PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

From genes to individuals: developmental genes and the generation of the phenotype

Diethard Tautz and Karl J. Schmid

Phil. Trans. R. Soc. Lond. B 1998 **353**, 231-240
doi: 10.1098/rstb.1998.0205

The Royal Society

From genes to individuals: developmental genes and the generation of the phenotype

Diethard Tautz* **and Karl J. Schmid†**

Zoologisches Institut der Universität München, Luisenstr. 14, 80333 München, Germany
(tautz@zi.biologie.uni-muenchen.de)

The success of the genetic approach to developmental biology has provided us with a suite of genes that are involved in the regulation of ontogenetic pathways. It is therefore time to ask whether and how such genes might be involved in the generation of adaptive phenotypes. Unfortunately, the current results do not provide a clear answer. Most of the genes that have been studied by developmental biologists affect early embryonic traits with significant effects on the whole organism. These genes are often highly conserved which allows us to do comparative studies even across phyla. However, whether the same genes are also involved in short-term ecological adaptations remains unclear. The suggestion that early acting ontogenetic genes may also affect late phenotypes comes from the genetic analysis of quantitative traits like bristle numbers in *Drosophila*. A rough mapping of the major loci affecting these traits shows that these loci might correspond to well known early acting genes. On the other hand, there are also many minor effect loci that are as yet uncharacterized. We suggest that these minor loci might correspond to a different class of genes. In comparative studies of randomly drawn cDNAs from *Drosophila* we find that there is a large group of genes that evolve fast and that are significantly under-represented in normal genetic screens. We speculate that these genes might provide a large, as yet poorly understood, reservoir of genes that might be involved in the evolution of quantitative traits and short-term adaptations.

Keywords: adaptive traits; evolution; developmental genetics; animal bauplan; fast-evolving genes

1. INTRODUCTION

The assumption that the understanding of ontogeny should also be the basis for understanding phylogeny can be traced back to Darwin's times and has become common place today. However, developmental and evolutionary disciplines have followed very different routes during the past decades and have generated their own paradigms. Often, these do not provide much room for overlap. This situation is currently changing. The progress in molecular genetic methods is starting to unite the two fields. We are beginning to understand how an organism is built at the molecular level and many of the genes involved in these processes appear to be highly conserved in evolution. This provides an immediate access to comparative studies in diverse organisms and is starting to create a new discipline: molecular comparative embryology. The early results from this discipline do already shed some light on the course of the main cladogenic events. However, as most of the developmental genetic approaches were not designed to solve evolutionary questions, their outcomes provide only a patchy picture of the possible routes of evolution of the organisms. The genes that are most heavily studied are those involved in early embryonic decisions, rather than those required for the differentiation of the adult morphology. Evidently, evolutionary biologists would be particularly interested in the latter class of genes, as these are the ones that should be most relevant for ecological adaptations and speciation. Characterizing such genes will therefore be an important task of the future for biologists working at the interface of development and evolution.

2. DEVELOPMENTAL GENETICS

One of the main breakthroughs in the field of developmental biology was the concept of using mutants with specific developmental defects to analyse ontogenetic processes (Lewis 1978; Nüsslein-Volhard & Wieschaus 1980). The organism of choice for such a genetic approach was *Drosophila*, as this allowed the most sophisticated types of genetic experiments. The developmental biology of *Drosophila*, on the other hand, was only poorly understood at that time, as the small size of the embryo as well as its specialized development presented large obstacles for experimental manipulation. In fact, some classic experimental manipulations on the *Drosophila* embryo were done only long after it was clear that the genetic approach was a success.

The systematic genetic screens have focused on two types of mutants: those that disrupt early embryonic pattern formation processes; and those that lead to transformations of body regions. The first group of genes are required to build the basic structure of the body: the bauplan. If one of them is mutant, specific regions of the body are missing. There is no generic name for this group

*Author for correspondence.
†Present address: Cornell University, 403 Biotechnology Building, Ithaca, NY 14853-2703, USA.

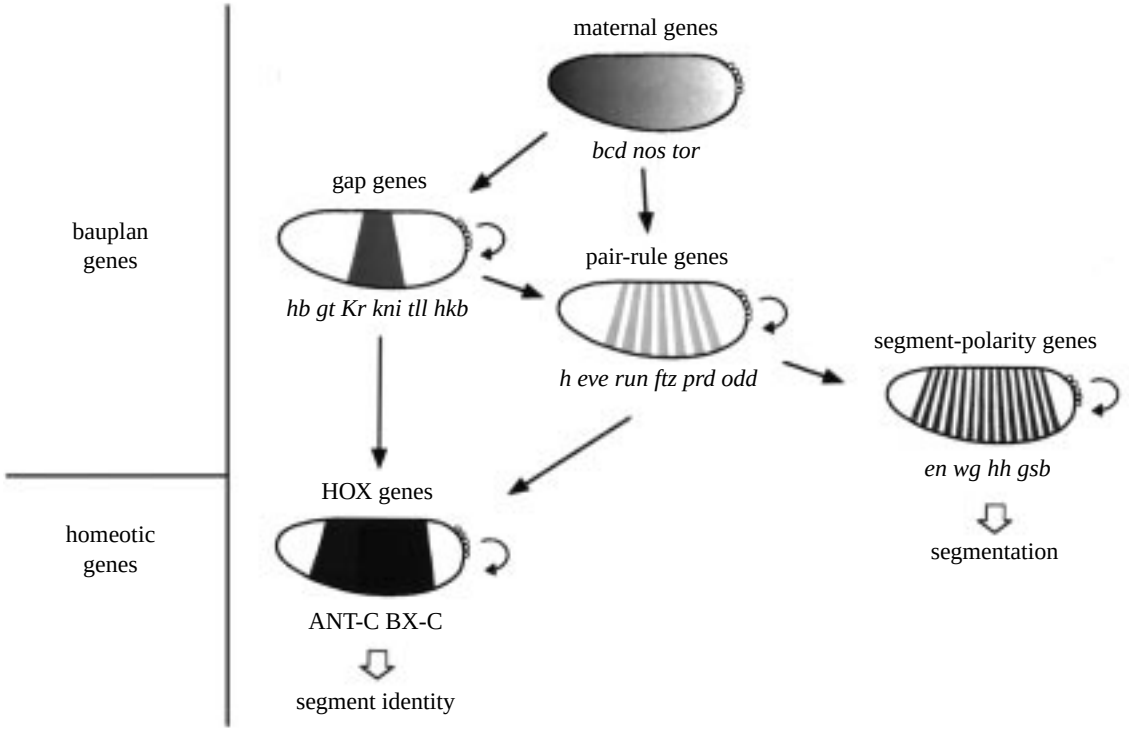

Figure 1. Schematic representation of the genetic hierarchy organizing the anterior–posterior axis in *Drosophila*. Some essential genes at each step of the hierarchy are represented by their respective abbreviations (see Pankratz & Jäckle (1993) for more details). The maternal genes are those that are required for providing positional information in the egg. Their products become localized during embryogenesis and act as long-range gradients to regulate the positioning of the gap genes. These are then expressed in broad domains (the example shown is the expression of *Krüppel*), which generate a series of overlapping short-range gradients. The combinatorial interaction of these gradients generates then the transient expression of the pair-rule genes. These are expressed in seven stripes, whereby each stripe is regulated independently by a certain combination of the gap gene and maternal gene products. The combinatorial interaction of the pair-rule genes is required to regulate the segment-polarity genes, which eventually determine the segment boundaries. The gap genes and the pair-rule genes are also required to regulate the genes of the two homeotic complexes which eventually specify segment identity. A similar cascade of gene interactions is required to specify the subdivision of the dorso-ventral axis into different subregions (Chasan & Anderson 1993).

of genes, but the term 'bauplan genes' might be fitting. The second group of genes is required to determine the identity of body regions. Their generic name is 'homeotic genes'. If they are mutant, one gets a transformation of certain body regions into other regions, but not loss of regions as is the case for mutants in bauplan genes.

The bauplan genes include also those genes that are necessary to generate the initial maternally determined asymmetries in the early embryo, which specify the anteriorposterior and dorso-ventral axis (St Johnston & Nüsslein-Volhard 1992). These asymmetries are then interpreted by the genes that are expressed in the developing embryo (Pankratz & Jäckle 1993) (figure 1). They generate the segmental subdivision, as well as the regionalization of the dorso-ventral axis (Chasan & Anderson 1993). Most of the interactions between the gene products occurring at these early stages are well understood in *Drosophila*. Both the initial dorso-ventral and the anterior–posterior asymmetries are ultimately caused by the asymmetry of the cytoskeleton in the early oocyte (Grünert & St Johnston 1996). A cascade of gene functions is then required to transform this information into several morphogenetic gradients of transcription factors. These factors regulate the expression of the zygotic segmentation genes, which act again via a cascade of regulatory molecules to achieve the functional subdivision of the embryo. Most of these early gene functions are short and transient, as they are only required for setting-up a set of stably expressed genes. Among these stably expressed genes are the homeotic genes that provide the different body regions with an identity, in particular in the anteriorposterior axis. If the homeotic genes were missing, the whole body would be transformed into a series of identical segments. One of the intriguing findings in this context was that the homeotic genes are clustered and that the anteriorposterior order in which they act reflects their linear position on the chromosome (Lewis 1978; McGinnis & Krumlauf 1992). The reason for this is still not understood, but the fact that this arrangement is highly conserved in evolution suggests that there must be some underlying regulatory principle that is required for this type of function.

Most of the early acting genes code for regulatory molecules which are either involved in transcriptional or translational regulation or in signalling processes. Many belong to well-known gene families, which include protein domains that may be conserved even between prokaryotes and eukaryotes. Generally, one can say that almost all types of regulatory molecules that are known

in cell biology are also involved in some aspect of early development. Still, mainly for historical reasons, one type of regulator is frequently singled out, namely the homeo-box containing transcription factors. They were first identified in the homeotic gene clusters (hence the name 'homeo box'; Gehring 1994), but occur also outside the cluster and may have very different types of functions in the developmental gene hierarchy. Also, not all homeotic genes code for homeo-box proteins, as, for example, those that are required for providing the identity to the terminal structures (Jürgens & Hartenstein 1993). To differentiate between the different types of genes, it has become customary to call genes that contain a homeo box and that are considered to function in the homeotic gene cluster 'HOX genes'.

It now seems clear that all animals contain at least one HOX gene cluster and that these are largely expressed in an anterior–posterior direction according to their position on the chromosome (McGinnis & Krumlauf 1992). This finding has been considered as so significant that it was even suggested as a defining character for all animals (the 'zootype'; Slack *et al*. 1993). However, one should note that though the HOX genes play an important role, they still constitute only a small part of the whole developmental gene hierarchy. Moreover, it is only their clustered organization that makes them special, not their degree of conservation, as other developmental genes may be equally or even more conserved.

In contrast to *Drosophila*, the developmental analysis of vertebrate embryos has proceeded in a more traditional direction, using embryo manipulation and biochemical methods to understand the formation of the early embryonic structures. A systematic genetic approach has only recently been undertaken for the zebrafish (Granato & Nüsslein-Volhard 1996), but the genes identified in these screens still have to be characterized. However, the traditional approach has, by now, also provided a very detailed understanding of the basic principles of early pattern formation in a vertebrate embryo. For *Xenopus* in particular, it was possible to identify genes that are involved in axis specification and tissue determination through the early organizer regions (Harger & Gurdon 1996). One of the fascinating outcomes of these studies is that the dorso-ventral patterning cascade uses two antagonistic genes (*BMP4* and *chordin*) that have complimentary homologues in the *Drosophila* dorso-ventral patterning cascade (*decapentaplegic* and *short gastrulation*). However, whereas *BMP4* specifies a ventralizing signal, its *Drosophila* homolog *decapentaplegic* is required for a dorsalizing signal (Ferguson 1996). This finding corroborates at the molecular level the long-held view that chordates and arthropods have a reversed dorso-ventral axis with respect to each other (Arendt & Nübler-Jung 1997).

Another emerging system for evolutionary comparisons in developmental biology are the nematodes. *Caenorhabditis elegans* has become one of the best studied organisms, because many of its developmental cell fate decisions can be traced to their molecular origins. This profound knowledge serves now as a basis for comparative studies in different nematode species. The structures of particular interest are currently the generation of the vulva (Sommer *et al*. 1994) and the tail (Fitch 1997). In the case of the vulva in particular, it has been shown that genetic screens for appropriate phenotypes can easily be done, also for other nematode species (Sommer *et al*. 1994). In the long run this should provide a very strong basis for inferences on the evolution of cell–cell communication pathways.

In contrast to animals, comparative developmental biology in plants is only at its beginning. Again, it is mainly the genetic screens for specific phenotypes in either floral development or embryonic development, that are about to provide a very detailed picture of the regulatory decisions that are at the basis of pattern formation in plants. The genes involved in developmental decision are also currently being molecularly characterized and it seems again that conserved gene families of regulatory genes play a role (Coen & Nugent 1994; Theißen & Saedler 1995).

3. THE PHYLOTYPIC STAGE

All animals (and to a certain degree also plants) seem to go through a particularly stereotypic phase during their ontogenetic development, in which the basic outline of the body pattern is generated before it is further modified to form the adult individual. This phase looks morphologically very similar, even among very distantly related taxa and is nowadays called the 'phylotypic stage' (Sander 1983). Interestingly, the morphological diversity of embryos before they reach this stage can be considerable. In particular, the first cleavage events and early embryonic development can differ markedly, even between closely related taxa. This observation of early and late morphological diversity with an intermediate stereotypic stage has become known as the hourglass model of development (Raff 1996). Furthermore, because of the observation that the expression of the HOX-genes is at its peak during the phylotypic stage, it was suggested that there may be an underlying molecular principle which necessitates this stage. It was suggested that the HOX complex acts as a timing device which links anterior–posterior patterning with growth control (Duboule 1994). Such an idea seems attractive if one considers the possibility that such a timing device may have already existed in the unicellular precursors of metazoans, where it might have acted to control different stages of a life cycle. This could then have been coopted for an anterior–posterior patterning device in the first multicellular organisms. Still, the spatial deployment of the HOX genes in *Drosophila* depends heavily of the regulation by the bauplan genes. Whether this is only the case for *Drosophila* is as yet unclear, since so little is known about these genes outside of insects. Another argument for explaining the constraints on the evolvability of the phylotypic stage might be the assumed multitude of modular developmental interactions that occur at this stage (Raff 1996). The large diversity before and after this stage would be explained by a lower number of modular interactions and by a requirement for ecological adaptations to the respective environments (Raff 1996; Sander 1983). However, these models do not shed light on the question of how the phylotypic stage should have evolved in the first place.

The hourglass model assumes implicitly that ontogeny has a defined start and end point. However, ontogeny is clearly a cyclical process and the asymmetric-wheel

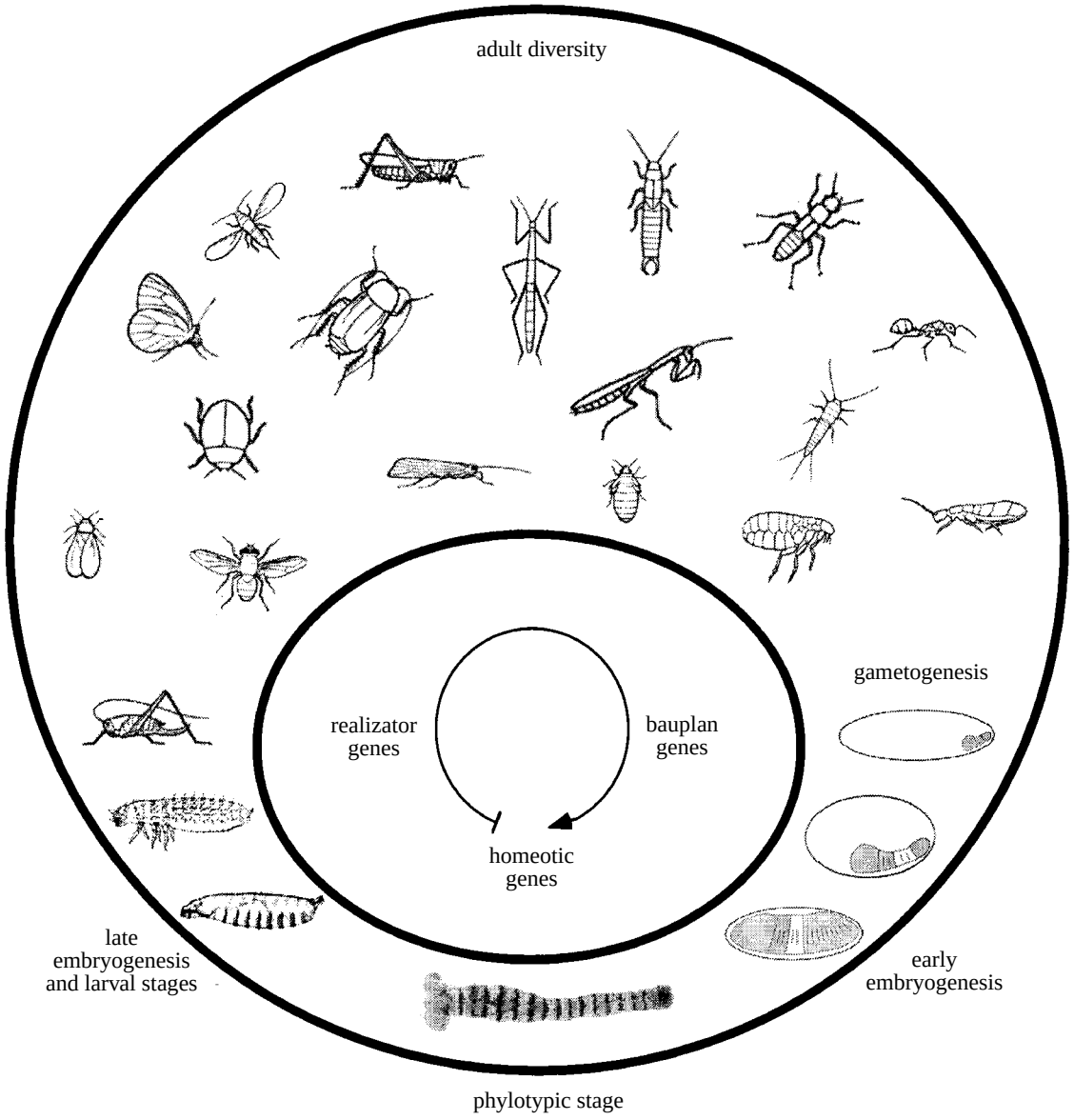

Figure 2. Representation of the phylotypic stage of insects within the asymmetric-wheel model. The phylotypic stage is represented by an embryo from a beetle (the flour beetle *Tribolium castaneum*), which has been stained with an antibody against the segment-polarity gene *engrailed* to visualize the segment boundaries that have been established at this stage. This stage looks very similar for all the different insects shown. Further development proceeds via larval stages (depicted as the larvae from *Drosophila* and *Tribolium*) which are already more diverse, but still show more similarities between the different species than the respective adults. The full diversity of the species is only realized at the adult level. The adults go on to produce sperm and eggs (gametogenesis), which are different between the species, though the range of possibilities is somewhat reduced. There are essentially three categories of eggs and early embryos in insects: the short-germ, the intermediate-germ, and the long-germ embryos which differ with respect to their yolk content and the size of their primary germ rudiment. Short-germ embryos (for example, the grasshopper *Schistocerca*) develop at blastoderm stage only a very small germ rudiment which consists essentially of headlobes and a growth zone, intermediate-germ embryos (for example, *Tribolium*) include the head lobes, thoracic segments and a growth-zone, whereas long-germ embryos (for example, *Drosophila*) generate all their segments at blastoderm stage (see Tautz *et al.* (1994) for a discussion of these types). Though the examples shown are restricted to insects, all arthropods, i.e. also the crustaceans, millipeds and spiders show easily comparable phylotypic stages to those of insects.

model depicted in figure 2 would therefore better represent the situation. As in the hour-glass model, it takes account of the fact that the adult morphology diversifies because of ecological adaptations. The adults then go on to produce eggs which themselves may be subject to adaptations. They may, for example, be designed for external or internal development or may be rich or poor in yolk content, or may be tuned for fast or slow development. However, the options available at that stage will already be fewer than for the adult individuals and the wheels are therefore merging again (figure 2). During early ontogeny, the germ-line cells will have to redeploy their genetic bauplan information to produce a three-dimensional embryo. This has to be done in different egg environments and will therefore go along somewhat different routes, depending on this environment. However, we assume that

a more or less conserved set of bauplan genes should act at these stages which eventually lead up to the correct deployment of the HOX genes. From this stage onwards, a new set of genes may become more prevalent to regulate the details of the formation of the adult morphology. These genes may be called the 'realizator genes' (Garcia-Bellido 1975) and we suspect that they are the genes that are required for ecological adaptations.

In the asymmetric-wheel model, it would seem sensible to use the phylotypic stage as the starting point of the cycle and not the point of fertilization, as it is usually done. In *Drosophila*, for example, pattern formation starts well before fertilization, namely during oogenesis. The main axes of the future embryo become specified by cell–cell signalling interactions between the oocyte and the surrounding nurse cells. After fertilization and egg laying, this prelocalized positional information is merely interpreted by the zygotic genome (St Johnston & Nüsslein-Volhard 1992). In contrast, the axis specification in chordates are apparently not achieved during oogenesis, but only after fertilization (Eyal-Giladi 1997). These examples suggest that the initiation of pattern formation does not need to be triggered by fertilization or egg laying. This has to be taken into account when making inferences about the possible conservation of developmental pathways between taxa.

4. DEVELOPMENTAL PATHWAYS AND MODULES

Early development in *Drosophila* proceeds along genetically and molecularly defined pathways. As yet, it is unclear to what extent these pathways are evolutionarily conserved. They might also reflect special adaptations to the long-germ mode of embryogenesis found in *Drosophila*, though we have argued that the system may be more or less conserved, even in less derived insects which show a somewhat different form of embryogenesis (Tautz & Sommer 1995). Furthermore, the fact that the correct regulation of the conserved HOX gene expression domains depends on these early pathways in *Drosophila* would suggest that at least some essential components of the pathways should also be conserved. There are indeed examples of early acting genes whose homologues show persuasively similar expression patterns in chordates, in particular genes involved in head development (Bally-Cuif & Boncinelli 1997). Interestingly, even some genes found to be primarily involved in segmentation in *Drosophila* were found to be similarly expressed during somitogenesis in chordates, even though these structures can not as yet be considered to be homologous (Müller *et al.* 1996; Holland *et al.* 1997). However, it is still largely unclear how much of the regulatory interactions among them are also conserved. Furthermore, there is an alternative explanation, namely that these genes are parts of versatile developmental modules which can be redeployed in different contexts.

One good example for such a redeployment, which is even involved in the generation of an adaptive adult character, is the involvement of the *distal-less* (*dll*) gene in the generation of butterfly eyespots (Carroll *et al.* 1994). The *dll* gene is an important gene required for the generation of the proximo-distal axis in *Drosophila* appendages and qualifies therefore as a bauplan gene. Moreover, its expression in appendages throughout the animal kingdom (Panganiban *et al.* 1997) suggests that this function is highly conserved. However, as vertebrate and invertebrate legs are clearly not homologous structures, one has to conclude that the proximo-distal specification pathway in which *dll* is involved, has frequently been recruited in different phyla for patterning body wall outgrowths (Panganiban *et al.* 1997). It therefore seems that this pathway acts as a developmental subroutine, which can be put into different contexts. The expression of *dll* in the eye spots in butterfly wings may reflect this versatility. Although it is not an appendage which is generated in this case, it is still an axis that is specified, namely an outer to inner axis of a planar field.

There is also another type of redeployment of modules during limb formation of vertebrates. It appears that some genes of the HOX cluster have become duplicated and are then specifically used to regulate limb development in a temporal–spatial progression (Sordino & Duboule 1996). Also, signalling molecules from the *wingless* gene family, which was first identified in the segmentation gene pathway in *Drosophila*, turn out to be involved in many developmental decisions at various stages of development (Nusse 1997), including leg formation in *Drosophila* (Lecuit & Cohen 1997). In fact, most of the segmentation genes in *Drosophila* are also re-expressed at later stages in multiple organs and may be functionally involved in specifying these.

Such considerations make it difficult to predict the role of bauplan genes in the generation of adult characters. Although it seems unlikely that changes in the early pathways leading to the phylotypic stage are required for late adaptive characters, it still seems possible and likely that regulatory modules are reused for the generation of specific structures at later stages. On the other hand, the reuse of genes in different pathways would make them bad candidates for driving adaptive evolution, as mutations in them would always be expected to have multiple, probably maladaptive effects. A more specialized set of genes would be less problematic in this respect.

5. SCREENS FOR ADAPTIVE TRAIT GENES

Morphological traits usually show a certain degree of quantitative variation. Part of this may be environmentally induced, but it is also clear that there is a genetic basis for this variation. This is best shown by the fact that one can take almost any quantitative morphological character and subject it to artificial selection for high and low values. One of the most intensively studied traits in this context is sensory bristle numbers on adult *Drosophila* flies, which have served as a model system to understand the genetics of quantitative variation (Mackay 1996).

There are basically two opposing models of how the genetics of quantitative traits can be understood. The first assumes a large number of genes with each having only small effects on the trait, whereas the second assumes the existence of a few loci with major effects, but possibly complemented by other loci with minor effects (Barton & Turelli 1989; Orr & Coyne 1992; Mitchell-Olds 1995). The first model is mainly supported by population genetical reasoning. It assumes that spontaneous mutations with small effects should be more frequent than those with

large effects. Also, mutations with large effects on phenotypic traits would be considered to have pleiotropic consequences on other characters. Moreover, the fixation of mutations with small effects should be easier than those with large effects. The latter would be expected to require strong selection over many generations to become fixed, because the associated pleiotropic and usually maladaptive effects must be overcome. However, at least this condition is met in artificial selection experiments and genetic analysis of the genes involved in bristle number variation in *Drosophila* has turned up several major effect loci. As predicted, some of these have pleiotropic effects on other fitness components, i.e. effect viability. In addition, both additive and epistatic effects were found for the different loci (Long *et al.* 1995). Similar conclusions were reached in another extensive study that has looked for morphological shape differences in male genitalia of *Drosophila* (Liu *et al.* 1996). In this case, it was not artificial selection that was employed, but hybridization between very closely related species that showed evolutionary fixed differences for this morphological trait. In this respect, the experiment can be considered as more 'natural' because the differences were not generated by artificial selection. Still, the analysis suggests again that there could be a small number of large effect loci which act mainly additively. On the other hand, the alternative, namely a clustering of multiple small effect loci in small chromosome regions could not be ruled out, as the mapping of the chromosomal intervals bearing the loci was relatively crude (Liu *et al.* 1996). A similar study has been done for two monkeyflower (*Mimulus*) species, which occur sympatrically, but are reproductively isolated. Still, they are fully fertile when artificially mated and it was thus possible to map eight floral traits that distinguish them. For each of the traits a chromosomal region could be identified that accounted for more than 25% of the variance (Bradshaw *et al.* 1995). Thus, major genes have likely played a role in the diversification of these plants, but there is still plenty of room for the influence of minor genes as well.

What are the genes that cause these quantitative trait differences? Unfortunately, at least for *Drosophila*, this is not clear yet, as the mapping methods will have to become much more refined (the main obstacle being the limited number of analyses that can be done for a single recombinant fly). However, at least candidate loci could be identified for the bristle number variation experiment (Long *et al.* 1995). Several of the known neurogenic genes turned out to be located in the intervals mapped for large effects. This seems sensible, as sensory bristles are products of the neurogenic pathway. Furthermore, the genes involved are known to act at multiple stages and in different regulatory contexts during development, which would also explain their pleiotropic effects. Unfortunately, this would make them less likely candidates for genes that effect bristle number traits under natural conditions, and it therefore remains open whether there are multiple small effect loci after all that could be the basis for natural selection.

In maize it has been possible to identify major loci that have caused a particular morphological change. The domestication of maize from its wild ancestor teosinte has involved strong selection for apical dominance, i.e. the concentration of the resources in the main stem of the plant instead of the axillary branches. Quantitative trait loci mapping has allowed the identification of two epistatically acting major genes involved in this, one of which is called the *teosinte branched1* locus (Doebley *et al.* 1995). The latter gene could subsequently be cloned and the molecular analysis suggests that it might be a conserved regulatory gene (Doebley *et al.* 1997). Interestingly, it appears that the differences between maize and teosinte are caused by different levels of expression rather than primary amino acid changes (Doebley *et al.* 1997). However, the transformation from teosinte into maize was again an artificial selection experiment. In this sense it corresponds to the examples discussed for *Drosophila*. Thus, the fact that only few major effect loci were identified does not need to imply that minor effect loci too might not play a role under more natural evolutionary conditions.

6. A SCREEN FOR NATURALLY SELECTED LOCI

Ideally, one would like to identify genes that have played a role in adaptations under natural conditions to better understand the functions of such genes. This makes it necessary to study wild-type populations and to avoid laboratory experiments, at least until a locus is identified. A pilot study in *Drosophila* shows an example of how this might be done (Schlötterer *et al.* 1997). *Drosophila melanogaster* has only recently colonized multiple habitats around the world. It seems to have originated in tropical Africa, but has adapted to very different climates and ecological conditions within only a few thousand years. Thus, the comparison of different wild-type populations from different climatic regions should allow the detection of genes that have been under positive selection to achieve the adaptations. Positive selection would result in a loss of polymorphism in the chromosomal region where the gene resides (Taylor *et al.* 1995). This hitchhiking effect should allow identification of the respective chromosomal regions. The practical approach is to survey highly polymorphic anonymous markers in different populations to see whether population and locus-specific loss of heterozygosity can be detected. Indeed, in an analysis of ten hypervariable microsatellite loci in *D. melanogaster* populations, we could identify at least one locus with a significant loss of heterozygosity, suggesting that it might be linked to a gene that has undergone a selective sweep (Schlötterer *et al.* 1997). Again, it will be necessary to conduct a much more refined study to identify a candidate gene for this effect, but our pilot study at least suggests that the initial approach might be feasible and that the chromosomal regions affected by adaptations to local environments might be tractable.

7. FAST-EVOLVING GENES

Another, as yet rather indirect approach to find genes involved in adaptations, is to look for genes that evolve very fast. The bauplan genes discussed above usually belong to well-known protein families whose sequence and structure are highly conserved over large evolutionary distances. Usually this correlates with their functional conservation in diverse taxa. Thus, if one would look specifically for genes which are not conserved, one might

be able to identify some that are responsible for novel or adaptive phenotypes in specific taxa. In fact, Rice & Holland (1997) have argued that genes might evolve particularly fast under conditions where they are involved in an 'interlocus contest' which could occur under a number of scenarios dealing with ecological and sexual adaptations.

Currently, the origin of evolutionary novelties is generally thought to be caused by two different processes, either evolutionary changes in regulatory networks or gene duplications with subsequent diversification of the genes. The latter process in particular would suggest that there is only a limited number of ancestral protein domains, that have become duplicated and reshuffled during evolution to make up the genes that we find today in complex organisms (Dorit *et al*. 1990; Orengo *et al*. 1994). The data from the current genome-sequencing projects suggest that this expectation may indeed be true. It was estimated that the number of naturally occurring protein domains which make up the universe of proteins is only between 1 000 and 7 000, most of which are thought to be already represented in the sequence databases (Orengo *et al*. 1994; Chotia 1994). On the other hand, in genome projects of higher eukaryotes one finds a relatively high number of proteins (up to about 40%) that appear to have no similarity to known sequences and whose identity and function remains unknown (Goffeau 1994; Wilson *et al*. 1994; Dujon 1996). One could assume that these 'orphans' are the first known members of as yet undiscovered protein families. However, as an alternative explanation we propose that the sequence of these proteins evolves so fast that their homologues cannot be identified in distant species by molecular methods. If such a class of genes exists, it might also be a source of evolutionary novelties.

To address this question, we have devised a screen to estimate the proportion of rapidly evolving genes in *Drosophila* (Schmid & Tautz 1997). A set of about 100 different cDNAs were randomly isolated from an embryonic *D. melanogaster* cDNA library, and their sequence evolution was examined on three different timescales: between distantly related species (60–250 Ma), closely related species (15 Ma) and different populations of sibling species (<0.2 Ma).

In the first step, the conservation of these clones was analysed by filter hybridization against genomic DNA from three insect species with increasing evolutionary distance. The hybridization conditions were chosen such that genes that evolve neutrally or close to a neutral rate would not hybridize to the genomic DNA of any of these species. Surprisingly, we found that more than one-third of the cDNAs did not even cross-hybridize with the genomic DNA from a distantly related *Drosophila* species under these conditions. To yield a signal in this assay, the probe would have had to include at least one stretch of about 100 b.p. with more than 65% sequence indentity. Thus, genes including one or more of the known conserved protein domains (see above) should have lit up. Partial sequencing of the clones and searching of sequence data bases showed that a similar proportion did not result in any matches with previously known genes or sequence motifs. Most interestingly, among the non-matching genes, there was only one that was already known from *D. melanogaster*, whereas 19 already known *D. melanogaster* genes were found in the more conserved class of clones (table 1). This suggests that there is a substantial proportion of fast-evolving genes in the *Drosophila* genome which are significantly under-represented among the genes that have been studied so far.

Table 1. *Numbers of randomly drawn cDNA clones in the conserved and fast-evolving classes and comparison with previously known* Drosophila *genes*

(The exact numbers provided are corrected for clones that turned out to be composed of non-coding 3′-ends only.)

	all cDNAs	exact *Drosophila* matches
conserved clones	49	19
fast evolving clones	46	1
G-test: $G=15.6$, $p<0.001$		

To study the nature of the fast-evolving sequences in more detail, their homologues were isolated from the closely related species *D. yakuba*, which has an evolutionary distance of about 15 Ma from *D. melanogaster*. This is a distance where one would expect that even neutrally evolving sequences should be recovered. For ten fast-evolving and one conserved clone, the homologous sequences were obtained and the cDNAs were completely sequenced. A total of 9 out of the 11 gene pairs contained an homologous open-reading frame, indicating that they code for functional proteins. The calculation of the substitution rates between these pairs showed that they are indeed among the fastest-evolving genes known from this species pair. A total of four of them showed amino-acid replacement rates that were only half as fast as the corresponding replacement rates at third codon positions, which are considered to evolve close to the neutral rate (table 2).

The third step of our analysis adressed the question of whether the high evolutionary rates are caused by neutral evolution due to low functional constraints on the protein sequence or by continuous adaptive selection for new variants as it would be predicted in the Rice & Holland (1997) scenarios. McDonald & Kreitman (1991) have suggested that a comparison of polymorphisms within populations with fixed replacements between species, should allow to assess whether fixation of amino acids has mainly occurred by drift or by selection. We have analysed the fastest-evolving genes in this test and have found no evidence for selection so far. In other words, the sequences showing the high replacement rates show also a high within species-polymorphism rate. However, the test is rather conservative in several respects and a more detailed analysis of the polymorphisms detected has still to be done. Another test which could point to selection is to check whether there are more positions that result in amino-acid replacements between species, than there are replacements at non-coding positions (compare with Tsaur & Wu (1997)). Again, we have not found evidence for this in our sequences if one uses them as a whole. However, there are some particularly fast-evolving subregions which show this effect. Thus, it will be necessary to better understand the details of their structure and function before solid statements can be made.

As yet, it's unclear what the function of the fast-evolving genes may be. We find that they are expressed in similar

Table 2. *Comparison of non-synonymous (K_a) and synonymous (K_s) substitutions per site of different genes and cDNAs from* D. melanogaster *and* D. yakuba *(CI is confidence interval)*

(The four fastest evolving cDNAs from the randomly drawn pool are at the top of the list. The other comparisons are based on previously published data (see Schmid & Tautz (1997) for details). Clone 2A12 was identified as a conserved clone in the genomic southern blot screening experiment and codes for a kinesin-like protein.)

gene	codons	K_a	95% CI	K_s	95% CI
1G5	347	0.168	0.252–0.441	0.346	0.134–0.199
1E9	393	0.116	0.092–0.139	0.279	0.205–0.356
2D9	280	0.108	0.070–0.143	0.233	0.157–0.303
1A3	227	0.092	0.063–0.120	0.196	0.119–0.279
anon–3B1.2	260	0.079	0.056–0.104	0.284	0.191–0.367
period	1233	0.032	0.026–0.040	0.309	0.257–0.340
Amylase	494	0.020	0.012–0.028	0.110	0.072–0.142
G–S–T	208	0.017	0.002–0.030	0.114	0.058–0.168
Adh	256	0.015	0.005–0.026	0.156	0.094–0.211
G–6–P	558	0.011	0.005–0.017	0.198	0.151–0.240
hunchback	829	0.011	0.005–0.015	0.183	0.143–0.220
CO I	498	0.007	0.002–0.0012	0.315	0.251–0.396
2A12	498	0.006	0.001–0.011	0.325	0.260–0.401

ways as the slow evolving ones (Schmid & Tautz 1997), indicating that they may be involved in both general cellular functions as well as specific developmental ones. However, we suspect that mutants in them would not cause obvious phenotypes, as the cDNAs that we have recovered are significantly under-represented among those *Drosophila* genes that have been studied because they have a defined genetic effect (table 1). Interestingly, there is one well-studied example of a *Drosophila* gene which has no apparent genetic effect. The gene was cloned because it was considered to be the homeotic gene *spalt*. It codes for a rather short protein with a functional signal sequence and it showed a very high evolutionary rate between closely related species (Reuter *et al*. 1989). However, it turned out that the real *spalt* gene is located about 20 k.b.p. away and codes for a highly conserved zinc-finger protein. The former gene was therefore called *spalt-adjacent* (Reuter *et al*. 1996). Genetic analysis of *spalt-adjacent* showed that it does not affect viability if it carries a premature stop codon. In fact, some standard laboratory strains are homozygous for this mutation. Furthermore, even an artificial overexpression construct did not yield any recognizable phenotype (Reuter *et al*. 1996). Thus it seems likely that *spalt-adjacent* has only a small function, which may only be relevant for flies under natural conditions.

There are other examples of fast-evolving genes in *Drosophila* as well. Among them are the *period* gene, which is involved in the species-specific song rhythms of the fly (Thackeray & Kyriacou 1990), the *transformer* gene, which is involved in the sex-determination cascade (O'Neil & Belote 1992) and the male ejaculatory protein gene *Acp-26A* (Tsaur & Wu 1997). It seems significant that all three of these genes are positioned in pathways where one would expect adaptive evolution to take place. Rice & Holland (1997) list more such examples from other species. There is also evidence for a substantial number of possibly fast-evolving genes from the yeast (Dujon 1996) and the nematode (Wilson *et al*. 1994) genome projects which are called 'orphans' in these cases. The evolutionary origins of these orphan genes remain as yet unknown, because most of the sequences from eukaryotes that are available in databases come from a few model organisms separated by large evolutionary distances. Studying the evolution of such genes over close evolutionary distances should help to understand them better in the future.

8. CONCLUSIONS

The genetics of the evolution of phenotypic adaptive traits still remains a puzzle. However, at least the framework within which it occurs is now becoming clearer. There is, on the one hand, a huge and increasing knowledge about the genes involved in embryogenesis and, on the other hand, there are feasible approaches to study quantitative trait loci involved in generating adaptive characters of adult individuals. Intuitively, one would expect that the former group of genes is mainly involved in the generation of major evolutionary novelties or changes in the bauplan, whereas the quantitative trait loci might be more important for ecological adaptations and speciation events. In other words, there might be a group of genes involved in the generation of macro-evolutionary novelties and a second group of genes required for micro-evolutionary adaptations (see Orr & Coyne (1992) for a relevant discussion of these terms). Accordingly, one would expect that changes in the former group of genes are relatively infrequent and that one should not normally find polymorphisms in them which have phenotypic consequences. However, even this expectation is not born out. Gibson & Hogness (1996) showed that there is a natural polymorphism in the *Ultrabithorax* locus of wild-type *Drosophila melanogaster* populations that results in the transformation of a whole body segment under environmental stress conditions. In fact this polymorphism behaves like a typical major quantitative trait loci locus, as it responds to artificial selection and the magnitude of its effect is influenced by other non-linked minor loci. Evidently, as segment transformations are not normally seen among wild-type flies, one would have to suspect that this polymorphism has a different role for late

expressed characters under natural conditions. Still, this result together with the results from the bristle number experiments, suggests that the major quantitative trait loci might indeed uncover genes that are otherwise primarily involved in early developmental decisions. On the other hand, it is also evident that there are many minor loci influencing any particular quantitative trait. Interestingly, one possible reason for this might be that developmental decisions need to be safeguarded by redundant pathways (Tautz 1992; Nowak *et al*. 1997). Modelling of such safeguarding effects shows that more and more genes would become recruited during evolution for any developmental process which has a certain chance of making errors (Nowak *et al*. 1997). It would then be expected that each of the genes might have a small effect on the process as a whole. Evidently, it will be difficult to identify unequivocally such genes with small or redundant effects, but candidates for them might be found among the fast-evolving genes discussed above. Though the evidence for this is rather indirect as yet, these genes would at least provide a large source of polymorphisms that could be exploited by natural selection on phenotypic traits and we believe therefore that they should be studied more closely in the future.

REFERENCES

Arendt, D. & Nübler-Jung, K. 1997 Dorsal or ventral: similarities in the fate maps and gastrulation patterns in annelids, arthropods and chordates. *Mech. Dev.* **61**, 7–21.

Bally-Cuif, L. & Boncinelli, E. 1997 Transcription factors and head formation in vertebrates. *BioEssays* **19**, 127–135.

Barton, N. H. & Turelli, M. 1989 Evolutionary quantitative genetics—how little do we know? *A. Rev. Genet.* **23**, 337–370.

Bradshaw, H. D., Wilbert, S. M., Otto, K. G. & Schemske, D. W. 1995 Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* **376**, 762–765.

Carroll, S. B., Gates, J., Keys, D. N., Paddock, S. W., Panganiban, G. E., Selegue, J. E. & Williams, J. A. 1994 Pattern formation and eyespot determination in butterfly wings. *Science* **265**, 109–114.

Chasan, R. & Anderson, K. V. 1993 Maternal control of dorsal-ventral polarity and pattern in the embryo. In *Development of* Drosophila melanogaster (ed. M. Bate & A. Martinez-Arias), pp. 387–424. New York: Cold Spring Harbor Laboratory.

Chotia, C. 1994 Protein families in the metazoan genome. *Dev. Suppl*. 27–33.

Coen, E. S. & Nugent, J. M. 1994 Evolution of flowers and inflorescences. *Dev. Suppl*. 107–116.

Doebley, J., Stec, A. & Gustus, C. 1995 *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of domonance. *Genetics* **141**, 333–346.

Doebley, J., Stec, A. & Hubbard, L. 1997 The evolution of apical dominance in maize. *Nature* **386**, 485–488.

Dorit, R. L., Schoenbach, L. & Gilbert, W. 1990 How big is the universe of exons? *Science* **250**, 1377–1382.

Duboule, D. 1994 Temporal colinearity and the phylotypic progression: a basis for the stability of the vertebrate bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl*. 135–142.

Dujon, B. 1996 The yeast genome project: what did we learn? *Trends Genet.* **12**, 263–270.

Eyal-Giladi, E. 1997 Establishment of the axis in chordates: facts and speculations. *Development* **124**, 2285–2296.

Ferguson E. L. 1996 Conservation of dorsal–ventral patterning in arthropods and chordates. *Curr. Opin. Genet. Dev.* **6**, 424–431.

Fitch, D. H. A. 1997 Evolution of male tail development in rhabditid nematodes related to *Caenorhabditis elegans*. *Syst. Biol*. **46**, 145–179.

Garcia-Bellido, A. 1975 Genetic control of wing disc development in *Drosophila*. In *Cell patterning* (ed. R. Porter & K. Elliott), pp. 161–178. Amsterdam: Elsevier.

Gehring, W. 1994 A history of the homeobox. In, *Guidebook to the homeobox genes* (ed. D. Duboule), pp. 1–10. Oxford University Press.

Gibson, G. & Hogness, D. S. 1996 Effect of polymorphism in the *Drosophila* regulatory gene *Ultrabithorax* on homeotic stability. *Science* **271**, 200–203.

Goffeau, A. 1994 Genes in search of functions. *Nature* **369**, 101–102.

Granato, M. & Nüsslein-Volhard, C. 1996 Fishing for genes controlling development. *Curr. Opin. Genet. Dev.* **6**, 461–466.

Grünert, S. & St Johnston, D. 1996 RNA localization and the development of asymmetry during *Drosophila* embryogenesis. *Curr. Opin. Genet. Dev.* **6**, 395–402.

Harger, P. L. & Gurdon, J. B. 1996 Mesoderm induction and morphogen gradients. *Sem. Cell Dev. Biol.* **7**, 87–93.

Holland, L. Z., Kene, M., Williams, N. A. & Holland, N. D. 1997 Sequence and embryonic expression of the Amphioxus *engrailed* gene (*AmphiEn*): the metameric pattern of transcription resembles that of its segmentpolarity homolog in *Drosophila*. *Development* **124**, 1723–1732.

Jürgens, G. & Hartenstein, V. 1993 The terminal regions of the body pattern. In *Development of* Drosophila melanogaster (ed. M. Bate & A. Martinez-Arias), pp. 687–746. New York: Cold Spring Harbor Laboratory.

Lecuit, T. & Cohen, S. M. 1997 Proximal–distal axis formation in the *Drosophila* leg. *Nature* **388**, 139–145.

Lewis, E. B. 1978 A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 761–769.

Liu, J., Mercer, J. M., Stam, L. F., Gibson, G. C., Zeng, Z.-B. & Laurie, C. 1996 Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *D. mauritiana*. *Genetics* **142**, 1129–1145.

Long, A. D., Mullaney, S. L., Reid, L. A., Fry, J. D., Langley, C. H. & Mackay, T. F. C. 1995 High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* **139**, 1273–1291.

Mackay, T. F. C. 1996 The nature of quantitative variation revisited: lessons from *Drosophila* bristles. *BioEssays* **18**, 113–121.

McDonald, J. H. & Kreitman, M. 1991 Adaptive protein evolution at the ADH locus in *Drosophila*. *Nature* **351**, 652–654.

McGinnis, W. & Krumlauf, R. 1992 Homeobox genes and axial patterning. *Cell* **68**, 283–302.

Mitchell-Olds, T. 1995 The molecular basis of quantitative genetic variation in natural populations. *Trends. Ecol. Evol*. **10**, 324–328.

Müller, M., von Weizäcker, E. & Campos-Ortega, J. A. 1996 Expression domains of a zebrafish homologue of the *Drosophila* pair-rule gene *hairy* correspond to primordia of alternating somites. *Development* **122**, 2071–2078.

Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. 1997 Evolution of genetic redundancy. *Nature* **388**, 167–171.

Nusse, R. 1997 A versatile transcriptional effector of *wingless* signaling. *Cell* **89**, 321–323.

Nüsslein-Volhard, C. & Wieschaus, E. 1980 Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801.

Orengo, C. A., Jones, D. T. & Thornton, J. M. 1994 Protein superfamilies and domain superfolds. *Nature* **372**, 631–634.

Orr, H. A. & Coyne, J. A. 1992 The genetics of adaptations—a reassessment. *Am. Nat*. **140**, 725–742.

O'Neil, M-T. & Belote, J. M. 1992 Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics* **131**, 113–128.

Panganiban, G. (and 13 others) 1997 The origin and evolution of animal appendages. *Proc. Natn. Acad. Sci. USA* **94**, 5162–5166.

Pankratz, M. & Jäckle, H. 1993 Blastoderm segmentation. In *Development of* Drosophila melanogaster (ed. M. Bate & A. Martinez-Arias), pp. 467–516. New York: Cold Spring Harbor Laboratory.

Raff, R. A. 1996 *The shape of life: genes, development, and the evolution of animal form*, pp. 173 ff. University of Chicago Press.

Reuter, D., Schuh, R. & Jäckle, H. 1989 The homeotic gene *spalt* (*sal*) evolved during *Drosophila* speciation. *Proc. Natn. Acad. Sci. USA* **86**, 5483–5486.

Reuter, D., Kühnlein, R. P., Frommer, G., Barrio, R., Kofatos, F. C., Jäckle, H. & Schuh, R. 1996 Regulation, function and potential origin of the *Drosophila* gene *spalt adjacent*, which encodes a secreted protein expressed in the early embryo. *Chromosoma* **104**, 445–454.

Rice, W. R. & Holland, B. 1997 The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav. Ecol. Sociobiol.* **41**, 1–10.

Sander, K. 1983 The evolution of patterning mechanisms: gleanings from insect embryogenesis and spermatogenesis. In *Development and evolution: the sixth symposium of the British Society for Developmental Biology* (ed. B. C. Goodwin, N. Holder & C. C. Wylie), pp. 137–159. Cambridge University Press.

Schlötterer, C., Vogl, C. & Tautz, D. 1997 Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* **146**, 309–320.

Schmid, K. J. & Tautz, D. 1997 A screen for fast evolving genes from *Drosophila*. *Proc. Natn. Acad. Sci. USA* **94**, 9746–9750.

Slack, J. M. W., Holland, P. W. H. & Graham, C. F. 1993 The zootype and the phylotypic stage. *Nature* **361**, 490–492.

Sommer, R. J., Carta, L. K. & Sternberg, P. W. 1994 The evolution of cell lineage in nematodes. *Dev. Suppl.* 85–94.

Sordino, P. & Duboule, D. 1996 A molecular approach to the evolution of vertebrate paired appendages. *Trends Ecol. Evol.* **11**, 114–119.

St Johnston, D. & Nüsslein-Volhard, C. 1992 The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201–219.

Tautz, D. 1992 Redundancies, development and the flow of information. *BioEssays* **14**, 263–266.

Tautz, D. & Sommer, R. J. 1995 Evolution of segmentation genes in insects. *Trends Genet.* **11**, 23–27.

Tautz, D., Friedrich, M. & Schröder, R. 1994 Insect embryogenesis—what is ancestral and what is derived? *Dev. Suppl.* 193–199.

Taylor, M. F. J., She, Y. & Kreitman, M. E. 1995 A population genetic test of selection at the molecular level. *Science* **270**, 1497–1499.

Thackeray, J. R. & Kyriacou, C. P. 1990 Molecular evolution in the *Drosophila yakuba period* locus. *J. Molec. Evol.* **31**, 389–401.

Theißen, G. & Saedler, H. 1995 MADS-box genes in plant ontogeny and phylogeny: Haeckel's biogenetic law revisited. *Curr. Opin. Genet. Dev.* **5**, 628–639.

Tsaur, S. C. & Wu, C.-I. 1997 Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. *Molec. Biol. Evol.* **14**, 544–549.

Wilson, R. (and others) 1994 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368**, 32–38.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES