the estimator has an upper bound close to the haploid number of chromosomal arms (Barton and Turelli 1989). Estimates approaching this upper bound are often cited as supporting the infinitesimal model. Nonetheless it is often overlooked that epistasis or dominance can lead to estimates of \( N_e \) greater than \( N \) (Barton and Turelli 1989). This may be especially pertinent to estimates of the effective number of factors in crosses between highly selected lines where epistasis may be important (e.g., the mapping results of this thesis).

1.2.6 Results from artificial selection experiments: Two arguments based on the theory of limits to selection (Robertson 1960) suggest the number of factors contributing to selection response may not be large. Robertson's theory predicts that the half life (the number of generations required to get half way to a plateau in selection response) of the theoretical maximal response to selection will occur in \( 1.4 \times N_e \) generations under the infinitesimal model (\( N_e \) is now the effective number of breeding individuals). Yet observed half lives are much shorter, often on the order of \( 0.5 \times N_e \) generations (Falconer 1989, pp. 217-22). This suggests the actual number of contributing loci may not
be that large or the distribution of effects is such that a few loci of large effect account for
much of the response. This is a weak argument though, as it is often difficult to define
when the limit has been reached because of new mutations and the generation of linkage
disequilibrium (Hill and Caballero 1992). A related set of experiments 'bottleneck' a
population and then carry out a selection experiment on bottlenecked and non-bottlenecked
lines. If the infinitesimal model is correct or selection response is due to a few loci of large
effect at intermediate frequency in the base population, then bottlenecks should not result in
greatly reduced responses to selection. The observation that bottlenecks only reduce
selection response by about 30% (Robertson 1967; Frankham 1990) does not support the
prediction of the House of Cards model that standing variation is due to genes of large
effect at extreme frequencies (Turelli 1984).

1.3 THEORETICAL EXTENSIONS OF THE INFINITESIMAL MODEL

1.3.1 Maintenance of genetic variation: Lewontin has proposed that the
central question of population genetics is what maintains genetic variation within
populations. This question is perhaps even more pertinent for quantitative traits, where
stabilizing selection on trait means is commonly inferred. One popular explanation for
standing variation is that new variation from mutations may balance the reduction in
variation from stabilizing selection. One class of models assumes that many additive loci
contribute to a trait, allelic effects are normally distributed within and between loci, and
there is no epistasis or dominance (Kimura 1965; Lande 1975). Turelli has often pointed
out the important fact that a normal distribution of breeding values in no way implies that
allelic effects are normally distributed, nor is there any empirical evidence on which this
assumption is based (Turelli 1988). The equilibrium genetic variance under these Gaussian
(or normal) models is \( V_s = \sqrt{2nV_mV_s} \), where \( n \) is the number of loci contributing to the
color, \( V_m \) is the genome wide mutational variance per generation, and \( V_s \) is the variance
of the Gaussian fitness function. This model requires that the effects of new mutations are small relative to standing variation per locus, the number of loci contributing to a character is large, and that all loci equally contribute. An alternate class of models assumes that new mutants are large relative to standing variation, and hence the equilibrium genetic variance is \( V_s = 4n\mu V_s \), where \( \mu \) is the mutation rate per generation (Barton and Turelli 1987).

Although this model is not dependent on the distribution of allelic effects, it does depend on the number of loci contributing to the character and the mutation rate per locus - both parameters have not been empirically determined with a large degree of accuracy. This model predicts that most standing variation will be due to alleles of large effect at extreme frequencies (Turelli 1984).

Models for the maintenance of variation based on single traits are likely unrealistic if the loci contributing to quantitative trait variation affect more than one trait. One way of dealing with this is to consider pleiotropic models in which a number of loci affect a number of characters. The problem with such models is that they quickly become intractable for even the simplest assumptions, they make assumptions about the nature of pleiotropy for which we have very little empirical evidence, and their dynamics are very difficult to predict unless we know the nature of selection acting on all traits (Kondrashov and Turelli 1992). Keightley and Hill (Keightley and Hill 1988; Barton 1990; Keightley and Hill 1990; Kondrashov and Turelli 1992) proposed an eloquent solution to this dilemma with a model which considered the effects of alleles on a trait under consideration and their pleiotropic effects on fitness. Early analyses of this model showed that a great deal of quantitative genetic variation can be maintained if new mutants are allowed that have a large effect on the trait and a very small fitness effect (i.e., the case where the trait is effectively neutral). A different model in which all new mutants are unconditionally deleterious, and fitness is a concave up function of the number of mutants an individual
harbours, showed there will be apparent stabilizing selection on quantitative characters even if there really is none. This model further predicts that the equilibrium genetic variance will be similar to that maintained under earlier House of Cards models and that the alleles contributing to quantitative variation will be rare (Kondrashov and Turelli 1992). The applicability of pleiotropic models, as with earlier models, depends on parameters for which we do not have a good empirical estimates. Ideally estimates of the joint distribution of mutant effects on the trait and fitness would be desirable. The frequency of alleles combined with the above estimates of allelic effects will allow us to determine if alleles at QTL: are rare as the pleiotropic and House of Cards models predict, are of small effect as Gaussian models predict, or of possibly large effect and at intermediate frequency which may imply other forms of balancing selection (Gillespie and Turelli 1989). Tests of competing models without knowledge of the underlying architecture of the traits under consideration are weak as they invariably must make assumptions about gene action, epistasis, distributions of allelic effects, and the frequency and number of alleles contributing to quantitative variation (Barton and Turelli 1989; Kondrashov and Turelli 1992).

1.3.2 Response to artificial selection: The standard response equation of quantitative genetics follows from the definition of heritability and is given as \( R = h^2 S \) (\( R \) is response and \( S \) is selection differential). Although this equation makes no assumptions about the underlying genetics of quantitative characters the properties of this predictive equation over selective episodes of more than a few generations will depend on the underlying genetics of the character. In particular, different underlying genetic architectures will result in changes in the heritability of the trait over time. In general, if allelic effects are not normally distributed the additive variance will change over time even if a large number of loci contribute to the trait (Turelli 1988). Of course if the trait is
determined by a few loci of large effect with possible epistatic relationships, the change in additive variance over the course of a selection experiment can only be understood in terms of the underlying genetics of the character (Crow and Kimura 1970; Falconer 1989). As the predictive equation is rarely accurate over more than a few generations it is apparent that empirical studies of the genetic basis of quantitative traits will help to define the role theory will play in predicting how the additive variance will change over the course of selection experiments. There are clearly many situations where detailed knowledge of the underlying genetics of a character will lead to improved methods of artificial selection. Often these changes will occur through marker assisted selection (Gimelfarb and Lande 1994), changes in breeding structure (such as family selection) to take advantage of non-additive gene action (Falconer 1989), and optimizing selection intensity to maximize gain while minimizing the chance of losing an allele through drift (Robertson 1960).

1.3.3 Long-term response to selection: Lande (1979) popularized the application of the standard response equation (extended to the multivariate case) to evolutionary time scales in order to determine the characters on which selection have acted and the relative intensity of selection over time. Turelli (1988) has been a vocal critic of this approach claiming that the assumption of the constancy of the genetic variance/covariance over long time scales is dubious. If standing variation is due to a small number of loci with unpredictable patterns of dominance and epistasis, the predictive equation can not be extrapolated over more than a few generations at best. Identification and cloning of the actual loci involved in standing variation or response to artificial selection will allow quantitative predictions of short term selective response in terms of the observed genetic parameters. An understanding of the role of that quantitative trait loci play in long term selection will likely only be possible by comparing the DNA sequence of such loci once identified among closely related species (Hudson et al. 1987; McDonald and
Kreitman 1991; Long and Langley 1993), and ultimately manipulative experiments involving gene transformation between species. It is possible that the molecular genetic events responsible for most standing variation are not related to adaptive differences between species at the molecular level.

1.4 HOW TO TEST THE ASSUMPTIONS

1.4.1 What is the architecture of quantitative traits: In earlier sections I have shown that there is not a great deal of empirical support for the infinitesimal model and that extensions of this model depend to some extent on the validity of these assumptions. In particular it will be important to determine the number of loci contributing to variation for a quantitative trait, the relative variation contributed by each locus, the nature of gene action both within (i.e., dominance) and between loci (epistasis), and estimates of effects associated with alleles at QTL on both the trait and fitness. New techniques discussed below will in theory provide a means to estimate some of these parameters with a high degree of confidence (e.g., gene action, numbers, variation attributable to each, and effects on the character), some will much more difficult to estimate but are achievable goals in the near future (e.g., frequencies of alleles at QTL), and some will likely remain impenetrable for some time (e.g., fitness effects of biological importance). If quantitative genetic variation is ultimately due to a large number of loci of small and approximately equal effect, as the infinitesimal model assumes, then quantitative variation may only be explainable in statistical terms. Whereas if quantitative variation is primarily due to a few loci of large effect, then it may be possible to explain the underlying variation in terms of the effects, frequencies, gene action, and epistasis associated with these loci. Eventually it will also be important to consider such estimates separately for the two sexes and assess the reliability of the estimates in different environment the genotypes are likely to encounter. Although such descriptions will, in theory, allow precise
descriptions of responses to short term selection, this is neither practical nor the goal of experiments designed to estimate such parameters. If the nature of standing variation is elucidated for even a few characters it will allow an assessment of the applicability of statistical models to predicting short term and long term responses to artificial and natural selection and the maintenance of quantitative variation. Ultimately, evolution may only be understandable in terms of the underlying loci diverged between populations and species. Given that estimates of the genetic parameters underlying quantitative traits are desirable and feasible, below I describe experimental approaches to their estimation.

1.4.2 Testing the assumptions: It is clear that models attempting to explain both the maintenance of quantitative genetic variation and the response to selection are dependent on genetical assumptions for which there is very little empirical evidence. In order to understand the true nature of quantitative genetic variation we will have to extend the statistical models of the last 70 years to include the nature of the underlying variation at the loci which contribute to continuous variation. Many of the recently developed tools in molecular, cell, and developmental biology may make this feat possible. Mapping these loci is an important first step. QTL mapping experiments will provide initial estimates of the numbers and relative effects of loci contributing to continuous variation, as well as estimates of effects and gene action of alleles at QTL, and potentially the levels of epistasis between QTL. Experiments designed to accomplish this goal for bristles in Drosophila are described in Chapter two of this thesis.

Further questions, such as the frequency of alleles at QTL, and the molecular nature of the variants contributing to quantitative variation may require cloning and characterizing QTL loci at the molecular level. Understanding the molecular nature of quantitative variation will be important if we wish to determine if the types of variants contributing to standing variation are the same as those causing divergence between species. This goal
will be greatly simplified if candidate loci harbour variation which is important for quantitative traits. Thus the next step in characterizing quantitative traits at the molecular level will be to determine if candidate loci are indeed important in standing variation. This has already been accomplished for one locus influencing bristle number in Drosophila (Mackay and Langley 1990), and Chapter three of this thesis describes the association of molecular variation at a second locus, *scabrous*, with standing variation in bristle number.
Chapter 2

High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*

Anthony D. Long†, Susan L. Mullaney*, Lucy A. Reid†, James D. Fry†, Charles H. Langley† and Trudy F. C. Mackay†

This work has been submitted for publication in the journal GENETICS

* Center for Population Biology, Storer Hall, University of California at Davis, Davis, CA 95616.

† Department of Biology, McMaster University, 1280 Main St. W., Hamilton, ON Canada, L8S 4K1.

‡ Department of Genetics, Box 7614, North Carolina State University, Raleigh NC 27695-7614.
Corresponding Author: Anthony D. Long
Center for Population Biology
Storer Hall
University of California at Davis
Davis, California, 95616

Phone: (916) 752-4253
Fax: (916) 752-1449
E-mail: tdlong@ucdavis.edu

Running Head: Mapping Bristle QTLs in Drosophila

Keywords: Drosophila, quantitative genetics, quantitative trait loci, mapping, bristle number, artificial selection, peripheral nervous system
ABSTRACT

Factors responsible for response to selection for abdominal bristle number and correlated responses in sternopleural bristle number were mapped to the X and third chromosome of *Drosophila melanogaster*. Lines divergent for high and low abdominal bristle number were created by 25 generations of artificial selection from a large base population, with an intensity of 25 individuals of each sex selected from 100 individuals of each sex scored per generation. Isogenic chromosome substitution lines in which the high (H) X or third chromosome were placed in an isogenic low (L) background were derived from the selection lines, and 93 recombinant isogenic (RI) HL X and 67 RI Chromosome 3 lines constructed from them. Highly polymorphic, neutral *roo* transposable elements were hybridized *in situ* to the polytene chromosomes of the RI lines to create a set of cytogenetic markers. This techniques yielded a dense map with an average spacing of 4 cM between informative markers. Factors affecting bristle number, and relative viability of the Chromosome 3 RI lines, were mapped using a multiple regression interval mapping approach, conditioning on all markers 10 cM or greater from the tested interval. Two factors with large effects on abdominal bristle number were mapped on the X chromosome and five factors on the third chromosome. Two factors with large effects on sternopleural bristle number were mapped to the X and two to the third chromosome; one of the X and both of the Chromosome 3 factors with sternopleural effects corresponding to those with effects on abdominal bristle number. Two of the Chromosome 3 factors with large effects on abdominal bristle number were also associated with reduced viability. Significant sex-specific effects and epistatic interactions between mapped factors of the same order of magnitude as the additive effects were observed. All factors mapped to the approximate positions of likely candidate loci (*ASC, bb, emc, h, mab, Dl*, and *E(spl)*), previously characterized by mutations with large effects on bristle number.
INTRODUCTION

The implications of the molecular biology revolution have been great, yet at present very little is known about the molecular genetic basis of standing variation in quantitative traits that are important for plant and animal breeding, medicine, and evolutionary theory (WEIR et al. 1988). Estimates of the distribution of effects at quantitative trait loci (QTLs) and the number of loci involved in standing variation and response to selection are essential to theoretical models of evolution, but are generally unknown (BARTON AND TURELLI 1989). Knowledge of the nature of molecular variation, and its relationship to the phenotype of the organism, will be important in understanding molecular evolution at these loci, and phenotypic evolution itself. The first objective of this work was to identify factors that respond to artificial selection and that are likely to be segregating in natural populations, and to estimate the genetic effects associated with these factors. A second objective of this study was to map factors with sufficient accuracy that candidate loci (characterized loci known from laboratory mutants of large effect) might be identified. If response to selection is due to variation at candidate loci the transition from mapping to molecular characterization of quantitative trait loci will be greatly simplified.

The number of bristles on adult Drosophila melanogaster are ideal traits for which to determine the genetic basis of quantitative variation. For more than 40 years bristles have been used in quantitative genetic studies because they are easy to count, and are highly heritable with primarily additive genetic variation (FALCONER 1989). More recently, the development of the peripheral nervous system (i.e., bristles are sensilla) has emerged as a powerful model for understanding the molecular genetic basis of pattern formation and cell signalling (LAWRENCE 1992; CAMPOS-ORTEGA and JAN 1991). Many of the genes involved in bristle patterning and bristle formation have been identified, cloned and
characterized. These genes are candidate loci at which naturally occurring allelic variation contributes to quantitative variation in bristle number and response to selection. The well-developed understanding of the roles these loci play in the development of the nervous system may make it possible to move beyond mapping QTLs to the association of quantitative variation with variation at the molecular level.

SAX (1923) showed that QTLs could be mapped by linkage to visible markers, and variations of this approach have been used to map factors affecting Drosophila bristle number. BREESE and MATHER (1957) used a multiply marked third chromosome and chromosomes selected for high (H) and low (L) abdominal bristle number to construct a series of recombinant chromosomes in which a segment from the H chromosome was introgressed into the L chromosome background. Effects were associated with each of the six chromosomal segments created in this manner. THODAY (1961) combined interval mapping with progeny testing to obtain estimates of map positions and effects of factors affecting sternopleural bristle number. In general, all intervals considered in the experiments using this method contained factors affecting bristle number, but some factors had large effect and contributed most of the variation between the tested and tester chromosomes (GIBSON and THODAY 1962; WOLSTENHOLME and THODAY 1963; THODAY et al. 1964; SPICKETT and THODAY 1966; DAVIES 1971; SHRIMPTON and ROBERTSON 1988a,b).

Previous work mapping bristle QTLs suffers from several problems (THODAY 1979). THODAY's method assumes there are no alleles on the multiply marked tester stock with greater effects on the character than the chromosome from the selection line (i.e., all alleles increasing bristle score are on the high selected chromosome), and that the marker loci do not themselves affect the trait. Failure of these assumptions leads to serious errors including finding loci that do not exist, overestimating the effects associated with
mapped factors, and assigning loci to the incorrect map positions (MCMILLAN and ROBERTSON 1974). Most of the published experiments have used marker chromosomes of unknown background (GIBSON and THODAY 1962; WOLSTENHOLME and THODAY 1963; THODAY et al., 1964; SPICKETT and THODAY 1966; DAVIES 1971), and an allele of \( h \), which itself has an effect on bristle number, was often used in the analysis of third chromosomes. The exception is the analysis of SHRIMPTON and ROBERTSON (1988a,b) in which marker loci were substituted by backcrossing into a chromosome derived from selecting in the opposite direction of the chromosome to be tested. A further problem with THODAY's (1961) method of QTL mapping is that recessive alleles on the selected chromosome will not be detected as heterozygotes against the marker stock. Finally, it is not clear to what extent the conclusion of previous studies, that a few loci of large effect account for most of the variation in bristle number, is applicable to loci segregating in a natural population. The high and low selected strains used in past experiments stemmed either from a small base population with a limited sample of naturally segregating variation (BREASE and MATHER 1957; GIBSON and THODAY 1962; WOLSTENHOLME and THODAY 1963), from long term selection lines in which new mutations contributing to selection response can not be ruled out (BREASE and MATHER 1957; THODAY et al. 1964; SPICKETT and THODAY 1966), or from different populations (SHRIMPTON and ROBERTSON 1988a,b).

We report here the results of an experiment to map factors affecting response to 25 generations of divergent selection for abdominal bristle number from a large base population: most alleles at intermediate frequency that contribute to standing variation in the base population are likely to be at high frequency in the selected lines, and new mutations are not likely to contribute to selection response. We have extended the approach of BREASE and MATHER (1957) to construct recombinant isogenic (RI) chromosomes
from lines selected in opposite directions from the same base population, so factors with recessive effects will be detected. Recombination breakpoints of the RI lines were inferred precisely using the cytological insertion sites of polymorphic roo transposable elements (MEYEROWITZ and HOGNESS 1982; also called B104, SCHERER et al. 1982) as originally suggested by SHRIMPTON and ROBERTSON (1988b). roo insertions are likely to be neutral with respect to bristle number and yield a high density map.

MATERIALS AND METHODS

Selection lines: The base population was a sample of 62 isofemale lines collected from Raleigh Farmer's Market in May, 1988. To equalize the contribution of each isofemale line, four females from line 1 (f1) were crossed to four males of line 2 (σ), 4 σ2 x 4 σ3, ..., 4 σn x 4 σn, in separate vials. After 48 hours, one male and one fertilized female from each cross were placed in a bottle, and discarded after three days. One hundred virgin females and males were collected from the offspring of the initial crosses (G0) and scored for abdominal bristle number (the number of bristles on the most posterior abdominal sternite; segment 6 in females and segment 5 in males). The 25 highest scoring individuals of each sex were selected as parents of the high line, and the 25 lowest as parents of the low line. The high and low lines were maintained each generation by selecting the 25 most extreme individuals of 100 scored of each sex, for a total of 25 generations of selection. At generation 25, the correlated response of sternopleural bristle number (the sum of the numbers of bristles on the left and right sternopleural plates) to selection for abdominal bristle number was measured from a sample of 100 males and 100 females from the two selection lines.

Isogenic sublines: At generation 26 the twenty most extreme males from the high and low lines were used to create isogenic sublines suitable for mapping. All chromosome balancer stocks referred to below are described in LINDSLEY and ZIMM
(1992). The multiple balancer stocks are of P cytotype for the P-M system of hybrid
dysgenesis (KIDWELL et al. 1977). All crosses and rearing were carried out at 25° C on
10 ml of cornmeal-agar-molasses medium in shell vials unless otherwise noted.

Single males from the high and low selection lines (genotype S1/Y; S2i/S2j;
S3i/S3j, where S indicates a high or low selected chromosome; I, 2, and 3 refer to the X,
second and third chromosomes, respectively; and i and j indicate non-isogenic
homologues) were crossed to females of a stock with C(1)DX and T(2,3) ap^x (abbreviated
Xa below) in an inbred Harwich background (G1). At G2 a single male progeny of
genotype S1/Y; Xa/S2i; S3i was crossed to females of the balancer stock Basc; T(2,3)
ap^x/SM5,Cy; TM6B, Tb (abbreviated B; Xa/Cy; Tb below). Male (B/Y; Cy/S2i; Tb/S3i)
and female (B/S1; Cy/S2i; Tb/S3i) progeny carrying isogenic copies of single X, second
and third chromosomes of the selection line were intercrossed (G3), and at G4 S1/Y;
Cy/S2i; Tb/S3i males were mated to their B/S1; Cy/S2i; Tb/S3i sibs. At G5 five males and
five females of each of the resultant eight genotypic classes (less if S2i or S3i harbored
recessive lethals) were scored for abdominal and sternopleural bristle number. Two
isogenic high (IH1, IH2) and low (IL1, IL2) lines with bristle numbers near those of the
mass-mated selection lines at the end of the selection experiment were retained for mapping
the genetic factors contributing to the high-low divergence in bristle number.

Quantitative genetic analysis: Generation means analysis (MATHER and
JINKS 1971) was used to estimate the additive effects and average degree of dominance
for autosomal and X-linked loci contributing to the difference in bristle number between the
pairs of selection lines IH1, IL1 and IH2, IL2. For each pair of high-low isogenic lines,
mean abdominal and sternopleural bristle numbers were estimated from a sample of 60
males and 60 females of the two parental lines, reciprocal F1 and reciprocal F2 crosses of
the parental lines, and the eight backcrosses derived by crossing the reciprocal F1's to the
parent lines in all possible combinations (i.e., HL x H, HL x L, H x HL, L x HL, LH x H, LH x L, H x LH and L x LH, where H (high) and L (low) are the isogenic parental selection lines, and HL and LH are the reciprocal F1's). Line means of the parental strains were estimated from a sample of 20 individuals of each sex in each of three replicate vials, and for all other generation means from a sample of 30 males and females in each of two replicate vials. The analysis was performed by fitting regression models for males and females separately to the complete set of 14 generation means. Terms in the model include additive autosomal ($d_a$) and X-linked ($d_x$) effects, and average degree of dominance for autosomal ($h_a$) and X-linked ($h_x$, for females only) loci (see FRY et al. 1994 for details). Unweighted least-squares regression was used to obtain parameter estimates and their standard errors, assuming no epistatic interactions among bristle loci and equal viability of all genotypes in the segregating crosses.

A lower limit to the number of loci contributing to the divergence in abdominal bristle number between the two pairs of high and low isogenic lines was estimated as $n^*_a = D^2/8V_a$, where $n^*_a$ is the effective number of loci, $D$ is the difference in means between the two parental lines, and $V_a$ the additive genetic variance segregating in the cross (WRIGHT 1968; LANDE 1981; FALCONER 1989; ZENG et al. 1990). This method is biased towards underestimating the true number of loci contributing to the difference between the parental lines because the restrictive assumptions, that the parental lines are fixed for all alleles increasing (decreasing) the value of the trait and that all effects are equal, strictly additive and unlinked, usually are not correct. In this case, however, the first requirement is met, and $V_a$ was estimated from response to selection from reciprocal F2 crosses of the high and low isogenic selection lines, which partly reduces the confounding effect of linkage when $V_a$ is estimated from the F2 variance only. The ten highest (lowest) scoring pairs of flies were selected from 20 scored from each reciprocal F2, and divergent selection
was continued for four generations. \( V_A \) was estimated from \( V_p \), \( h^2_{\text{REML}} \), where \( V_p \) is the average phenotypic variance from the F2 and of the high and low selection lines from the F2, and \( h^2_{\text{REML}} \) is the restricted maximum likelihood estimate of realized heritability from divergent selection, calculated using the method of MACKAY et al. (1994).

**Chromosome substitution lines:** Chromosome substitution lines of genotype \( H1; L2; L3 \) and \( L1; L2; H3 \) (where \( H \) indicates the high and \( L \) the low parental isogenic selection line, and 1, 2, and 3 the \( X \), second and third chromosome, respectively) were synthesized for each pair (IH1, IL1 and IH2, IL2) of isogenic selection lines. Females from the \( B; Xa/Cy; Tb \) balancer stock were crossed in small mass matings to males from either the high or low isogenic lines (G1). From the high line cross, G2 females of genotype \( B/H1; Cy/H2; Tb/H3 \) were crossed to \( B/Y; Cy/H2; Tb/H3 \) males. From the low line cross, G2 \( B/L1; Cy/L2; Tb/L3 \) females were backcrossed to males of the low isogenic selection line.

To produce the \( L1; L2; H3 \) chromosome substitution lines, G3 females of genotype \( L1; L2; Tb/L3 \) emerging from the low line cross were mated to \( B/Y; Cy/H2; H3/H3 \) G3 males from the high line cross. G4 \( B/L1; Cy/L2; Tb/H3 \) female and \( L1/Y; Cy/L2; Tb/H3 \) male progeny were mated *inter se*, and at G5 \( L1; L2; H3 \) females and males were selected.

To produce the \( H1; L2; L3 \) chromosome substitution lines, G3 females from the low line cross of genotype \( B/L1; L2; L3 \) were mated to G3 \( H1/Y; Cy/H2; Tb/H3 \) males from the high line cross. At G4 \( B/H1; Cy/L2; Tb/L3 \) females and \( B/Y; Cy/L2; Tb/L3 \) males were crossed *inter se*, and at G5 females of genotype \( B/H1; L2; L3 \) were crossed to \( H1/Y; L2; L3 \) males. At G5 \( H1; L2; L3 \) females and males were selected.

**Recombinant isogenic lines:** High/low abdominal bristle number RI lines for the \( X \) and third chromosome were constructed in a low isogenic selection line background. To produce Chromosome 3 RI lines, \( L1; L2; H3 \) females were crossed to \( L1; L2; L3 \) males
from the low isogenic selection line (G1). At G2 \(L1; L2; H3/L3\) female progeny were crossed to males from a stock in which the \(TM6B, Tb\) balancer chromosome and the dominant third chromosome marker \(Sb\) had been placed in the low isogenic \(X\) and second chromosome background \((L1; L2; Tb/Sb)\). Single \(L1; L2; Tb/H3L3\) male progeny of this cross (where \(H3L3\) represents a potential recombination between the third chromosomes from the isogenic high and low selection lines) were backcrossed to females of the \(L1; L2; Tb/Sb\) marker stock, and in G4 \(L1; L2; Tb/H3L3\) females and males were mated *inter se*. At G5 \(L1; L2; H3L3\) females and males were selected to establish a single potential RI3 line. A total of 74 such lines were synthesized.

To produce Chromosome 1 RI lines, \(H1; L2; L3\) females were crossed to \(L1; L2; L3\) males (G1). G2 \(H1/L1; L2; L3\) females were backcrossed to \(L1; L2; L3\) males, and at G3, single \(H1L1/Y; L2; L3\) males (where \(H1L1\) represents a potential recombination between isogenic high and low \(X\) chromosomes) were crossed to females in which the \(Basc\) \(X\) chromosome had been substituted into the low second and third isogenic background \((B; L2; L3)\). G4 \(B/H1L1; L2; L3\) female and \(B/Y; L2; L3\) male progeny were mated *inter se* and at G5 \(B/H1L1; L2; L3\) females were crossed to \(H1L1; L2; L3\) males. G6 \(H1L1; L2; L3\) females and males were selected to establish a single potential RI1 line. In total, 97 such lines were constructed.

For each of the RI lines, abdominal and sternopleural bristle number was scored on a sample of 10 males and 10 females from each of two replicate vials. Viability relative to the \(Tb\) balancer chromosome was also estimated for the RI3 lines. For each line, two replicate vials of small mass matings of six \(L1; L2; Tb/H3L3\) females and males were set up (Day 1), and the total number of \(Tb/H3L3\) and \(H3L3\) flies emerging to Day 18 recorded. Viability of the \(H3L3\) chromosome relative to \(Tb\), \(v\), was estimated as \(2(1-r)/r\), where \(r\) is the observed proportion of \(Tb/H3L3\) flies.
Determination of recombination breakpoints: The \textit{roo} transposable element (MEYEROWITZ and HOGNESS 1982; and \textit{B104}, SCHERER \textit{et al.} 1982) has a haploid copy number of approximately 60 in natural populations of \textit{D. melanogaster} (MONTGOMERY \textit{et al.} 1987). Like other transposable elements, it has a low occupancy rate in natural populations, such that on average one element will occupy a given cytological location (site) only once in a sample of individuals (MONTGOMERY and LANGLEY 1983). Therefore different \textit{roo} insertions were expected to be fixed in the high and low isogenic lines. The positions of these markers were used to infer crossover positions in the RI lines.

\textit{roo} insertion sites were determined by \textit{in situ} hybridization of biotin-labelled \textit{roo} DNA to polytene salivary chromosomes of third instar larvae reared at 18° C, according to the procedure of SHRIMPTON \textit{et al.} (1986). Phage probes containing a complete copy of \textit{roo} (SCHERER \textit{et al.} 1982) were labelled with biotinylated dATP (bio-7-dATP, BRL) by nick translation. Hybridization was detected using the VECTASTAIN ABC kit (Vector Laboratories) and visualized with horseradish peroxidase/ 3,3′-diaminobenzidine. Since the expected RI pattern of insertion sites is a mixture of the two known parental patterns and since the large and conserved \textit{roo} element gives a consistent and strong signal, in general only one high quality preparation was scored for each RI line. Element locations were determined at the level of cytological band subdivisions on the standard Bridge's map (LEFEVRE 1976) using a phase contrast 100X oil objective. The genetic map positions of the markers were inferred from the map position of genes at or near the same cytological position of a given \textit{roo} element insertion using the 'cytogenetic map' of LINDSLEY and ZIMM (1992). This results in a genetic map in which the map position of each marker is known within 2 cM for most regions and the order of markers is known.
**SSCP markers:** As the X chromosome contained no informative markers from 9C1-2 to 17D1-2 oligonucleotide primers were made to amplify the introns of 3 sequenced genes in this interval: *Lsp1α* (11A7-9; GeneBank X03872; SMITH et al.1981), *rutabaga* (12F-13A; GeneBank M81887; LEVIN et al.1992), and *non-A* (14C1-2; GeneBank X55902; JONES and RUBIN, 1990). The primers were 5' -

AAAACGGAATAGAGTGATGCC - 3' and 5' - CTCCATGTACAGATCGAAGTGCC - 3' for *Lsp1α* (expected product 814 bp), 5' - TGGAAACGTTCAGGTGTCG - 3' and 5' - TGCAAGATCTTTTGCAGACG - 3' for *rutabaga* (observed product 1.4 kb as the primers flank an unsequenced intron; L. LEVIN per. comm.), and 5' -

AACCAAACCGGACTTTTCC - 3' and 5' - TCTTAGCAAGTGCATTAGGGC - 3' for *nonA* (expected product 1392 bp). Amplified products of these genes, averaging 1 kb in length, were phenol extracted, ethanol precipitated, and restricted with a series of 4 and 6 cutter restriction enzymes. The product of these restrictions were then run on non-denaturing acrylamide gel and scored for Single Strand Conformational Polymorphisms (SSCP) between the high and low parental chromosomes (AGUADE et al.1994). No polymorphism was observed for *non-A* or *rutabaga* between parental chromosomes for any enzymes or enzyme combinations (*Cfo I, Hae III + Hinf I*, and *Sau3A I* for *non-A*; *Acc I, Cfo I, Hae III, Hinf I, Sau3A I, Acc I + Hinf I, Acc I + Sau3A I, and Hinf I + Sau3A I* for *rutabaga*). A polymorphism was observed for *Lsp1α* cut with *BamHI*, so each X line was assayed for its genotype with respect to *Lsp1α* using SSCP.

**Mapping algorithm:** The genotype of each line was examined to determine if the observed pattern of *roo* element presence and absence in each line was consistent with a small number of crossover events. *roo* markers that were fixed at the same cytological position in both the high and low lines were ignored in the analysis. Every interval between the remaining adjacent markers was then scored as either High, Low, or Non-
informative. If both markers flanking an interval were High (Low) the interval between them was considered High (Low); otherwise the interval was classed as Non-informative. Considering lines with a crossover between adjacent markers Non-informative has minor ramifications in the dense marker, small number of recombinants, situation of this study (see Discussion).

For each interval the data was analysed using the model:

\[ Y = \mu + I_i + \sum_j S_j + \text{sex} * I_i + \sum_j \text{sex} * S_j + \text{line} (I_i) + \text{sex} * \text{line} (I_i) \]

\[ + \text{rep} (\text{line} I_i) + \text{sex} * \text{rep} (\text{line} I_i) + \varepsilon \]

where \( I_i \) is the interval being tested, \( S_j \) is the effect of the \( j \)th single marker for all single markers which are greater than 10 cM away from both of the interval's defining flanking markers, \( \text{rep} \) refers to replicate vials, parentheses represent the nesting of an effect, and all other factors are as described above. The standard ANOVA model subscripts, indicating levels of each factor, are not specified in the above model to avoid confusion with the subscripts which refer to interval and single marker numbers. The effects of Sex, Interval, and Single marker are all fixed, as are interactions involving only these effects, and all other effects are random. Mean Square Errors (MSE) were estimated for each factor in the above model using the GLM procedure of SAS (SAS INSTITUTE, Inc. 1988). The probability of no factor(s) being present within a tested interval (i) was estimated by constructing a F statistic using the MSE's for \( I_i \) and line(\( I_i \)) as the numerator and denominator respectively and deriving a p-value using the inverse F distribution. A similar p-value for sex*\( I_i \) was derived using the estimate of MSE for the effect of sex*line(\( I_i \)) as a denominator. These p-values represent the relative 'support' for a locus affecting bristle number existing in a given interval, smaller p-values indicating greater support. This mapping method combines interval mapping (LANDER and BOTSTEIN 1989) and conditioning on flanking markers (ZENG 1994; JANSEN and STAM 1994) within a
conceptually simplified framework. This method differs from previously published methods in that information from intervals containing a crossover are ignored; RI lines are used; a least squares approach is utilized; and additional factors are easily incorporated into the model.

**Estimation of allelic effects:** The effect \((a)\) of a given interval was estimated as the difference between the least square means of lines with a High vs. a Low score for that interval after fitting a model that includes all other candidate intervals likely to contain a locus affecting bristle number. Other intervals were included in the model to avoid over-estimating \(a\) by including the effects of linked loci affecting bristle number. Specifically, \(a\) for interval \(i\) was estimated separately for each sex as the difference in marker class least square means from the following model in SAS: 

\[
Y = \mu + I_i + \sum_{j \neq i} S_j + \epsilon,
\]

where \(j\) is the set of all candidate intervals for a given chromosome (see Results and Table 4). Standard errors (s.e.) of estimates of \(a\) were made for the \(i^{th}\) interval by adding the terms line(\(I_i\)) and rep(line \(I_i\)) to the above model and estimating the s.e. as 

\[
\sqrt{\frac{MSE_{\text{line}(I_i)} \times \left(\frac{1}{n_H} + \frac{1}{n_L}\right)}{N_{\text{ind}/\text{line}}}}.
\]

Interactions between the \(i^{th}\) and \(j^{th}\) interval were estimated for each sex separately from the least square means for each combination of candidate intervals from the following model:

\[
Y = \mu + I_i + I_j + I_i \times I_j + \sum_{k \neq i,j} S_k + \epsilon,
\]

where \(k\) is from the same set as \(j\) above. The significance of the interaction effect was assessed by constructing a F ratio using the interaction term as the numerator and the line within interaction term as the denominator. The estimate of the interaction effect between two intervals is 

\[
((\bar{X}_{11} + \bar{X}_{22})-(\bar{X}_{12} + \bar{X}_{21})),
\]

where the first subscript is 1 if interval \(i\) is low and 2 if interval \(i\) is high, and the second subscript takes on the same values for interval \(j\). With no epistasis between markers this term should be zero. Standard errors of
Figure 1 - Response to 25 generations of divergent selection for abdominal bristle number. The standard errors on the generation means are very small as they are based on bristle counts from 100 individuals of each sex.

interaction effects were computed in a manner analogous to those for the main effects by adding terms corresponding to line(I_i*I_j) and rep(line I_i*I_j) to the above model and estimating the s.e. as

$$\sqrt{\frac{\text{MSE}_{\text{line}}(I_i*I_j)}{N_{\text{ind}}/\text{line}}} \times \left( \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right).$$

RESULTS

Selection lines: The response to divergent selection for abdominal bristle number from the Raleigh population is shown in Figure 1. The response had started to plateau after
25 generations of selection, with little change of mean in either line from G22-25. The mean abdominal bristle numbers in the high and low lines over this period were 33.5 and 2.4, respectively. Sternopleural bristle number, which averaged 17.4 in the base population, changed as a correlated response to selection for abdominal bristle number. Mean sternopleural bristle numbers recorded at G25 were 26.4 in the high line and 12.8 in the low line. The purpose of constructing the divergent selection lines was not to estimate quantitative genetic parameters, for which replication is necessary, but rather to create material suitable for mapping the loci responsible for the divergence between lines. Nevertheless, the selection response is typical for single segment abdominal bristle number. The estimated realized heritability from ordinary (unweighted) least squares regression of cumulative response on cumulative selection differential over the first 10 generations is 0.29 (c.f. FRANKHAM et al. 1968), and the estimate of additive genetic variance of the trait is 1.65.

Isogenic sublines were derived from the high and low selection lines using balancer chromosomes. Two isogenic high (IH1, IH2) and low (IL1, IL2) sublines with abdominal bristle numbers closest to those of the random-bred selection lines were retained to use as potential lines for mapping. The first pair, IH1 and IL1, had mean abdominal bristle scores of 33.3 and 6.5, respectively; and average sternopleural bristle scores of 25.4 and 15.6, respectively. Average abdominal bristle numbers of IH2 and IL2 were 32.6 and 5.2, respectively; and respective mean sternopleural bristle numbers were 27.1 and 15.6. The phenotypic scores for the parental lines used to generate the RI lines are given in Table 1.
Table 1. Phenotypic values of the parental lines used to generate the recombinant isogenic lines.

<table>
<thead>
<tr>
<th>Line</th>
<th>Sex</th>
<th>Abdominal Bristle Number</th>
<th>Sternoteral Bristle Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean        Variance</td>
<td>Mean        Variance</td>
</tr>
<tr>
<td>LLL1</td>
<td>♀</td>
<td>5.450       7.8436</td>
<td>16.900       3.3744</td>
</tr>
<tr>
<td></td>
<td>♂</td>
<td>9.950       6.2026</td>
<td>15.775       1.8712</td>
</tr>
<tr>
<td>LLH</td>
<td>♀</td>
<td>20.650      6.7462</td>
<td>21.400       2.4000</td>
</tr>
<tr>
<td></td>
<td>♂</td>
<td>22.225      5.5122</td>
<td>20.350       2.4897</td>
</tr>
<tr>
<td>LLL2</td>
<td>♀</td>
<td>8.325       6.0712</td>
<td>15.575       2.4045</td>
</tr>
<tr>
<td></td>
<td>♂</td>
<td>9.700       3.4974</td>
<td>14.600       0.9641</td>
</tr>
<tr>
<td>HLL</td>
<td>♀</td>
<td>19.225      4.7942</td>
<td>17.300       4.0615</td>
</tr>
<tr>
<td></td>
<td>♂</td>
<td>16.675      4.4301</td>
<td>17.450       3.5872</td>
</tr>
</tbody>
</table>

Means of parental LLL, LLH and HLL lines used to contract RI lines. 40 individuals of each sex were scored per line, at the same time the bristle measurements were made on the RI lines. LLL1 is equivalent to IL1 and is the low line used to generate the RI lines for the third chromosome and LLL2 is equivalent to IL2 and is the low line used for the X chromosome (see text). Similarly the high third chromosome in LLH came from IH1 and the high chromosome in HLL came from IH2. Averaged over sexes, LLL1 and LLH differed by 5.36 and 2.96 standard deviations for abdominal and sternopleural bristle number respectively. Similarly, LLL2 and HLL differed by 4.16 and 1.56 standard deviations for abdominal and sternopleural bristle number respectively.

Quantitative genetic analyses: Generation means analysis was used to estimate additive effects and average degree of dominance for the X chromosome and autosomes for abdominal bristle number, the selected trait, and for correlated response in sternopleural bristle number for these two pairs of isogenic high and low sublines. The estimated parameters and their standard errors are given in Table 2.
Table 2. Generation means analysis of abdominal (AB) and sternopleural (ST) bristle number of isogenic abdominal bristle selection sublines.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Parameter</th>
<th>AB</th>
<th>ST</th>
<th>AB</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>$D$</td>
<td>31.20</td>
<td>8.80</td>
<td>32.67</td>
<td>10.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.59)</td>
<td>(0.58)</td>
<td></td>
<td>(1.25)</td>
</tr>
<tr>
<td></td>
<td>$[d_X]$</td>
<td>0.86</td>
<td>-0.64</td>
<td>8.71</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.35)</td>
<td>(0.50)</td>
<td></td>
<td>(1.07)</td>
</tr>
<tr>
<td></td>
<td>$[h_A]$</td>
<td>-5.69</td>
<td>-1.54</td>
<td>-2.59</td>
<td>-2.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.34)</td>
<td>(0.86)</td>
<td></td>
<td>(1.85)</td>
</tr>
<tr>
<td></td>
<td>$[h_X]$</td>
<td>2.49</td>
<td>-0.13</td>
<td>-0.80</td>
<td>-0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.66)</td>
<td>(0.61)</td>
<td></td>
<td>(1.31)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.98</td>
<td>0.95</td>
<td>0.99</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Fit (Prob)</td>
<td>0.03</td>
<td>0.14</td>
<td>0.04</td>
<td>0.18</td>
</tr>
<tr>
<td>O</td>
<td>$D$</td>
<td>22.28</td>
<td>10.48</td>
<td>22.15</td>
<td>12.22</td>
</tr>
<tr>
<td></td>
<td>$[d_A]$</td>
<td>8.49</td>
<td>4.22</td>
<td>6.92</td>
<td>5.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.45)</td>
<td>(0.40)</td>
<td></td>
<td>(1.03)</td>
</tr>
<tr>
<td></td>
<td>$[d_X]$</td>
<td>2.27</td>
<td>0.50</td>
<td>3.17</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.32)</td>
<td>(0.28)</td>
<td></td>
<td>(0.73)</td>
</tr>
<tr>
<td></td>
<td>$[h_A]$</td>
<td>-0.12</td>
<td>-1.99</td>
<td>-1.71</td>
<td>-2.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.63)</td>
<td>(0.57)</td>
<td></td>
<td>(1.46)</td>
</tr>
<tr>
<td></td>
<td>$[h_X]$</td>
<td>0.99</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td></td>
<td></td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Fit (Prob)</td>
<td>0.63</td>
<td>0.04</td>
<td>0.32</td>
<td>0.24</td>
</tr>
</tbody>
</table>

*: 0.01 < P < 0.05; **: 0.001 < P < 0.01; ***: P < 0.001 $D$ is the average divergence between the high and low selection sublines. $[d_A]$ and $[d_X]$ estimate one-half the difference between parental means contributed by additive autosomal and X-linked effects, respectively; $[h_A]$ and $[h_X]$ estimate the summed dominance deviations of autosomal and X-linked loci, respectively. Standard errors of the estimates are given in parentheses. $R^2$ is the proportion of variation explained by the regression model, with corresponding overall significance. The lack of fit test is described in the text.

The two pairs of lines differ in the relative contributions of additive effects on the autosomes and X chromosome: the difference in mean abdominal bristle number between IH1 and IL1 is explicable in terms of additive autosomal effects, but 58% of the difference
in mean between IH2 and IL2 females can be attributed to additive X-linked effects. For males, additive X-linked effects account for 21% of the mean difference in abdominal bristle number between IH1 and IL1 and 31% of the mean difference between IH2 and IL2. The discrepancy between females and males could reflect sex dimorphism in expression of loci affecting abdominal bristle number, as has been found previously for naturally occurring variation (MACKAY and LANGLEY 1990) and for spontaneous (MACKAY et al. 1992a, 1994; FRY et al. 1994) and P-element-induced (MACKAY et al. 1992b) mutations affecting this trait. Alternatively, estimates of abdominal bristle effects in females may be biased since for both pairs of sublines these data show significant lack of fit to the regression model, possibly attributable to linked or epistatically interacting bristle loci or to viability selection against some genotypes in segregating crosses. On the basis of these results, IH1 and IL1 were chosen as the parental lines for mapping third chromosome bristle loci, and IH2 and IL2 for X-linked loci.

Significant autosomal directional dominance for abdominal bristle number was found only for IH1, IL1 females but in all cases the estimates were negative, indicating a tendency for dominance of low alleles. The correlated response of sternopleural bristle number for both pairs of isogenic sublines was caused by loci with additive autosomal effects, low alleles being dominant.

From the generation means analysis, the expectation is that the mapped bristle loci will have significant additive effects on abdominal and sternopleural bristle number. In addition, genetic features of bristle loci contributing to the divergence between the high and low lines may include differences in abdominal bristle effects in males and females, close linkage of loci affecting abdominal bristle number or of loci each affecting one of the two bristle traits; pleiotropic effects of loci affecting abdominal bristle number on viability or on
Table 3. Estimates of effective number of loci

<table>
<thead>
<tr>
<th>Cross (♂ x ♀)</th>
<th>D</th>
<th>V_p</th>
<th>h^2_REML (C. L.)</th>
<th>V_A</th>
<th>n_e</th>
</tr>
</thead>
<tbody>
<tr>
<td>IH1 x IL1</td>
<td>26.74</td>
<td>32.49</td>
<td>0.248 (0.145, 0.406)</td>
<td>8.053</td>
<td>11.1</td>
</tr>
<tr>
<td>IL1 x IH1</td>
<td>27.41</td>
<td></td>
<td>0.339 (0.215, 0.523)</td>
<td>9.293</td>
<td>9.6</td>
</tr>
<tr>
<td>IH2 x IL2</td>
<td>27.41</td>
<td>21.57</td>
<td>0.455 (0.307, 0.663)</td>
<td>9.806</td>
<td>9.6</td>
</tr>
<tr>
<td>IL2 x IH2</td>
<td>27.12</td>
<td></td>
<td>0.397 (0.263, 0.589)</td>
<td>10.763</td>
<td>8.7</td>
</tr>
</tbody>
</table>

D is the divergence in mean abdominal bristle number between the high and low isogenic selection sublines; V_p is the phenotypic variance averaged over the F2 and four generations of divergent selection, h^2_REML (C. L.) are heritability estimates (confidence limits) from REML analysis of response to selection. Additive genetic variance (V_A) is estimated from V_p h^2_REML, and n_e is the effective number of loci.

sternopleural bristle number; and/or epistatic interactions among loci affecting abdominal bristle number.

Estimates of effective number of loci from response to four generations of selection from the F2 of reciprocal crosses of the two pairs of high and low isogenic selection lines are given in Table 3. At least 10 loci contribute to the divergence in abdominal bristle number between the high and low sublines.

RI lines: The 74 Chromosome 3 and 97 X chromosome RI lines were scored for abdominal and sternopleural bristle number, and for roo insertion sites. In addition, viability of the Chromosome 3 RI lines was estimated relative to the TM6B, Tb balancer chromosome. Of the 74 RI3 lines, one was not included in the analysis because it was not homozygous viable, and six were excluded because their roo patterns were not simple combinations of the high and low parental lines. The former may be a synthetic lethal and the latter are likely the consequence of recombination with the balancer chromosome. Four
of the 97 X chromosome RI lines were discarded as they had apparently recombined with the balancer chromosome during extraction.

Mean bristle numbers and roo patterns are given for all X and Chromosome 3 RI lines used in the analysis in Appendices A and B, respectively. There were 22 roo markers in total on the high and low X chromosomes, of which 16 were informative (i.e., polymorphic and reliably scored cytologically). Of the 42 roo insertions on the parental third chromosomes, 29 were informative. The average spacing of informative X chromosome markers was 4.1 cM, and of the third chromosome markers was 3.8 cM. The largest region of the X chromosome without an informative marker was 28.5 cM and on the third chromosome as 13.5 cM. Several of the RI lines had no detectable recombination based on the roo markers: 12 of the X chromosome lines had the low and 11 the high parental patterns; 7 of the Chromosome 3 lines had the low and 9 the high parental marker pattern.

There were no informative markers between cytological positions 9C1-2 and 17D1-2 on the X chromosome, and 4 roo elements were present at the same positions in both the high and low lines, so it is possible that this region may have been fixed for the same segment of chromosome. Oligonucleotide primers were designed to amplify the introns of genes in this interval. The amplified gene product of the high and low lines were then restricted with a series of enzymes and run under SSCP conditions. Using this method no polymorphisms were found for rutabaga or non-A, suggesting at least a portion of the region from 9C1-2 and 17D1-2 is likely the same in the high and low lines. Lsp1α was not scored for all lines, but is informative in the lines for which it was scored. The section of the X chromosome common to both the high and low lines may be from the balancer, or represent the same section of the X chromosome which has gone to fixation in both the high and low lines.
Figure 2 - The probability that there is no genetic factor in a given interval as a function of the X chromosome map position of the interval midpoint. Effects of all single markers greater than 10 cM from the endpoints of the tested interval were statistically removed before fitting the effect of each interval (see text). Filled circles are values for the main effect of the tested interval and hollow squares are for the sex by interval interaction. The horizontal line is the probability value required to reject the hypothesis of no genetic factor, and is equal to 0.05 divided by the number of intervals tested. Thus the higher the 'peak' the greater the support for a genetic factor existing in the tested interval. Tests for abdominal bristle number are given in the top panel and tests on the correlated response in sternopleural bristle number in the bottom panel.

Figure 3 - The probability that there is no genetic factor in a given interval as a function of the map position of the interval midpoint for the chromosome three RI lines. The top panel depicts abdominal bristle number and the bottom panel the correlated response on sternopleural bristle number.

Estimates of the map position of genetic factors: The panels in Figure 2 depict the probability of no factor existing for each interval as a function of interval midpoint. Figure 2 indicates that there is a locus or loci near the tip of the X chromosome (see below) that had an affect on both abdominal and sternopleural bristle number. A second large factor that affected only abdominal bristle number was located near the centromere, in either Interval 16 or 17 with an interval midpoint of 1-62.5 (17D1-2 to 19F). Both these regions also display significant sex*character interactions, indicative of a region at which a substitution has a strong affect in one sex, but not the other. An additional interval whose midpoint was at 1-20.5 (cyotgenetically 7B to 7C) had an effect on sternopleural bristle number, but not on abdominal bristles.

Figure 3 shows a region of large effect on both abdominal and sternopleural bristle number was located near the left tip of the third chromosome at interval midpoint 3-6.8
(cytogenetically 61A1 to 64C1). There is an additional region that affected abdominal bristle number at interval midpoint 3-47.5 (cytogenetically 75C1 to 85E1), and a region that affected sternopleural bristle number at interval midpoint 3-64.2 (cytogenetically 89D1 to 92E1). None of the significant factors mapped on the third chromosome showed significant sex*interval effects. Although not formally significant, additional factors that affected abdominal bristle number are suggested by the smaller peaks at interval midpoints 3-26.0, 3-64.2, 3-87.5 (cytogenetically 64D1 - 66A1, 89D1 - 92E1 and 96B5 - 96F5 respectively). An additional sternopleural bristle number factor was suggested by a non-significant peak at interval midpoint 3-103.5 (cytogenetically 99F1 to 100C2).

The method of statistically removing the effects of flanking single markers before fitting each interval is a powerful approach, but it is not without shortcomings (ZENG 1994; JANSEN and STAM 1994). Figure 2 demonstrates such a problem. The probability of a genetic factor existing in Interval 6 appears two orders of magnitude higher than such a factor existing in Interval 1. We excluded lines recombinant between adjacent markers, but only lines 6 and 71 had a recombination event in Intervals 1 through 6 (i.e., 1A1 to 4C1-2). Thus, these intervals were not recombinantly differentiated in our set of RI lines, and it follows that the apparent significance of Interval 6 is an artifact of the conditioning on flanking markers. When a model was used that did not remove the effect of linked markers, and the effect of marker 4C1-2 (for abdominals $F_{1,91} = 152.43$ and $a = 4.92$ bristles) is compared to marker 6A1-2 (for abdominals $F_{1,91} = 35.49$ and $a = 3.21$ bristles), it is apparent that the effect observed at the tip is likely in the interval from 1A1 to 4C1-2.

The gain in precision in mapping the positions of putative genetic factors by conditioning on flanking markers results in a corresponding decrease in the power of detecting loci (ZENG 1994; JANSEN and STAM 1994). All the peaks discussed earlier
that are not significant can be made so by conditioning on markers further away than the 10
cM used in this work (data not shown), at the expense of resolving different peaks. It
follows that the significance levels presented here are likely to be too conservative. Thus,
the significance threshold line in Figures 2 and 3 indicating the threshold for statistical
significance should not be considered as a rigorous hypothesis tests, so much as indicating
a promising region for further genetic work.
Columns give estimates of the allelic effect for each interval that appeared likely to contain a factor(s) affecting bristle number from Figures 2 and 3. The text describes the methods for estimating effects and their associated standard errors. The columns labelled Interval and Sex*Interval are the significance tests for the main effect of interval and the Sex*Interval interaction effects, respectively. These significance tests were performed on a model that included the interval being tested and all single markers corresponding to the left-most marker for each of the other intervals in the table, thus the p-values may not correspond those in Figures 2 and 3. ** denotes significance at p < 0.05 divided by the total number of intervals tested (i.e., 0.0033 and 0.0018 for the X and chromosome 3 respectively), and * denotes significance at p < 0.01 not corrected for multiple tests.

**Estimates of average effects:** Table 4 gives estimates of the effects associated with intervals that appeared to contain factors affecting bristle number in Figures 2 and 3. Both effects and significance were assessed using a model that contained only the tested interval and all other single markers that corresponded to the left-most marker of the other promising intervals (i.e., the other intervals in Table 4). Not all intervals presented are

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Interval Name</th>
<th>Cytogenetic Interval</th>
<th>m</th>
<th>s.e.</th>
<th>l</th>
<th>s.e.</th>
<th>m</th>
<th>s.e.</th>
<th>l</th>
<th>s.e.</th>
<th>Interval</th>
<th>Sex*Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>11</td>
<td>1E3-4 : 4C1-2</td>
<td>3.10</td>
<td>0.153</td>
<td>5.41</td>
<td>0.175</td>
<td>**</td>
<td>0.68</td>
<td>0.033</td>
<td>0.23</td>
<td>0.101</td>
<td>**</td>
</tr>
<tr>
<td>19</td>
<td>7B1-2 : 7C1-2</td>
<td>0.34</td>
<td>0.149</td>
<td>0.28</td>
<td>0.171</td>
<td>0.50</td>
<td>0.600</td>
<td>0.48</td>
<td>0.008</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>19C1-2 : 19E1-2</td>
<td>0.77</td>
<td>0.149</td>
<td>3.58</td>
<td>0.171</td>
<td>**</td>
<td>0.01</td>
<td>0.000</td>
<td>0.04</td>
<td>0.008</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>61A1 : 64C1</td>
<td>2.37</td>
<td>0.356</td>
<td>2.40</td>
<td>0.615</td>
<td>*</td>
<td>3.09</td>
<td>0.200</td>
<td>3.48</td>
<td>0.221</td>
<td>**</td>
</tr>
<tr>
<td>14</td>
<td>66A1 : 67C4</td>
<td>1.44</td>
<td>0.349</td>
<td>4.20</td>
<td>0.620</td>
<td>**</td>
<td>0.23</td>
<td>0.191</td>
<td>-0.11</td>
<td>0.207</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>75C1 : 85E1</td>
<td>2.00</td>
<td>0.323</td>
<td>5.55</td>
<td>0.559</td>
<td>**</td>
<td>-0.96</td>
<td>0.184</td>
<td>-1.45</td>
<td>0.206</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>80D1 : 82E1</td>
<td>3.62</td>
<td>0.317</td>
<td>2.41</td>
<td>0.560</td>
<td>*</td>
<td>1.23</td>
<td>0.194</td>
<td>1.58</td>
<td>0.206</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>96E5 : 96F5</td>
<td>2.74</td>
<td>0.298</td>
<td>2.77</td>
<td>0.550</td>
<td>**</td>
<td>0.66</td>
<td>0.188</td>
<td>0.77</td>
<td>0.212</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>99F1 : 100C2</td>
<td>-0.00</td>
<td>0.339</td>
<td>-0.43</td>
<td>0.569</td>
<td>-0.72</td>
<td>0.195</td>
<td>-0.05</td>
<td>0.212</td>
<td>**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
formally significant, and estimates are given for both abdominal and sternopleural bristle number even if the interval is significant in only one character. It is clear from this table that the power to detect genetic factors is higher for the X chromosome lines than the third chromosome lines. Among the X chromosome lines a factor in Interval 9 was significant for sternopleural bristles that had an affect of only 0.50 bristles in males and 0.46 bristles in females. The smallest significant effect detected on the third chromosome was in Interval 20 which had an effect of 1.23 bristles in males and 1.58 bristles in females. A factor in Interval 12 on the third chromosome had an effect of -1.45 sternopleural bristles in females, but was not significant. Similarly, Interval 20 had an effect of 3.62 and 2.41 abdominal bristles in males and females respectively, but was only nominally significant. Although Interval 27 was not significant for sternopleural bristle number, it approached significance with a p-value of 0.027. Smaller effects can be detected in the X lines because the number of RI lines for the X chromosome was almost twice that for the third chromosome. Regions with smaller effects would be detected as significant by not conditioning on flanking markers, but this would result in failure to differentiate among linked loci.

It is apparent from Table 4 that allelic effects are often different in males and females. Depending on the interval tested and character considered the low to high allelic substitution can have a larger effects in males (Interval 1 on the X for sternopleural bristles) or in females (Intervals 1 on the X for abdominal bristles). The estimates of effects associated with the third chromosome factors suggest sex-specific allelic effects, but these effects were not formally statistically significant because of the reduced power associated with mapping factors on the third chromosome. Such sex-specific effects are not trivial as they are of the same order of magnitude as the main effects.
The sum of the sternopleural bristle number effects associated with the mapped factors on the X chromosome accounted for 41% and 38% of the difference between the parental lines, in males and females respectively. Similarly, the X chromosome abdominal bristle number effects accounted for 60% and 85% of the difference between parental lines, in males and females respectively. For the third chromosome factors, abdominal bristle effects accounted for 99% and 110%, and sternopleural effects accounted for 77% and 87% of the difference in parental lines, in males and females respectively. As some large factors affecting sternopleural bristle number were negative in value (i.e., the low chromosome had factors increasing sternopleural bristle number), the simple sum of the factors for sternopleural bristle number, and hence the difference between the high and low chromosomes, obscures the large factors effecting this character. Although there may be other smaller factors that were not mapped in this experiment, it is clear that the mapped factors of large effect account for much of the difference between the parental lines.
Table 5: Estimates of effects associated with tests of epistasis

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Interaction Tested</th>
<th>m</th>
<th>s.e.</th>
<th>f</th>
<th>s.e.</th>
<th>Epistasis</th>
<th>Sex</th>
<th>Epistasis</th>
<th>m</th>
<th>s.e.</th>
<th>f</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>II*19</td>
<td>0.67</td>
<td>0.362</td>
<td>-1.48</td>
<td>0.414</td>
<td>0.30</td>
<td>0.221</td>
<td>0.21</td>
<td>0.226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II*117</td>
<td>0.72</td>
<td>0.313</td>
<td>0.74</td>
<td>0.399</td>
<td>0.30</td>
<td>0.192</td>
<td>0.23</td>
<td>0.205</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19*117</td>
<td>0.13</td>
<td>0.306</td>
<td>-0.70</td>
<td>0.361</td>
<td>-0.18</td>
<td>0.185</td>
<td>0.14</td>
<td>0.202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>II*14</td>
<td>1.43</td>
<td>1.23</td>
<td>2.73</td>
<td>2.20</td>
<td>-0.48</td>
<td>0.68</td>
<td>-0.35</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II*112</td>
<td>1.51</td>
<td>0.75</td>
<td>1.96</td>
<td>1.32</td>
<td>1.13</td>
<td>0.41</td>
<td>1.47</td>
<td>0.45 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II*120</td>
<td>1.54</td>
<td>0.73</td>
<td>2.88</td>
<td>1.29 *</td>
<td>1.52</td>
<td>0.41</td>
<td>1.57</td>
<td>0.42 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II*125</td>
<td>0.46</td>
<td>0.66</td>
<td>2.44</td>
<td>1.20 *</td>
<td>0.72</td>
<td>0.41</td>
<td>0.41</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II*127</td>
<td>-0.19</td>
<td>0.74</td>
<td>1.00</td>
<td>1.27</td>
<td>0.25</td>
<td>0.41</td>
<td>-0.18</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14*112</td>
<td>1.74</td>
<td>1.07</td>
<td>2.22</td>
<td>1.94</td>
<td>1.82</td>
<td>0.55</td>
<td>1.49</td>
<td>0.62 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14*120</td>
<td>2.32</td>
<td>0.81</td>
<td>4.57</td>
<td>1.49 **</td>
<td>2.08</td>
<td>0.45</td>
<td>1.95</td>
<td>0.50 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14*125</td>
<td>0.87</td>
<td>0.69</td>
<td>3.71</td>
<td>1.26 **</td>
<td>0.94</td>
<td>0.41</td>
<td>0.61</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14*127</td>
<td>0.04</td>
<td>0.75</td>
<td>1.36</td>
<td>1.30</td>
<td>0.36</td>
<td>0.40</td>
<td>0.29</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112*120</td>
<td>2.16</td>
<td>1.54</td>
<td>6.66</td>
<td>2.62 *</td>
<td>2.02</td>
<td>0.92</td>
<td>2.29</td>
<td>0.97 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112*125</td>
<td>-0.43</td>
<td>0.67</td>
<td>1.51</td>
<td>1.24 *</td>
<td>0.06</td>
<td>0.43</td>
<td>-0.13</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112*127</td>
<td>-0.03</td>
<td>0.68</td>
<td>1.06</td>
<td>1.16</td>
<td>0.09</td>
<td>0.38</td>
<td>-0.04</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120*125</td>
<td>0.06</td>
<td>0.74</td>
<td>1.81</td>
<td>1.42</td>
<td>0.06</td>
<td>0.52</td>
<td>-0.59</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120*127</td>
<td>0.15</td>
<td>0.68</td>
<td>0.84</td>
<td>1.19</td>
<td>0.09</td>
<td>0.41</td>
<td>-0.25</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125*127</td>
<td>0.23</td>
<td>0.92</td>
<td>-1.03</td>
<td>1.68</td>
<td>0.55</td>
<td>0.55</td>
<td>0.01</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Columns give estimates of the epistatic effect for every pair of intervals tested. The text describes the methods for estimating epistatic effects and their associated standard errors. The columns labelled Epistasis and Sex*Epistasis are the significance tests for the specified Epistasis between two intervals and the Sex*Epistasis effects respectively. *** denotes significance at p < 0.001, ** denotes significance at p < 0.01, and * denotes significance at p < 0.05.
Estimates of epistasis between mapped genetic factors: Pairs of intervals likely to contain factors that affected abdominal or sternopleural bristle number (i.e., the intervals of table 4) were fitted in a model that included the pair of intervals, their interaction, and all other single markers corresponding to the left-most marker of the remaining candidate intervals. The results of these tests and the associated interaction effects by sex are presented in Table 5. None of the tests were significant for the X chromosome intervals. For the third chromosome intervals four tests are significant for abdominal bristle epistasis, four tests for abdominal bristle sex*epistasis interaction were significant, five tests were significant for sternopleural bristle epistasis, and zero tests were significant for the sternopleural bristle sex*epistasis effects. All 13 tests of epistasis and sex*epistasis out of 60 tests in total were significant at $p < 0.05$; assuming the tests are independent the probability of observing this many tests significant by chance is less than $10^{-1}$. Interval 20 was involved in 6 of the 9 significant tests of epistasis, even though it did not have the largest main effect. It is apparent from an examination of the effects that epistasis is an important component of the variation observed in these experiments as the significant epistatic effects ranged from 0.87 to 6.66 bristles. In addition, 4 of the 60 tests of epistasis were significant for an epistasis*sex interaction. The effect of sex on epistasis can be quite large: Interval 12 by 20 interaction had an effect of 2.16 abdominal bristles in males and 6.66 abdominal bristles in females. That is, epistatic effects are of the same order of magnitude as the average allelic effects for single intervals, and the modifying effect of sex on epistasis between factors can also be as large as the allelic effects for single intervals.
Figure 4 - The probability that there is no genetic factor affecting viability relative to TM6B, Tb balancer chromosome in a given interval as a function of the map position of the interval midpoint for the chromosome three RI lines.

Factors affecting viability: Viability relative to the TM6B, Tb balancer chromosome was mapped in a fashion similar to the two bristle characters for the chromosome 3 RI lines, except that only one estimate of viability for each sex was available for each vial. Figure 4 depicts the probabilities of no factor affecting viability existing in each interval as a function of interval midpoint. No factors were formally significant, although two factors were marginally significant (p < 0.01) at interval midpoints 3-6.8 and 3-96.0, and two were suggestive at interval midpoints 3-47.5 and 3-77.8. Of these four peaks in probability, two (Intervals 1 and 12) correspond to the
locations of the two most probable mapped factors affecting abdominal bristle number (Figure 3). This suggests that two of the factors mapped to chromosome three had a reduction in viability associated with their fixation. Two other potential factors affecting viability were not associated with mapped factors affecting abdominal or sternopleural bristle number. In addition, the viability factor in Interval 12 had a large sex*interval interaction effect, suggesting that the viability effect was much stronger in one sex.

DISCUSSION

Selection response is due to a few factors of large effect: Using a dense map of neutral markers at least 3 factors affecting bristle number were mapped on the X chromosome and 5 factors were detected on the third chromosome. These factors cumulatively accounted for much of the difference between the parental lines, and were of relatively large effect, ranging from approximately 0.5 to 6 bristles in magnitude (or 0.2 to 3.0 σp, where σp is the phenotypic standard deviation of the base population). It is possible that mapped factors represent clusters of closely linked loci of individually smaller effect, but this can not be determined in the context of the current experiment. These results do not rule out the existence of other factors of small effect, but it is clear that a few factors of large effect were responsible for the major portion of the observed selection response. The selection intensities, number of generations selected, and number of individuals selected are consistent with the factors of this study existing in the base population at intermediate frequencies (ROBERTSON 1960; ROBERTSON 1967).

Published estimates of the effects associated with mapped factors are overestimates of the actual effects if there are loci contributing to the high/low line divergence linked to the interval under consideration (MCMILLAN and ROBERTSON 1974; EDWARDS et al.1987; PATERSON et al.1991). One solution to this problem may be to reestimate effects after further fine-scale recombination experiments; this has the added advantage of
more precisely mapping the locus of interest (e.g., PATERSON et al. 1990). Although not as powerful as fine-scale recombination we conditioned our estimates of average effects on those of other linked intervals to which we mapped factors. This will remove some of the upward bias associated with estimates of allelic effects, but still can not distinguish between the effects of one major locus and many loci of small effect clustered together.

Much sexual dimorphism in the size of the estimated effects on bristle number were detected, particularly for abdominal bristle number. Differences in \( a \) between males and females were often between 1 and 3 bristles - of the same order of magnitude as the main effects. Sex dimorphism in the allelic effects for bristle loci has been observed previously for naturally occurring variation (MACKAY and LANGLEY 1990), for spontaneous (MACKAY et al. 1992a; 1994; FRY et al. 1994), and for \( P \)-element-induced (MACKAY et al. 1992b) mutations. Sex-specific effects of mapped factors affecting bristle number had not often been previously reported because males and females were usually pooled for analysis to increase power (but see BREESE and MATHER 1957). Different allelic effects in the sexes may have important evolutionary implications for the maintenance of quantitative variation, since sex-specific effects can be considered a special case of a Genotype by Environment interaction in which an allele always finds itself in one of two environments (male or female) with probability one half (GILLESPIE and TURELLI 1989).

Significant epistasis was detected between mapped intervals and significant differences in epistasis was observed between the two sexes for a given pair of factors. Estimates of the effects associated with the epistatic interactions were as large as 6.6 bristles and the differences in epistatic effects between the sexes as large as 4.5 bristles. Epistasis has been observed in previous studies in Tomato (PATERSON et al. 1991), and Drosophila (SPICKETT and THODAY 1966; SHRIMPTON and ROBERTSON 1988a),
but the estimates of epistatic effects were small in these studies compared to the present study. Part of the reason that epistasis has not been generally observed in the past may be that the power to detect such effects is much smaller than the power to detect single locus effects of the same size. Given the practical limits on the number of progeny that can be scored, and that are necessary to detect effects of biological significance, it is not surprising that the mapping methods currently in use have not noted high levels of epistasis (LANDER and BOTSTEIN 1989).

Past evidence from both estimates of variance components and chromosomal substitution experiments shows that genetic variation for abdominal and sternopleural bristle number is primarily additive, with little detectable dominance or epistatic variation (FALCONER 1989). In this context the large epistatic effects of factors reported here are surprising. However, variation from additive*additive epistasis is always confounded with additive variation, and experiments to estimate variance components are designed to maximize the additive variance and have little power to detect interactions. If large epistatic interactions between factors affecting a typical "additive" trait like abdominal bristle number are an important feature of the genetic architecture of the trait, then the practise of ignoring epistasis in quantitative genetic models (e.g., FALCONER 1989; BARTON and TURELLI 1989) must be questioned.

Sternopleural bristle number showed a correlated response to selection on abdominal bristle number. In most cases intervals with significant effects on sternopleural bristle number also affected abdominal bristle number, consistent with either pleiotropic effects of loci within the intervals affecting both traits, or linked loci within the intervals separately affecting each trait. In two cases (Interval 9 on the X chromosome and Interval 27 on third Chromosome 3) effects were detected for the correlated trait alone; possibly the significance of the effects on abdominal bristle number was underestimated for these
intervals, or the correlated response was attributable to linkage. Two of the mapped viability differences between the high and low third chromosomes were associated with mapped abdominal bristle factors, but the other two viability factors were not associated with bristle number factors. The mapped factors affecting viability that corresponded to those affecting abdominal bristle number were among those factors with the largest abdominal bristle number effects, suggesting that alleles of larger effect are more likely to have pleiotropic deleterious fitness effects. Viability effects not associated with factors affecting bristle number may be attributable to fixation of recessive alleles with deleterious fitness effects by isogenizing the parental chromosomes, or combinations of recessive alleles brought together by recombination as suggested by the observed synthetic lethal chromosome.

Candidate loci: It has been hypothesized that the factors which are responsible for response to selection are alleles of moderate effect at loci which have been identified through laboratory mutants as potentially having a large effect on the trait in question. Such loci are referred to as candidate loci. This hypothesis is supported by the mapping data of this study as all the regions to which we have mapped factors affecting bristle number contain candidate genes. On the X chromosome both Notch and the ASC complex are located near the tip; similarly the bobbed locus is very close to the centromere of the X chromosome. On the third chromosome likely candidate genes in the intervals in which we have mapped factors are extramachrochaetae (61D1-2), hairy (66D9-11), abdominal (66D9-E1), malformed abdomen (84B), Delta (92A1-2), and Enhancer of Split (96F10-14) (LINDSLEY and ZIMM 1992). Both achaete-scute and all the candidate genes on the third chromosome, with the exception of malformed abdomen, are members of the well characterized neurogenic loci (CAMPOS-ORTEGA and JAN 1991). These loci are known
to be involved in determining the location and spacing patterns of sensory bristles on the fly.

In many cases the phenotypes of the mapped factors in our study have similar genetic properties to characterized mutants at their corresponding candidate loci. The factor mapped to the tip of the X chromosome has a larger effect in females than in males as is typical of mutants in the ASC complex (LINDSLEY and ZIMM 1992). The role of the ASC complex in response to selection on bristle number is also supported by the observation of YOO (1980) that a mutation at the scute locus was responsible for a renewed response to long term selection after a plateau had occurred, and that naturally occurring variation in the ASC region is associated with phenotypic variation in bristle number (MACKAY and LANGLEY 1990). It is likely that the factor mapped near the centromere of the X chromosome was an allele at the bobbed locus as new "mutations" at the bobbed locus have appeared in previous selection experiments (FRANKHAM 1988) and the bobbed locus has an inherently high "mutation" rate ($3 \times 10^4$) due to X-Y exchanges in the tandemly repeated rDNA that comprises the locus (FRANKHAM 1988). Allelic effects of the bobbed locus are expected to be larger in our experiment in females than in males due to our failure to extract the selected Y chromosome (i.e., males are $bb^y/bb^r$).

Another prediction from the list of candidate loci is that the postulated ASC allele at 1-0.1 may interact genetically with the factors at 3-0.0 (emc) and 3-26.5 cM (hairy) as both hairy and emc are DNA binding factors which are believed to be negative regulators of transcription in the ASC complex (VAN DORAN et al. 1990). The observed epistasis between whole chromosomes in the generation means anaysis supports this hypothesis, although further experiments are desirable. At least two other studies have mapped factors close to the position of hairy (THODAY et al. 1964; SHRIMPTON and ROBERTSON 1988b), which suggests alleles of large effect at this locus are not rare. It is not clear that
alleles at *hairy* should have sex-specific effects, nor should they be epistatic to other candidate genes on the third chromosome. This suggests that the factor mapped to the region near *hairy* may actually be *abdominal*, which is an interesting candidate as its phenotype is more pronounced in females than males. *mab* was chosen as a candidate for the factor mapped to Interval 75C1 to 85E1 because of the observation that the low line used in the experiments had cuticular defects and the factor mapped to this position had little effect on sternopleural bristle number as is characteristic of the available *mab* alleles. Previous studies have consistently mapped large factors to the region that includes *mab* (3-49.0)(BREESE and MATHER 1957; WOLSTENHOLME and THODAY 1963; SHRIMPTON and ROBERTSON 1988b), this may not be surprising as this is a region with a compressed genetic map relative to its physical map (ASHBURNER 1989 pp. 452-457). *Delta* (3-66.0) is a likely candidate for the factor mapped to the interval from 89D1 to 92E1. This locus is known to interact with the *E(spl)* locus (3-89.0), a good candidate for the factor mapped to the interval from 96F5 to 99F1, and is a region to which factors have been mapped in the past (SHRIMPTON and ROBERTSON 1988b). Future experiments are necessary to address the allelism of the mapped factors with candidate loci though complementation testing and deficiency mapping. In some cases our candidate loci may be incorrectly assigned, or we have mapped loci that have not been characterized yet.

**Mapping approach:** Data were analysed using a simplified interval mapping procedure combined with conditioning on flanking markers (ZENG 1994; JANSEN and STAM 1994), conditioning on all single markers greater than 10 cM from the endpoints of the interval being tested. This approach allowed the inclusion of nested effects, and analysis of the data using pre-existing software (i.e., SAS). Given the high density of markers and replication of genotypes in the present study, 'coding' intervals as non-informative that included a recombinant should not have resulted in a great loss in power.
Conditioning on markers closer than 10 cM from the interval midpoints resulted in very few recombinants separating the interval being tested and flanking markers and hence loss of power to detect effects. Conditioning on markers further than 10 cM resulted in higher significance values for a given interval, but a corresponding loss of resolution. Based on these analyses it is clear that the significance level chosen to reject the null hypothesis of no genetic factor in a given interval is largely arbitrary. Theory must still be developed to address the significance threshold as a function of the markers chosen for conditioning.

A potentially misleading aspect of interval mapping combined with conditioning on flanking markers is that conditioning can shift the position of a probability peak. Figures 2a and b provide an example of this. The peak in probability of the five left-most intervals is at Interval 5 as opposed to Interval 1, yet there are only two recombinants separating these intervals. It follows that these intervals should give very similar p-values, and the difference between them is an artifact of the conditioning on flanking markers. This is confirmed by analysing single markers 1 and 6 without conditioning. A better way to analyse the data would be to consider intervals that are recombinantly indistinguishable as one large interval, in our study of the X chromosome it is impossible to say where in the interval from 1A1 to 4C1-2 the mapped factor lies.

The nature of variation in quantitative traits: The artificial selection experiment which provided the base material for the mapping of this study was designed to maximize the probability that the mapped factors were not newly arising mutants that occurred during the course of the selection experiment (MACKAY 1989). The selection was carried out at relatively low intensity (25%) for a small number of generations (25) from a large base population of 62 isofemale lines, and the response to selection did not contain any jumps. Thus it is likely that many of the mapped factors were segregating in the base population. Although it is impossible to determine the exact frequencies of these
factors in natural populations, replicate selection experiments from the same starting material may provide some information regarding the frequency of these factors in the base population. Observing the same factors repeatedly going to fixation may suggest a few loci of large effect at intermediate frequency are responsible for standing variation in quantitative traits, whereas different factors of large effect in each replicate may suggest a large number of factors of large effect at low frequency (i.e., the house-of-cards model, TURELLI 1984). Alternately, if alleles at the proposed candidate loci are responsible for the differences between the selection lines, and the molecular variants determining the differences between the selected lines could be identified, then the base population could be resampled and the frequency of the factors estimated. The hypothesis that selection response was due to a few loci with alleles at intermediate frequency is supported by past experiments which have apparently mapped factors to the same locations as this study. This suggests that the factors of this study may account for significant amounts of standing variation in the natural populations.

Acknowledgements: This work was supported by NIH grants GM 45146 and GM 45344, and NSERC and OGS predoctoral fellowships to A.D.L. We thank Shao-Bang Zeng for sharing details of his multiple regression interval mapping procedure prior to publication and Lonny Levin for sharing information regarding unpublished *rutabaga* intron positions.
REFERENCES

conformation polymorphism analysis coupled with DNA sequencing reveals
reduced sequence variation in the su(s) and su(w') regions of the Drosophila

Spring Harbor Laboratory Press.

BARTON, N., and M. A. TURELLI, 1989 Evolutionary quantitative genetics: How little

BREESE, E. L., and K. MATHER, 1957 The organization of polygenic activity within a

CAMPOS-ORTEGA J. A., and Y. N. JAN, 1991 Genetic and molecular bases of

DAVIES, R. W., 1971 The genetic relationship of two quantitative characters in
Drosophila melanogaster. II. Location of the effects. Genetics 69: 363-375.

EDWARDS, M. D., C. W. STUBER, and J. F. WENDEL, 1987 Molecular-marker-
facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic

& Sons, Inc.

FRANKHAM, R., 1988 Exchanges in the rRNA multigene family as a source of genetic
variation, pp. 236-242 in Proceedings of the Second International Conference on
Quantitative Genetics, edited by B. S. WEIR, E. J. EISEN, M. M. GOODMAN,


LANDE, R., 1981 The minimum number of genes contributing to quantitative variation between and within populations. Genetics 99: 541-553.


Genetic and Biology of Drosophila, edited by M. ASHBURNER and E.


LEVIN L. R., P. L. HAN, P. M. HWANG, P. G. FEINSTEIN, R. L. DAVIS, and R.
R. REED, 1992 The Drosophila learning and memory gene rutabaga encodes a


MACKAY, T. F. C., 1989 Mutation and the origin of quantitative variation. pp 113-119
in Evolution and Animal Breeding edited by W. G. HILL and T. F. C. MACKAY.

MACKAY, T. F. C., J. D. FRY, R. F. LYMAN AND S. V. NUZHDIN, 1994

Polygenic mutation in Drosophila melanogaster: Estimates from response to

MACKAY T. F. C., and C. H. LANGLEY, 1990 Molecular and phenotypic variation in

MACKAY, T. F. C., R. F. LYMAN, M. S. JACKSON, C. TERZIAN AND W. G.
HILL., 1992a Polygenic mutation in Drosophila melanogaster: Estimates from
divergence among inbred strains. Evolution 46: 300-316.

MACKAY, T. F. C., R. F. LYMAN AND M. S. JACKSON, 1992b Effects of P element
insertions on quantitative traits in Drosophila melanogaster. Genetics 130: 315-
332.

MATHER, K., and J. L. JINKS, 1971 Biometrical Genetics. Cornell University Press,
Ithaca, N.Y.

McMILLAN I., and A. ROBERTSON, 1974 The power of methods for the detection of
MEYEROWITZ, E. M. and D. S. HOGNESS, 1982 Molecular organization of a

Mendelian populations. II. Distribution of three copia-like elements in a natural

MONTGOMERY E., B. CHARLESWORTH, and C. H. LANGLEY, 1987 A test for the
role of natural selection in the stabilization of transposable element copy number in

mapping of quantitative trait loci using selected overlapping recombinant

PATERSON, A. H., S. DAMON, J. D. HEWITT, D. ZAMIR, H. D. RABINOWITCH,
S. E. LINCOLN, E. S. LANDER and S. D. TANKSLEY, 1991 Mendelian
factors underlying quantitative traits in tomato: comparison across species,

B. 153: 234-249.

ROBERTSON, A., 1967 The nature of quantitative genetic variation, pp. 265-280 in
Heritage from Mendel, edited by A. BRINK. The University of Wisconsin Press,
Madison, Wis.


SAX, K., 1923 The association of size differences with seed-coat pattern and
pigmentation in Phaseolus vulgaris. Genetics 8: 552-556.
SCHERER G., C. TSCHUDI, J. PERERA, H. DELIUS, and V. PIRROTTA, 1982
B104, a new dispersed repeated gene family in Drosophila melanogaster and its

SHRIMPTON, A. E., AND A. ROBERTSON, 1988a The isolation of polygenic factors
controlling bristle score in Drosophila melanogaster. I. Allocation of third
chromosome bristle effects to chromosome sections. Genetics 118: 437-443.

SHRIMPTON, A. E., AND A. ROBERTSON, 1988b The isolation of polygenic factors
controlling bristle score in Drosophila melanogaster. I. Distribution of third
chromosome bristle effects within chromosome sections. Genetics 118: 445-459.

SHRIMPTON, A. E., E. A. MONTGOMERY AND C. H. LANGLEY, 1986 Om
mutations in Drosophila ananassae are linked to insertions of a transposable

SMITH D. F., A. MCCLELLAND, B. N. WHITE, C. F. ADDISON, and D. M.
GLOVER, 1981 The molecular cloning of a dispersed set of developmentally
regulated genes which encode the major larval serum protein of D. melanogaster.
Cell 23: 441-449.

SPICKETT, S. G., and J. M. THODAY, 1966 Regular responses to selection. 3.


THODAY, J. M., 1979 Polygene mapping: Uses and limitations, pp. 219-233 in
Quantitative Genetic Variation, edited by J. N. THOMPSON and J. M. THODAY.

THODAY, J. M., J. B. GIBSON, and S. G. SPICKETT, 1964 Regular responses to


LEDEND FOR APPENDICES

The marker patterns and mean bristle numbers for each RI line are provided for the X and third chromosome RI lines in Appendices A and B respectively. In both appendices bold lettering of the cytological position of a roo insert indicates that the element is present in the high parental chromosome, and normal font indicates the element is present in the low chromosome. There were a number of roo markers at either the same position in the high and low lines or too close together cytologically discern. On the X chromosome non-informative roo elements were at cytogenetic positions 3A, 13B1-2, 14B1-2, 16D, 17C, 20A1-2. Similarly the third chromosome had non-informative markers at cytogenetic positions 61C1, 62D2, 78E1, 79C-D, 83D2, 84B1, 84D1, 84D2, 85A1, 87D1, 94D-E, 98B1, 98C1. The band(s) at 20A1-2 on the X chromosome, and from 78E1 to 85A1 on the third chromosome are in regions that are difficult cytologically because of poor banding and quality of spreads near the centromere. In addition, both the X and third chromosome have the occasional roo insertion present in one or a few of the RI lines. Such inserts were observed at positions 9E (in RI1-lines 32, 57, 59, 78, 80, 85), 10A (in RI1-lines 32, 57, 59, 78, 80, 85), and 12E (RI1-lines 4, 8, 21, 29) on the X chromosome, and at positions 63E1 (RI3-line 58), 65D (RI3-line 12), 67C2 (RI3-line 82), 73C (RI3-line 84), 90A (RI3-lines 67, 69), 93C-D (RI3-lines 29, 81), and 95E1 (RI3-line 79) on the third chromosome. These bands may represent new insertions or crossovers with the balancer chromosome. These bands were ignored in the analysis and the lines included in the data set analysed.
Chapter 3

Naturally Occurring Variation in Bristle Number Associated with DNA Sequence Polymorphisms at the scabrous Locus of Drosophila melanogaster

Chaoqiang Lai*, Richard F. Lyman†, Anthony D. Long*‡, Charles H. Langley* and Trudy F.C. Mackay†

* Center for Population Biology, University of California at Davis, CA95616.
† Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC27695-7614. ‡ Department of Biology, McMaster University, 1280 Main St. West, Hamilton, Ontario, Canada, L8S 4K1

Submitted for publication in the journal SCIENCE

Corresponding Author: Chaoqiang Lai
Center for Population Biology
Storer Hall
University of California at Davis
Davis, California, 95616

Phone: (916) 752-4253
Fax: (916) 752-1449
E-mail: cqlai@ucdavis.edu

Key words: quantitative genetics, natural variation, molecular polymorphism, Drosophila, bristle number, peripheral nervous system and scabrous
Preface to Chapter 3

Chapter three of this thesis revolves around the significance of observed associations between polymorphic restriction sites at *scabrous* (a candidate gene) and bristle number variation among a series of second chromosomes extracted from the wild. This is the first such study in a series (i.e., *emc, h, Dl, E(spl), da, etc.* to follow) in which I am involved. Studies which associate molecular variation at candidate loci with phenotypic variation in a quantitative trait are important as they implicate the candidate locus as harbouring allelic variants affecting the trait and they provide a rough estimate of the frequency of such variants in the population being surveyed. QTL mapping studies accomplish neither of these objectives. C. Lai is the principle author of the *scabrous* manuscript, and as such collected, and analyzed much of the data. An unforeseen problem with the data though, is that it involved a large number of F tests on sites that were largely independent of one another. Thus, unlike previous studies a single test on haplotypes could not be carried out, and although a number of the molecular marker sites were associated with bristle differences at a marginal significance level, a correction for multiple tests (i.e., the Bonferroni test) resulted in no sites being globally significant. My role was to develop a statistical method which would allow analysis of the data and assess its significance. To this end a new permutation test was created and a computer program written by me which would implement this test and assess the significance of associations between restrictions sites and phenotypic data (see Appendix A). This program has a number of features which will be used in future studies of this sort. Without the permutation test employed in this program, the results of this study would not have reached statistical significance and we would not have been able to conclude that *scabrous* harbours alleles of large effect at intermediate frequency affecting bristle number variation. Thus, although not a first author on this work I feel I had a very definable, critical, and substantial
role in completing this work.
We examined the association between quantitative genetic variation in bristle number in *Drosophila melanogaster* and molecular variation at a candidate neurogenic locus, *scabrous*. Approximately 32% (21%) of the genetic variation in abdominal (sternopleural) bristle number among 47 second chromosomes from a natural population was correlated with DNA sequence polymorphisms at this locus; several polymorphic sites associated with large (= 0.6 phenotypic standard deviation) phenotypic effects occurred at intermediate frequency. These results indicate that quantitative genetic variation caused by alleles of large effects at a few loci, segregating at intermediate frequencies, should be accommodated in models for the maintenance of quantitative genetic variance.

Despite the importance of quantitative characters in medicine, improvement of domestic species and evolution, little is known of the particular Mendelian variants that give rise to the heritable component of these traits. One hypothesis amenable to experimental investigation is that allelic variation at loci involved in the development of a particular trait are a major source of quantitative differences in that trait. The numbers of abdominal and sternopleural bristles of *Drosophila melanogaster* are typical quantitative characters (1). Because Drosophila bristles are sensilla (sensory organs) of the peripheral nervous system, likely candidate genes for these traits are the 10-20 proneural and neurogenic loci that determine the presence and absence of sensory hairs (2). Mutations at some of these loci (3) have contributed to response to artificial selection for high and low bristle numbers, and insertional polymorphisms in the proneural *achaete-scute* complex have been associated with naturally occurring genetic variation in bristle number (4). The *scabrous* (*sca*) locus encodes a signal protein involved in lateral inhibition of the developing nervous system,
with mutant alleles having large effects on bristle numbers and eye morphology (5). Here we test the hypothesis that allelic variation at sca contributes to quantitative genetic variation in natural populations of D. melanogaster by associating molecular polymorphisms at this locus with genetic variation in bristle number.

Table 1 gives the mean abdominal and sternopleural bristle numbers for each sex for each of 47 independent second chromosome lines extracted from a natural population and placed in an isogenic genetic background (6). The overall means were 16.30 ± 0.42 (S.E.) abdominal bristles and 16.15 ± 0.22 (S.E.) sternopleural bristles. Assuming additivity, the total additive genetic variance (σ²a) of the second chromosome was estimated as 1.23 for abdominal bristle number and 0.64 for sternopleural bristle number (7). The additive genetic covariance between both characters was 0.39 with a genetic correlation coefficient of 0.43 (7). These estimates are consistent with previous observations on these characters (8).

Table 1. Mean bristle numbers of the homozygous second chromosome lines and their molecular variation in the scabrous region. For diallelic polymorphic sites (restriction enzyme polymorphism and deletion/insertion), + = present, - = absent. Rare polymorphic sites listed in the fourth column from the right are: 1 = InA(-14.1), 6 = PstI(-5.2), 10 = PstI(-1.6), 12 = InG(1.2), 13 = InH(1.5), 14 = InI(3.1), 16 = InK(4.7), 17 = BamHI(4.7), 22 = EcoRI(11.5)-1.3, 26 = DelN(13.5), 27 = InO(14.7), 29 = InP(15.6), 30 = InQ(20.0), 31 = InR(21.2), 33 = InT(20.2), 35 = InU(25.3), 36 = InV(24.7), 39 = PstI(31.0), 42 = InX(30.6). "AB(1)Haplotype" are the abdominal bristle number haplotypes based on EcoRI(6.5) and PstI(26.6) only; "AB(2)Haplotype" are the abdominal bristle number haplotypes based on EcoRI(6.5), PstI(26.6) and SSCP1447. "SB-Haplotype" are the sternopleural bristle number haplotypes based on PstI(-12.0), EcoRI(-1.5) and InL(5.3). Unique haplotypes were pooled into one haplotype class. Diallelic polymorphism frequencies are given in the last row of the Table. The 9 SSCP1447 alleles were grouped into 6 classes: 1(22), 2(5), 3(8), 4(6), 8(2), 5(4), where the number in parentheses is the number of chromosomes of each allelic class and the last class listed is composed of all unique alleles. Likewise, the 8 SSCP1683 alleles were grouped into 6 classes (1(26), 2(6), 3(7), 4(3), 6(2), 5(3)) and the 14 SSCP1836 alleles into 8 classes (0(4), 1(23), 3(3), 4(2), 5(3), 6(3), 10(2), 2(7)).
<table>
<thead>
<tr>
<th>Chromosome Line</th>
<th>Abdominal Bristle</th>
<th>Sternopleural Bristle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>1</td>
<td>17.00</td>
<td>18.10</td>
</tr>
<tr>
<td>2</td>
<td>15.25</td>
<td>16.30</td>
</tr>
<tr>
<td>3</td>
<td>15.80</td>
<td>16.00</td>
</tr>
<tr>
<td>4</td>
<td>16.75</td>
<td>18.05</td>
</tr>
<tr>
<td>5</td>
<td>15.35</td>
<td>17.30</td>
</tr>
<tr>
<td>6</td>
<td>13.75</td>
<td>15.70</td>
</tr>
<tr>
<td>7</td>
<td>16.30</td>
<td>17.45</td>
</tr>
<tr>
<td>8</td>
<td>15.60</td>
<td>17.55</td>
</tr>
<tr>
<td>9</td>
<td>16.10</td>
<td>18.30</td>
</tr>
<tr>
<td>10</td>
<td>15.40</td>
<td>15.90</td>
</tr>
<tr>
<td>11</td>
<td>14.40</td>
<td>17.25</td>
</tr>
<tr>
<td>12</td>
<td>13.65</td>
<td>16.55</td>
</tr>
<tr>
<td>13</td>
<td>13.50</td>
<td>15.70</td>
</tr>
<tr>
<td>14</td>
<td>16.90</td>
<td>17.80</td>
</tr>
<tr>
<td>15</td>
<td>17.15</td>
<td>19.55</td>
</tr>
<tr>
<td>16</td>
<td>20.05</td>
<td>21.15</td>
</tr>
<tr>
<td>17</td>
<td>13.25</td>
<td>14.25</td>
</tr>
<tr>
<td>18</td>
<td>16.25</td>
<td>17.95</td>
</tr>
<tr>
<td>19</td>
<td>14.30</td>
<td>14.80</td>
</tr>
<tr>
<td>20</td>
<td>15.85</td>
<td>17.10</td>
</tr>
<tr>
<td>21</td>
<td>18.55</td>
<td>19.10</td>
</tr>
<tr>
<td>22</td>
<td>15.65</td>
<td>15.70</td>
</tr>
<tr>
<td>23</td>
<td>15.75</td>
<td>17.35</td>
</tr>
<tr>
<td>24</td>
<td>14.55</td>
<td>16.20</td>
</tr>
<tr>
<td>25</td>
<td>15.35</td>
<td>15.65</td>
</tr>
<tr>
<td>26</td>
<td>13.95</td>
<td>15.80</td>
</tr>
<tr>
<td>27</td>
<td>14.60</td>
<td>16.25</td>
</tr>
<tr>
<td>28</td>
<td>16.90</td>
<td>18.00</td>
</tr>
<tr>
<td>29</td>
<td>15.70</td>
<td>16.90</td>
</tr>
<tr>
<td>30</td>
<td>18.60</td>
<td>18.20</td>
</tr>
<tr>
<td>31</td>
<td>14.65</td>
<td>15.45</td>
</tr>
<tr>
<td>32</td>
<td>12.75</td>
<td>16.20</td>
</tr>
<tr>
<td>33</td>
<td>16.30</td>
<td>18.10</td>
</tr>
<tr>
<td>34</td>
<td>15.90</td>
<td>18.70</td>
</tr>
<tr>
<td>35</td>
<td>19.60</td>
<td>22.10</td>
</tr>
<tr>
<td>36</td>
<td>15.50</td>
<td>18.05</td>
</tr>
<tr>
<td>37</td>
<td>15.80</td>
<td>16.75</td>
</tr>
<tr>
<td>38</td>
<td>16.40</td>
<td>18.60</td>
</tr>
<tr>
<td>39</td>
<td>15.75</td>
<td>17.05</td>
</tr>
<tr>
<td>40</td>
<td>10.35</td>
<td>10.70</td>
</tr>
<tr>
<td>41</td>
<td>14.35</td>
<td>15.15</td>
</tr>
<tr>
<td>42</td>
<td>14.95</td>
<td>16.55</td>
</tr>
<tr>
<td>43</td>
<td>15.55</td>
<td>15.50</td>
</tr>
<tr>
<td>44</td>
<td>13.90</td>
<td>17.65</td>
</tr>
<tr>
<td>45</td>
<td>16.10</td>
<td>18.45</td>
</tr>
</tbody>
</table>

Polymorphism Frequency

| 0.02 | 0.17 | 0.15 | 0.06 | 0.06 | 0.13 | 0.05 | 0.11 | 0.06 | 0.03 | 0.28 | 0.06 | 0.06 | 0.15 |
Restriction map variation of a 45-kilobase region including the \textit{sca} locus was quantified among the 47 chromosomes (Fig.1 and Table 1). There were 18 restriction site polymorphisms and 25 length polymorphisms (insertions and deletions). Single Stranded Conformation Polymorphism (SSCP) was determined (9) for 3 fragments covering the last intron and parts of the third and fourth exons (Fig.1); 31 classes of SSCP were observed. The estimated nucleotide diversity ($\pi$) for the entire 45-kb region was 0.010 (10). Some small insertions and deletions (<0.5kb) were at intermediate frequencies, whereas large insertions (>0.5kb) were rare, consistent with previous surveys (11). Levels of pair-wise linkage disequilibria among the 27 polymorphic sites with frequencies between 3/47 and 44/47 were analyzed and are given in the lower part of Fig.1. This amount and pattern of molecular variation is typical for \textit{D. melanogaster} loci (11).

\textbf{Fig.1: Upper Part}: Restriction map variation in the \textit{scabrous} locus region of \textit{D. melanogaster} among 47 second chromosomes. DNA was isolated from adult flies from each homozygous chromosome line and digested separately with three 6-cutter restriction enzymes (\textit{PstI, BamHI}, and \textit{EcoRI}). Southern blot hybridization was performed using 9 probes completely covering the 45 kb region. The probes were labeled using Digoxigenin-dUTP (Boehringer Mannheim) (22). Polymorphic sites of 6-cutter restriction enzymes are shown above the map. Monomorphic sites of 6-cutter restriction enzymes and insertion/deletion polymorphisms are indicated above the map. Solid squares = \textit{PstI}, open squares = \textit{BamHI}, open circles = \textit{EcoRI}. The open portions of the map are the \textit{sca} coding regions with the 5' end to the left, the 3' end to the right, and the start of transcription at 0 (indicated by the arrow). The approximate sizes of insertions (In) / deletions (Del) are as follows: A = 0.2kb; B = 0.2kb; C = 0.1kb; D = 0.1kb; E = 0.1kb; F = 50kb; G = 2.0kb; H = 0.1kb; I = 2.3kb with an \textit{EcoRI} site; J = 0.3kb; K = 1.3kb; L = 0.2kb; M = 0.2kb; N = 0.1kb; O = 1.0kb; P = 3.6kb; Q = >10kb; R = 0.6kb with an \textit{EcoRI} site; S = 0.4kb; T = 3.2kb; U = 1.9kb with an \textit{EcoRI} site; V = 3.9kb with an \textit{EcoRI} site; W = 0.2kb; X = 4.9kb with an \textit{EcoRI} site; Y = 0.1kb. SSCP was investigated for three DNA fragments (9): 1683(186 n.t.), 1447(300 n.t.) and 1836(290 n.t.), in the region of the third intron and parts of the third and fourth exons. \textbf{Lower Part}: Pair-wise linkage disequilibria among the 27 polymorphisms with frequencies between 3/47 and 44/47. The corresponding names of each of these sites are listed diagonally and their corresponding map positions are indicated by the dashed line. Significance levels are depicted by shading: black P<0.001, darker shading P<0.01, light shading P<0.05 and no shading P>0.05. Polymorphic sites indicated by solid diamonds are the eleven sites that show significant association with variation in bristle number. Among them, DelE(-2.6) is in significant linkage (P<0.05) disequilibrium with \textit{EcoRI}(12.0)-1.8 and \textit{PstI}(26.6), and almost significantly (Fisher's exact test = 0.054) with \textit{EcoRI}(6.5), and alleles at SSCP1447 and SSCP1683 are in a strong (P<0.001) linkage disequilibrium with each other. \textit{EcoRI}(12.0)-1.8 is in significant (P<0.005) linkage disequilibrium with \textit{EcoRI}(14.5).
Table 2 Polymorphic Sites Significantly Associated with Bristle Number Variation and Estimates of Their Contribution to Additive Genetic Variation in Natural Populations

<table>
<thead>
<tr>
<th>Polymorphic Site</th>
<th>P Value of F test</th>
<th>Effect [X(+) - X(-)]</th>
<th>Unit (σ_P)</th>
<th>Additive Genetic Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dele(-2.6)</td>
<td>0.014</td>
<td>2.35</td>
<td>0.45</td>
<td>0.140</td>
</tr>
<tr>
<td>InF(-2.0)*</td>
<td>0.043</td>
<td>1.06</td>
<td>0.46</td>
<td>0.39</td>
</tr>
<tr>
<td>EcoRI(6.5)</td>
<td>0.026</td>
<td>1.47</td>
<td>0.22</td>
<td>0.53</td>
</tr>
<tr>
<td>PstI(26.6)</td>
<td>0.038</td>
<td>1.10</td>
<td>0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>SSCP1447</td>
<td>0.004</td>
<td></td>
<td></td>
<td>0.299</td>
</tr>
<tr>
<td>SSCP1683</td>
<td>0.018</td>
<td></td>
<td></td>
<td>0.209</td>
</tr>
<tr>
<td>AB1(1) Haplotype</td>
<td></td>
<td></td>
<td></td>
<td>0.388</td>
</tr>
<tr>
<td>AB2(2) Haplotype</td>
<td></td>
<td></td>
<td></td>
<td>0.126</td>
</tr>
<tr>
<td>PstI(-12.0)</td>
<td>0.030</td>
<td>0.98</td>
<td>0.10</td>
<td>0.51</td>
</tr>
<tr>
<td>EcoRI(-1.5)</td>
<td>0.018</td>
<td>1.31</td>
<td>0.12</td>
<td>0.68</td>
</tr>
<tr>
<td>InF(6.3)</td>
<td>0.042</td>
<td>1.04</td>
<td>0.11</td>
<td>0.55</td>
</tr>
<tr>
<td>EcoRI(12.0)-1.8*</td>
<td>0.019</td>
<td>0.47</td>
<td>0.19</td>
<td>0.24</td>
</tr>
<tr>
<td>EcoRI(14.5)*</td>
<td>0.039</td>
<td>0.39</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>SB Haplotype</td>
<td></td>
<td></td>
<td></td>
<td>0.132</td>
</tr>
</tbody>
</table>

Table 2. Polymorphic sites significantly associated with bristle number variation and estimates of their contribution to additive genetic variation in a natural population. *indicates a polymorphic site with a significant sex*site interaction effect, but not a significant main effect. The differences between the means of two classes of diallelic polymorphic sites were estimated as |X(+) - X(-)| for main effects, where X(+) and X(-) are the means of the chromosome lines in which the polymorphism is present (+) and absent (-). The sex*site interaction effects were estimated as (X(M) - X(F)) - (X(+M) - X(+F)), where (X(M) - X(F)) is the difference in males and (X(+M) - X(+F)) the difference in females. S.E. is the standard error of |X(+) - X(-)|. Unit (σ_P) is |X(+) - X(-)| expressed as a fraction of the phenotypic standard deviation (σ_P). The additive genetic variance of each site was estimated by using the formula described in (14). Unique SSCP classes and haplotypes were pooled into one class or one haplotype when estimating additive genetic variance (Table 1). % is the percentage of the total additive genetic variance of the second chromosome explained by the estimated σ_A^2. The additive genetic (σ_A) and phenotypic (σ_P) standard deviation are the square root of the additive genetic and phenotypic variance of bristle number for the second chromosome (7), which are respectively 1.11 and 2.77 abdominal bristles and 0.80 and 1.91 sternopleural bristles.
We tested whether 27 polymorphic sites with frequencies between 3/47 and 44/47 (12) and the presence or absence of large insertions (> 0.5kb) (Table 1) were associated with bristle number variation using a linear model (13). Eleven sites showed a significant \((P < 0.05)\) or a highly significant \((P < 0.01)\) association with variation in bristle numbers among lines or between sexes for one of the two bristle traits (Table 2). The probability of observing 11 significant sites out of 112 tests was assessed by randomly permuting the molecular haplotypes among phenotypes. In 1000 such permutations, 957 had 10 or fewer significant associations of polymorphic sites or sex*site interactions with bristle characters, indicating that some molecular polymorphisms in the \(sca\) region are significantly associated with variation in abdominal and/or sternopleural bristle number. The main effects of polymorphic sites DelE(-2.6), \(EcoRI(6.5)\), \(PstI(26.6)\), SSCP1447 and SSCP1683, and the sex*site interactions of InF(-2.0), were significantly \((P < 0.05)\) correlated with variation in abdominal bristle number, whereas the main effects of \(PstI(-12.0)\), \(EcoRI(-1.5)\) and InL(5.3), and the sex*site interactions of \(EcoRI(12.0)-1.8\) and \(EcoRI(14.5)\) showed significant associations with variation in sternopleural bristle number. The mean bristle number of the lines with large insertions (> 0.5kb) in the \(sca\) region are not significantly deviated from those of the lines without large insertions \((P > 0.93\) and \(P > 0.77\) for sternopleural and abdominal bristle numbers, respectively). The sites associated with variation in abdominal bristle number were different from those associated with variation in sternopleural bristle number, suggesting that different molecular variants in the \(sca\) region might contribute to quantitative genetic variation in these two bristle traits.

Fig.2: The difference between means of chromosome lines containing (+) or lacking (-) a polymorphic site. A) Mean abdominal bristle numbers of 47 chromosome lines were ranked and classified into two groups based on presence (filled) and absence (open) of the \(EcoRI(6.5)\) restriction site. The mean of each line (ordinate) is plotted against the cumulative percentile in the group (abscissa). The width of the gap between two curves reflects the difference between the means of two groups. (B) The effect of presence or absence of InL(5.3) on sternopleural bristle number plotted as in (A).
A

Abdominal Brisle Number

Percent

22
20
18
16
14
12
10
8
6
4
2
0

B

Stemopleural Brisle Number

Percent

20
19
18
17
16
15
14

1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Effects attributable to polymorphisms on the bristle traits were estimated for the 9 diallelic polymorphic sites (Table 2); they range from 0.98 to 2.35 bristles for the main effect, and 0.39 to 1.06 bristles for sex*site interaction effect. In terms of additive genetic ($\sigma_A$) and phenotypic ($\sigma_P$) standard deviation units (Table 2), the main effects of the three diallelic polymorphic sites on abdominal bristle number averaged $1.48\sigma_A$ and $0.59\sigma_P$, and the main effect of three sites on sternopleural bristle number were on average $1.39\sigma_A$ and $0.58\sigma_P$. The sex dimorphic effect of InF(-2.0) on abdominal bristle number was $0.95\sigma_A$ and $0.38\sigma_P$, and that of EcoRI(12.0)-1.8 and EcoRI(14.5) on sternopleural bristles was on average $0.54\sigma_A$ and $0.22\sigma_P$. The extent of the divergence in bristle number between lines with alternative alleles at a polymorphic site is illustrated in Fig.2 for the effect of EcoRI(6.5) on abdominal bristles, and the effect of InL(5.3) on sternopleural bristle number. It should be emphasized that the polymorphic sites associated with large phenotypic effects were not rare, but occurred at intermediate frequencies. Although alleles with large effects were previously observed in some lines selected for extreme bristle number (3), this may be the first direct evidence that alleles with large effects on bristle numbers exist at intermediate frequencies in natural populations.

How much genetic variation in bristle number of the natural population can be attributed to the sco locus? Assuming additivity of allelic effects, the additive genetic variance ($\sigma_s^2$) attributable to each significant polymorphism at sco is $1/2[1-\Sigma(p_i^2)]\sigma_c^2$ (14), where $p_i$ is the frequency of the $i$th class at a polymorphic site and $\sigma_c^2$ is the estimated variance component due to the main effect or site*sex interaction of each polymorphism. The additive genetic variance or genetic variance due to sex*site interaction associated with each of the 11 significant polymorphic sites are given in Table 2. The genetic variance for the three polymorphic sites with sex-specific effects was negligible. The estimates of the genetic variance contributed by all other polymorphic sites were large,
varying from 8 to 24% of the total second chromosome genetic variance of abdominal bristle number, and from 7 to 11% of the total second chromosome genetic variance of sternopleural bristle number.

As some of the polymorphic sites significantly associated with bristle number variation were in significant linkage disequilibrium with each other (see Fig. 1), the overall genetic variance of both bristle traits contributed by sca is not equal to the sum of the estimated additive genetic variances associated with individual sites. The total genetic variance in bristle number contributed by the sca locus was estimated by constructing haplotypes for each trait from the sites that showed a significant association with bristle number variation, but were not in significant linkage disequilibrium with each other; and computing the additive genetic variance based on these haplotypes. Taking into account the linkage disequilibria among sites and excluding the three sites associated with only sex-specific effects, three sites were used to construct haplotypes associated with variation in abdominal (EcoRI(6.5), PstI(26.6) and SSCP1447), and sternopleural (PstI(-12.0), EcoRI(-1.5) and InL(5.3)) bristle number (Table 1). The additive genetic variance attributable to abdominal and sternopleural bristle haplotypes was estimated as for individual sites (14), pooling all unique haplotypes into a single class (Table 2). The overall additive variance in abdominal and sternopleural bristle number contributed by the sca locus was 0.388 (0.126 without SSCP1447) and 0.132, respectively, or 32% (10% without SSCP1447) and 21%, of the second chromosome genetic variance of the two traits. The second chromosome is about 40% of the D. melanogaster genome, thus the estimated contribution of the sca locus is 13% and 8% of the total additive genetic variance for abdominal bristle number and sternopleural bristle number, respectively.
Fig. 3: Association of haplotypes with bristle number variation. (A) The correlation of AB(2) Haplotypes (Table 1, but here without pooling the unique haplotypes) with abdominal bristle number. The mean abdominal bristle number of each chromosome line is plotted against its rank and the haplotype the chromosome carries is indicated. The polymorphic sites are arranged in the order EcoRI(6.5) and PsI(26.6), and the haplotypes are labeled as: open square = + - , solid square = + + , solid triangle = - + , solid star = - - . The outside gray (dashed) lines indicate the mean bristle number of the common haplotype on the ordinate and the median bristle number of the common haplotype on the abscissa, while the inner gray lines indicate the mean and the median of the remaining haplotypes.

(B) Plotted as (A) but for SB-Haplotype (Table 1, but here without pooling the haplotypes). The polymorphic sites are in the order PsI(-12.0), EcoRI(-1.5) and InL(5.3) (14), and the haplotypes are labeled as: open square = - + - , solid star = + + - , solid diamond = - + + , solid cross = - - - , solid square = + - - , solid triangle = + + + , solid circle = - - + .

Fig. 3 depicts the relationship between the constructed haplotypes and the genetic variance in bristle number among the 47 chromosome lines. The mean bristle number of each chromosome line was plotted against its rank and the haplotype the chromosome carries is indicated, for abdominal bristle haplotypes based on EcoRI(6.4) and PsI(26.6) (Fig. 3A) and for sternopleural bristle haplotypes based on PsI(-12.0), EcoRI(-1.5) and InL(5.3) (Fig. 3B). The most common haplotypes for each trait are distributed to one side of the curve, while the remaining haplotypes cluster together at the other.

These results have important implications regarding the number of loci, and effects and frequencies of alleles causing naturally occurring quantitative variation. Quantitative genetic variation is generally thought to be caused by many loci, alleles at each of which have small effects on the traits (1, 15). We have shown here that 8-13% of the total genetic variation in bristle number can be attributed to variation at the sca locus, and previously 5% of the genetic variation in these traits was associated with insertional variation in the achaete-scute complex (4). Clearly, the infinitesimal model (many loci, alleles at each of which have small effects on the traits) is not an accurate reflection of genetic basis of quantitative variation in bristle number. These results further support the hypothesis (4) of a strong correlation between loci identified as important and specific in developmental
genetic studies and those contributing most to the phenotypic variation for that trait in natural populations.

If a few loci account for a substantial fraction of genetic variation in a quantitative trait, allelic effects at these loci must be large. Segregation of alleles with large effects on quantitative traits, at low frequencies, is predicted by theoretical models in which the equilibrium distribution of quantitative variation reflects a balance between the input of new mutations affecting the trait and their elimination by either stabilizing selection on the trait (15) or because the mutations have deleterious pleiotropic effects on fitness (16). For statistical reasons, we did not include rare polymorphic sites in our analysis and so can not exclude the possibility that rare alleles with large effects contribute a significant fraction of the genetic variation in bristle number. However, we did detect diallelic polymorphisms at the \( sca \) locus associated with large phenotypic effects, segregating at intermediate frequencies. This observation is consistent with maintenance of quantitative genetic variation by mutation-selection balance only if some mutations of large effect at the \( sca \) locus are selectively neutral (17), but most of the available evidence suggests that mutations affecting bristle number have deleterious pleiotropic effects on fitness (18). Our results indicate that quantitative genetic variation caused by alleles of large effects at a few loci, segregating at intermediate frequencies, should be accommodated in models for the maintenance of quantitative genetic variance. Further work is necessary to explore empirically and in theory the implications of these observations.

Our estimates of genetic variation of bristle number associated with molecular polymorphism at \( sca \) are potentially biased by our assumptions that allelic effects are strictly additive and that effects are causally associated with the observed molecular variation. For a given gene frequency and effect at a diallelic locus, the minimal and maximal genetic variance associated with that locus is for the case of complete dominance; the minimal
variance is when the recessive allele has the lower frequency, and the maximal variance is when the recessive allele has the higher frequency (19). For example, estimates of genetic variance in abdominal bristle number attributable to *EcoRI*(6.5) polymorphism with complete dominance range from 0.038 to 0.35 (compared to 0.112 assuming strict additivity). For the range of the gene frequencies and the effects observed, estimates of variance contributed assuming additivity could be over- or under-estimates by a factor of three. Furthermore, it is possible that some of the significant associations are spurious, this may bias upwards the estimates of genetic variance attributable to variation at *sca*. On the other hand, it is not likely that the observed polymorphic sites are the direct causes of the differences in bristle number, but rather that they are in linkage disequilibrium with the actual molecular variants causing the effects. In this case, we have underestimated the true allelic effects and consequently also the genetic variance contributed by the *sca* locus (20).

The sites associated with variation in abdominal bristle number are different from those associated with variation in sternopleural bristle number. This may be a reflection of the intricate regulation of the expression of *sca* (5, 21). For example, the third intron and fourth exon are proposed as regulatory regions with binding sites for the homeobox transcription factors coded by *Ubx*, *abA* and *AbdB*(21). These regions may participate in negative regulation of *sca* expression in abdominal segments. Indeed, we found polymorphic sites (SSCP1447 and SSCP1683) in this region were strongly associated with variation in abdominal bristle number. A more mechanistic interpretation of the impact of *sca* on quantitative genetic variation will rest on both further molecular and developmental studies and more focused analyses of specific alleles from natural populations.

References and Notes


6. Forty-seven independent second chromosomes extracted from flies caught at Raleigh Farmer's Market in May, 1988 were substituted into a Samarkand X; ry506 chromosome 3 isogenic genetic background. The homozygous effect of each chromosome on abdominal and sternopleural bristle numbers was estimated from 10 male and 10 female flies scored in each of two replicate vials.

7. The additive genetic variance ($\sigma^2_a$) of the second chromosome =1/2 the variance component due to homozygous chromosome lines, assuming strict additivity (1). Variance components were estimated from a linear model (H. R. Lindman, Analysis of Variance in Experimental Design, Springer-Verlag, New York, 1992), $Y_{jklm} = u + S_j + L_k + (S^*L)_{jk} + R(L)_{(k)} + S^*R(L)_{j(k)} + e_{jklm}$, where $Y_{jklm}$ is the bristle number of individual m, of the jth sex, the ith replicate, and the kth line; $j = 1$ and 2; $i = 1$ and 2; $k = 1, 2, ..., 47; m = 1, 2, ...10$. Variance was partitioned into sources due to the independent fixed effects of sex ($S_j$), the random effects of chromosome line ($L_k$), sex*line interaction ($S^*L_{jk}$), replicates
nested within line \((R(L)_{ij})\), sex*replicate interaction nested within line \((S*R(L)_{ij(k)})\), and error \((e_{jkm})\). The estimate of additive genetic variance of the second chromosome was 1.23 for abdominal bristle number, and 0.64 for sternopleural bristle number. The phenotypic variance was 7.678 for abdominal bristle number and 3.631 for sternopleural bristle number. Using the above model, the additive genetic covariance between abdominal and sternopleural bristle number was estimated as 0.39, and genetic correlation coefficient is \(0.39/\sqrt{(1.266\times0.64)} = 0.43\) (1). All statistical analyses were performed with "SAS" programs (SAS/STAT User’s Guide, Release 6.03 ed., SAS Institute Inc., Cary, NC, USA).


9. SSCP in the three DNA fragments indicated in Fig.1 were investigated as described in M. Aguadé et al., Proc. Natl. Acad. Sci. U.S.A. 91, 4658 (1994). For SSCP1447, the primers are 5'-TGGATCGATTGCAGTCTGGTC-3' and 5'-TCTGATGATG-TGGTTCATTCTGTGC-3'; for SSCP1683, the primers are 5'-GTTCGAGCATGCTCAG-ATCACCCTC-3' and 5'-GACCAGCGACTGCAATCG-3'; for SSCP1836, the primers are 5'-TCAACCACCTGAACAAAGC-3' and 5'-AGTCTGCACTGATCAGACCC-3'. These primers were designed based on the sequences published in (4). The entire third intron was sequenced and determined to be 199 n.t.


12. The \(P\)-values associated with a given F statistics do not accurately reflect the probability of a positive association between molecular and genetic variation for sites with a frequency less than 3 out of 47. This was demonstrated by generating the lesser of 1000 or all
possible "dummy" molecular data sets for polymorphic site frequencies 2, 3, 4, 5, and 7; reanalysing each data set and calculating F statistics; and then comparing the observed and expected distribution of the F statistics. We therefore focused our analysis on 27 polymorphisms with frequencies of 3/47-44/47: 24 restriction site and length polymorphisms and 3 SSCP loci, and the presence or absence of large insertions greater than 0.5kb (see Table 1). All rare alleles of multiple-allelic SSCP loci were pooled into single classes (i.e., SSCP1447 alleles 5, 6, 7, 8 and 9 were pooled into class 5; SSCP1683 alleles 5, 6 and 8 were pooled into class 5; SSCP1836 alleles 2, 4, 7, 8, 9, 10, 11, 12 and 13 were pooled into class 2.)

13. The following linear model was used to test whether individual polymorphic sites at the *sca* region was associated with variation in bristle number: \(Y_{ijklm} = u + M_i + S_j + (M*S)_{ij} + L(M)_{k(i)} + S*L(M)_{jk(i)} + R(L*M)_{l(ki)} + S*R(L*M)_{ij(ki)} + e_{ijklm}\). The notation for each variable is as described above (7), with the addition of the variable \((M)\), representing the polymorphic site tested, where \(i = 1\) and 2 for diallelic polymorphic sites, and \(i = 1, 2, 3, \ldots, r\), for SSCP loci and derived haplotypes. All effects were considered random, except sex, polymorphic site, and their interaction. F tests were conducted by dividing the mean square \((MS_M)\) for polymorphic site by the mean square \((MS_{L(M)})\) of lines nested within polymorphic site for the main effect, and the mean square \((MS_{S*M})\) of sex*site interaction by the mean square \((MS_{S*L(M)})\) of sex*line interaction within site for the sex*site interaction effect.

14. The additive genetic variance \((\sigma_A^2)\) associated with each polymorphic site was estimated from the variance component due to the main effect or sex*site interaction of a polymorphic site. The variance component due to the main effect of a polymorphic site is:

\[
\sigma_c^2 = \frac{(MS_M-MS_{L(M)})/(NJ)}{(n-\Sigma(J_i)^2/n)}
\]

where \(N\) is the number of observations per chromosome line, and \(J = 1/(r-1)[n-\Sigma(J_i)^2/n]\) is the standardized coefficient when there is an unequal number
of lines \( J_i \) for each class \( i \) of a polymorphic site, \( r = \sum i \), \( n \) is the total number of chromosome lines, \( MS_M \) and \( MS_L(M) \) are specified as (13). Since \( MS_M = N/(r-1)\sum \{J_i[X_i - \Sigma(p_iX_i)]^2 \}, \) where \( p_i, X_i \) are the frequency and the mean of chromosome lines homozygous for class \( i \) of a polymorphic site, and \( J_i = np_i, \) then \( \Sigma \{p_i[X_i - \Sigma(p_iX_i)]^2 \} = [1-\Sigma(p_i)^2] \sigma_e^2 + MS_L(M)/(NJ). \) To account for the variance \( (MS_L(M)) \) due to the chromosome line effect and sampling error, \( \Sigma \{p_i[X_i - \Sigma(p_iX_i)]^2 \} = [1-\Sigma(p_i)^2] \sigma_e^2. \) Assuming strictly additive allelic effects, the additive genetic variance associated with a polymorphic site of \( r \) classes (i.e., \( r \) alleles at one locus) is \( \sigma_A^2 = 1/2 \Sigma(p_i (X_i - \Sigma(p_iX_i))^2). \) (O. Kemphorne, *Introduction to Genetic Statistics*, p318 (1957)). Using the above relationship we can show \( \sigma_A^2 = 1/2 \Sigma(p_i (X_i - \Sigma(p_iX_i))^2). \) where \( \sigma_e^2 \) was estimated based on the model described in (13). The same procedure was used to estimate the additive genetic variance associated with haplotypes and the sex*site interaction effect of a polymorphic site.


19. With complete dominance, the genetic variance contributed by a diallelic polymorphic site is \( \sigma_0^2 = 2q(1+q)\sigma_A^2, \) where \( q \) is the gene frequency of the recessive allele and \( \sigma_A^2 \) is the additive genetic variance. \( \sigma_0^2 \) is minimal when \( q \) is the rarer of the two alleles and maximal when the recessive allele has the higher frequency (1).

20. The correlation between the effect \( (a) \) attributable to a polymorphic site and the effect
($\alpha$) due to the actual molecular variant with which the site is associated is $\alpha = \frac{[p(1-p)]/D}{a}$, where $p$ is the frequency of the polymorphic site and $D$ is the linkage disequilibrium between the molecular variant directly causing the bristle number effect and the polymorphic site associated with it. Since $D \leq p(1-p)$, $\alpha \geq a$; that is the estimated effect attributable to a polymorphic site is an underestimate of the effect of the actual variant with which the site is associated.


23. We thank W. G. Hill and M. Turelli for suggestions and comments on methods for estimating the genetic variance, and J. D. Fry and H. Zhou for advices on statistical analysis. We also thank G. M. Rubin, M. Mlodzik, and N. E. Baker for kindly providing us with the probes and information on sca. This research was supported by NIH grants GM45344 and GM45146 and OGS Fellowship to A.D.L.
Chapter 4

DISCUSSION

4.1 QTL Mapping

4.1.1 Methodology: If continuous variation is often due to a few genes of large effect, as opposed to a large number of genes of small effect, it should be possible to genetically map and characterize these factors. This idea was first exploited by Sax (1923), who was able to detect a factor affecting seed size in beans by virtue of linkage with a discretely inherited Mendelian marker. Thoday (1961) initiated a more systematic study of the association between quantitative characters and Mendelian markers in Drosophila. Relevant Drosophila work is reviewed in Chapter two of this thesis. Polygenes are generally mapped from a cross of two parental lines which differ in their means for a trait of interest and are presumably homozygous for marker genes and genes affecting the quantitative trait. The F1 from such a cross is heterozygous and in complete linkage disequilibrium for all polygene loci and marker loci. Recombinants from the F1 individuals are then generated by an F1 intercross, a backcross to one of the parental lines, or some other manner to create a mapping population. This mapping population will have the phenotype of every individual (or line) determined and then regressed on some function of the marker loci genotypes scored for all individuals (or lines). In its simplest form this regression is a t-test on the phenotypic means with the genotype of every marker used in turn as the independent variable (Tanksley et al. 1982; Edwards et al. 1987). Extensions
include: 1) interval mapping whereby an interval flanked by informative markers can be used as an indicator of the genotype of each individual for that interval (Paterson et al. 1988; Lander and Botstein 1989), and 2) interval mapping with conditioning on single flanking markers that do not define the interval in question to reduce the 'noise' in the model (Jansen and Stam 1994; Zeng 1994). Although simple in principle, QTL mapping has only become common relatively recently, mainly because large numbers of highly polymorphic, neutral, Mendelian markers (e.g., RAPDs, RFLPs, SSRs) have only recently become available in most species ( Tanksley 1993). QTL mapping replaces statistical descriptions of quantitative variation in terms of variance components with estimates of factor positions, effects and gene action, and epistasis. Although such mapping is now technically feasible results are still relatively scant because the effort and costs required for QTL mapping are fairly large. For example, mapping based on individual recombinants (as opposed to recombinant inbred lines) requires upwards of 500 individuals, each genotyped at enough genetic loci to cover the genome at a density of at least one marker per 20 cM, for even crude estimates of position, gene effects, and epistasis. Nonetheless, for the few who have persevered the rewards have been tremendous ( Jacob et al. 1991; Paterson et al. 1991; Stuber et al. 1992; Chapter two of this thesis).

4.1.2 Estimates of effects associated with QTLs: Interval mapping provides better estimates of allelic effects than regression on single markers, but such estimates are not without sources of bias, despite claims otherwise (Lander and Botstein 1989; Tanksley 1993). Sources of bias will come from linked factors with effects in the same direction as the interval for which an effect is being estimated, and epistasis between mapped factors which are not controlled for. Inaccurate estimates of gene action can come from closely linked factors of opposite effect alternately fixed (pseudo-overdominance)
(Tanksley 1993), or from alternately fixed factors of opposite effect mapping to
approximately the same position (apparent additivity). The prevalence of these biases will
not be determined until further characterization of QTLs has occurred. The role epistasis
plays in determining the phenotype for continuous traits has not been determined. Current
methods have little power to detect epistatic effects as a large number of genotypic classes
(some of low frequency) potentially exist even for just two diallelic loci (9 classes). Given
a realistic number of genotyped offspring, tests of epistasis lack much statistical power
(Lander and Botstein 1989; Tanksley 1993). The experiments described in Chapter 2 of
this thesis provides a unique opportunity to test for the role of epistasis in the architecture
of quantitative traits.

4.1.3 Inferences about the base population: QTL mapping provides a
wealth of information over traditional variance component based methods of inferring the
nature of continuous variation, but this method is not without shortcomings. QTL mapping
does not answer the question of what is the nature of standing variation for continuous
traits. A question variance component estimation at least attempts to answer. Instead it
changes the problem into the essentially Mendelian problem of mapping the factors
responsible for the differences between two lines, strains, or species. This approach tells
us nothing about the frequency of the mapped variants in the base population (or in some
cases if a 'base' population even exists). This is not a problem for animal and plant
breeders who are primarily interested in identifying useful variation, but it is a shortcoming
to evolutionary and medical geneticists whose goal is to determine how much phenotypic
variation in some population is attributable to a characterized factor
(remember $V_e = 2pq[a + d(q - p)]^2 + [2pqd]^2$).

An additional and related problem with mapping from divergent strains is that it is
not clear that the mapped factors are at any appreciable frequency in the base population
(i.e., they may be rare deleterious mutants), or that they even existed in the base population
(i.e., they may be new mutants that arose during the selection experiment or domestication
of the species) (Mackay 1989). It is plausible that mutants are likely to have large effects
relative to standing variation, and it is known that mapping experiments are biased toward
detecting factors of large effect. Thus, it is not clear how often QTL mapping experiments
are mapping mutants that have occurred during the course of the selection experiment
carried out in order to generate the high and low parental lines. This is expected to be
common in selection experiments carried out for more than 50 generations (Mackay 1989),
and indeed such newly arising mutants have been observed on a number of occasions (Yoo
1980; Frankham 1988). Replicated short term selection experiments initiated from small
samples from the same base population followed by QTL mapping, although involving
considerable effort, would address some of these concerns about the frequency of mapped
factors.

4.1.4 Positional cloning of QTLs: In order to address the question of the
frequency of the alleles corresponding to mapped factors it would seem necessary to clone
the factors and find the actual molecular changes causing the difference between lines
(Quantitative Trait Nucleotide or QTN) or find molecular markers in complete linkage
disequilibrium with the QTN in the base population - a formidable task at any rate.
Positional cloning of such loci is often suggested (by non-molecular biologists) (Hill and
Caballero 1992), but may be impractical without physical lesions associated with mutants at
the locus of interest as is common in such walking experiments. An alternate method
includes covering the region with a series of phage clones or Yeast Artificial Chromosomes
(YACs) and screening cDNA libraries for RNA's expressed in appropriate tissues, or at
appropriate developmental stages (Martin et al. 1993). This is by no means trivial as the
locus would have to be very well localized to even attempt this method, and proving a
cloned locus corresponds to the actual locus of interest will not be simple for loci of small
effect.

4.2 Candidate Genes

4.2.1 Systems without well characterized mutants: One promising way
to move from mapped factors to cloned loci is via candidate loci identified through mutants
of large effect on the trait of interest. In species like Drosophila, maize, mice, and humans
in which a large number of mutants of large effect have been cloned and characterized this
approach may be practical. Other systems will have to await the characterization of factors
of large effect on the trait of interest. Lai (1994) suggested that as transposable element
systems are developed in agricultural species they will allow efficient 'tagging' of candidate
quantitative trait loci. Tagging such loci will lead to their immediate characterization and
cloning (Lai and Mackay 1993). It appears as though in some species quantitative trait loci
will be mapped before mutants of large effect on the same character are readily available. It
may be possible to choose appropriate candidate loci from a genetically well characterized
closely related species if there is well defined synteny between the species. This is one of
the goals of comparative genomic mapping projects (Tanksley 1993), and will certainly
play a role in the future quantitative genetics of species closely related to the model species
of Drosophila, zebrafish, mice, Arabidopsis, maize, and tomato.

4.2.2 Quantitative complementation testing: Systems with well
characterized mutants of large effect on quantitative characters will be the most effective at
utilizing the candidate gene approach. Candidate loci that share the approximate position of
mapped QTLs are potential candidates for this approach. It should be evident that similar
map positions do not guarantee correspondence, and in many cases a chosen candidate
gene will be lead to false hopes. One method of assessing whether or not a candidate locus
is allelic to a mapped QTL is though a modified complementation testing procedure. Long
and his colleagues (unpublished results) are currently using a method called 'quantitative complementation testing' (QCT) to determine if mapped factors are indeed allelic to candidate loci. The test involved crossing the high (H) and low lines (L) (or lines that only differ for a small region including the candidate locus) from a mapping experiment to a control (C) and null mutant at a candidate locus (M) and measuring the means in replicate lines generated in this manner. \( \frac{H}{M} - \frac{L}{M} = \alpha_{\text{hemi}} \) (i.e., the mean of the lines that are High over Mutant minus the mean of the lines that are Low over Mutant) measures the average effect of this locus as a hemizygote and \( \frac{H}{C} - \frac{L}{C} = \alpha_{\text{het}} \) measures the average effect as a heterozygote. If \( \alpha_{\text{hemi}} - \alpha_{\text{het}} \neq 0 \), the null allele at the candidate locus interacts with the difference between the high and low alleles. This is analogous to a 'failure to complement' the QTL and implies allelism. Factors of large effect that appeared in long term selection experiments have been shown to be allelic to candidate genes by a method similar to the QCT (Yoo 1980; Frankham 1988). But these tests, like the standard complementation test used by geneticists, differ from the QCT by not including a control for the additive effect of the QTL proposed to be allelic to the candidate gene. Such a control is not needed if effects are 'obvious' (i.e., having recessive mutant phenotypes), but will prove to be necessary when effects are more subtle (Long et al., unpublished data).

Mapping experiments described in this thesis found a number of factors whose map positions closely corresponded with candidate loci. In addition, preliminary QCT data suggests that the factors of the selection experiment are allelic to the candidate loci.

4.2.3 Gene transformation: Candidate loci may allow a straight-forward transition from mapped factors to genetic loci, but proving that the candidate loci are allelic to mapped QTLs will be more difficult. QCT provides one approach, but can only be carried out in species with appropriate mutants available, and are uninformative if the effect of the QTL is additive over the control and mutant alleles. This will be particularly
disappointing if the gene action at most mapped QTLs turns out to be additive ( Tanksley 1993). If the effects of QTLs are large and sources of background variation can be controlled it may be possible to transform an inbred line with the candidate locus from both the high and low lines used as parents for the mapping population in an attempt to recover the mapped effect. These methods are already available in Drosophila, maize, and mice, but may not be effective if the locus suffers from position effects, or the gene region that must be transformed is very large (e.g., maybe QTNs are 5' or 3' transcriptional enhancers) (Ashburner 1989; Lai 1994).

4.2.4 Associating molecular variation with phenotypic variation: A third widely applicable method, even outside model organisms, is to associate polymorphic molecular markers in the region surrounding the candidate locus with phenotypic variation over a series of wild chromosomes sampled from nature (Sing et al. 1988; Mackay and Langley 1990; Chapter 3 of this thesis). In most species linkage disequilibrium is not expected to extend over large physical distances (Langley et al. 1974; Langley et al. 1977) (polymorphic inversions and domesticated species may provide exceptions), so significant associations between markers and phenotypes imply the actual QTNs are physically quite close to the markers. If there is a great deal of linkage disequilibrium in the region being surveyed, haplotypes or derivatives thereof can be tested for an association with the continuous phenotype (Sing et al. 1988; Templeton et al. 1988; Mackay and Langley 1990). If the region is characterized by linkage equilibrium between molecular markers each marker can be separately tested, and the statistical validity of the number of significant single marker tests assessed using resampling techniques (see Chapter 3 and Appendix A of this thesis). In humans relatively high levels of linkage disequilibrium within a genetic locus appears common (Trowsdale 1993; Hoffman 1994), whereas in Drosophila it varies over loci. For example Adh (Langley et al. 1982; Aquadro et al.
1986) and ASC (Aguadé et al. 1989) show high levels of linkage disequilibrium, whereas white (Miyashita and Langley 1988) and Adh in D. pseudoobscura (Schaeffer and Miller 1993) show low levels of linkage disequilibrium. It follows that the statistical method which will commonly be used to test associations of molecular markers at candidate loci with continuous variation will vary over loci and study organisms.

4.3 Quantitative Trait Nucleotides: In order to fully understand the nature of quantitative variation at the molecular level the search for 'Quantitative Trait Nucleotides' or QTNs will eventually be necessary. If linkage disequilibrium is high at a candidate locus and allelic effects can be associated with a haplotype, the frequency of QTNs at this locus can be estimated without recourse to characterization of the QTN causing the difference. In this situation characterization of a prospective QTN will be very difficult as the different haplotypes may have multiple differences separating them and the physical region potentially containing a QTN may be quite large. Alternately, at loci harbouring variants causing phenotypic variation in a continuous character with low levels of linkage disequilibrium, characterization of QTNs will be necessary in order to accurately estimate the frequency of alleles contributing to variation at the locus. High levels of natural disequilibrium will simplify characterization of QTNs as transformation of an inbred strain can be carried out with 'naturally occurring constructs' (as opposed to laboratory generated constructs), which differ for candidate QTNs and have the background throughout the rest of the construct randomized (Choudhary and Laurie 1991; Laurie and Stam 1994). Regardless, characterization of QTLs to the level of a set of candidate QTNs will require an in-depth understanding of the molecular biology of the candidate locus.

4.4 Progress In Identifying Candidate Genes

4.4.1 Candidate genes in non-Drosophilids: There are a number of instances where attempts are being made to associate molecular variation in candidate genes
with standing variation in a continuous phenotype. The majority of such attempts are being made in human medical genetics or in model systems for the study of polygenic human diseases. The most successful work to date has been in the association of variants at a number of apolipoprotein genes with atherosclerosis and risk of heart disease (Breslow 1992). The associations between apolipoprotein candidate genes and arteriosclerosis in humans were strong and repeatable enough to motivate the development of a mouse model system using gene targeting and knockouts of these genes (Rubin and Smith 1994). Although such animal model systems have been useful they are not without problems. angiotensin-converting enzyme (ACE) (Jacob et al. 1991) and the SA gene (Iwai and Inagami 1992) were originally associated with hypertension in a rat model system, but only the SA gene (Iwai et al. 1994) and not ACE (Harrap et al. 1993) were associated with hypertension in humans. Other candidate genes, such as the insulin receptor and dipeptidyl carboxypeptidase-I, were examined because of their role in controlling blood pressure, and found to be associated with hypertension in humans without recourse to an other animal system (Morris et al. 1993). The candidate gene approach has met with some success for other traits as well and include the association of ApoE and ACE alleles with longevity (Schachter et al. 1994), and the association of haplotypes of the MHC complex with insulin-dependent diabetes mellitus (Dorman et al. 1991; Tait and Harrison 1991). There is increasing evidence that a single base pair causing an amino acid substitution at HLA-DQβ (in the MHC complex) may put individuals at a higher risk of insulin-dependent diabetes (a non-Asp at amino acid position 57 increases the risk of IDDM) (Dorman et al. 1991) - this is the first identified Quantitative Trait Nucleotide (QTN)! In humans the candidate gene approach has been applied to a number of polygenic behavioural traits with little success. Of particular note are the many failed attempts to associate variation at any of a number of dopamine receptors with alcoholism, schizophrenia, and depression (Holmes et al. 1991;
Coor. et al. 1993; Suarez et al. 1994). More work must be done in the area of human
behavioural genetics with other candidate loci. In terms of understanding the genetic basis
of standing variation in quantitative traits, two problems still exist with human candidate
gene approaches. Studies have focused on detecting candidate effects in affected versus
control individuals. This sort of design does not provide estimates of the frequencies of the
alleles responsible for continuous variation nor does it allow unbiased estimates of the
effects associated with an allele at a candidate locus (i.e., affected individuals may have
'increasing' alleles at other loci effecting the trait). The second problem is that the traits
studied are not continuous (or at least rarely described as such) and individuals are often
considered affected (or at risk) as opposed to unaffected (or not at risk). It follows that
estimates are rarely provided of the additive and dominance effects associated with such
QTLs, parameters of importance in evolutionary models.

4.4.2 Candidate genes in Drosophila

4.4.2.1 Candidate genes for bristle number: Although bristle number in
Drosophila has been central to the understanding of quantitative genetics over the last 40
years (Falconer 1989), very little of the quantitative genetics literature has delved into the
developmental biology of bristle formation. This may be because knowledge of the nature
of the genes involved in the development of the Drosophila peripheral nervous system has
only been an active area of research in the last ten years (Campos Ortega and Jan 1991). If
bristle QTLs are allelic to candidate loci characterized through mutants of large effect and
we wish to understand quantitative variation in terms of the underlying genetics of the loci
contributing to a trait it will be essential to learn the developmental biology of bristle
formation. The review will address bristle development at a superficial level in order to
demonstrate that knowledge about the candidate genes involved in bristle formation will
help us to understand the nature of quantitative variation. I refer extensively to four recent
reviews which cover this rapidly changing field in far greater detail than I do here (Held 1991; Campos Ortega 1993; Ghysen and Dambly Chaudiere 1993; Ghysen et al. 1993). I will briefly identify some of the major players in bristle formation, specifically those loci involved in identifying regions which will contain bristle mother cells (BMCs), the loci involved in lateral inhibition and the determination of a dominant BMC, and those loci involved in cell fate decisions of the daughter cells of the BMC.

4.4.2.2 Defining equivalence groups: After general positioning information is determined in the fly through early acting segment polarity genes (Lawrence 1992) a set of loci, referred to as the proneural genes, define proneural clusters. These clusters, also referred to a equivalence groups, are areas composed of a number of cells each with the ability to eventually form a BMC. The achaete-scute complex (ASC) encodes a number of different transcripts whose expression are necessary to define BMC equivalence groups (i.e., loss of function mutants in the ASC lack stereotypic sets of bristles). The genes that make up this complex encode somewhat functionally redundant Helix-Loop-Helix (HLH) proteins and have characteristically large regulatory regions. At least three other factors interact genetically and likely physically with genes of the ASC complex to further define bristle equivalence groups: hairy (h), extramacrochaetae (emc), and daughterless (da). h, emc, da and ASC contain HLH domains which allow the proteins coded by these genes to form homodimers with themselves or heterodimers with one another. Some of these loci contain additional basic domains presumed to be important in binding DNA as regulators of transcription. It is believed expression of the ASC complex is regulated through complex auto- and cross-regulation with these other three loci and heterodimers formed between these loci. A fascinating aspect of the biology of da and h is that they have roles in development that are completely independent of their role in regulating ASC expression: da is involved in sex-determination and h is a 'pair-rule' gene
with an early role in defining segments. If these candidate loci harbour alleles important in standing variation for quantitative traits, previously unimaginable forms of pleiotropy may exist at the level of the QTL.

4.4.2.3 Lateral inhibition: The bristles in equivalence groups generally show very regular forms of spacing, likely too regular for the positions of every bristle to be determined from gradients which supply complete positional information. Thus it seems likely that another mechanism operates to position a BMC within an equivalence group. A likely mechanism is lateral inhibition, whereby one cell in an equivalence group 'decides' to become a BMC and produces a local inhibitory signal which will suppress a similar fate in neighbouring cells. A number of loci have been implicated in the determination of the cell which will eventually form the BMC. Notch (N), Delta (Dl), and Enhancer of Split (E(spl)) appear to be the major players and are characterized by a series of complex genetic interactions. Loss of function mutants at these loci cause an over commitment to a neural fate and hence they are collectively referred to as the neurogenic loci. N and Dl both encode proteins with a transmembrane domain and Epidermal Growth Factor (EGF) like repeats and hence are believed to be involved in cell to cell communication. Mosaic studies (Lawrence 1992) in Drosophila show N and E(spl), but not Dl, to be cell autonomous (i.e., only cells mutant for N in an otherwise wild type patch of cells show the N phenotype whereas cells mutant for Dl (nonautonomous) can cause a mutant phenotype in neighbouring wild type cells). The conclusion from these studies is that Dl is the signal and N the receptor. E(spl) is cell autonomous, but appears to be required in cells which will assume an epidermal fate, and likely acts further downstream of N and Dl as the basic HLH structure of the protein encoded by this gene implies involvement in transcriptional regulation.
Figure 1 (synthesized from Held 1991; Campos Ortega 1993; Ghysen and Dambly Chaudiere 1993; Ghysen et al. 1993) depicts the current model for the role of these genes in forming the BMC. This model involves \( Dl \) from one cell signalling \( N \) in another through physical contact which in turn down regulates \( ASC \) in the nucleus (possibly through unknown second messengers) which in turn up regulates \( Dl \). Eventually, one of the neighbouring communicating cells surpasses a threshold and assumes a dominant BMC fate at which time \( ASC \) begins to up-autoregulate itself resulting in higher levels of \( Dl \) activity which further down regulates the communicating cell. It is possible that at this time \( E(spl) \) begins to play an active role in further down regulating \( ASC \) activity in the cell which will now assume an epidermal fate. There appear to be other factors involved in lateral inhibition which have mutant phenotypes resembling \( N, Dl, \) and \( E(spl) \), but do not seem to be as well incorporated into the molecular picture. The structure of \textit{scabrous (sca)} suggests that it may be a diffusable factor somehow involved in lateral inhibition, but this does not fit nicely with the \textit{N} - \textit{Dl} / \textit{sev} - \textit{boss} paradigm implicating cell-cell communication through direct contact as being the dominant force in cell fate decisions. The molecular events that result in a single cell assuming the role of the BMC and the other(s) assuming an epidermal fate are still far from understood, but it is already clear that the loci affecting this process show intricate patterns of epistasis between one another and other loci, and

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{A model for lateral inhibition}
\end{figure}
second that these loci are involved in a large number of processes independent of bristle formation (e.g., oogenesis, wing venation, eye formation, and muscle development).

4.4.2.4 The daughters of BMC: Once the BMC is defined it undergoes a number of cell division with its daughter cells finally forming a bristle structure. External sensory (es) organs undergo two rounds of cell divisions to produce four daughter cells. The first division produces an inner lineage and an outer lineage, with the inner lineage going on to produce a neuron and accessory cell (neuronal sheath in the case of es organs), and the outer lineage a bristle and a socket. A number of loci affect this process. Loss of function mutants at the cut locus act in a 'homeotic' fashion by transforming an BMC with an es fate to a chordotonal (ch) (internal stretch sensory organs), with a resulting loss in es bristles. Other genes, such as numb, act after the first cell division of the BMC, with loss of function resulting in the offspring of both cells assuming the fate of the outer lineage, that is the formation of two bristles and two sockets that are not innervated. Conversely, loss of function mutants at the locus oversensitive (osn), result in the offspring of the first division cell both assuming a neuronal fate at the expense of the bristle and socket. Mutants at the Hairless (H) locus cause transformation of shafts into sockets resulting in double sockets with no bristles. Loss of function mutants at the Notch locus have such a startling neurogenic phenotype that it was originally thought that the product of this locus acted only at the lateral inhibition stage of bristle development. Yet loss of function temperature sensitive alleles inactivated only later in development show that loss of N function results in an over commitment to neurogenic fate in the BMC offspring, with each BMC forming four neurons with no sheath, bristle, or socket. As in the earlier case of lateral inhibition it appears that daughter cells of the BMC communicate with one another in a manner such that each offspring assumes a different cell fate. Although these late acting loci can clearly result in changes in bristle number, it is not clear that they will harbour
variants causing standing variation in bristle number at appreciable frequencies as variants at these loci tend to produce dysfunctional ectopic bristles, or neurons without an external structure capable of detecting environmental stimuli. This argument is supported by the observation that naturally occurring bristle variation rarely involves sockets without bristles or double bristles in one socket.

4.4.2.5 The nature of candidate loci and the nature of quantitative variation: The development of the peripheral nervous system is far from being understood at the cellular level. There are many other loci involved in bristle development with more complex interactions than I have mentioned in the summary above (Held 1991; Campos Ortega 1993; Ghysen and Dambly Chaudiere 1993; Ghysen, Dambly Chaudiere et al. 1993). Nonetheless, a number of important features of the candidate loci are understood well enough to be of use to Drosophila quantitative geneticists. The direction of the phenotype of null mutants can sometimes help identify an appropriate candidate locus. For example we may expect selection response in the low direction of a bristle selection experiment to be due to partial loss of function mutants at loci such as ASC, cut, and H, whereas response in the high direction due to partial loss of function alleles at the neurogenic loci. The other noteworthy observation, at the level of characterized mutants, is that epistasis is the rule not the exception. Almost every locus involved in bristle formation interacts with at least one other, either at the genetic or the biochemical level. The interactions are often multidimensional and involve a number of factors (Shepard et al. 1989). Thus it should not come as a surprise to quantitative geneticists that epistasis may be far more common than previous analyses have suggested. Finally, the pleiotropy observed at the level of candidate loci is not what one might have naively expected. In the past quantitative geneticists likely thought of pleiotropic effects in related structures (say arm and leg length), but if the candidate loci involved in bristle development harbour alleles
with quantitative effects we must be prepared for pleiotropic effects on seemingly unrelated traits. A lesson learned from molecular developmental biology is that signally pathways which are effective tend to be re-used in many different contexts (e.g., *N, Dl, da*, and *h*).
Chapter 5

CONCLUSIONS

What is the underlying genetic basis of continuous variation? The introduction of this thesis demonstrates that although continuous characters have been studied since the turn of the century we still do not know the underlying genetic basis of variation in continuous characters. It was shown in the introduction that theoretical models of the response to selection, the maintenance of quantitative genetic variation, and the evolution of continuous characters depend critically on the answer to the above question. The goal of the work described in this thesis was to begin to answer this question. Two approaches were taken. Chapter two describes a QTL mapping based approach to understanding the nature of continuous variation, and chapter three a candidate gene approach used to answer the above question. In the following conclusion, progress made in this thesis towards answering the above question will be summarized. In addition, an attempt will be made to assess the significance of the findings of this thesis with respect to future work in the field of quantitative genetics.

Chapter two of this thesis describes experiments carried out in order to map factors responsible for differences in abdominal bristle number between lines selected for high and low bristle number. This experiment mapped two factors of large effect on abdominal bristle number on the X chromosome and 5 factors on the third chromosome. Some of the mapped factors contributed to differences in sternopleural bristle number, and viability,
between the high and low lines. The sum of the mapped allelic effects accounted for nearly one hundred percent of the difference between the parental high and low chromosomes. These results indicate that, although the existence of many genes of small effect can not be ruled out, a few factors of large effect account for much of the difference between selected lines. The selection experiment that generated the high and low lines was carried out for a very short time (25 generations) at a relatively small population size (200 census individuals). A selection experiment with these parameters makes it unlikely that many of the mapped factors were mutants that occurred during the course of the selection experiment. The observation that much of the difference in a quantitative character between lines, varietals, or subspecies is due to a few loci is not unique (Tanksley 1993). But, this is the first study in which newly arising mutants are unlikely to correspond to a large percentage of the mapped factors. Thus, the results of this mapping study show that there are alleles of large effect contributing to standing variation in quantitative characters in natural populations.

The mapping study also showed that there are large and significant sexual differences in allelic effects and that there are significant epistatic effects attributable to pairs of mapped factors. In the past sex specific effects have not been observed, often because phenotypic measures were 'sex corrected' and pooled to increase power. In hindsight, sex specific effects are not surprising as mutants at candidate loci show sex specific effects (Lindsley et al. 1992) and newly arising mutants causing continuous variation are known to have sex specific effects (Fry et al. 1994). Future studies aimed at mapping QTLs in animals should consider sex specific allelic effects. Theory currently does not deal with this empirical observation even though it may contribute significantly to standing genetic variation. Sex specific allelic effects can be incorporated into theoretical models, as they are really a special case of a genotype by environment interaction where an allele always
finds itself in one of two environments with probability one half. Theoretical models of quantitative variation also ignore the role of epistasis in contributing to selection response and standing variation. This is based on the assumption that epistatic effects are small relative to allelic effects, so are negligible. Yet this work observed epistatic (and sex by epistatic) effects of the order of the main effects. Epistasis should be incorporated into future models of standing variation and selection response. Further empirical work is also necessary in order to better quantify the nature of epistasis for quantitative traits.

The move from mapped factors to genes affecting quantitative traits will be greatly simplified if the factors contributing to continuous variation are minor alleles at previously characterized candidate loci. In Drosophila a number of candidate genes for bristle number variation are well characterized. This is by virtue of mutants in these genes giving phenotypes which disrupt the normal development of the peripheral nervous system (i.e., bristles are sensilla). A very promising result of the mapping experiments of this thesis is that in the majority of cases the mapped factors corresponded in map position to well characterized candidate loci. In particular factors were mapped to the approximate positions of the *achaete-scute* complex, *Notch*, and *bobbed* on the *X* chromosome, and the approximate positions of *extramacrochaetae*, *hairy*, *malformed abdomen*, *Delta*, and *Enhancer of split* on the third chromosome. If the factors responsible for response to selection can be proven to be allelic to the proposed candidate loci, through methods such as the Quantitative Complementation Test (QCT), the task of characterizing quantitative variation at the level of single loci will be greatly simplified. Preliminary results from the QCT, prior work on naturally occurring phenotypic and molecular variation in the *achaete-scute* complex (Mackay et al. 1990), and the *scabrous* results of this thesis suggest that molecular variation at candidate loci may contribute substantial variation to continuous characters.
Mapping QTLs is a powerful, but also limited approach to understanding the underlying genetics of quantitative traits. In particular, mapping does not identify a single gene as harbouring quantitative variation, but a region containing a locus or loci contributing variation. This can be a problem if one wishes to move from mapped factors to effects at actual genetic loci, as QTLs are not often mapped very accurately. A second problem is that mapping can not provide estimates of the frequency of alleles causing quantitative variation in the base population, only the effects of such alleles. If the frequency of alleles of large effect mapped in such experiments is actually very low in the base population, then they may not contribute much to standing variation. This is because the genetic variation at a locus is a function of allele frequency. In order to understand quantitative variation at the level of the genes causing such variation it is important to identify actual genes and to estimate the effects and frequencies of alleles at these loci. Associating molecular variation at candidate loci with phenotypic variation for a trait of interest is one means to this end.

Chapter three of this thesis describes an association between molecular variation in the region of the *scabrous* locus and variation in bristle number among a series of wild chromosomes extracted from a natural population. Unlike a previous study of the *achaete-scute* region (Mackay et al. 1990), the type of molecular variation associated with phenotypic variation was not insertions and deletions. In addition, the molecular variants associated with phenotypic variation were at intermediate frequency which suggests that the molecular variants that actually cause the phenotypic differences are at similar frequencies. Thus, alleles of large effect at intermediate frequency at the *scabrous* locus contribute on the order of 10% of the standing genetic variation in bristle number. This observation is not consistent with current models for the maintenance of quantitative genetic variation by mutation selection balance, unless bristle number variation is effectively neutral. The
neutrality of bristle number variation is questionable as the pattern and number of bristles appears conserved between Drosophila taxa. This observation suggests that some form of balancing selection may be responsible maintaining considerable variation in nature. This need not be the case for all loci though, as some appear to harbour alleles of large effect at low frequency (e.g., *achaete-scute*), or have a high intrinsic 'mutation' rate (e.g., *bobbed*), which are consistent with mutation selection balance models.

A major technical problem plagues studies that attempt to associate molecular variation with phenotypic variation is statistical significance. If a candidate gene region is characterized by high levels of linkage disequilibrium, then only one statistical test of the association of haplotypes with phenotypic data need be performed. But, if the region is characterized by low or intermediate levels of linkage disequilibrium, there are potentially as many haplotypes as lines, and so tests must be performed on each molecular marker. This leads to a statistical problem of multiple, non-independent tests, and it is thus difficult to assess the significance of associations between molecular markers and phenotypic values. The approach that was developed to assess the significance of the results for *scabrous*, a locus with intermediate levels of linkage disequilibrium, was to record the total number of (marginal) significant associations at $p < 0.05$ over all molecular markers and characters, and then repeat this process 1000 times with a data set in which the entire molecular haplotypes was shuffled with respect to the phenotypic values for each line. This permutation test gives the distribution of the number of tests marginally significant under the null hypothesis of no association between markers and phenotypes to which the observed number of tests marginally significant can be compared. In the case of *scabrous* the observed number of tests marginally significant was significant as $p < 0.05$. This is a new approach to assessing the significance of molecular variation at candidate loci with respect to phenotypic variation and should be applicable to other systems. A problem with
this approach is that it is difficult to determine which of the marginally significant molecular markers is the marker in linkage disequilibrium with the actual molecular variant causing the observed phenotypic differences, and which markers represent false positives. Thus, although both sex specific allelic effects and different molecular variants associated with abdominal versus sternopleural bristle number were observed to be marginally significant, it is difficult to determine if these observations represent actual significance.

The association of molecular variation at candidate loci with phenotypic variation is potentially a powerful approach to characterizing the nature of quantitative variation at the level of the individual loci contributing to the trait, but it does have limitations. If no significant associations are found, one can not conclude the candidate locus does not contribute to standing variation. Lack of significance can be due to loss of power due to multiple tests or failure to find a molecular variant that is in linkage disequilibrium with the molecular variant actually causing a quantitative difference. In this sense candidate gene association studies are not a replacement for pedigree (or mapping) based studies. QTL mapping combined with fine scale mapping and/or genetic complementation tests can provide good evidence that a candidate gene locus is important in contributing to quantitative variation. QTL mapping studies can also provide evidence that a candidate gene is not the cause of differences between the lines used for the mapping experiment. A replicable lack of association between candidate gene and phenotypic variation over a number of mapping studies suggests that the candidate locus is not important in standing variation. Thus mapping studies will continue to be of use in suggesting candidate regions or loci, which will provide a basis for addition work with candidate loci.

The characterization of the genes contributing to standing variation in quantitative traits is far from complete. The work of this thesis, and other observations, suggests that continuous variation in many case may be due to a few loci of large effect at intermediate
frequency. Additional evidence suggests that candidate loci may harbour alleles important in standing variation, and that at the level of the genes contributing to continuous variation sex specific effects and epistasis may be important. Clearly, the confidence that can be placed on these assertions depends on additional empirical work of the type described in this thesis. If the above assertions are true it is possible that the relationship between standing variation and differences between species must be reassessed. That is, standing variation may be due to both alleles of large effect at mutation selection balance, and alleles of large effect maintained by undocumented balancing selection. Both these types of variation may not contribute to differences between species. This question can only be answered through molecular characterization of alleles at Quantitative Trait Loci within and between species.
Appendix A

This appendix gives a listing of the program used to estimate quantitative genetic parameters associated with molecular markers in the scabrous paper. Following the program listing is an example of the use of the program to demonstrate its utility. The program was written by Tony Long between January and April of 1994. The framework is general enough that with minor modification it will be of great use in studies attempting to associate phenotypic variation with molecular variants at candidate loci.

*****************************************************************************
*****
* The following is a program designed to assess the significance of associations between
* molecular markers and phenotypic variation. It is designed for use with molecular markers
* covering a candidate locus. The program assesses the effect of every marker independently,
* thus in no sense is this a mapping program. The program makes extensive use of resampling
* statistics in order to determine the statistical significance of results and provide variance
* on estimates of parameters (such as Va: the additive variance attributable to a given
* molecular marker). The program can only handle one statistical design (see the ANOVA
* function) and is still not completely user friendly. I have tried to comment the code
* throughout in an attempt to make it usable by others.
* 
* Tony Long, Center for Population Biology, U.C. Davis: July 16/1994
*
*****************************************************************************
*****

#include<stdio.h>
#include<math.h>
#include<stdlib.h>

#define NSEX 2.0 /* number of sexes */
#define NCROSS 2.0 /*
* number of crosses
* this is the number of levels of the cross effect
* e.g., hemizygoites and heterozygoites
*/
#define ZZ 9 /* INTEGER: max value of alleles (defined by SSCP) */
/*
* this is a potential source of grief:
* The program uses the value of the allele to place squares and means

- 112 -
* in correct registers. Thus ZZ must equal the largest VALUE an
* allele takes plus 1 !!
*
#define NLINE 47.0  /* number of lines as double */
#define NLINEI 47  /* INTEGER: number of lines as integer */
#define MARK 31  /* INTEGER: number of molecular markers */
#define XSEED 12345678
#define SIX 6
#define BULL 100  /* INTEGER: number of shuffles to make per boot rep */

double drand(void);
long int xseed;
void nrand(void);
double square(double x);
double gammln(double xx);
double beta(double a, double b, double x);
double betacf(double a, double b, double x);
double pol(double F, double v1, double v2);
void Shuff(int type, int N_line, int N_mark, int six, double *MA, double *MS, int *MM);
void Read_data(int N_line, int N_mark, int six, double *MA, double *MS, int *mark_name, int *MM);
void ANOVA (int Print, int N_line, int N_mark, int N_last, double *MA, double *MS, int *MM, int *mark_name, int *Nsize);
void Remove_effect(int rmeff, int N_line, int N_mark, double *MA, double *MS, int *MM);
void VAR_add (int noise, int boot, int N_line, int N_mark, double *MA, double *MS, int *MM, double *V);

/***************************************************************
* square : will square a variable sent to it
* and will return 0.0 if x is 0.0
* written by Tony Long in Jan. of 1994
***************************************************************/

double square(double x)
{
    double z;

    if (x == 0.0) {
        z = 0.0;
    }
    else {
        z = x * x;
    }

    return z;
}

/***************************************************************
* RAND - Is "Minimal standard" random number generator
*
* synopsis
* int rand(void);
* void srand(int seed);
* double drand(void)
*
* generators: Good ones are hard to find. Commun. ACM 31:1192-
* 1201. The integer routine on page 1195 is used. To improve,
* read D.G. Carta 1988. Two fast implementations of the
* "minimal standard" random number generator. Commun. ACM
void mrand(void)
{
    long int a, m, q, r, lo, hi, test;
    a = 16807;
    m = 2147483647;
    q = 127773;
    r = 2836;
    hi = (double) xseed/(double) q;
    lo = fmod((float) xseed, (float) q);
    test = (float) a * (float) lo - (float) r * (float) hi;
    if (test > 0) xseed = test;
    else xseed = test + m;
}

#define TEST

double drand(void)
{
    mrand();
    return ((double) xseed/RM);
}

#define RM 2147483647.0

/* pof is the p-value associated with an F statistic with v1 and v2 degrees of freedom

the probability of observed F statistic based on relationship
between F and incomplete beta function described in Numerical
Recipes in C pp 181 (1988)

this routine uses functions
betal
betacf
gammin
*/

double pof(double F, double v1, double v2)
{
    double X;
    if( v2/(v2 + v1 * F) > 1.0 || v2/(v2 + v1 * F) < 0.0){
        printf("%g %g %g %g\n", v1, v2, F, v2/(v2 + v1 * F));
        X = 0.999;
    } else{
double gammln(double xx)
{
    double x,tmp,ser;
    static double cof[6] = {76.18009173, -86.50532033, 24.01409822, -1.231739516, 0.120858003, -0.536578212, 0.183750032} ;
    int j;
    x=xx-1.0;
    tmp=x+5.5;
    tmp = (x+0.5)*log(tmp);
    ser=1.0;
    for (j=0;j<5;j++) {
        x += 1.0;
        ser += cof[j]/x;
    }
    return -tmp+log(2.50662827465*ser);
}

#define ITMAX 100
#define EPS 3.0e-7

double betacsf(double a, double b, double x)
{
    double qap, qam, qab, em, tem, d;
    double bx,bm=1.0,bp,bpp;
    double az=1.0,am=1.0,ap,app,old;
    int m;
    qab=a+b;
    qap=a+1.0;
    qam=a+1.0;
    bz=1.0-qab*x/qap;
    for (m=1;m<=ITMAX;m++) {
        em=(double)m;
        tem=em+em;
        d = em*(b-em)*x/((qam+tem)*(a+tem));
        ap=az+d*am;
        bp=bz+d*bm;
        d = -(a+em)*(qab+em)*x/((qap+tem)*(a+tem));
        app=ap+d*az;
        bpp=bp+d*bp;
        old=az;
        am=ap/bpp;
        bm=bp/bpp;
        az=app/bpp;
        bz=1.0;
        if (fabs(az-old) < (EPS*fabs(az))) return az;
    }
    return az;
}
```c
} 
printf("a or b too big, or ITMAX too small in BETACPn");
}

#undef ITMAX
#undef EPS

/**************************************************************************
 * beta distribution subroutine (p.d.f.), from Numerical Recipes in C
 */

double betai(double a, double b, double x) 
{
    double bt;
    /* double gammln(), beta0(); */
    if (x < 0.0 || x > 1.0) printf("Bad x in routine BETANu");
    if (x == 0.0 || x == 1.0) bt=0.0;
    else
        bt = exp(gammln(a+b)-gammln(a)-gammln(b)+a*log(x)+b*log(1.0-x));
    if (x < (a+1.0)/(a+b+2.0))
        return bt*betacf(a,b,x)/a;
    else
        return 1.0-bt*betacf(b,a,1.0-x)/b;
}

/**************************************************************************
 * Subroutine to shuffle _OR_ take a bootstrap sample form the data set
 * 
 * Tony Long: Feb. of 1994
 */

void Shuff(int type, int N_line, int N_mark, int six, double *MA, double *MS, int *MM)
{
    int i, j, k, l;
    int *MMtemp, *IND;
    double *MAtemp, *MStemp;
    /*
    * if type = 1 bootstrap data
    * that is take NLINE's from the data set with replacement
    *
    * if type = 2 randomize molecular marker arrays
    * that is randomize the molecular haplotypes w.r.t. the phenotypic data
    */
    MMtemp = (int *) calloc(N_line * N_mark, sizeof (int));
    MAtemp = (double *) calloc(N_line * six, sizeof (double));
    MStemp = (double *) calloc(N_line * six, sizeof (double));
    IND = (int *) calloc(N_line, sizeof (int));

    if (type == 2)
    {
        for (i = 0; i < N_line; i++)
        {
            *(IND + i) = i;
        }
        j = N_line;
        while (j-- > 1)
        {
            k = (int)(double) (j) * drand4();
            l = *(IND + k);
            *(IND + k) = *(IND + j);
            j = l;
        }
    }
```
*(IND + j) = i;
}
for (i=0; i < N_line; i++){
    for (j=0; j < N_mark; j++){
        *(MMtemp + j + i* N_mark) = *(MM + j + *(IND + i) * N_mark);
    }
    for (j=0; j < six; j++){
        *(MAtemp + j + i*six) = *(MA + j + i*six);
        *(MStemp + j + i*six) = *(MS + j + i*six);
    }
}
}
if (type == 1){
    for (i=0; i < N_line; i++){
        k = (int)(double) N_line * drand4();
        for (j=0; j < N_mark; j++){
            *(MMtemp + j + i*N_mark) = *(MM + j + k*N_mark);
        }
        for (j=0; j < six; j++){
            *(MAtemp + j + i*six) = *(MA + j + k*six);
            *(MStemp + j + i*six) = *(MS + j + k*six);
        }
    }
}
for (i=0; i < N_line; i++){
    for (j=0; j < N_mark; j++){
        *(MM + j + i*N_mark) = *(MMtemp + j + i*N_mark);
    }
    for (j=0; j < six; j++){
        *(MA + j + i*six) = *(MAtemp + j + i*six);
        *(MS + j + i*six) = *(MStemp + j + i*six);
    }
}
free(MMtemp);
free(MAtmp);
free(MStemp);
free(IND);
return;
}

/**************************************************************************
 * Read_data: will read in the raw data from an external file
 * this routine could be made much better in subsequent versions,
 * right now the code actually has to be changed to change the name
 * of the input file, I was a little worried about portability here
 *
 * Tony Long: March 1994
**************************************************************************/

void Read_data(int N_line, int N_mark, int six, double *MA, double *MS, int *mark_name, int *MM)
{
    int i, j, Idummy;
    double crap;
    FILE *in;

    /* INFORMATION ON FORMATING THE INPUT FILE
     *
     * input file consists of:
if(!(in = fopen("scab_in.dat","r")) == NULL)
    printf("can't open input file\n");
}

/* this function reads in the data */

for (i=0; i< N_line; i++)
    for (j=0; j < six; j++)
        if(!fscanf(in,"%li",&crap)) != 1)
            printf("the horror of all horrors scanf not working\n");
        else
            *(MA + j + i*six) = (double) crap;

    for (i=0; i< N_line; i++)
        for (j=0; j < six; j++)
            if(!fscanf(in,"%li",&crap)) != 1)
                printf("the horror of all horrors scanf not working\n");
            else
                *(MS + j + i*six) = (double) crap;

for (i=0; i < N_mark;i++)
    if(!fscanf(in,"%d",&Idummy)) != 1)
        printf("the horror of all horrors scanf not working\n");
    else
        *(mark_name + i) = Idummy;

for (i=0; i< N_line; i++)
    for (j=0; j < N_mark; j++)
        if(!fscanf(in,"%d",&Idummy)) != 1)
            printf("the horror of all horrors scanf not working\n");
else{
    *(MM + j + i*N_mark) = Idummy;
}
}
}
fclose(in);
return;
}

******************************************************************************
* ANOVA: written by Tony Long; Feb. 1994
* many of the algorithms come from:
*    Springer-Verlag.
*    
* this subroutine does a number of things things:
* 1. it estimates F statistics based on two models that are fitted to the data
   for a balanced design. SEE BELOW
* 2. it calculates the p-values associated with these F statistics
* 3. it prints out the mean bristle number for associated with each molecular marker
   (if not bootstrapping)
* 4. it prints out the p-value associated with every test (if not bootstrapping)
* 5. if prints out the number of tests significant for certain combinations of tests and
   probability levels (when bootstrapping and not)
*    
* The tests and models are described below:
* Generally only polymorphic markers are submitted to the program where polymorphism
* is defined as having a frequency of 3 or greater in the total sample. Tests with the
* phenotypic data and 'dummy' molecular markers, gave a distribution of F statistics which
* did NOT fit the expected distribution of F. That is 1000 bootstrap reps with dummy
* variables gave a cumulative distribution of F statistics which did NOT match the expected
* distribution. Two models were fitted one to the homozygous phenotypic data and
* one to the hemizygous and heterozygous data. Specifically the model was
*    Y = u + m + s + m*s + l(m) + s*l(m) + e
* where u = mean, m = molecular marker, s = sex, l = line, and () is a nesting (there was
* also the effect of reps in the design but these do not effect significant tests of the
* effects of marker. Two F statistics were estimated from this design
*    MS(m)/MS[l(m)] and
*    MS(m*s)/MS[s*l(m)]
* The second design was:
*    Y = u + m + c + s + m*c + m*s + c*s + m*c*s + l(m) + c*l(m) + s*l(m) + c*s*l(m) + e
* where notation is as before and c = cross (hemizygous vs. homozygous data). Again
* two F statistics were estimated
*    MS(m*c)/MS[c*l(m)] and
*    MS(m*s*c)/MS[s*c*l(m)]
******************************************************************************

void ANOVA (int Print, int N_line, int N_mark, int N_last, double *MA, double *MS, int *MM, int *mark_name,
    int *Nsig)
{
    int mark, line, zz, i, j, k, l, Zmax;
double Nmark, dummy, test;
double *FA, *FS;
double *SSA, *SSS, *SSSA, *SSSS;
FILE *out2, *out3;

out2 = fopen("scab_pvalues.out","a");
out3 = fopen("scab_means.out","a");

FA = (double *) calloc(4, sizeof(double));
FS = (double *) calloc(4, sizeof(double));
SSA = (double *) calloc(8*ZZ, sizeof(double));
SSS = (double *) calloc(8*ZZ, sizeof(double));
SSA = (double *) calloc(12, sizeof(double));
SSS = (double *) calloc(12, sizeof(double));
RSA = (double *) calloc(6, sizeof(double));
RSS = (double *) calloc(6, sizeof(double));
TTA = (double *) calloc(10, sizeof(double));
TTS = (double *) calloc(10, sizeof(double));
SUMA = (double *) calloc(8, sizeof(double));
SUMS = (double *) calloc(8, sizeof(double));

for (i = 0; i < N_line; i++){
    *(TTA + i) = 0.0;
    *(TTS + i) = 0.0;
}

for (i = 0; i < N_line; i++){
    /* TOTALS */
    /* AB */
        *(TTA + 0) += *(MA + i*SIX);
        *(TTA + 1) += *(MA + 1 + i*SIX);
        *(TTA + 2) += *(MA + 2 + i*SIX);
        *(TTA + 3) += *(MA + 3 + i*SIX);
        *(TTA + 4) += *(MA + 4 + i*SIX);
        *(TTA + 5) += *(MA + 5 + i*SIX);
    /* SB */
        *(TTS + 0) += *(MS + i*SIX);
        *(TTS + 1) += *(MS + 1 + i*SIX);
        *(TTS + 2) += *(MS + 2 + i*SIX);
        *(TTS + 3) += *(MS + 3 + i*SIX);
        *(TTS + 4) += *(MS + 4 + i*SIX);
        *(TTS + 5) += *(MS + 5 + i*SIX);

    *(TTA + 6) = *(TTA + 0) + *(TTA + 1);
    *(TTA + 7) = *(TTA + 2) + *(TTA + 3);
    *(TTA + 8) = *(TTA + 4) + *(TTA + 5);
    *(TTA + 9) = *(TTA + 7) + *(TTA + 8);
    *(TTS + 6) = *(TTS + 0) + *(TTS + 1);
    *(TTS + 7) = *(TTS + 2) + *(TTS + 3);
    *(TTS + 8) = *(TTS + 4) + *(TTS + 5);
    *(TTS + 9) = *(TTS + 7) + *(TTS + 8);

    /* RAW SUM SQUARES (RS?) not involving marker */

    *(RSA + 0) = (square(*(TTA + 0)) / (NSEX * NLINE));
    *(RSA + 1) = (square(*(TTA + 1)) / (NSEX * NLINE));
    *(RSA + 2) = (square(*(TTA + 2)) / (NSEX * NLINE));
    *(RSA + 3) = (square(*(TTA + 3)) / (NSEX * NLINE));
    *(RSA + 4) = (square(*(TTA + 4)) / (NSEX * NLINE));
    *(RSA + 5) = (square(*(TTA + 5)) / (NSEX * NLINE));
    *(RSS + 0) = (square(*(TTA + 0)) / (NSEX * NLINE));
    *(RSS + 1) = (square(*(TTA + 1)) / (NSEX * NLINE));
    *(RSS + 2) = (square(*(TTA + 2)) / (NSEX * NLINE));
    *(RSS + 3) = (square(*(TTA + 3)) / (NSEX * NLINE));
    *(RSS + 4) = (square(*(TTA + 4)) / (NSEX * NLINE));
    *(RSS + 5) = (square(*(TTA + 5)) / (NSEX * NLINE));
*\(RSS + 1\) = (sqrt((TTS + 0)) + sqrt((TTS + 1))) / NLINE; /* RSS S \{homo\} * /
*\(RSS + 2\) = sqrt((TTS + 9)) / (NSEX * NCRoss * NLINE); /* RSS SSS \{meu \{h-h\} * /
*\(RSS + 3\) = (sqrt((TTS + 2)) + (TTS + 4)) + (sqrt((TTS + 3)) + (TTS + 5))) / (NCRoss * NLINE);
/* RSS S \{h-h\} * /
*\(RSS + 4\) = (sqrt((TTS + 7)) + sqrt((TTS + 8))) / (NSEX * NLINE); /* RSS C \{h-h\} * /
*\(RSS + 5\) = (sqrt((TTS + 2)) + sqrt((TTS + 3)) + sqrt((TTS + 4)) + sqrt((TTS + 5))) / NLINE;
/* RSS SC \{h-h\} * /
/
/* initialize counters at level of replicate */
for(i=0; i<6; i++){
    *(Msig + i) = 0;
}
/
/* loop for molecular markers */
for (mark = 0; mark < N_mark - N_last; mark++){
    /* initializing within line counters */
    for (i = 0; i < 4; i++){
        *(F + i) = 0.0;
        *(F + i) = 0.0;
    }
    for (i = 0; i < 8; i++){
        for (j = 0; j < ZZ; j++){
            *(SSA + i + (j*ZZ)) = 0.0;
            *(SSS + j + (j*ZZ)) = 0.0;
        }
    }
    for (i = 0; i < 12; i++){
        *(SSS + i) = 0.0;
        *(SSS + i) = 0.0;
    }
    Nmark = 0.0;
/
/* preliminary calculations of sum of squares (not summed over alleles) */
for (line = 0; line < N_line; line++){
/* abdominal bristles Sums of Squares */
    zz = *(MM + mark + line*N_mark);
    *(SSA + zz + ZZ*1) += *(MA + line*SIX);
    *(SSA + zz + ZZ*2) += *(MA + line*SIX);
    *(SSA + zz + ZZ*3) += *(MA + 2 + line*SIX);
    *(SSA + zz + ZZ*4) += *(MA + 3 + line*SIX);
    *(SSA + zz + ZZ*5) += *(MA + 4 + line*SIX);
    *(SSA + zz + ZZ*6) += *(MA + 5 + line*SIX);
    *(SSA + zz + ZZ*7) += 1.0;
    /* freq of allele zz */
/* second counter */
    *(SSS + 0) += square(*(MA + line*SIX) + *(MA + 1 + line*SIX))/ NSEX; /* Xjk^2 \{homo\} */
    *(SSS + 1) += square(*(MA + line*SIX) + *(MA + 1 + line*SIX)); /* Xjkl^2 \{h-h\} */
    *(SSS + 2) += square(*(MA + 2 + line*SIX) + *(MA + 4 + line*SIX) + square(*(MA + 3 + line*SIX) +
    *(MA + 5 + line*SIX)) / NCRoss; /* Xjkl^2 \{h-h\} */
    *(SSS + 3) += square(*(MA + 2 + line*SIX) + square(*(MA + 3 + line*SIX) + square(*(MA + 4 +
    line*SIX) + square(*(MA + 5 + line*SIX)); /* Xjkl^2 \{h-h\} */
* (SSSA + 4) += \text{(square}*(MA + 2 + line*SIX) + *(MA + 3 + line*SIX)) + \text{square}*(MA + 4 + line*SIX) + *(MA + 5 + line*SIX))/NSEX; /* X.jh.k^2 \{h-h\} */
dummy = *(MA + 2 + line*SIX) + *(MA + 3 + line*SIX);
dummy += *(MA + 4 + line*SIX) + *(MA + 5 + line*SIX);
* (SSSA + 5) += \text{square}*(dummy)/NSEX; /* X.jk^2 \{h-h\} */

/* sternopleural bristles Sums of Squares */

zz = *(MM + mark + line*N.mark);
* (SSS + zz + ZZ*1) += *(MS + line*SIX); /* XMj. \{homo\} */
* (SSS + zz + ZZ*2) += *(MS + 1 + line*SIX); /* XfJ. \{homo\} */
* (SSS + zz + ZZ*3) += *(MS + 2 + line*SIX); /* XMjHTk \{h-h\} */
* (SSS + zz + ZZ*4) += *(MS + 3 + line*SIX); /* XfJHTk \{h-h\} */
* (SSS + zz + ZZ*5) += *(MS + 4 + line*SIX); /* XMjHMk \{h-h\} */
* (SSS + zz + ZZ*6) += *(MS + 5 + line*SIX); /* XfJHMk \{h-h\} */
* (SSS + zz + ZZ^7) += 1.0; /* freq of allele zz */

/* second counter */

* (SSSS + 0) += \text{square}*(MS + line*SIX) + *(MS + 1 + line*SIX))/NSEX; /* X.jk^2 \{homo\} */
* (SSSS + 1) += \text{square}*(MS + line*SIX) + *(MS + 1 + line*SIX)); /* Xjk^2 \{h-h\} */
* (SSSS + 2) += \text{square}*(MS + 2 + line*SIX) + *(MS + 4 + line*SIX)); /* Xjk^2 \{h-h\} */
* (SSSS + 3) += \text{square}*(MS + 2 + line*SIX) + *(MS + 3 + line*SIX)); /* Xjk^2 \{h-h\} */
* (SSSS + 4) += \text{square}*(MS + 2 + line*SIX) + *(MS + 3 + line*SIX)); /* Xjk^2 \{h-h\} */
* (SSSS + 5) += \text{square}*(MS + 3 + line*SIX)); /* Xjk^2 \{h-h\} */

} /* next line */

/* loop over alleles to total up Sums of Squares */
for (zz = 0; zz < ZZ \{zz\}; zz++)

/* second set of Sums of Squares for abdominal bristles */

if (* (SSSA + zz + ZZ^7) > 0.0) {

* (SSSA + 6) += \text{square}*(SSSA + zz + ZZ*1) + *(SSSA + zz + ZZ*2))/NSEX; /* SSSA + ZZ*7); /* Sum Xij.\{h\} */
* (SSSA + 7) += \text{square}*(SSSA + zz + ZZ*1) + \text{square}*(SSSA + zz + ZZ*2))/NSEX; /* SSSA + ZZ*7); /* Sum Xij.\{h\} */
dummy = *(SSSA + zz + ZZ*3) + *(SSSA + zz + ZZ*4);
dummy += *(SSSA + zz + ZZ*5) + *(SSSA + zz + ZZ*6);
* (SSSA + 8) += \text{square}*(dummy)/NSEX; /* NCRoss \{SSSA + zz + ZZ*7}); /* Sum Yij.\{h\} */

} /* SSSA + ZZ*7) */

/* second set of Sums of Squares for sternopleural bristles */

if (* (SSSS + zz + ZZ^7) > 0.0) {

* (SSSS + 6) += \text{square}*(SSSS + zz + ZZ*1) + *(SSSS + zz + ZZ*2))/NSEX; /* SSSS + ZZ*7); /* Sum Xij.\{h\} */
* (SSSS + 7) += \text{square}*(SSSS + zz + ZZ*1) + \text{square}*(SSSS + zz + ZZ*2))/NSEX; /* SSSS + ZZ*7); /* Sum Xij.\{h\} */

} /* SSSS + ZZ*7) */
dummy = *(SSS + zz + ZZ*3) + *(SSS + zz + ZZ*4);  
dummy += *(SSS + zz + ZZ*5) + *(SSS + zz + ZZ*6);
*(SSS + 8) += square(dummy) / (NSEX * NCROSS * *(SSS + zz + ZZ*7));  
  /* Sum Y.j.*2  [h-h] */  
  *(SSS + 9) += square(*(SSS + zz + ZZ*3) + *(SSS + zz + ZZ*5) + square(*(SSS + zz + ZZ*4) + *(SSS + zz + ZZ*6))) / (NCROSS * *(SSS + zz + ZZ*7));  
  /* Sum Y.j.*2  [h-h] */  
  *(SSS + 10) += square(*(SSS + zz + ZZ*3) + *(SSS + zz + ZZ*5) + *(SSS + zz + ZZ*6)) / (NSEX * *(SSS + zz + ZZ*7));  
  /* Sum Y.h.*2  [h-h] */  
  *(SSS + 11) += square(*(SSS + zz + ZZ*3) + square(*(SSS + zz + ZZ*5)) + square(*(SSS + zz + ZZ*4)) + square(*(SSS + zz + ZZ*6))) / *(SSS + zz + ZZ*7);  
  /* Sum Y.h.*2  [h-h] */  
}
Nmark += (*(SSA + zz + ZZ*7) > 0.0) ? 1.0 : 0.0;  
  /* total number of alleles at molecular marker locus zz */

}  
  /* next allele within marker */
/* Calculate SUM OF SQUARES from RAW SUM SQUARES */

*(SUMA + 0) = *(SSS + 6) - *(RSA + 0);  
  /* SS M [homo] */  
*(SUMA + 1) = *(SSS + 7) - *(RSA + 1) - *(SSS + 6) + *(RSA + 0);  
  /* SS SSM [homo] */  
*(SUMA + 2) = *(SSS + 0) - *(SSS + 6);  
  /* SS LM [homo] */  
*(SUMA + 3) = *(SSS + 1) - *(SSS + 0) - *(SSS + 7) + *(SSS + 6);  
  /* SS SLM [homo] */  
*(SUMA + 4) = *(SSS + 10) - *(SSS + 8) - *(RSA + 4) - *(RSA + 2);  
  /* SS MC [h-h] */  
*(SUMA + 5) = *(SSS + 11) - *(SSS + 9) - *(RSA + 5) - *(RSA + 3) + *(SSS + 8) + *(RSA + 4) - *(RSA + 2);  
  /* SS SSMC [h-h] */  
*(SUMA + 6) = *(SSS + 4) - *(SSS + 10) - *(SSS + 5) + *(SSS + 8);  
  /* SS CLM [h-h] */  
*(SUMA + 7) = *(SSS + 3) - *(SSS + 11) - *(SSS + 2) - *(SSS + 4) + *(SSS + 5) + *(SSS + 10) + *(SSS + 9) - *(SSS + 8);  
  /* SS SCLM [h-h] */  

/* Calculate F statistics, p values, and tabulate results */

/* abdominal F statistics */
/* F for effect of marker */
*(FA + 0) = (*SUMA + 0) * (NLINE - Nmark) / (*SUMA + 2) * (Nmark - 1.0));  
/* F for effect of marker*sex */
*(FA + 1) = (*SUMA + 1) * (NLINE - Nmark) / (*SUMA + 3) * (Nmark - 1.0));  
/* F for effect of marker*cross */
*(FA + 2) = (*SUMA + 4) * (NCROSS - 1.0) * (NLINE - Nmark) / (*SUMA + 6) * (Nmark - 1.0));  
/* F for effect of marker*sex*cross */
*(FA + 3) = (*SUMA + 5) * (NLINE - Nmark) / (*SUMA + 7) * (Nmark - 1.0));  

/* sternopleural F statistics */
/* F for effect of marker */
*(FS + 0) = (*SUMS + 0) * (NLINE - Nmark) / (*SUMS + 2) * (Nmark - 1.0));  
/* F for effect of marker*sex */
*(FS + 1) = (*SUMS + 1) * (NLINE - Nmark) / (*SUMS + 3) * (Nmark - 1.0));  
/* F for effect of marker*cross */
*(FS + 2) = (*SUMS + 4) * (NCROSS - 1.0) * (NLINE - Nmark) / (*SUMS + 6) * (Nmark - 1.0));  
/* F for effect of marker*sex*cross */
*(FS + 3) = (*SUMS + 5) * (NLINE - Nmark)) / (*SUMS + 7) * (Nmark - 1.0));

/* P values */

/* MINIMIZE CALLS TO F SUBROUTINE */

*(FA + 0) = pof(*(FA + 0), Nmark - 1.0, NLINE - Nmark);
*(FA + 1) = pof(*(FA + 1), Nmark - 1.0, NLINE - Nmark);
*(FA + 2) = pof(*(FA + 2), Nmark - 1.0, NLINE - Nmark);
*(FA + 3) = pof(*(FA + 3), Nmark - 1.0, NLINE - Nmark);
*(FS + 0) = pof(*(FS + 0), Nmark - 1.0, NLINE - Nmark);
*(FS + 1) = pof(*(FS + 1), Nmark - 1.0, NLINE - Nmark);
*(FS + 2) = pof(*(FS + 2), Nmark - 1.0, NLINE - Nmark);
*(FS + 3) = pof(*(FS + 3), Nmark - 1.0, NLINE - Nmark);

/* save results if not jack-knifing */
if (Print == 1){

  fprintf(out2,"%d\n%g\n%g\n%g\n%g\n%g\n%g\n%g
",
      *(mark_name + mark), *(FS + 0), *(FS + 2), *(FS + 3), *(FA + 0), *(FA + 2), *(FA + 1), *(FA + 3));

/* print out means by marker */

  if (Nmark > 2.0)
    Zmax = (int) Nmark;
  else
    Zmax = 2;
  for (zz = 0; zz < Zmax; zz++){
    fprintf(out3,"%d\n%g\n%g\n%g\n%g\n%g\n%g\n%g
",
            *(mark_name + mark), zz, *(SSA + zz + ZZ*7));
    for(j = 0; j < 6; j++)
      test = *(SSA + zz + ZZ*(j+1)) / *(SSA + zz + ZZ*7);
      if(test < 0.00001 && test > - 0.00001){
          fprintf(out3,"0.00000\n");
      }
      else{
          fprintf(out3,"%g\n",test);
      }
    for(j = 0; j < 6; j++)
      test = *(SSS + zz + ZZ*(j+1)) / *(SSS + zz + ZZ*7);
      if(test < 0.00001 && test > - 0.00001){
          fprintf(out3,"0.00000\n");
      }
      else{
          fprintf(out3,"%g\n",test);
      }
    fprintf(out3,"\n");
  }
}

/*
  this section of the function keeps track of the number of tests significant
  grouped by interesting criteria. In the end 6 columns are printed out. The numbers
  in these columns are:
  1. the number of markers (m) and markers*sex (m*s) significant at p < 0.05 for the
     homozygous data set for both AB and SB
  2. same as (1.) but at p < 0.01
*/
3. the number of markers*cross [m*c] and markers*cross*sex [m*c*s] significant at p < 0.05 for the hemizygous and heterozygous data sets for both AB and SB
4. same as [3.] but at p < 0.01
5. the number of m [from homo's] and m*c [form hemi, hetero's] for which both were significant at at least p < 0.10 for both characters (note: effects including sex excluded)
6. same as [5.] but at p < 0.05
/

/*
number of p < 0.05 and p < 0.01 for ABs and SBs
for homozygous data set ONLY
--- note it is relatively straight forward to keep track of abs and sb separately
*/

*(Nsig + 0) += (* (FA + 0) <= 0.050) ? 1 : 0;
*(Nsig + 1) += (* (FA + 0) <= 0.010) ? 1 : 0;
*(Nsig + 0) += (* (FA + 1) <= 0.050) ? 1 : 0;
*(Nsig + 1) += (* (FA + 1) <= 0.010) ? 1 : 0;
*(Nsig + 0) += (* (FS + 0) <= 0.050) ? 1 : 0;
*(Nsig + 1) += (* (FS + 0) <= 0.010) ? 1 : 0;
*(Nsig + 0) += (* (FS + 1) <= 0.050) ? 1 : 0;
*(Nsig + 1) += (* (FS + 1) <= 0.010) ? 1 : 0;

/*
number of p < 0.05 and p < 0.01 for ABs and SBs
for hemizygous versus heterozygous data set ONLY
--- note it is relatively straight forward to keep track of abs and sb separately
*/

*(Nsig + 2) += (* (FA + 2) <= 0.050) ? 1 : 0;
*(Nsig + 3) += (* (FA + 2) <= 0.010) ? 1 : 0;
*(Nsig + 2) += (* (FA + 3) <= 0.050) ? 1 : 0;
*(Nsig + 3) += (* (FA + 3) <= 0.010) ? 1 : 0;
*(Nsig + 2) += (* (FS + 2) <= 0.050) ? 1 : 0;
*(Nsig + 3) += (* (FS + 2) <= 0.010) ? 1 : 0;
*(Nsig + 2) += (* (FS + 3) <= 0.050) ? 1 : 0;
*(Nsig + 3) += (* (FS + 3) <= 0.010) ? 1 : 0;

/*
number of p < 0.10 and p < 0.05 significant for in both the homo [marker] and hemi vs. hetero [marker*cross] for abs and sbs
ALL sex cross effect are excluded here
*/

*(Nsig + 4) += (* (FA + 0) <= 0.10 && * (FA + 2) <= 0.10) ? 1 : 0;
*(Nsig + 5) += (* (FA + 0) <= 0.050 && * (FA + 2) <= 0.050) ? 1 : 0;
*(Nsig + 4) += (* (FS + 0) <= 0.10 && * (FS + 2) <= 0.10) ? 1 : 0;
*(Nsig + 5) += (* (FS + 0) <= 0.050 && * (FS + 2) <= 0.050) ? 1 : 0;

} /* next molecular marker */
free(FA);
free(FS);
free(SSA);
free(SSS);
free(SSSA);
free(SSSS);
free(RSA);
free(RSS);
free(TTA);
free(TTS);
free(SUMA);
free(SUMS);
cfclose(out2);
cfclose(out3);
return;
}

*******************************************************************************
* Remove_effect: will remove the average sex*marker effect of the
* marker specified in the first argument of the call
 *
* routine to remove marker by sex by cross effects
* rneff is the number of the marker to be removed
* the number must be 1 (NOT ZERO) to NMARK and correspond to the
* column number of the marker NOT any other designation
* rneff must be set equal to zero if no effect is to be dropped
 *
* Tony Long: March 1994
*******************************************************************************

void Remove_effect(int rneff, int N_line, int N_mark, double *MA, double
*MS, int *MM)
{
int i, j, line, zz;
double *SSA, *SSS;

SSA = (double *) calloc(8*ZZ, sizeof(double));
SSS = (double *) calloc(8*ZZ, sizeof(double));

/* initialize counters */

for (i = 0; i < B; i++)
    for (j = 0; j < ZZ; j++)
        *(SSA + j + i*ZZ) = 0.0;
    *(SSS + j + i*ZZ) = 0.0;
}

/* calculate sums for each marker * sex * cross */

for (line = 0; line < N_line; line++)
{

zz = *(MM + rneff - 1 + line*MARK);

*(SSA + zz + ZZ*1) += *(MA + line*SI6);
*/ XMj. (homo) */

*(SSA + zz + ZZ*2) += *(MA + 1 + line*SI6);
*/ XFj. (homo) */

*(SSA + zz + ZZ*3) += *(MA + 2 + line*SI6);
*/ XMjHTk [h-h] */

*(SSA + zz + ZZ*4) += *(MA + 3 + line*SI6);
*/ XFjHTk [h-h] */

*(SSA + zz + ZZ*5) += *(MA + 4 + line*SI6);
*/ XMjHMk [h-h] */

*(SSA + zz + ZZ*6) += *(MA + 5 + line*SI6);
*/ XFjHMk [h-h] */

*(SSA + zz + ZZ*7) += 1.0;
*/ freq of allele zz */

*(SSS + zz + ZZ*1) += *(MS + line*SI6);
*/ XMj. (homo) */

*(SSS + zz + ZZ*2) += *(MS + 1 + line*SI6);
*/ XFj. (homo) */

*(SSS + zz + ZZ*3) += *(MS + 2 + line*SI6);
*/ XMjHTk [h-h] */

*(SSS + zz + ZZ*4) += *(MS + 3 + line*SI6);
*/ XFjHTk [h-h] */

*(SSS + zz + ZZ*5) += *(MS + 4 + line*SI6);
*/ XMjHMk [h-h] */

*(SSS + zz + ZZ*6) += *(MS + 5 + line*SI6);
*/ XFjHMk [h-h] */

*(SSS + zz + ZZ*7) += 1.0;
*/ freq of allele zz */
 knocking effect of marker rmeff from MA and MS matrices */

for (i = 0; i < N_line; i++) {
    for (j = 0; j < SIX; j++) {
        zz = *(MM + rmeff - 1 + i*MARK);
        *(MA + j + i*SDX) = *(SSA + zz + ZZ*(j+1))/(*(SSA + zz + ZZ*7);
        *(MS + j + i*SDX) = *(SSS + zz + ZZ*(j+1))/(*(SSS + zz + ZZ*7);
    }
}

/* end routine to remove effect of marker rmeff */
free(SSA);
free(SSS);
return;
}

******************************************************************************
* VAR_add: this subroutine calculates the 'additive' variance attributable to
* a defined molecular marker and send it back to the main program. It calculates:
* Va(m AB), Va(f AB), Va(m SB), Va(f SB). It will print the actual values to a file
* [scab_VA.out] but send back a pointer for shuffled reps. Va is estimated as:
* Va = 0.5*SUM[pi(Xi - u)^2]
* where u = SUM[piXi], Xi is the mean bristle lines over all lines of molecular
* marker type i and pi is the simple frequency of that molecular marker
* from the homozygous data. It is assumed that heterozygotes are intermediate (i.e., the
* trait is additive).
* 
* Tony Long: March 1994
******************************************************************************/

void VAR_add (int noise, int boot, int N_line, int N_mark, double *MA,
              double *MS, int *MM, double *V)
{
    int i, j, line, zz;
    double *SSA, *SSS, TTA1, TTA2, TTS1, TTS2;
    FILE *out4;

    out4 = fopen("scab_VA.out", "a");
    SSA = (double *) calloc(3*ZZ, sizeof(double));
    SSS = (double *) calloc(3*ZZ, sizeof(double));

    /* initializing counters */

    for (i = 0; i < 4; i++) {
        *(V + i) = 0.0;
    }

    for (i = 0; i < 3; i++) {
        for (j = 0; j < ZZ; j++) {
            *(SSA + j + i*ZZ) = 0.0;
            *(SSS + j + i*ZZ) = 0.0;
        }
    }

    TTA1 = 0.0;
    TTA2 = 0.0;
    TTS1 = 0.0;
    TTS2 = 0.0;
for (line = 0; line < N_line; line++)
{
    TTA1 <= *(MA + line*SIX);
    /* Sum Y1_ [homo] sex, mark, line */
    TTA2 <= *(MA + 1 + line*SIX);
    /* Sum Y2_ [homo] */
    TTS1 <= *(MS + line*SIX);
    /* Sum Y1_ [homo] sex, mark, line */
    TTS2 <= *(MS + 1 + line*SIX);
    /* Sum Y2_ [homo] */

    zz = *(MM + boot - 1 + line*N_mark);
    *(SSA + zz + ZZ*0) += *(MA + line*SIX);
    /* XMj. [homo] */
    *(SSA + zz + ZZ*1) += *(MA + 1 + line*SIX);
    /* XFj. [homo] */
    *(SSA + zz + ZZ*2) += 1.0;
    /* freq of allele zz */
    *(SSS + zz + ZZ*0) += *(MS + line*SIX);
    /* XMj. [homo] */
    *(SSS + zz + ZZ*1) += *(MS + 1 + line*SIX);
    /* XFj. [homo] */
    *(SSS + zz + ZZ*2) += 1.0;
    /* freq of allele zz */
}
/
for (z = 0; z < ZZ; z++)
{
    if(*(SSA + zz + ZZ*2) > 0.0)
    {
        *(V + 0) += square(*(SSA + zz + ZZ*0)/(*(SSA + zz + ZZ*2)))*(*(SSA + zz + ZZ*2)/(double) N_line);
        *(V + 1) += square(*(SSA + zz + ZZ*1)/(*(SSA + zz + ZZ*2)))*(*(SSA + zz + ZZ*2)/(double) N_line);
        *(V + 2) += square(*(SSS + zz + ZZ*0)/(*(SSS + zz + ZZ*2)))*(*(SSS + zz + ZZ*2)/(double) N_line);
        *(V + 3) += square(*(SSS + zz + ZZ*1)/(*(SSS + zz + ZZ*2)))*(*(SSS + zz + ZZ*2)/(double) N_line);
    }
    if(*(SSA + zz + ZZ*2) == (double) N_line - 1.0 || *(SSA + zz + ZZ*2) == (double) N_line)
    {
        printf(out4,"The Freq of one Allele is %d\n",*(SSA + zz + ZZ*2));
    }
    /
}
*(V + 0) = (* (V + 0) - square(TTA1/(double) N_line)) * 0.5;
*(V + 1) = (* (V + 1) - square(TTA2/(double) N_line)) * 0.5;
*(V + 2) = (* (V + 2) - square(TTS1/(double) N_line)) * 0.5;
*(V + 3) = (* (V + 3) - square(TTS2/(double) N_line)) * 0.5;

if (noise == 0)
{
    printf(out4,"%,%g,%v,%v,%v,%v",*(V + 0), *(V + 1), *(V + 2), *(V + 3));
}
fclose(out4);
free(SSA);
free(SSS);
return;
}

/**************************************************************************
 * This is the main segment of the program designed to assess the significance of
 * various associations between molecular markers and quantitative traits. Currently
 * the program is not very general -- it only can handle one statistical design,
 * two characters, and three different 'types' of measures on the characters (see
 * description of function ANOVA). A second problem is that the input and output files
 * only have 'hard-coded' names. Hopefully this will also be [easily] addressed in future
 * versions of the program.
 * 
 * Tony Long: April 1994
 * 
 * INPUT
 * 
 * The program will ask for a number of variables to be inputed and depending on the
 * values supplied will perform different analyses. Since the input comes from the
STDIN the program can be used interactively or with pipes on a unix based machine
(this can be particularly useful as many analyses are computer intensive).
Specifically:
Number of REPS to Carry Out: The value supplied here will be used to determine
   EITHER the number of permutation tests to carry out in order to assess the locus
   wide significance of the analysis OR the number of bootstrap replicates to carry
   out in order to get a distribution on an estimate of Va for a given molecular marker.
RANK Number of a MARKER to Remove from Analysis: The value supplied here allows one to
statistically remove the effect of a marker and then reanalyse the data. The number
supplied must correspond to the rank number of the marker in the list of molecular
markers NOT the 'name' of the marker one wishes to remove. This is
useful if one wishes to consider the significance of the data, the effect of a marker,
or the variance contributed by a marker CONDITIONAL on the effect of another marker.
More specifics of how this is accomplished is supplied in the Remove effect function.
N, where N will remove the Last N markers form consideration: The value supplied here
allows one to remove a series of N markers form the analysis. Nothing fancy here, the
program just never reads in these LAST N molecular markers and hence they are removed
from any analyses. This function is useful if the last few molecular markers are
'synthetic' -- for example a haplotype constructed from a number of other markers or
'all large inserts'.
RANK number of a MARKER to bootstrap Va for: The value supplied here allows one to
specify the molecular marker one wishes a set of bootstrap estimates of Va for. These
set of Va estimates serve the purpose of allowing one to put a confidence interval
on the point estimate of Va (achieved using this option with REPS set to 0). The
Significance of Va is assessed by determining the number of estimates of Va greater
than zero. The comments in the function VAR_add explains how Va is estimated and below
under OUTPUT explain how actual estimates of Va are made (in particular
how one can get NEGATIVE estimates of Va!!).
QUIT?: O for no and 1 for yes. These parameter values must be provided, even in piped
UNIX jobs.

OUTPUT
Four output files are created. In future versions these will not be 'hard-wired'.

scab_permute.out: Output for the permutation testing of the locus wide significance
of the results. A header is printed giving the input parameters the analysis is for
and then 6 columns: 1) the number of markers [m] and markers*sex [m*s] significant
at p < 0.05 for the homozygous data set for both AB and SB; 2) same as (1.) but
at p < 0.01; 3) the number of markers*cross [m*c] and markers*cross*sex [m*c*s]
significant at p < 0.05 for the hemizygous and heterozygous data sets for both AB
and SB; 4) same as (3.) but at p < 0.01; 5) the number of m [from homo's] and
m*c [from hemi, hetero's] for which both were significant at least p < 0.10 for
both characters (note: effects including sex excluded; and 6) same as (5.) but at
p < 0.05.
scab_pvalues.out: Outputs the p-values associated with every molecular marker tested. Nothing
is outputed unless REP=0. Columns are: 1) marker name, 2) SB homo marker, 3) SB homo
marker*sex, 4) SB hemi/het cross*marker, 5) SB hemi/het cross*marker*sex, 6 through 9
are the same as 2 through 5 but for ABS.
scab_means.out: Outputs the means for analyses in which REP = 0. Columns are: 1)
marker name, 2) allele value (i.e., 0, 1, etc.), 3) number of lines having allele value
in (2), 4) AB homo male mean, 5) AB homo female mean, 6) AB hemi male mean, 7) AB hemi
female mean, 8) AB hetero male mean, 9) AB hetero female mean, 10 through 15 are the
same as 4 through 9 but for SB's
scab_VA.out: Outputs the estimates of Va associated with a given marker (Note: header
gives the rank of the marker for which the bootstrapping is carried out not the marker
name. Columns are: 1) Va male ABs, 2) Va female ABs, 3) Va male SBs, 4) Va female
SBs, 5) Va male ABs, 6) Va female ABs, 7) Va male SBs, 8) Va female SBs. It is important
to note that the first four estimates are the actual estimates of Va based on the actual
data [REP = 0] OR a bootstrap sample of the data [REP > 0], and the last four values are
averages based on BULL shufflings of either the actual data [REP = 0] or a particular bootstrap
sample [REP > 0]. BULL is defined to be equal to 100 but can be changed. Thus, the
last set of four values are the estimates of Va under the hypothesis of no effect of the
marker and can be subtracted from the first four (1 - 5, 2 - 6, etc.) to get unbiased
estimates of the additive variance contributed by a given molecular marker. Further notes
provided in the VAR_add function.

HAVE FUN, BE YOUNG, COUNT BRISTLES

**********************************************

main(){

int rep, mark, line, i, j, Idummy, marboot, noise, bull;
int REP, rmeff, lastN, QUIT;
int *MSttrue, *MM, *Nsig, *mark_name;

double crap;

MA = (double *) calloc(NLINEI*SIX, sizeof(double));
MS = (double *) calloc(NLINEI*SIX, sizeof(double));
MAttrue = (double *) calloc(NLINEI*SIX, sizeof(double));
MSttrue = (double *) calloc(NLINEI*SIX, sizeof(double));
MM = (int *) calloc(NLINEI*MARK, sizeof(int));
MSttrue = (int *) calloc(NLINEI*MARK, sizeof(int));
mark_name = (int *) calloc(MARK, sizeof(int));
Nsig = (int *) calloc(6, sizeof(int));
Va = (double *) calloc(4, sizeof(double));
VVa = (double *) calloc(4, sizeof(double));

out = fopen("scab_permute.out","a");
QUIT = 0;
xseed = XSEED;
while (QUIT == 0){

printf("Enter Number of REPS to Carry Out [ 0 if not doing\n permutation tests or bootstrapping Va ]nn");
scanf("%d",&REP);
printf("Enter RANK Number of a MARKER to Remove from Analysis [ 0 if raw data ]nn");
scanf("%d",&rmeff);
printf("Enter N: where N will remove the LAST N markers from consideration\n when jack-knifing [O if all markers are to be considered]\n");
scanf("%d",&lastN);
printf("Enter RANK Number of a MARKER to bootstrap Va for [ 0 if no marker ]n");
printf("Note: you must enter zero here if doing permutation tests of p-values\n");
scanf("%d",&marboot);
fprintf(out, "\nThe Following Analysis has Marker %d Removed\n",rmeff);
fprintf(out, "\nThe following Analysis is Bootstrapped for Marker %d\n",marboot);
out2 = fopen("scab_pvalues.out","a");
out3 = fopen("scab_means.out","a");
out4 = fopen("scab_VAs.out","a");
fprintf(out2, "\nThe Following Analysis has Marker %d Removed\n",rmeff);
fprintf(out3, "\nThe Following Analysis has Marker %d\n",rmeff);
fprintf(out4, "\nThe Following Analysis has Marker %d\n",rmeff);
fprintf(out4, "\nThe following Analysis is Bootstrapped for Marker %d\n",marboot);
fclose(out2);
fclose(out3);
fclose(out4);
/* Read in data */
    Read_data(NLINE, MARK, SIX, MAtrue, MStrue, mark_name, MMtrue);

/* Remove effects if requested */
    if (rmeff != 0){
        Remove_effect(rmeff, NLINEI, MARK, MAtrue, MStrue, MMtrue);
    }

/* Routines for REP = 0: actual p-values, means, and requested Va estimates */
    if (REP == 0){

/* Do a significance test on each marker and output results plus means */
        if (marboot == 0){
            ANOVA(1, NLINEI, MARK, lastN, MAtrue, MStrue, MMtrue, mark_name, Nsig);
            fprintf(out,"%dV%dV%dV%dV%dV%dV", *(Nsig + 0), *(Nsig + 1), *(Nsig + 2), *(Nsig + 3), *(Nsig + 4), *(Nsig + 5));
        }

/* Get an estimate of Va associated with a given marker */
        else{
            for (noise = 0; noise < 2; noise++){
                if (noise == 0){
                    VAR_add(noise, marboot, NLINEI, MARK, MAtrue, MStrue, MMtrue, Va);
                }
            }

/* 'Noise' associated with estimate of Va (averaged over BULL reps) */
        else{
            for (i = 0; i < 4; i++){
                *(VVa + i) = 0.0;
            }
            for (bull = 0; bull < BULL; bull++){
                Shuff(2, NLINE, MARK, SIX, MAtrue, MStrue, MMtrue);
                VAR_add(noise, marboot, NLINEI, MARK, MAtrue, MStrue, MMtrue, Va);
                for (i = 0; i < 4; i++){
                    *(VVa + i) += (1.0 / (double) BULL) * *(Va + i);
                }
            }
            out4 = fopen("scab_VA.out","a");
            fprintf(out4,"%gV%gV%gV%gV%gV", *(Va + 0), *(Va + 1), *(Va + 2), *(Va + 3));
            fclose(out4);
        }

/* Routines for REP > 0: bootstrapped p-values, and bootstrapped Va estimates */
        if (REP != 0){
            for (rep = 0; rep < REP; rep++){

/* set true matrices equal to experimental matrices */
                for (i = 0; i < NLINEI; i++){
                    for (j = 0; j < MARK; j++){
                        *(MM + j + i*MARK) = *(MMtrue + j + i*MARK);
                    }
                    for (j = 0; j < SIX; j++){
                        *(MA + j + i*SIX) = *(MAtrue + j + i*SIX);
                        *(MS + j + i*SIX) = *(MStrue + j + i*SIX);
                    }
                }
            }
        }
}
/* Permute haplotypes w.r.t. phenotypic data and estimate number of tests significant */
if (marboot == 0)
  Shuff(2, NLINE, MARK, SIX, MA, MS, MM);
  ANOVA (0, NLINEI, MARK, lastN, MA, MS, MM, mark_name, Nsig);
  fprintf(out,"%d%d%d%d%d%d",*(Nsig + 0), *(Nsig + 1), *(Nsig + 2), *(Nsig + 3), *(Nsig + 4), *(Nsig + 5));
}

/* Do REP bootstrap reps of the data set estimating Va for each one */
else
  for(noise = 0; noise < 2; noise++)
    
/* Actual estimate of Va (for this bootstrap rep) */
if (noise == 0)
  Shuff(1, NLINE, MARK, SIX, MA, MS, MM);
  VAR_add(noise, marboot, NLINEI, MARK, MA, MS, MM, Va);
else
  for (i = 0; i < 4; i++)
    *(VVa + i) = 0.0;

/* 'Noise' associated with this bootstrap rep (averaged over BULL reps) */
for (bull = 0; bull < BULL; bull++)
  Shuff(2, NLINE, MARK, SIX, MA, MS, MM);
  VAR_add(noise, marboot, NLINEI, MARK, MA, MS, MM, Va);
  for (i = 0; i < 4; i++)
    *(VVa + i) += (1.0 / (double) BULL) * *(Va + i);

out4 = fopen("scab_VA.out","a");
fprintf(out4,"%g%g%g%g%g%g",*(Va + 0), *(Va + 1), *(Va + 2), *(Va + 3));
fclose(out4);

}

/* Do You wish to Quit?? [1 if Yes, O otherwise]n*/
scanf("%d",&Idummy);
QUIT = Idummy;
}
free(MA);
free(MS);
free(MAtrue);
free(MStrue);
free(MM);
free(MMtrue);
free(mark_name);
free(Nsig);
free(Va);
free(Va);
fclose(out);
The following is a sample run of the program designed to estimate quantitative parameters associated with candidate loci. Three things are provided: 1) an example input file, 2) example of the running of the program as seen at the terminal (your screen) for an interactive run, and 3) the four files created as output from the program. Information generated by the computer (terminal or file) and the input file is in 9 point courier font, information supplied by the user during the interactive session (or as a unix pipe) is in 9 point courier italic, file names are in 12 point courier bold, and all comments are in 12 point times. This is to avoid confusion with the output of the program, the description of the program, and user supplied components of the program.

The input file is presented below and consists of the four components described in the Read_data function of the program. The mean AB bristle numbers for the first line for male homozygotes, female homozygotes, male hemizygotes, female hemizygotes, male heterozygotes and female heterozygotes followed by the same for the other lines (there should be NLINEs in all). The number of lines is 'hard wired' and will be easily (it is a defined constant) changed in future versions. Next there are NLINEs of data the same as above but for SB bristle numbers. Note that the program does not have a problem with blank spaces between the AB and SB data. The following line is a list of all the molecular marker names -- unfortunately at present these names must be integer variables! Finally there are NLINE sets of haplotypes. That is the marker genotype for each line at each of the molecular markers. These must also be integer variables, but DO NOT have to be only 0 and 1. In this data set the last few columns are synthetic variables that represent the haplotypes created by combining significant markers. It is important to note that although the values in this section of the input file are only limited to being integer, THE VALUE OF ZZ IN THE DEFINE SECTION OF THE PROGRAM must reflect the largest integer plus one used in this section (NOT THE MAXIMUM NUMBER OF ALLELES). Thus keeping the numbers small in this section reduces the memory required to run the program.

scab_in.dat

| 17 | 18.1 | 17.1 | 17.95 | 16.75 | 17.35 |
| 16.8 | 17.6 | 16.05 | 17.65 | 16.45 | 17.2 |
| 16.75 | 18.85 | 16.2 | 18.25 | 15.6 | 18.45 |
| 15.35 | 17.3 | 16.05 | 17.55 | 16.05 | 16.6 |
| 13.75 | 15.7 | 16.45 | 18 | 16.15 | 18 |
| 16.3 | 17.45 | 16.8 | 16.8 | 16.9 | 17.2 |
| 13.6 | 17.55 | 15.5 | 17.2 | 14.95 | 17.5 |
| 16.1 | 18.3 | 15.95 | 17.25 | 17.15 | 18.4 |
| 15.4 | 15.8 | 16.25 | 17.65 | 16.55 | 16.5 |
| 14.4 | 17.25 | 15.95 | 17.05 | 15.5 | 17.3 |
| 13.85 | 16.55 | 16.2 | 17.8 | 16.7 | 17.7 |
| 13.3 | 15.7 | 15.65 | 16.7 | 13.8 | 16.65 |
| 16.9 | 17.8 | 17.95 | 19.8 | 17.15 | 19.6 |
| 17.15 | 19.55 | 17.05 | 19 | 16.95 | 18.6 |
| 20.25 | 21.15 | 17.4 | 18.75 | 16.85 | 19 |
| 15.35 | 17.4 | 16.9 | 16.75 | 16.7 | 15.6 |
| 15.1 | 18 | 16 | 18.25 | 16.45 | 18.05 |
| 13.2 | 14.25 | 14.35 | 15.75 | 13.5 | 16.5 |
| 16.25 | 17.95 | 16.1 | 17.75 | 14.65 | 17.2 |
| 14.3 | 14.8 | 15.25 | 16.45 | 14.65 | 17.35 |
| 15.85 | 17.1 | 15.95 | 16.6 | 15.8 | 16.7 |
| 18.55 | 19.1 | 99 | 99 | 99 | 99 |
| 15.65 | 15.7 | 14.7 | 14.6 | 15.5 | 16.05 |
| 15.75 | 17.35 | 99 | 99 | 99 | 99 |
| 17.1 | 17.55 | 17.65 | 19.1 | 16.9 | 17.65 |
| 13.95 | 15.75 | 15.2 | 16.85 | 14.95 | 16.65 |
| 16.2 | 15.85 | 16.95 | 17.4 | 16.35 | 17.1 |
| 14.7 | 15.15 | 16.0 | 16.95 | 15.7 | 16.1 |
| 15.5 | 16.45 | 15.95 | 17.55 | 16.45 | 17.25 |
| 15.4 | 16.7 | 14.6 | 16.4 | 15.9 | 16.7 |
| 16.2 | 16.8 | 19.3 | 20.35 | 16.8 | 17.7 |
| 17.45 | 17.15 | 16.1 | 17.65 | 16.1 | 17.35 |
| 17.9 | 18.45 | 15.3 | 17.65 | 16.45 | 17.6 |

| 2 | 3 | 4 | 5 | 7 | 8 | 9 | 11 | 15 | 18 | 19 | 20 | 21 | 23 | 24 | 25 | 28 | 32 | 34 | 37 | 38 | 40 | 41 | 43 | 44 | 45 | 46 | 99 | 100 | 101 | 102 |
| 000011001020111110010100011101011101111 |
| 00001010010010010101111001000010111101110 |
| 00001001001001011110011000011111120111 |
| 000010010111011001110111112502112 |
| 000010110100100010110001101103113 |
| 0010000010010100111001011010100110353 |
| 10001010100011011001110111011501533 |
| 00000001011001000101100111001004330415 |
| 100000111000010110011010010010032202112 |
| 1000000001001000101000110111111111 |
| 0000000001000010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
| 0000000000100010011001100111111111 |
Below I outline a run of the program which demonstrates its utility. The following is the terminal session created from an interactive run of the program. In order to run the program in the batch mode on a unix machine it would be invoked with a command like:

```
 a.out <parameters.in >screen_messages.out
```

`parameters.in` would contain the program parameters in italics listed below and `a.out` is an executable version of the program. The program is totally portable and this method is highly recommended as runs using realistic bootstrap parameters (e.g., REP = 1000) can take a great deal of time on my (admittedly slow) desktop MAC. For space reasons, I only show the terminal session output for two of the sets of parameters used. Below I provide a table of the entire set of parameter values used in the following example:

<table>
<thead>
<tr>
<th>parameter set</th>
<th>Number of Replicates</th>
<th>Rank Mark to Remove</th>
<th>Last N to Remove</th>
<th>Rank Mark to Estimate Va for</th>
<th>QUIT??</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

In the example I do not remove the effect of the last few 'synthetic' molecular markers as would be appropriate for doing the permutation test on the significance of the data. In addition I have only used 10 bootstrap replicates to bootstrap estimates of Va and do the permutation test on the data, clearly a larger number would be appropriate here. I have chosen to examine RANK marker number 25 and 6, these are SSCP 1447 and Del E (-2.6) respectively in the scabrous experiment, and were both described as having a significant effect on abdominal bristle number in the paper (i.e., Table 2). In the output sections I describe more fully the results based on these markers.

Enter Number of REPS to Carry Out [ 0 if not doing permutation tests or bootstrapping Va ]
0
Enter RANK Number of a MARKER to Remove from Analysis [ 0 if raw data ]
0
Enter N, where N will remove the LAST N markers from consideration when jack-knifing [0 if all markers are to be considered]
0
Enter RANK Number of a MARKER to bootstrap Va for [ 0 if no marker ]
Note: you must enter zero here if doing permutation tests of p-values
0
Do You wish to Quit?? [1 if Yes, 0 otherwise]
0

[...six sets of parameters deleted...]

Enter Number of REPS to Carry Out [ 0 if not doing permutation tests or bootstrapping Va ]
10
Enter RANK Number of a MARKER to Remove from Analysis [ 0 if raw data ]
6
Enter N: where N will remove the LAST N markers from consideration when jack-knifing [0 if all markers are to be considered]
0
Enter RANK Number of a MARKER to bootstrap Va for [ 0 if no marker ]
Note: you must enter zero here if doing permutation tests of p-values
25
Do You wish to Quit?? [ 1 if Yes, 0 otherwise]
1

The output from the program follows. I will present the output files in the order of `scab_pvalues.out`, `scab_means.out`, `scab_permute.out`, and `scab_VA.out` with comments within the output and at the end of each file's output. In many cases the output has been partially formatted to increase readability and some cases have been deleted to decrease the size of this example.

**scab_pvalues.out**

The Following Analysis has Marker 0 Removed

<table>
<thead>
<tr>
<th>Rank</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.723248</td>
</tr>
<tr>
<td>3</td>
<td>0.0302524</td>
</tr>
<tr>
<td>4</td>
<td>0.239307</td>
</tr>
<tr>
<td>5</td>
<td>0.221337</td>
</tr>
<tr>
<td>6</td>
<td>0.665679</td>
</tr>
<tr>
<td>7</td>
<td>0.653797</td>
</tr>
<tr>
<td>8</td>
<td>0.184717</td>
</tr>
<tr>
<td>9</td>
<td>0.913436</td>
</tr>
</tbody>
</table>

(...p-values associated with a number of markers deleted...)

<table>
<thead>
<tr>
<th>Rank</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>0.524165</td>
</tr>
<tr>
<td>45</td>
<td>0.831912</td>
</tr>
<tr>
<td>46</td>
<td>0.584965</td>
</tr>
<tr>
<td>99</td>
<td>0.931627</td>
</tr>
<tr>
<td>100</td>
<td>0.271189</td>
</tr>
<tr>
<td>101</td>
<td>0.036419</td>
</tr>
<tr>
<td>102</td>
<td>0.608258</td>
</tr>
</tbody>
</table>

The p-values associated with the test for every marker for parameter set A. The marker names are in the first column and subsequent columns are as described in the main routine of the program. Note that the marker ranked number 6 (R6) is labelled number 8 (M8)(the marker’s name), similarly the marker ranked number 25 (R25) is labelled number 44 (M44). It is apparent that both marker’s, M8 and M44 are significant (p <
0.014 and $p < 0.004$ respectively) for the effect of marker for abdominal bristle number (column 5).

It is possible that M8 and M44 are in linkage disequilibrium and really measure the same thing, so below I repeat the analysis after statistically removing the effect of M8 (R6) (Parameter set B)...

The Following Analysis has Marker 6 Removed

<table>
<thead>
<tr>
<th></th>
<th>0.709991</th>
<th>0.798882</th>
<th>0.753617</th>
<th>0.0233813</th>
<th>0.643204</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.607511</td>
<td>0.488589</td>
<td>0.399879</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0170452</td>
<td>0.314731</td>
<td>0.931822</td>
<td>0.38227</td>
<td>0.200452</td>
</tr>
<tr>
<td>4</td>
<td>0.0577689</td>
<td>0.848145</td>
<td>0.908412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.782022</td>
<td>0.589447</td>
<td>0.965789</td>
<td>0.315246</td>
<td>0.987366</td>
</tr>
<tr>
<td>6</td>
<td>0.720536</td>
<td>0.848642</td>
<td>0.919824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.16195</td>
<td>0.205711</td>
<td>0.803346</td>
<td>0.425118</td>
<td>0.555269</td>
</tr>
<tr>
<td>8</td>
<td>0.704849</td>
<td>0.989115</td>
<td>0.788556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.707512</td>
<td>0.233975</td>
<td>0.89480</td>
<td>0.0366384</td>
<td>0.84966</td>
</tr>
<tr>
<td>0.473239</td>
<td>0.771303</td>
<td>0.729669</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(...p-values associated with a number of markers deleted...)

<table>
<thead>
<tr>
<th></th>
<th>0.342305</th>
<th>0.284292</th>
<th>0.999449</th>
<th>0.6748.3</th>
<th>0.00455826</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>0.472326</td>
<td>0.989878</td>
<td>0.712178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>0.752399</td>
<td>0.361211</td>
<td>0.999682</td>
<td>0.448059</td>
<td>0.0137728</td>
</tr>
<tr>
<td>46</td>
<td>0.326568</td>
<td>0.99419</td>
<td>0.394307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>0.431748</td>
<td>0.0474689</td>
<td>0.999855</td>
<td>0.2949698</td>
<td>0.0597054</td>
</tr>
<tr>
<td>99</td>
<td>0.652779</td>
<td>0.379159</td>
<td>0.581701</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We can see that although M8 is not significant anymore, M44 still is ($p < 0.005$). Apparently, M8 and M44 are independent of one another. Further analyses (Parameter sets C through H) involved permutation testing OR Variance component estimates so p-value tables were not printed out, although indicators were.

The Following Analysis has Marker 0 Removed

The Following Analysis has Marker 6 Removed

The Following Analysis has Marker 0 Removed

The Following Analysis has Marker 0 Removed

The Following Analysis has Marker 6 Removed

The Following Analysis has Marker 0 Removed

**scab_means.out**

The Following Analysis has Marker 0 Removed
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15.5667</td>
<td>16.5333</td>
<td>16.6</td>
<td>17.45</td>
<td>15.04...</td>
<td>[... means for a number of markers deleted ...]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>44</th>
<th>0</th>
<th>ONAN04</th>
<th>NAN04</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>NAN04</th>
<th>ONAN04</th>
<th>NAN04</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>N...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16.3295</td>
<td>17.4841</td>
<td>16.0636</td>
<td>17.0909</td>
<td>44</td>
<td>2</td>
<td>5</td>
<td>16.65</td>
<td>17.97</td>
<td>15.67</td>
<td>17.39</td>
<td>15.58</td>
<td>17.71</td>
<td>16.4</td>
<td>17.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.76</td>
<td>17.08</td>
<td>15.79</td>
<td>16.99</td>
<td>44</td>
<td>3</td>
<td>8</td>
<td>16.1938</td>
<td>17.9562</td>
<td>26.6875</td>
<td>27.7688</td>
<td>26.3313</td>
<td>27.8</td>
<td>15.725</td>
<td>16.7875</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.9667</td>
<td>17.05</td>
<td>16.475</td>
<td>17.375</td>
<td>[... means for a number of additional markers deleted ...]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>102</th>
<th>0</th>
<th>ONAN04</th>
<th>NAN04</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>NAN04</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>N...</th>
<th>ONAN04</th>
<th>N...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>102</td>
<td>1</td>
<td>15</td>
<td>15.3367</td>
<td>16.5767</td>
<td>15.6633</td>
<td>16.9367</td>
<td>15.81</td>
<td>17.0933</td>
<td>15.7933</td>
<td>16.7447</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.9444</td>
<td>17.4556</td>
<td>16.1399</td>
<td>17.0778</td>
<td>102</td>
<td>4</td>
<td>4</td>
<td>14.325</td>
<td>15.2</td>
<td>15.6</td>
<td>17</td>
<td>15.5375</td>
<td>16.8125</td>
<td>15.7675</td>
<td>16.3622</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.0875</td>
<td>17.1375</td>
<td>16.7375</td>
<td>17.4375</td>
<td>Above are the means associated with Parameter set A. Like the case of the p-values, means are only printed out for Parameter set A and B (i.e., both REPS and ESTIMATE VA must be zero). We can see that for molecular markers with more than 2 alleles, the mean for each allele is printed out. In this particular data set the user did not use the value 0 as an indicator for the multi-allelic markers (i.e., he used 0 and 1 for restriction sites, but 1,..., n for multialelic loci), as a result the means for allele 0 for marker 102 are given as ONAN04, this is not a problem and merely reflects that this allele does not exist.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The means below are for the case were M8 was removed. Two things are important to note: 1) the effects did not change much [mean of allele 1 - mean of allele 2], but 2) the 'origin' did change. That is the means are all centered on zero now as opposed to the natural scale. In order to return the values back to their original scale the grand mean (by sex and 'cross') would have to be added back to each case.

The Following Analysis has Marker 6 Removed

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.08445</th>
<th>0.910701</th>
<th>-0.968845</th>
<th>-0.954616</th>
<th>-0.885156</th>
<th>-0.00679451</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.15</td>
<td>0.199394</td>
<td>0.106717</td>
<td>2.12894</td>
<td>2.03561</td>
<td>2.31879</td>
<td>2.04737</td>
</tr>
<tr>
<td>3</td>
<td>0.39</td>
<td>0.149767</td>
<td>0.11049</td>
<td>-1.36002</td>
<td>-1.34336</td>
<td>-1.25047</td>
<td>-1.22622</td>
</tr>
<tr>
<td>4</td>
<td>1.8</td>
<td>-0.730114</td>
<td>-0.538636</td>
<td>6.63011</td>
<td>6.54886</td>
<td>6.09602</td>
<td>5.97784</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.8426146.75511</th>
<th>6.5892</th>
<th>6.42216</th>
<th>6.41875</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.41</td>
<td>0.0179047-0.015160</td>
<td>0.287278</td>
<td>0.261669</td>
</tr>
<tr>
<td>6</td>
<td>0.0209534</td>
<td>0.318971</td>
<td>0.277772</td>
<td>0.278409</td>
</tr>
<tr>
<td>7</td>
<td>0.15</td>
<td>0.123406</td>
<td>-0.013598</td>
<td>-1.96307</td>
</tr>
</tbody>
</table>

0.143162-2.17348 | -1.89811 | -1.90246 | -1.84792|

|   | 0.40 | 0.0544860 | 0.507393 | -1.43938   | -1.40209   | -1.42602   | -1.37347     |
| 8  | 0.0813636 | -1.44011 | -1.42858   | -1.36622   | -1.38281   | -0.15764   |

|   | 0.974356 | -2.49322   | 8.16532    | 7.80698    | 7.90179    |
| 9  | 0.11 | -0.196594 | 0.0413912-3.05103 | -2.77376   | -3.17975   | -2.99325   |
| 10 | -0.1573 | -3.39573  | -3.06171   | -3.48326   | -3.46136   | 1.05764   |

|   | 0.36 | 0.0601010-0.0126473 | 0.93226   | 0.847558   | 0.971591   |
| 11 | 0.048064 | 1.03758   | 0.935522   | 1.06433    |

|   | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.000000   |
| 12 | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.000000   |

{... means for a number of markers deleted ...}

| 44 | 0   | 0.976941 | 0.96431   | -3.92841   | -3.9125    | -4.01364   | -4.08409   |
| 45 | 2   | 2.15 | 1.42091   | 1.25242   | -3.17091   | -2.74091   | -3.09818   |
| 46 | 3   | 2.25 | 1.24424   | 2.30503   | -3.42727   | -3.15      | -0.244606  |
| 47 | 4   | 2.15 | 0.582366  | 0.680114  | 7.01136    | 6.84261    | 6.75852    |
| 48 | 5   | 0.991136 | 6.28011   | 6.37045    | 6.67841    |
| 49 | 6   | 1.003535 | -2.46465   | -2.48081   | -2.30606   |

{... means for a number of additional markers deleted ...}

| 102 | 0   | 0.976941 | 0.96431   | -3.92841   | -3.9125    | -4.01364   | -4.08409   |
| 103 | 2   | 2.15 | 1.42091   | 1.25242   | -3.17091   | -2.74091   | -3.09818   |
| 104 | 3   | 2.25 | 1.24424   | 2.30503   | -3.42727   | -3.15      | -0.244606  |
| 105 | 4   | 2.15 | 0.582366  | 0.680114  | 7.01136    | 6.84261    | 6.75852    |
| 106 | 5   | 0.991136 | 6.28011   | 6.37045    | 6.67841    |
| 107 | 6   | 1.003535 | -2.46465   | -2.48081   | -2.30606   |

| 0.310985 | -1.62765 | -1.64394 | -1.4983   | -1.69375 |
The Following Analysis has Marker 0 Removed
The Following Analysis has Marker 6 Removed
The Following Analysis has Marker 0 Removed
The Following Analysis has Marker 6 Removed
The Following Analysis has Marker 0 Removed
The Following Analysis has Marker 6 Removed

**scab_permute.out**

The Following Analysis has Marker 0 Removed
The following Analysis is Bootstrapped for Marker 0

11  3  11  1  4  2

The Following Analysis has Marker 6 Removed
The following Analysis is Bootstrapped for Marker 0

8  2  10  1  2  1

The output for **scab_permute.out** gives the NUMBER of tests significant at different p-values and in different statistical tests (see comments in main routine of program). It is apparent that removing the effect of M8 (rank marker 6) results in the loss of some of the total number of significant sites and some of the pairs significant. Note that because the data set includes a few 'synthetic' molecular markers this example is only illustrative.

Below are two results in which Va was estimated for a marker, thus no permutation tests are printed out.

The Following Analysis has Marker 0 Removed
The following Analysis is Bootstrapped for Marker 25

The Following Analysis has Marker 6 Removed
The following Analysis is Bootstrapped for Marker 25

The Following Analysis has Marker 0 Removed
The following Analysis is Bootstrapped for Marker 0

1   0   3   1   0  0
3   0   9   2   0  0
2  1   6  1  1  0
4   0   6  2  0  0
5   0  7  1  0  0
2   0  10  1  1  0
3   1  13  2  2  0
6   1   7  2  0  0
6   1   9  2  0  0
6   2   9  5  0  0

The Following Analysis has Marker 6 Removed
The following Analysis is Bootstrapped for Marker 0

5   1   4  2  1  0
2  0  15  1  1  0
4   2  6  1  1  0
Above two sets of ten analyses, similar to those in the first section, were carried out on a permuted data set with the NUMBER of tests significant printed out. Although, it is clear that more replicates would be desirable, in no cases were the number of tests significant in the first column as large as those actually seen. In the scabrous paper this test, but with 1000 bootstrap replicates, is the best evidence that scabrous has variants affecting bristle number. Even after the effect of marker 8 was removed, the results still appear significant.

Below are two more analyses in which permutation tests were not carried out (Parameter sets G and H).

The following Analysis has Marker 0 Removed
The following Analysis is Bootstrapped for Marker 25

The following Analysis has Marker 6 Removed
The following Analysis is Bootstrapped for Marker 25

**scab VA.out**

The following Analysis has Marker 0 Removed
The following Analysis is Bootstrapped for Marker 0

The following Analysis has Marker 6 Removed
The following Analysis is Bootstrapped for Marker 0

The following Analysis has Marker 0 Removed
The following Analysis is Bootstrapped for Marker 25

0.423636  0.364976  0.0468036  0.0558704  0.10145  0.174633
0.0510782 0.0368086

The following Analysis has Marker 6 Removed
The following Analysis is Bootstrapped for Marker 25

0.365774  0.325512  0.05991  0.0722417 0.234509  0.119587
0.0664734 0.054411

The analyses for **scab VA.out**, which only prints out results if the value in the column labelled Rank of Marker to Estimate Va for is non-zero. Immediately above are the estimates of Va by sex and character as outlined in the comments of the main routine. One set is on the raw data and one set is the estimate with marker 8 removed. The first four estimates are the actual estimates and the second four the 'noise'. It follows that the values should be pasted into a spreadsheet to calculate the actual estimates. These are provided here as:

<table>
<thead>
<tr>
<th>Marker</th>
<th>R25, R0 removed</th>
<th>R25, R6 removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.322</td>
<td>0.190</td>
<td>-0.004</td>
</tr>
<tr>
<td>0.131</td>
<td>0.206</td>
<td>-0.007</td>
</tr>
</tbody>
</table>
It is apparent that the estimates of Va for abdominals are large and likely significant but those for sternopleurals are small and likely not significant. There is some suggestion that the Va estimates for males and females are different and it is possible that the removal of marker 8 (R6) had an effect on the estimate of Va.

The following analysis has Marker 0 Removed
The following analysis is bootstrapped for Marker 0

The following analysis has Marker 6 Removed
The following analysis is bootstrapped for Marker 0

The following analysis has Marker 0 Removed
The following analysis is bootstrapped for Marker 25

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.610216</td>
<td>0.701421</td>
<td>0.0558436</td>
<td>0.0621538</td>
<td>0.218574</td>
<td>0.371945</td>
<td></td>
</tr>
<tr>
<td>0.0362306</td>
<td>0.0676418</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.384931</td>
<td>0.182415</td>
<td>0.0891093</td>
<td>0.0452159</td>
<td>0.149763</td>
<td>0.0861319</td>
<td></td>
</tr>
<tr>
<td>0.0940951</td>
<td>0.0415188</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.225462</td>
<td>0.194129</td>
<td>0.0550038</td>
<td>0.0501219</td>
<td>0.173672</td>
<td>0.165015</td>
<td></td>
</tr>
<tr>
<td>0.0897383</td>
<td>0.0666292</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.329296</td>
<td>0.41153</td>
<td>0.228492</td>
<td>0.270713</td>
<td>0.0512556</td>
<td>0.109675</td>
<td></td>
</tr>
<tr>
<td>0.0989618</td>
<td>0.0801263</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.327708</td>
<td>0.197151</td>
<td>0.0528007</td>
<td>0.045318</td>
<td>0.060996</td>
<td>0.151152</td>
<td></td>
</tr>
<tr>
<td>0.0272687</td>
<td>0.0141332</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0421574</td>
<td>0.208625</td>
<td>0.0317842</td>
<td>0.0236943</td>
<td>0.254708</td>
<td>0.18844</td>
<td></td>
</tr>
<tr>
<td>0.0579899</td>
<td>0.0496062</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.311406</td>
<td>0.128301</td>
<td>0.0379204</td>
<td>0.0409203</td>
<td>0.100337</td>
<td>0.0848625</td>
<td></td>
</tr>
<tr>
<td>0.0957806</td>
<td>0.10849</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.26971</td>
<td>0.204394</td>
<td>0.0420643</td>
<td>0.0478577</td>
<td>0.0538098</td>
<td>0.0491147</td>
<td></td>
</tr>
<tr>
<td>0.0442916</td>
<td>0.0462005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.893932</td>
<td>1.00228</td>
<td>0.0376418</td>
<td>0.0122639</td>
<td>0.140695</td>
<td>0.152886</td>
<td></td>
</tr>
<tr>
<td>0.00576306</td>
<td>0.0211329</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.10426</td>
<td>0.75177</td>
<td>0.0690573</td>
<td>0.111074</td>
<td>0.337988</td>
<td>0.203421</td>
<td></td>
</tr>
<tr>
<td>0.0823901</td>
<td>0.0779042</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following analysis has Marker 6 Removed
The following analysis is bootstrapped for Marker 25

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.723026</td>
<td>0.49993</td>
<td>0.01994</td>
<td>0.0858522</td>
<td>0.170961</td>
<td>0.0741652</td>
<td></td>
</tr>
<tr>
<td>0.0466411</td>
<td>0.0310698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.325493</td>
<td>0.238214</td>
<td>0.114966</td>
<td>0.120384</td>
<td>0.193911</td>
<td>0.0951576</td>
<td></td>
</tr>
<tr>
<td>0.027818</td>
<td>0.0288007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.424886</td>
<td>0.609342</td>
<td>0.0826637</td>
<td>0.127642</td>
<td>0.0969876</td>
<td>0.118951</td>
<td></td>
</tr>
<tr>
<td>0.0139111</td>
<td>0.0359959</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.383976</td>
<td>0.367316</td>
<td>0.0506117</td>
<td>0.0963547</td>
<td>0.0724123</td>
<td>0.0532086</td>
<td></td>
</tr>
<tr>
<td>0.0351486</td>
<td>0.0434659</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.399646</td>
<td>0.485186</td>
<td>0.144831</td>
<td>0.132777</td>
<td>0.185364</td>
<td>0.219711</td>
<td></td>
</tr>
<tr>
<td>0.0639536</td>
<td>0.0829797</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.647288</td>
<td>0.581675</td>
<td>0.0244114</td>
<td>0.078728</td>
<td>0.050608</td>
<td>0.0210424</td>
<td></td>
</tr>
<tr>
<td>0.0318603</td>
<td>0.0769675</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.013118</td>
<td>0.459895</td>
<td>0.10232</td>
<td>0.0768486</td>
<td>0.138253</td>
<td>0.20472</td>
<td></td>
</tr>
<tr>
<td>0.0940853</td>
<td>0.0913328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.377721</td>
<td>0.280322</td>
<td>0.194304</td>
<td>0.214632</td>
<td>0.0826932</td>
<td>0.139498</td>
<td></td>
</tr>
<tr>
<td>0.313698</td>
<td>0.161272</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.718901</td>
<td>0.79206</td>
<td>0.189429</td>
<td>0.173518</td>
<td>0.0321898</td>
<td>0.0364812</td>
<td></td>
</tr>
<tr>
<td>0.0824091</td>
<td>0.0642703</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.638752</td>
<td>0.571678</td>
<td>0.103697</td>
<td>0.0802818</td>
<td>0.0362852</td>
<td>0.0130353</td>
<td></td>
</tr>
<tr>
<td>0.0499504</td>
<td>0.0596878</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In a manner similar to the actual estimates one can obtain a number of bootstrapped estimates of Va (Parameters sets G and H). These are shown above with actual estimates (after subtracting C5 form C1, etc) provided below. The first rows of each, in bold, are the actual estimates copied from above.

<table>
<thead>
<tr>
<th>Marker R25, R0 removed (bootstrap)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.322</td>
<td>0.190</td>
<td>-0.004</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>0.392</td>
<td>0.329</td>
<td>0.020</td>
<td>-0.005</td>
<td></td>
</tr>
<tr>
<td>0.235</td>
<td>0.096</td>
<td>-0.005</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>0.051</td>
<td>0.029</td>
<td>-0.035</td>
<td>-0.017</td>
<td></td>
</tr>
<tr>
<td>0.278</td>
<td>0.302</td>
<td>0.130</td>
<td>0.190</td>
<td></td>
</tr>
<tr>
<td>0.267</td>
<td>0.046</td>
<td>0.026</td>
<td>0.031</td>
<td></td>
</tr>
<tr>
<td>0.167</td>
<td>0.020</td>
<td>-0.026</td>
<td>-0.026</td>
<td></td>
</tr>
<tr>
<td>0.211</td>
<td>0.043</td>
<td>-0.058</td>
<td>-0.068</td>
<td></td>
</tr>
<tr>
<td>0.216</td>
<td>0.155</td>
<td>-0.002</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>0.753</td>
<td>0.849</td>
<td>0.032</td>
<td>-0.009</td>
<td></td>
</tr>
<tr>
<td>0.766</td>
<td>0.548</td>
<td>-0.013</td>
<td>0.033</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker R25, R6 removed (bootstrap)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.131</td>
<td>0.206</td>
<td>-0.007</td>
<td>0.018</td>
<td></td>
</tr>
<tr>
<td>0.552</td>
<td>0.426</td>
<td>-0.027</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>0.132</td>
<td>0.143</td>
<td>0.087</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>0.328</td>
<td>0.490</td>
<td>0.069</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>0.312</td>
<td>0.314</td>
<td>0.025</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>0.214</td>
<td>0.265</td>
<td>0.081</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>0.597</td>
<td>0.561</td>
<td>-0.007</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>0.675</td>
<td>0.255</td>
<td>0.007</td>
<td>-0.014</td>
<td></td>
</tr>
<tr>
<td>0.295</td>
<td>0.141</td>
<td>-0.119</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>0.687</td>
<td>0.756</td>
<td>0.107</td>
<td>0.109</td>
<td></td>
</tr>
<tr>
<td>0.602</td>
<td>0.559</td>
<td>0.054</td>
<td>0.021</td>
<td></td>
</tr>
</tbody>
</table>

It is apparent from the above, for the case with no marker removed, that Va attributable to marker 44 (R25) for abdominal bristles (first two columns) is likely significantly different from zero. Clearly more bootstrap replicates are needed. But, the earlier observation that there may be sex differences is likely not significant, because of the apparent overlap in the bootstrap estimates between the sexes. It is also apparent that the Va's estimated for sternopleurals are likely not significantly greater than zero, as many replicates are less than zero. Similarly, the second set of values, with M8 removed, although M8 originally appeared to result in different estimates of Va is not likely significantly so. This is because the bootstrap estimates appear to overlap a great deal. It is apparent that the variances on estimates of Va are generally very large as would be expected. The estimates are also very close to those presented in the *scabrous* paper (i.e., 0.299 in table 2 for SSCP 1447 ABs vs. (0.322 (male) + 0.190(female))/2 = 0.256 above), even though they were estimated by different methods.
Bibliography


