Minireview

Getting a buzz out of the bee genome

Michael Ashburner* and Charalambos P Kyriacou†

Addresses: *Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK. †Department of Genetics, University of Leicester, Leicester LE1 7RH, UK.

Correspondence: Michael Ashburner. Email: ma11@gen.cam.ac.uk

Published: 26 October 2006

Genome ***Biology*** 2006, **7:**239 (doi:10.1186/gb-2006-7-10-239)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2006/7/10/239

Abstract

The honey bee *Apis mellifera* displays the most complex behavior of any insect. This, and its utility to humans, makes it a fascinating object of study for biologists. Such studies are now further enabled by the release of the honey-bee genome sequence.

We have long looked forward to the sequencing of the genome of the honey bee, for now we may uncover the genetic basis of divination: Bees "have too the power of divination, so that they know in advance when rain or frost are coming" (Aelian, *On Animals* I, 11). Unfortunately, the Honey Bee Genome Sequencing Consortium (HBGSC) has not yet discovered the divination gene in the 236 megabases of the clonable bee genome [1]. But much that is fascinating has been discovered, and this paper will be a landmark, not only in genomics, but also in bee research. Honey bees have been exploited by humans for millennia, and their extraordinary behavior and biology have always intrigued and puzzled us. The achievement of sequencing the bee genome, by a team at the Baylor College of Medicine collaborating closely with the honey-bee research community, will provide an enormous boost to our understanding of some fascinating biology.

Surprises from the genome

The genome of the honey bee will inevitably be compared to that of the fruit fly *Drosophila melanogaster*. Inevitably, because so far we have the genomes of only two other orders - Diptera (*Drosophila*) and Lepidoptera (the silkworm *Bombyx mori*) - of the 30 or so orders of insects (the honey bee belongs to the Hymenoptera). Members of three other orders - Coleoptera (beetles), Anoplura (lice) and Heteroptera (bugs) - will soon join this group. At a coarse level, the genomes of fly and bee are quite different: that of the bee is relatively AT-rich, a fact that posed a technical problem to the sequencers, and, even more remarkably, the genes themselves are in regions that average 71% AT; in *Drosophila* the genes are on average 56% AT. The HBGSC suggests that this difference may be a consequence of cytosine methylation in the honey bee, as unlike *Drosophila*, the bee genome contains members of all three known families of cytosine-5-methyltransferase genes; indeed, it has two genes from the Dnmt1 family of genes. The presumption is that high levels of cytosine methylation, which tend to repress gene expression, have led to the preferential selection of AT-rich regions as a more favorable context for genes. If so, one might expect the bee genome to be deficient in the dinucleotide CpG; the paradox is that this genome has the highest CpG over-representation (by 1.67-fold) of any known genome. Although there is direct experimental evidence for some CpG methylation in honey bees [2], neither its extent, nor its significance, is yet known.

Another surprise of the bee genome is its complement of transposable elements, which comprise only 1% of the sequenced genome - in contrast to 5.3% of the euchromatic genome of *D. melanogaster* [3]. Even more surprising is that this 1% is almost entirely made up of members of the *mariner* family, which transpose by simple excision and reintegration. Retrotransposable elements, a common feature of most metazoan genomes, are represented by only a small number of very degraded sequences. Whether or not this is a consequence of the haploidy of male bees, as suggested by the HBGSC, is an open question. The other group in which

retrotransposable elements are known to be absent are the fully parthenogenetic bdelloid rotifers (see [4]).

Sex determination in Hymenoptera

Like most Hymenoptera, honey bees have an extraordinary sex-determining mechanism known as haplo-diploidy: females are normally diploid and a product of sexual congress; males are haploid and develop parthenogenetically from unfertilized eggs [5]. The study of the genetic basis of this mechanism of sex determination in honey bees had to await the development of artificial insemination; otherwise it is impossible to do controlled crosses, a fact that, despite his efforts, defeated Gregor Mendel [6]. It was the great, but much underappreciated, geneticist P.W. Whiting who, working with a more tractable hymenopteran, *Bracon hebetor*, discovered this mechanism. There is a sex-determining locus with many alleles; heterozygous zygotes develop as females, hemizygous or homozygous zygotes develop as males [7]. This hypothesis was confirmed for honey bees by Woyke [8] and the *complementary sex determiner* (*csd*) gene was cloned by Beye and colleagues in 2003 [9]. The product of *csd* is an RNA-binding protein and it may, like the Transformer protein in *Drosophila,* control sex by determining the splicing pattern of the *doublesex* gene. Population studies of the sequence of *csd* show that polymorphism of this gene, essential for sex determination, is maintained by balancing selection [10].

The development of diploid honey-bee zygotes may follow one of two paths: to sterile workers who devote their lives to collecting nectar and pollen and taking care of the next generation; or to queens who, after a brief mating flight, have a life of leisure laying eggs. The genome sequence of the honey bee will provide a valuable resource for the detailed analysis of differences in gene expression between these castes. Early data from relatively small cDNA libraries already indicate major differences in intermediary metabolism between workers and queens (for example, see [11]). The role of nutrition in determining caste development in honey bees has been known for over 200 years (see [12]), and Wheeler *et al.* [13] have used the official gene list from the HBGSP [1] to implicate the insulin-signaling pathway in this developmental decision.

Shedding light on bee behavior

The rich behavioral repertoire of social bees compared to that of the Diptera has often been invoked to explain the long-established observation that the hymenopteran brain has a dramatic expansion of the mushroom body region. This paired protocerebral structure has 170,000 intrinsic neurons (called Kenyon neurons) per hemisphere in the adult honey bee [14], compared to a mere 2,500 in *Drosophila* [15]. In fact, about 15% of bee neurons are dedicated to the mushroom bodies compared to only around 1% in the fly, underscoring the enhanced role of these neural structures in bee behavior. The mushroom bodies have been much studied in *Drosophila,* and appear particularly important for integrating sensory information, especially in the context of olfaction [16].

Making and strengthening connections between unconditioned and conditioned stimuli during olfactory learning is a major role of the mushroom bodies in *Drosophila* [17], and so it seems reasonable to assume that much of the seemingly more complicated social behavior of *Apis* may be mediated by this brain center. In support of this view is the observation that odorant receptors are among the gene families most over-represented in *Apis* compared with the fly [1]. Thus we might guess that the duplication of odorant receptor genes provided a driving force for an exponential enlargement of the brain regions that deal with the extra demands of the huge increase in potential olfactory associations. This enhanced neural plasticity may have led to the retention in Hymenoptera of genes such as *Mahya*, which is also found in vertebrates but has been lost from Diptera and Lepidoptera. This gene encodes a secreted protein that is expressed in the bee mushroom bodies and antennal lobes, and in vertebrates is present in the olfactory bulb, the structure that shares the same function as antennal lobes in bees, namely the processing and integration of olfactory information. These observations provide an intriguing association between the presence of this gene, its anatomical site of expression, and species with higher cognitive functions [18].

In contrast, the gene *foraging* (*for*), which encodes a cGMP-dependent protein kinase (PKG), is found in both flies and bees and, as its name suggests, is implicated in behavioral strategies for food searching in both organisms [19,20]. In bees, *for* is expressed in the lamina of the optic lobes and also in a region of the mushroom bodies that receives visual information. Nurse bees age to become foragers when levels of *for* rise significantly in these brain regions, and these (now) foraging worker bees become positively phototactic. They then leave the darkness of the hive to become *bona fide* foragers [21]. In flies, however, ablation of the mushroom bodies in the larva does not affect food searching [19], so an additional level of regulation via these structures has clearly been recruited in the honey bee, further underscoring their critical neurogenic role at the interface between genome evolution and complex social behavior.

Rhythms in evolution

The honey bee also misled one of us (C.P.K.) for several years about how one of the canonical circadian clock genes evolved. In 2000, it was revealed that flies and moths have two '*timeless*' genes - the one first discovered and called *timeless* (*tim*), which has a cardinal role in the 24-hour clock, and *tim2* (or *timeout*), which apparently was the only

tim-like sequence found in mammals, nematodes, and other animals [22,23]. Thus it appeared that a relatively recent duplication had occurred in the ancestors of Lepidoptera and Diptera around 300 million years ago, and that *tim* had evolved rapidly to take on a dedicated circadian role. This view was further strengthened by the fact that mutations in *tim2* in mammals or nematodes were lethal [24,25], whereas mutating *tim* in *Drosophila* led to healthy, albeit arrhythmic, flies, revealing *tim* to be a dedicated 'behavioral' rather than a 'developmental' gene [26]. As the years crept by, peeking at the emerging bee genome did not reveal *tim*, but did reveal *tim2* - the ancestral form of *tim*. This was consistent with a scenario of a relatively recent duplication of *tim2* to generate the clock-relevant *tim* in the ancestors of Lepidoptera and Diptera. This cosy story has been rudely demolished, however, as the *tim* sequences have recently been identified in the beetle *Tribolium* and, even more surprisingly, in sea urchins [27]. This puts back the date for the duplication of *tim* to pre-Cambrian times.

The genes that we presume encode the circadian clockworks of honey bees show a number of other interesting features, apart from *tim* evolution, in that their genes seem to be more mouse-like than fly-like. For example, in flies and mice, the *Clock* (*Clk*) and *cycle* (*cyc*, also called *Bmal1*) genes encode positive transcription factors that directly regulate the negative autoregulators encoded by *period* and *tim*. In flies, the abundance of *Clk* mRNA cycles with a circadian rhythm but *cyc* is expressed constitutively, whereas in the mammal, *cyc* cycles and *Clk* does not [28]. As if to highlight this species difference, the carboxy-terminal transactivation domain found in fly Clk protein has been transposed to mouse Cyc.

Flies also have a dedicated circadian photoreceptor, encoded by the *cryptochrome* (*cry*) gene, whereas mammals have two *Cry* genes, which act as negative transcriptional regulators, not photoreceptors [28]. Nevertheless, the single copy of *Cry* in the bee encodes sequences more reminiscent of the mammalian than the fly protein, suggesting that the bee Cry protein also functions as a negative regulator, not a photoreceptor [1]. In fact, Lepidoptera have two copies of *Cry*; one acts as a negative regulator, the other probably acts as a photoreceptor [29]. Thus basal lineages probably had two types of *Cry* and two types of *tim,* and different organisms appear to have mixed, matched and eliminated one or other copy of these two genes according to their needs. Lepidoptera kept both types for each of their *tim* and *cry* genes, with both types of functions apparent for each gene [29]. Bees, on the other hand, have the stripped-down version, and have lost one copy of each gene, maintaining obligatory *tim* developmental, and non-photoreceptor *Cry* function [27]. Mammals kept developmental *tim,* but both *Cry* genes lost photoreceptor function [28]. *Drosophila* kept both *tim* genes, but only the photoreceptor *cry* [22,29]. Evolution surely plays tricks on the unwary biologist.

The sting in the tail

Most of us have, at one time or another, been stung by a honey bee. Reading the account of the venoms predicted from the genome sequence [1] makes it quite clear why these stings are so painful: bee venom contains perhaps 20 different allergens including "several homologues of scorpion and snake venoms". The domesticated European honey bee (*Apis mellifera ligustica*) is not, thankfully, very aggressive, but the African *A. mellifera scutellata*, introduced to Brazil by Warwick Kerr 40 years ago [30], is (see Bill Hamilton's amusing account of their attack [31]). One of the consequences of the honey bee genome project is a very dense map of single-nucleotide polymorphisms (SNPs), with nearly 5,500 SNPs already identified and mapped [1]. These have already been used to study the four major groups of subspecies of *A. mellifera*, with the surprising result that the Eastern (*A. mellifera ligustica*) and Western (*A. mellifera mellifera*) European populations result from independent colonizations of Europe by African populations.

Bee researchers, like their colleagues who work with *Drosophila,* will now distinguish the BG (Before the Genome) and AG (After the Genome) epochs. We can confidently predict that honey-bee research will now be even more vibrant and interesting than BG, with great consequences for both fundamental and applied biology.

Acknowledgements

C.P.K. thanks the Royal Society for a Wolfson Research Merit Fellowship. M.A. thanks the MRC for a quarter of a century's continuous research funding. We both thank George Weinstock (Baylor College of Medicine) for a preprint of the Honey Bee Genome Sequencing Consortium's paper and Gene Robinson (University of Illinois at Urbana-Champaign) and Richard Gibbs (Baylor College of Medicine) for critical comments on our manuscript. We also thank Diana Wheeler (University of Arizona, Tucson) for a preprint of her paper.

References

1. Honey Bee Genome Sequencing Consortium*:* **The genome of a highly social insect, the honey bee *Apis mellifera*.** *Nature* 2006, **443:**931-949.
2. Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE: **Functional CpG methylation system in a social insect.** *Science*, in press.
3. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D: **Combined evidence annotation of transposable elements in genome sequences.** *PLoS Comput Biol* 2005, **1:**166-175.
4. Arkhipova I, Meselson M: **Deleterious transposable elements and the extinction of asexuals.** *BioEssays* 2005, **27:**76-85.
5. Dzierzon J: **Gutachten über die von Herrn Direktor Stöhr im ersten und zweiten Kapitel des General-Gutachtens aufgestellten Fragen.** *Eichstädter Bienenzeitung* 1845, **1:**109-113; 119–121.
6. Kühne F: **Miscellen**. *Ungarische Bien,* Temesvar 1881, **1881:**2-8, 16-22, 25-26.
7. Whiting PW: **Multiple alleles in complementary sex determination of *Habrobracon*.** *Genetics* 1943, **28:**365-382.
8. Woyke J: **Sex determinatio**n. In *Bee Genetics and Breeding*. Edited by Rinderer TE. Orlando, FL: Academic Press; 1986: 91-119.
9. Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW: **The gene *csd* is the primary signal for sexual development in the honey bee and encodes an SR-type protein.** *Cell* 2003, **114:** 419-429.

10. Hasselmann M, Beye M: **Signatures of selection among sex-determining alleles of the honey bee.** *Proc Natl Acad Sci USA* 2004, **101:**4888-4893.
11. Evans JD, Wheeler DE: **Expression profiles during honeybee caste determination.** *Genome Biol* 2000, **2:**research0001.1-00001.6.
12. Huber F: *New Observations on the Natural History of Bees.* Edinburgh: 1821. Third English edition of *Nouvelles Observations sur les Abeilles.* Geneva, 1802.
13. Wheeler DE, Buck N, Evans JD: **Expression of insulin pathway genes during the period of caste determination in the honeybee, *Apis mellifera.*** *Insect Mol Biol*, in press.
14. Witthoft W: **Absolute Anzahl und Verteilung der Zellen im Him der Honigbiene.** *Z Morph Tiere* 1967, **61:**160-184.
15. Balling A, Technau GM, Heisenberg M: **Are the structural changes in the adult *Drosophila* mushroom bodies memory traces? Studies on biochemical learning mutants.** *J Neurogenet* 1987, **4:**65–73.
16. Margulies C, Tully T, Dubnau J: **Deconstructing memory in *Drosophila*.** *Curr Biol* 2005, **15:**R700-R713.
17. Davis RL: **Olfactory memory formation in *Drosophila*: from molecular to systems neuroscience.** *Annu Rev Neurosci* 2005, **28:**275-302.
18. Tsuchimoto M, Yasuo S, Funada M, Aoki M, Sasagawa H, Yoshimura T, Tadauchi O, Cameron SA, Kitagawa Y, Kadowaki T: **Conservation of novel *Mahya* genes shows the existence of neural functions common between Hymenoptera and Deuterostomes.** *Dev Genes Evol* 2005, **215:**564-574.
19. Osborne KA, Robichon A, Burgess E, Butland S, Shaw RA, Coulthard A, Pereira HS, Greenspan RJ, Sokolowski MB: **Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*.** *Science* 1997, **277:**834–836.
20. Osborne KA, de Belle JS, Sokolowski MB: **Foraging behaviour in *Drosophila* larvae: mushroom body ablation.** *Chem Senses* 2001, **26:**223-230.
21. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE: **Influence of gene action across different time scales on behavior.** *Science* 2002, **296:**741-744.
22. Benna C, Scannapieco P, Piccin A, Sandrelli F, Zordan M, Rosato E, Kyriacou CP, Valle G, Costa R: **A second *timeless* gene in *Drosophila* shares greater sequence similarity with mammalian *tim*.** *Curr Biol* 2000, **10:**R512-R513.
23. Gotter AL, Manganaro T, Weaver DR, Kolakowski LF Jr, Possidente B, Sriram S, MacLaughlin DT, Reppert SM: **A time-less function for mouse *Timeless*.** *Nat Neurosci* 2000, **3:**755-756.
24. Li Z, Stuart RO, Qiao J, Pavlova A, Bush KT, Pohl M, Sakurai H, Nigam SK: **A role for Timeless in epithelial morphogenesis during kidney development.** *Proc Natl Acad Sci USA* 2000, **97:**10038–10043.
25. Chan RC, Chan A, Jeon M, Wu TF, Pasqualone D, Rougvie AE, Meyer BJ: **Chromosome cohesion is regulated by a clock gene paralogue TIM-1.** *Nature* 2003, **423:**1002-1009.
26. Sehgal A, Price, JL Man B, Young MW: **Loss of circadian behavioral rhythms and *per* RNA oscillations in the *Drosophila* mutant *timeless*.** *Science* 1994, **263:**1603-1606.
27. Rubin R, Shemesh Y, Cohen M, Elgavish S, Robertson HM, Bloch G: **Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock.** *Genome Res* 2006, **16:**1352-1365.
28. Reppert SM, Weaver DR: **Coordination of circadian timing in mammals.** *Nature* 2002, **418:**935-941.
29. Zhu H, Yuan Q, Briscoe AD, Froy O, Casselman A, Reppert SM: **The two CRYs of the butterfly.** *Curr Biol* 2005, **15:**R953-R954.
30. Kerr WW: **The history of the introduction of African bees to Brazil.** *S Afr Bee J* 1967, **39:**3-5.
31. Hamilton WD: **My intended burial and why.** *Ethol Ecol Evol* 2000, **12:**111-122. Reprinted in *Narrow Roads of Gene Land,* Vol. 3. Edited by Ridley M. Oxford: Oxford University Press; 2005.