Ann. Zool. Fennici 43: 423–442 ISSN 0003-455X
Helsinki 29 December 2006

Determining the molecular basis of sociality in insects: progress, prospects and potential in sociogenomics

Seirian Sumner

Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

Received 1 Nov. 2005, revised version received 27 Oct. 2006, accepted 29 Aug. 2006

Sumner, S. 2006: Determining the molecular basis of sociality in insects: progress, prospects and potential in sociogenomics. — *Ann. Zool. Fennici* 43: 423–442.

How complex biological diversity can arise from seemingly simple strands of DNA is one of the most pertinent of questions confronting 21st century biologists. With the increasing availability and cost effectiveness of genomic techniques, there has been a rapid expansion in the taxonomic breadth of species that can now be studied at the molecular level of the genes. Consequently, behavioural ecologists are now able to examine the behaviours of their study organisms at an entirely new level. Here I review the current progress made in sociogenomics — the study of the molecular basis of sociality — with particular emphasis on what genome-level studies can reveal about social evolution. First I discuss the evolutionary interactions that occur between the genome and sociality. Next I review the current literature on how genes underlie queen and worker caste evolution: I identify 19 genes that are likely to be of importance in the evolution of caste systems across eusocial taxa; I make predictions on how gene expression patterns might orchestrate the evolution of social complexity, and make preliminary tests using the available data. Finally, I outline major questions in social evolution that can be addressed for the first time using a sociogenomic approach, highlight practical considerations in sociogenomics and discuss suitable model systems for future research on the molecular basis of sociality.

Introduction

It has long been recognised that biological diversity evolves at the level of the genes (Darwin 1859, Dawkins 1976, 1982). Recent advances in molecular biology have revealed that even the most complex organisms are made up of a surprisingly small number of genes. For example, the human genome consists of a mere 20 000–25 000 genes (Li *et al*. 2001). How does such a seemingly limited genetic toolkit produce complex biological diversity? Clearly, a single gene must have many functions and interact with other genes in complex networks. Determining how variation in gene regulation brings about phenotypic novelties (e.g. in terms of an individual's morphology, physiology and/or behaviour) is crucial for understanding how the diversity of life evolves (Carroll *et al*. 2005).

Genes are regulated in response to the environment. The environment of a social animal is especially variable, as any individual is influenced by its social interactions with other group members in addition to the usual biotic and abiotic influences experienced by non-social animals. This may explain why the phenotypic

diversity displayed by the social insects (bees, wasps, ants and termites) is amongst the most impressive in the animal kingdom. The morphological, physiological and behavioural variation displayed between individuals within a single colony or amongst species within a taxonomic family provides an unrivalled opportunity to explore the relationship between the genome and biological diversity at a range of different unit levels (Keller 1999). Studying social insects at the molecular level across different levels of sociality provides critical insights into how complex, highly derived, social systems evolved from simpler, ancestral ones. Determining the genetic basis of sociality has become a rapidly growing field of research, as techniques for studying the genome have become more accessible and affordable (Robinson 1999, 2002b, Evans & Wheeler 2001). The integration of molecular genetics with the study of sociality — 'sociogenomics' *sensu* Robinson (Robinson 1999) — therefore leads the classical study of social behaviour to new levels, making it one of the most promising fields for studying how novel gene functions evolve, as well as revealing how variation in transcript abundance of certain genes reflects behaviour and influences social evolution (Fitzpatrick *et al*. 2004). Questions raised include whether the same genes underlie similar behaviours in different taxa, and identifying genes that have been crucial in influencing social evolution, or which have been completely lost or gained. These gene-level insights into sociality enable us to build a more realistic picture of how the genome orchestrates diversity, in the form of sociality.

There are several excellent reviews of sociogenomics, which in particular describe the ground-breaking progress that is being made in understanding how the genomic properties of the honeybee (*Apis mellifera*) underlie its sociality (Evans & Wheeler 2001, Robinson 2002a, 2002b, Robinson *et al*. 2005). In this paper, I explore the realms of sociogenomics beyond the honeybee. I draw together current data on social insect genomes in order to discuss the complex interaction between sociality and genome evolution. In particular, I focus on the genetic basis of queen and worker castes, the topic in sociogenomics that has received the most attention to date. I review the literature in order to identify common genes and expression patterns that underpin different caste systems, and provide a synthesis of our understanding of caste evolution at the level of the genome. Finally, I discuss future directions in sociogenomics, outlining key questions in social evolution that can now be addressed, and discuss the potential of social insects as genetic-model organisms of the future.

Social evolution and genome evolution: cause or consequence?

There may be something special about the genome of social animals that predisposes them to become social. For example, they may have genome 'signatures for eusociality', or the ancestors of social lineages may have had a rich complement of genes coding for recognition or communication receptors which facilitated social evolution. Conversely, being social might affect the way the genome evolves. For example, there may have been major genome rearrangements through chromosome breakage and exchange of genes during the transition from non-social to social. Disentangling the causes of eusocial evolution from those that arise as a consequence of it is important if we are to develop realistic models on the molecular basis of social evolution, as well as understand how the (social) environment influences genome evolution. A major obstacle in disentangling cause from consequence is the feedback between the genome and sociality. A likely scenario is outlined in Fig. 1: Haplodiploidy, a property of most social insect (all Hymenoptera) genomes, can predispose individuals to altruistic behaviour by altering the social structure of a group such that some individuals help others reproduce at the expense of their own direct reproduction. In this way, the genome has the potential to influence social evolution. The resulting change in social structure (i.e. reduction in the ratio of breeders to non-breeders) means that the effective population size is reduced (Crozier 1979), which might alter the rate of molecular evolution (Bromham & Leys 2005). Any resulting gene losses, gains and duplications, nucleotide substitutions and deletions could contribute to the evolution of

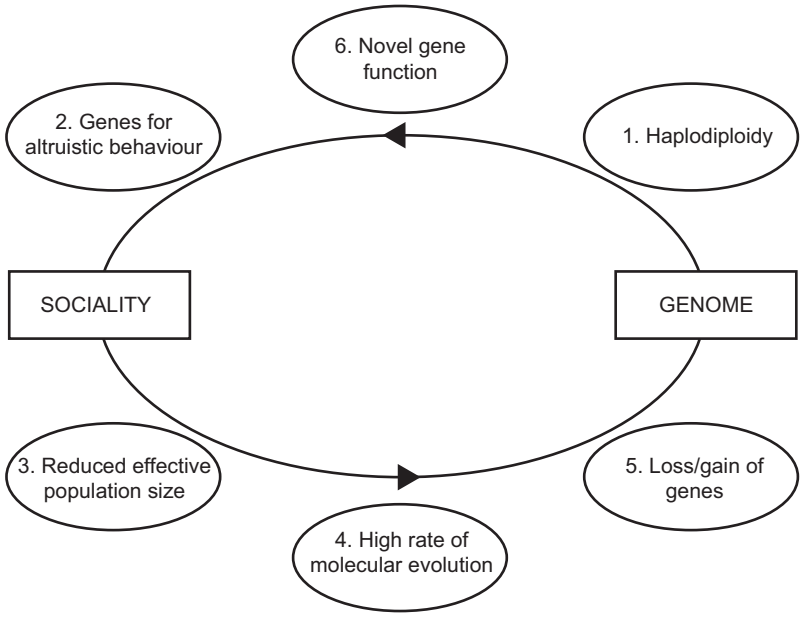

Fig. 1. A hypothetical example of how the genome may influence the evolution of sociality and sociality influence genome evolution. Examples of feedback are numbered 1–6 in the order in which they are likely to take effect. *See* text for further details.

novel genes or gene functions, which in turn may alter the behaviour of individuals and their social structure. In sum, sociality has the potential to act as a vehicle for genome evolution through a constant feedback: changes at the gene level alter behaviour which alters how natural selection acts on the genes, and so on. Determining how sociality evolves ultimately demands a better understanding of this feedback.

The genomes of many social insects exhibit unique properties that appear to be associated with sociality. For example the haplodiploid genetic system, found in all Hymenoptera (bees, wasps and ants), can provide the basis for eusocial evolution by providing relatedness incentives for sib-rearing. Females are diploid, developing from fertilised eggs, but males are haploid, developing from unfertilised eggs. The asymmetry in a female's relatedness to her sisters (r = 0.75), brothers (r = 0.25) and own offspring (r = 0.5) has the potential to predispose haplodiploid females to raise sisters instead of their own offspring (Hamilton 1964, Trivers & Hare 1976). Under certain conditions (e.g. single-mating and female-biased sex ratios) this can form the basis for a reproductive division of labour between queens and workers as well as overlapping mother–daughter generations, both of which are conditions for eusociality (Wilson 1971). However, haplodiploidy is not a prerequisite for eusocial evolution. Neither is it likely to have evolved as a consequence of eusociality. Evidence for this is found amongst the eusocial taxa that are diplodiploid (e.g. naked mole rats, termites and aphids, shrimps and an ambrosia beetle), the many large taxonomic groups that are haplodiploid but lack any eusociality (e.g. some mites, beetles, white fly and scale insects (Normark 2003)).

Cause and consequence cannot be distinguished so easily for other genome properties found in social insects. For example, there is some evidence that social insects tend to have higher chromosome numbers than their non-social counterparts (Sherman 1979, Crozier 1987). Also, worker specialisation in ants roughly correlates positively with chromosome number, and primitively eusocial wasps appear to have lower chromosome numbers than advanced eusocial wasps, suggesting that chromosome number most likely became elevated in response to sociality, rather than the other way round (Sherman 1979). High chromosome number is likely to reduce the variance in sib-sib relatedness, but it is not clear how this might benefit sociality. Sherman suggested that because workers would be less able to determine the relative fractions of genes carried by differ-

ent siblings, they would raise siblings unselectively, which would be the queen's preference. However, this requires that workers can identify with which siblings they share the most alleles, for which there is little evidence (Queller *et al*. 1990, Keller 1997). Moreover, any change in selection pressure that might arise as a result of a mutant queen with high chromosome number would take many generations to exert an effect (Dawkins 1982). High chromosome number, that for whatever reason evolved, may instead be a pre-adaptation to eusociality (Dawkins 1982). This conclusion would also account for the many non-social species which have high chromosome numbers, and the eusocial species which have a very low chromosome number (e.g. the primitive ant genus *Mymecia* has only 1 pair of chromosomes (Crosland & Crozier 1986)). The interplay in the evolution of chromosome number and sociality evidently requires more rigorous testing of the data with phylogenetic correction and a more considered explanation of the mechanism.

There has been some discussion as to whether high rates of molecular evolution may have driven social evolution. The honeybee (*Apis mellifera*) has a very high recombination rate and exhibits great colony-level genetic diversity (Hunt & Page 1995), whilst more primitively-eusocial species (e.g. *Bombus terrestris*) and solitary species (e.g. *Nasonia vitripennis*) have much lower recombination rates. Although limited, these data suggest that social complexity and recombination rates are positively correlated (Gadau *et al*. 2000). High recombination rates may produce greater genotypic, and hence probably phenotypic, diversity amongst workers through multi-gene traits where variable loci are linked on chromosomes. A related issue is whether the small effective population sizes that arise as a result of some individuals becoming workers affects the rate of molecular evolution in social animals. Bromham and Leys (2005) conducted a phylogenetic analysis to see if there is a link between sociality and the rate of molecular evolution across the Hymenoptera, termites, shrimps and naked mole rats, by comparing related non-social and social species. In contrast to an earlier study that used fewer taxa (Schmitz & Moritz 1998), they found no generally significant relationship. However, advanced eusocial species did generally have a higher nucleotide substitution rate than non-social species, suggesting that over long evolutionary periods complex sociality may increase overall rates of molecular evolution. Their analysis was based on 1–3 genes per social/non-social comparison, and so it is possible that more sequence data might reveal further relationships between sociality and molecular evolution.

Whether a high rate of molecular evolution is a cause or consequence of sociality, it will certainly give rise to the evolution of new genes or new gene functions through gene duplication, single nucleotide substitutions, insertions or deletions. There is already evidence for this as some conserved genes have novel functions in social animals (Robinson & Ben-Shahar 2002). For example, major royal jelly protein (MRJP) is used by honeybee workers in the production of royal jelly (Whitfield *et al*. 2003), whilst in other social insects it is associated with queen-specific functions (Tian *et al*. 2004, Sumner *et al*. 2006). Moreover, Krieger and Ross (2002) recently identified for the first time a gene that determines a complex social behaviour, illustrating that simple nucleotide mutations can strongly influence sociality. They showed that allelic differences in the gene *Gp-9* in the fire ant *Solenopsis invicta* determine whether a colony will be single- or multi-queened. Workers that are heterozygous at the locus (*Bb*) will tolerate multiple queens (polygyny) if queens carry the *b* allele, but will kill *BB* queens. Thus, the social structure of a colony depends on the presence or absence of the *b* allele. The *Gp-9* sequence is similar to pheromone-binding proteins in moths, and is used in chemical recognition by *S. invicta,* prompting the authors to suggest that it is used as a phenotypic signature, or "green-beard" by workers to distinguish between queens. Interestingly, the two alleles differ only at 9 nucleotides, which code for 8 amino acids (1 substitution is synonymous, causing no change in amino acid production). Moreover, the authors were able to identify 3 amino acids shared uniquely by the *b*-like allele, of which one or more is likely to play an essential role in producing the gene product that induces polygyny. It is unclear whether the role of *Gp-9* in controlling polygyny is con-

served across taxa beyond the genus *Solenopsis*, as the authors were unable to amplify the gene in a related myrmicine genus, *Monomorium*. This suggests that the role of *Gp-9* as a social regulator evolved as a consequence of sociality in *Solenopsis invicta* rather than being a cause, although more extensive comparative studies are required to confirm this. The most probable scenario is that the rate of molecular evolution arises as a consequence of sociality as well as being a cause.

That gene regulation and function can be influenced by sociality, strongly suggests that unique genes (or gene networks) that are absent in non-social animals are likely to evolve. A candidate for this might be a gene(s) for altruism (Wade 1980, Michod 1982, Queller 1992). In order for an altruistic gene to spread it must be present in the beneficiary (i.e. the queen) as well as the altruist (i.e. the sterile worker). Either it must be facultatively expressed (e.g. some adults help, and others reproduce) or it must be obligately expressed at different life stages (e.g. young adults help and old ones reproduce) (Charlesworth 1978, Crozier 1979, Bourke & Franks 1995). If the altruism gene is newly evolved as a product of sociality it will be absent in the solitary relatives of the social lineage. Linksvayer and Wade (2005) elaborate on this by explaining eusocial evolution in terms of maternal carers (queens) and sib-social carers (workers) whose behaviours are derived via differential expression of ancestral maternal care genes. The authors make predictions which can be — and to some extent have already been (*see* below) — tested using sociogenomic techniques, namely that queen (maternal care) and worker (sib-social care) behaviours arise through heterochrony (temporal expression differences) of ancestral maternal-care genes. There is already good evidence that queen and worker castes from a range of social insect societies differentially express shared genes, both at the developmental and adult stages. Studies on the molecular basis of castes in social insects have figured prominently in sociogenomics to date, providing the best picture yet of social evolution at the level of the genome. I review this blossoming research topic in the next section.

The genetic basis of queen and worker castes

One of the major drivers in the evolution of social complexity is task partitioning between group members. In social insects it is taken to extremes in the form of reproductive (queen) and non-reproductive (worker, soldier) castes. Caste specialisation in social insects correlates strongly with ecological success (Bourke 1999). Thus, studying the genetic basis of castes in extant social insects provides insights into the molecular machinery central to eusociality. More broadly, such studies can also reveal how a single genome can give rise to phenotypic diversity, what and how novel genes and gene functions evolve, and what the relationship is between the transcriptome and behavioural, physiological and morphological diversity.

Genes underpinning queen and worker castes

Queens and workers are different phenotypes arising from the same genome. With few exceptions (Julian *et al*. 2002, Volny & Gordon 2002, Helms-Cahan & Keller 2003), whether an individual becomes a queen or worker depends on how it responds to environmental stimuli at critical periods in caste determination, rather than genotypic differences (Wilson 1971, Wheeler 1986). Phenotypic variation exhibited by castes therefore usually arises through differential expression of shared genes. Insights into the evolutionary origins of caste are best obtained from studying simple societies, which exhibit the most primitive caste systems (where queens and workers differ only in behaviour) and are likely to represent the ancestral state of advanced castes (where queens and workers may differ in morphology) found in more complex insect societies. The over-arching question is whether the genes underlying behavioural differences in castes of simple societies (i.e. those associated with the origins of caste, from an ancestral, solitary state) are the same as those underlying morphological differences in castes of complex societies (i.e. those associated with the maintenance of castes, in the derived state) (Evans &

Wheeler 1999). Thus, comparisons of behavioural castes in simple societies with morphological castes in more complex societies can reveal the identity of some of the molecular drivers of caste evolution. It is a fortunate coincidence that the studies in which gene expression differences between castes have been studied cover a sizable complement of social complexity, ranging from simple behavioural castes in paper wasps (*Polistes canadensis*) to morphological castes in honeybees (*Apis mellifera*), fire ants (*Solenopsis invicta*) and termites. Moreover, the studied species represent amongst them at least 6 independent origins of eusociality, in each of which queen and worker castes have evolved independently (Wilson 1971). Here I summarize the findings of these studies and identify genes that appear to be ubiquitously differentially expressed with respect to caste across taxa.

Differential gene expression in castes of social insects

Genome-level research in social insects is currently best established in the honeybee *Apis mellifera*, and with a fully sequenced genome near completion in this species it will be a powerful resource for comparative genome research in social insects (Robinson *et al*. 2005). Evans and Wheeler (1999, 2000) paved the way for honeybee genomics by identifying differentially expressed genes in queen and worker larvae. These authors were the first to empirically illustrate that differential expression of many genes underlie queen and worker polyphenisms in social insects. Subsequent development of a microarray of the honeybee brain has facilitated large-scale transcriptome screening of adult honeybees, and some inspiring studies on the genes underlying adult worker behaviours in honeybees have recently been published. For example, honeybee workers change their behaviour from nursing to foraging with increasing age. Kucharski and Maleszka (2002a) illustrated the spatial and temporal changes in gene expression that take place during this behavioural development by comparing expression in the brains and abdomens of naïve worker bees, nurses and foragers. More recently, Whitfield *et al*. (2003) used the expression profiles of 50 predictive genes to accurately determine whether a bee was a nurse or forager. These studies illustrate very elegantly how complex behaviours that are coordinated by multiple genes in undefined networks can be studied in a genetic non-model organism.

Aside from the honeybee, queen and worker castes have been studied at the gene level in several other species that exhibit morphological caste differences. Genes that are differentially expressed amongst castes in adults and brood of the bumblebee *Bombus terrestris* have been isolated (Pereboom *et al*. 2005), revealing some intriguing patterns of expression. Pereboom *et al*. (2005) found that larval caste development in the bumblebee is associated with temporal variation in gene expression, such that larvae which up-regulate genes early in development become queens, but those that up-regulate these same genes late in development become workers. Temporal regulation of gene expression is increasingly recognised as a way in which a seemingly small number of genes can produce such biological complexity (Arnosti 2003). Pereboom *et al*. (2005) made the first comparisons between expression patterns of brood and adults within a species. They found that different genes appear to underlie caste differences in brood and adults. This study also facilitated the first comparisons of the genetic basis of castes across species. Four genes were identified as differentially expressed in both *B. terrestris* and *A. mellifera*, but only one of them (hexamerin II) shared similar expression patterns between the two species. Five differentially expressed genes have also been identified amongst newly-emerged adult queens and workers of the stingless bee *Melipona quadrifasciata,* and one of these (cytochrome P450) was also differentially expressed in *A. mellifera* (Judice *et al*. 2004). The pioneering of transcriptome studies on species other than *A. mellifera* opens avenues for conducting comparative genomic analyses on many other social insects.

Ants exhibit more derived caste systems than do bees and wasps, with extreme caste polymorphism amongst workers and in some species the evolution of a soldier caste (Hölldobler & Wilson 1990). To date, there has only been one study on gene-level caste differences in ants. Tian *et al*. (2004) isolated differentially expressed genes in

dealate queens (mated 24 hrs prior to collection, who have shed their wings) and alate queens (virgin queens, yet to mate and shed their wings) of the imported fire ant *Solenopsis invicta*. They identified 7 genes that are up-regulated in dealate queens, suggesting these genes are important in queen maturation. Tian *et al*. (2004) also produced the most comprehensive data so far on how expression levels of single genes change in the different brood and adult developmental stages, which includes eggs, larvae, pupae, adult workers, alate and dealate queens, although only 9 genes were examined. Because there are no solitary or even primitively eusocial species of ants, studies on ants can only reveal how castes are maintained rather than how they evolved (Crozier 1982). Nonetheless, these studies are important since the highly complex societies (and hence extreme caste differentiation) observed in ants have not been achieved by wasps or bees.

To date, there has only been one study on the genes underlying queen and worker castes in a truly primitively eusocial insect, where queen and worker behavioural castes are determined during adulthood. Genes associated with adult development in the paper wasp *Polistes canadensis* show gradual changes in expression with social status, such that queens (high-ranked females) and newly emerged (low-ranked) females exhibit distinct patterns of gene expression, with workers (mid-ranked females) exhibiting intermediate patterns (Sumner *et al*. 2006). Sumner *et al*. (2006) looked for similarities amongst the genes and patterns of expression displayed by adult castes in the 4 species mentioned above (*P. canadensis*, *B. terrestris*, *A. mellifera* and *S. invicta*) and identified 9 genes that appear to be differentially expressed with respect to caste across 2 or more of these species. This comparison suggested that although the same genes may be differentially expressed with respect to caste across taxa, the patterns of expression appear to be widely divergent. Research is to be strongly encouraged on species with behavioural caste systems, especially through comparisons with extant solitary relatives, which are likely to represent the ancestral non-social state. Such comparisons may be crucial in revealing the genome changes that accompany the first steps in eusocial evolution.

Termites are phylogenetically distant from the hymenopterans, being more closely related to cockroaches and mantids. All termites are eusocial and represent an origin of caste evolution that is independent of the Hymenoptera (Thorne 1997). Comparisons of termite caste systems with hymenopteran caste systems are complicated for several reasons. Firstly, termites are hemimetabolous and so immature stages have a similar body form to adults such that a single individual can pass through worker (pseudogate) and reproductive or soldier castes within its lifetime. Since all termites are eusocial there is no intermediate, primitively eusocial termite and so a direct comparison of caste evolution with the Hymenoptera cannot be made. However, because an individual termite can change caste over time, there are elements of caste development in termites that may be considered analogous to the temporal, behavioural castes observed in extant primitively eusocial Hymenoptera (e.g. *Polistes* as discussed above). Another problem is that sterile castes (workers and soldiers) in termites are composed of both males and females, whilst in the Hymenoptera the sterile castes are always female. However, sex-specific caste specialisations are restricted to the most derived termite groups, for which there are currently no gene expression data. Despite these complications, gene-level comparisons made with queen and worker pathways in the social Hymenoptera may reveal important insights into convergent evolution of caste systems amongst the social insects (Miura 2004). Much of the gene expression research to date has focussed on the lower termite *Reticulitermes flavipes,* where workers, soldiers, alates (winged reproductives) and supplementary reproductives (wingless) have been examined (Scharf *et al*. 2003, 2004). Genes that are exclusively expressed in the mandibular glands of soldiers of another species, *Hodotermopsis japonica,* have also been isolated (Miura *et al*. 1999, 2003, Miura 2001, Hojo *et al*. 2005). Comparisons of these data with soldier castes in hymenopteran species (e.g. *Atta* leafcutting ants) may prove insightful when such data is available, since soldiers in the Hymenoptera are modified workers whilst soldiers in the termites are thought to have evolved independently of workers.

Identifying genes associated with caste across eusocial taxa

Here, I compare the genes and expression patterns that have been identified as differentially expressed with respect to caste across 8 species of social insect. These are 3 bees (*Bombus terrestris*, *Apis mellifera* and *Melipona quadrifasciata*), 1 wasp (*Polistes canadensis*), 1 ant (*Solenopsis invicta*) and 3 termites (*Reticulitermes flavipes*, *Nasutitermes takasagoensis* and *Hodotermopsis japonica*). To my knowledge, these are the only species in which castes have been examined at the level of the transcriptome. Unfortunately, the current diversity of species studied is too limited and the number of genes studied too small to determine whether the genes underlying behavioural and morphological castes are the same (*see* 'Genes underpinning queen and worker castes'). Thus, my purpose here is to simply identify any genes that have been conserved across developmental stages and across evolutionary time with respect to caste. Accordingly, I (1) identify genes that underlie caste differences in larvae and adults of any one species; (2) identify any genes that underlie caste differences in any one life-stage across different species. Finally, I (3) compare expression patterns of specific genes within these groups and discuss any functional divergence of caste-associated genes that has occurred over evolutionary time.

In order to identify genes that are ubiquitously expressed across taxa with respect to caste, I looked for similarities in the published sequences (available on Genbank) rather than used the gene category given to each expressed sequence tags (EST — sub-section of a gene) at the time of the publication. This was necessary because many of the similarity matches recorded at the time of publication are now out of date as new sequence data are constantly submitted to Genbank. I looked for similarities amongst cDNA ESTs available for *A. mellifera* $n = 156$ ESTs (Corona *et al.* 1999, Evans & Wheeler 2000, Toma *et al.* 2000, Ben-Shahar *et al.* 2002, Kucharski & Maleszka 2002a, 2002b, Piulachs *et al.* 2003, Whitfield *et al.* 2003), *B. terrestris* $n = 12$ (Pereboom *et al.* 2005), *M. quadrifasciata* $n = 16$ (Judice *et al.* 2004), *P. canadensis* $n = 43$ (Sumner *et al.* 2006), *S. invicta* $n = 10$ (Tian *et al.* 2004) and the 3 termites $n = 17$ (Miura *et al.* 1999, Scharf *et al.* 2003, 2004, Hojo *et al.* 2005). These sequences were aligned into overlapping regions (contigs) using the sequence viewing program Sequencher, using either large gap alignment or 'dirty' realignment with 60%–90% minimum match and 70%–90% minimum overlap. Contig sequences that had not scored a similarity match on Genbank at the time of publication were subjected to BLASTX in order to verify homology and infer gene function.

I identified 19 ESTs that are differentially expressed with respect to caste in 2 or more species (listed in Appendix 1) and hence may have an evolutionary important role in caste systems in social insects. Although differentially expressed ESTs from *M. quadrifasciata* had similarity matches with *A. mellifera* sequences (Judice *et al.* 2004), none of these are currently known to be differentially expressed with respect to caste in *A. mellifera*, which explains the absence of *M. quadrifasciata* in Appendix 1. No single EST was consistently expressed by the same caste in all species for which there are data, suggesting that the specific roles of these genes have changed over evolutionary time, in terms of either their specific function or how they interact with other genes or gene networks.

I used these data to determine whether there are any genes that underlie caste differences in both larvae and adults within species. Although some genes are unlikely to be expressed by brood and adults (e.g. those involved in egg production), genes that are expressed by both life-stages may have important roles in caste regulation. Pereboom *et al.* (2005) found in the bumblebee that 8 out of 12 genes were differentially expressed with respect to castes in both brood and adults, and concluded that the same genes do not always underlie caste differences between adults and larvae. In addition to this study, I found comparable data for the honeybee and fire ant, and was able to identify 7 genes that were differentially expressed in larvae and adults of one or more of these three species (*see* Table 1). Expression patterns in adults and either early or late instar larvae (but not in both instars) were the same for 4 genes (Table 1, column 1), e.g. Cytochrome oxidase I was upregulated in both adult workers and late-instar worker-destined

larvae, but not early-instar larvae of *B. terrestris*. These 4 genes may be important in regulating the switch from one caste to another. Five genes had different expression patterns in larvae and adults of any one species, and 2 of these were expressed only in larvae (cuticle protein and hexamerin II). The remaining 3 genes (cytochrome oxidase I & II and ATP synthase) were all up-regulated by one caste in larvae and another caste in adults (Table 1, column 2). In conclusion, some genes are differentially expressed with respect to caste in both brood and adults, but many genes appear to have different roles in the different life-stages. Direct comparisons of expression patterns between queens and workers at brood and adult stages for more genes, and studies on the specific functions of genes at different life-stages are evidently required.

My second aim was to identify any genes underlying caste differences across species. These genes may shed light on the molecular processes involved in caste evolution as well as reveal mechanisms of gene evolution. I examined larvae and adults separately, since the relevance of examining the same genes at different life-stages is unclear (*see* previous paragraph). In the larvae of the honeybee and bumblebee (fire ant is excluded as there are too few data) 6 genes were differentially expressed with respect to caste (Table 2). Three genes have similar patterns of expression across species: peroxiredoxin, ribosomal protein and hexamerin II are all up-regulated in late-instar worker-destined larvae relative to queen-destined larvae (Table 2, column 1). The other 3 genes exhibit different expression patterns across species: cytochrome oxidase I, ATP synthase and cuticle protein are all up-regulated in late-instar worker-destined larvae of the bumblebee and late-instar queen-destined larvae of the honeybee. Amongst adults (which includes all species listed in Appendix 1), five genes appear to be consistently up-regulated in workers (Table 3, column 1). This suggests that worker behaviours in even distantly related species (such as the paper wasp and honeybee) are underpinned not only by some of the same (or closely related) genes, but also by similar expression patterns. These genes may be particularly fruitful to pursue in future studies on adult castes in social insects. A further 6 genes clearly do not share expression patterns across species (Table 3, column 2). These genes may still be of evolutionary importance in the caste systems

Table 1. Genes that are differentially expressed with respect to caste in both adults and larvae of *B. terrestris* (*Bombus*), *A. mellifera* (*Apis*) or *S. invicta*. Genes are listed as those that have similar patterns of expression in adults and larvae (column 1: i.e. up-regulated in the same caste in both life-stages) and those that differ in their patterns of expression (column 2: i.e. up-regulated in different castes at different life-stages).

Same expression patterns in adults and larvae	Different expression patterns in adults and larvae
Cytochrome oxidase I (*Bombus*)	Cytochrome oxidase I (*Apis*)
ATP synthase (*Bombus*)	ATP synthase (*S. invicta*, *Apis*)
Peroxiredoxin (*Bombus*)	Cytochrome oxidase II (*S. invicta*)
Ribosomal protein (*Bombus*)	Cuticle protein (*Bombus*)
	Hexamerin II (*Bombus*)

Table 2. Genes that are differentially expressed with respect to caste in larvae of both *B. terrestris* and *A. mellifera*. Genes listed in column 1 have similar patterns of expression, and those in column 2 differ in their patterns of expression.

Same expression patterns in larvae across species	Different expression patterns in larvae across species
Hexamerin II	Cytochrome oxidase I
Peroxiredoxin	ATP synthase
Ribosomal protein	Cuticle protein

Table 3. Genes that have similar expression profiles (column 1) and mixed expression profiles (column 2) amongst adults of different species.

Genes up-regulated in adult workers of different species	Genes varying in expression patterns across adults of different species
Cytochrome oxidase I	Peroxiredoxin
ATP synthase	Ribosomal protein
Heat shock protein	IDGF
Lectin like	Transferrin
SPARC	Vitellogenin
	MRJP

of social insects, but their functions are likely to have diverged.

It is clear now that some of the same genes are associated with caste determination across several social insect species. There is currently little overlap between the data available for both termites (Isoptera) and Hymenoptera (only 6 genes) and so it is difficult to determine to what extent the same genes underlie castes in these two distantly related lineages. However, there is evidence of convergent evolution in the genetic basis of caste systems in both the Hymenoptera and Isoptera, since 4 genes (18S ribosomal protein, cytochrome oxidase 1, Hexamerin I and 28S ribosomal protein) are differentially expressed with respect to caste in both termites and at least one other hymenopteran species (Appendix 1).

Functional divergence in genes associated with queen and worker castes

In the previous section, I identified genes that are likely to have evolutionary and/or mechanistically important roles in some of the caste systems displayed in social insects. More research is required in order to understand the functions of these genes in the different castes, life-stages and species, as patterns of expression with respect to caste are clearly poorly understood at present. In this section I review what we know about the putative functions of a few of these genes in order to illustrate the point that many of these genes have multiple functions within species and divergent functions amongst species or life stages. Our understanding of gene regulation and multi-functionality is limited and this may explain why the genetic patterns underlying caste evolution (reviewed in 'Identifying genes associated with caste across eusocial taxa') are currently difficult to interpret.

Vitellogenin is a gene that is primarily associated with egg production and as expected is up-regulated in queens relative to workers in honeybees (Piulachs *et al*. 2003), paper wasps (Giray *et al*. 2005, Sumner *et al*. 2006) and termites (Scharf *et al*. 2004). Recent studies, however, have suggested that its role in the more advanced eusocial insects may have diverged. For example, in honeybees it is also involved in the production of royal jelly and this suggests that vitellogenin may have helped direct the evolution of age polyethism in the honeybee, since young nurse bees up-regulate it relative to older, forager bees (Amdam *et al*. 2003a, Piulachs *et al*. 2003).

Major royal jelly proteins (MRJP) are a family of proteins with multiple functions and are of particular interest because of the novel functions they seem to have evolved in some social insects. In some species they are evidently associated with active queen behaviour: they are expressed by queens but not workers in the paper wasp (Sumner *et al*. 2006), and a related protein (yellow protein, from the same family as MRJP) is expressed by dealate but not alate queens of the fire ant, *S. invicta* (Tian *et al*. 2004). However, in the honeybee MRJPs are up-regulated by foragers (Kucharski & Maleszka 2002a) where they play a fundamental role in worker production of royal jelly (Kucharski & Maleszka 1998). MRJPs also appear to play an important role in the honeybee brain where 8 different proteins have already been identified, indicating high rates of gene duplication and diversification (Albert & Klaudiny 2004). These functions of MRJPs are evidently divergent and complex in some socially advanced species of social insect. It remains to be seen how widespread divergent evolution in this gene family has been in other eusocial lineages.

Another gene that has a clearly complex functional role in caste regulation is transferrin, an iron-binding protein that is thought to be involved in innate immunity and vitellogenesis in insects. In the honeybee it is upregulated in the brains of mature adult foragers and in the abdomens of virgin queens (Kucharski & Maleszka 2003). Such temporal and context-dependent fluctuations in expression of transferrin indicate that it has a dual function in the honeybee. In contrast, its role in the paper wasp appears to be more simple, where it is up-regulated strongly in queens, with very weak expression in workers (Sumner *et al*. 2006). The multi-functionality of transferrin is well documented in vertebrates (Espinosa-Jeffrey *et al*. 2002), and so its complexity in insects is of little surprise.

The genes discussed here represent a small fraction of the genes actually involved in the

evolution and maintenance of castes in social insects. The emerging pattern to date is that many genes have been co-opted into regulating caste systems — both across different life-stages within a species, and also amongst species separated by varying phylogenetic distances — but that gene function and regulation are clearly not ubiquitously conserved. It is likely that a key mechanism for social plasticity in many social animals, invertebrates and vertebrates, is spatial and temporal variation in gene expression. The design of experiments in future studies on the genes involved in sociality should bear this in mind. Moreover, it is important to note that most studies are biased towards finding convergence, since researchers tend to choose to study ESTs that have similarity matches with known protein sequences on Genbank, or candidate genes that have been studied in other animals. This means a great many ESTs that we know nothing about are omitted because they fall below the similarity threshold and remain unclassified. One such gene is listed in Appendix 1 (Unknown 1). Unknown genes are of particular interest as they may have evolved specifically in response to sociality and are likely to perform unique functions.

A model of caste evolution at the transcriptome level

A single-gene level approach to studying the molecular basis of sociality, however, may obscure any large-scale patterns of differential expression. Here I make predictions about the large-scale genomic changes that are likely to accompany the processes involved in caste evolution and discuss whether the gene expression data currently available fit these predictions.

Predictions

Extant eusocial species likely evolved from solitary species in which all females would have carried out both queen tasks (i.e. egg-laying) and worker tasks (i.e. foraging, nest building and brood rearing) (West-Eberhard 1987, Amdam *et al*. 2004, Hunt & Amdam 2005). Subsequent evolution of complex societies (in which morphological castes are determined during larval development) are likely to have evolved from simple societies (in which behavioural castes are determined during adulthood). Understanding how queen and worker castes evolve from a solitary state and how subsequent caste polymorphisms evolve are fundamental questions in the study of social evolution. It has been suggested that as castes evolve, the reproductive and non-reproductive parts of the solitary phenotype are decoupled, such that the two alternatives (queens and workers) are mutually dependent, complementary morphs (West-Eberhard 1996, Gadagkar 1997, Giray *et al*. 2005, Hunt & Amdam 2005). Thus, in addition to changes at the level of individual genes (discussed in 'Genes underpinning queen and worker castes'), the evolution of queen and worker castes from a solitary state may have involved redirecting transcriptome regulation into caste-specific roles. This could take place through suppression or induction of particular genetic pathways in one or both castes such that they express different, complementary sets of genes, the null hypothesis being that queens and workers are equally divergent. It has also been suggested that there are likely to be more genes with caste-specific expression in the complex (advanced) eusocial species than in the simple (primitively) eusocial species (West-Eberhard 1996, Gadagkar 1997, Linksvayer & Wade 2005). Thus, transcriptome patterns in queens and workers are predicted to diverge such that more genes become caste-specialised with increasing social complexity (*see* Fig. 2: Predictions).

Observations

These predictions can be tested by comparing the large-scale expression patterns of genes associated with queen and worker castes in simple societies with those in more complex societies. One way of doing this is to compare the relative frequency of genes that are equally expressed in queens and workers of species with different caste systems. This is a crude assessment and assumes that genes have equally important effects on the developing phenotype, but it is the only feasible way to address these predictions with the currently available data. I exam-

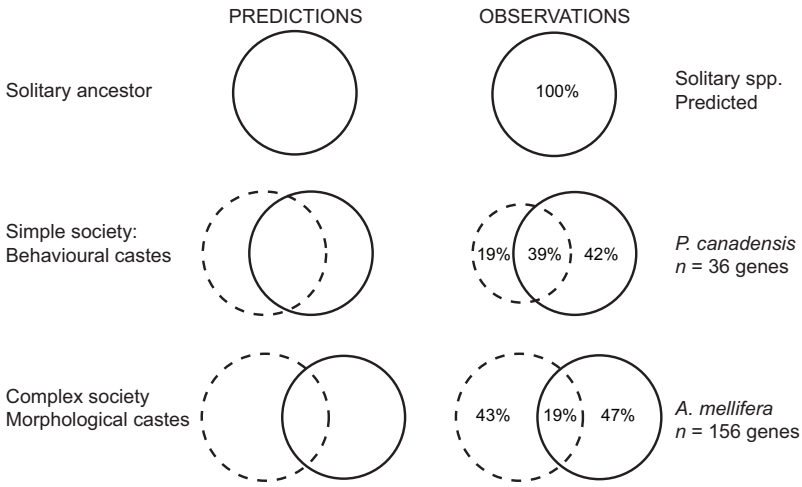

Fig. 2. Predictions of how the transcriptome changes with increasing social complexity. Observations of transcriptome differences between simple and complex societies. Patterns are based on the limited available gene expression data for simple societies of *P. canadensis* (adults) and more complex societies of *A. mellifera* (brood). Circles with dotted lines represent worker transcriptome, and those with solid lines represent the queen transcriptome. Overlap of the two circles indicates the proportion of genes that are equally expressed (less than 2-fold differences in observed expression); non-overlapping indicates the proportion of genes up-regulated in one caste.

ined published data on expression patterns in adult queens and workers of *P. canadensis* (from Sumner *et al*. 2006) and queen- and worker-destined larvae of *A. mellifera* (from supplementary data in Evans and Wheeler (2000)), which represent two extremes of sociality. I found some preliminary support for these predictions. I regarded genes as being differentially expressed between queens and workers if there was at least a 2-fold difference in expression level. In the behavioural castes of *P. canadensis,* a large proportion of the genes studied were equally expressed (less than 2-fold difference in expression level) by workers and queens (39%, 14/36). Of the genes that were differentially expressed, fewer were upregulated in workers with respect to queens (19%, 6/36), than in queens with respect to workers (42%, 15/36; Fisher's exact: $p = 0.037$). This suggests that the transcriptomes that give rise to queens and workers in *P. canadensis* have been unequally decoupled. Workers express fewer unique genes or expression patterns than queens. One interpretation of this is that workers are primarily queen-like but with part of the queen pathway suppressed (*see* Fig. 2, Observations): this would suggest that workers in *Polistes* evolved primarily by delaying the development of the reproductive state.

When queen and worker-destined larvae of *A. mellifera* are examined in the same way it is clear that they differ strikingly from *P. canadensis* (Fig. 2, Observations). First, there is no significant difference between the number of genes up-regulated in queens and workers of *A. mellifera* (74 and 68 genes respectively, Fisher's exact: $p = 0.55$). Second, the difference in the degree of decoupling in *A. mellifera* and *P. canadensis* is highly significant: queens and workers in *A. mellifera* expressed only 10% (16/158) of the sampled genes equally, whilst *P. canadensis* equally expressed 39% (14/36) of sampled genes (Fisher's exact: $p = 9.73 \times 10^{-5}$). Third, there is a significant difference in the proportion of genes up-regulated by workers relative to queens in *A. mellifera* and *P. canadensis* (43% and 19% respectively Fisher's exact: $p = 0.0083$). This comparison suggests that the degree of decoupling between *A. mellifera* queens and workers has been equal and more extensive relative to *P. canadensis*.

Although this analysis is necessarily preliminary and crude, the patterns that appear to be emerging are that evolving from a solitary lifestyle to a society with simple, behavioural castes involves suppression of queen pathways in order to evolve workers, rather than workers becoming

specialised. Increasing social complexity and the evolution of morphological castes, by whatever mechanism, is apparently accompanied by true decoupling of the transcriptome such that queens and workers exhibit separate and equally complementary phenotypes. Evidently, more rigorous tests of this model are required, which demand large-scale screening of random cDNA microarrays for a range of species representing different levels of social complexity, as well as a better understanding of the relative effects of different genes on phenotypic traits. Although the data available were obtained from macroarrays of selective libraries for both *P. canadensis* and *A. mellifera,* the patterns outlined in Fig. 2 are unlikely to be artefacts. This is because the genes examined were not all obtained from the same queen/worker subtractive libraries. For example, almost half the genes used in the *P. canadensis* study were obtained from a library of newly emerged females which are neither queen nor worker (Sumner *et al*. 2006). I repeated the analysis using just these genes and similar patterns were obtained, indicating that the pattern observed for *P. canadensis* in Fig. 2 is unlikely to be an artefact (Queens upregulate 4 genes whilst workers upregulate 1; another 4 genes are equally expressed). Likewise, the *A. mellifera* study included genes that were isolated in young, bipotent larvae that had not fully embarked on queen or worker pathways. Future analyses of larger, random datasets using quantitative real-time PCR to quantify expression across more species, including solitary ones, are however needed in order to confirm the patterns suggested here.

Future prospects in sociogenomics

The current achievements of sociogenomics (reviewed in this article) have already secured a promising future for social insects as model systems in genomic research. Firstly, they demonstrate that viable and informative gene-level studies can be performed on species whose genomes have been little studied (non genetic-model organisms). This is facilitated by the increasingly affordable techniques available for analysing genomes that are now transferable across research disciplines and taxonomic groups. Secondly, they illustrate how sociogenomics allows us to address questions beyond the scope of current genetic-model organisms like *Drosophila,* which are not social. For example, we are beginning to understand the complexity of the interactions that occur between the genome evolution and sociality, and the genetic basis of reproductive division of labour. Thirdly, studying key issues in social evolution at the gene level reveals novel insights into the patterns and processes involved in sociality, invoking new ideas and clarifying theoretical models of both social and genome evolution. In this section I identify key areas for future research in sociogenomics, with a particular emphasis on the evolution of sociality. I summarise some practical considerations for the future of sociogenomics and discuss the potential for social insects as genetic-model organisms of the future.

Major questions for sociogenomic studies

A very general question in biology is how the genome, with its limited genetic toolkit, produces the diversity of life observed around us (Carroll *et al*. 2005). The morphological, behavioural and social diversity displayed by social insects makes them ideal for addressing this question, although it is important to distinguish between studies on the origins of sociality from those on its maintenance. Here I outline key questions on the genetic basis of the origin and maintenance of sociality in social insects that can now be addressed using sociogenomic techniques.

How does the genome restrict the evolution of eusociality?

Eusociality evolves through the interplay of the environment and genes. But it appears to be difficult to become eusocial, as it evident by the rare and clustered nature of eusocial clades (Wilson 1971). Unlike the role of the environment in eusocial evolution, there has been little discussion of what genomic properties might restrict social evolution, partly because it is

impossible to locate failed attempts of social evolution. However, we can gain an insight into the changes that occur in the transition between the two life styles by comparing the genomes of primitively eusocial species with solitary relatives and identifying the genomic changes that occur. For example, is the regulation of certain genes uniquely altered in eusocial animals as compared to solitary ones? Have eusocial animals evolved novel genes (e.g. for altruism) or gene functions? Have they lost gene families or gene functions (Robinson & Ben-Shahar 2002)? Genes that are lost in eusocial species may be some of the key inhibitors of social evolution. We have begun to address these issues (reviewed in 'Social evolution and genome evolution: cause or consequence?'), but a more directed, comparative approach (Wei *et al*. 2002) is to be encouraged. Examining how genomes change after apparent reversions from eusociality to solitary living (Wcislo & Danforth 1997) would be particularly informative.

Are polyphenisms that have evolved in several different social lineages maintained by the same genes and expression patterns?

The major polyphenisms that characterise social insects have evolved several times in different lineages. The evolution of queen and worker castes is a prime example. Whether castes in different lineages of the Hymenoptera are underlain by the same genes, despite their independent evolutionary pathways, was discussed in 'The genetic basis of queen and worker castes'. Other phenomena that have evolved multiple times in the eusocial insects and would be good candidates for genomic studies include social parasitism, colony founding strategies, multiple mating and nest-mate communication and recognition systems. A phylogenetic approach of studying the genes that underlie the same polymorphisms in different social lineages will reveal whether these genes/regulatory patterns are products of convergent evolution or whether they are conserved genes, inherited from a shared ancestor.

Do the same genes/regulatory networks underlie both the origin and maintenance of sociality?

The genome properties responsible for the onset of eusocial evolution are not necessarily identical to those that maintain it. There is growing evidence that a relatively small number of genes produces vast biological diversity through temporal and spatial variation in expression (Arnosti 2003 and 'The genetic basis of queen and worker castes'). The morphological, behavioural and social diversity found amongst social insects provides an excellent opportunity to examine how phenotypic diversity evolves at the level of the genes (West-Eberhard 1987, 1989). Primitively eusocial species are the best representation of a species at the onset of eusociality as few of their traits are likely to be evolutionarily derived. Within-genus/family comparisons of species (i.e. which share a recent common ancestor) occupying different levels of sociality can reveal how genes and/or their expression patterns change in order to produce different levels of sociality. A good candidate group for such a study is the bumblebee genus *Bombus* which displays a range of caste mechanisms that may reflect different levels of sociality (Röseler 1991).

Practical limitations and opportunities in sociogenomics

It is important that sociogenomics progresses away from simply correlating gene expression profiles with social attributes, towards manipulating genes and altering phenotypic traits. A strong reason for doing this is that much of the gene expression detected in adult social insects may be a consequence of the behaviour rather than a cause (i.e. down-stream rather than up-stream genes). An indication of this is that so far we have not found any cis-regulatory elements, which regulate gene expression and are the real clues to the interaction between genes and social evolution. Moreover, the functions of isolated genes are almost entirely inferred rather than proven. Studying gene expression has its own problems. Firstly, it is essentially the study of the set of reactions that control the abundance

of gene products. Thus, although gene transcript abundance reflects behavioural variation, it does not necessarily predict gene product (e.g. protein) abundance. Protein targeting may be a more accurate approach to studying the relationship between the genome and diversity (Gerlai 2001), although proteomic techniques are currently less transferable between species and less affordable than genomic techniques. Secondly, gene expression studies are cluttered with noise. Distinguishing meaningful patterns from noise is a challenge as we currently do not fully understand the sources of noise (Raser & O'Shea 2005). Moreover, standardising expression levels between samples is problematic as many housekeeping genes are not (as their name suggests) equally expressed. Some of these problems can be solved by manipulating transcript abundance and observing the altered phenotype. Honeybee genomics is already testing the waters in this exciting new research field (Robinson 2002a): gene expression has been localised to certain body regions and even specific brain regions (Whitfield *et al*. 2003, Cash *et al*. 2005), and RNAinterference (RNAi) techniques have been employed to inactivate/activate target genes (Amdam *et al*. 2003b). However, it is a steep learning curve and likely riddled with unexpected problems. Behavioural traits are likely to be influenced by large suites of genes as well as environmental factors (e.g. Whitfield *et al*. 2003) and our understanding of how genes interact with each other in general is still very basic, even for genetic-model organisms (Brazhnik *et al*. 2002). This leads to problems when trying to use gene targeting to determine the function of a specific gene: for example, inactivating one gene may instigate compensatory responses by other genes, such that no phenotypic effect will be observed.

Comparative genomics is a large-scale, holistic approach to studying genome evolution, which looks for similarities and differences in 2 or more genomes (or parts of genomes), at any level, e.g. different species, subspecies or strains of the same species (Wei *et al*. 2002). It rests on the premise that the two genomes under examination share common ancestry such that every base pair in each organism can be explained as the combination of the original ancestral genome and the action of evolution. Such comparisons can provide insights into the evolution of a species or particular trait (e.g. sociality), for example in terms of gene birth and death, phylogenies, species/trait origins and adaptations. Comparative genomics usually involves looking for similarities in sequences that infer homology (i.e. that they have a common evolutionary ancestor). Genome comparisons allow the reconstruction of genetic changes (or genetic 'footprints') that occurred at crucial stages in the evolution of sociality.

Recent sequence data suggests that the honeybee shares over 95% of its orthologs with *D. melanogaster* (Whitfield *et al*. 2003). The long history of developmental genetics in *Drosophila* therefore provides a very lucrative spring-board for sociogenomics. Researchers are now exploiting this resource in ingenious ways. For example, it has been shown that *Apis mellifera* workers up- or down-regulate an ortholog of the *Drosophila* foraging gene (*for*) depending on whether they are foragers or nurses (Ben-Shahar *et al*. 2002). This indicates that gene regulation of specific behaviours can be highly conserved, even over the 300 millions of years of evolutionary history that separate *Drospophila* from *Apis* — *see* also Toma *et al*. (2000), Abouheif and Wray (2002), Ben-Shahar *et al*. (2002) for more examples of how *Drosophila* orthologs have been used to study social evolution.

Social insects as genetic-model organsims

The Hymenoptera lend themselves to comparative genomics because they offer an unrivalled range of behaviours, social adaptations and social complexity, with extant representatives for each of the steps likely to have been taken during social evolution (Evans & Wheeler 2000, 2001). The future of comparative sociogenomics rests on careful selection of species from across the spectrum of sociality. There are practical considerations as well as scientific ones to be made when selecting future model species for sociogenomics (Evans & Gundersen-Rindal 2003). Behavioural phenotypes are especially difficult to study because they are sensitive to

social and environmental fluctuations and they must be studied in a semi-natural context (Robinson & Ben-Shahar 2002). Ideal model species for sociogenomics would be easy to rear under controlled laboratory conditions, as well as easy to study in their natural environments. They would be common and cosmopolitan in their distributions in order to maximise accessibility for an international body of researchers, and their natural history should be well understood. The honeybee satisfies all these requirements and consequently is well on its way to being the first eusocial genetic-model organism. There is already a wealth of quantitative genetic data available for the honey bee, e.g. from QTL analyses (Hunt & Page 1995, Page *et al*. 2002) moreover, the annotation of the honey bee's genome sequence is now complete (The Honeybee Genome Sequencing Consortium 2006). The honeybee has emerged as an ideal candidate for sociogenomics, facilitating the study of social evolution at one of its most complex states. This is excellent news for students of sociality, as social insects now have a credible reputation for genomic studies. But insights into the origins of sociality are best obtained through complementary studies of simpler social organisations than those exhibited by the honeybee.

Polistes paper wasps are probably the best-studied genus of primitively eusocial insects (Turillazzi & West-Eberhard 1996). They are cosmopolitan in their distribution, can be reared relatively easily in natural, semi-natural (e.g. nest boxes) and lab environments, facilitating manipulation experiments. Their genome is relatively small (300Mb, Johnson *et al*. (2004)), and transcriptome studies (i.e. on mRNA) have already commenced on one species (Sumner *et al*. 2006). Colony-level social life in *Polistes* is highly dynamic as queen and worker castes are behavioural and determined during adulthood such that nest-mates compete to realise their reproductive potential (Reeve 1991). Moreover, reproductive options for females are varied: a female may remain as a worker all her life, or she may succeed the existing queen or found her own colony (West-Eberhard 1969, Hughes *et al*. 1987, Peters *et al*. 1995). Thus, *Polistes* provides opportunities to study the genetic basis of behavioural traits not found in honeybees, such as dominance hierarchies, queen supercedure, conflicts over egg-laying and choice of reproductive strategy. *Polistes* is the basal genus (most deep branching in a phylogeny) of a large clade (the Polistinae), making the study of these wasps especially informative because they can divulge the state of genes before the evolution and radiation of more recent and divergent groups (Arevalo *et al*. 2004, Carroll *et al*. 2005). The Polistinae is the largest family of wasps and displays almost the entire range of sociality, from the simplest forms of true eusociality (e.g. in independent nest founders like *Polistes* and *Mischocyttarus*), through to complex eusocial species (e.g. swarm founders like *Agelaia* and *Polybia*). The potential for exploring the origins, evolution and maintenance of sociality using comparative genomics within this group is therefore immense.

Acknowledgements

I would like to thanks the participants in the symposium "*Polistes* paper wasps: emergence of a model system" for helpful discussions, and Andrew Bourke, Bill Jordan and Zjef Pereboom for insightful comments on earlier drafts of the manuscript.

References

Abouheif, E. & Wray, G. A. 2002: Evolution of the gene network underlying wing polyphenism in ants. — *Science* 297: 249–252.

Albert, S. & Klaudiny, J. 2004: The MRJP/YELLOW protein family of *Apis mellifera*: identification of new members in the EST library. — *J. Insect Physiol*. 50: 51–59.

Amdam, G. V., Norberg, K., Fondrk, M. K. & Page, R. E. 2004: Reproductive ground plan may mediate colony-level selection effects on individual foraging behaviour in honey bees. — *P. Natl. Acad. Sci. USA* 101: 11350–11355.

Amdam, G. V., Norberg, K., Hagen, A. & Omholt, S. W. 2003a: Social exploitation of vitellogenin. — *P. Natl. Acad. Sci. USA* 100: 1799–1802.

Amdam, G. V., Simoes, Z. L. P., Guidugli, K. R., Norberg, K. & Omholt, S. W. 2003b: Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. — *BMC Biotechnology* 3, art. 1. doi:10.1186/1472-6750-3-1.

Arevalo, E., Zhu, Y., Carpenter, J. M. & Strassmann, J. E. 2004: The phylogeny of the social wasp subfamily Polistinae: evidence from microsatellite flanking sequencing, mitochondrial CO1 sequence and morphological char-

acters. — *BMC Evol. Biol.* 4, art. 8. doi:10.1186/1471-2148-4-3.

Arnosti, D. N. 2003: Analysis and function of transcriptional regulatory elements: Insights from *Drosophila*. — *Annu. Rev. Entomol.* 48: 579–602.

Ben-Shahar, Y., Robichon, A., Sokolowski, M. B. & Robinson, G. E. 2002: Influence of gene action across different time scales on behaviour. — *Science* 296: 741–744.

Bourke, A. F. G. & Franks, N. R. 1995: *Social evolution in ants*. — Princeton University Press, Princeton.

Bourke, A. G. F. 1999: Colony size, social complexity and reproductive conflict in social insects. — *J. Evol. Biol.* 12: 245–257.

Brazhnik, P., de la Fuente, A. & Mendes, P. 2002: Gene networks: how to put the function in genomics. — *Trends Biotechnol.* 20: 467–472.

Bromham, L. & Leys, R. 2005: Sociality and the rate of molecular evolution. — *Mol. Biol. Evol.* 22: 1393–1402.

Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. 2005: *From DNA to diversity*, 2nd ed. — Blackwell Publishing, Oxford.

Cash, A. C., Whitfield, C. W., Ismail, N. & Robinson, G. E. 2005: Behaviour and the limits of genomic plasticity: power and replicability in microarray analysis of honeybee brains. — *Genes, Brain Behav.* 4: 267–271.

Charlesworth, B. 1978: Some models of the evolution of altruistic behaviour between siblings. — *J. Theor. Biol.* 72: 297–319.

Corona, M., Estrada, E. & Zurita, M. 1999: Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee, *Apis mellifera*. — *J. Exp. Biol.* 202: 929–938.

Crosland, M. W. J. & Crozier, R. H. 1986: *Myrmecia pilosula*, an ant with only one pair of chromosomes. — *Science* 231: 1278.

Crozier, R. H. 1979: Genetics of sociality. — In: Hermann, H. R. (ed.), *Social insects*, vol. 1: 223–286. Academic Press, New York.

Crozier, R. H. 1982: On insects and insects: twists and turns in our understanding of the evolution of eusociality. — In: Breed, M., Michener, C. & Evans, H. (eds.), *The biology of social insects*: 4–9. Westview Press, Boulder, Colorado.

Crozier, R. H. 1987: Towards a sociogenetics of social insects. — In: Eder, J. & Rembold, H. (eds.), *Chemistry and biology of social insects*: 325–328. Verlag J. Peperny, Munich.

Darwin, C. 1859: *On the origin of species*. — Penguin Books, London.

Dawkins, R. 1976: *The selfish gene*. — Oxford University Press, Oxford.

Dawkins, R. 1982: *The extended phenotype*. — W.H. Freeman and Co., Oxford.

Espinosa-Jeffrey, A., Kumar, S., Zhao, P. M., Awosika, O., Agbo, C., Huang, A., Chang, R. & De Vellis, J. 2002: Transferrin regulates transcription of the MBP gene and its action synergizes with IDF-1 to enhance myelinogenesis in the md rat. — *Dev. Neurosci.* 24: 227–241.

Evans, H. & Gundersen-Rindal, D. 2003: Beenomes to *Bombyx:* future directions in applied insect genomics. — *Genome Biol.* 4: 107.

Evans, J. D. & Wheeler, D. E. 1999: Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. — *P. Natl. Acad. Sci. USA* 96: 5575–5580.

Evans, J. D. & Wheeler, D. E. 2000: Expression profiles during honeybee caste determination. — *Genome Biol.* 2: 1–6.

Evans, J. D. & Wheeler, D. E. 2001: Gene expression and the evolution of insect polyphenisms. — *BioEssays* 23: 62–68.

Fitzpatrick, M. J., Ben-Shahar, Y., Smid, H. M., Vet, L. E. M., Robinson, G. E. & Sokolowski, M. B. 2004: Candidate genes for behavioural ecology. — *Trends Ecol. Evol.* 20: 96–104.

Gadagkar, R. 1997: The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution. — *J. Genet.* 76: 167–179.

Gadau, J., Page, R. E., Werren, J. H. & Schmid-Hempel, P. 2000: Genome organisation and social evolution in Hymenoptera. — *Naturwissenschaften* 87: 87–89.

Gerlai, R. 2001: Gene targeting: technical confounds and potential solutions in behavioural brain research. — *Behav. Brain Res.* 125: 13–21.

Giray, T., Giovanetti, M. & West-Eberhard, M. J. 2005: Juvenile hormone, reproduction & worker behaviour in the neotropical social wasp *Polistes canadensis*. — *P. Natl. Acad. Sci. USA* 102: 3330–3335.

Hamilton, W. D. 1964: The genetical evolution of social behaviour. I. — *J. Theor. Biol.* 7: 1–16.

Helms-Cahan, S. & Keller, L. 2003: Complex hybrid origin of genetic caste determination in harvester ants. — *Nature* 424: 306–309.

Hojo, M., Morioka, M., Matsumoto, T. & Miura, T. 2005: Identification of soldier caste-specific protein in the frontal gland of nasute termite *Nasutitermes takasagoensis* (Isoptera: Termitidae). — *Insect Biochem. Molec. Biol.* 35: 347–354.

Hölldobler, B. & Wilson, E. O. 1990: *The ants*. — Springer Verlag, Berlin.

Hughes, C. R., Beck, M. O. & Strassmann, J. E. 1987: Queen succession in the social wasp, *Polistes annularis*. — *Ethology* 76: 124–132.

Hunt, G. J. & Page, R. E. 1995: Linkage map of the honeybee, *Apis mellifera,* based on RAPD markers. — *Genetics* 139: 1371–1382.

Hunt, J. H. & Amdam, G. V. 2005: Bivoltinism as an antecedent to eusociality in the paper wasp genus *Polistes*. — *Science* 308: 264–267.

Johnston, J. S., Ross, L. D., Beani, L., Hughes, D. P. & Kathirithamby, J. 2004: Tiny genomes and endoreplication in Strepsiptera. — *Insect Mol. Biol.* 13: 581–585.

Judice, C., Hartfelder, K. & Pereira, G. A. G. 2004: Caste-specific gene expression in the stingless bee *Melipona quadrifasciata* — are there common patterns in highly eusocial bees? — *Insectes Soc.* 51: 352–358.

Julian, G. E., Fewell, J. H., Gadau, J., Johnson, R. A. & Larrabee, D. 2002: Genetic determination of the queen caste in an ant hybrid zone. — *P. Natl. Acad. Sci. USA* 99: 8157–8160.

Keller, L. 1997: Indiscriminate altruism: unduly nice parents and siblings. — *Trends Ecol. Evol.* 12: 99–103.

Keller, L. (ed.) 1999: *Levels of selection.* — Princeton University Press, Princeton.

Krieger, M. J. B. & Ross, K. G. 2002: Identification of a major gene regulation complex social behaviour. — *Science* 295: 328–332.

Kucharski, R. & Maleszka, R. 1998: A royal jelly protein in expresses in a subset of Kenyon cells in the mushroom bodies of the honey bee brain. — *Naturwissenschaften* 85: 343–346.

Kucharski, R. & Maleszka, R. 2002a: Evaluation of differential gene expression during behavioural development in the honeybee using microarrays and northern blots. — *Genome Biol.* 3 art. 2. doi:10.1186/gb-2002-3-2-research0007.

Kucharski, R. & Maleszka, R. 2002b: Molecular profiling of behavioural development: differential expression of mRNAs for inositol 1,4,5-triphosphate 3-kinase isoforms in naive and experienced honeybees (*Apis mellifera*). — *Mol. Brain Res.* 99: 92–101.

Kucharski, R. & Maleszka, R. 2003: Transcriptional profiling reveals mutilfunctional roles for transferrin in the honeybee, *Apis mellifera*. — *J. Insect Sci.* 3: 1–8.

Li, W.-H., Gu, Z., Wnag, H. & Nekrutenko, A. 2001: Evolutionary analyses of the human genome. — *Nature* 409: 847–849.

Linksvayer, T. A. & Wade, M. J. 2005: The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects and heterochrony. — *Q. Rev. Biol.* 80: 317–336.

Michod, R. E. 1982: The theory of kin selection. — *Annu. Rev. Ecol. Syst.* 13: 23–55.

Miura, T. 2001: Morphogenesis and gene expression in the soldier-caste differentiation of termites. — *Insectes Soc.* 48: 216–223.

Miura, T. 2004: Proximate mechanisms and evolution of caste polyphenism in social insects: from sociality to genes. — *Ecol. Res.* 19: 141–148.

Miura, T., Koshikawa, S. & Matsumoto, T. 2003: Winged presoldiers induced by a juvenile hormone analog in *Zootermopsis nevadensis:* implications for plasticity and evolution of caste differentiation in termites. — *J. Morphol.* 257: 22–32.

Miura, T., Kamikouchi, A., Takeuchi, H., Natori, S., Kubo, T. & Matsumoto, T. 1999: Soldier caste-specific gene expression in the mandibular glands of *hodotermopsis japonica* (Isoptera: Termopsidae). — *P. Natl. Acad. Sci. USA* 96: 13874–13879.

Normark, B. B. 2003: The evolution of alternative genetic systems in insects. — *Annu. Rev. Entomol.* 48: 397–423.

Page, R. E., Gadau, J. & Beye, M. 2002: The emergence of Hymenoptera genetics. — *Genetics* 160: 375–379.

Pereboom, J. J. M., Jordan, W. C., Sumner, S., Hammond, R. L. & Bourke, A. F. G. 2005: Differential gene expression in queen-worker caste determination in bumblebees. — *Proc. R. Soc. Lond. B* 272: 1145–1152.

Peters, J. M., Queller, D. C., Strassmann, J. E. & Solis, C. R. 1995: Maternity assignment and queen replacement in a social wasp. — *Proc. R. Soc. Lond. B* 260: 7–12.

Piulachs, M. D., Guidugli, K. R., Barchuk, A. R., Cruz, J., Simoes, Z. L. P. & Belles, X. 2003: The vitellogenin of the honey bee, *Apis mellifera*: structural analysis of the cDNA and expression studies. — *Insect Biochem. Molec. Biol.* 33: 459–465.

Queller, D. C. 1992: A general model for kin selection. — *Evolution* 46: 376–380.

Queller, D. C., Hughes, C. R. & Strassmann, J. E. 1990: Wasps fail to make distinctions. — *Nature* 344: 388.

Raser, J. M. & O'Shea, E. K. 2005: Noise in gene expression: origins, consequences and control. — *Science* 309: 2010–2013.

Reeve, H. 1991: Polistes. — In: Ross, K. G. & Matthews, R. W. (eds.), *The social biology of wasps*: 99–148. Cornell University Press, Ithaca.

Robinson, G. E. 1999: Integrative animal behaviour and sociogenomics. — *Trends Ecol. Evol.* 14: 202–205.

Robinson, G. E. 2002a: Genomics and integrative analysis of division of labour in honeybee colonies. — *Am. Nat.* 160: S160–S172.

Robinson, G. E. 2002b: Sociogenomics takes flight. — *Science* 297: 204–205.

Robinson, G. E. & Ben-Shahar, Y. 2002: Social behaviour and comparative genomics: new genes or new gene regulation? — *Genes, Brain Behav.* 1: 197–203.

Robinson, G. E., Grozinger, C. M. & Whitfield, C. W. 2005: Sociogenomics: social life in molecular terms. — *Nat. Rev. Genet.* 6: 257–270.

Röseler, P.-F. 1991: Roles of morphogenetic hormones in caste polymorphism in bumblebees. — In: Gupta, A. P. (ed.), *Morphogenetic hormones of arthropods. Roles in histogenesis, organogenesis and morphogenesis*: 384–399. Rutgers University Press, New Brunswick.

Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R. & Bennett, G. W. 2003: Caste- and development-associated gene expression in a lower termite. — *Genome Biol.* 4, R62. doi:10.1186/gb-2003-4-10-r62.

Scharf, M. E., Wu-Scharf, D., Zhou, X., Pittendrigh, B. R. & Bennett, G. W. 2004: Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. — *Insect Mol. Biol.* 14: 31–44.

Schmitz, J. & Moritz, R. F. A. 1998: Sociality and the rate of rDNA sequence evolution in wasps (Vespidae) and honeybees (Apis). — *J. Mol. Evol.* 47: 606–612.

Sherman, P. W. 1979: Insect chromosome numbers and eusociality. — *Am. Nat.* 113: 925–935.

Sumner, S., Pereboom, J. J. M. & Jordan, W. C. 2006: Differential gene expression and phenotypic plasticity in behavioural castes of the primitively eusocial wasp, *Polistes canadensis*. — *Proc. R. Soc. Lond B* 273: 19–26.

The Honeybee Genome Sequencing Consortium 2006: Insights into social insects from the genome of the honeybee *Apis mellifera*. — *Nature* 443: 931–949.

Thorne, B. L. 1997: Evolution of eusociality in termites. — *Annu. Rev. Ecol. Syst.* 28: 27–54.

Tian, H., Vinson, B. & Coates, C. J. 2004: Differential gene expression between alate and dealate queens in the red imported fire ant, *Solenopsis invicta* Buren (Hymenoptera: Formicidae). — *Insect Bioche. Molec. Biol.* 34:

937–949.

Toma, D. P., Bloch, G., Moore, D. & Robinson, G. E. 2000: Changes in period mRNA levels in the brain and division of labor in honeybee colonies. — *P. Natl. Acad. Sci. USA* 97: 6914–6919.

Trivers, R. L. & Hare, H. 1976: Haplodiploidy and the evolution of the social insects. — *Science* 191: 249–263.

Turillazzi, S. & West-Eberhard, M. J. 1996: *Natural history and evolution of paper wasps*. — Oxford University Press, Oxford.

Volny, V. P. & Gordon, D. M. 2002: Genetic basis for queen-worker dimorphism in a social insect. — *P. Natl. Acad. Sci. USA* 99: 6108–6111.

Wade, M. J. 1980: Kin selection: its components. — *Science* 210: 665–667.

Wcislo, W. T. & Danforth, B. N. 1997: Secondarily solitary: the evolutionary loss of social behaviour. — *Trends Ecol. Evol.* 12: 468–474.

Wei, L., Liu, Y., Dubchak, I., Shon, J. & Park, J. 2002: Comparative genomics approaches to study organism similarities and differences. — *J. Biomed. Inform.* 35: 142–150.

West-Eberhard, M. J. 1969: The social biology of polistine wasps. — *Misc. Publ. Mus. Zool. Univ. Michigan* 140: 1–101.

West-Eberhard, M. J. 1987: Flexible strategy and social evolution. — In: Ito, Y., Brown, J. L. & Kikkawa, J. (eds.), *Animal societies: theories and facts*: 35–51. Japan Scientific Societies Press, Tokyo.

West-Eberhard, M. J. 1989: Phenotypic plasticity and the origins of diversity. — *Annu. Rev. Ecol. Syst.* 20: 249–278.

West-Eberhard, M. J. 1996: Wasp societies as microcosms for the study of development and evolution. — In: Turillazzi, S. & West-Eberhard, M. J. (eds.), *Natural history and evolution of paper-wasps*: 290–317. Oxford University Press.

Wheeler, D. E. 1986: Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. — *Am. Nat.* 128: 13–34.

Whitfield, C. W., Cziko, A. M. & Robinson, G. E. 2003: Gene expression profiles in the brain predict behaviour in individual honey bees. — *Science* 302: 296–299.

Wilson, E. O. 1971: *The insect societies*. — Harvard University Press, Cambridge, MA.

Appendix 1. A summary of the genes expressed in 2 or more species of social insect with respect to caste. The comparison made is indicated for each species. Direction of expression is indicated by named caste; Equal (Eq) = no difference in expression detected between castes; 0 = no expression detected; N = no data available. Reference sources given in text. Q = queen, W = worker, F = forager, S = soldier, Y = Young.

Putative gene	Larvae				Adults								
	B. terrestris	*A. mellifera*	*S. invicta*	*P. canadensis*	*B. terrestris*		*A. mellifera* (workers)		*S. invicta*		*R. flavipes*		
	Q *vs.* W (early & late instars)	Q *vs.* W	Q *vs.* W	Q *vs.* F *vs.* Young female	Q *vs.* W	Reproductive W *vs.* non-reproductive W	Nurse *vs.* F	Naïve *vs.* F	Q *vs.* W	Delate *vs.* alate Q	S *vs.* W	Alate Q *vs.* W	Sub-reproductive Q *vs.* W
18S ribosomal protein	N	N	N	N	N	N	N	Naïve	N	N	W	Eq	Sub
28S ribosomal protein	N	Q	N	N	N	N	N	N	N	N	Eq	Alate	Sub
Ribosomal protein	Q(early)/W(late)	W	N	Q/Young	W	Reprod	N	N	N	N	N	N	N
ATP synthase	Q(early)/W(late)	Q	W	F	W	N	N	Naïve	Eq	Eq	N	N	N
CG31605	0	N	N	Young	0	0	N	N	N	N	N	N	N
Cuticle protein	Q(early)/W(late)	Q	N	N	0	N	N	N	N	N	N	N	N
COX I	Q(early)/ W(late)	Q	Eq	N	W	Non-reprod	N	Naïve	Eq	Eq	S	N	N
COX II	N	Q	W	F/Q	N	N	N	N	Eq	Dealate	N	N	N
Heat shock protein 70	N	Eq	N	F	N	N	F	Naïve	N	N	Eq	Eq	Eq
Hexamerin I	N	N	N	N	N	N	N	N	N	N	W	W	Sub
Hexamerin II	W(late)	W	N	Q	0	0	N	N	N	Eq	W	W	W
Imaginal disc	0	N	N	F/Q	0	0	F	Naïve	N	N	N	N	N
Lectin like	N	N	N	N	N	N	Nurse	Naïve	N	N	N	N	N
MRJP	N	N	0	Q	N	N	N	F	Q	Dealate	N	N	N
Peroxiredoxin	Q(early)/W(late)	W	N	F/Young	Q	Non-reprod	N	N	N	N	N	N	N
SPARC	N	N	N	N	N	N	Nurse	Naïve	N	N	N	N	N
Transferrin	N	N	N	Q	N	N	F	Virgin Q	N	N	N	N	N
Unknown 1	N	Q	N	Q	N	N	N	N	N	N	N	N	N
Vitellogenin	N	N	Eq	Q	N	N	Nurse	N	Dealate	Dealate	Eq	N	N

I included only the 50 most predictive ESTs for nurse and foraging honeybees from Whitfield *et al.* 2003. COX I & COX II = cytochrome oxidase I & II respectively. MRJP = major royal jelly protein.