QUANTITATIVE GENETIC ANALYSES OF COMPLEX BEHAVIOURS IN *DROSOPHILA*

Robert R. H. Anholt‡ and Trudy F. C. Mackay‡*

Abstract | Behaviours are exceptionally complex quantitative traits. Sensitivity to environmental variation and genetic background, the presence of sexual dimorphism, and the widespread functional pleiotropy that is inherent in behavioural phenotypes pose daunting challenges for unravelling their underlying genetics. *Drosophila melanogaster* provides an attractive system for elucidating the unifying principles of the genetic architectures that drive behaviours, as genetically identical individuals can be reared rapidly in controlled environments and extensive publicly accessible genetic resources are available. Recent advances in quantitative genetic and functional genomic approaches now enable the extensive characterization of complex genetic networks that mediate behaviours in this important model organism.

PLEIOTROPY
The phenomenon in which a single gene is responsible for several different phenotypic effects.

**Department of Zoology, and ‡Department of Genetics and W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, North Carolina 27695-7617, USA.*
Correspondence to R.R.H.A.
e-mail: anholt@ncsu.edu
doi:10.1038/nrg1472

Behaviours are typical quantitative traits. They are driven by epistatic networks of multiple segregating genes with PLEIOTROPIC effects, which necessitates the use of statistical analyses in describing behavioural phenotypes[1,2]. Behaviours are also exquisitely sensitive to environmental variation, enabling a wide range of behavioural phenotypes to be produced by genetically identical individuals. Understanding the genetic architecture of a given behaviour requires the identification of all contributing genes and knowledge of how these genes interact as functional ensembles. In addition to identifying the genes that are essential for manifestation of a behaviour, we must also determine which of these genes contribute to naturally occurring phenotypic variation in behaviour, as polymorphisms that account for this variation form the substrate for evolutionary change.

Until recently, hopes of accomplishing these daunting tasks might have seemed unrealistic. However, recent advances in quantitative genetic and functional genomic technologies are now rapidly bringing these goals within reach. Special attention must be paid to controlling the genetic background and accounting for the effects of environmental variation and sexual dimorphism in behavioural phenotypes.

Drosophila melanogaster presents an excellent model for studying the genetic architectures of complex traits, including behaviours. The *D. melanogaster* genome has been sequenced[3], the organism is extremely amenable to genetic manipulation, and extensive public databases and genetic resources are available (for example, see the FlyBase website). Of great importance for the study of behaviour, *D. melanogaster* also provides the ability to rapidly generate large numbers of genetically identical individuals and rear them under controlled environmental conditions.

In the last few decades, mutagenesis approaches have identified several key genes that control behaviours in *D. melanogaster*, including the pioneering discoveries of *dunce*[4] and *rutabaga*[5], which are implicated in memory and learning, and the *period* gene, which regulates circadian locomotor activity[6]. Behaviours, however, are orchestrated through ensembles of many interacting genes, which contribute to the manifestations of behaviours to different extents. Early studies in which individual genes with large effects on specific behaviours were identified have provided isolated pieces of a puzzle, the completion of which requires the synthesis of traditional molecular genetic and quantitative genetic analyses. The advent of the genomic era has enabled this

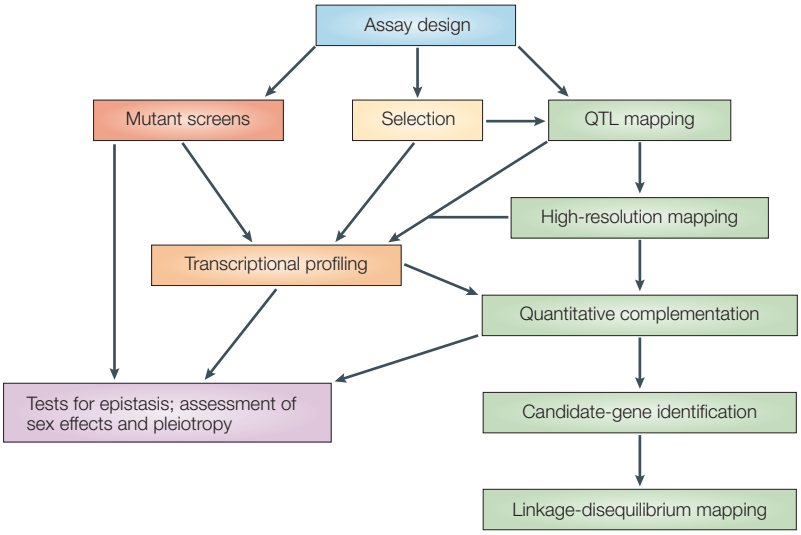


Figure 1 | **Experimental strategies used in quantitative genetic analyses of behavioural traits in *Drosophila melanogaster*.** The first step is the design of a simple and reproducible behavioural assay. One strategy is then to carry out large-scale mutant screens for genes that contribute to the behaviour. Epistatic interactions, effects attributable to sexual dimorphism and pleiotropy can subsequently be evaluated, and effects of mutations in these genes on the transcriptome can be assessed. As an alternative to screening, mapping studies that capitalize on phenotypic variation can be used to identify regions containing quantitative trait loci (QTLs) that contribute to variation in behaviour. QTL intervals can be refined by high-resolution mapping, and quantitative complementation testing can then be used to identify candidate genes. Linkage disequilibrium analysis can then identify polymorphisms within the candidate gene that contribute to the observed phenotypic variation. In a third strategy, artificial selection can be used to generate populations that differ widely for the behavioural trait. The genes that give rise to these differences in the selected populations can be identified either by QTL mapping or by transcriptional profiling.

integration by facilitating the large-scale discovery of novel genes that contribute to behaviours, and has made it possible to analyse how these genes are organized in genetic networks.

Here, we review methods for the genetic analysis of complex behaviours in *D. melanogaster*, and provide guidelines for the design of mutant screens, the mapping of QUANTITATIVE TRAIT LOCI (QTLs), and other genetic analyses of complex behaviours. We first discuss the design of behavioural assays, and then examine how genetic background, environmental factors and sexual dimorphism should be taken into account. We also outline strategies for identifying genes that contribute to behaviours, using the complementary approaches of mutagenesis and mapping of genes that contribute to naturally occurring variation. Finally, we describe genetic and genomic approaches to organize behavioural genes into functional networks, and discuss the prominence of epistasis and pleiotropy.

Assay design

FIGURE 1 shows an overview of experimental strategies used for studying the genetics of complex behaviours in *D. melanogaster*. The starting point of any genetic analysis of behaviour is to develop a quantitative assay. Whereas such assays might sometimes, out of necessity, be complex, ideally they should be simple enough to enable large and rapid screens. Simple assays have been developed to quantify many complex behaviours in *D. melanogaster* (FIG. 2). For example, courtship behaviour can be quantified by measuring courtship latency (the time to initiate courtship) and copulation latency (the time to copulation)[7,8]; larval foraging behaviour can be quantified by measuring foraging track lengths[9] (FIG. 2a); olfactory avoidance behaviour can be measured by the number of animals that migrate away from a repellent odourant[10] (FIG. 2b); and aggression can be measured by quantifying the number of aggressive encounters during a defined time period[11]. Classic experiments in behavioural genetics have involved clever apparati designed to quantify PHOTOTACTIC behaviour[12], GEOTACTIC behaviour[13], CIRCADIAN LOCOMOTOR BEHAVIOUR[6], learning and memory[4,5,14], and ethanol sensitivity[15–17] (FIG. 2c).

Complex behaviours are modular; for example, courtship and mating behaviour require recognition, orientation, locomotion, wing vibration and, ultimately, copulation[9,18]. Phenotypic variation in any one of these components will affect courtship and mating behaviour. Many assays measure only one aspect of a behaviour, which might not always be sufficient for correct evaluation of the phenotype. For example, flies that show aberrant avoidance responses to repellent odourants[10] could be impaired not only in olfaction but also in locomotion, and this should be evaluated using a different assay. Similarly, 'learning and memory mutants' identified using OLFACTORY CONDITIONING PARADIGMS[4,5,14,19,20] might be impaired for olfaction, rather than being specifically impaired for learning and memory. Investigators who carry out these studies are well aware of this. So, experiments that measure a behavioural trait within the context of a particular assay are always designed with this caveat in mind; behavioural components that are not directly measured might nonetheless affect the animal's performance.

Controlling genetic background

Differences in genetic background can profoundly affect behavioural phenotypes. For example, CANTON-S FLIES will consistently jump in response to a small amount of odourant[21], whereas jump responses are seldom observed with Samarkand flies[10]. Furthermore, introduction of a single *P*-ELEMENT in a defined genetic background causes widespread transcriptional alterations[22], indicating that single polymorphisms might profoundly affect transcriptional profiles, and emphasizing the importance of controlling genetic background when analysing genes and genetic networks that mediate behaviours. So, the most powerful experimental designs for genetic analyses of behaviours in *D. melanogaster* involve the construction of genetically identical lines: for example, CO-ISOGENIC single *P*-element insertion lines for mutational analyses, and RECOMBINANT INBRED LINES (RILs) for QTL mapping. For mutations resulting from the insertion of a *P*-element, PHENOTYPIC REVERSION and TRANSGENIC RESCUE experiments to confirm that the *P*-element causes the mutant phenotype by disrupting the candidate gene must be conducted in the same co-isogenic background.

QUANTITATIVE TRAIT LOCI (QTLs). One of many genetic loci affecting variation in a complex trait (for example, behaviours, some aspects of morphology, resistance to environmental stress). QTLs have individually small effects, are sensitive to environmental variation, and are initially identified by statistical association of trait phenotypes with polymorphic molecular markers.

PHOTOTAXIS
Movement towards a light source.

GEOTAXIS
Movement upwards or downwards, which requires the perception of and response to gravity.

CIRCADIAN LOCOMOTOR BEHAVIOUR
Variation in endogenous locomotor activity that depends on time of day.

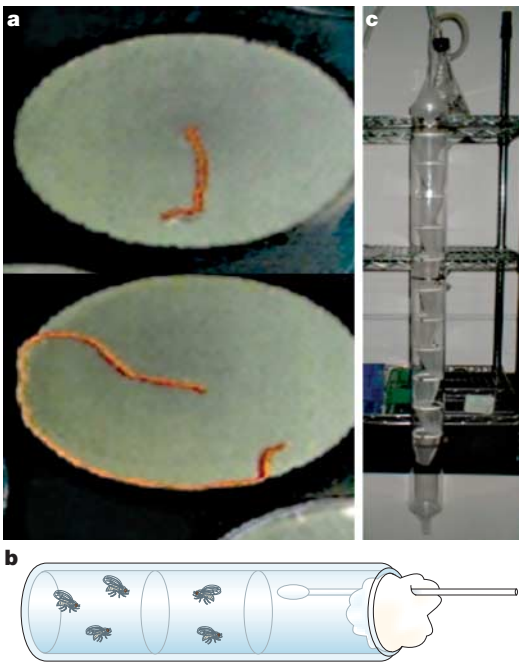


Figure 2 | **Simple behavioural assays that are amenable to high-throughput screening. a** | Naturally occurring polymorphisms in larval foraging behaviour can be readily quantified by measuring track lengths of feeding larvae. 'Rovers' (lower image) will migrate long distances on their food source, whereas 'sitters' (upper image) feed in a restricted area. These two distinct behaviours are due to polymorphisms in the *foraging* gene, which encodes a cyclic GMP-dependent protein kinase[9,84] (image courtesy of M. B. Sokolowski, University of Toronto, Canada). **b** | Assay for olfactory avoidance behaviour. Single-sex groups of five flies are placed in a vial with marked compartments. A repellent odourant is introduced on a cotton wool swab, the vial is placed horizontally, and after a 15 second period the number of flies in the compartment farthest from the odour source is counted at 5-second intervals. Ten counts are averaged to obtain an avoidance score. This 1-min assay can then be repeated several times to obtain a reliable quantitative statistical measurement of the behaviour for a given genotype[10]. **c** | Alcohol sensitivity can be quantified by measuring knock-down time in an 'inebriometer'. This is a 122cm-long vertical glass column, which contains a series of slanted mesh partitions to which flies can attach. Flies are introduced in the top of the column and exposed to ethanol vapours. As they lose postural control, they fall through the column. The elution time can be used as a quantitative measure of alcohol intoxication[15].

OLFACTORY CONDITIONING PARADIGM
An experimental procedure in which subjects learn to avoid a particular odour through the pairing of exposure to that odour with an aversive stimulus, such as an electric shock.

CANTON-S FLIES
A standard wild-type *D. melanogaster* strain, genetically different from Samarkand (another standard wild-type *D. melanogaster* strain)

P-ELEMENTS
A family of transposable elements that are widely used as the basis of tools for mutating and manipulating the *D. melanogaster* genome.

CO-ISOGENIC LINES
Homozygous lines that differ only by the presence of a single mutation.

RECOMBINANT INBRED LINES
A population of fully homozygous lines derived by 20 or more generations of full-sibling mating from the F_2 derived from a cross between two different inbred lines. Each line comprises ~50% of each parental strain in different combinations.

PHENOTYPIC REVERSION
Demonstration of restoration of the wild-type phenotype by construction of an allele in which a *P*-element associated with a mutation has been precisely excised.

TRANSGENIC RESCUE
Restoration of the wild-type phenotype in a mutant background by a construct containing the wild-type copy of the gene that has been introduced into the genome by germ-line transformation.

Transgenesis in *D. melanogaster* often involves the use of a binary expression system, such as the Gal4–UAS (upstream activating sequence) system. This system involves transactivation by the yeast Gal4 transcription factor of a transgene that has been cloned behind the *GAL4* promoter UAS[23]. Flies expressing the transgene are derived by crossing a line expressing Gal4 in a specific pattern (the Gal4 driver line) to a second line that carries the UAS-transgene construct. Hybrid stocks containing both the transgene and the *GAL4* driver are often derived from parental stocks with different genetic backgrounds. The effects of this can be controlled by comparing the phenotypes of the Gal4–UAS offspring with the phenotypes of F_1 individuals derived from the *GAL4* parental line crossed to the original transposon-free host strain in which the UAS-transgene construct was introduced (for example, REF. 25).

Sample sizes

Behaviours are highly sensitive to environmental variation — even genetically identical individuals reared under controlled environmental conditions will exhibit variation in their behaviours (see below). This necessitates measuring many individuals from each line to give a statistically accurate estimate of the mean genotypic effect.

How large is the required sample size? The answer must take into account the phenotypic variation — which is composed of genetic and environmental variation (or, in the case of isogenic lines, only environmental variation)[24] — and the probabilities of detecting spurious associations (false positives) or discarding significant associations (false negatives). Let us assume that we want to detect a quantitative difference in behaviour between a line that is homozygous for a mutation and another line that is homozygous wild type[24]. The relationship between the number of individuals that must be tested and the size of the difference in behaviour to be detected, for standard false-positive and false-negative significance thresholds, is illustrated in FIG. 3. Note that as the magnitude of the effect falls, the number of individuals needed to detect it rises exponentially. Most segregating QTLs have effects that require samples of 500–2,000 individuals in order for them to be detected.

Confounding environmental variation

As noted above, studies of *D. melanogaster* behavioural genetics typically involve screening hundreds to thousands of individuals. The design of such labour-intensive experiments requires careful consideration of the effect of common environmental variation on the expression of behaviours. The quantitative-genetic concept of common environment refers to any environmental condition that causes phenotypes of individuals reared within it to be more similar to each other than to individuals reared in different common environments. *D. melanogaster* reared in the same vial experience a common spatial environment, whereas those measured at the same time experience a common temporal environment. Different observers or handlers represent another source of common environmental variation.

As an example, consider a case in which variation in behaviour among 500 *D. melanogaster* lines is to be analysed, measuring 20 individuals per line for a total of 10,000 measurements. It might seem sufficient to simply determine how many flies could be accurately measured in a typical working day, and score all 20 flies from '*x*' lines per day until all were measured. However, daily fluctuations in temperature, humidity and barometric pressure contribute to variation in behaviour that would be erroneously ascribed to genetic differences between lines. Furthermore, most behaviours require locomotion, and

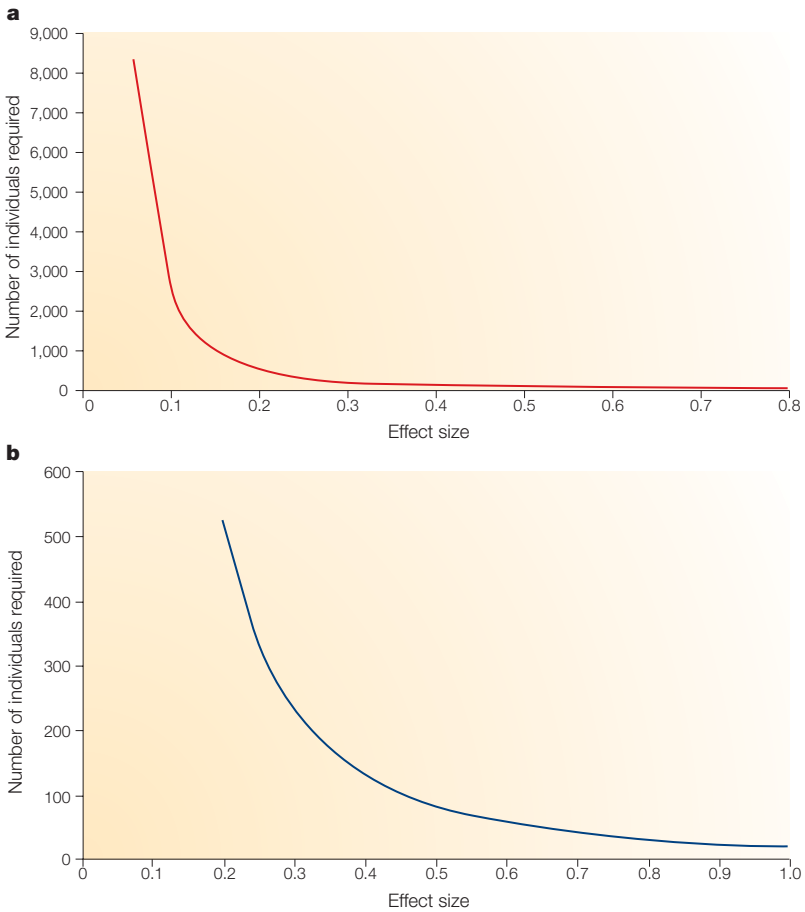


Figure 3 | **Sample sizes required in genetic analyses of complex behaviours.** The minimum numbers of individuals (*n*) that are required to detect a quantitative difference in a behavioural trait between two genotypes are shown, as calculated using standard statistical theory[82]. The false-positive and false-negative error rates used in the calculations were 0.05 and 0.1, respectively. Effect sizes are given in within-genotype standard deviation units (σ). Values of *n* are shown for effect sizes in the ranges 0.05σ–0.8σ (**a**) and 0.2σ–1.0σ (**b**). Consider mating behaviour as an example. If the within-genotype variance of copulation latency is 16 min^2, one would need a sample size of 84 individuals per genotype to detect a 2-min difference (0.5σ) in copulation latency between them. However, if the within-genotype variance in copulation is 100 min^2, a total of 525 individuals per genotype would be required to detect a 2-min difference (0.2σ) in copulation latency between them.

locomotor activity varies throughout the day, with peaks early in the morning and in the late afternoon[6,26–29]. In addition, two independent studies found circadian patterns in the expression of more than a hundred genes[30,31], showing that networks of genes under circadian control overlap with those that subserve many biological processes, including behaviours. This emphasizes the importance of controlling for variation in circadian rhythms in the quantitative analysis of behaviour, which can be achieved by restricting behavioural measurements to one of the peak periods of activity.

Common spatial and temporal environmental variation can be accounted for by strict randomization across environments, or by direct estimation. The former strategy requires that each of the 20 individuals tested per line in the example above comes from a different replicate vial (the spatial environment), and that only one individual per line is measured on any one day, randomizing lines across days. Alternatively, common environmental variation can be explicitly accounted for by taking replicate measurements of flies of the same genotype emerging from different vials, and at different times. For example, one could measure *x* lines per week (a block), with 4 individuals per replicate vial, per line, for each of 5 days. BOX 1 describes how the statistical technique of analysis of variance (ANOVA) can be used to analyze the results from such experimental designs and to determine how much of the observed variation is due to common environmental effects. Another common design is to include multiple measurements for a standard control genotype for each block, and to express scores for the genotype as deviations from the mean obtained for the control[32].

Genotype by environment interactions

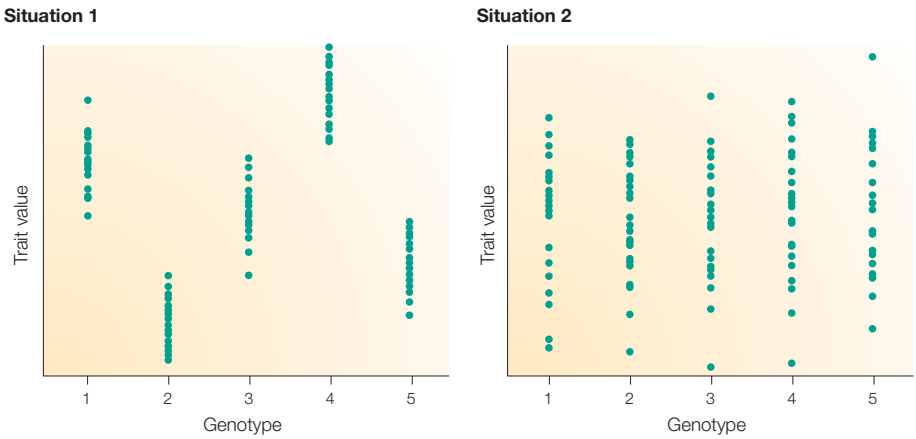
The variation in mean performance across a range of environments (for example, different temperatures) in a behavioural assay of individuals of a single genotype is called the environmental sensitivity of the genotype[24] (FIG. 4). Different genotypes often vary in their environmental sensitivities, a phenomenon called genotype by environment interaction[24] (FIG. 4), and this can lead to difficulties in replicating the results of behavioural assays. A classic demonstration of this comes from an experiment in which eight different inbred strains of mice were tested in six behavioural assays at three different locations[33]. Despite extensive efforts to use exactly the same strains and apparati, there were striking differences in performance of the different strains in the different test environments[33]. It should be noted that many robust behaviours are reproducible in different laboratories. In addition, when different results are obtained in different laboratory environments these discrepancies do not necessarily cast doubt on their validity, but rather establish genotype by environment interaction as a sensitive property of the trait under study.

Sexual dimorphism in behavioural traits

A significant source of phenotypic variation in behaviour is the effect of sex. Male and female flies have generally not been separated in behavioural assays, or in the subsequent analyses of their results, except when sex-specific behaviours, such as courtship and mating behaviour, have been analysed. Differences between males and females (sexual dimorphism) in mean phenotypic values of quantitative traits, including behavioural traits, are common[10,34–37], both at the level of mutational[10] variation and naturally segregating variation[34]. Furthermore, differences at the level of trait phenotype are re-capitulated at the level of transcription, and 50% or more of the *D. melanogaster* transcriptome exhibits sex-biased expression[38–40]. Whereas in whole flies such observations might be attributed to differential expression in the gonads, significant sexual dimorphism is also evident when RNA is extracted from *D. melanogaster* heads[22,41].

Variation in sexual dimorphism is equivalent to genotype-by-sex interaction, and can occur if a mutation or QTL affects only one sex (sex-specific effects),

Box 1 | **Analysis of variance**



General principles

Analysis of variance (ANOVA) is a key statistical tool in quantitative genetics. In essence, it is an expansion of the standard Student's *t*-test, which assesses the statistical significance of differences between the mean values of two data sets relative to the variance within each. ANOVA allows multiple comparisons and divides the observed variance among its contributing sources; that is, it quantifies to what extent variation attributable to each factor — for example genotype, age, sex, experimental treatment — contributes to the overall variation around the mean.

The principle of ANOVA can be illustrated with reference to the figure above, which depicts two different scenarios of variation in a behavioural trait within and between five distinct genotypes. The question of interest is whether there is a significant difference between the five genotypes. The ANOVA model used to determine this is $Y = \square + G + Er$, where Y is the observed value for each individual, $\square$ is the overall mean value, G is the effect of genotype, and Er is the variance within each genotype. The effect of genotype will be significant if the variance between the mean value of the behaviour between genotypes is greater than the variance in the behaviour within genotypes. In other words, G will be significant if the behaviour of individuals of a given genotype is more similar to each other than to individuals of a different genotype. In the example shown in the figure, the effect of genotype would be significant in situation 1, but not in situation 2.

Examples of ANOVA models used in quantitative behavioural genetics

Example 1. An ANOVA model designed to take account of common spatial and temporal environmental variation when evaluating behavioural differences for a single sex between multiple genotypes is $Y = \square + B + G(B) + R(G) + Er$. The other terms of the model partition the environmental variance between blocks of time (B), different genotypes within blocks (G) and replicate vials within genotypes (R). Here, Er represents the average environmental variance within replicate vials. This model evaluates the effects of the different genotypes as deviations from the block means, taking account of the common temporal environmental variation, and averaged over the mean of all replicate vials, taking account of the common spatial environmental variation.

Example 2. The magnitude of a genotype by environment interaction effect (FIG. 4) can be estimated in an experimental design in which the behaviour of each of several genotypes is assessed in several defined environments (for example, temperature, different locations). The simplest model for a single sex is: $Y = \square + G + E + G{\times}E + Er$. This model evaluates whether there are differences in behaviour between genotypes (G), averaged over all environments, and whether there are differences in behaviour between environments (E), averaged over all genotypes. The genotype by environment interaction ($G{\times}E$) is the extent to which the performance of a particular genotype in a given environment deviates from the performance predicted by the average performance of that genotype across all environments, and the average effect of that environment across all genotypes. Er is the average variance within genotypes and environments. Sex can also be considered a specific environment in which the genome operates. The sex environment and genetic interactions with sex (S) can be evaluated in the analogous model: $Y = \square + S + G + G{\times}S + Er$.

Example 3. ANOVA models can be constructed for any number of effects. For example, whole-genome expression arrays can be assessed for males and females (S), between different genotypes (G) and in different environmental treatments or tissues (T). The ANOVA model for such an experiment, applied to each gene on the array, would partition variation among the main effects and their interactions as follows: $Y = \square + S + G + T + S{\times}G + S{\times}T + G{\times}T + S{\times}G{\times}T + Er$. In this way, we can evaluate not only the extent to which transcript abundance varies between sexes (averaged over genotypes and treatments), genotypes (averaged over sexes and treatments) and treatments (averaged over sexes and genotypes), but also the extent to which there is genotype by sex and genotype by environment interaction for transcript abundance.

For a comprehensive discussion of ANOVA models, see REF. 82.

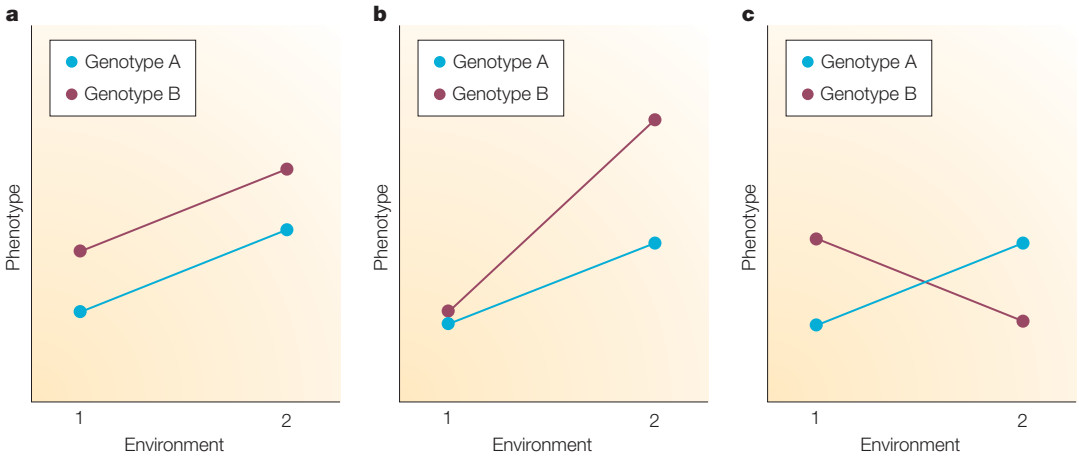


Figure 4 | **Genotype by environment interaction.** If different genotypes vary in the magnitude or direction of their environmental sensitivities, this is considered genotype by environment interaction. **a** | Phenotypic values for genotypes A and B in two different environments. The phenotypic value for B is greater than for A, but the relative differences in the two environments are the same, giving rise to parallel reaction norms (the lines connecting the phenotypic values in each environment) that indicate an absence of genotype by environment interaction. **b** | Both genotypes perform equally well in environment 1, but the phenotypic value of genotype B is substantially greater than that of genotype A in environment 2, indicating genotype by environment interaction. **c** | This shows a more extreme scenario in which the reaction norms cross, as genotype A performs better than genotype B in one environment, but in the other environment genotype B outperforms genotype A. The presence of genotype–environment interaction in panels **b** and **c** would be detected by a significant genotype by environment interaction term in a factorial ANOVA of the data (BOX 1). Note that the same analysis pertains to sexual dimorphism, as each sex can be considered a different environment.

affects both sexes but to different degrees (sex-biased effects), or affects both sexes but in opposite directions (sex-antagonistic effects)[42]. Sex-antagonistic effects for behaviours that are components of fitness are of particular interest, as opposite effects in males and females lead to the maintenance of genetic variation at such loci in natural populations[43]. Clearly, an appreciation of the genetic architecture of behaviours requires that the sexes be measured separately, where appropriate (FIG. 4).

Mutant screens

Mutant screens for behavioural genes are far more laborious than those for mutations with large qualitative effects, such as homozygous-lethal or sterile mutations. Dissection of genetic networks that drive behaviours requires the analysis of hypomorphic mutations that often elicit subtle phenotypic effects, which are sensitive to environmental variation[1,2,44]. This necessitates the measurement of behavioural phenotypes for multiple individuals carrying the same mutation, which first requires establishing stocks for each homozygous viable mutation.

Until recently, most publicly available mutant lines have come from various genetic backgrounds and often harbour several additional mutations in addition to the mutation that is under investigation. As the segregating variation within these strains is often of the same order of magnitude as the mutational effect under analysis, this precludes or complicates screens for behavioural effects. At the very least, it imposes a limit to the magnitude of the mutational effect that can be reliably detected. However, highly sensitive screens for mutations with subtle, quantitative effects on behaviour can be done if the mutations are introduced in co-isogenic backgrounds. This approach is highly effective, both for the identification of first mutations in predicted genes that affect behaviour, and for uncovering novel pleiotropic effects of known genes on behavioural phenotypes[10,45,46]. The large Exelixis collection of co-isogenic *P* and *piggyBac* insertions will be an invaluable resource to the fly community in this regard[47].

Sample sizes of 50 or fewer individuals in primary screens of thousands of lines might not be large enough to detect effects of less than 0.7 standard deviation units (FIG. 3, inset). It is advisable to perform preliminary screens to determine which mutant lines exceed prescribed confidence intervals from the overall mean[32]; these lines can then be re-tested with larger sample sizes to remove false positives.

Mapping genes affecting natural variation

QTL mapping presents a complementary approach to mutant screens that exploits naturally occurring allelic variation to identify genes that affect behaviour. These studies also give insights into the evolutionary forces responsible for maintaining natural genetic variation for behavioural traits. Regions defined by an initial genome scan can be further refined by high-resolution recombination mapping or deficiency COMPLEMENTATION TESTS. After narrowing the QTL intervals, candidate genes can be identified and validated through quantitative complementation tests. Finally, LINKAGE DISEQUILIBRIUM (LD) mapping can identify the molecular polymorphisms that define QTL alleles (FIG. 1). These mapping strategies are described in more detail in this section.

COMPLEMENTATION TESTS
In classical genetics, two mutations with the same phenotype are said to complement if their F_1 hybrid is wild type, and fail to complement if the F_1 hybrid exhibits the mutant phenotype. Failure to complement can arise if the mutations are alleles of the same locus, or are alleles of different loci that interact epistatically in the same genetic pathway.

LINKAGE DISEQUILIBRIUM
Non-random association of gene frequencies at two or more polymorphic loci in a population; that is, alleles of two different genes are not present together in gametes in the frequencies predicted by the product of their frequencies.

LINKAGE MAPPING
Markers that are physically close to a locus of interest segregate 'tightly' with the locus and will statistically be more closely associated with the observed variance of a trait. This property can be used to detect association in a population between a genetic marker and a locus that contributes to a particular phenotype.

F_2
A segregating generation of an intercross between F_1 individuals derived from two parental lines.

PERMUTATION TESTING
A method for obtaining appropriate significance thresholds for data sets in which multiple statistical tests are performed. The original analysis is repeated many times on data sets generated by appropriate random scrambling of the original data, generating an empirical distribution of the test under the null hypothesis.

INBREEDING DEPRESSION
The reduction in viability and fertility of inbred offspring compared with outbred offspring.

GENETIC DRIFT
Also known as random drift. A phenomenon whereby the frequency of a gene in a population changes over time owing to random sampling in finite populations.

RECOMBINATION MAPPING
Determining the order and location of genes on a chromosome in terms of the rate of recombination between them. When applied to QTL mapping, the position and effect of the QTL is inferred by linkage disequilibrium of the trait phenotype with the genotype of flanking molecular markers.

BACKCROSS
A segregating generation in which F_1 hybrids derived from two inbred parents are crossed to one of those parents.

ADVANCED INTERCROSS POPULATION
A population derived from several generations of crosses among F_2 individuals to maximize recombination for high resolution QTL mapping.

DEFICIENCY CHROMOSOME
A chromosome in which a defined region has been deleted.

QTL mapping. The first step is to map genetic regions containing one or more QTLs that affect variation in behaviour. This can be carried out by LINKAGE MAPPING using polymorphic molecular markers in mapping populations derived from two genetically divergent inbred strains[1,2]. Given the low heritability of most behaviours, the most efficient experimental design is to allow recombination to occur between two parental strains by generating an F_2 and then creating RILs by inbreeding the F_2 for 20 generations or more. As for mutagenesis screens, the ability to obtain replicate measurements from multiple genetically identical individuals from each RIL greatly increases statistical power for mapping QTLs. In the past, we and others have used this method to map QTLs affecting variation in olfactory behaviour[48], courtship song[49] and male mating behaviour[8]. Statistical methods that account for multiple QTLs[50] and that use PERMUTATION to determine the appropriate experiment-wise false-positive error rates[51] are now regarded as standard.

There are two challenges for mapping QTLs that affect complex behaviours. The first challenge relates to genetic sampling, as the QTLs that have been mapped so far encompass only a tiny fraction of naturally occurring variation. When carrying out QTL mapping, it is necessary to use a genetic base that is as broad as possible to build a complete picture of the variation affecting a particular trait. This can be accomplished by deriving a large number of inbred lines from a natural population and screening them for the trait of interest to identify extreme phenotypes. Alternatively, one could construct lines that have undergone artificial selection for increased or decreased expression of the behaviour. Selection has the advantage that mean trait values are likely to be pushed far beyond the range seen in the initial population. Guidelines for the design of artificial selection experiments are well-established[24] and include: starting with a large base population that has been recently derived from nature, to ensure that most common alleles affecting the trait have been sampled; replicating all selection lines; including replicate control lines to assess INBREEDING DEPRESSION and to account for fluctuations in the environment at each generation; minimizing the effects of GENETIC DRIFT during selection by maximizing population size; and stopping selection after approximately 25 generations, after which spontaneous mutations are expected to contribute to the response to selection[52–54].

The second challenge is to map QTLs to the level of contributing genetic loci. Achieving this by standard RECOMBINATION MAPPING alone requires vast numbers of recombinants, and an ultra-high-density map of polymorphic molecular markers. In the past, the number of available markers has been limiting, but large numbers of easily scored SNPs and length variants have been identified recently[55], and additional polymorphisms can be readily found by direct sequencing. Increasing the number of recombination events can be achieved by constructing very large F_2 or BACKCROSS mapping populations (10,000 individuals or more), or large ADVANCED INTERCROSS POPULATIONS, from which new and large populations of RILs can then be developed.

Complementation analyses. Once QTL regions have been identified, complementation analyses can be used to narrow the QTL interval and identify candidate genes. *D. melanogaster* geneticists are in the privileged position to be able to quickly nominate candidate genes for further study by using quantitative complementation tests to deficiencies and mutations. Quantitative deficiency mapping involves crossing the two parental lines used to map the QTLs using publicly available deficiencies that span each QTL region. Failure of a deficiency to complement the QTL alleles occurs when the difference in behaviour between the two parental strains is greater in the genetic background of the DEFICIENCY CHROMOSOME than in the background of the BALANCER CHROMOSOME[56]. Failure to complement can occur if the deficiency exposes a QTL in the two parental strains with different allelic effects on the trait, or if there is epistasis between the deficiency chromosome and other QTLs affecting the trait in the parent strains. In either case, the deficiency can be said to uncover a QTL affecting the trait, but the mode of action (allelic or epistatic) will not be known until further tests are done. The use of overlapping deficiencies can map QTLs to sub-centimorgan intervals[48,56].

Until recently, the interpretation of these tests was compromised because the available deficiencies were generated in different genetic backgrounds, which often contained visible morphological mutations, and the breakpoints were only defined cytologically. The use and precision of deficiency complementation has been greatly advanced by the DrosDel and Exelixis collections of small deletions with molecularly defined breakpoints that are induced in isogenic backgrounds[57].

After the QTL interval has been narrowed and the number of possible candidate genes has been reduced by complementation analyses to deficiencies, complementation tests to all genes in the region for which mutants are available can identify candidate genes for further study. Quantitative complementation tests to mutations follow the same procedure and logic as deficiency complementations[48,58]. Again, the Exelixis collection of co-isogenic *P* and *piggyBac* insertions will facilitate this approach[47].

Linkage disequilibrium mapping. Linkage disequilibrium mapping capitalizes on historical recombination and can be used both to determine whether candidate genes are associated with variation in a behaviour in a natural population and to identify polymorphisms that contribute to this variation[24]. The method is based on establishing statistically significant associations between sequence variants in candidate genes and phenotypic values. The resolution of the method depends on local recombination rates, the number of generations following the mutation event, population demography, and the density of polymorphic markers surveyed. In regions of normal recombination in *D. melanogaster*, LD decays rapidly within several hundred bp[59], which provides enough resolution to identify candidate genes and, ultimately, QUANTITATIVE TRAIT NUCLEOTIDES (QTNs)[58].

Box 2 | **Using permutation tests for linkage disequilibrium mapping**

Typically, linkage disequilibrium mapping studies involve testing many polymorphisms per candidate gene for associations with phenotype. Therefore, one cannot use the standard $p<0.05$ criterion for statistical significance, as 5% of the tests would be expected to be significant by chance alone. Two kinds of permutation tests, which address two subtly different questions, can be used to solve the problem of multiple testing. The first question is whether the candidate gene affects the behavioural phenotype. If so, one would expect more significant genotype-phenotype associations than expected by chance. To test this, the phenotypic data are randomly permuted among the genotypes for each allele, that is, the phenotypes are randomly associated with the genotypes. Tests for genotype-phenotype association are conducted on the permuted data, and the total number of associations at $p<0.05$ is recorded. The process of scrambling and random association of genotype and phenotype, followed by recording the total number of significant associations, is repeated 1,000 times. This yields an empirical distribution of numbers of significant associations under the null hypothesis where there is no significant genotype-phenotype association. If the number of significant associations observed in the analysis of the real data exceeds the 950th highest number in the permuted data set, we can infer that the candidate gene is associated with naturally occurring variation in behaviour[83].

The second question is whether particular molecular polymorphisms in the candidate gene are associated with naturally occurring variation in the trait. To assess whether any single association is more significant than expected by chance, the same permutation is done, but only the lowest *p*-value for each permuted data set is recorded. Any single association with a *p*-value lower than the 950th lowest *p*-value in the permuted data set is associated with variation in behaviour, and is either the causal quantitative trait nucleotide, or is in linkage disequilibrium with the causal quantitative trait nucleotide[59].

BALANCER CHROMOSOME
A chromosome with one or more inverted segments that suppress recombination, ideally over the length of the chromosome. It is usually identified in crosses by a dominant marker, and carries at least one recessive lethal mutation. They allow lethal mutations to be maintained without selection, as the only offspring that will be viable from an intercross will be those that carry the mutation and are heterozygous for the balancer chromosome.

QUANTITATIVE TRAIT NUCLEOTIDE
Molecular polymorphisms(s) associated with naturally occurring variation in a quantitative trait.

CHROMOSOME SUBSTITUTION LINE
A stock in which a single homozygous chromosome from one strain is introduced into the homozygous genetic background of a second, unrelated strain. It is possible to construct chromosome substitution lines in a few generations using *D. melanogaster* balancer chromosomes.

The principle of LD mapping resembles that of standard QTL mapping. The requirements are: a sample of individuals from a population for which quantitative measurements of the behavioural phenotype have been obtained; and data on molecular polymorphisms at a candidate gene — which can be identified from *P*-element mutagenesis screens, QTL mapping or transcriptional profiling (see below). For each polymorphism, a statistical test (ANOVA or student's *t*-test) is performed to determine whether differences in the trait phenotype are associated with a molecular marker genotype. If so, the true causal polymorphism is in LD with the molecular marker.

Typically, LD mapping studies involve the testing of many polymorphisms per candidate gene for associations with a phenotype. This poses a multiple testing problem, as the probability of finding associations by chance increases when the same data set is sampled multiple times. Permutation tests can be used to determine an appropriate statistical threshold for the detection of significant associations (BOX 2).

Linkage disequilibrium mapping studies require large sample sizes. If a polymorphism in a single candidate gene were to account for 1% of the total phenotypic variation in a trait, a sample of 2,100 individuals would be needed for each homozygous marker genotype to detect this effect (FIG. 3). If the alternate marker polymorphisms associated with the trait both occurred with a frequency of 0.5, this would require a total sample of 8,400 individuals; the total sample size required would be even larger for more extreme marker allele frequencies. However, using *D. melanogaster* as a model system we can employ strategies to reduce the phenotypic variation that take advantage of the ability to make replicate measurements of homozygous genotypes. Constructing inbred lines and measuring *n* individuals from each reduces the within-line environmental standard error by the square root of *n*. Genetic variation unlinked to the candidate gene can be eliminated by substituting homozygous wild-derived chromosomes into a common inbred background (chromosome substitution). The power of these approaches in *D. melanogaster* is illustrated by a study in which only about 200 CHROMOSOME SUBSTITUTION LINES were required to detect an association at a candidate gene that accounted for 2% of the total phenotypic variation[58]. Finally, constructing near-isoallelic lines of the wild-type candidate-gene alleles considerably increases the power to detect associations with small effects[60,61].

It is rare to be able to conclude definitively that a polymorphism in a candidate gene associated with variation in a trait actually causes the variation, as spurious LD could arise from recent ADMIXTURE. In addition, using only a subset of polymorphic markers within the candidate gene leaves open the possibility that the true causal polymorphism is in LD with the candidate site, but actually lies outside it[1,2]. To alleviate the latter problem, it will be necessary in the future to obtain full DNA sequences of the candidate genes for each allele to identify all polymorphisms and to extend the analysis to neighbouring genetic regions to document the extent of LD precisely.

Transcriptional profiling

Whole-genome transcriptional profiling is another approach that can be used to identify genes that regulate complex behaviours. Specifically, changes in the expression of a gene in the background of single mutations that affect a trait[22], or between lines selected for different phenotypic values of the trait[62], implicate the involvement of the gene in a particular behaviour. However, transcript abundance itself is a quantitative trait, so the design and analysis issues discussed above also pertain to whole-genome expression analysis. Moreover, RNA extraction, labelling and hybridization during profiling experiments, as well as analysis of the signal, add additional sources of variation.

Controlling for biological variation. To control for biological variation between individuals in microarray experiments — that is, for confounding environmental variation and genotype by environment interactions — all genotypes compared on an array must be reared simultaneously under standard conditions. Pools of 50 or more flies of each genotype are required to produce sufficient RNA for a single hybridization, and care should be taken to ensure that each replicate vial in which the animals are reared is equally represented in the sample. As the abundances of many transcripts show circadian variation[30,31], all samples should be prepared at the same time of day. At least two independent pools of individuals, RNA extractions and hybridizations are required for each combination of factors to be tested in the statistical analysis (BOX 1).

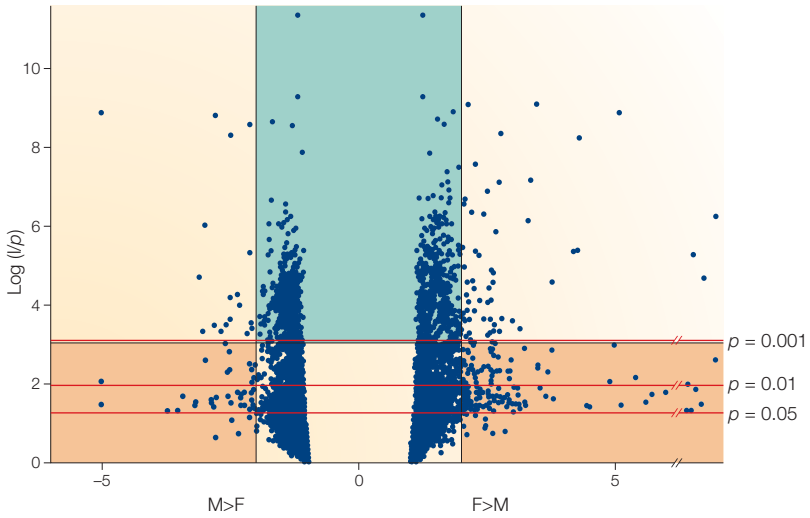


Figure 5 | **Microarray analysis of sex-biased transcripts in *Drosophila melanogaster*.** Twelve replicates of RNA extracted from male (M) and female (F) heads of co-isogenic *D. melanogaster P*-element insertion lines were hybridized to Affymetrix GeneChip microarrays[22]. The ratio of transcript abundance for each probe set on the array for males and females was determined. Data were analysed by ANOVA to establish *p*-values for determining significant differences in expression between males and females for each probe set. The graph shows a 'volcano' plot, which compares *p*-values for differences in transcript abundance (*y*-axis), with an arbitrary significance threshold based on a fold-change in transcript abundance between each sex (*x*-axis). Probe sets are indicated by circles. The bottom, central and top horizontal red lines denote thresholds of $p = 0.05$, $p = 0.01$ and $p = 0.001$, respectively. The vertical red lines denote two-fold thresholds. M>F and F>M designate regions of the graph where the ratio of expression of a gene in males is greater than in females, and vice versa, respectively. A total of 666 probe sets are significantly different at $p<0.001$ (those in the area above the line corresponding to this value); of these, only 90 had differences of two-fold or greater. The green area contains the 576 probe sets that would have been false negatives based on a two-fold significance threshold criterion. The orange areas would have been false positives, which do not reach formal statistical significance at $p<0.001$, despite a two-fold or greater difference in transcript abundance. Adapted, with permission, from *Nature Genetics* REF. 22 © (2003) Macmillan Magazines Ltd.

Statistical analysis of microarray data. Many array analyses of changes in transcript abundance focus on pair-wise comparisons[62] and commonly use a fold-change criterion for significance. However, experiments often have multiple possible comparisons, for example between males and females, different genotypes and, possibly, different environmental conditions. ANOVA models are therefore appropriate for the analysis of expression microarray data (BOX 1).

Because partitioning of the variance in transcript abundance for analysis by ANOVA requires that the experimental design is balanced ($2ijn$ arrays are needed for two sexes, i genotypes, j treatments and n replicates per sex, genotype and treatment), the number of arrays needed to perform even a small experiment can become large. For example, an experiment using high-density oligonucleotide microarrays (Affymetrix GeneChips) comparing males and females of two genotypes, with two replicates per sex for each genotype, requires eight arrays. Sample sizes for EST arrays must be at least doubled because the intensity of the signal is based on competition between samples labelled with different fluorescent dyes; to control for dye effects — that is, differential labelling of particular cDNAs with each of the two dyes — the experiments must be repeated with the samples labelled with the reciprocal dyes[63]. Various 'loop designs' have been proposed, in which each member of a set of samples is compared with a different member of the set in such a way that each sample serves as a reference for at least one other sample[63]. Whereas such designs reduce the numbers of samples required, these analyses are not balanced and do not allow an orthogonal partitioning of the variance. They are, therefore, not recommended unless samples are difficult to procure and cost is prohibitive.

Accounting for multiple tests. The results of statistical analyses of transcript abundance generate a *p*-value for each term in the ANOVA (for example, for sex, genotype, treatment and their interaction terms, see BOX 1) for each of the elements on the array. The *p*-value is the probability that nonsignificant effects will be called significant — the false-positive rate. Given that there are about 14,000 *D. melanogaster* genes to evaluate[3], it is clear that adopting the standard $p<0.05$ false-positive rate is too lenient, as 700 genes would be deemed significant by chance alone. One method to control the false-positive rate is to reduce the *p*-value at which terms are considered to be significant to 0.001 or less[22]; only 14 tests are expected to be this extreme by chance.

Another possible way of identifying genuine positive signals would be to assign significance to arbitrary fold-change cut-offs. However, as shown in FIG. 5, this method is not suitable for this type of analysis as it results in the identification of an extremely large number of false positives and the incorrect designation of a large number of false negatives. Evaluating the data according to statistical significance levels greatly increases the reliability and sensitivity of detecting changes in transcript abundance. Reducing the *p*-value to an arbitrary level might be overly conservative, however, and controlling the false discovery rate (FDR) is gaining acceptance as the optimal method for assessing significance of whole-genome expression data[64]. The FDR is the proportion of false positives among all significant effects and, in practical terms, represents the rate at which significant effects fail to replicate in subsequent experiments[64].

Assessment of candidate genes from profiling experiments. Expression microarrays can reveal the extent of transcriptional disruption that results from single mutations affecting behaviours in co-isogenic backgrounds, and also provide insights into changes in gene expression that accompany artificial selection for behavioural traits (FIG. 1). For example, studies on olfactory behaviour in co-isogenic *P*-element insertion lines of *D. melanogaster* identified 14 lines that showed an aberrant olfactory avoidance response[10]. Transcriptional profiling of five of these mutations in *smell impaired* (*smi*) genes and their control showed that 530 genes were significantly up- or downregulated in one or more *smi* mutant backgrounds[22]. Similarly, 250 genes showed two-fold or greater differences in expression between two lines selected for high and low geotaxis behaviour[62].

ADMIXTURE
The mixture of two or more genetically distinct populations.

Table 1 | ***Drosophila melanogaster* behavioural genes with pleiotropic effects**

Gene	Protein encoded	Traits affected	References
Shaker	Voltage-gated potassium channel	Courtship; lifespan gustation; olfactory behaviour	19,22
Cryptochrome	Cryptochrome	Circadian rhythms; geotaxis	62,74
Pigment dispersing factor	Pigment dispersing factor	Circadian rhythms; geotaxis	62,74
cAMP-dependent protein kinase type II	cAMP-dependent protein kinase type II	Circadian rhythms; sensitivity to cocaine and ethanol; ovarian development	75
dunce	cAMP-specific phosphodiesterase	Learning and memory; locomotor rhythms; ethanol tolerance; olfaction; food search; aggregation behaviour; female fertility; lifespan	16,20,76–81

In the former case, because changes in gene expression are observed in a co-isogenic background, these changes are attributable to genes co-regulated by the *smi* mutation. In the latter case, the genetic backgrounds of the selected lines have diversified through changes in allele frequencies. The observed expression differences could therefore be due to genetic variation in genes affecting geotaxis or other genes that differ between the lines as a result of random drift, or genes that are co-regulated by genetically variable loci.

Genes with differential transcript abundance in genotypes exhibiting different behavioural phenotypes are candidate genes affecting the behaviour under study. However, genes that show transcriptional co-regulation do not necessarily contribute to the behavioural phenotype. If mutations in these genes are available, one can test whether the co-regulated candidate gene has a direct effect on the behaviour by comparing the behaviour of mutant and wild-type flies[62]. Although differences in genetic backgrounds of most existing mutations compromise this approach, these can be minimized by chromosome substitution and backcrossing to a common inbred stock[62]. Three of the four mutations in candidate genes tested in the study of geotactic behaviour described above had direct effects on the trait[62].

To address the question of whether genes whose expression is co-regulated by the mutation also interact epistatically with the mutation, quantitative complementation tests can be performed with existing mutant stocks[22] by crossing the mutations at the candidate genes to the tested mutations and their co-isogenic control strains[22]. In the study of the *smi* mutations described above, 14 of 21 mutations in candidate genes tested (67%) exhibited epistasis for olfactory behaviour — that is, they were smell-impaired as double heterozygotes (transheterozygotes) with a recessive *smi* mutation, but not as heterozygotes in the control background[22]. So, transcriptional profiling studies present an effective tool for the discovery of new candidate behavioural genes.

Profiling of transgenic lines. Expression microarrays can also be used to assess the consequences of targeted overexpression of transgenes. Crossing a homozygous *GAL4* driver line with a line homozygous for the UAS-transgene construct results in hybrid F_1 offspring in which the transgene is expressed in cells that express Gal4 (REF 23). Often, the parental driver line and the line containing the UAS transgene have different genetic backgrounds. To account for this, transcriptional profiles can be analyzed by comparing F_1 values to the expected midparent values[65] — that is, the predicted average values of the two homozygous parental strains. In this case, however, resolution becomes limited to the detection of large additive effects, as epistasis in the F_1 as a consequence of the mixed genetic background from the parental strains remains a confounding factor.

Epistasis

In quantitative genetics terminology, epistasis refers to non-additive interactions between loci that affect a trait[24]. Identifying epistatic interactions is important for gaining insights in genetic pathways or networks of genes affecting a trait. The simplest quantitative genetic approach to this in *D. melanogaster* is to evaluate epistatic interactions between co-isogenic mutations that affect a trait by generating all possible transheterozygotes. This method was first applied to the set of co-isogenic *P*-element *smi* mutations described above, revealing an ensemble of interacting genes with defined enhancer or suppressor effects[66].

A simple ANOVA of the behavioural data from such transheterozygotes will reveal whether there is variation among the different genotypes. If so, the data are analysed as a HALF-DIALLEL CROSS. For each mutation, its average phenotypic value as a transheterozygote compared with all other mutations is calculated and expressed as the deviation from the overall mean to give a value known as the general combining ability (GCA). The specific combining ability (SCA) of each double heterozygote[67,68] — defined as the difference between the observed phenotypic value of the genotype from that expected from the sum of the corresponding GCAs of contributing mutations — is also calculated. The SCA, therefore, estimates the magnitude of epistatic interactions.

This half-diallel design can only be used to detect epistasis if the mutations are co-isogenic, as the SCA would otherwise estimate both dominance and epistasis. Another drawback to this design is that the number of crosses necessary to evaluate n mutations is $n(n–1)/2$, which increases exponentially as n increases. For example, analysis of epistasis among 12 *smi* mutations required constructing 66 transheterozygous genotypes[66].

More extensive analyses of epistasis require constructing all nine possible two-locus genotypes for two

HALF-DIALLEL CROSS
Construction of all possible $n(n–1)/2$ heterozygous genotypes between n homozygous lines, excluding reciprocal crosses.

potentially epistatic loci with two alleles at each locus — instead of only the transheterozygotes required in the diallel method — and computing additive, dominance and all possible epistatic effects[69]. Care must be taken when applying this method to *P*-element insertional mutations, as variation in the numbers of *P*-element copies (from 0 in the *P*-element-free host strain to four in the double mutant homozygote) could be potentially confounding[70].

Pleiotropy

Screens for quantitative effects of *P*-element insertions[10,71,72], as well as transcriptional profiling[22,30,31,62], indicate that large numbers of interacting genes affect behaviour. This, in turn, implies massive pleiotropy, challenging the view that complex behaviours are specified by specific dedicated regulatory genes[73]. Behavioural and developmental pleiotropy is apparent from the emergence of the same loci in independent mutant screens, and also from transcriptional profiling studies that show that most co-regulated transcripts in mutant backgrounds represent gene products that would not be expected *a priori* to contribute to the behavioural phenotype. Some examples of behavioural genes with pleiotropic effects are given in TABLE 1.

Future perspectives

The emerging picture for the genetic architecture of complex behaviours is of dynamic, overlapping, epistatic networks of pleiotropic genes. This realization raises a number of central questions. First, what is the extent of epistasis of each gene in a network, and how plastic are these interactions? Second, how do genetic networks that drive behaviour change during development? Third, how are genetic networks modified through interactions with the physical and social environments, and to what extent do these networks differ between the sexes? Finally, how do genetic networks that drive behaviours evolve? These challenges define future areas of research for behavioural geneticists and can be met only by careful quantification of phenotypes, an awareness that genetic backgrounds and sexually dimorphic effects are crucial parameters, and an appreciation of the pleiotropy of behavioural genes.

Microarray expression studies have great promise for determining the basis of complex behavioural traits, but experiments performed so far have relied on RNA from whole flies or whole fly heads. So, the observed transcriptional profiles are composites that are summed over different heterogeneous neuronal subpopulations. An important challenge for the emerging field of quantitative neurogenomics is to develop methods to analyse transcriptional profiles in single neurons or in homogeneous neuronal populations. Placing cell-specific genetic networks in the context of neural connectivity is the next frontier in our understanding of neural and genetic principles that drive behaviours.

1. Mackay, T. F. C. Quantitative trait loci in *Drosophila*. *Nature Rev. Genet.* **2**,11–20 (2001).
2. Mackay, T. F. C. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).
3. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
4. Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G. & Benzer, S. *dunce*, a mutant of *Drosophila* deficient in learning. *Proc. Natl Acad. Sci. USA* **73**, 1684–1688 (1976).
5. Quinn, W. G., Harris, W. A. & Benzer, S. Conditioned behavior in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **71**, 708–712 (1974).
6. Konopka, R. J. & Benzer, S. *Clock* mutants of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
7. Greenspan, R. J. & Ferveur, J.-F. Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**, 205–232 (2000).
8. Moehring, A. J. & Mackay, T. F. C. The quantitative genetic basis of male mating behavior in *Drosophila melanogaster*. *Genetics* **167**, 1249–1263 (2004).
9. Osborne, K. A. *et al.* Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* **277**, 834–836 (1997).
A classic paper, reporting the behavioural effects of naturally occurring polymorphisms in the *foraging* gene, which encodes a cGMP-dependent protein kinase. The authors show that subtle alterations in enzyme activity profoundly alter larval foraging behaviour in *D. melanogaster*.
10. Anholt, R. R. H., Lyman, R. F. & Mackay, T. F. C. Effects of single *P*-element insertions on olfactory behavior in *Drosophila melanogaster*. *Genetics* **143**, 293–301 (1996).
11. Chen, S., Lee, A. Y., Bowens, N. M., Huber, R. & Kravitz, E. A. Fighting fruit flies: a model system for the study of aggression. *Proc. Natl Acad. Sci. USA* **99**, 5664–5668 (2002).
12. Benzer, S. Behavioral mutants of *Drosophila* isolated by countercurrent distribution. *Proc. Natl Acad. Sci. USA* **58**, 1112–1119 (1967).
13. Hirsch, J. & Erlenmeyer-Kimling, L. Sign of taxis as a property of the genotype. *Science* **143**, 835–835 (1961).
Reports one of the earliest experiments in fly behavioural genetics. The authors demonstrate that artificial selection can be used to generate flies that prefer to migrate either upwards or downwards in a geotactic maze. The selection lines reported in this paper were used 40 years later in the transcriptional profiling studies described in reference 62.
14. Quinn, W. G., Sziber, P. P. & Booker, R. The *Drosophila* memory mutant *amnesiac*. *Nature* **277**, 212–214 (1979).
15. Weber, K. E. An apparatus for measurement of resistance to gas-phase reagents. *Drosoph. Inf. Serv.* **67**, 91–93 (1988).
16. Moore, M. S. *et al.* Ethanol intoxication in *Drosophila*: genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* **93**, 997–1007 (1998).
17. Rothenfluh, A. & Heberlein, U. Drugs, flies, and videotape: the effects of ethanol and cocaine on *Drosophila* locomotion. *Curr. Opin. Neurobiol.* **12**, 639–645 (2002).
18. Hall, J. C. The mating of a fly. *Science* **264**, 1702–1714 (1994).
19. Dubnau, J. & Tully, T. Gene discovery in *Drosophila*: new insights for learning and memory. *Annu. Rev. Neurosci.* **21**, 407–444 (1998).
20. Waddell, S. & Quinn, W. G. Flies, genes and learning. *Annu. Rev. Neurosci.* **24**, 1283–1309 (2001).
21. McKenna, M., Monte, P., Helfand, S. L., Woodard, C. & Carlson, J. A simple chemosensory response in *Drosophila* and the isolation of *acj* mutants in which it is affected. *Proc. Natl Acad. Sci. USA* **86**, 8118–8122 (1989).
22. Anholt, R. R. H. *et al.* The genetic architecture of odor-guided behavior in *Drosophila*: epistasis and the transcriptome. *Nature Genet.* **35**,180–184 (2003).
Quantifies the extent of transcriptional co-regulation among a set of co-isogenic mutant lines with single *P*-element insertions that affect olfactory avoidance behaviour and illustrates how transcriptional profiling together with quantitative complementation tests for epistasis can be used for gene discovery.
23. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
24. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* (Longman, Essex, 1996).
25. McGuire, S. E., Le, P. T. & Davis, R. L. The role of *Drosophila* mushroom body signaling in olfactory memory. *Science* **293**,1330–1333 (2001).
26. Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**, 1091–1119 (2001).
27. Ben-Shlomo, R. & Kyriacou, C. P. Circadian rhythm entrainment in flies and mammals. *Cell. Biochem. Biophys.* **37**,141–156 (2002).
28. Ashmore, L. J. & Sehgal, A. A fly's eye view of circadian entrainment. *J. Biol. Rhythms* **18**, 206–216 (2003).
29. Cirelli, C. Searching for sleep mutants of *Drosophila melanogaster*. *Bioessays* **25**, 940–949 (2003).
30. McDonald, M. J. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
This paper, together with reference 31, presents the first expression microarray analyses of a behavioural trait in *D. melanogaster*. Although the observed transcriptional profiles differ between these papers, they both demonstrate that circadian time affects not only expression of genes that are directly implicated in the biological clock, but has far reaching implications for the transcription of many genes, including those implicated in several behaviours.
31. Claridge-Chang, A. *et al.* Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron* **32**, 657–671 (2001).
32. Norga, K. K. *et al.* Quantitative analysis of bristle number in *Drosophila* mutants identifies genes involved in neural development. *Curr. Biol.* **13**, 1388–1396 (2003).
33. Crabbe, J. C., Wahlsten, D. & Dudek, B. C. Genetics of mouse behavior: interactions with laboratory environment. *Science* **284**, 1670–1672 (1999).
34. Mackay, T. F. C., Hackett, J. B., Lyman, R. F., Wayne, M. L. & Anholt, R. R. H. Quantitative genetic variation of odor-guided behavior in a natural population of *Drosophila melanogaster*. *Genetics* **144**, 727–735 (1996).
35. Devaud, J. M. Experimental studies of adult *Drosophila* chemosensory behaviour. *Behav. Processes* **64**, 177–196 (2003).
36. Belgacem, Y. H. & Martin, J. R. Neuroendocrine control of a sexually dimorphic behavior by a few neurons of the pars intercerebralis in *Drosophila*. *Proc. Natl Acad. Sci. USA* **99**, 15154–15158 (2002).
37. Helfrich-Forster, C. Differential control of morning and evening components in the activity rhythm of *Drosophila melanogaster* — sex-specific differences suggest a different quality of activity. *J. Biol. Rhythms* **15**, 135–154 (2000).

38. Jin, W. *et al.* The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genet.* **29**, 389–395 (2001).
39. Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745 (2003).
40. Parisi, M. *et al.* Paucity of genes on the *Drosophila X* chromosome showing male-biased expression. *Science* **299**, 697–700 (2003).
41. Fujii, S. & Amrein, H. Genes expressed in the *Drosophila* head reveal a role for fat cells in sex-specific physiology. *EMBO J.* **21**, 5353–5363 (2002).
42. Rice, W. R. Sexually antagonistic genes: experimental evidence. *Science* **256**, 1436–1439 (1992).
43. Levene, H. Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953).
44. Greenspan, R. J. A kinder, gentler genetic analysis of behavior: dissection gives way to modulation. *Curr. Opin. Neurobiol.* **7**, 805–811 (1997).
45. Ganguly, I., Mackay, T. F. C. & Anholt, R. R. H. *Scribble* is essential for olfactory behavior in *Drosophila*. *Genetics* **164**, 1447–1457 (2003).
46. Kulkarni, N. H., Yamamoto, A. H., Robinson, K. O., Mackay, T. F. C. & Anholt, R. R. H. The DSC1 channel, encoded by the *smi60E* locus, contributes to odor-guided behavior in *Drosophila melanogaster*. *Genetics* **161**, 1507–1516 (2002).
47. Thibault, S. T. *et al.* A complementary transposon tool kit for *Drosophila melanogaster* using *P* and *piggyBac*. *Nature Genet.* **36**, 283–287 (2004).
48. Fanara, J. J., Robinson, K. O., Rollmann, S. M., Anholt, R. R. H. & Mackay, T. F. C. *Vanaso* is a candidate quantitative trait gene for olfactory behavior in *Drosophila melanogaster*. *Genetics* **162**, 1321–1328 (2002).
49. Gleason, J. M., Nuzhdin, S. V. & Ritchie, M. G. Quantitative trait loci affecting a courtship signal in *Drosophila melanogaster*. *Heredity* **89**, 1–6 (2002).
50. Zeng, Z. B. Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468 (1994).
51. Doerge, R. W. & Churchill, G. A. Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294 (1996).
52. Hill, W. G. Predictions of response to artificial selection from new mutations. *Genet. Res.* **40**, 255–278 (1982).
53. Mackay, T. F., Fry, J. D., Lyman, R. F. & Nuzhdin, S. V. Polygenic mutation in *Drosophila melanogaster*: estimates from response to selection of inbred strains. *Genetics* **136**, 937–951 (1994).
54. Rodriguez-Ramilo, S. T., Perez-Figueroa, A., Fernandez, B., Fernandez, J. & Caballero, A. Mutation-selection balance accounting for genetic variation for viability in *Drosophila melanogaster* as deduced from an inbreeding and artificial selection experiment. *J. Evol. Biol.* **17**, 528–541 (2004).
55. Berger, J. *et al.* Genetic mapping with SNP markers in *Drosophila*. *Nature Genet.* **29**, 475–481 (2001).
56. Pasyukova, E. G., Vieira, C. & Mackay, T. F. C. Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics* **156**, 1129–1146 (2000).
Introduces the method of quantitative complementation mapping using deficiencies, and demonstrates how this method can be used to narrow QTL regions to a small number of candidate genes that affect variation in lifespan in *D. melanogaster*.
57. Parks, A. L., *et al.* Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. *Nature Genet.* **36**, 288–292 (2004).
58. DeLuca, M. *et al. Dopa decarboxylase (Ddc)* affects variation in *Drosophila* longevity. *Nature Genet.* **34**, 429–433 (2003).
Demonstrates how LD mapping can be used as a powerful method to identify not only a candidate gene, but also polymorphisms (QTNs) and haplotypes within such genes that contribute to natural variation in a complex trait, in this case lifespan.
59. Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. C. Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017 (1998).
60. Lyman, R. F. & Mackay, T. F. C. Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in *Drosophila melanogaster*: the *Delta-Hairless* gene region. *Genetics* **149**, 983–998 (1998).
61. Robin, C., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. C. *hairy*: a quantitative trait locus for *Drosophila* sensory bristle number. *Genetics* **162**, 155–164 (2002).
62. Toma, D. P., White, K. P., Hirsch, J. & Greenspan, R. J. Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait. *Nature Genet.* **31**, 349–353 (2002).
63. Kerr, M. K. & Churchill, G. A. Experimental design for gene expression microarrays. *Biostatistics* **2**, 183–201 (2001).
Contains excellent advice on the design and analysis of expression microarrays.
64. Storey, J. D. & Tibshirani, R. Statistical significance for genome-wide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
Introduces assessment of the false discovery rate to estimate false positives in transcriptional profiling experiments. The statistical method described will probably become the standard procedure for microarray analysis.
65. Borrás, T. *et al.* Transcription profiling in *Drosophila* eyes that overexpress the human glaucoma-associated trabecular meshwork-inducible glucocorticoid response protein/myocilin (TIGR/MYOC). *Genetics* **163**, 637–645 (2003).
66. Fedorowicz, G. M., Fry, J. D., Anholt, R. R. H. & Mackay, T. F. C. Epistatic interactions between smell-impaired loci in *Drosophila melanogaster*. *Genetics* **148**, 1885–1891 (1998).
Describes the first application of the half-diallel cross design for the analysis of epistasis among behavioural genes, and reports an unexpectedly large extent of epistasis among 12 independently isolated co-isogenic *P*-element insertion lines that affect olfactory avoidance behaviour. This paper formed the basis for the subsequent transcriptional profiling study described in reference 22.
67. Sprague, G. F. & Tatum, L. A. General *vs.* specific combining ability in single crosses of corn. *J. Amer. Soc. Agron.* **34**, 923–932 (1942).
68. Griffing, B. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* **9**, 463–493 (1956).
69. Cockerham, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).
70. Clark, A. G. & Wang, L. Epistasis in measured genotypes: *Drosophila P*-element insertions. *Genetics* **147**, 157–163 (1997).
71. Eberl, D. F., Duyk, G. M. & Perrimon, N. A genetic screen for mutations that disrupt an auditory response in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **94**, 14837–14842 (1997).
72. Ejima, A., Nakayama, S. & Aigaki, T. Phenotypic association of spontaneous ovulation and sexual receptivity in virgin females of *Drosophila melanogaster* mutants. *Behav. Genet.* **31**, 437–444 (2001).
73. Baker, B. S., Taylor, B. J. & Hall, J. C. Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* **105**, 13–24 (2001).
74. Williams, J. A. & Sehgal, A. Molecular components of the circadian system in *Drosophila*. *Annu. Rev. Physiol.* **63**, 729–755 (2001).
75. Park, S. K., Sedore, S. A., Cronmiller, C. & Hirsh, J. Type II cAMP-dependent protein kinase-deficient *Drosophila* are viable but show developmental, circadian, and drug response phenotypes. *J. Biol. Chem.* **275**, 20588–20596 (2000).
76. Kauvar, L. M. Defective cyclic adenosine 3′:5′-monophosphate phosphodiesterase in the *Drosophila* memory mutant *dunce*. *J. Neurosci.* **2**, 1347–1358 (1982).
77. Levine, J. D., Casey, C. I., Kalderon, D. D. & Jackson, F. R. Altered circadian pacemaker functions and cyclic AMP rhythms in the *Drosophila* learning mutant *dunce*. *Neuron* **13**, 967–974 (1994).
78. Martin, F., Charro, M. J. & Alcorta, E. Mutations affecting the cAMP transduction pathway modify olfaction in *Drosophila*. *J. Comp. Physiol. A* **187**, 359–370 (2001).
79. Tinette, S, Zhang, L. & Robichon, A. Cooperation between *Drosophila* flies in searching behavior. *Genes Brain Behav.* **3**, 39–50 (2004).
80. Ueyama, M. & Fuyama, Y. Enhanced cost of mating in female sterile mutants of *Drosophila melanogaster*. *Genes Genet. Syst.* **78**, 29–36 (2003).
81. Dauwalder, B. & Davis, R. L. Conditional rescue of the *dunce* learning/memory and female fertility defects with *Drosophila* or rat transgenes. *J. Neurosci.* **15**, 3490–3499 (1995).
82. Sokal, R. R. & Rohlf, F. J. *Biometry* 2nd edn (W. H. Freeman and Company, New York, 1981).
83. Lai, C., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. C. Naturally occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila melanogaster*. *Science* **266**, 1697–1702 (1994).
84. Sokolowski, M. B. *Drosophila*: genetics meets behaviour. *Nature Rev. Genet.* **2**, 879–890 (2001).

Acknowledgement

Work in the authors' laboratories is supported by grants from the National Institutes of Health.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:
FlyBase: http://flybase.bio.indiana.edu/
dunce | *period* | *rutabaga*

FURTHER INFORMATION

Anholt's homepage: http://www.cals.ncsu.edu/beh_bio/anholt
DrosDel *Drosophila* Isogenic Deficiency Kit:
http://www.drosdel.org.uk
Mackay's homepage:
http://www.cals.ncsu.edu/genetics/mackay/mackay.html
Access to this interactive links box is free online.