Review TRENDS in Ecology and Evolution Vol.20 No.2 February 2005

Candidate genes for behavioural ecology

Mark J. Fitzpatrick[1], Yehuda Ben-Shahar[2], Hans M. Smid[3], Louise E.M. Vet[3,4], Gene E. Robinson[5] and Marla B. Sokolowski[1]

[1]Department of Biology, University of Toronto at Mississauga, Mississauga, ON, Canada, L5L 1C6
[2]Howard Hughes Medical Institute, 500 EMRB, University of Iowa, College of Medicine, Iowa City, IA 52242, USA
[3]Laboratory of Entomology, Wageningen University, Binnenhaven 7, NL-6709 PD Wageningen, the Netherlands
[4]Netherlands Institute of Ecology (NIOO KNAW), Villa Vijverhof, Rijksstraatweg 6, Nieuwersluis, PO Box 1299, 3600 BG Maarssen, the Netherlands
[5]Department of Entomology & Neuroscience Program, University of Illinois at Urbana Champaign, 320 Morrill Hall, 505 S. Goodwin Ave, Urbana, IL 61801, USA

In spite of millions of years of evolutionary divergence, the conservation of gene function is common across distant lineages. As such, genes that are known to influence behaviour in one organism are likely to influence similar behaviours in other organisms. Recent studies of the evolution of behaviour and morphological adaptation support this notion. Thus, the candidate gene approach offers great potential to expand our understanding of behavioural ecology. Changes in the expression of candidate genes can reveal their contribution to behavioural variation and/or phenotypic plasticity. Knowledge of gene function also enables experimental manipulation of behaviour in the lab and in the field. The candidate gene approach provides an accessible and useful tool for generating insights about animals that are not typically associated with genetic experimentation.

Bridging the Gap

These are exciting times for interdisciplinary research, where the enquiries of biologists interested in mechanisms are converging with those of researchers interested in evolution and ecology [1–3]. Genes have become a major focus of this convergence because they provide mechanistic as well as evolutionary insights [3,4].

Classic examples of adaptation have recently been expanded on by the use of candidate genes. A 'candidate gene' is any gene that has been identified in one organism that is hypothesized to influence a similar phenotype in another organism. The adaptive radiation of beak morphology in Darwin's finches *Geospiza* spp. has been attributed to differences in the expression pattern (see Glossary) of the candidate gene *Bmp4* [5], based on its previously known effects on beak morphology in the chicken *Gallus gallus* [6,7]. Similarly, adaptive colour polymorphisms in the rock pocket mouse *Chaetodipus intermedius* [8] and reptiles [9], phenotypes that are reminiscent of Kettlewell's *Biston betularia* moths, arise from nucleotide changes in the candidate gene *Mc1r*, which was first known to affect pigmentation in the mouse *Mus musculus* [10]. These success stories highlight a role for candidate genes in our understanding of the molecular basis of evolutionary adaptation. They also support the idea that the genetic and molecular underpinnings of

Glossary

Alternative splicing: producing different RNAs from a single gene. During the maturation process of RNA, different portions are removed to produce the different variants.
Complimentary DNA (cDNA): a synthetic strand of DNA that is transcribed from mRNA. cDNA is constructed using a reverse transcriptase. It provides a stable form of mRNA that can be used for various purposes in the laboratory including cloning and the design of probes.
Degenerate primers: a pool of primers where some of the nucleotide positions contain more than one possible base. Degenerate primers are useful because they can accommodate for evolutionary divergence at these positions and are often used when cloning candidate genes.
Expressed sequence tag (EST): small unique section from the coding region of a gene (often identified using cDNA). Because this string of nucleotides is not found anywhere else, it can be used as a tag to map the genomic location of the gene and to reveal the entire sequence of the gene (introns and exons).
Expression pattern: the determination of where and when the gene of interest is found within the organism. This is often visualized by staining (e.g. for RNA by *in situ* hybridization and for protein by immunohistochemistry using antibodies).
Messenger RNA (mRNA): a complimentary copy of a string of DNA that is used to encode a polypeptide (thymine is replaced by uracil in RNA.)
Ortholog: two or more sequences that share similarity owing to common ancestry
PCR-based subtractive hybridization: used to amplify selectively genes that are differentially expressed in a pool of mRNA that was isolated from two biological samples.
Pleiotropy: when a single locus influences multiple phenotypes.
Quantitative real-time PCR (qRT-PCR): a PCR-based method of detecting differences in the levels of mRNA between two or more samples. Amounts of initial RNA present in the sample are estimated by the latency of the reaction to reach the exponential phase. Shorter times to reach the exponential phase indicate higher initial RNA concentrations.
Quantitative trait locus (QTL): a region in the genome that houses natural genetic variation in a quantitative trait. A single QTL can contain one gene or several genes that contribute to the trait.
Rapid amplification of cDNA ends (RACE): a procedure used to uncover the full sequence of cDNA (ultimately the sequence of mRNA). This method enables the 3′ and 5′ ends to be determined because the synthesis of cDNA from mRNA rarely yields fragments of full length.
RNA-mediated interference (RNAi): small fragments of double-stranded RNA whose sequence matches the transcribed sequence of a gene. This technique is used to decrease the expression of a gene by disabling the transcribed mRNA.

Corresponding author: Sokolowski, M.B. (msokolow@utm.utoronto.ca).
Available online 15 December 2004

 0169-5347/$ - see front matter doi:10.1016/j.tree.2004.11.017

natural variation in behaviour can be advanced by using a candidate gene approach (CGA) (Box 1, Figure 1).

Variation in quantitative traits provides the raw material used by natural selection in adaptation. How genetic variation in quantitative traits is maintained in natural populations and whether the same quantitative trait locus (QTL) affects variation within and between populations and species, are questions of great interest to evolutionary biologists. Knowledge of the genes underlying quantitative traits and how these genes function to influence phenotypic variation in natural populations enables us to begin to formulate comprehensive answers to these complex questions.

Recent evidence shows that the distribution of allelic effects of quantitative traits is exponential [11]. A few loci with large effects (major genes) influence most of the genetic variation and an increasingly large number of loci with increasingly smaller effects (minor genes) influence the remaining variation. The CGA described here relies on choosing genes that have major effects on the phenotype, such as those used in the morphological examples that we describe above. Numerous genes with large effects on behaviour have been identified by mutation [12], Mendelian analysis of behavioural variants [12], QTL mapping [13] and the identification of differences in RNA or protein expression between behavioural variants [14]. These are all candidate genes for natural variation in behaviour.

Until recently, analysing the genetic underpinnings of behaviour was restricted mainly to the 'genetic model organisms', prized for their genetic manipulability in the laboratory. These include, but are not restricted to, the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, and the mouse *M. musculus* (Table 1). Genes identified in these organisms can now serve as 'candidate genes' for similar behaviours in additional organisms.

Our purpose here is to discuss the potential of the CGA to broaden our understanding of behavioural ecology and evolution. Together with genomics [2] (Box 2) and QTL mapping [13], the CGA enables the identification of genes involved in ecologically relevant behaviours. It builds on the wealth of information available mainly from genetic model organisms to provide a means for studying the relationships between genes and behaviour in other organisms. Here, we review several examples that support the use of candidate genes in behavioural ecology.

Promising insights for candidate genes in behavioural ecology

Genes exert their actions in a dynamic fashion, interacting with many other genes in the genome (e.g. epistasis) and in response to complex interactions with the environment [15]. Candidate genes can influence behaviour in several ways. Some genes exhibit allelic variation that affects behaviour [16,17]. Others do not vary genetically, but change their expression within an individual over time, resulting in changes in behaviour (e.g. plasticity) [18]. Examples are discussed below (Figure 2); however, this list is not exhaustive. Examples were chosen that: (i) illustrate different routes taken using the CGA; and (ii) are at varying levels of progression through the CGA. Some examples use the CGA based on the similarity of behaviours in unrelated organisms (e.g. the *for* gene in the honey bee *Apis mellifera* and fruit fly). Others begin with observations of protein expression differences between individuals that differ in behaviour (e.g. expression of the vasopressin receptor in voles).

Box 1. The candidate gene approach

Candidate genes are 'nominated' by knowledge of how they influence similar behaviours in other organisms. These genes can be sourced using the databases listed in Table 1, main text. Once chosen, the candidate is cloned, sequenced and measurements of expression (mRNA and protein encoded by the gene) are made in animals that differ in their behaviour. Gene expression can be artificially manipulated and the observed concomitant changes in behaviour can be used to link functionally the gene to the behaviour in a causal rather than correlative manner.

Isolation and sequencing is accomplished using PCR. The selection of the PCR primers is the most crucial step because evolutionary divergence is expected to have occurred between the organism under investigation and the genetic model organism. Conserved regions of protein sequence can be identified by comparing multiple sequence alignments constructed from available orthologs of the candidate gene. As more orthologs become available, the ease of determining these conserved regions increases. Conserved regions (e.g. a string of amino acids found in all orthologs) are identified from an alignment of the protein sequence of the gene of interest. Primers are then designed by reverse transcription because it is the mRNA that will be amplified by PCR for sequencing. A cocktail of primers containing all possible nucleotide combinations to encode the conserved region is obtained. These are termed 'degenerate primers' because, owing to codon degeneracy, most amino acids can be encoded by more than one nucleotide triplet [67,68]. Thus, degenerate primers account for any evolutionary divergence in a conserved region. However, a higher degree of degeneracy often leads to greater difficulty when cloning the candidate gene. Degenerate primer design is important because suboptimal primers can result in the amplification of non-specific sequences. Rose *et al.* [69] describe strategies to reduce the level of degeneracy and name various downloadable software programs that can help in the design of degenerate primers.

The fragment is then sequenced and compared to orthologs from other organisms to confirm that it is indeed a fragment of the candidate gene. New primers are developed from this fragment, and the entire coding region of the gene can be cloned from the cDNA using methods such as RACE (rapid amplification of cDNA ends) [70]. Once the first fragment is cloned, northern blots can be used to determine the number and size of the RNA transcript(s) of the gene. If only one transcript is found then expression analyses using qRT-PCR [71] can proceed without having to clone the entire gene. If multiple, alternatively spliced transcripts are found then transcript-specific PCR primers should be used to quantify the abundance of each RNA transcript, otherwise the expression data obtained might be confounded. For example, if only one transcript is important for the behaviour but all transcripts are measured simultaneously then the wrong conclusions might be drawn.

The speed and success with which the first fragment of the gene of interest is found is not predictable. Thus, it is useful to begin with two to three candidate genes if possible. This is especially important if the candidate gene belongs to a multigene family where the involvement of each can be examined with respect to its influence on behaviour. New candidate genes can also be identified by way of genomic analyses using cDNA microarrays (Box 2). These microarrays can also be used in the cases where convergent behaviours have evolved different genetic bases.

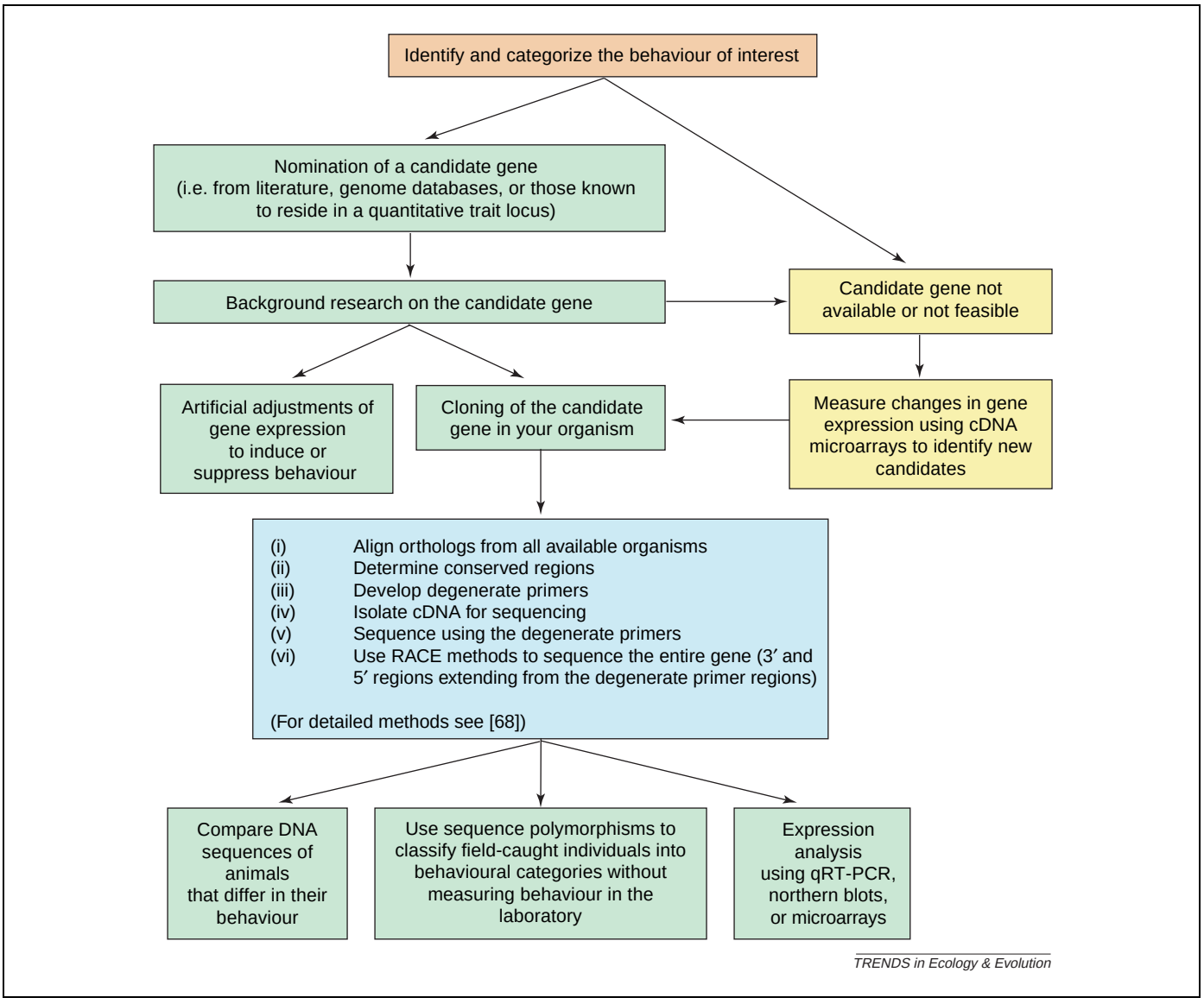

Figure 1. Flowchart depicting the candidate gene approach for behavioural ecology. Genes previously characterized in organisms such as the fruit fly *Drosophila melanogaster*, nematode *Caenorhabditis elegans*, and mouse *Mus musculus*, are, on the basis of their role in these organisms, elected as candidate genes for similar roles in additional organisms.

Table 1. Examples of current genome databases

Taxon	Database	URL
Chicken *Gallus gallus*	ChickBASE	http://www.genome.iastate.edu/chickmap/
	US Poultry Genome Project	http://poultry.mph.msu.edu/
Fruit fly *Drosophila melanogaster*	FlyBase	http://www.flybase.net
	Berkeley *Drosophila* Genome Project	http://www.fruitfly.org
Honey bee *Apis mellifera*	Honey Bee Genome Project	http://www.hgsc.bcm.tmc.edu/projects/honeybee/
House mouse *Mus musculus*	Mouse Genome Resources	http://www.ncbi.nlm.nih.gov/genome/guide/mouse
	Mouse Genome Informatics	http://www.informatics.jax.org
Human *Homo sapiens*	GDB Human Genome Database	http://gdbwww.gdb.org/gdb/
	National Human Genome Research Institute	http://www.genome.gov/
Japanese puffer fish *Fugu rubripes*	*Fugu* Genomics Project	http://fugu.hgmp.mrc.ac.uk/
Mosquito *Aedes aegypti*	Mosquito Genomics WWW	http://mosquito.colostate.edu/tikiwiki/
Anopheles gambiae	Mosquito Genome Browser	http://www.ensembl.org/Anopheles_gambiae/
Nematode worm	WormBase	http://www.wormbase.org
Caenorhabditis elegans		
	Caenorhabditis elegans WWW	http://elegans.swmed.edu
Rat *Rattus norvegicus*	Ratmap	http://ratmap.gen.gu.se/
	Rat Genome Database	http://rgd.mcw.edu/
Silkworm moth *Bombyx mori*	SilkBase	http://www.ab.a.u-tokyo.ac.jp/silkbase/
Zebrafish *Danio rerio*	Zebrafish Information Network	http://zfin.org

Box 2. Genomics: powerful approaches to generate new candidate genes

The behaviours of interest to behavioural ecologists are more diverse and often more complex than those routinely studied in genetic model organisms. Sometimes candidate genes are not readily available and it is not always possible to make a reasonable analogy between the behaviour of two or more different organisms. In these cases, one can look for changes in the expression of suites of genes that correlate with changes in behaviour. Genes are identified that have different levels of expression in individuals performing different behaviours or different forms of the behaviour of interest; typically the differences are in the brain or localized to a brain region.

cDNA microarrays are the method of choice for uncovering new candidates for further study without *a priori* knowledge of specific genes. cDNA microarrays are best used after an expressed sequence tag (EST) project, because the array can then comprise partially sequenced, characterized and annotated genes to ensure high-quality results. A single microarray contains thousands of spots of cDNA sequences that are printed on a standard glass slide. Each spot is a fragment of a single gene. When RNA that has been isolated from one individual (or a pool of individuals that behave in a similar manner) is applied to this slide, individual RNAs bind to their corresponding cDNA and the amount of RNA expressed is visualized by the intensity of a fluorescent dye.

The EST-microarray approach is beginning to be applied to organisms of interest to behavioural ecologists [21,72–74]. However, although sequencing costs continue to decrease, large-scale sequencing projects are still expensive. An economical approach involves constructing a microarray before sequencing any genes; spots that show differential expression are then selected for sequencing. This approach does not provide the experimenter with the ability to create an array of high-quality sequences, but it is more economical. Sometimes, arrays can be made that are already enriched for differentially expressed genes [75], using PCR-based subtractive hybridization [76]. In any of these approaches, problems associated with limited amounts of RNA in the brains of small animals can be overcome by amplification of RNA, starting from as little as nanogram quantities [77]. The genomic approaches outlined here offer a useful set of tools with which to select candidate genes in a manner that is roughly comparable to 'forward genetic' screens that are commonly used in the genetic model organisms, even without previous knowledge of the genetics and/or molecular biology of those organisms.

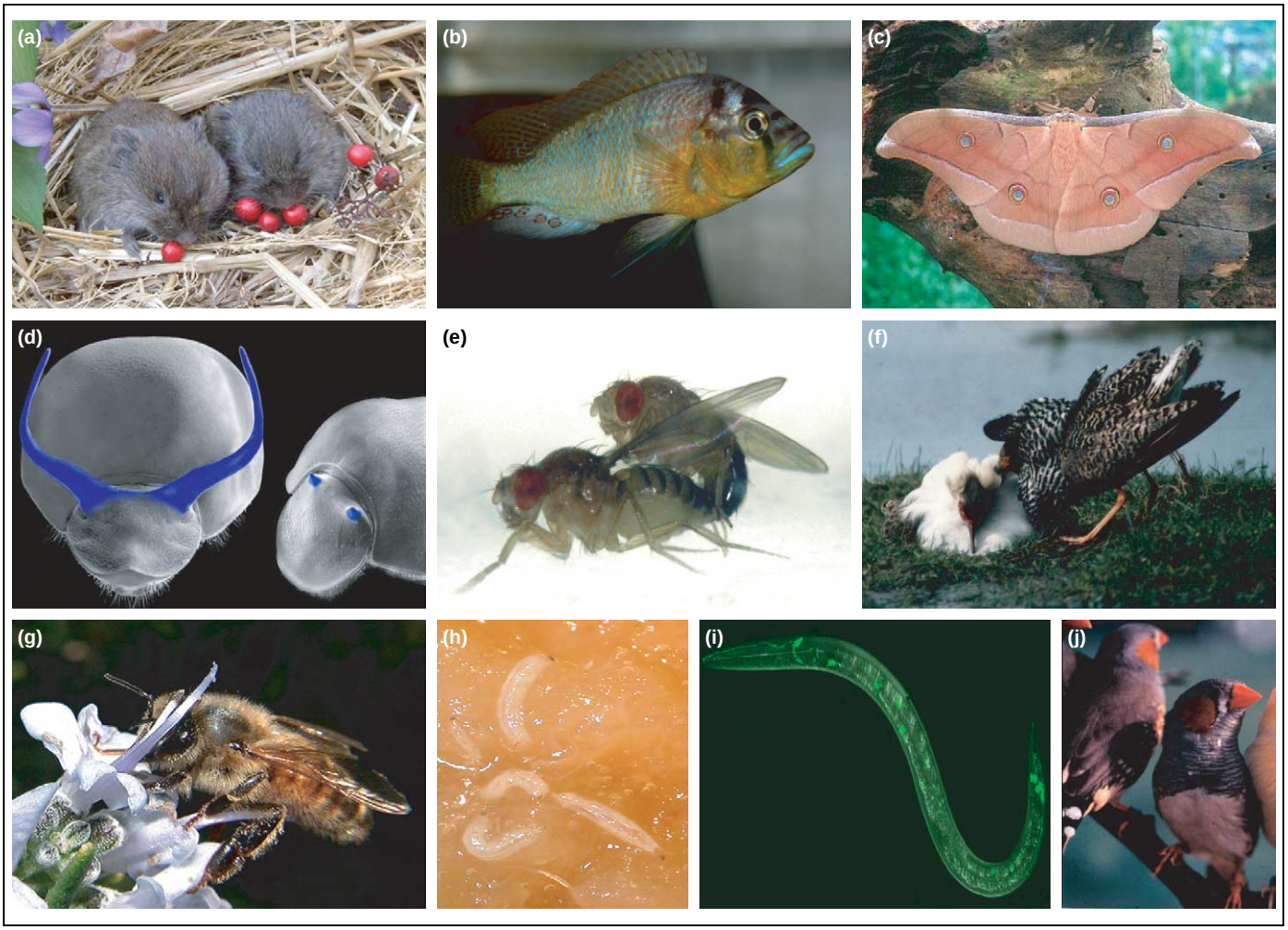

Figure 2. Promising support for the candidate gene approach in behavioural ecology. **(a)** Both vasopressin and its receptor influence partner affiliation in voles *Microtus* spp. **(b)** Gonadotropin-releasing hormone (GnRH) influences dominance behaviour and territoriality in Burton's mouthbreeder cichlid *Astatotilapia (Haplochromis) burtoni*. **(c)** Transferring *per* from the giant silkmoth *Antheraea pernyi* restores normal circadian function in the fruit fly. **(d)** Juvenile hormone influences horn size and alternative mating strategies in the beetle *Onthophagus taurus*. **(e)** *nonA* influences species-specific components of the courtship song in fruit flies *Drosophila* spp. **(f)** Testosterone influences plumage colouration and alternative mating strategies in the ruff *Philomachus pugnax*. Orthologs of *for* influence food-dependent behaviours in **(g)** the honey bee *Apis mellifera*, **(h)** the fruit fly *Drosophila melanogaster* and **(i)** the nematode *Caenorhabditis elegans*. **(j)** *FoxP2* influences song learning in the zebra finch *Taeniopygia guttata*. Photo credits: (a) Miranda Lim, Elizabeth Hammock, and Larry Young, (b) Christian Landry and Hans Hofmann, (c) Thibaud Decaens, (d) Doug Emlen, (e, h) Mark Fitzpatrick, (f) Oene Moedt courtesy of Dov Lank, (g) Alex Wild, (i) Benny Cheung courtesy of Mario de Bono, (j) Erich Jarvis.

Foraging

In spite of the observed variation in foraging behaviours among species [19], conservation in the genetic underpinnings of such behaviours has received attention only recently. For example, Osborne *et al.* [16] found that the cGMP-dependent protein kinase (PKG) encoded by the *for* gene influences foraging behaviour in larval and adult *D. melanogaster*. Rover larvae, which have long foraging trails and a higher propensity to leave a food patch, have higher PKG activity levels than do sitters, which have relatively short foraging trails and a low propensity to leave a food patch [12]. Naturally occurring allelic variants (rover, for^{R} and sitter, for^{s}) are found in orchard populations of *D. melanogaster*. The rover allele is phenotypically dominant to the sitter and they occur naturally at 70:30 phenotypic ratios. Sokolowski *et al.* [20] showed that the allele frequencies respond to density-dependent selection; larvae from crowded environments are more rover-like (selection for for^{R}), whereas larvae from uncrowded environments are more sitter-like (selection for for^{s}).

Using *for* as a candidate gene, Ben-Shahar *et al.* [18] found that the regulation of division of labour in the honey bee is influenced by *Amfor*, the bee ortholog of the fruit fly *for* gene (87% similarity). Adult worker honey bees spend the first 2–3 weeks working inside the hive at tasks such as caring for the brood ('nurses') and then switch to foraging, leaving the hive in search of pollen and nectar. The authors chose *for* as a candidate because: (i) it shows natural variation in *D. melanogaster*; (ii) the allelic variants show differences in gene expression, making it easier to envisage a role for this gene in the behavioural changes that occur within the lifetime of the worker bee; and (iii) a loose but provocative analogy was made between *D. melanogaster* and honey bees: forager bees are similar to rover flies and nurse bees are similar to sitter flies. Forager bees have higher expression of *Amfor* RNA in their brain than do nurse bees [18]. Foragers also have higher PKG enzyme activity than do nurses. An activator of PKG (8-Br-cGMP, Sigma, B-1381) was used to extend this correlation to causation. When treated with this PKG activator, young bees were induced to forage precociously [18]. Although many genes are involved in the behavioural transition from working in the hive to foraging outside the hive [21], manipulating the expression of *Amfor* is sufficient to cause the striking behavioural transition from nurse to forager. This use of the CGA provides the first evidence for a gene involved in the regulation of division of labour in honey bee colonies.

The *for* gene is ripe for consideration as a candidate gene for other animals that vary in food-related behaviours. Using *C. elegans*, Fujiwara *et al.* [22] found that changes in the ortholog of *for*, called *egl-4*, influences food-dependent locomotion. The differences in PKG expression between these 'roamer' and 'dweller' nematodes, along with rover and sitter fruit flies and forager and nurse bees, probably result in changes in the expression of other genes downstream from PKG; however, in all cases, *for* has a significant effect on observed behavioural differences. Using a phylogeny constructed from all available *for*-PKG protein sequences, Fitzpatrick and Sokolowski [23] developed several hypotheses for an evolutionarily conserved association between PKG and food-related behaviours. In spite of the broad taxonomic breadth (e.g. nematode–human), several of the hypotheses suggest a widespread, conserved association between PKG and behaviour. This encourages the further use of *for* as a candidate gene for food-related behaviours.

Learning and memory

Learning evolves in laboratory populations [24] and retaining the ability to learn might have important evolutionary and ecological tradeoffs [25,26]. The cellular and genetic mechanisms underlying learning and memory are remarkably conserved [27–30].

Songs used by oscine songbirds such as the zebra finch *Taeniopygia guttata* are largely learned [31], as is language acquisition in humans [32]. Mutations in the gene *FoxP2*, a member of the forkhead box transcription factor family, are associated with severe abnormalities in human speech and language [33]. Haesler *et al.* [34] report a higher level of *FoxP2* expression in the basal ganglia vocal nucleus of vocal-learning birds during the crucial developmental stages when song is acquired. This basal ganglia nucleus is required for song acquisition at this stage. High expression of *FoxP2* is not found in non-vocal basal ganglia areas of vocal learning or non-vocal-learning birds. These results, although still correlative, suggest that *FoxP2* is a candidate for further studies of song learning.

Social interactions

Candidate genes that arose from knowledge of endocrine and neuropeptide function have provided important insights into the evolution of social interactions, such as promiscuity and social status. For example, vole species are naturally either monogamous or polygamous. The prairie vole *Microtus ochrogaster* is monogamous and exhibits long-term pair bonding and biparental care. Conversely, male montane voles *M. montanus* provide parental care but are promiscuous. Because many aspects of male vertebrate reproduction and parental care are mediated by the neuropeptide vasopressin, Insel and Young [14] hypothesized that differences in some aspect of vasopressin signalling might underlie the species differences in mating habits. A key difference was found in the distribution of vasopressin V1a receptors in the male vole brain. Relative to polygamous male montane voles, male prairie voles have a higher density of V1a receptors in the ventral pallidal area, which is part of the brain that responds to various types of appetitive reward [14]. Similar differences in the spatial distribution of V1a receptors are seen in other species of voles, and are also correlated with monogamy and polygamy. Young *et al.* [35] found that monogamous voles typically have a microsatellite in the regulatory region of the V1a receptor gene (*V1aR*) that is absent in polygamous species. Their experiments with transgenic mice and prairie voles demonstrated that differences in the distribution of V1a receptors in the brain causally affect mating habits [35,36]. Artificially expressing V1a receptors in the ventral forebrain of a promiscuous vole species was sufficient to increase affiliative behaviour of males to

their female partners [37]. This work shows that interspecific differences in the mating patterns of voles stems from allelic variation in the regulatory region of the *V1aR*, resulting in differences in the patterns of gene expression.

Another example comes from Burton's mouthbreeder cichlid *Astatotilapia (Haplochromis) burtoni*, which has male hierarchies. Dominant males are big, aggressively territorial, brightly coloured, have high levels of circulating testosterone and enjoy significant reproductive success, whereas subordinate males lack all of these attributes (reviewed in [38]). Hofmann and Fernald [38] hypothesized that changes in gonadotropin-releasing hormone (GnRH) expression influence dominance behaviour in *A. burtoni*, because this neuropeptide is known to coordinate reproduction in other vertebrates. Indeed, GnRH expression levels in the preoptic region of the brain are tightly correlated with changes in social status; subordinate *A. burtoni* have low levels whereas dominants have higher levels [39]. It is not yet known whether allelic variation in the gene encoding GnRH contributes to the differences in the behaviour of dominant versus subordinate males.

Alternative mating strategies

Distinct mating strategies exist in males of many species, including the marine isopod *Paracerceis sculpta*, bluegill sunfish *Lepomis macrochirus*, and side-blotched lizard *Uta stansburiana* [40,41]. At present, candidate genes for alternative mating strategies have not been identified, although several recent studies suggest that genes encoding hormones are involved.

For example, territorial male ruffs *Philomachus pugnax* have dark plumage whereas non-territorial satellite males have light plumage [42]. Lank *et al.* [43] found that testosterone implantation in females induces the expression of normally male-limited behaviour and plumage alternatives. The resultant 'male' behaviour and plumage in females is dependent on the pedigree of the female. A single locus with two alleles is believed to underlie this polymorphism [42,43], although the gene has not yet been identified.

Large territorial males of the beetle *Onthophagus taurus* use their horns in fights for mating access to females [44,45], whereas small hornless males sneak copulations. Horn development requires a threshold body size, the absence of the steroid ecdysteroid, and the presence of juvenile hormone (JH) during crucial periods of development [46,47]. Small juvenile males, which generally develop into hornless adults, grow horns when treated with a JH analogue. JH is known to have pleiotropic effects on both morphology and sexual behaviour [48], which is a common correlation in species with alternative mating strategies. Thus, genes involved in the JH pathway would make excellent candidates for investigating alternative mating strategies in dung beetles and perhaps other insects.

Hormones such as JH and testosterone have a crucial role in the development of alternative mating strategies in some species; however, whether this is a general phenomenon remains to be determined. Further studies of the evolution of alternative mating studies could investigate whether variation in the levels of JH and testosterone naturally correlate with different mating strategies and, if so, how the genes that underlie these behavioural polymorphisms influence the structure, function and expression of these hormones.

Applications of the CGA

Manipulating traits

Once a gene that is associated with behavioural variation has been identified, its expression can be manipulated in transgenic organisms (by overexpression, knocking out or altering gene expression) or by pharmacological interventions as discussed above. Although transgenic technologies are currently available for only a few organisms, results of experiments using transgenic animals and RNAi are instructive.

Chapman *et al.* [49] used RNA-mediated interference (RNAi) to further understand the role of sex peptides (transferred to females from males during copulation) in the reduction of mating receptivity and the increase of egg production in female *D. melanogaster*. Using RNAi, males were generated that lacked sex peptides. Females mated with these males have increased mating receptivity and reduced egg production relative to females that are mated with control males (which transferred sex peptides). This reduction in female receptivity post-copulation, once thought to result from the sperm [50], is thus shown to result from an effect of the sex peptides.

To demonstrate experimentally the connection between a morphological trait in males and mating preferences in females, the male trait is often physically manipulated (e.g. preferences for extreme tail length in the long-tailed widowbird *Euplectes progne* [51]). This becomes difficult when the trait to be manipulated is behaviour. However, with knowledge of the underlying genes, behaviours can be altered with the various techniques listed above. The experimental manipulation of behaviour might be particularly beneficial for studies of sexual selection: consider this hypothetical example.

Females preferentially mate with males having more chirps in their courtship song. Offspring fitness is positively correlated with chirp number. It is unknown, however, whether males with more chirps transfer better sperm (i.e. genetic quality [52]) and/or whether females invest more in the offspring sired by these males (i.e. differential allocation). The CGA followed by manipulation of the trait is a suitable method with which to disentangle these hypotheses. Candidate genes influencing characteristics of the courtship song can be identified and treatments, such as pharmacological agents or RNAi, which can alter the expression of the gene, can be used to influence chirp number in males. Males that would generally have fewer chirps can be induced to have more chirps. If the differential allocation hypothesis holds true, offspring from a treated male should have higher fitness than should the offspring sired from his untreated full-sib brother (a control for genetic quality). Conversely, males having fewer than normal chirps can be also generated, whereby the opposite results would be expected (higher offspring fitness for the untreated brother).

Comparative approaches

Much can be learned about the mechanism and evolution of behaviour by comparing divergent populations, related species and phylogenetically similar groups of species. For example, comparative studies highlight the evolutionarily conserved role of a gene in behaviour and its potential role in evolutionary processes, such as speciation. The use of transgenics (the transfer of genes within and between species) and phylogenetic comparisons might clarify the evolutionary connections between genes and behaviour.

The *period* (*per*) gene was first implicated in circadian rhythmicity of *D. melanogaster* [53]. Orthologs of *per* have been identified in numerous additional organisms (reviewed in [54–57]), including the giant silkmoth *Antheraea pernyi* [58]. To demonstrate evolutionary conservation for the circadian function of *per*, Levine *et al.* [59] restored normal circadian rhythmicity in *per* mutant *D. melanogaster* by inserting a functional copy from the silkmoth. These results suggest that *per* function is maintained between evolutionarily distinct orders of insects: the Lepidoptera and Diptera are thought to have diverged between 330 and 350 million years ago [60].

Behavioural mechanisms causing reproductive isolation are believed to contribute to speciation [61]. The gene *no-on-transient A* (*nonA*) has pleiotropic effects on courtship song and vision in *D. melanogaster*. Campesan *et al.* [62] inserted the *nonA* ortholog from a related fruit fly, *D. virilis*, into a *D. melanogaster* strain lacking a functional copy of *nonA* (i.e. knockout). This transgenic strain (*D. melanogaster* expressing *nonA* from *D. virilis*), in addition to having restored vision, included components that were reminiscent of *D. virilis* in its courtship song. These results suggest that *nonA* (probably in concert with other genes) conveys species-specific information in the courtship song of fruit flies. Similar approaches should shed light on the role of courtship in the divergence of populations and their subsequent speciation.

Phylogenies can be used to understand the evolutionary connections between genes and behaviour [23]. Once a gene has been identified, cloned and sequenced using CGA, questions of the role(s) that the gene had in behavioural evolution and whether different species use different genes for similar behaviours can be facilitated using these phylogenetic comparisons.

Predicting behaviour using markers

In cases where allelic differences in the candidate gene directly influence behaviour, allele-specific sequence differences (e.g. single nucleotide polymorphisms) that are causally linked to the behaviour can be used as markers. Such an approach will be especially beneficial when the behavioural classes are morphologically indistinguishable (e.g. rover and sitter fruit flies). This accurate genotypic reflection of behaviour allows for: (i) predicting behaviour of field-caught individuals without having to measure their behaviour; (ii) predicting adult behaviour if juveniles are collected from the field; and (iii) monitoring the frequency and density of behavioural classes in time and space. By reducing the time and effort required to assess behaviour, and circumventing any issues of day effects on behavioural performance, this method can expedite further studies of the ecological and evolutionary significance of natural variation in behaviour.

Some challenges to consider when using the CGA

Genes that underlie complex quantitative traits such as behaviour provide us with several challenges. First, the relationship between the genetic variant and behaviour is not deterministic. Having a pair of alleles gives the animal a predisposition to behave a certain way, in a certain environment. When animals with one pair of alleles respond differently to a range of environments than do animals with an alternate pair of alleles, they are said to have different norms of reaction. A case of gene-by-environment (GxE) interaction occurs when the environmental sensitivity curves are non-additive [11,13]. Differences in gene expression that are necessary to elicit a change in behaviour might only occur in particular environments and in animals of a particular sex, age or developmental state. Thus, the success of the CGA might depend on the subjects chosen for gene expression measurements.

Second, most genes have pleiotropic effects [12,63] and these might or might not be related to the behaviour under study. Related pleiotropic effects would include when a gene contributes to the development of legs and locomotory behaviour. Non-related pleiotropic effects would include when a gene is important for vision and for courtship song (see description of *nonA* above). The consequences of pleiotropy for the CGA arise when gene expression levels are manipulated using pharmacological or genetic interventions. When the gene has an important role in development, then feeding young animals an inhibitor or activator of the protein during development might kill them or render them too sick to exhibit their 'normal' behaviour. Nevertheless, adjusting the dose and timing of the application might circumvent this issue.

Third, genes interact regularly with other genes to affect the phenotype [11]. Epistatic interactions on behaviour are evident when crossing mutants, natural genetic variants or QTLs into different genetic backgrounds. Different backgrounds commonly have different effects on the behaviour [13,64] suggesting that particular genetic backgrounds interact with the gene of interest to affect the phenotype. Epistasis might have important consequences for a CGA because its expression is often studied in animals sampled from natural populations where the genetic background is heterogeneous and not controlled. If the genes that are epistatic with the candidate gene segregate at significant frequencies in the population, it will be difficult to demonstrate consistent effects of the candidate gene without further genetic analyses involving crosses. However, many studies in development and behaviour have successfully used the CGA, suggesting that epistasis might not be as prominent for genes with major effects.

Closing remarks

Recent use of candidate genes has expanded our understanding of evolutionary adaptation [5,8,9] and behaviour [18,34–36] in organisms that are, for the most part, not

known for extensive genetic experimentation. We anticipate numerous additional success stories such as these in the near future. The CGA, in concert with genomics [2] and QTL mapping [13], provides an accessible and useful tool for providing a detailed understanding of the role of genes in behavioural ecology.

Efforts are being made to develop 'universal microarrays' [65], which will enable extensive gene expression profiling for all species. There is also great interest in developing innovative new methods to shrink sequencing costs; the US National Institutes of Health (http://www.nih.gov) has released a Request for Applications for the development of 're-sequencing technologies' for the '$1000 human genome' [66]. If achieved, costs for *de novo* sequencing will decrease, enabling more species to be studied with new genome-enabled resources.

How the environment has shaped the evolution of animal behaviour is a primary focus of behavioural ecology. Recent advances in genome analysis and genetic cloning make it timely to integrate these advances into studies of behavioural ecology and evolution by use of the CGA. Comparative methods can elucidate whether and how the gene influences species-specific behavioural differences and when the association between a gene and a behaviour evolved. This integrated view will surely broaden our understanding of the ecology and evolution of behaviour.

Acknowledgements

We thank H.A. Hofmann, E.D. Jarvis, K.A. Judge, C.D. Kelly, D.B. Lank, J.D. Levine, F. Mery, T.F.C. Mackay, F.H. Rodd and L. Rowe for helpful comments and discussions. We also thank two anonymous reviewers for their insightful suggestions. Thanks to the following people for permission to use their images in Figure 2: B. Cheung, M. de Bono, T. Decaens, D.J. Emlen, E.A.D. Hammock, H.A. Hofmann, E.D. Jarvis, C. Landry, D.B. Lank, M.M. Lim, O. Moedt, A. Wild, L.J. Young. Research cited in this paper that was conducted by the authors has been supported by grants from the Canada Research Chair Program, NSERC and CIHR (to M.B.S.); NIH, NSF, and Burroughs Wellcome Fund (to G.E.R.); and the Council for Earth and Life Sciences (A.L.W.) and NWO (to L.E.M.V.).

References

1 Stearns, S.C. and Magwene, P. (2003) The naturalist in a world of genomics. *Am. Nat.* 161, 171–180
2 Thomas, M.A. and Klaper, R. (2004) Genomics for the ecological toolbox. *Trends Ecol. Evol.* 19, 439–445
3 Robinson, G.E. (1999) Integrative animal behaviour and sociogenomics. *Trends Ecol. Evol.* 14, 202–205
4 Boake, C.R.B. *et al*. (2002) Genetic tools for studying adaptation and the evolution of behaviour. *Am. Nat.* 160, S143–S159
5 Abzhanov, A. *et al*. (2004) *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* 305, 1462–1465
6 Hu, D. *et al*. (2003) A zone of frontonasal ectoderm regulates patterning and growth in the face. *Development* 130, 1749–1758
7 Wu, P. *et al*. (2004) Molecular shaping of the beak. *Science* 305, 1465–1466
8 Nachman, M.W. *et al*. (2003) The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5268–5273
9 Rosenblum, E.B. *et al*. (2004) Adaptive reptile color variation and the evolution of the *Mc1r* gene. *Evolution* 58, 1794–1808
10 Robbins, L.S. *et al*. (1993) Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter receptor function. *Cell* 72, 827–834
11 Mackay, T.F.C. (2001) Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* 2, 11–20
12 Sokolowski, M.B. (2001) *Drosophila*: genetics meets behaviour. *Nat. Rev. Genet.* 2, 879–890
13 Anholt, R.R.H. and Mackay, T.F.C. (2004) Quantitative genetic analyses of complex behaviours in *Drosophila*. *Nat. Rev. Genet.* 5, 838–849
14 Insel, T.R. and Young, L.J. (2000) Neuropeptides and the evolution of social behaviour. *Curr. Opin. Neurobiol.* 10, 784–789
15 Robinson, G.E. (2004) Beyond nature and nurture. *Science* 304, 397–399
16 Osborne, K.A. *et al*. (1997) Natural behaviour polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* 277, 834–836
17 de Bono, M. and Bargmann, C.I. (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* 94, 679–689
18 Ben-Shahar, Y. *et al*. (2002) Influence of gene action across different time scales on behaviour. *Science* 296, 741–744
19 Kamil, A.C. *et al*., eds (1986) *Foraging Behaviour*, Plenum Press
20 Sokolowski, M.B. *et al*. (1997) Evolution of foraging behaviour in *Drosophila* by density-dependent selection. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7373–7377
21 Whitfield, C.W. *et al*. (2003) Gene expression profiles in the brain predict behaviour in individual honey bees. *Science* 302, 296–299
22 Fujiwara, M. *et al*. (2002) Regulation of body size and behavioural state of *C. elegans* by sensory perception and the EGL-4 cGMP-dependent protein kinase. *Neuron* 36, 1091–1102
23 Fitzpatrick, M.J. and Sokolowski, M.B. (2004) In search of food: exploring the evolutionary link between cGMP-dependent protein kinase (PKG) and behaviour. *Integr. Comp. Biol.* 44, 28–36
24 Mery, F. and Kawecki, T.J. (2002) Experimental evolution of learning ability in fruit flies. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14274–14279
25 Mery, F. and Kawecki, T.J. (2003) A fitness cost of learning ability in *Drosophila melanogaster*. *Proc. R. Soc. Lond. Ser. B* 270, 2465–2469
26 Vet, L.E.M. *et al*. (1995) Parasitoid foraging and learning. In *Chemical Ecology of Insects* (Cardé, R.T. and Bell, W.J., eds), pp. 65–101, Chapman & Hall
27 Alberini, C.M. (1999) Genes to remember. *J. Exp. Biol.* 202, 2887–2891
28 Silva, A.J. *et al*. (1998) CREB and memory. *Annu. Rev. Neurosci.* 21, 127–148
29 Tully, T. *et al*. (2003) Targeting the CREB pathway for memory enhancers. *Nat. Rev. Drug Discov.* 2, 267–277
30 Roman, G. and Davis, R.L. (2001) Molecular biology and anatomy of *Drosophila* olfactory associative learning. *Bioessays* 23, 571–581
31 Marler, P. (1990) Song learning: the interface between behaviour and neuroethology. *Philos. Trans. R. Soc. Lond. Ser. B* 329, 109–114
32 Doupe, A.J. and Kuhl, P.K. (1999) Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631
33 Lai, C.S.L. *et al*. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523
34 Haesler, S. *et al*. (2004) *FoxP2* expression in avian vocal learners and non-learners. *J. Neurosci.* 24, 3164–3175
35 Young, L.J. *et al*. (1999) Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* 400, 766–768
36 Pitkow, L.J. *et al*. (2001) Facilitation of affiliation and pair-bond formation by vasopressin receptor gene transfer into the ventral forebrain of a monogamous vole. *J. Neurosci.* 21, 7392–7396
37 Lim, M.M. *et al*. (2004) Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature* 429, 754–757
38 Hofmann, H.A. and Fernald, R.D. (2000) Social status controls somatostatin neuron size and growth. *J. Neurosci.* 20, 4740–4744
39 White, S.A. *et al*. (2002) Social regulation of gonatotropin-releasing hormone. *J. Exp. Biol.* 205, 2567–2581
40 Gross, M.R. (1996) Alternative reproductive strategies and tactics: diversity within sexes. *Trends Ecol. Evol.* 11, 92–98
41 Shuster, S.M. and Wade, M.J. (2003) *Mating Systems and Strategies*, Princeton University Press
42 Lank, D.B. *et al*. (1995) Genetic polymorphism for alternative mating behaviour in lekking male ruff *Philomachus pugnax4*. *Nature* 378, 59–62
43 Lank, D.B. *et al*. (1999) Testosterone-induced male traits in female ruffs (*Philomachus pugnax*): autosomal inheritance and gender differentiation. *Proc. R. Soc. Lond. Ser. B* 266, 2323–2330
44 Emlen, D.J. (1997) Alternative reproductive tactics and male-dimorphism in the horned beetle *Onthophagus taurus* (Coleoptera: Scarabaeidae). *Behav. Ecol. Sociobiol.* 41, 335–341

45 Moczek, A.P. and Emlen, D.J. (2000) Male horn dimorphisms in the scarab beetle *Onthophagus taurus*: do alternative reproductive tactics favour alternative phenotypes? *Anim. Behav.* 59, 459–466
46 Emlen, D.J. and Nijhout, H.F. (1999) Hormonal control of male horn length dimorphism in *Onthophagus taurus* (Coleoptera: Scarabaeidae). *J. Insect Physiol.* 45, 45–53
47 Emlen, D.J. and Nijhout, H.F. (2001) Hormonal control of male horn length dimorphism in *Onthophagus taurus* (Coleoptera: Scarabaeidae): a second critical period of sensitivity to juvenile hormone. *J. Insect Physiol.* 47, 1045–1054
48 Ringo, J.M. (1996) Sexual receptivity in insects. *Annu. Rev. Entomol.* 41, 473–494
49 Chapman, T. *et al*. (2003) The sex peptide of *Drosophila melanogaster*: female post-mating responses analysed by using RNA interference. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9923–9928
50 Manning, A. (1962) A sperm factor affecting the receptivity of *Drosophila melanogaster* females. *Nature* 194, 252–253
51 Andersson, M. (1982) Female choice selects for extreme tail length in a widowbird. *Nature* 299, 818–820
52 Hunt, J. *et al*. (2004) What is genetic quality? *Trends Ecol. Evol.* 19, 329–333
53 Konopka, R.J. and Benzer, S. (1971) Clock mutants of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 68, 2112–2116
54 Wager-Smith, K. and Kay, S.A. (2000) Circadian rhythm genetics: from flies to mice to humans. *Nat. Genet.* 26, 23–27
55 Albrecht, U. (2002) Regulation of mammalian circadian clock genes. *J. Appl. Physiol.* 92, 1348–1355
56 Klarsfeld, A. *et al*. (2003) Circadian rhythms of locomotor activity in *Drosophila*. *Behav. Process.* 64, 161–175
57 Stanewsky, R. (2003) Genetic analysis of the circadian system in *Drosophila melanogaster* and mammals. *J. Neurobiol.* 54, 111–147
58 Reppert, S.M. *et al*. (1994) Cloning of a structural and functional homolog of the circadian clock gene *period* from the giant silkmoth *Antheraea pernyi*. *Neuron* 13, 1167–1176
59 Levine, J.D. *et al*. (1995) PERIOD protein from the giant silkmoth *Antheraea pernyi* functions as a circadian clock element in *Drosophila melanogaster*. *Neuron* 15, 147–157
60 Gaunt, M.W. and Miles, M.A. (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* 19, 748–761
61 Panhuis, T.M. *et al*. (2001) Sexual selection and speciation. *Trends Ecol. Evol.* 16, 364–371
62 Campesan, S. *et al*. (2001) The *nonA* gene in *Drosophila* conveys species-specific behavioral characteristics. *Genetics* 158, 1535–1543
63 Fitzpatrick, M.J. (2004) Pleiotropy and the genomic location of sexually selected genes. *Am. Nat.* 163, 800–808
64 de Belle, J.S. and Heisenberg, M. (1996) Expression of *Drosophila* mushroom body mutations in alternative genetic backgrounds: a case study of the *mushroom body miniature* gene (*mbm*). *Proc. Natl. Acad. Sci. U. S. A.* 93, 9875–9880
65 Roth, M.E. *et al*. (2004) Expression profiling using a hexamer-based universal microarray. *Nat. Biotechnol.* 22, 418–426
66 Shendure, J. *et al*. (2004) Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* 5, 335–344
67 Innis, M.A. *et al*., eds (1990) *PCR Protocols: A Guide to Methods and Applications*, Academic Press
68 Sambrock, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual* (3rd edn), Cold Spring Harbor Press
69 Rose, T.M. *et al*. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* 26, 1628–1635
70 Schaefer, B.C. (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* 227, 255–273
71 Bustin, S.A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25, 169–193
72 Whitfield, C.W. *et al*. (2002) Annotated expressed sequence tags and cDNA microarrays for studies of brain and behaviour in the honey bee. *Genome Res.* 12, 555–566
73 Clayton, D.F. (2004) Songbird genomics: methods, mechanisms, opportunities and pitfalls. *Ann. N.Y. Acad. Sci.* 1016, 45–60
74 Renn, S.C.P. *et al*. (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* 5:42
75 Dougherty, J.D. and Geschwind, D.H. (2002) Subtraction-coupled custom microarray analysis for gene discovery and gene expression studies in the CNS. *Chem. Senses* 27, 293–298
76 Diatchenko, L. *et al*. (1996) Reliable cDNA subtraction starting from small amounts of total RNA. *Mol. Biol. Cell* 7, 932–932
77 Ginsberg, S.D. and Che, S.L. (2002) RNA amplification in brain tissues. *Neurochem. Res.* 27, 981–992