

# From complex traits to complex alleles

Several recent studies using natural populations of *Drosophila* show that one must be very careful when sorting among the large number of molecular polymorphisms found at most loci to identify the nucleotide changes responsible for phenotypic variation in complex traits. Indeed, several mutations within a single allele can interact to generate the overall observed effect. The results are instructive both for those interested in the genetics of evolutionary change and for those attempting to ferret out the genetic basis of complex human diseases.

As more and more research is conducted aimed at dissecting the genetic basis of complex traits, the hope is that we will eventually be able to get our hands on the genes underlying these traits and not have them run through our fingers as complexity mounts upon complexity. In particular, the genetic variants or alleles that ultimately lead to different trait or disease states need to be well defined for us to proceed towards a complete characterization of the molecular genetics involved. It is not difficult to slip comfortably into the mindset that each allele that we might find is the result of a single mutation. After all, this is the way that well-behaved alleles act in experiments involving induced mutations. Alleles in natural populations, however, are unlikely to have been so recently derived and are the products of an unknown evolutionary past. The results from several recent studies suggest that the mapping between mutation and allele might, indeed, be more complex than one would like, at least for natural populations of *Drosophila melanogaster*.

## A *Delta* in bristle number

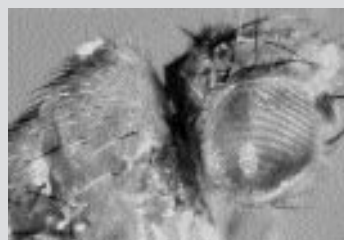
The last decade has seen a tremendous increase in studies investigating the genetic basis of quantitative traits displaying continuous variation. One emerging paradigm for addressing this problem is to use marker-based mapping to locate the chromosomal regions or quantitative trait loci (QTL) of interest<sup>1</sup>. A problem with this approach is that, although the mapping can be quite informative, the regions identified can be quite large, encompassing dozens or even hundreds of genes. Therefore, an alternative approach is to focus on likely candidate loci and then directly assess their impact on quantitative variation. In the latest of a series of papers on this topic<sup>2</sup> from the laboratories of Trudy Mackay and Charles Langley, Long *et al.*<sup>3,4</sup> studied the effect of the gene *Delta* on bristle number variation in 55 naturally derived chromosomes (Fig. 1). Two lines of evidence led these researchers to the *Delta* locus. First, the results of previous mapping experiments indicated a QTL in the chromosomal region that includes *Delta* (Ref. 4). Second, *Delta* is known to be involved in the development of the neurogenic system of which bristles are a part<sup>5,6</sup>. *Delta* encodes a ligand that interacts with the Notch receptor to generate cell-cell signaling during bristle formation<sup>7</sup> (Fig. 2). Complementation studies focusing on *Delta* verify that this locus can generate quantitative differences among lines<sup>8,9</sup>.

Knowing that a locus has an effect on a quantitative character is not enough; we want to know which specific changes within this fairly large gene are responsible for generating variation (Fig. 2). Understanding the functional nature of allelic differences will be very important in studies of disease, as it is difficult to know how to cure something if you do not know what has gone wrong. Long *et al.* used an association test<sup>10</sup> based on linkage-disequilibrium to map the correspondence between phenotype and molecular variation within *Delta*. Using RFLP markers that spanned the entire 57 kb region of the gene, they found that two markers were associated with a substantial fraction of the variation in sternopleural bristles in males and sternopleural and abdominal bristles in females. Interestingly, both of these markers were in introns. Although it is possible that these markers reflect linked effects of nearby exons, it is known that introns within *Delta* can contain transcriptional enhancers<sup>11</sup>. Tracking these marker associations directly to the actual nucleotides leading to the trait differences (quantitative trait nucleotides, QTN; Ref. 3) will depend ultimately on a sufficiently high density of markers, as well as a solid understanding of the functional basis of the molecular effects (Fig. 2).

## Intralocus interactions in ADH activity

This overall approach is supported very strongly by similar research on alcohol dehydrogenase (ADH) activity in *D. melanogaster*. In this case, Stam and Laurie<sup>12</sup> found even more striking intralocus interactions when looking at allelic variation within and around the *Adb* locus. Using engineered alleles and genetic transformation, they showed that three different regions within *Adb* can combine and interact to form 'superalleles' with large effects on transcription rates. Again, two of the three sites were in noncoding regions, including an intron and the 3' untranslated region, although the final site, the well-known fast/slow allelomorphs of *Adb*, was in a coding region. Other work has suggested that there could be long-range interactions between the 5' and 3' untranslated regions that might be involved in the regulation of gene expression<sup>13</sup>. While certainly consistent with our current understanding of gene regulation, the possibility that noncoding regions might play an important role in quantitative variation points out how complex these complex traits might actually be.

FIGURE 1. Bristles on a *Drosophila* head and thorax



The *Delta* gene has quantitative effects on bristle numbers in *Drosophila*.

Patrick C. Phillips  
pPhillips@uta.edu

Biology Department,  
Box 19498,  
University of Texas at  
Arlington, Arlington,  
TX 76019-0498, USA.

Even if we have the allele in hand, which of the many potential differences within that allele are responsible for shaping the phenotype? How do we identify the QTN within the QTL? The implication from this work is that the most important differences between alleles might not always be the most obvious. Careful experiments along the lines of the genetic transformations performed by Stam and Laurie might be necessary to find out what really makes an allele 'different' from the standpoint of the trait under study. Further, finding a single locus with a large effect on a particular trait does not preclude the possibility that the effect is generated by multiple mutational substitutions. An allele of large effect cannot be equated with a mutation of large effect<sup>3</sup>, as is often unconsciously done in discussions of how variation and adaptations arise<sup>14</sup>.

### How to build a sterility factor

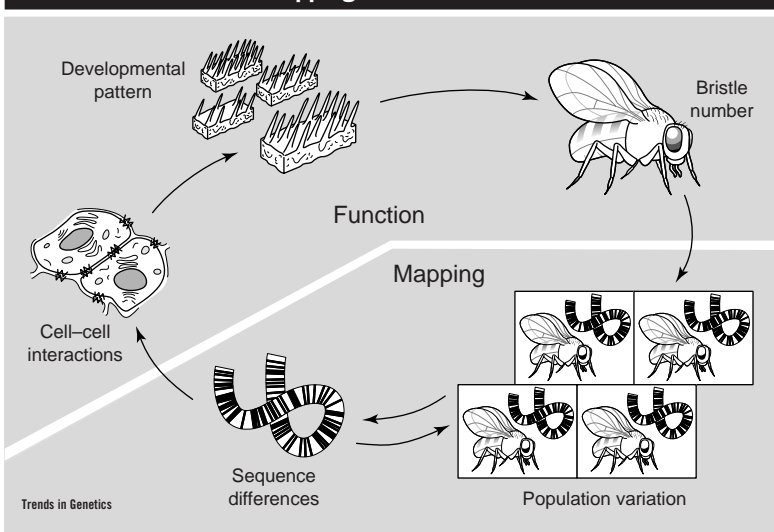
That potential complexity should not be taken lightly is well illustrated by the genetic dissection of sterility factors in *Drosophila*. Crosses between *D. mauritiana* and *D. simulans* result in hybrid male sterility. Analysis of the genetic basis of this sterility has revealed a complex set of interacting factors underlying the differences between these species<sup>15</sup>. In a study of what they initially thought was one of these factors on the X chromosome of *D. mauritiana*, Davis and Wu<sup>16</sup> used fine-scale mapping to show that this single region could, in fact, be broken into three separate factors. Turning to a 200 kb fragment containing one of these smaller factors, they found that further mapping turned this single factor into yet two more factors. The authors use the metaphor of the broom of the sorcerer's apprentice, which can forever be splintered to form more and more brooms, to describe this situation. Initial uniformity and simplicity dissolves under closer scrutiny and an attempt to identify the precise nature of the difference (it is almost as though there is an 'uncertainty principal' for the mapping of sterility factors). The lesson here is that mapping results alone, without a finer-scale analysis, might not reveal the multiple changes underlying what otherwise appears to be a single 'factor'.

### Making simple genetic effects complex

When the Long *et al.*<sup>3</sup> and Stam and Laurie<sup>12</sup> results are combined with the extremely complex pattern of interactions emerging from studies of hybrid sterility in *Drosophila* (Ref. 15), a striking pattern emerges. In the best-studied systems to date, multiple changes within a single locus or small chromosomal region are responsible for natural variation, and these changes are often cryptic with respect to what one might traditionally define as an allele. Therefore, what we view as an allele in the current context of a population might have been pieced together historically over time from multiple mutations. Studies seeking to understand the molecular basis of quantitative variation need to focus on the proper locus of variation, including mutational input. In these cases, this input might include complex interactions within a single 'gene'.

Even if much of the variation of a trait is caused by differences at a single locus, does this mean that its genetics are simple? Simplicity and complexity are relative terms that should be related to the number of genetic changes needed to reach a particular state, regardless of whether or not these changes happen to be concentrated within a particular coding region that we would normally call a gene.

**FIGURE 2. Association mapping within a candidate locus**



Localizing an allelic effect to a particular region or set of regions within a locus requires two fundamental elements. First, a suitable candidate locus must be found. In the case of Long *et al.*<sup>3</sup>, the *Delta* gene is known to influence sensory bristle development by serving as a ligand that mediates cell-cell interactions with the Notch receptor<sup>7</sup>. These interactions then affect bristle spacing during development<sup>6</sup>, which, in turn presumably leads to bristle number differences in adult flies. Second, mapping sequence differences among a large number of flies is accomplished by looking for an association between the phenotype of the fly and its genotype at a number of markers within the locus. Significant associations are generated by linkage disequilibrium between the marker and the true sequence difference that leads to the effect. For the *Delta* locus, Long *et al.*<sup>3</sup> found two regions with a strong association between marker and phenotype. However, many sequence differences might not lead to functional differences. Therefore, the next step, yet to be taken, is to identify how these intralocus effects translate into functional differences, allowing the circle of causation to be closed. Knowledge from many levels of analysis will need to be combined to fully understand the connection between sequence differences and population variation.

### Implications for human genetics

In many of the human genetic diseases studied thus far [e.g. cystic fibrosis<sup>17</sup> and TP53 (p53)-based carcinomas<sup>18</sup>], thousands of different disease alleles at each locus have been recognized. These studies have the advantage that the disease state is readily determined, whereas with complex traits it might be difficult to distinguish a QTN from a neutral difference between individuals. For example, in the gene that encodes lipoprotein lipase, a candidate susceptibility locus for cardiovascular disease, 88 variable sites were found in a sample of 71 individuals, yielding an average of one variable site every 500 bp in a 9.7 kb region of the gene<sup>19</sup>. The total level of variability in this gene will make assessment of the disease risk associated with any particular genetic change difficult, at best.

If multiple intralocus effects and interactions turn out to be the rule for QTL, then mapping based on association tests<sup>10</sup> will need to be at a fairly fine scale and sample sizes will need to be large enough to allow for intralocus recombination. Actually demonstrating the functional basis of an allelic effect will be quite difficult, as the precision possible in the *Drosophila* studies is precluded in investigations involving humans. Model systems will play an important role, as always.

### Conclusion

Understanding the molecular genetic basis underlying quantitative variation will continue to be a growing concern among geneticists. Even if we map to a single locus

and have the gene in our hands, understanding the true genetic basis of the variation might depend on how well our fingers grasp the potential complexities of intralocus effects. It is sobering that the best studied systems to date question our simple notion of allelic variation at a single locus.

**Acknowledgements**

I thank N. Johnson and two reviewers for providing very helpful comments, B. McAllister for supplying the fly, and C. Gates for creating the artwork for the figures. Supported by grants from the National Institutes of Health (GM54185) and the National Science Foundation (DBI-9722921).

**References**

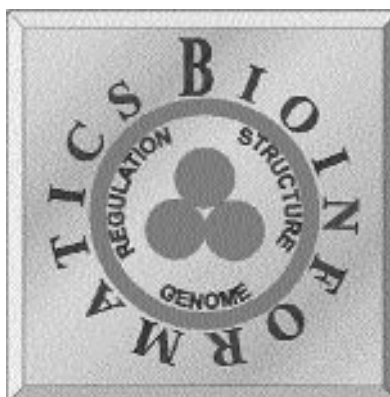
<p>1 Tanksley, S.D. (1993) Mapping polygenes. <i>Annu. Rev. Genet.</i> 27, 205–233</p> <p>2 Mackay, T.F.C. (1995) The genetic basis of quantitative variation: numbers of sensory bristles of <i>Drosophila melanogaster</i> as a model system. <i>Trends Genet.</i> 11, 464–470</p> <p>3 Long, A.D. <i>et al.</i> (1998) Two sites in the <i>Delta</i> gene region contribute to naturally occurring variation in bristle number in <i>Drosophila melanogaster</i>. <i>Genetics</i> 149, 999–1017</p> <p>4 Long, A.D. <i>et al.</i> (1995) High resolution mapping of genetic factors affecting abdominal bristle number in <i>Drosophila melanogaster</i>. <i>Genetics</i> 144, 1497–1510</p> <p>5 Jan, Y.N. and Jan, L.Y. (1993) The peripheral nervous system, in <i>The Development of Drosophila melanogaster</i>, Vol 2 (Bate, M, and Martinez Arias, A., eds), pp 1207–1244, Cold Spring Harbor Laboratory Press</p> <p>6 Parks, A.L. and Muskavitch, M.A. (1993) <i>Delta</i> function is required for bristle organ determination and morphogenesis in <i>Drosophila</i>. <i>Dev. Biol.</i> 157, 484–496</p>	<p>7 Artavanis-Tsakonas, S. <i>et al.</i> (1995) Notch signaling. <i>Science</i> 268, 225–232</p> <p>8 Long, A.D. <i>et al.</i> (1996) Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in <i>Drosophila melanogaster</i>. <i>Genetics</i> 144, 1497–1510</p> <p>9 Lyman, R.F. and Mackay, T.F.C. (1998) Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in <i>Drosophila melanogaster</i>: the <i>Delta-Hairless</i> gene region. <i>Genetics</i> 149, 983–998</p> <p>10 Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. <i>Science</i> 273, 1516–1517</p> <p>11 Haenlin, M. <i>et al.</i> (1994) Genomic regions regulating early embryonic expression of the <i>Drosophila</i> neurogenic gene <i>Delta</i>. <i>Mech. Dev.</i> 47, 99–110</p> <p>12 Stam, L.F. and Laurie, C.C. (1996) Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in <i>Drosophila melanogaster</i>. <i>Genetics</i> 144, 1559–1564</p>	<p>13 Parsch, J. <i>et al.</i> (1997) Site-directed mutations reveal long-range compensatory interactions in the <i>Adh</i> gene of <i>Drosophila melanogaster</i>. <i>Proc. Natl. Acad. Sci. U. S. A.</i> 94, 928–933</p> <p>14 Orr, H.A. and Coyne, J.A. (1992) The genetics of adaptation: a reassessment. <i>Am. Nat.</i> 140, 725–742</p> <p>15 Wu, C.I. and Palopoli, M.F. (1994) Genetics of postmating reproductive isolation in animals. <i>Annu. Rev. Genet.</i> 27, 283–308</p> <p>16 Davis, A.W. and Wu, C.I. (1996) The broom of the sorcerer's apprentice: the fine structure of a chromosomal region causing reproductive isolation between two sibling species of <i>Drosophila</i>. <i>Genetics</i> 143, 1287–1298</p> <p>17 Cystic Fibrosis Genetic Analysis Consortium (1998) Cystic fibrosis genetic database. <a href="http://www.genet.sickkids.on.ca/cftr/">www.genet.sickkids.on.ca/cftr/</a></p> <p>18 Beroud, C. and Soussi, T. (1998) <i>p53</i> gene mutation: software and database. <i>Nucleic Acids Res.</i> 26, 200–204</p> <p>19 Nickerson, D.A. <i>et al.</i> (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. <i>Nat. Genet.</i> 19, 233–240</p>
--	--	---



# Bioinformatics in Siberia

First International Conference on Bioinformatics of Genome Regulation and Structure, Novosibirsk, Siberia, Russia, 24–27 August 1998.

As genome sequencing progresses with an ever increasing pace, the need to perform genomic data storage, retrieval, and analysis is getting more and more pressing. As a result, a new interdisciplinary field has emerged: bioinformatics, involving people with various backgrounds, including biology, computer science, mathematics and physics. As the first achievements of the community of bioinformaticians as a whole, one can invoke the development of a wealth of general and specialized databases, as well as various software packages for sequence assembly, gene finding, homology search, etc.<sup>1</sup>



Up to now, most of the work of bioinformaticians has focused on structural properties of individual genes and their products. However, among the most challenging tasks that bioinformaticians are facing are those of inte-

grating the various types of information on gene regulation, and developing dynamic representations of the genome and its expression. These challenges were the focus of the conference in Siberia convened by Nicolay Kolchanov (Inst. of Cytology and Genetics, Novosibirsk, Russia) and Victor Solovyev (The Sanger Centre, Cambridge, UK). The conference included more than 50 oral presentations, that covered topics including: databases on regulatory genomic sequences, regulatory protein and gene networks; gene finding; and evolutionary genomics.

Although the conference encompassed most of the themes usually included in the emerging field of Bioinformatics, a more novel aspect of the conference was the recurrent references made to the global regulative and

Denis Thieffry  
denis@mpiwg-berlin.mpg.de

\*Gerald Z. Hertz  
hertz@colorado.edu

Planck Institute for the History of Science; Wilhelmstrasse 44, D-10117 Berlin, Germany.

\*Department of Molecular, Cellular and Developmental Biology; University of Colorado, Campus Box 347, Boulder, CO 80309, USA.