Proposal for the Sequencing of a New Target Genome: White Paper for a Honey Bee Genome Project

The Honey Bee Genome Sequencing Consortium[A]

[A]Represented by the authors of this document: G.E. Robinson[B], K. Aronstein, J.E. Evans, S.E. Fahrbach, J.S. Johnston, R. Maleszka, R.E. Page, H.M. Robertson, D.B. Weaver[C]

[B]Contact information: Ph. 217-265-0309; Fax 217-244-3499; generobi@life.uiuc.edu

Summary. *Homo sapiens* is a highly social species and social interactions are critical determinants of human mental and physical health. We propose to sequence the genome of another highly social species, the honey bee, *Apis mellifera*. Though phylogenetically distant, honey bees live in societies that rival our own in complexity, internal cohesion, and success in dealing with the myriad challenges posed by social life, including those related to communication, aging, social dysfunction and infectious disease. A honey bee genome sequencing project (HBGP) will benefit human health and medicine in diverse areas, including venom toxicology, allergic disease, mental illness, infectious disease, parasitology and gerontology. In addition, the HBGP will improve human nutrition by enabling enhanced pollination of food plants and accelerated delivery of hymenopteran parasitoids for biological control of pests. The HBGP will also improve honey bee sentinel function, providing enhanced capabilities for detection and location of chemical and biological agents of harm. Sequencing the genome of the honey bee, a beneficial, non-dipteran, insect endowed with a small brain but cognitive sophistication, with complex social organization but amenable to molecular, genetic, neural, and ecological manipulation, will provide important tools and unique models to improve human health. When these benefits are balanced against the costs of sequencing a 270MB genome, the HBGP promises to provide a valuable and economical resource.

Table of Contents

A. Specific biological rationales for the utility of new sequence data

1. Improving human health: *Novel antibiotics.* Increased drug resistance by pathogenic bacteria has created an urgent demand for new antibiotics. Insects are among the more promising sources of novel antibiotics[1] and honey bees likely offer a rich source because of their sociality. Like humans, honey bees live in a social environment with nearly ideal conditions for growth and transmission of pathogens. Food is constantly shared among individuals, the beehive is maintained at a temperature of 33ºC (93ºF) and 95% relative humidity, and population densities are mind-boggling (as many as 50,000 adults and 50,000 juveniles at densities equivalent to ca. 15 adult humans in a 6 x 4 m apartment). Although afflicted with many diseases, honey bees must have evolved many powerful antibacterial peptides to cope with the huge number of pathogens that would thrive in such conditions. Interest in this topic is increasing[2], but a HBGP is necessary for efficient genomic bio-prospecting.

Infectious disease. Humans show both antigen-specific and innate immune responses to important pathogens including *Mycobacterium tuberculosis* and *Streptococcus pneumoniae*[3]. Better understanding of innate immunity can help counter these diseases, especially when vaccines have limited effectiveness[4]. Non-human models, especially insects, are very useful; immunity is phylogenetically ancient, and defensive strategies are highly conserved at the molecular leve[5]. Genomic surveys of *Drosophila* are used to understand innate and cellular immune-responses[6] and pharmacogenomic companies are exploiting this approach[7]. With a sequenced genome, honey bees would be an excellent model for elucidating immune system function; the natural pathogens and parasites of honey bees are well known because they cause substantial economic loss[8]. Prior knowledge of bee pathogens and parasites, and the possibility of generating sequence information for them, could provide a platform for genomic analyses of natural co-evolving species communities. "Community genomics" promises to provide new epidemiological and mechanistic insights into human infectious diseases. *A* HBGP also will provide information on parasite resistance, as the DNA source for the HBGP is a strain resistant to *Varroa destructor,* a serious bee parasite[9]. This selected bee strain suppresses *Varroa* reproduction via as yet unknown mechanisms. Identification of genes related to malaria transmission in the recently completed *Anopheles gambiae* genome will also be facilitated by a HBGP. Mosquitoes and *Drosophila* are both Diptera; sequence from the more distantly related honey bee (Hymenoptera) will add power to comparative genomic analyses.

Bee venom, anaphylaxis and human allergic disease. Honey bees defend their hive aggressively with both sophisticated behavioral and biochemical mechanisms. Bee venom has a wide range of medically important and pharmacologically active compounds. Several of them already have been identified, notably melittin and apamin, with outstanding therapeutic potential for cancer, sleep disorders, learning and memory enhancement, Parkinson's disease, HIV and AIDS associated dementia, schizophrenia, and novel non-viral vector development for gene therapy[10]. But other venom components remain to be identified[11]. Because honey bees have had intense evolutionary pressure from mammalian predators, it is likely that bee venom contains other compounds with similar human therapeutic potential.

Bee venom also elicits sting-induced anaphylaxis and other serious allergic reactions in susceptible humans. Though the incidence of mortality from bee stings is low, bee sting allergies remain a common and expensive health care issue. When considered together, allergies and other hyperinflammatory immune responses constitute one of the major problems of modern medicine[12]. The response to bee stings has historically been a preeminent model in the study of allergic disease (see Hamilton letter), which affects ca. 1/3 of the general population[13]. Some bee venom proteins have been described but many components remain obscure, particularly the high molecular weight fraction that elicits strong B and T cell responses[14]. Identification of putative human atopy genes, the description of core molecular features of allergic disease etiology, and the search for therapeutic compounds also would benefit from a complete characterization of honey bee venom. Genomic sequence information can lead to efficient bio-prospecting, even after a system has been subjected to extensive protein analyses[15].

Bee sting health problems will increase in the southern and western parts of the US that have been invaded by the more aggressive Africanized Honey bee (AHB), especially as it increases in population density in urban areas[16]. A HBGP will aid programs to identify and reduce AHB populations and mitigate AHB impacts, perhaps by providing a tool to identify the genes responsible for a complex behavioral trait that affects human health – colony defense and stinging. In addition, massive envenomation by bees and related species causes a newly recognized syndrome of rhabdomyolysis, renal failure and hemolysis[17]. Its pathophysiology is poorly understood but may be relevant to other common disorders that share important features such as bacterial sepsis, hemolytic-uremic syndrome, and fibrinolysis.

Nutrition. Honey bees are the premier beneficial insect worldwide. While best known for honey, the honey bee's more critical contribution to human nutrition is crop pollination, valued at nearly $15 billion/year in the US[18]. Pollination increases the quantity and quality of fruits, nuts, and seeds, many of them increasingly recognized as sources of nutraceuticals. But parasites and pathogens compromise bee health and pollination activities. Adding to the problem, exotic parasites have decimated feral honey bees, and increasing insecticide use and ecosystem disturbance have reduced native pollinator populations. These problems threaten to decrease insect pollination and reduce food quantity and quality. A HBGP will help to breed bees that resist disease and insecticides, pollinate more efficiently, but sting less. Another benefit will be information to enhance the use of parasitic wasps (Hymenoptera, like bees) in the biological (non-pesticide) control of agricultural pests, an industry currently worth over $20 billion/year[19] (see Appendix).

Mental health. Some forms of mental illness, such as autism, involve problems with social integration[20]. The basics of how individuals respond to their social environment (sensory structures, signal transduction cascades, various forms of neural plasticity) are highly conserved across phyla. Bees show a high degree of social integration, and their activities are highly dependent upon their ability to read social cues; identification of several well-defined sets of social cues make for unusually tractable experimental social systems[21]. Combined with a HBGP and the highest known genetic recombination rate of any animal, this provides the platform for mapping complex behavioral traits, including those related to social integration.

Biosensors. A HBGP also may enhance use of honey bees as environmental sentinels. Honey bees evolved as efficient explorers, canvassing and exploiting areas of several square miles around their hive. As such, honey bees function as a comprehensive array of autonomous biosensors, capable of reporting the presence, location and concentration of environmental toxins[22]. Preliminary evidence[23] suggests bees can be trained to locate substances used in various types of warfare, and bees have been deployed in ongoing DARPA research to detect biological and chemical weapons. These security-related activities might be aided by "tuning" bee detection capabilities with information obtained from the identification of genes involved in olfaction, e.g., olfactory receptor genes, which are very difficult to find without extensive genome sequence information due to rapid evolutionary sequence divergence[24].

X chromosome diseases. Mutations on the X-chromosome are responsible for many serious conditions, including Turner's syndrome, Trisomy-X, Kleinfelter's syndrome, hemophilia, colorblindness, and fragile-X syndrome, the leading cause of mental retardation[25]. These are thought to be due in part to unique features of X chromosome biology, among them the demands of dosage compensation and sex determination. Honey bees are "haplo-diploid;" in a sense, each bee chromosome is an X-chromosome, i.e., one copy in the male and two copies in the female. A HBGP will enable comparative analyses to address questions such as: What control regions are important in gene expression, sexual development, and dosage compensation on the X? What role, if any, do orthologs of dosage compensation and DNA repair genes play in a haplo-diploid? No haplo-diploid animal has yet been sequenced.

2. Informing human biology: *Instincts.* The societies of honey bees and other social insects occupy Wilson's[26] second "pinnacle of social evolution," with complexity that rivals our own. Among the provocative similarities are: extensive communication systems (including the only non-primate symbolic language);

socially modulated patterns of behavioral maturation (via inhibitory pheromones); intricate systems of division of labor; highly organized defense and warfare; complex architecture (including the insect equivalent of skyscrapers – 4 m high termite nests in Africa); and expressions of personal sacrifice unheard of in most of the rest of the animal kingdom. Many of these traits are instincts or have strong instinctual components, suggesting that it should be possible to identify genes in humans that are involved in similar traits. In bees these traits are amenable to experimental molecular analysis; the full range of behavioral maturation unfolds in a lifespan of about one month and the natural social environment can be manipulated extensively[27]. In addition, knowledge of bee biogeography and speciation gained from molecular evolution studies by Consortium members Sheppard and Smith and others[28] provide ecological and evolutionary grounding, in a manner analogous to the role of anthropology. Recent studies by Consortium members are starting to implicate genes in complex social instincts[29]. We predict that interest and productivity in this area are poised for an explosive increase, especially with the help of a HBGP. For example, the cover story in the 11 January issue of *Science*[30] reported the discovery of a gene involved in determining whether a colony of fire ants is "governed" by one or more queens. Understanding the regulation of behavior by elucidating how nature/nurture interactions act at the molecular level is a pressing question in human biology; with a sequenced genome, honey bees will deliver some important answers.

Cognition. Bees collect food from flowers, a highly ephemeral food source, and have evolved sophisticated cognitive abilities to maximize foraging success[31]. They are excellent at associative learning, based on the need to associate a color, shape, scent, or location with a food reward. These talents are elegantly captured in laboratory assays pioneered by Consortium member Menzel[32]; Menzel and Consortium member Erber have localized learning phenomena to discrete brain regions and recently extended their analyses to the molecular level[33]. Honey bees also can learn abstract concepts such as "similar" and "dissimilar," and are able to negotiate complex mazes by using visual stimuli as direct or abstract "signposts" or by recognizing path irregularities[34]. Bee behavior also has inspired the construction of robots in DARPA-funded research[35]. Other noteworthy cognitive abilities include the famous bee dance language (which encodes distance and direction information into symbolic movements and sounds) and the ability of guard bees to safeguard their colony's food stores by distinguishing hivemates from unrelated, foreign intruders (also studied in the lab[36]). Bees thus display "vertebrate-like" cognitive abilities, with a brain with only 4X > neurons than *Drosophila*. A HBGP will allow for the identification of "cognition genes" and gene networks and, through comparative genomic analyses involving *Drosophila*, also allow us to explore the relationship between behavioral complexity, neuron number, and gene number. Understanding this relationship is important for understanding human cognition and the treatment of neurodegenerative disease.

Gerontology. Queens and their workers have identical genotypes but queens live two orders of magnitude longer. Moreover, this difference is natural; while it has been relatively easy to select for extended longevity in *Drosophila* and other laboratory models, long-lived strains have not been observed in the wild[37]. Identification of all differentially expressed genes responsible for these striking differences in lifespan, facilitated by a HBGP, undoubtedly has important implications for human longevity and aging; these issues are beginning to be addressed by Consortium members[38]. Preliminary results[39] reveal striking queen-worker differences, including a 10-fold expression difference in a key antioxidant gene, *MnSOD*.

3. Informing the human sequence: *Gene regulation.* The differences between primates and humans are generally thought to be due to differences in gene regulation. Likewise, *Drosophila* and the honey bee probably share the vast majority of genes; therefore their striking differences in neural and behavioral complexity (including sociality) are likely due to differences in gene regulation. Yet despite ca. 300 MY of divergence, flies and bees retain many similarities. With a HBGP, comparative genomics can be used to identify conserved regulatory sequences (and then regulatory networks) that underlie fly/bee similarities and differences[40]. This will involve new techniques that can identify regulatory sequences from the honey bee genome using the *Drosophila* sequence, despite their 300 MY of evolutionary divergence. For example, a new approach[41] focuses on 5-8 bp transcription factor binding sites, which are highly conserved during

evolution[42], and assembly of first-order weight matrices representing the probabilities that each of the four nucleotides will be found at each position in the site. Similarly, enhancer prediction algorithms can be applied to identify novel candidate *cis*-regulatory sequences in the honey bee genome. Candidate bee enhancers, chimeric bee/fly enhancers, and other modified enhancers will then be tested in transgenic constructs introduced into *Drosophila*. Comparing transcriptional regulation in *Drosophila* and honey bee promises to reveal both conserved and diverged functional elements, with insights of broad general interest.

4. Providing a better connection between human and non-human sequences: *Orthologs of human disease genes*. There are known *Drosophila* orthologs for 77% of 929 human disease genes[43]. A HBGP will improve the identification and characterization of additional insect orthologs of human disease genes. Consortium member Robertson, working with about 20,000 EST sequences from a survey of genes expressed in the bee brain from Robinson's lab, discovered that a sizable fraction of bee protein sequences (25% of ca. 3500 assembled EST contigs and singletons) were more similar to human orthologs than to *Drosophila* orthologs[44]. In addition, while most of these predicted proteins do have identifiable *Drosophila* orthologs, ca. 125 do not, indicating that the fly ortholog has been lost during evolution. No orthologs are identified in the *Drosophila* genome sequence or in the large *Drosophila* EST collections, so this is not simply a matter of genes not being sequenced. Extrapolating from these data, we estimate that ca. 500 genes shared by bees and vertebrates are absent from *Drosophila*. Some of these are members of major families e.g., protein kinases, phospholipases, and ubiquitin conjugases. This may appear to be surprising, given that flies and bees diverged about 300MY ago while insects and mammals last shared a common ancestor at least 600MY ago. However, given what is now known about patterns of genome evolution, these results are expected; rates of gene divergence and loss vary in a gene- and lineage-specific fashion. The honey bee genome sequence will thus identify orthologs of human genes that are not conserved in *Drosophila*, and the honey bee is a model system for functional studies of these genes. The *Anopheles* genome will likely also reveal some of the "missing" fly orthologs of human genes, but a substantial number of insect orthologs of human genes will likely be revealed only by genome sequencing of a more distantly related species such as the honey bee.

5. Expanding our understanding of biological processes relevant to human health: *Developmental biology.* Social insects are known for their striking developmental polymorphisms. The best studied case is the honey bee queen/worker polymorphism. Female larvae develop into queens or workers on the basis of larval nutrition and endocrine signaling. Developmental switches render workers almost entirely sterile while queens develop into one the most fecund animals known (they can lay 1 egg/sec for several years). An understanding of the genetic architecture behind this "caste determination" can help answer fundamental questions related to reproduction, nutrition and growth rates, and aging, and provide insight into how genes and the environment interact during development. Gene expression analyses have begun[45]; complete characterization, and studies of relevant *cis* and *trans* regulators require a HBGP. Tractability (e.g., the ability to follow individuals throughout development), combined with genomic knowledge, will establish honey bees as important models in developmental biology.

Neurobiology. Exciting opportunities exist to learn about the molecular basis of neural plasticity, especially in the mushroom bodies (MB), an insect brain region associated with learning and memory. Consortium members Fahrbach and Robinson discovered vertebrate-like structural plasticity in the MB, i.e., changes in volume during behavioral maturation that reflect changes in dendritic arborization[46]. Among the topics relevant to human brain function that are addressed by molecular studies of the bee brain are: the causes and consequences of parallel arrays of neurons found in the MB (as in hippocampus and cerebral cortex); cytoarchitectural plasticity of adult neurons as a function of aging and environmental effects; the function of nuclear hormone receptors in an adult non-reproductive (worker bee) context; and caste determination as a model for organizational effects on the CNS, comparable to sex differences in vertebrates.

Complex systems analysis. One unifying paradigm in biology and medicine is the self-organization of complex systems, with profound implications for understanding diverse phenomena including cardiac regulation, embryonic development, brain function, genome organization, and the behavior of the stock market[47]. Complex systems analysis is in its infancy, but already the bee society has been used as a model because the numerous activities of individuals are integrated into a cohesive and adaptable colony without centralized control. Determining genome structure and function, and the genes and regulatory elements that control the emergence of honey bee self-organization, may facilitate the elucidation of a general theoretical framework to explain how individual interactions give rise to global order.

6. Additional surrogate systems for human experimentation*: Infectious disease.* Honey bees show great promise as models for studies of infectious disease (see A1). Their sociality gives them relevance for human disease transmission not found in other disease models. A colony under stress, due to disease, environmental perturbations, famine or other factors, is much more vulnerable to the effects of other such perturbations[8], analogous to the impact of social and cultural upheaval on human society. Bee responses to pathogens are reminiscent of human's, including a hive-level "fever" to combat fungal pathogens[48] and hygienic behavior that is directed at specific pathologies[49]. Disease resistance in termites (another social insect) increases when individuals are grouped[50], suggesting novel social influences on immunity.

7. Facilitating the ability to do experiments ("direct" genetics, positional mapping). Construction of detailed linkage maps with DNA markers[51] revealed that the honey bee has an unusually high recombination rate, about 52kb per cM (or about 19cM per MB, 20X the human rate), the *highest known* recombination rate of any higher Eukaryote. A high recombination rate was critical for the mapping[52] and isolation[53] of the major bee sex determining gene by Consortium members Beye, Page, and Hunt. Linkage maps and artificial selection (facilitated by artificial insemination of queens[54]) also have been used to identify QTLs for several complex behavioral traits by Consortium members including differences in aggression, food gathering, and learning[55]. Honey bees can be used to map and identify genes for complex traits such as aggression or disease resistance. This can be accomplished with a strategy based on: 1) whole genome screening via SNPs to be generated from a HBGP; 2) the presence of natural genetic variation for many such complex traits; 3) an extremely high recombination rate; 4) a relatively short generation time; and 5) artificial insemination for breeding.

Genomic manipulations are possible; Consortium member Smith has preliminary success making transgenic bees by inseminating a queen with semen mixed with a DNA construct[56], the Menzel lab has manipulated gene expression in a specific region of the brain via antisense[33], and the Smith lab has preliminary results with RNAi[57]. Single-gene mutations affecting eye and body color, wing morphology, and social behavior have been identified[58], and their allelic basis has been confirmed with the help of artificial insemination[59]. Orthologs for the *Drosophila* eye-pigment genes white, claret, and yellow are present in the Bee Brain EST set. Bee mutants have been used to study color vision and social communication[58]. Complete cDNAs and flanking regions for them will help determine their similarity to *Drosophila* orthologs, facilitate identification and preservation of future mutant strains, and enhance their use in state-of-the art functional genomic assays such as those being developed for *Drosophila* and *C. elegans*[60].

Another novel aspect of a HBGP is that it will accomplish important research as an integral component of the sequencing effort. The HBGP will provide "proof-of-concept" for the new Clone Array Pooled Shotgun Sequencing (CAPSS) strategy[61]. It also will identify haplotypes, SNPs and other sequence differences for the Africanized honey bee (and their dramatically heightened aggression). This will involve 7-8X coverage (CAPSS) using the mite-resistant strain followed by 1X coverage (Whole Genome Shotgun) with AHB. This approach, coupled with backcrossing, phenotypic testing, and the honey bee's high rate of recombination, will lead to rapid identification of candidate genes responsible for two complex traits (aggression and parasite resistance) as a primary research mission of the HBGP.

8. Expanding our understanding of evolutionary processes. Altruism is the social glue, the trait that enables a complex society to evolve and function. But altruism has long been an evolutionary enigma, inconsistent with basic Darwinian theory. This is seen in its starkest form in the insect societies: Most members spend their lives helping the queen to reproduce rather than increase their direct fitness by generating their own progeny. Efforts to solve this puzzle have had profound effects. They have led to the development of many of the most widely accepted theories of social evolution in all organisms, including humans, i.e., kin selection and reciprocal altruism, and have spawned "evolutionary psychology," a controversial subdiscipline that assumes that aspects of human sociality are evolved traits, and therefore have biological bases[62]. Molecular analyses of bee social behavior can contribute to our understanding of social evolution. While ants and termites are all highly social, there are bee species that span the range of possible social phenotypes from solitary to primitively social and on up to those with the most advanced societies. In addition, within the Order Hymenoptera (ants, bees, and wasps), it is estimated that sociality evolved independently at least 11 times[63]. A comparative genomic approach, spearheaded by a HBGP, can use these natural experiments to gain insights into the molecular basis of sociality.

B. Strategic issues in acquiring new sequence data

1. Demand for honey bee genome sequence data. The primary intended beneficiary communities for a HBGP include labs that study honey bees, social insects, parasitoid wasps, insects in general, and a much broader community interested in comparative genomics, especially those who study *Drosophila*. The honey bee community numbers about 150 laboratories worldwide. A special feature of this community is that many members are interested in a comprehensive and interdisciplinary understanding of honey bee biology, especially behavior. Consequently a great deal is known of bee biology and natural history, information that promises to add context and meaning to the new molecular information to come from a HBGP (We *know* what pathogens affect the bee and so can conduct focused studies on coevolved relationships to uncover sophisticated mechanisms of immunity; we *know* which members of a bee society are likely to sting, and so can predict the best source of bioactive venom components; we *know* when and why bees use certain types of learning in their life, and so can pick contexts appropriate for experimentation, etc.) It is anticipated that many of these labs will be drawn to molecular analyses by the HBGP. There are ca. 30 laboratories already focused on molecular analyses of honey bees, and members of this group form the core of the Consortium that have prepared this proposal. Their recent publications (see References) in journals such as *Nature*, *PNAS, Genetics, J Neurosci, Genome Res,* and *Genome Biol*, suggest that HBGP results will be used well.

The bee community has met three times to mobilize for a HBGP (#1 organized by Robinson in Bellagio, Italy[64]; #2 organized by Menzel in Berlin; #3 held as part of a USDA Comparative Insect Genomics workshop). At all three meetings a HBGP was recognized as a high priority; at #3 the honey bee emerged as one of the top two insect prospects for genome sequencing[65]. The community that studies social insects and other Hymenoptera, which includes ca. 1500 investigators, also is very enthusiastic about a HBGP (see Appendix). Scientists interested in comparative genomics, especially those that study *Drosophila*, also are interested in having the genome sequence of a distantly related insect (see Appendix). E. Green's success in mammalian comparative genomics[66] provides a model for what can be accomplished with carefully selected insect species. An improved understanding of gene function and gene evolution will inform similar studies in mammals. We predict that the HBGP, coupled with the compelling attributes of bees outlined here, will lead to an influx of young scientists from some of the more "crowded" sectors of biology looking for opportunities in genomics and proteomics (already two Stanford- and Harvard-trained Ph.Ds from the *C. elegans* and yeast communities, respectively, have elected to do postdoctoral research in one Consortium member's laboratory, in part due to the availability of EST and microarray resources).

Plans for genome annotation also indicate the depth and breadth of interest in a HBGP. It will be spearheaded by Consortium member Robertson and will involve a team of bee collaborators and fly experts assembled with the aid of G. Rubin of BDGP and W. Gelbart of FlyBase. All data will be freely available on a web site set up and maintained under the direction of Consortium member Maleszka. It will be linked with

FlyBase and developed with new NIH "generic module" database tools. There will be full and careful coordination with Flybase. Leland Ellis, head of bioinformatics at USDA-ARS, has indicated a willingness to fund bee informatics initiatives with FlyBase for annotation and broad dissemination of the results. Dr. Ellis has offered to organize a workshop to make detailed arrangements for the annotation of the honey bee genome. The genome sequence web site will be part of a comprehensive web site that integrates genomic information with information on bee neurobiology, behavior, ecology, evolutionary biology, and applied aspects of bee biology. This site already has begun to be developed by Consortium member Z. Huang; enhancement (NSF) funding will be sought to incorporate information from the HBGP.

2. Suitability of honey bees for experimentation: Table 1.

Table 1. Summary of features

Haploid chromosome #	16
Genome size	270 Mb (est. 20-40% heterochromatin, based on fly genome)
Genes/ORFs	14,000-16,000 (est. based on *Drosophila*)
Cellular organization	Complex multicellular
Ploidy	Male-haploid, Female-diploid
Generation time (egg->egg)	30 d
Cultivation	Requires space, outside
Genetic resources/tools	
Fraction sequenced	<5%
cDNA/EST resources	28,000 (~10,000 genes) from brain, other specific tissues/life stages
Gene expression	Microarray from bee brain (6,200 genes)
BAC libraries	TAMU-Baylor- 25x (DeJong), Clemson-Purdue - 15x, end-sequences
Transfect/transgenic	Yes
Gene inactivation	RNAi, antisense
Mutants	Many morphological
Germplasm storage	Yes, sperm and embryos
Special Strengths	Agricultural & environmental importance; comprehensive knowledge of bee biology and behavior
	QTL and linkage maps (RAPD and >1000 STS)
	Extreme recombination rate (52 kb/cM) facilitates pos. cloning
	Haploid males aid positional cloning, functional analyses, imprinting studies
	Routine artificial insemination
	Behavioral, aging, and disease model
	Brain atlas (anatomy and gene expression);
	Tissue culture, including neurons
	Population genetics well developed
Weaknesses	Mutant stock maintenance difficult
	No immortal cell culture
	Some inbreeding, no isogenics
	Heterologous gene expression
Databases	Gene-expression data, physical maps, EST resource, EM resource
Web Resources	U. Illinois, http://titan.biotec.uiuc.edu/bee/honeybee_project.htm
	TIGR Gene Index, http://www.tigr.org/tdb/amgi/
	USDA Beenome Site, http://www.barc.usda.gov/psi/brl/beenome.html
	General Bee Biology Site, http://www.cyberbee.net/
	http://www.neurobiologie.fu-berlin.de/honeybeeALatlas/
Number of Labs	Honey bee: ca. 150 (genetics, physiology, pathology, behavior; 30 genetics/molecular biology); ca. 350 social insect/Hymenoptera labs
Number of investigators	1500 (members of international social insect society, IUSSI; 300)

3. Rationale for complete sequence. A full genome sequence is required to use honey bees to address a broad range of human health-related problems: efficient genomic prospecting for novel antibiotics, therapeutic venom components and genes involved in innate immunity; mapping and identifying genes for complex traits including those with relevance to social behavior, mental health, gerontology, development, and nutrition; insights into X chromosome diseases; identification of olfactory receptors (and other important and rapidly evolving proteins) for basic and perhaps security-related purposes; understanding the relationship between behavioral complexity, neuron number, and total gene number (including regulatory genes typically expressed at low levels and difficult to obtain via other approaches); identification of regulatory elements to study social regulation of gene expression (e.g., are regulatory elements for genes affecting longevity and complex behavior more complex in bees vs. flies?); and enhanced identification of orthologs of human disease genes (see Part A).

4. Costs and readiness. Weaver's contacts with R. Gibbs, Director, Baylor College of Medicine Human Genome Sequencing Center (BHGSC), led to the following: 1) BHGSC expressing strong interest in pursuing a HBGP (see Appendix); 2) identification of a suitable strain for sequencing; 3) preparation of high-molecular weight DNA for a BAC library; 4) a 20X BAC library (Texas A&M/Industry-funded) from P. de Jong (BAC PAC Resources & Children's Hospital, Oakland Research Institute), made from *one haploid* (male) bee to facilitate genomic sequence assembly; 5) 1800 sequence reads; 6) budgetary estimates for a HBGP; and 7) a White Paper planning meeting hosted by BHGSC (20-21 December, 2001, organized by Weaver).

BHGSC has performed preliminary sequencing in anticipation of a HBGP. A short insert plasmid library (1-3 Kb, randomly sheared) was prepared from a single (haploid) drone from the mite-resistant bee strain that will be used in the HBGP. Inserts were cloned into pUC 18 using a double adaptor method, and transformed into XL10 gold cells before being picked, grown, and sequenced from both ends (other libraries will be made to more precise specifications for the HBGP; e.g. with prior removal of mitochondrial DNA and with tighter control over insert length). A total of 1800 sequence reads were made. Results indicate ca. 40% GC content, which will not pose problems for more extensive cloning, sequencing, and assembly of the complete euchromatic genome. Although this is a very small sample, very few repeat structures were encountered, there were no matches to retrotransposons, and only four matches to short DNA transposons of the mariner family that have previously been described from honey bees[67]. These results suggest that honey bees are relatively depauperate in transposons (as speculated 10 years ago[68]), and that most of the non-coding sequences might be unique, facilitating final assembly of the complete genome. As expected, BLASTX analyses revealed numerous significant matches to genes from *Drosophila*. The "hit rate" (ca. 10%) is about the same as for *Anopheles* (BAC end) matches to *Drosophila*. We thus expect a similar bee/mosquito gene density, which is consistent with their similar genome size estimates. More detailed analyses with the bee brain (20,000) EST set[44] reveal that of 3449 assembled sequences with an ORF $\geq$ 450 bp, 2616 (76%) had matches in the nr database. Additional DNA from haploid drones is readily available, which will facilitate assembly at only modest levels of coverage. SNPs will be obtained by obtaining sequence information from a 1X sequencing of DNA from Africanized honey bees. Preliminary analyses by BHGSC indicate no technical impediments to rapid and efficient completion of a HBGP.

The HBGP strategy developed by BHGSC will aim for a 7-8X coverage of the 270MB genome, but not full finishing. This will employ a mixed approach of low coverage sequencing of BAC clones (1-2x coverage BAC skims) with the rest of the sequence coming from whole genome shotgun reads. Assuming inserts of 150kb (from de Jong's libraries), BAC end sequences (BES) will be generated from 7200 clones, corresponding to 4X clone coverage. These clones also will be sequenced using a pooled clone array (CAPSS) approach to minimize the number of BAC DNA preparations and shotgun libraries needed. In the extreme form of this method, the BACs will be arranged in an 85x85 array and the 170 rows and columns will be pooled separately. DNA preparations and shotgun libraries will be prepared from the pools and sequenced to an average coverage of 1-2x per BAC. The sequences from each row and column will be mixed in each pairwise combination. The mixtures of reads will be co-assembled and contigs with both row

and column reads will be judged to come from the BAC at the intersection of the row and column. Thus the reads will be assigned to each BAC. This method reduces the number of DNA preparations and shotgun libraries from 7200 to 170. A more conservative approach, that may be more manageable technically, will use 18 arrays of 20x20 clones, requiring 720 DNA preparations and libraries, which is still a 10-fold reduction. Experiments already underway will determine the best-sized array.

Whole genome shotgun reads, to about 4X genome coverage, will come from 3kb, 10kb, and 40kb libraries. All sequencing will be in plasmids and end pairing maintained at high (>90%) fidelity. The whole genome shotgun reads are binned into appropriate BACs, using ATLAS genome assembly software at BHGSC. The enriched BACs are assembled by deriving their overlaps based on sequence comparison of the enriched contigs and placement of the BES to more accurately define the overlaps. The contigs at the end of the insert are anchored by identifying read pairs with one read in the insert and one in the vector. This strategy is being used successfully to pick clones for the rat genome project. This will give an initial assembly of the genome, at around 8X sequence coverage. Some gaps are expected in the map; they will be filled with BACs identified by overgo hybridization to the remainder of the BAC library. Hybridization probes are derived from the BAC sequences, which will be sequenced to corresponding coverage. In general, all methodology proposed for the HBGP is proven and is now being used in the rat genome project.

In addition to the construction of the draft DNA sequence, other activities are of high interest. The first is obtaining more information about the Africanized Honey Bee. A 1x shotgun sequencing of the AHB genome will be performed and these reads will be used to identify differences with the assembled sequence. Finally, finishing regions of biological interest and full-length cDNAs will also be considered.

Sequencing an entire 270MB genome requires ca. 5 million successful reads, which would cost $10M. But subtracting an assumed 30% heterochromatin, the cost would be ~$7M. The cost of the project will be $5-10M with our best estimate currently at $7M. This could be performed at BHGSC in 4 months.

5. Other (partial) sources of funding. Funding for the BAC library to be used in the HBGP was obtained from Texas A&M University, the Texas Beekeepers Association and private industry donations. Funding for bioinformatics has been promised by USDA-ARS, and will involve leverage of Genome Browser (J. Kent, UC-Santa Cruz) and EBI's Ensembl (supported by the Wellcome Trust). BHGSC currently is using both of these tools. An important component of the bioinformatics strategy is to transition the genome sequence from the sequencing center, and public databases, into the community. There is a significant opportunity to leverage existing tools, FlyBase and GadFly, together with reusable "generic modules" that can be readily ported and used for a HBGP. This was discussed at the recent USDA-sponsored international Comparative Insect Genomics Workshop, along with follow-up discussions between Dr. Ellis and Michael Ashburner and Suzi Lewis of FlyBase. USDA/ARS is especially interested in working with the HBGP and other funding agencies to help support the further development of such resources.

Letters of Support: Roster (see Appendix 1)

The following individuals have written in support of the HBGP, as prominent scientists with general perspective, representatives of specific communities, or both. Permission was granted from Dr. Jane Peterson to include letters of support as an appendix.

G. Weinstock, Co-Director, BHGSC --directed preliminary data collection; American Beekeeping Federation; T. Cech, President, HHMI; C. Fraser, President, TIGR; R. Hamilton, Director, Johns Hopkins Laboratory of Dermatology, Allergy and Clinical Immunology; L. Hood, President, Institute for Systems Biology; R. Menzel (Berlin), leading honey bee neuroscientist and Consortium member; K. Ross, leading ant geneticist; G. Rubin, Director, BDGP; B. Thorne, leading termite biologist; J. Tower, (fly) gerontologist, interested in transgenic bees with HBGP as resource; R. Van Driesche, leading biological control expert; D. Weaver, industry liaison and Consortium member; E.O. Wilson, leading social insect biologist; endorses honey bee as first social insect for genome sequencing. (Additional letters available upon request.)

References

[1]Otvos L (2000) Antibacterial peptides isolated from insects. J Pept Sci 6: 497-511; Sajjan US, Tran LT, Sole N, Rovaldi C (2001) P-113D, an antimicrobial peptide active against *Pseudomonas aeruginosa* retains activity in the presence of sputum from cystic fibrosis patients. Antimicrob Agents Chemother 45: 3437-44

[2]Schmid-Hempel (1998) *Parasites in Social Insects.* Princeton Univ Press, Princeton, NJ; Moret Y, Schmid-Hempel P (2001) Immune defence in bumble-bee offspring. Nature 414: 506

[3]Cooke GS, Hill A (2001) Genetics of susceptibility to human infectious disease. Nature Rev Gen 2: 967-977

[4]Andersen P (2001) TB vaccines: progress and problems. Tr Immunol 22: 160-8

[5]Fellermann K, Stange EF (2001) Defensins-innate immunity at the epithelial frontier. Eur J Gastroenterol Hepatol 13: 771-6; Mushegian A, Medzhitov R (2001) Evolutionary perspective on innate immune recognition. J Cell Biol 155:705-10; Fortini ME, Skupski MP, Boguski MS, Hariharan IK. (2000) A survey of human disease gene counterparts in the *Drosophila* genome. J Cell Biol. 24: 150-160; De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. (2001) Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. Proc Natl Acad Sci 98: 12590-12595

[6]Michelle T, Reichhart J-M, Hoffmann JA, Royet J (2001) Drosophila Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature 414: 756-759; Williams MJ (2001) Regulation of antibacterial and antifungal innate immunity in fruitflies and humans. Adv Immunol 79: 225-59

[7]Irving P, Troxler L, Heuer TS, Belvin M, Kopczynski C, Reichhart JM, Hoffmann JA, Hetru C (2001) A genome-wide analysis of immune responses in *Drosophila*. Proc Natl Acad Sci 98: 15119-24

[8]Morse RA, Flottum K, eds (1997) *Honey Bee Pests, Predators, and Diseases*. 3rd Ed. AI Root, Medina, OH

[9]Harbo JR, Harris JW (2001) Resistance to Varroa destructor (Mesostigmata: Varroidae) when mite-resistant queen honey bees (Hymenoptera: Apidae) were free-mated with unselected drones. J Econ Entomol 94: 1319-23

[10]Chung I et al (2002) Mononuclear phagocyte biophysiology influences brain transendothelial and tissue migration: implication for HIV-1 associated dementia. J Neuroimmunol 122: 40-54; Ogris M. et al (2001) Melittin enables efficient vesicular escape and enhanced nuclear access of nonviral gene delivery vectors. J Biol Chem 276: 47550-47555; Roger T et al. (2001) MIF regulates innate immune response through modulation of Toll-like receptor 4. Nature 414: 920-924; Wachinger M et al. (1998) Antimicrobial peptides melittin and cecropin inhibit replication of human immunodeficiency virus 1 by suppressing viral gene expression. J Gen Virol 79: 731-740; Finlayson K et al (2001) Characterization of [125I]-apamin binding sites in rat brain membranes and HEK293 cells transfected with SK channel subtypes. Neuropharmacol 41: 341-350

[11]Konno K (2001) Anoplin, a novel antimicrobial peptide from the venom of the solitary wasp *Anoplius samariensis*. Biochem Biophys Acta – Protein Struct & Molec Enzymol 1550: 70-80

[12]Ono SJ (2000) Molecular genetics of allergic diseases. Annu Rev Immunol 18: 347-366

[13]Golden DB, Marsh DG, Freidhoff LR, Kwiterovich KA, Addison B, Kagey-Sobotka A, Lichtenstein LM (1998) Natural history of Hymenoptera venom sensitivity in adults. J Allergy Clin Immunol 1 02: 702-3; Kolecki P (1999) Delayed toxic reaction following massive bee envenomation. Ann Emerg Med 34: 411-412

[14]Kettner A et al (1999) IgE and T-cell responses to high-molecular weight allergens from bee venom. Clin Exp Aller 29: 394-401

[15]Bulag G et al. (2001) Delta-conotoxin structure/function through a cladistic analysis. Biochem 40: 13201-13208

[16]Schumacher MJ et al (1995) Significance of Africanized bees for public health. Arch Int Med 155: 2038-2043

[17]Vetter RS et al (1999) Mass envenomation by honey bees and wasps. West J Med 170: 223-227

[18]Morse RA, Calderone NW (2000) The value of honey bee pollination in the United States. Bee Culture 128: 1-15

[19]Hackett K, USDA-ARS, pers comm

[20]Insel, TR, Young, LJ (2001)The neurobiology of attachment. Nat Rev Neurosci 2: 129-36

[21]Seeley TD (1995) *The wisdom of the hive: The social physiology of honey bee colonies.* Harvard Univ Press; Robinson GE (1998) From society to gene with the honey bee. Am Sci 86: 456-462

[22]Bromenshenk JJ, Carlson SR, Simpson JC, Thomas JM (1985) Pollution monitoring of Puget Sound, Washington USA with honey bees. Science 227: 632-634; Lighthart B, Prier K, Loper GM, Bromenshenk J (2000) Bees scavenge airborne bacteria. Microb Ecol 39:314-321

[23]Bromenshenk J, pers comm

[24]Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR (1999) A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila. Neuron 22: 327-38

[25]Amos W and Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. Phil Trans R Soc Lond B 353:177-86

[26]Wilson EO (1975) *Sociobiology: A new synthesis.* Belknap Press of Harvard Univ Press

[27]Robinson, G.E., Fahrbach, S.E. and M.L. Winston. (1997) Insect societies and the molecular biology of social behavior. *BioEssays* 19: 1099-1108

[28]Hall HG, Smith DR (1991) Distinguishing African and European honey bee matrilines using amplified mitochondrial DNA. Proc Natl Acad Sci USA 88: 4548-4552; Sheppard WS, Rinderer, TE, Mazzoli, J, Stelzer, JA, Shimanuki H (1991) Gene flow between African- and European-derived honey bee populations in Argentina. Nature 349: 782-784

[29]Page RE Jr, Fondrk MK, Hunt GJ, Guzman-Novoa E, Humphries MA, Nguyen K, Greene AS (2000) Genetic dissection of honeybee (*Apis mellifera* L.) foraging behavior. J Hered 91: 474-9; Hunt GJ,

Guzman NE, Fondrk MK, Page RE (1998) Quantitative trait loci for honey bee stinging behavior and body size. Genetics 148:1203-1213; Toma DP, Bloch G, Moore D, Robinson GE. (2000) Changes in *period* mRNA levels in the brain and division of labor in honey bee colonies. Proc Natl Acad Sci :6914-9; Kucharski R, Maleszka R (2002) Evaluation of differential gene expression during behavioral development in the honeybee using microarrays and northern blots. Genome Biology 3: 7.1-7.9

[30]Krieger MJ, Ross KG (2002) Identification of a major gene regulating complex social behavior. Science 295: 328-32

[31]von Frisch K (1967) Dance language and orientation of the honey bee. Harvard Univ Press; Seeley 1995, op cit; Capaldi EA, Smith AD, Osborne JL, Fahrbach SE, Farris SM, Reynolds DR, Edwards AS, Martin A, Robinson GE, Poppy GM Riley JR (2000) Harmonic radar reveals ontogeny of orientation flight in the honeybee. Nature 403: 537-540

[32]Menzel R, Mueller U (1996) Learning and memory in honeybees: from behavior to neural substrates. Ann Rev Neurosci 19:379-404

[33]Fiala A, Mueller U, Menzel R. (1999) Reversible downregulation of protein kinase A during olfactory learning using antisense technique impairs long-term memory formation in the honeybee, *Apis mellifera*. J Neurosci 19: 10125-34

[34]Srinivasan MV, Zhang SW,. and Zhu H (1998): Honeybees link sights to smells. Nature (Lond.) 396, 637-638; Zhang SW, Mizutani A, Srinivasan MV (2000) Maze navigation by honeybees: learning path regularity. Learning and Memory 7, 363-374; Giurfa M, Zhang SW, Jenett A, Menzel R, and Srinivasan MV (2001): The concepts of 'sameness' and 'difference' in an insect. Nature 410: 930-933

[35]Srinivasan MV, Zhang S, Chahl JS (2001) Landing strategies in honeybees, and possible applications to autonomous airborne vehicles. Biol Bull 200: 216-21

[36]Breed MD, Rogers KB (1991) The behavioral genetics of colony defense in honeybees: genetic variability for guarding behavior. Behav Genet 21:295-303

[37]Vettraino J, Buck S, Arking R (2001) Direct selection for paraquat resistance in *Drosophila* results in a different extended longevity phenotype. J Gerontol A 56: B415-25

[38]Evans JD, Wheeler DE (1999) Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. Proc Nat Acad Sci 96: 5575-5580; Evans JD, Wheeler DE (2000) Expression profiles during the honey bee caste program. Genome Biology 2: 1-6; Page RE, Peng CY (2001) Aging and development in social insects with emphasis on the honey bee, *Apis mellifera* L. Exp Gerontol 36: 695-711

[39]Corona M, Robinson GE unpubl

[40]Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288: 136-140

[41]Berman BB, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levin M, Rubin GM, Eisen. Exploiting transcription binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. PNAS. In press

[42]Schmid KJ, Tautz D (1999) A comparison of homologous developmental genes from Drosophila andTribolium reveals major differences in length and trinucleotide repeat content. J Mol Evol 49: 558-66

[43]Reiter LT, Potocki L, Chien S, Gribskov M, Bier E (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. Genome Res 11: 1114-25; http://homophila.sdsc.edu

[44]Whitfield, CW Band M, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Res. In press; http://titan.biotec.uiuc.edu/bee/honeybee_project.htm

[45]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: 929-938Evans JD, Wheeler DE (1999) Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. Proc Nat Acad Sci 96: 5575-5580; Evans JD, Wheeler DE (2000) Expression profiles during the honey bee caste program. Genome Biology 2: 1-6; Hepperle C, Hartfelder K (2001) Differentially expressed regulatory genes in honey bee caste development. Naturwiss 88: 113-116

[46]Withers GS, Fahrbach SE, Robinson GE (1993) Selective neuroanatomical plasticity and division of labour in the honeybee. Nature 364: 238-240; Farris SM, Robinson GE,. Fahrbach SE (2001) Experience- and age-related outgrowth of intrinsic neurons in the mushroom bodies of the adult worker honey bee. J Neurosci 21: 6395-6404

[47]Camazine S, et al (2001) *Self-organization in biological systems*. Princeton Univ Press

[48]Starks PT, Blackie CA, Seeley TD (2000) Fever in honeybee colonies. Naturwiss 87: 229-31

[49]Arathi HS, Burns I, Spivak M (2000) Ethology of hygienic behaviour in the honey bee. *Apis mellifera* L. (Hymenoptera: Apidae): Behavioural repertoire of hygienic bees. Ethol 106: 365-379

[50]Traniello J, pers comm

[51]Hunt GJ, Page RE (1995) Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. Genetics 139: 371-82

[52]Hasselman M, Fondrk MK, Page RE, Beye M (2001) Fine scale mapping in the sex locus region of the honey bee (*Apis mellifera*). Insect Molec Biol 10: 605-608

[53]Unpubl

[54]Laidlaw HH (1944) Artificial insemination of the queen bee (*Apis mellifera* L.): morphological basis and results. J Morph 74: 429-465

[55]Hunt GJ, Guzman-Novoa E, Fondrk MK, Page RE (1998) Quantitative trait loci for honey bee stinging behavior and body size. Genetics 148: 1203-13; Hunt GJ, Page RE Jr, Fondrk MK, Dullum CJ (1995) Major quantitative trait loci affecting honey bee foraging behavior. Genetics 14: 1537-45; Chandra SB, Hunt GJ, Cobey S, Smith BH (2001) Quantitative trait loci associated with reversal learning and latent inhibition in honeybees (*Apis mellifera*). Behav Genet 31: 275-85

[56]Robinson KO, Ferguson HJ, Cobey S, Vaessin H, Smith BH (2000) Sperm-mediated transformation of the honey bee, *Apis mellifera*. Insect Mol Biol 9: 625-34

[57]Unpubl

[58]Laidlaw HH, Green MM, Kerr WE (1953) Genetics of several eye color mutants in the honey bee, *Apis mellifera* L. J Hered 44: 246-250; Gribakin FG (1988) Photoreceptor optics of the honeybee and its eye color mutants: The effect of screening pigments on the long-wave subsystem of color vision. J Comp Physiol A 164:123-140; Kirchner WH, Sommer K (1992) The dance language of the honeybee mutant diminutive wings. Behav Ecol Sociobiol 30:181-184; Montague CE, Oldroyd BP (1998) The evolution of worker sterility in honey bees: Investigation into a behavioral mutant causing failure of worker policing. Evol 52: 1408-1415

[59]Laidlaw HH, Tucker KW (1965) Compound inseminations to abbreviate tests for allelism in honey bee queens. J Hered 56: 127-130

[60]Kalidas S, Smith DP (2002) Novel genomic cDNA hybrids produce effective RNA interference in adult Drosophila. Neuron 33: 177-184

[61]Cai WW, Chen R, Gibbs RA, Bradley A (2001) A clone-array pooled shotgun strategy for sequencing large genomes. Genome Res 1: 1619-23

[62]Wilson EO (1998) *Consilience*. Knopf

[63]Wilson EO (1971) *The insect societies.* Belknap Press of Harvard Univ Press

[64]Maleszka R (2000) Molecules to behaviour in the honeybee – the emergence of comparative neurogenomics. 23: 513-514

[65]Pennisi E (2001) Science 294: 1261-1262

[66]Green ED, Chakravarti A (2001) The human genome sequence expedition: views from the "base camp". Genome Res 11: 645-51

[67]Ebert PR, Hileman JP, Nguyen HT (1995) Primary sequence, copy number, and distribution of mariner transposons in the honey bee. Insect Mol Biol 4: 69-78; Robertson HM (1993) The mariner element is widespread in insects. Nature 362: 241-245

[68]Bigot Y, Lutcher F, Hamelin MH, Periquet G (1992) The 28S ribosomal RNA-encoding gene of Hymenoptera: inserted sequences in the retrotransposon-rich regions. Gene 121: 347-352

[c]Institutional Affiliations of White Paper Authors and Acknowledgements: G.E. Robinson, Dept. Entomology and Neuroscience Program, Univ. Illinois at Urbana-Champaign; K. Aronstein, USDA Bee Lab., Weslaco, TX; J.E. Evans, USDA Bee Lab., Beltsville, MD; S.E. Fahrbach, Dept. Entomology and Neuroscience Program, Univ. Illinois at Urbana-Champaign; J.S. Johnston, Dept. Entomology, Texas A&M Univ.; R. Maleszka, Dept. Biology, Australian Natl. Univ.; R.E. Page, Dept. Entomology, Univ. California Davis; H.M. Robertson, Dept. Entomology and Neuroscience Program, Univ. Illinois at Urbana-Champaign; D.B. Weaver, B Weaver Apiaries, Inc., Navasota, TX. Thanks to J. Belmont, A. Collins, R. Gibbs, H. Lewin, R. Hoskins, S. Richards, and G. Weinstock for advice and information.