The Journal of Experimental Biology 210, 1632-1640
Published by The Company of Biologists 2007
doi:10.1242/jeb.002691

Integrative physiology, functional genomics and the phenotype gap: a guide for comparative physiologists

Julian A. T. Dow
Division of Molecular Genetics, University of Glasgow, Glasgow, G11 6NU, UK
e-mail: j.a.t.dow@bio.gla.ac.uk

Accepted 23 January 2007

Summary

Classical, curiosity-led comparative physiology finds itself at a crossroads. Major funding for classical physiology is becoming harder to find, as grant agencies focus on more molecular approaches or on science with more immediate strategic value to their respective countries. In turn, this shift in funding places Zoology and Animal Science departments under enormous stress: student numbers are buoyant, but how can research funding be maintained at high levels?

Our research group has argued for the redefinition of integrative physiology as the investigation of gene function in an organotypic context in the intact animal. Implicit in this definition is the use of transgenics and reverse genetics to manipulate gene function in a cell-specific manner; this in turn implies the use of a genetically tractable 'model organism'. The significance of this definition is that it aligns integrative physiology with functional genomics. Again, functional genomics draws heavily on reverse genetics to elucidate the function of novel genes. The phenotype gap (the mismatch between what a genetic model organism's genome encodes and the reasons that it has historically been studied) emphasises the need to attract and empower functional biologists: can *all* 13 500 genes in *Drosophila* really be explained in terms of developmental biology? So, by embracing the integrative physiology manifesto, comparative physiologists can not only accelerate their own research, but their functional skills can make them indispensable in the post-genomic endeavour.

Glossary available online at
http://jeb.biologists.org/cgi/content/full/210/9/1632/DC1

Key words: *Drosophila melanogaster*, *Caenorhabditis elegans*, *Danio rerio*, mouse, Malpighian tubule, bioinformatics.

Introduction

This paper is aimed at students and academics who are contemplating whether to attempt real, integrative physiology – particularly in the fruit fly, *Drosophila melanogaster*. It sets out the reasons why this may be desirable and outlines the basic steps needed to obtain the resources required and to start experimenting.

Integrative physiology and functional genomics

Physiology is the study of how a living organism works. Classical physiologists are becoming an endangered species, just as the need for their skills is growing once again. This shortfall is for several reasons, not least the stringency of regulations for animal work but also because non-molecular work is now discouraged by many grant agencies. Nonetheless, cellular and molecular approaches to physiology have proved potent and cost-effective paradigms. Ultimately, however, these reductionist or analytical methodologies can prove restricting: how can we draw on the strengths of these methodologies, while addressing function in the whole organism – the new, 'integrative physiology'?

Our group has argued that, rather than abandoning molecular biology in a return to classical techniques, it is possible to integrate a gene-based approach into studies of the whole organism (Dow and Davies, 2003). Integrative physiology is seen – as in the present motto of the American Physiological Society – as the move '*from gene to tissue to organism*'. How can this be achieved? We have argued that transgenic organisms, in which expression of a particular gene of interest is manipulated upward or downward in a cell-specific manner, provide a potent link between gene and organism (Dow and Davies, 2003). In essence, physiologists are exhorted to adopt the skills of the geneticist, by trying to understand the function of specific genes by disrupting them or modulating them – what is known as 'reverse genetics'. This philosophy is not merely

academic: it was adopted by the UK's Biotechnology and Biological Sciences Research Council in their recent 'Genomics in Animal Function (GAIN)' Initiative.

The attraction of this working definition of integrative physiologist is that it aligns comparative physiology with one of the most important research fields today; that of functional genomics – the understanding of the function of all the genes encoded in a genome. Perhaps one-third of the genes in any genome (so perhaps 5000–10 000) are sufficiently novel that their function cannot be predicted *in silico*; for such genes, a reverse genetic work-up is considered one of the most powerful approaches.

For functional genomics, however, there is a major log jam in the reverse genetic pipeline; there is no point in mutating a particular gene unless the experimenter is able to recognize and study any resulting phenotype. Functional genomics thus requires the broadest range of functional biologists to align themselves with their endeavour. The problem is that model organisms have come to prominence, not for their physiological tractability but for highly focused studies – typically of development. The 'phenotype gap' is therefore the mismatch between the functions encoded by an organism's genome and what it has historically been used to study (Brown and Peters, 1996; Bullard, 2001; Dow, 2003; Wang et al., 2004). The scale of this mismatch for most model organisms emphasizes the opportunities available to comparative physiologists flexible enough to take on a new organism.

It is actually possible to quantify the phenotype gap. By the time the *Drosophila* genome was sequenced in 2000, I estimate that at least 300 000 researcher-years had been devoted to *Drosophila* – the large majority spent on studies of development. *Drosophila* genes are given (more or less witty) names as they are first encountered, and so it is possible to distinguish known, studied genes from those that are merely computer annotated. Of the 13 500 or so genes identified in the first release of the genomic sequence, only 20% were named (Wang et al., 2004). We can thus assume that developmental biology, as a screen for new genes, has now neared saturation. For the other 80% of genes, we need to seek new phenotypes.

The need for model organisms

Once the case for transgenics as a physiological tool is accepted, then the choice of organism is immediately circumscribed. Transgenics is only feasible in the small set of organisms (like mice, fruit flies, worms or zebra fish) known as genetic models (Table 1). Typically, these organisms have fully sequenced genomes and a wealth of freely available mutant stocks, or the resources to create more as required. However, this appears to run contrary to the Krogh principle, beloved of comparative physiologists; that for any physiological question, there is a species in which it can best be studied (Krogh, 1929). When there are perhaps 30 million species of animal (mostly insect) in the world, how can it be sensible to restrict oneself to less than a few tens of species?

The answer is twofold. Firstly, integrative physiology is not doctrinaire; it remains possible to continue working in a target species, while dipping occasionally into the remarkable resources associated with the model organism. For example, an insect physiologist working on an agricultural pest organism could look up the sequence of a particular neuropeptide receptor from the *Drosophila* genome project and use it to design primers that would be likely to pull the gene out of the target organism. Alternatively, if a gene was identified by an advanced molecular technique like suppression subtractive hybridization in the target species, mutants could be sought in the phylogenetically closest model species, and studied there, so as to understand the gene of interest rather better. The second answer is that, just because a model organism has a sequenced genome and a wealth of genetic resources, it is not intrinsically less interesting than a non-model; indeed, if one considers that the Krogh principle applies both to an organism that exemplifies a trait '*and the ease with which it can be studied experimentally*', then model organisms can acquire Krogh status for a surprising range of studies.

Real physiology

So far, these arguments could be seen as philosophical. However, taking *Drosophila* as an example, it is possible to identify several examples of real physiology interacting closely with genetics in order to provide powerful, fundamental insights.

(1) Our understanding of the circadian clock is based on pioneering (and painstaking) screens of insects that failed to eclose from their pupae at the normal time of day, implying that they had lost track of time (Konopka and Benzer, 1971). Mutants discovered in *Drosophila*, like *period* and *timeless*, have influenced the whole field.

(2) Similarly, mutants identified in simple learning paradigms have implicated the cyclic AMP signalling pathway; dunce is a cAMP phosphodiesterase, while rutabaga is an adenylate cyclase (Dudai et al., 1976; Dudai and Zvi, 1985). The work in *Drosophila* thus closely aligns with Kandel's Nobel prize-winning studies on habituation of the gill withdrawal reflex in the non-model sea hare, *Aplysia californica* (Kandel and Schwartz, 1982).

(3) The suspicion that the 24-transmembrane pass voltage-gated ion channels were actually derived from two gene duplication events acting on a 6-transmembrane pass ancestor was dramatically confirmed when a line of flies with legs that continued to shake under ether anaesthesia (Catsch, 1944) were shown to be mutants in a potassium channel gene, named *Shaker* (Kamb et al., 1987). At first sight, one might think it only appropriate that a primitive channel had been identified in a 'primitive' organism, but this would show a grave misunderstanding of evolution. *Drosophila* has continued to evolve in the 400 million years since it diverged from the common ancestor of humans, so it is *different*, rather than *primitive*. Indeed, the Shaker channel of *Drosophila* triggered the discovery of a ubiquitous family of quarter-sized channels that were found even in humans (Salkoff et al., 1992).

Of course, these physiological examples are all drawn from neuroscience. Our group's work, however, has shown that renal function can also be studied to great advantage in *Drosophila* (Dow and Davies, 2001; Dow and Davies, 2003; Dow and Davies, 2006). The sequenced genome allowed the rapid identification of genes encoding diuretic neuropeptides (Cabrero et al., 2002; Coast et al., 2001; Kean et al., 2002; Terhzaz et al., 1999) and their receptors (Johnson et al., 2005; Radford et al., 2002), often before it proved possible in non-model organisms, and indeed these studies paved the way for similar work in other insects (Radford et al., 2004).

Using the tools

The key genetic tool for *Drosophila* physiology is the GAL4 enhancer trap. This has been described in detail many times (Brand and Perrimon, 1993; Sentry et al., 1994) but, in essence, it provides the ability to express genetic constructs of choice in specific cells in an otherwise normal organism – exactly the technological requirements for integrative physiology (Dow and Davies, 2003). Targeted ectopic expression is possible in a wide range of genetic models, but perhaps nowhere as simply as in *Drosophila*. The GAL4/UAS system is binary; that is, a fly is generated by crossing together a 'driver' line (one in which the yeast transcription factor GAL4 is expressed in a desired pattern) and a UAS line (one in which the genetic payload is placed downstream of five copies of the UAS binding site recognized by GAL4). In such flies, the genetic payload is switched on strongly in those cells in which GAL4 is being expressed (Fig. 1).

Where do such useful lines come from? GAL4 drivers are either derived from enhancer trap screens or are made by inserting a gene's control regions next to the *GAL4* gene in a plasmid, which is then incorporated into the germ line of *Drosophila*. Our lab participated in the Kaiser screens (Kaiser, 1993) of the Brand constructs (Brand and Perrimon, 1993), and from 750 lines obtained 50 with patterned expression in the alimentary canal, of which about 10 provided informative and useful expression patterns in the Malpighian tubules (Sözen et al., 1997). This panel of driver lines has been of great use to us, but have also been distributed around the world to other tubule groups. Similarly, most *Drosophila* labs around the world now keep a panel of GAL4 drivers of utility to their research areas.

However, there is still greater sophistication available. What if it is desirable to express a gene of interest not just in a particular spatial pattern but at a particular time? For example, it might be important to express a deleterious construct only in adult *Drosophila* just before study, to prevent pleiotropic effects (or even lethality) in the embryo or larva. This can be achieved with a temperature-sensitive GAL4-binding protein, GAL80. In flies transgenic for a GAL4 driver, a UAS-driven transgene and GAL80, GAL4 is expressed in a spatially restricted pattern, but bound by GAL80, thus preventing it from binding to UAS and activating the transgene. However, when the temperature is raised to 30°C, GAL80 dissociates, allowing GAL4 to bind to UAS and so activate the transgene (McGuire et al., 2004; Suster et al., 2004).

Another strategy is to provide the genetic equivalent of a 'latching' switch for the GAL4/UAS system. If a fly is generated containing a GAL4 driver and a UAS-driven payload, but in the additional presence of a GAL4 transgene downstream of a UAS promoter, then wherever GAL4 is transiently expressed, it will activate the GAL4 transgene, so providing high levels of GAL4 expression indefinitely in that cell (Hassan et al., 2000).

Sometimes, it can be hard to study the effects of a gene because available mutants are lethal. The traditional *Drosophila* genetic technique is to generate X-ray-induced mosaics, in which the mutation is only homozygous in a small population of cells (Becker, 1975). Clever experimental design can allow these cells to be visibly marked; these mutant patches can then be studied in the context of an otherwise normal animal. Yeast technology adds a modern twist to these classical experiments; if the lethal mutation is crossed onto a chromosome that has a yeast flippase recombination target (FRT) site near its centromere, and then yeast flippase (FLP) is driven transgenically, there is a finite chance that the chromosomes will recombine (Golic and Lindquist, 1989; Xu and Harrison, 1994). This has the effect of producing clones of cells carrying two mutant (or two wild-type) chromosomes. Of course, if UAS–FLP is driven with GAL4, it is possible to target specific populations of cells, rather than throughout the organism.

RNA interference (RNAi) by double-stranded RNA, originally employed in *Caenorhabditis elegans* (Fire et al., 1998), has proved a potent means of obtaining hypomorphic

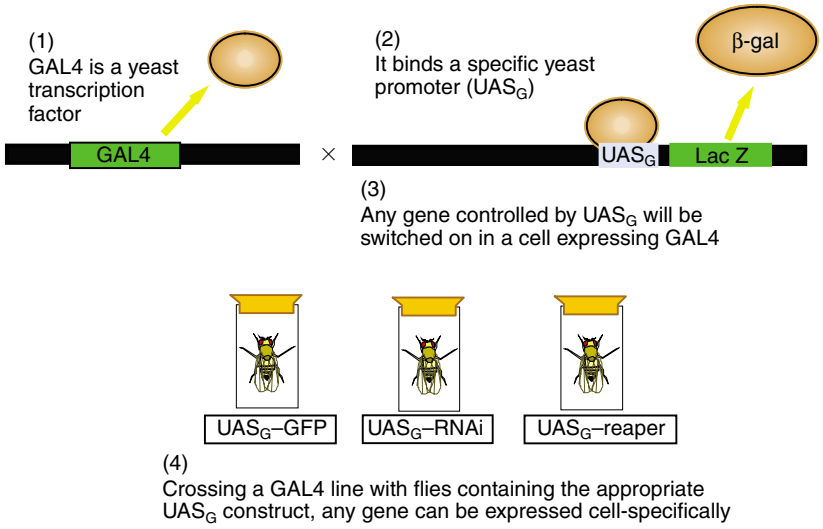

Fig. 1. The GAL4/UAS system. Modified from Dow (Dow, 2007).

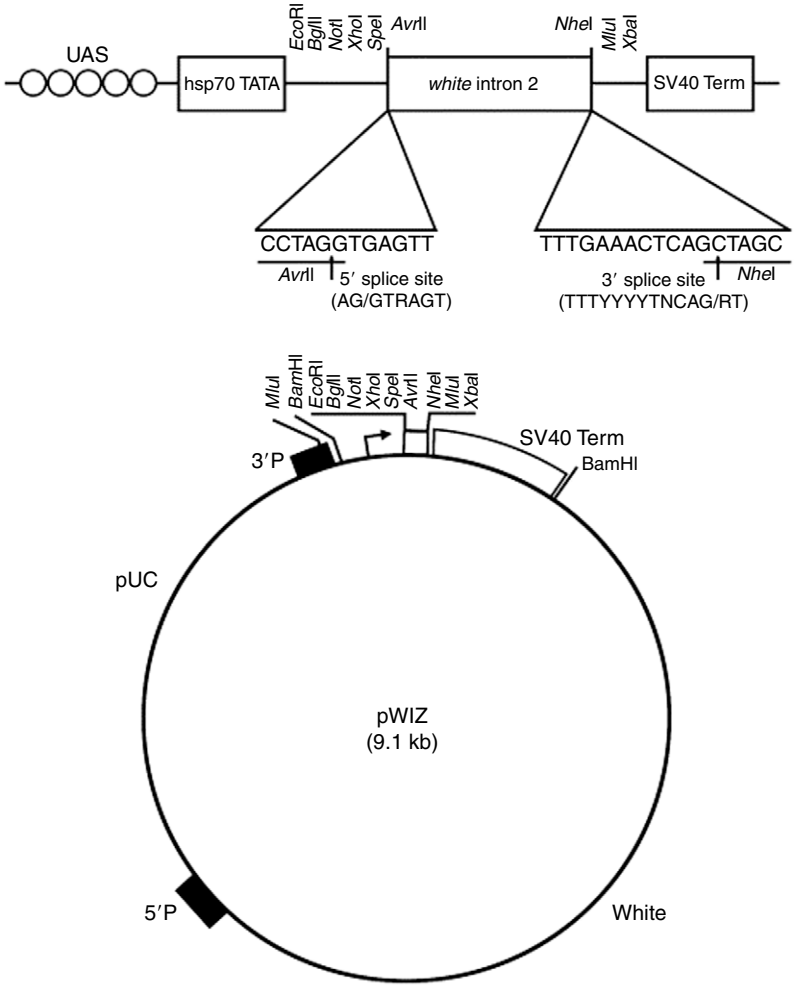

Fig. 2. Schematic of the pWiz vector. Reproduced with permission from Lee and Carthew (Lee and Carthew, 2003).

alleles of genes, without resorting to screening for new mutants. Although many means of expressing double-stranded RNA (dsRNA), the precursor of short interfering RNA (siRNA), have been developed, the favoured means at present is to use Carthew's pWiz vector (Lee and Carthew, 2003). This has a complete pUC plasmid with antibiotic resistance marker (so that it can be grown up in *Escherichia coli*), P-element ends (so that it can transform into *Drosophila* at high efficiency), a *white* minigene (to allow transgenic flies to be identified by their red eyes) and rare-cutter sites (downstream of UAS) that allow DNA fragments to be dropped into the vector in opposite orientations (Fig. 2). The cloning sites are on either side of an efficiently spliced *white* intron, believed to help in the generation of dsRNA. This makes the generation of flies transgenic for UAS–RNAi constructs a matter of a week's cloning, followed by a three-month period of intermittent fly husbandry.

The weakness of many model organisms is the difficulty or impossibility of targeted homologous recombination. Rather than just adding a transgene, this is the ability to replace a particular genomic sequence with another sequence; for example, to replace the open reading frame of a gene with green fluorescent protein (GFP), so inactivating the gene while putting the reporter under precisely the combination of promoters and enhancers that control the wild-type gene. In *Drosophila*, this is now possible, although not trivial. Golic's lab realized that the limitation was in producing a linear targeting construct capable of recombination with the *Drosophila* germ-line (Rong and Golic, 2000; Rong et al., 2002). The procedure they designed involves inserting the targeting construct stably into the germline by conventional transformation, then excising it at FRT sites with a transgenic FLP recombinase enzyme. The targeting construct is then able to synapse with its target, allowing recombination to take place with reasonable efficiency (Rong and Golic, 2000; Rong et al., 2002).

The jewel in the crown of *Drosophila* genetics, however, is its long history of community-generated resources. The largest stock centre, in Bloomington (http://flystocks.bio.indiana.edu/), holds over 20 000 mutant fly lines, available for a nominal charge by e-mail: requests are processed within a couple of weeks. Not only does this imply that there is a good chance that an informative allele for a particular gene may already exist but there is also the possibility of handling the direct descendents of flies identified by the *Drosophila* pioneers, Morgan, Bridges and Sturtevant. For example, we were recently able to document the recapitulation of the human renal disease xanthinuria type I in mutants at the *Drosophila* rosy locus, using the direct descendents of the *rosy*^{2} mutants first identified around 1916 (Wang et al., 2004). The utility of these mutants is increased by the ease with which they can be identified; the freely accessible Flybase website (http://www.flybase.net/) is exemplary in its ease of use.

Of course, similar results could be achieved in other organisms; it is the speed and cost that makes the *Drosophila* version of the technology so attractive.

Drosophila and other models

I suggest that there are certain minimum requirements on a model organism for it to be useful for integrative physiology. It must be possible to introduce transgenes, to produce mutations of specific genes and to intervene in a tissue-specific manner. It must also be possible to perform physiology on the organism and it is highly desirable that there should be a sequenced genome, as this makes many aspects of the work far easier.

In this context, it becomes clear that the trade-off between biomedical relevance and genetic power does not automatically lead to the mouse as uniformly best choice (Table 1).

Put simply, a transgenic mouse line takes several years, and perhaps in excess of $100 000 to realize, and then $10 000/year

Table 1. *Comparison of important properties of some model organisms*

Organism	Genome?	Generation time?	Cost/year?	Transgenics?	Targeted mutagenesis?	Tissue specific?	Available mutants?	Physiology?
Human	Yes	16–40 years	Very high	No!	No!	No	Yes	Very limited
Mouse	Yes	3 months	Very high	Yes	Yes	Yes	Many	Yes
Zebra fish	End 2008	3–4 months	Moderate	No	No	Yes	Yes	Some
Drosophila	Yes	1 week	Very low	Yes	Yes	Yes	Many	Some
Caenorhabditis	Yes	4 days	Very low	Yes	Yes	Yes	Many	Very little
Saccharomyces	Yes	90 min	Very low	Yes	Yes	Yes	Many	Not as we know it
Escherichia	Yes	20 min	Very low	Yes	Yes	Yes	Many	Not as we know it

to maintain thereafter; whereas a transgenic fruit-fly line can be made for $500 in three months and then costs $30/year to maintain. Clearly, an investigator can gain time and flexibility, while saving considerable amounts of money, if their questions can be addressed in a simpler model.

Neither should it be assumed that biomedical research is the only kind worth doing! Integrative physiology has perhaps as many as 30 million species with which to work; and the demands of both basic, curiosity-led research and perceived 'usefulness' do not have to be antagonistic. For example, many millions of lives are lost annually to parasitic diseases (notably, but not exclusively, malaria) that are carried by Diptera – phylogenetically close relatives of *Drosophila* (Butler, 2003).

More generally, is there thus an adequate phylogenetic spread of model species such that integrative biology can be evoked for a good proportion of problems in comparative physiology? So far, the match is not perfect, according to the NCBI's genome page (http://www.ncbi.nlm.nih.gov/Genomes/). Mammals are well-represented for genome projects, and mouse is the mainstay model organism. The position is similar for insects, with *Drosophila melanogaster* as the constituency champion (and sequencing underway or complete for a total of 50 species). For birds, the chicken genome is now released, but the transgenic resources are not yet as potent as for the first models. For this reason, much is made of a particularly easy-to-transform lymphoid cell line with high rates of recombination, DT40 (Buerstedde et al., 1990) – although this of course is hardly 'integrative' as we have defined it. Other species, such as the zebra finch, *Taeniopygia guttata*, are being sequenced. For fish, fugu (*Takifugu rubripes*), the related pufferfish *Tetraodon nigroviridis* and zebra fish (*Danio rerio*), and *Oryzias latipes* (Japanese medaka) sequences are nearly complete; but targeted mutagenesis has been announced several times, rather than been deployed as a routine tool, in fish. Elasmobranchs are represented by *Leucoraja erinacea* (little skate). The Reptilia are conspicuously unsequenced, and the Amphibia are represented by *Xenopus tropicalis* (western clawed frog).

Among the simpler animal phyla, the nematode worm *C. elegans* has remarkably potent genetic tools, especially for making transgenics; it is possible to microinject embryos with plasmids that then replicate as episomal 'rafts', so providing effectively stable transformants overnight. By generating worms transgenic for fluorescent reporters, mutagenising them and passing them through essentially a modified FACS (fluorescence activated cell sorting) machine, it has proved possible to screen 300 000 worms in a weekend – a throughput which is the envy of the fly community (Strange, 2003; Strange, 2007). As well as several other members of the genus *Caenorhabditis*, the nematode roundworm *Trichinella spiralis* (the cause of human trichinosis) is being sequenced.

Outside of these 'hotspots', however, things get patchier. The echinoderm species *Strongylocentrotus purpuratus* is being sequenced, and there is an expressed sequence tag (EST) sequencing project underway for the crab *Carcinus maenus*, and a genome project for the water flea *Daphnia pulex*, the tick *Ixodes scapularis*, the hemichordate acorn worm *Saccoglossus kowalevskii*, the freshwater planarian *Schmidtea mediterranea* and the pig tapeworm *Taenia solium*. For the molluscs, sequencing of the Atlantic surf clam (*Spisula solidissima*) and *Biomphalaria glabrata* (the freshwater snail host for schistosomiasis) is underway, and the sequence for *Aplysia californica* (California sea hare) is being assembled. In the even simpler cnidarians, assembled genomic sequence is available for two sea squirts (*Cionia* spp.), and a sea anemone (*Nematostella vectensis*) and the hydrozoan polyp *Hydra magnipapillata* are being sequenced. Perhaps most exotically, the tunicate *Oikopleura dioica* and the simplest known animal (and only known member of the Placozoa), *Trichoplax adhaerens*, are being sequenced. These data are summarized in Table 2.

Of course, genomic sequence is only one criterion for integrative biology, and it must be conceded that there are no publications describing transgenic technologies for the large majority of the organisms listed above. Nor can many of the species listed above be considered to be established physiological models. So, the list of models compatible with integrative physiology in Table 1 remains fairly definitive for the time being, until the genetics catches up with the genomics.

The future for model organisms

If one considered genomic sequence and transgenic technology to be sufficient for integrative biology, then the days of the models might be limited. For example, now that genomic sequence is available for the malaria mosquito *Anopheles gambiae* and the yellow fever mosquito *Aedes aegypti*, and germ line transformation with transposon-based vectors has shown to be feasible for both, has *Drosophila* had

Table 2. *The most advanced sequencing project for a range of animal phyla*

Phylum	Species	Genome status
Vertebrates		
Mammals	*Mus musculis*	Released
Birds	*Gallus gallus*	Released
Reptiles	None	
Amphibians	*Xenopus tropicalis*	Underway
Fishes	*Danio rerio*	Underway
Invertebrates		
Arthropods	*Drosophila melanogaster*	Released
Crustacea	*Daphnia pulex*	Underway
Arachnids	*Ixodes scapularis*	Underway
Molluscs	*Aplysia californica*	Assembly
Echinoderms	*Strongylocentrotus purpuratus*	Assembly
Worms	*Caenorhabditis elegans*	Released
Tunicates	*Cionia intestinalis*	Assembly

Based on a manual search of NCBI genomes (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome) on 5 December 2006.

its day? Interestingly, perhaps the opposite holds true; comparative genomics approaches make it much easier to drill through to a model such as *Drosophila* and to access the extant mutant stocks and long-term research database available for the model species. The Bloomington stock centre alone holds over 20 000 stocks; whereas the technical difficulty of keeping mosquitoes means that even an active lab could not hope to maintain more than perhaps six – and then only while grant funding continued. So the 'model organism package' needs to be seen as a whole.

RNAi – does it promote 'target' species to 'model' status?

If the utility of mutants is accepted as a key property of a model, can RNAi be used to provide mutants in non-model species? Since its discovery, RNAi has proved a powerful experimental tool, particularly in *C. elegans*, where it suffices to feed worms on *E. coli* harbouring a plasmid encoding an RNAi hairpin construct (Wang and Barr, 2005). In *Drosophila*, there are transformation vectors available that make generation of dsRNA particularly easy (Fig. 2), and there are genome-wide screens underway both in whole flies and cell lines.

However, problems remain; RNAi relies on the subversion of a mechanism probably intended to attack invading viruses, and a common side-effect of RNAi treatment is the non-specific shutdown of transcription. Care must be taken to control also for the knockdown of closely related transcripts (so-called 'off-target effects'). In addition, most RNAi alleles are hypomorphs, rather than nulls. This can be an advantage, as lethal mutations can be hard to study! However, we find that driving high levels of dsRNA inside cells with the GAL4/UAS system still only produces detectable knockdowns in about half of all cases (J.A.T.D., unpublished).

Given that RNAi is only partially effective when actually generated inside a cell, this might make one sceptical of the miraculous properties ascribed to RNAi in non-model organisms. dsRNA has been applied in the food, or injected into the haemocoel of larval and adult insects, and found to produce complete knockdowns that last for days, and in some cases have been heritable. These results imply that dsRNA is able to tunnel across basement membranes as well as plasma membranes, and sometimes even the blood–brain barrier. So, although RNAi (or morpholinos) will prove important in non-model species, it must be seen as a maturing technology that requires strict controls to be respected. For example, western blotting with a specific antibody to show protein knockdown, accompanied by counterstaining for a related protein that is not affected, should be seen as a minimum requirement.

Diversity and the insects

Model organisms are vital tools, but do they actually have any relevance for target species? More generally, do data in any given species have relevance to other species? This question is particularly acute for the insects, where there are perhaps as many as 30 million species in existence. Is it possible to estimate the extent to which *Drosophila* is a model insect, rather than a model fly? And is it possible to design a rational sampling strategy to cover this wide diversity with finite resources?

In the case of the *Drosophila* tubule, the evidence is promising. The importance of the apical V-ATPase is known from many species (Bertram et al., 1991; Dow et al., 1994; Garayoa et al., 1995; Maddrell and O'Donnell, 1992; Pietrantonio and Gill, 1995), and the indications that a basolateral, glibenclamide and barium-sensitive potassium channel is important are also well known (Beyenbach and Masia, 2002; Evans et al., 2005; Masia et al., 2000; Weltens et al., 1992; Wiehart et al., 2003a).

Neuroendocrine control shows great commonality, implying that neuropeptides have a common origin in insects. In all insects studied so far, cyclic AMP is diuretic and can be raised by either the corticotropin releasing factor-like DH_{44} or the calcitonin-like DH_{31} (Coast, 1998). Although the similarities with the vertebrate peptides are very faint, it could be argued that signalling through these peptides is conserved beyond insects.

The case can be made much more strongly, however, for the leucokinin family. Indeed, although the first leucokinins were characterized in an insect (Holman et al., 1984), the first gene for a leucokinin receptor, and its cognate peptide, were identified in the pond snail *Lymnaea stagnalis* in a single, thoroughly impressive paper (Cox et al., 1997). Similarly, leucokinin signalling is known to occur in the Acari (mites and ticks) (Holmes et al., 2000). In insects, leucokinins are uniformly myogenic and diuretic (Coast et al., 2002). *Drosophila* has proved useful in elucidating the mode of action of these neuropeptides; not only were the first insect leucokinin and leucokinin receptor genes identified in *Drosophila*, but the mode of action (through intracellular calcium) was established

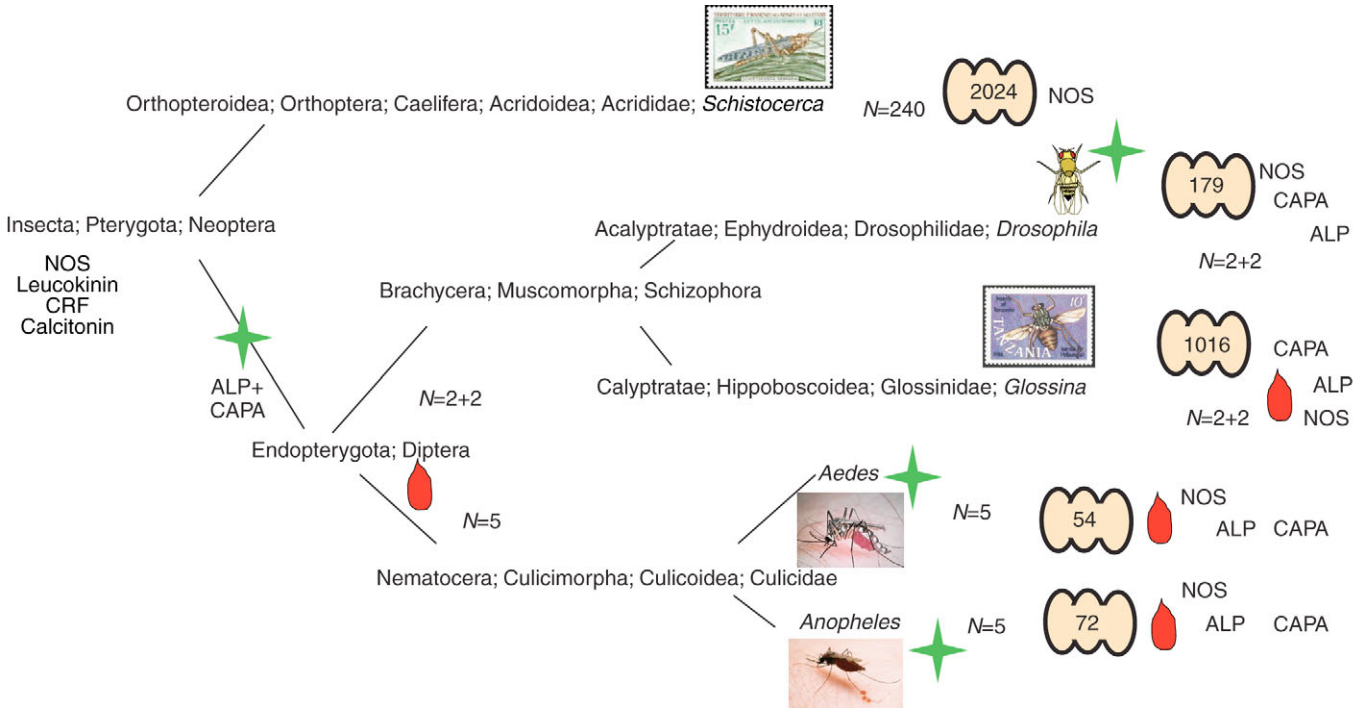

Fig. 3. Reconciling tubule physiology with phylogeny. Five species are represented; the muscomorph flies *Drosophila melanogaster* and the tsetse fly *Glossina morsitans*; the mosquitoes *Aedes aegypti* and *Anopheles gambiae*; and the more primitive orthopteran *Schistocerca gregaria*. Some recent comparative results from our lab are overlaid with classical data; the number of tubules per animal (*N*=xx), the approximate number of cells per tubule (brown shading), the presence of nitric oxide synthase in the tubule (NOS) (Pollock et al., 2004), the diuretic activity of Capa (CAPA) (Pollock et al., 2004), the existence of a defined alkaline phosphatase domain in the lower (proximal) tubule (ALP) (Cabrero et al., 2004), the presence of stellate cells (green stars) (Cabrero et al., 2004) and the haematophagous habit (red drops). The diagram shows that some properties can be considered to be common to insects whereas others seem to be associated with the Diptera. Conspicuously, tsetse flies, although closely related to *Drosophila*, lack stellate cells: parsimony suggests this is a secondary loss, perhaps associated with the degenerate lifestyle of these unusual flies.

with transgenic calcium reporter technology (Radford et al., 2002; Terhzaz et al., 1999). Leucokinin signalling is sufficiently well conserved that neuropeptides can be active across wide phylogenetic distances in the insects.

A further neuropeptide family highlights differences, as well as similarities. The Capa neuropeptides [the prototype was identified in a lepidopteran, *Manduca sexta* (Davies et al., 1994), and the first gene in *Drosophila* (Kean et al., 2002)] signal through intracellular calcium (Rosay et al., 1997). However, in *Drosophila* tubule, the cell type that receives the Capa signal is loaded with nitric oxide synthase (Davies et al., 1997), a calcium/calmodulin-sensitive enzyme. Capa peptides thus raise calcium, nitric oxide and ultimately cyclic GMP (cGMP) in the same cell (Davies et al., 1997). Both calcium and cGMP have diuretic effects in the cell; cGMP through its protein kinase (MacPherson et al., 2004), and calcium by activating mitochondria to increase the ATP supply to the apical V-ATPase (Terhzaz et al., 2006). However, although cGMP is diuretic in *Drosophila* and other Diptera, cGMP is antidiuretic in some other orders (Quinlan and O'Donnell, 1998; Wiehart et al., 2003b), and Capa can be either without effect or antidiuretic (Predel and Wegener, 2006).

So, overall, it looks as if data from *Drosophila* Malpighian tubules rolls out across all Diptera with very few caveats, and indeed the broad pattern of tubule function and control is recognizable across all insects (Fig. 3). So, the model organism/integrative physiology agenda seems to survive this relatively severe test (up to 30 million species in over 20 orders separated by 150 million years of rapid divergent evolution).

Conclusion

This review has shown that there is considerable promise in the judicious use of model organisms to establish general principles of function. It should be clear that the ease and power with which some of these organisms can be manipulated does elevate them to 'Krogh status' for several interesting biological questions. The *Drosophila* Malpighian tubule is in no way an exceptional tissue in an exotic organism; it has simply been studied in some detail. Clearly, it is vital for functional genomics that more experimental biologists take an active interest in specific tissues in these valuable organisms. This will lay the ground not just for a better understanding of what genes do but for a systems approach to the function of the whole organism. Surely this should be the ultimate goal of integrative physiology?

Our group is generously supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC). I am most grateful to Dr Shireen Davies for her critical reading of the manuscript.

References

Becker, H. J. (1975). X-ray- and TEM-induced mitotic recombination in *Drosophila melanogaster*: unequal and sister-strand recombination. *Mol. Gen. Genet.* **138**, 11-24.

Bertram, G., Shleithoff, L., Zimmermann, P. and Wessing, A. (1991). Bafilomycin$_{A1}$ is a potent inhibitor of urine formation by Malpighian tubules of *Drosophila hydei* – is a vacuolar-type ATPase involved in ion and fluid secretion? *J. Insect Physiol.* **37**, 201-209.

Beyenbach, K. W. and Masia, R. (2002). Membrane conductances of principal cells in Malpighian tubules of *Aedes aegypti*. *J. Insect Physiol.* **48**, 375-386.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.

Brown, S. D. M. and Peters, J. (1996). Combining mutagenesis and genomics in the mouse – closing the phenotype gap. *Trends Genet.* **12**, 433-435.

Buerstedde, J. M., Reynaud, C. A., Humphries, E. H., Olson, W., Ewert, D. L. and Weill, J. C. (1990). Light chain gene conversion continues at high rate in an ALV-induced cell line. *EMBO J.* **9**, 921-927.

Bullard, D. C. (2001). Mind the phenotype gap. *Trends Mol. Med.* **7**, 537-538.

Butler, D. (2003). Gates steps up war on malaria with donation of 168 million dollars. *Nature* **425**, 331.

Cabrero, P., Radford, J. C., Broderick, K. E., Costes, L., Veenstra, J. A., Spana, E. P., Davies, S. A. and Dow, J. A. T. (2002). The *Dh* gene of *Drosophila melanogaster* encodes a diuretic peptide that acts through cyclic AMP. *J. Exp. Biol.* **205**, 3799-3807.

Cabrero, P., Pollock, V. P., Davies, S. A. and Dow, J. A. T. (2004). A conserved domain of alkaline phosphatase expression in the Malpighian tubules of dipteran insects. *J. Exp. Biol.* **207**, 3299-3305.

Catsch, A. (1944). Eine erbliche Storung des Bewegungsmechanismus bei *Drosophila melanogaster*. *Z. Indukt. Abstamm. Verebungsl.* **82**, 64-66.

Coast, G. M. (1998). Insect diuretic peptides: structures, evolution and actions. *Am. Zool.* **38**, 442-449.

Coast, G. M., Webster, S. G., Schegg, K. M., Tobe, S. S. and Schooley, D. A. (2001). The *Drosophila melanogaster* homologue of an insect calcitonin-like diuretic peptide stimulates V-ATPase activity in fruit fly Malpighian tubules. *J. Exp. Biol.* **204**, 1795-1804.

Coast, G. M., Orchard, I., Phillips, J. E. and Schooley, D. A. (2002). Insect diuretic and antidiuretic hormones. *Adv. Insect Physiol.* **29**, 279-409.

Cox, K. J., Tensen, C. P., Van der Schors, R. C., Li, K. W., van Heerikhuizen, H., Vreugdenhil, E., Geraerts, W. P. and Burke, J. F. (1997). Cloning, characterization, and expression of a G-protein-coupled receptor from *Lymnaea stagnalis* and identification of a leucokinin-like peptide, PSFHSWSamide, as its endogenous ligand. *J. Neurosci.* **17**, 1197-1205.

Davies, S. A., Dow, J. A. T., Huesmann, G. R., Maddrell, S. H. P., O'Donnell, M. J. and Tublitz, N. J. (1994). CAP$_{2b}$, a cardioactive peptide of lepidopteran insects, can also stimulate fluid production by Malpighian tubules of *Drosophila melanogaster* through a mechanism involving cGMP. *J. Physiol.* **482**, 26P.

Davies, S. A., Stewart, E. J., Huesmann, G. R., Skaer, N. J. V., Maddrell, S. H. P., Tublitz, N. J. and Dow, J. A. T. (1997). Neuropeptide stimulation of the nitric oxide signaling pathway in *Drosophila melanogaster* Malpighian tubules. *Am. J. Physiol.* **42**, R823-R827.

Dow, J. A. (2003). The *Drosophila* phenotype gap – and how to close it. *Brief. Funct. Genomic. Proteomic.* **2**, 121-127.

Dow, J. A. T. (2007). Model organisms and molecular genetics for endocrinology. *Gen. Comp. Endocrinol.* doi:10.1016/j.ygcen.2007.01.023.

Dow, J. A. T. and Davies, S. A. (2001). The *Drosophila melanogaster* Malpighian tubule. *Adv. Insect Physiol.* **28**, 1-83.

Dow, J. A. T. and Davies, S. A. (2003). Integrative physiology and functional genomics of epithelial function in a genetic model organism. *Physiol. Rev.* **83**, 687-729.

Dow, J. A. T. and Davies, S. A. (2006). The Malpighian tubule: rapid insights from post-genomic biology. *J. Insect Physiol.* **52**, 365-378.

Dow, J. A. T., Maddrell, S. H. P., Görtz, A., Skaer, N. V., Brogan, S. and Kaiser, K. (1994). The Malpighian tubules of *Drosophila melanogaster*: a novel phenotype for studies of fluid secretion and its control. *J. Exp. Biol.* **197**, 421-428.

Dudai, Y. and Zvi, S. (1985). Multiple defects in the activity of adenylate cyclase from the *Drosophila* memory mutant rutabaga. *J. Neurochem.* **45**, 355-364.

Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G. and Benzer, S. (1976). dunce, a mutant of *Drosophila* deficient in learning. *Proc. Natl. Acad. Sci. USA* **73**, 1684-1688.

Evans, J. M., Allan, A. K., Davies, S. A. and Dow, J. A. T. (2005). Sulphonylurea sensitivity and enriched expression implicate inward rectifier K^+ channels in *Drosophila melanogaster* renal function. *J. Exp. Biol.* **208**, 3771-3783.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.

Garayoa, M., Villaro, A. C., Klein, U., Zimmermann, B., Montuenga, L. M. and Sesma, P. (1995). Immunocytochemical localization of a vacuolar-type ATPase in Malpighian tubules of the ant *Formica polyctena*. *Cell Tissue Res.* **282**, 343-350.

Golic, K. G. and Lindquist, S. (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59**, 499-509.

Hassan, B. A., Bermingham, N. A., He, Y., Sun, Y., Jan, Y. N., Zoghbi, H. Y. and Bellen, H. J. (2000). atonal regulates neurite arborization but does not act as a proneural gene in the *Drosophila* brain. *Neuron* **25**, 549-561.

Holman, G. M., Cook, B. J. and Wagner, R. M. (1984). Isolation and partial characterization of five myotropic peptides present in head extracts of the cockroach, *Leucophaea maderae*. *Comp. Biochem. Physiol.* **77C**, 1-5.

Holmes, S., He, H., Chen, A., Ivie, G. and Pietrantonio, P. (2000). Cloning and transcriptional expression of a leucokinin-like peptide receptor from the southern cattle tick, *Boophilus microplus* (Acari: Ixodidae). *Insect Mol. Biol.* **9**, 457-465.

Johnson, E. C., Shafer, O. T., Trigg, J. S., Park, J., Schooley, D. A., Dow, J. A. T. and Taghert, P. H. (2005). A novel diuretic hormone receptor in *Drosophila*: evidence for conservation of CGRP signaling. *J. Exp. Biol.* **208**, 1239-1246.

Kaiser, K. (1993). Transgenic *Drosophila* – second generation enhancer traps. *Curr. Biol.* **3**, 560-562.

Kamb, A., Iverson, L. E. and Tanouye, M. A. (1987). Molecular characterization of *Shaker*, a *Drosophila* gene that encodes a potassium channel. *Cell* **50**, 405-413.

Kandel, E. R. and Schwartz, J. H. (1982). Molecular biology of learning: modulation of transmitter release. *Science* **218**, 433-443.

Kean, L., Pollock, V. P., Broderick, K. E., Davies, S. A., Veenstra, J. and Dow, J. A. T. (2002). Two new members of the CAP$_{2b}$ family of diuretic peptides are encoded by the gene *capability* in *Drosophila melanogaster*. *Am. J. Physiol.* **282**, R1297-R1307.

Konopka, R. J. and Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **68**, 2112-2116.

Krogh, A. (1929). The progress of physiology. *Am. J. Physiol.* **90**, 243-251.

Lee, Y. S. and Carthew, R. W. (2003). Making a better RNAi vector for *Drosophila*: use of intron spacers. *Methods* **30**, 322-329.

MacPherson, M. R., Lohmann, S. M. and Davies, S. A. (2004). Analysis of *Drosophila* cGMP-dependent protein kinases and assessment of their in vivo roles by targeted expression in a renal transporting epithelium. *J. Biol. Chem.* **279**, 40026-40034.

Maddrell, S. H. P. and O'Donnell, M. J. (1992). Insect Malpighian tubules: V-ATPase action in ion and fluid transport. *J. Exp. Biol.* **172**, 417-429.

Masia, R., Aneshansley, D., Nagel, W., Nachman, R. J. and Beyenbach, K. W. (2000). Voltage clamping single cells in intact Malpighian tubules of mosquitoes. *Am. J. Physiol. Renal Physiol.* **279**, F747-F754.

McGuire, S. E., Mao, Z. and Davis, R. L. (2004). Spatiotemporal gene expression targeting with the TARGET and gene-switch systems in Drosophila. *Sci. STKE* **2004**, pl6.

Pietrantonio, P. V. and Gill, S. S. (1995). Immunolocalization of the 17-Kda vacuolar H^+-atpase subunit-C in *Heliothis virescens* midgut and Malpighian tubules with an antipeptide antibody. *J. Exp. Biol.* **198**, 2609-2618.

Pollock, V. P., Mcgettigan, J., Cabrero, P., Maudlin, I. M., Dow, J. A. T. and Davies, S. A. (2004). Conservation of capa peptide-induced nitric oxide signalling in Diptera. *J. Exp. Biol.* **207**, 4135-4145.

Predel, R. and Wegener, C. (2006). Biology of the CAPA peptides in insects. *Cell. Mol. Life Sci.* **63**, 2477-2490.

Quinlan, M. C. and O'Donnell, M. J. (1998). Anti-diuresis in the blood-feeding insect *Rhodnius prolixus* Stal: antagonistic actions of cAMP and cGMP and the role of organic acid transport. *J. Insect Physiol.* **44**, 561-568.

Radford, J. C., Davies, S. A. and Dow, J. A. T. (2002). Systematic GPCR analysis in *Drosophila melanogaster* identifies a leucokinin receptor with novel roles. *J. Biol. Chem.* **277**, 38810-38817.

Radford, J. C., Terhzaz, S., Cabrero, P., Davies, S. A. and Dow, J. A. T. (2004). Functional characterisation of the *Anopheles* leucokinins and their cognate G-protein coupled receptor. *J. Exp. Biol.* **207**, 4573-4586.

Rong, Y. S. and Golic, K. G. (2000). Gene targeting by homologous recombination in *Drosophila*. *Science* **288**, 2013-2018.

Rong, Y. S., Titen, S. W., Xie, H. B., Golic, M. M., Bastiani, M., Bandyopadhyay, P., Olivera, B. M., Brodsky, M., Rubin, G. M. and Golic, K. G. (2002). Targeted mutagenesis by homologous recombination in *D. melanogaster*. *Genes Dev.* **16**, 1568-1581.

Rosay, P., Davies, S. A., Yu, Y., Sozen, M. A., Kaiser, K. and Dow, J. A. T. (1997). Cell-type specific calcium signalling in a *Drosophila* epithelium. *J. Cell Sci.* **110**, 1683-1692.

Salkoff, L., Baker, K., Butler, A., Covarrubius, M., Pak, M. D. and Wei, A. (1992). An essential "set" of K^+ channels conserved in flies, mice and humans. *Trends Neurosci.* **15**, 161-166.

Sentry, J. W., Goodwin, S. F., Milligan, C. D., Duncanson, A., Yang, M. and Kaiser, K. (1994). Reverse genetics of *Drosophila* brain structure and function. *Prog. Neurobiol.* **42**, 299-308.

Sözen, M. A., Armstrong, J. D., Yang, M. Y., Kaiser, K. and Dow, J. A. T. (1997). Functional domains are specified to single-cell resolution in a *Drosophila* epithelium. *Proc. Natl. Acad. Sci. USA* **94**, 5207-5212.

Strange, K. (2003). From genes to integrative physiology: ion channel and transporter biology in *Caenorhabditis elegans*. *Physiol. Rev.* **83**, 377-415.

Strange, K. (2007). Revisiting the Krogh Principle in the post-genome era: *Caenorhabditis elegans* as a model system for integrative physiology research. *J. Exp. Biol.* **210**, 1622-1631.

Suster, M. L., Seugnet, L., Bate, M. and Sokolowski, M. B. (2004). Refining GAL4-driven transgene expression in *Drosophila* with a GAL80 enhancer-trap. *Genesis* **39**, 240-245.

Terhzaz, S., O'Connell, F. C., Pollock, V. P., Kean, L., Davies, S. A., Veenstra, J. A. and Dow, J. A. T. (1999). Isolation and characterization of a leucokinin-like peptide of *Drosophila melanogaster*. *J. Exp. Biol.* **202**, 3667-3676.

Terhzaz, S., Southall, T. D., Lilley, K. S., Kean, L., Allan, A. K., Davies, S. A. and Dow, J. A. T. (2006). Differential gel electrophoresis and transgenic mitochondrial calcium reporters demonstrate spatiotemporal filtering in calcium control of mitochondria. *J. Biol. Chem.* **281**, 18849-18858.

Wang, J. and Barr, M. M. (2005). RNA interference in *Caenorhabditis elegans*. *Meth. Enzymol.* **392**, 36-55.

Wang, J., Kean, L., Yang, J., Allan, A. K., Davies, S. A., Herzyk, P. and Dow, J. A. T. (2004). Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol.* **5**, R69.

Weltens, R., Leyssens, A., Zhang, A. L., Lohhrmann, E., Steels, P. and van Kerkhove, E. (1992). Unmasking of the apical electrogenic H pump in isolated Malpighian tubules (*Formica polyctena*) by the use of barium. *Cell. Physiol. Biochem.* **2**, 101-116.

Wiehart, U. I., Klein, G., Steels, P., Nicolson, S. W. and Van Kerkhove, E. (2003a). K^+ transport in Malpighian tubules of *Tenebrio molitor* L.: is a K_{ATP} channel involved? *J. Exp. Biol.* **206**, 959-965.

Wiehart, U. I., Nicolson, S. W. and Van Kerkhove, E. (2003b). The effects of endogenous diuretic and antidiuretic peptides and their second messengers in the Malpighian tubules of *Tenebrio molitor*: an electrophysiological study. *J. Insect Physiol.* **49**, 955-965.

Xu, T. and Harrison, S. D. (1994). Mosaic analysis using FLP recombinase. *Methods Cell Biol.* **44**, 655-681.

The Journal of Experimental Biology 210, 1492-1496
Published by The Company of Biologists 2007
doi:10.1242/jeb.02783

Glossary of terms

This section is designed to help readers adapt to the complex terminology associated with contemporary molecular genetics, genomics and systems biology. Fuller descriptions of these terms are available at http://www.wikipedia.org/

Ab initio prediction	methods used to predict the potential genes encoded in the genome, which are trained on datasets made of known genes, and used computationally to predict coding regions out of genome without the aid of cDNA sequence. Although their performance is improving, these algorithms perform very poorly on non-protein coding genes.
Annotation	as applied to proteins, DNA sequences or genes. The storage of data describing these entities (protein/gene identities, DNA motifs, gene ontology categorisation, etc.) within a biological database. Active projects include FlyBase and WormBase. See Gene ontology.
Assembly	the process of aligning sequenced fragments of DNA into their correct positions within the chromosome or transcript.
cDNA	complementary DNA. This is DNA synthesised from a mature mRNA template by the enzyme reverse transcriptase. cDNA is frequently used as an early part of gene cloning procedures, since it is more robust and less subject to degradation than the mRNA itself.
ChIP	chromatin immunoprecipitation assay used to determine which segments of genomic DNA are bound to chromatin proteins, mainly including transcription factors.
Chip	see Microarray.
ChIP-on-chip	use of a DNA microarray to analyse the DNA generated from chromatin immunoprecipitation experiments (see ChIP).
cis-acting	a molecule is described as *cis*-acting when it affects other genes that are physically adjacent, on the same chromosome, or are genetically linked or in close proximity (for mRNA expression, typically a promoter).
Collision-induced dissociation	a mechanism by which molecules (e.g. proteins) are fragmented to form molecular ions in the gas phase. These fragments are then analysed within a mass spectrometer to provide mass determination.
Connectivity	a term from graph theory, which indicates the number of connections between nodes or vertices in a network. Greater connectedness between nodes is generally used as a measure of robustness of a network.
CpG islands	regions that show high density of ‘C followed by G’ dinucleotides and are generally associated with promoter elements; in particular, stretches of DNA of at least 200 bp with a C–G content of 50% and an observed CpG/expected CpG in excess of 0.6. The cytosine residues can be methylated, generally to repress transcription, while demethylated CpGs are a hallmark of transcription. CpG dinucleotides are under-represented outside regulatory regions, such as promoters, because methylated C mutates into T by deamination.
Edge	as in networks. Connects two nodes (or vertices) within a system. These concepts arise from graph theory.
Enhancer	a short segment of genomic DNA that may be located remotely and that, on binding particular proteins (*trans*-acting factors), increases the rate of transcription of a specific gene or gene cluster.
Epistasis	a phenomenon when the properties of one gene are modified by one or more genes at other loci. Otherwise known as a genetic interaction, but epistasis refers to the statistical properties of the phenomenon.

eQTL	the combination of conventional QTL analysis with gene expression profiling, typically using microarrays. eQTLs describe regulatory elements controlling the expression of genes involved in specific traits.
EST	expressed sequence tag. A short DNA sequence determined for a cloned cDNA representing portions of an expressed gene. The sequence is generally several hundred base pairs from one or both ends of the cloned insert.
Exaptation	a biological adaptation where the current function is not that which was originally evolved. Thus, the defining (derived) function might replace or persist with the earlier, evolved adaptation.
Exon	any region of DNA that is transcribed to the final (spliced) mRNA molecule. Exons interleave with segments of non-coding DNA (introns) that are removed (spliced out) during processing after transcription.
Gene forests	genomic regions for which RNA transcripts, produced from either DNA strand, have been identified without gaps (non-transcribed genomic regions). Conversely, regions in which no transcripts have ever been detected are called ‘gene deserts’.
Gene interaction network	a network of functional interactions between genes. Functional interactions can be inferred from many different data types, including protein–protein interactions, genetic interactions, co-expression relationships, the co-inheritance of genes across genomes and the arrangement of genes in bacterial genomes. The interactions can be represented using network diagrams, with lines connecting the interacting elements, and can be modelled using differential equations.
Gene ontology (GO)	an ontology is a controlled vocabulary of terms that have logical relationships with each other and that are amenable to computerised manipulation. The Gene Ontology project has devised terms in three domains: biological process, molecular function and cell compartment. Each gene or DNA sequence can be associated with these annotation terms from each domain, and this enables analysis of microarray data on groups of genes based on descriptive terms so provided. See http://www.geneontology.org
Gene set enrichment analysis	a computational method that determines whether a defined set of genes, usually based on their common involvement in a biological process, shows statistically significant differences in transcript expression between two biological states.
Gene silencing	the switching-off of a gene by an epigenetic mechanism at the transcriptional or post-transcriptional levels. Includes the mechanism of RNAi.
Genetic interaction (network)	a genetic interaction between two genes occurs when the phenotypic consequences of a mutation in one gene are modified by the mutational status at a second locus. Genetic interactions can be aggravating (enhancing) or alleviating (suppressing). To date, most high-throughput studies have focussed on systematically identifying synthetic lethal or sick (aggravating) interactions, which can then be visualised as a network of functional interactions (edges) between genes (nodes).
Genome	a portmanteau of gene and chromosome, the entire hereditary information for an organism that is embedded in the DNA (or, for some viruses, in RNA). Includes protein-coding and non-coding sequences.
Heritability	phenotypic variation within a population is attributable to the genetic variation between individuals and to environmental factors. Heritability is the proportion due to genetic variation usually expressed as a percentage.
Heterologous hybridization	the use of a cDNA or oligonucleotide microarray of probes designed for one species with target cRNA/cDNAs from a different species.
Homeotic	the transformation of one body part to another due to mutation of specific developmentally related genes, notably the *Hox* genes in animals and *MADS-box* genes in plants.
Hub	as in networks. A node with high connectivity, and thus which interacts with many other nodes in the network. A hub protein interacts with many other proteins in a cell.

Hybridisation	the process of joining (annealing) two complementary single-stranded DNAs into a single double-stranded molecule. In microarray analysis, the target RNA/DNA from the subject under investigation is denatured and hybridised to probes that are immobilised on a solid phase (i.e. glass microscope slide).
Hypomorph	in genetics, a loss-of-function mutation in a gene, but which shows only a partial reduction in the activity it influences rather than a complete loss (cf. hypermorph, antimorph, neomorph, etc).
Imprinting	a phenomenon where two inherited copies of a gene are regulated in opposite ways, one being expressed and the other being repressed.
Indel	insertion and deletion of DNA, referring to two types of genetic mutation. To be distinguished from a 'point mutation', which refers to the substitution of a single base.
Interactome	a more or less comprehensive set of interactions between elements within cells. Usually applied to genes or proteins as defined by transcriptomic, proteomic or protein–protein interaction data.
Intron	see Exon.
KEGG	The Kyoto Encyclopedia of Genes and Genomes is a database of metabolic and other pathways collected from a variety of organisms. See http://www.genome.jp/kegg
Metabolomics	the systematic qualitative and quantitative analysis of small chemical metabolite profiles. The metabolome represents the collection of metabolites within a biological sample.
Metagenomics	the application of genomic techniques to characterise complex communities of microbial organisms obtained directly from environmental samples. Typically, genomic tags are sequence characterised as markers of each species to inform on the range and abundance of species in the community.
Microarray	an arrayed set of probes for detecting molecularly specific analytes or targets. Typically, the probes are composed of DNA segments that are immobilised onto the solid surface, each of which can hybridise with a specific DNA present in the target preparation. DNA microarrays are used for profiling of gene transcripts.
Model species	a species used to study particular biological phenomena, the outcome offering insights into the workings of other species. Usually, the selection is based on experimental tractability, particularly ease of genetic manipulation. For the geneticist, it is an organism with inbred lines where sibs will be >98% identical (i.e. *Drosophila*, *Caenorhabditis elegans* and mice). For genomic science, it refers to a species for which the genomic DNA has been sequenced.
miRNA	a category of novel, very short, non-coding RNAs, generated by the cleavage of larger precursors (pri-miRNA). These short RNAs are included in the RNA-induced silencing complex (RISC) and pair to the 3′ ends of target RNA, blocking its translation into proteins (in animals) or promoting RNA cleavage and degradation (in plants).
mRNA	a protein-coding mRNA containing a protein-coding region (CDS), preceded by a 5′ and followed by a 3′ untranslated region (5′ UTR and 3′ UTR). The UTRs contain regulatory elements. A full-length cDNA contains the complete sequence of the original mRNA, including both UTRs. However, it is often difficult to assign the starting–termination positions for protein synthesis unambiguously. A cDNA containing the entire CDS is often considered acceptable for bioinformatic and experimental studies requiring full-length cDNAs.
ncRNA	non-coding RNA is any RNA molecule with no obvious protein-coding potential for at least 80 or 100 amino acids, as determined by scanning full-length cDNA sequences. It includes ribosomal (rRNA) and transfer RNAs (tRNA) and is now known to include various sub-classes of RNA, including snoRNA, siRNA and piRNA. Just like the coding mRNAs, a large proportion of ncRNAs are transcribed by RNA polymerase II and are large transcripts. A description of the many forms of ncRNA can be found at http://en.wikipedia.org/wiki/Non-coding_RNA.

Node	as in networks. Objects linked by edges to create a network.
PCR	polymerase chain reaction. A molecular biology technique for replicating DNA *in vitro*. The DNA is thus amplified, sometimes from very small amounts. PCR can be adapted to perform a wide variety of genetic manipulations.
piRNA	Piwi-interacting RNA. A class of RNA molecules (29–30 nt long) that complex with Piwi proteins (a class of the Argonaute family of proteins) and are involved in transcriptional gene silencing.
PMF	peptide mass fingerprinting. An analytical technique for protein identification in which a protein is fragmented using proteases. The resulting peptides are analysed by mass spectrometry and these masses compared against a database of predicted or measured masses to generate a protein identity.
Polyadenylation	the covalent addition of multiple A bases to the 3′ tail of an mRNA molecule. This occurs during the processing of transcripts to form the mature, spliced molecule and is important for regulation of turnover, trafficking and translation.
Post-source decay	in mass spectrometry. The fragmentation of precursor molecular ions as they accelerate away from the ionisation source of the mass spectrometer. All precursor ions leaving the ion source have approximately the same kinetic energy, but fragmentation results in smaller product ions that can be distinguished from precursor ions using a ‘reflectron’ by virtue of their lower kinetic energies.
Post-translational modification	the chemical modification of a protein after synthesis through translation. Some modifications, notably phosphorylation, affect the properties of the protein, offering a means of regulating function.
Principal component analysis (PCA)	a technique for simplifying complex, multi-dimensional datasets to a reduced number of dimensions, the principal components. This procedure retains those characteristics of the data that relate to its variance.
Promoter	a regulatory DNA sequence, generally lying upstream of an expressed gene, which in concert with other often distant regulatory elements directs the transcription of a given gene.
Proteome	the entire protein complement of an organism, tissue or cell culture at a given time.
Quantitative trait	inheritance of a phenotypic property or characteristic that varies continuously between extreme states and can be attributed to interactions between multiple genes and their environment.
qPCR	quantitative real-time PCR, sometimes called real-time PCR. A more quantitative form of RT-PCR in which the quantity of amplified product is estimated after each round of amplification.
QTL	quantitative trait loci. A region of DNA that contains those genes contributing to the trait under study.
RISC	RNA-induced silencing complex. A protein complex that mediates the double-stranded RNA-induced destruction of homologous mRNA.
RNAi	RNA interference or RNA-mediated interference. The process by which double-stranded RNA triggers the destruction of homologous mRNA in eukaryotic cells by the RISC.
RT-PCR	reverse transcription–polymerase chain reaction. A technique for amplifying a defined piece of RNA that has been converted to its complementary DNA form by the enzyme reverse transcriptase. See qPCR.
siRNA	small interfering RNA, or silencing RNA. A class of short (20–25 nt), double-stranded RNA molecules. It is involved in the RNA interference pathway, which alters RNA stability and thus affects RNA concentration and thereby suppresses the normal expression of specific genes. Widely used in biomedical research to ablate specific genes.

snoRNA	small nucleolar RNA. A sub-class of RNA molecules involved in guiding chemical modification of ribosomal RNA and other RNA genes as part of the regulation of gene expression.
SNP	single nucleotide polymorphism. A single base-pair mutation at a specific locus, usually consisting of two alleles. Because SNPs are conserved over evolution, they are frequently used in QTL analysis and in association studies in place of microsatellites, and in genetic fingerprinting analyses.
SSH	suppressive subtractive hybridisation. A powerful protocol for enriching cDNA libraries for genes that differ in representation between two or more conditions. It combines normalisation and subtraction in a single procedure and allows the detection of low-abundance, differentially expressed transcripts, such as those involved in signalling and signal transduction.
Structural RNAs	a class of non-coding RNA, long known to have a structural role (for instance, the ribosomal RNAs), transcribed by RNA polymerase I or III.
Systems biology	treatment of biological entities as systems composed of defined elements interacting in defined ways to enable the observed function and behaviour of that system. The properties of the systems are embedded in a quantitative model that guides further tests of systems behaviour.
TATA-boxes	sequences in promoter regions constituted by TATAAA, or similar variants, which were considered the hallmark of Promoters. Recent data show that they are present only in the minority of promoters, where they direct transcription at a single well-defined location some 30 bp downstream of this element.
trans-acting	a factor or gene that acts on another unlinked gene, a gene on a separate chromosome or genetically unlinked usually through some diffusible protein product (for mRNA expression, typically a transcription factor).
Transcript	an RNA product produced by the action of RNA polymerase reading the sequence of bases in the genomic DNA. Originally limited to protein-coding sequences with flanking UTRs but now known to include large numbers of products that do not code for a protein product.
Transcriptome	the full set of mRNA molecules (transcripts) produced by the system under observation. Whilst the genome is fixed for a given organism, the transcriptome varies with context (i.e. tissue source, ontogeny, external conditions or experimental treatment).
Transgene	a gene or genetic material that has been transferred between species or between organisms using one of several genetic engineering techniques.
Transinduction	generation of transcripts from intergenic regions. At least some such products do not relate to a definable promoter or transcriptional start site.
Transposon	sequences of DNA able to move to new positions within the genome of a single cell. This event might cause mutation at the site of insertion. Also called ‘mobile genetic elements’ or ‘jumping genes’.
Transvection	an epigenetic phenomenon arising from the interaction between one allele and the corresponding allele on the homologous chromosome, leading to gene regulation.
TUs	transcriptional units. Used to group all of the overlapping RNA transcripts that are transcribed from the same genomic strand and share exonic sequences.
UTR	untranslated region. Regions of the mRNA that lie at either the 3′ or 5′ flanking ends of the molecule (i.e. 3′ UTR and 5′ UTR). They bracket the protein-coding region and contain signals and binding sites that are important for the regulation of both protein translation and RNA degradation.