Cornell University
College of Agriculture and Life Sciences

Physiology and Toxicology
Department of Entomology
6134 Comstock Hall
Ithaca, NY 14853-2601
Fax: 607 255-0939
Phone: 607 255-7340
jgs5@cornell.edu

January 6, 2006

Coordinating Committee
National Human Genome Research Institute
National Institutes of Health

Dear Coordinating Committee Members:

As you may recall, the house fly (*Musca domestica*) white paper was submitted to NHGRI twice before. The last version was enthusiastically supported by the working group, but did not receive approval from the Coordinating Committee. I discussed this matter with Dr. Felsenfeld who suggested that we revise the proposal where we felt it was necessary, but that we focus our efforts on a letter to you detailing the strong case that we have established for sequencing of the house fly genome. These points are briefly detailed below with greater information being available in the white paper itself.

Why the house fly genome should be sequenced:

1. The first opportunity to study the interactions between a pest insect and its parasitoid at the whole genome level. Sequencing of the *Nasonia vitripennis* genome is currently underway. *Nasonia* is a parasitoid of the house fly (sold commercially for fly control), thus having the genome of both the parasitoid (*Nasonia*) and the host (*M. domestica*) will allow the first opportunity to study the interactions between a pest insect and its parasitoid at the whole genome level. How do *Nasonia* eggs evade the immune response of the house fly? What is the response of the house fly to *Nasonia* venoms? What immunological factors permit some flies, but not others to destroy the developing *Nasonia* embryo? Sequencing of the house fly genome is strongly supported by the *Nasonia* community (see letter from J. H. Werren).

2. Fundamental discoveries in numerous areas, such as insect toxicology and insecticide resistance, have relied heavily on studies of the house fly. In these studies house flies have been preferred over other organisms, such as *Drosophila*, for numerous reasons such as the abundance of insecticide resistant strains. Having access to the whole genome of the house fly would facilitate the rapid advancement of our understanding in these research areas (see letters from J. Casida, I Denholm, M. Williamson F. Collins, Y. Ozoe, T. Sparks and Tomita et al.). House flies have also led the way in fundamental discoveries of the endocrine regulation and biochemistry of cuticular hydrocarbon biosynthesis. Identification of the genes involved in these processes would greatly facilitate studies on how genes are regulated during pheromone biosynthesis and/or hydrocarbon production (see letter from G. J. Blomquist and C. Tittiger).

3. Sex determination in house flies is different from *Drosophila*, but is similar to many other insect species. In house flies, sex is determined by the presence or absence of M, which is normally found on the Y chromosome. In some populations of house fly M is found on one of the autosomes, and there appears to be a North/South cline in the frequency of Y^M vs. A^M males. In some populations males become homozygous (A^M/A^M). In these populations there is a mutation in an unknown gene (F) on autosome IV. Individuals with the mutant F are females, independent of the number of copies of M. Thus, in some populations males have become the homogametic sex and females are the heterogametic sex. Sequencing of the house fly genome will provide the tools necessary for the study of sex determination and understanding the evolution of A^M and F (see letters from M. Ashburner, D.Bopp and M. Heddiger, I. Denholm, M. Williamson and Tomita et. al.).

4. Sequencing of the house fly genome will lead to better control of this pest. Resistance to two new insecticides (spinosad and indoxacarb) has recently been documented in the house fly. To understand the

resistance it is critical to identify the gene responsible. Using *Drosophila/Musca* homology maps, combined with PCR amplification of the orthologous genes, failed to identify the genes involved in resistance to these two new and economically important insecticides. The whole genome sequence will allow us to rapidly identify candidate genes and determine their role in resistance. With allele specific probes we will be able to detect the frequency of resistant individuals (resistance is recessive, so we can only detect heterozygotes with an allele specific assay) in order to design strategies to delay the evolution of resistance. In addition, the genome will allow for the identification of potential target sites for the development of new insecticides (see letters from J. Casida and T.C. Sparks).

5. The house fly lives in a virtual sea of animal pathogens, yet we know very little about how the house fly is able to evade infection (many pathogens are eaten by the house fly, so lack of infection is not due to lack of internalization). Genome wide analysis of the immune response of house fly will present a unique opportunity to study the immune system of this important pest. Sequencing of the house fly genome is strongly supported by scientists interested in insect immunity (see letters for F. Kafatos, L. Zurek and B. Lazzaro).

6. House flies are phylogenetically well placed, such that sequencing of the house fly genome will aid in the analysis and annotation of mosquito and *Drosophila* genomes (see letters from B. Wiegmann and P. Simpson). Sequencing of the house fly genome is strongly supported by the *Drosophila* and mosquito communities (see letters from the *Drosophila* Board, M. Ashburner, M. Eisen, F. Kafatos, F. Collins and P. Simpson).

7. The house fly community is anxious, able and eager to have the house fly genome completed. In addition to numerous well characterized and highly inbred strains, there are large preserved collections, as well as numerous genomic, cDNA and EST libraries available. House flies are easy to rear and there are abundant genetic markers available. RNAi has been successfully used to knock out gene expression and mobile elements are available for transformation experiments.

8. Sequencing of the house fly genome is tractable. House flies are easily reared, inbred strains are available and sequencing from autosomes I and V has not revealed any highly repetitive sequences. The size of the house fly genome is only slightly larger than *D. melanogaster*.

9. The house fly is an important vector of human and animal diseases. While there are numerous scientific reasons for sequencing the house fly genome it is worth remembering that advances in our understanding of house fly biology have the potential to translate into novel methods for control of house flies, and thus improve human and animal health (see letters from M. Ashburner, A. Broce, S. S. Caglar, A. Giangaspero, D. Otranto, B. A. Mullens, and Tomita et al.).

With these points in mind I request (on behalf of Drs. N. Liu, M. Kristensen, and the house fly, mosquito, *Drosophila*, and *Nasonia* communities) that you approve sequencing of the house fly genome. Thank you for your time and consideration.

Sincerely,

Jeffrey G. Scott
Daljit S. and Elaine Sarkaria Professor of Insect Physiology and Toxicology

Rationale for sequencing the genome of the house fly, *Musca domestica*

Jeffrey G. Scott[1], Nannan Liu[2] and Michael Kristensen[3]

[1]Department of Entomology, Comstock Hall, Cornell University, Ithaca, NY 14853 USA, jgs5@cornell.edu, 607-255-7340 (phone), 607-255-0939 (fax), [2]Department of Entomology and Plant Pathology, Auburn University, AL 36849 USA, [3]Danish Pest Infestation Laboratory, Skovbrynet 14, DK-2800, Lyngby, Denmark

Overview: House flies are carriers of dozens of devastating diseases that have severe consequences for human and animal health. Despite the fact that it is a passive vector, a key bottleneck to progress in controlling the devastating human diseases transmitted by house flies is lack of knowledge of the basic molecular biology of this species. Sequencing of the house fly genome will provide important inroads to the discovery of novel target sites for house fly control, understanding of the immune response in this dung-living fly, rapid elucidation of insecticide resistance genes and understanding of numerous aspects of the basic biology of this insect pest. The ability of the house fly to prosper in a remarkably septic environment motivates analysis of its innate immune system. Its polymorphic sex determination system, with male-determining factors on either the autosomes or the Y chromosome, is ripe for a genomic analysis. Sequencing of the house fly genome would allow the first opportunity to study the interactions between a pest insect and its parasitoid (*Nasonia vitripennis*) at the whole genome level. In addition, the house fly is well placed phylogenetically to leverage analysis of the multiple Dipteran genomes sequenced (including several mosquito and *Drosophila* species), and would serve as an important bridge from *Drosophila* to expedite analysis and annotation of the *Anopheles gambiae* genome. The community of researchers investigating *Musca domestica* are well prepared and highly motivated to apply genomic analyses to their widely varied research programs (see appendix). Additionally, experienced *Drosophila* and mosquito communities will benefit from, and are eager to obtain, the house fly genome because it is phylogenetically well spaced between *Drosophila* and mosquitoes (see appendix).

Response to reviews of the January 2005 *Musca* white paper

We appreciate the thought and insight that went into the consideration of the review of our first effort to persuade the Comparative Genomics and Evolution committee that the genome of *Musca domestica* is a worthwhile target. We are very pleased that the Working Group was reasonably enthusiastic about the proposal and recommended it to the Coordinating Committee. It seemed there was agreement that the species is an important disease vector, that the genomic information could be important for controlling disease transmission, and that *Musca* is well placed for comparative genomic analysis with both *Drosophila* and *Anopheles*. Based on an e-mail and phone conversation with Dr. Adam Felsenfeld there were three concerns that lowered the Coordinating Committee's enthusiasm for *Musca.* The first is that *Musca* is primarily a passive vector, and the feeling was that genomic information of active vectors, specifically transmitting co-adapted pathogens, would be more useful for control. While it is tempting to believe that an understanding of the specific co-adaptation between vector and pathogen will be the key to

successful control of insect vectors, in most cases the best solution is to lower the vector population density to the point that the disease is below epidemic threshold. All locally successful attempts to reduce malarial burden, for example, have come from reducing *Anopheles* density and by reducing human contact through bed nets and repellents. While the research community is hopeful that *Anopheles* can be made refractory to malarial transmission, simple population control is far more likely to have a greater impact on malarial loads for the next decade. Similarly, the potential for *Musca* to vector devastating diseases increases with fly densities, which can become seemingly apocalyptic. Genomic information will help to identify targeted insecticides, repellents, pheromones, etc. to control this devastating disease carrier much more effectively than any current measures.

The second concern was the genome size, previously estimated at 900 Mbp. The first and only estimate of the house fly genome size was done by quantitative ultraviolet microscopy by Bier and Müller (1969). While this method has the potential for reasonably accurate results, no details of replication were given to assess the reliability of their particular estimate. In an effort to get a better estimate of the *Musca* genome size, we used the quantitative Real-Time PCR of Wilhelm et al. (2003). Doing the experiments in six replicates we get a revised estimate of genome size for *Musca domestica* of 308-312 Mbp, or only 1.7 times that of *Drosophila melanogaster*. Further details on our revised genome size estimates appear in Section B.

The final concern was that we had not secured an official ruling from the *Drosophila* Board stating the enthusiasm of the *Drosophila* community for this project. Instead, we just had letters from a few key individuals in the fly community (e.g. Michael Ashburner). We have shared our goals with the *Drosophila* Board, but the timing was insufficient to get an official vote of confidence from the whole board. Informally, Mark Krasnow, President of the *Drosophila* Board, did indicate that the *Musca* project is totally in line with the white paper being presented to NIH Council for support for research infrastructure for the fly community. In their white paper, they cite as a critical goal the acquisition of outgroup genome information to further improve annotation of the *Drosophila* genome.

Additional letters of support from Frank Collins (Notre Dame), Fotis Kafatos (Imperial College), Pat Simpson (Cambridge), Bradley Mullens (UC Riverside), Thomas Sparks (Dow AgroSciences) and Ludek Zurek (Kansas State) are now included.

A. Specific biological rationale for *Musca* genomic sequence data.

Improving human health, informing human biology and providing new model experimental systems. House flies (*Musca domestica)* are cosmopolitan, ubiquitous, and are the vectors of more than 100 human and animal intestinal diseases (Greenberg, 1965; Keiding, 1986; Scott and Lettig, 1962), including bacterial infections such as salmonellosis, anthrax ophthalmia, shigellosis, typhoid fever, tuberculosis, cholera and infantile diarrhea; protozoan infections such as amebic dysentery; helminthic infections such as pinworms, roundworms, hookworms and tapeworms; as well as viral and rickettsial infections. Recently house flies were shown to spread a deadly strain of *Escherichia coli* in Japan (Sasaki et al., 2000). Flies also transmit eye diseases such as trachoma and epidemic conjunctivitis, and infect wounds or skin with diseases such as cutaneous diphtheria, mycoses, yaws and leprosy (Keiding, 1986). Fly-transmitted trachoma alone causes 6 million cases of childhood blindness each year (Organization, 2004). Considering that house flies are highly mobile, come into contact with

excreta, carcasses, garbage and other septic matter, and that they are intimately associated with humans, our food and utensils, it is not surprising that they are involved in transmission of so many serious and widespread diseases (Keiding, 1986; Scott and Lettig, 1962). Most recently house flies have been shown to vector life threatening antibiotic resistant bacteria (L. Zudek, personal communication), which are an ever increasing problem in hospitals and other health care facilities (Graczyk et al., 2001; Maisnier-Patin and Andersson, 2004; Sundin, 1996).

House flies are always found in association with humans and human activities. In fact, house flies and humans have evolved together, with house flies following the spread of *Homo sapiens* across the planet (Mündi, 1994). House flies are also one of the most serious pests at dairy, horse, hog, sheep and poultry facilities worldwide. Exposure to debilitating disease-causing agents, public health and nuisance concerns, lowered levels of milk and egg production, reduced feed conversion, all result from house fly activity. Economic losses and the cost associated with fly suppression are difficult to quantify, but costs of pesticides for fly control at poultry facilities alone are estimated at over $200 million annually in the USA (Geden et al., 1994).

The house fly thrives in a virtual sea of animal pathogens. Sequencing of the house fly genome will shed light on the immune defense systems of this important species, and provide valuable information about how it is able to flourish, while living in intimate contact with such a multitude of pathogens. Comparison with the innate immune systems of *Musca* with *Drosophila* (and *Anopheles*), which face different ecological pressures and pathogens, will be informative, just as the *Drosophila-Anopheles* comparison has been (Christophides et al., 2002). The relatively close relationship to *Drosophila* has already greatly expedited this analysis, as over 30 individual genes in innate immunity have been sequenced in *Musca*. The advantage of a genome sequence is that it will allow discovery of genes unique to *Musca*, and regulatory systems that allow it to survive in a far more septic environment. Moreover, it has become clear that house fly biology is closely linked to microbes. Unlike Drosophila and several other Diptera, development of house fly larvae is strictly dependant on a live and active microbial community in a natural developmental habitat. Larvae cannot develop beyond the first instar in sterilized a natural or artificial substrate/medium. The principle of this symbiosis is unknown although it has been shown that different bacteria support the house fly development to different degrees. Moreover, it has become clear that house fly biology is closely linked to microbes.

Given the tremendous importance of house flies in the transmission of human and animal diseases, substantial effort has been made to control this pest. Availability of the house fly genome will identify important target sites and will allow for the development of selective new insect control agents. Identification of novel target sites in the house fly will also aid in the development of new insecticides for control of agricultural pests that limit the supply (and quality) of human foods. A genome sequence would also provide an opportunity to explore biological control in novel ways, including disruption of the unusual autosome-based sex determination system, sterile male release, confounding signals for mate recognition, etc. Such approaches may be safer than insecticides, given the proximity of house flies to humans, animals and many of their important food sources.

The biochemistry and genetics of insecticide resistance have been well studied in the house fly, arguably more widely than in any other insect, due to the economic importance of house flies, the relatively rapid rate at which they develop resistance, and because the house fly has proven to be a useful model for understanding and predicting resistance in other insect species. Availability of the house fly genome would allow for more rapid identification of the genes and regulatory sequences involved in resistance to insect control agents.

Sequencing of the parasitoid wasp *Nasonia vitripennis* is expected to be completed early this year. *Nasonia* is a parasitoid of the house fly (sold commercially for fly control). Having the genome of both the parasitoid (*Nasonia*) and the host (*M. domestica*) will allow *the first opportunity to study the interactions between a pest insect and its parasitoid at the whole genome level.* How do *Nasonia* eggs evade the immune response of the house fly? What is the response of the house fly to *Nasonia* venoms? What immunological factors permit some flies, but not others to destroy the developing *Nasonia* embryo? Sequencing of the house fly genome is strongly supported by the *Nasonia* community (see letter from J. H. Werren).

Not all aspects of house fly biology have a negative consequence to humans. Maggot therapy, the use of late instar larvae to control infection has been used for centuries by several human societies. These dipteran larvae (primarily blow flies), in addition to consuming bacteria and necrotic tissue, produce a "healing secretion" (a cocktail of enzymes) with potent antibiotic activities. Identification of the genes responsible for these antibiotics would be of enormous value to human health (Kokoza et al., 2000; Metlitskaia et al., 2004; van der Biezen, 2001), especially given the increase in antibiotic-resistance bacteria (Lerch et al., 2003; Wollina et al., 2000).

Many species of arthropods are the sources of potent allergens that sensitize and induce IgE-mediated allergic reactions in humans. Most of these arthropod allergens are proteins, and the allergic response mechanism to these allergens is the same as it is for allergens from other sources such as plant pollens, molds, and foods. A large number of people affected by allergic reactions to stinging insects, cockroaches, and dust mites. Allergies to house flies are rare, but cases of respiratory allergy from occupational exposure (farmers) have been reported (Focke et al., 2003; Wahl and Fraedrich, 1997). Identification of house fly allergens could lead to recombinant allergens with a potential use in diagnosis and immunotherapy.

Providing better annotation and understanding of the human genome. The completed genome sequences of *Drosophila melanogaster* and *Anopheles gambiae* have been extremely valuable for deductions about the evolutionary origins, structure, and even the function of many human genes (Kortschak et al., 2003). Nevertheless, a significant number of gene modifications and extensive gene loss has occurred in *Drosophila*. Although the genomes and proteomes of *An. gambiae* and *D. melanogaster*, which diverged about 220-240 million years ago (MYA) (Wiegmann et al., 2003), reveal considerable similarities, both lineages have experienced multiple gene acquisitions and losses, especially through expansions and contractions of gene families (Zdobnov et al., 2002). Sequences of orthologous genes in these two insect species have diverged to the point that synonymous positions are virtually randomized (Zdobnov et al., 2002). It was hoped that regulatory regions of genes would become clear by comparison of 5' regions of genes in *Drosophila* and *Anopheles*, but this has proven to be much more difficult. For example, the 5' and 3' regulatory flanking regions of many genes in house flies are virtually unalignable to the orthologous sequence in *Drosophila* (Shaw et al., 2001); the genetic cascades regulating sex determination of the house fly and *D. melanogaster* appear strikingly different, and the upstream regulators of sex determination genes are different between these two insect species (Dübendorfer et al., 2002). Furthermore, 24% of *Apis* ESTs showed better matches to Chordata than to *Drosophila* genes (Whitfield et al., 2002). Some *Apis* ESTs showed significant matches to human sequences, but no matches to the *Drosophila* genome, (inferred to be genes that were lost from *Drosophila*). Similar results have also been identified in the current house fly EST sequences (N. Liu, unpublished). While either *Drosophila* or *An. gambiae* (or both) homologs could be recognized for more than half of the house fly EST sequences, some of these EST

sequences showed better matches to other more distant species, such as *Plasmodium falciparum*, *Carassius auratus* (goldfish), and *Homo sapiens*, than to *Drosophila* and/or *An. gambiae* homologs. Some of the *Musca* EST sequences showed no matches to the *Drosophila* and/or *An. gambiae* genome. These results indicate that the genomic sequences from other insect species will be extremely important for linking human genes to their *Drosophila* or *An. gambiae* homologs.

Understanding the evolution of *cis*-regulatory sequences in *Drosophila* has proven difficult in some cases (e.g. *achaete-scute* genes), because the patterns of expression are not substantially different between *Drosophila species*, but are so extremely diverged in *Anopheles* that analysis is difficult. Thus, the house fly genome would provide a critical resource for the analysis of *cis*-regulatory sequences in *Drosophila*.

A completed house fly genome will provide new information to fill gaps in our understanding of animal genome evolution. For example, the house fly genome will provide a valuable out-group for analyses of *Drosophila* genomes, given their more recent common ancestor (compared to *Drosophila* vs. *Anopheles*). The deepest common ancestor to the set of *Drosophila* species whose genomes are being sequenced is estimated to be 60-40 MYA, and the common ancestor between *D. melanogaster* and *M. domestica* has been estimated to be approximately 100 MYA (Beverley and Wilson, 1984). This places it remarkably well in the gap between *Drosophila* and *Anopheles*, and will allow a very broad evolutionary analysis across the Dipteran order. Other dipteran genomes that are in initial stages are *Aedes aegypti*, the vector for yellow fever, and the tsetse fly, vector of sleeping sickness. *Aedes* has an extensive BAC-end and EST sequencing project running at TIGR, and it seems likely that a whole genome shotgun will be launched in the near future. Preliminary trial sequencing is under way for the tsetse fly genome, and plans for the full project are being made under by a consortium formed under the auspices of the United Nation's Tropical Disease Research program.

Suitable outgroups for genome comparisons. The concept of "best" outgroup(s) depends on the questions that one would like to address. House flies make a good choice in particular, because they likely share an ancestor with *Drosophila* within 60-40 MYA which is much deeper than *melanogaster* vs. *pseudoobscura* comparisons and is old enough for major changes to occur, but not so divergent as *Drosophila* vs. *Anopheles* which have very divergent genomes (240-220 MYA). Choosing an outgroup for mosquito genomes is a similarly complex issue. The house fly genome is an excellent choice here because it provides an alternative comparison of similar-aged divergence to the *Drosophila*/mosquito one. The house fly represents a separate major lineage of cyclorrhaphan flies (calyptratae) from *Drosophila*. Multiple, deeply divergent, comparisons within the order allows identification of lineage effects on rates and patterns of genomic diversity. These comparisons become more powerful in elucidating genome evolution as the phylogenetic context is broadened. Another point for consideration is that a single outgroup is much like an experiment with n=1. If the outgroup is “unusual” for the trait being investigated, then it does not serve the purpose of a suitable method for rooting a phylogenetic tree or examining patterns of genomic evolution. Multiple outgroups provide a more robust analysis. Some in the *Drosophila* community have called for a genome project for *Scaptodrosophila*, but *M. domestica* would serve this role as well (or better for some analyses) and has more compelling justification because of its serious pest status. Choice of an organism for genome sequencing requires several criteria be met, including an available colony (preferably inbred), known (and preferably small) genome size, known markers and linkage groups, etc. The

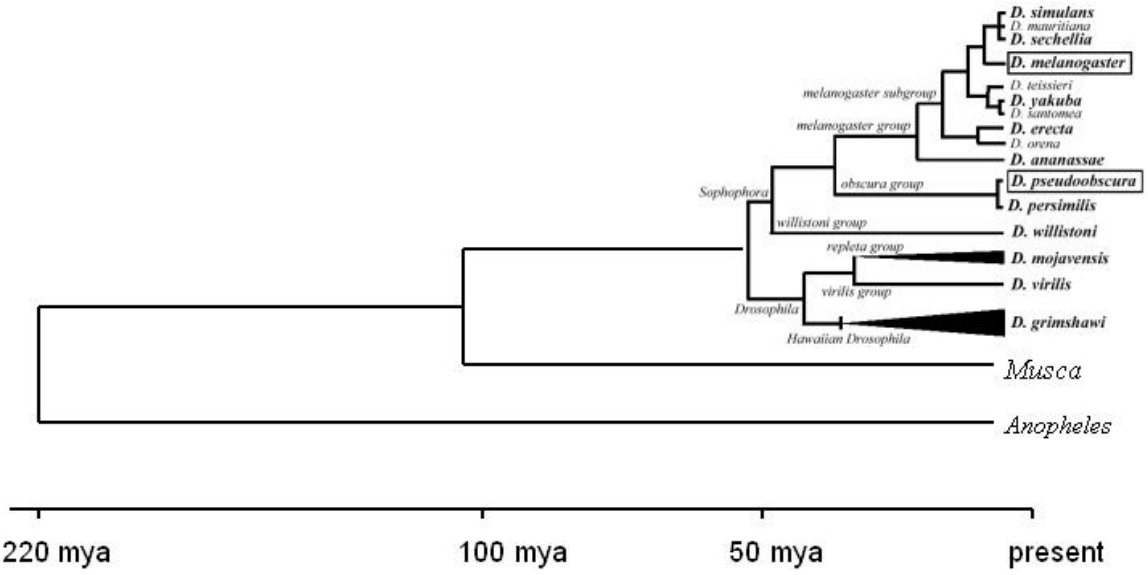


Figure 1. Phylogeny of the Diptera showing the placement of *M. domestica* relative to *Drosophila* and *Anopheles*.

Expanding our understanding of basic biological processes relevant to human health, including developmental biology and neurobiology. The house fly has been, and continues to be a major insect for studies of environmental toxicants. It has had a preeminent role in insect toxicology studies, especially with focus on insecticide resistance (Scott, 1999), comparative toxicity between insects and mammals, and in the development of new insecticides (Casida and Quistad, 2004). The house fly has been a model insect for these scientific areas of inquiry and completion of its genome will facilitate this research by the identification of novel target sites, further elucidation of differences in target sites (ion channels, neuroreceptors, hormones, etc.) between insects and mammals, and by facilitating the identification of genes involved in insecticide resistance. The neonicotinoids serve as an excellent example of the payoff that comes from this comparative biochemical approach. Neonicotinoids are the fastest growing class of insecticides, and were developed specifically by the process of selecting agents that interact with insect and not mammalian receptors (Matsuda et al., 2001; Nauen et al., 2003; Tomizawa and Casida, 2003; Wakita et al., 2003).

Completion of the *Drosophila* and *Anopheles* genomes provided unprecedented opportunities to study insect-pathogen interactions (Christophides et al., 2002; Lazzaro and Clark, 2003; Lazzaro et al., 2004; Osta et al., 2004; Schlenke and Begun, 2003; Srinivasan et al., 2004). The house fly will be potentially of even greater value for two reasons. First, house flies live in intimate association with vertebrate pathogens such as *Helicobacter pylori* (causative agent of gastric ulcer (Li and Stutzenberger, 2000)), *Salmonella*, *Campylobacter jejuni*, *E. coli* (including toxin producing strains E105 and O157:H7 that cause food poisoning (de Jesus et al., 2004; Moriya et al., 1999)) and trachoma (i.e. transmission of *Chlamydia trachomatis* (Emerson et al., 2000; Emerson et al., 1999)). Yet house flies are remarkably resilient to pathogens. Understanding the basis for their refractoriness to many pathogens would offer important insights into ways to improve human health. Second, house fly populations in temperate climates are occasionally decimated by an entomopathogenic fungus (Zygomycetes,

Entomophthoraceae). A genome sequence would expedite the investigation of why certain populations of *Musca* are sensitive to this fungus, while others are refractory. Microarray studies using the house fly genome to investigate genes associated with pathogen exposure will be a cornerstone in future studies in this field (Jensen et al., 2001; Kalsbeek et al., 2001; Zurek et al., 2002), and would become a model-system for biological control (entomopathogenic fungi; parasitic hymenoptera; microsporidia) of insects.

In the house fly, sex is determined by a dominant factor, M, which is located on the Y chromosome in "standard" populations. Thus, males are XY^M and females are XX (Dübendorfer et al., 2002; Hiroyoshi, 1964). This is believed to be the ancestral state of sex determination in house flies (Bull and Charnov, 1977; Denholm et al., 1983). However, there are "autosomal male" (A^M) strains in which the M factor is located on one or more of the five autosomes (I-V) (Franco et al., 1982; Inoue et al., 1983; Tomita and Wada, 1989) or even rarely on X (Schmidt et al., 1997). The M factor located on Y functions biologically in a way identical to the M located on any of the other autosomes (Schmidt et al., 1997; Tomita and Wada, 1989). In the A^M strains females are XX and males are also XX (or XO) (Denholm et al., 1983; Denholm et al., 1990; Franco et al., 1982; Hiroyoshi, 1964; Wagoner, 1969). It has been suggested that autosomal males may be causally related to the evolution of insecticide resistance (Hiroyoshi, 1980) or due to a consequence of tight linkage to resistance genes (Bull, 1983; Franco et al., 1982; Kence and Kence, 1992). However, this hypothesis now has very little support based on the strong North-South cline in autosomal males that was discovered in the USA (100% of the males being III^M in Florida,100% being XY^M in Maine, and intermediate values being found in New York and North Carolina) that had no correlation with insecticide resistance (Hamm et al., 2005). Some populations are found in which males are A^M/A^M (Tomita and Wada, 1989). Such populations have females with F (feminizing factor located on autosome 4), which is epistatic to M, as a means to produce equal ratios of male and female offspring. In these populations females have become the heterogametic sex.

Despite efforts by developmental biologists, the molecular identities of the M and F factors have remained elusive. In *Drosophila*, *Sex-lethal* (*Sxl*) integrates information about the dose of X and autosomes and provides the initial switch for the sex determination cascade. In *M. domestica*, F and M are not signals for *Sxl*, and in fact *Sxl* is not even involved in sex determination. Most intriguingly, in *Drosophila*, sex determination and dosage compensation are tied to the same pathway, whereas these processes are de-coupled in *Musca*. It has been speculated that it is this decoupling that gives *Musca* the impressive flexibility and polymorphism in sex determination mechanisms (Dübendorfer et al., 2002; Hediger et al., 2004). A full genome sequence of *M. domestica* would allow immediate identification of all homologs to the *Drosophila* sex determination cascade, and would greatly accelerate discovery of the genes that cause the radical divergence in the fundamental processes of sex determination and dosage compensation (Dübendorfer et al., 2002; Hediger et al., 2004).

The house fly has been a model system for studies of insect olfaction (Kelling et al., 2002; Kelling et al., 2003) and (Z)-9-tricosene plays in important role in inter-sex communication and mate selection in house flies. Sequencing of the house fly genome will identify receptor molecules (in antennal and palpal olfactory cells) that will aid olfaction studies in this model organism, and will facilitate development of attractants for house flies to baits in management systems (Darbro and Mullens, 2004; Hanley et al., 2004).

B. Strategic issues in acquiring new sequence data.

The demand for the new sequence data. Letters of support (see Appendix) eloquently demonstrate how researchers from diverse scientific areas (genomics, proteomics, developmental biology, population genetics, evolutionary biology, etc.) would make immediate use of the *M. domestica* genome sequence to accelerate their research programs on fundamental aspects of genetics (sex determination, dosage compensation, olfaction, immunology, etc.) as well as practical problems of pest control. It is clear that this community considers sequencing of the house fly genome to be an extremely high priority.

The suitability of the organism for experimentation. The house fly has many advantages as an experimental organism for laboratory studies necessary in functional genome annotation. It is easy to rear on standard media, and large numbers can be readily produced. Under normal laboratory conditions it takes about 10 days to develop from egg to adult. House flies are highly fecund, so that thousands of house flies can be produced in a matter of weeks. Most of the protocols for basic molecular biology of *Drosophila* work well with *Musca*. In addition to the well studied genetics of the house fly, genetic markers are available for each chromosome. We propose to use DNA from the wild-type CSYM4 strain for sequencing of the house fly genome. This strain was established from a single pair of house flies for four generations, starting with the highly inbred CS (Cornell-Susceptible standard reference strain; CS) strain of house fly. The CS strain was colonized in the 1960s in the USA (i.e. inbred for approximately 1000 generations). Preliminary sequencing of five genes (*Ace1*, *Rdl*, *CYP6D1*, *VSSC1* and *Gfi-1*) has found complete homozygosity of the strain, but pilot sequencing by the genome center is still recommended to test library quality and coverage.

Insect transgenesis is critically important for both practical applications and for addressing basic scientific questions. The generation of transgenic lines of insects has proven to be perhaps the most powerful method for demonstrating the functional role of genes, both by overexpression studies and by mutation-rescue studies. Transgenic insects have resulted in significant progress in understanding the genes involved in disease transmission, and in understanding the biological and physiological roles of numerous genes. Germ line transformation of house flies has been successfully carried out by various methods (Atkinson et al., 1993; Hediger et al., 2001; O'Brochta et al., 1994; Warren et al., 1994). More recently, one of the authors (N. Liu, Auburn Univ.) has carried out *Musca* transformations to study the genes associated with insecticide resistance. Similar studies will also facilitate the identification of new target sites that could lead to the development of novel insecticides with new modes of action and low toxicity to non-target species. The availability of transgenic technology and completion of the house fly genome will open numerous areas of investigation that were previously not approachable.

The rationale for generating the complete sequence of the organism. A thorough understanding of the biology of complex organisms requires complete sequencing information and identification of all functional elements from the genomes of these organisms. The "whole genome" approach has vastly improved comparative and evolutionary studies, as well as physical map building. It has addressed several important scientific questions about genome evolution, such as evolutionary rates, speciation, genome reorganization, and origins of variation. The approach has also been important for identification of conserved sequences involved in gene

regulation and other genomic functions, identification of specific functional sequences (i.e., those that have been substituted or modified during evolution, and which have undergone recent selection (Vandahl et al., 2004)) and elucidation of sequence variation in the population of organisms (such as alternative splicing in the regulation of gene function (Tan et al., 2002)). The "whole genome" approach will also be important for identification of sequences that are broadly conserved across insect genomes to provide insight into the unique features in the genome, and for obtaining a broader and more complete assessment of the extent of genetic variation in the population of organisms; identification of variation in gene expression and understanding the evolution (Yan et al., 2002). The whole genome is also necessary to understand the interactions of house flies with the parasitoid wasp *N. vitripennis* While some of the genes that are expected to be regulated by parasitoid venoms and egg laying could be inferred from other studies, only a whole genome will provide comprehensive insight into the genes involved in host/parasitoid interactions..

The calyptrate flies, with *M. domestica* as the most prominent experimental organism, includes a large number of important vectors of human and veterinary diseases, as well as important species for forensic entomology: dog dung fly (*Musca sorbens*), face fly (*Musca autumnalis*), blow flies (*Lucillia, Calliphora, Chrysomya*), flesh flies (*Sarcophaga*), screwworm (*Cochliomyia*), tsetse fly (*Glossina*), the little house fly (*Fannia*), warble flies (*Hypoderma*), yellow dung flies (*Scathophaga*), and the root maggot fly (*Anthomyiida*). By using genetic manipulations of *M. domestica* to place function of novel genes in its genome, we anticipate that it will be easy to transfer the knowledge gained to other synanthropic flies.

Many genes, especially regulatory genes, are often expressed at a very low level and they would be rare in EST libraries. The entire house fly genome sequence will, especially when compared with the *Drosophila* and *Anopheles* genome sequences, facilitate the identification of homologous genes expressed at low levels or in a specific tissue. Expression patterns can then be validated with high throughput real-time PCR systems for use in either general population or microevolutionary studies (e.g., the spread and fitness of resistance genes).

The cost of sequencing the genome and the state of readiness of the organism's DNA for sequencing. Recently, quantitative Real-Time PCR (qRT-PCR has been shown to be a reliable method for determination of genome size (Wilhelm et al., 2003). Using this method, we compared the genome size of *D. melanogaster* and *M. domestica* (Gao and Scott, unpublished). In two experiments, with six replicates per experiment, we have determined the size of the *D. melanogaster* genome to be 180-181 Mbp (in agreement with the published genome size (Adams et al., 2000)) and the size of the house fly genome to be 309-312 Mbp, or approximately 1.7-fold larger than *D. melanogaster* (slightly smaller than the size of *D. virilus*). Having run house fly and *Drosophila* side-by-side (using homologous single copy genes) and replicating the published size of the *Drosophila* genome we are highly confident of our results. Our estimate is smaller than the size of the house fly genome originally estimated by ultraviolet microscopy (Bier and Müller, 1969) that suggested the house fly had a genome roughly five times the size of the *Drosophila melanogaster*. Unfortunately, this paper provided no details about the number of replications or level of variation in their experiments. Thus, it is difficult to reconcile the current qRT-PCR results with the only other published estimate of genome size. We would propose that a pilot run of random whole genome shotgun reads should be generated as a test both of the quality of the library, and to acquire more information about the distribution of euchromatin. Although polytene chromosomes exist in *Musca* (Vecchi and Rubini, 1973), *in situ* hybridization

has not been routinely practiced. This procedure would be very useful in linking mapping and genomics efforts, and in identification of heterochromatic regions and localization of repetitive elements. From this pilot, we will be able to design the sequencing strategy in collaboration with whichever genome center takes on the task. At this time, it appears that some sort of hybrid of the hierarchical, clone-by-clone approach and whole genome shotgun may be the most effective. The hierarchical approach has been used for most eukaryotic genome sequences, including the yeast *S. cerevisiae*, the nematode *C. elegans* (Consortium, 1998), the mustard weed *A. thaliana* (Iniative, 2000), and the human (Consortium, 2001).

A strain of house fly, CSYM4, will be used as a source for genomic DNA construction of the BAC library(s), DNA sequencing, and molecular and physical maps. The homozygosity of this strain will be important for avoiding problems in contig misassembly, which could happen in genomes with heterozygous loci (Mongin et al., 2004). 7-8 X coverage of the house fly genome will be the most informative. BAC end sequences from both ends of approximately 30,000 - 40,000 BAC clones will be generated with an average insert size of 200 kb to yield 60,000 – 80,000 sequence-tagged connectors (STCs). An expanded EST sequencing project, with a good spread of libraries from different tissues and developmental periods would assist in gene annotation and would inform the functional analysis of the genome. The whole genome shotgun project would require about 8 million sequencing reads, so an EST project of 250,000 sequences would provide a reasonable balance of effort, cost, and utility.

Current status of genome efforts in *M. domestica*. Several resources that will facilitate the assembly and annotation of the house fly genome have been funded through a variety of sources. NIHGMS supported research has led to the production of both genomic and cDNA libraries (Scott lab). An embryonic genomic library was prepared as part of ongoing studies into sex determination (D. Bopp lab) and an antennal cDNA library was constructed for studies of olfaction. As a part of USDA NRI and Auburn University Biogrant funded project, a house fly normalized cDNA library has been constructed from the mRNA of house flies. In a pilot study of this library 300 ESTs have been generated (N. Liu lab), resulting in 292 high quality cDNA sequence reads. 39 ESTs were assembled into 8 contigs. The remaining 253 ESTs are unique, suggesting a 15% redundancy in the house fly sequence set. This EST sequencing effort, combined with other larger EST projects, will be excellent resources for the genomic library (BAC library) screening and building contig maps for comparative genomic studies. In addition, five house flies genomic DNA libraries have been generated by cloning genomic DNA fragments digested by HindIII, BamHI, ScaI, EcoRI, and SspI into pUC18 vectors. Again, these house fly genomic libraries will be resources for house fly WGS sequence reads. House flies can be readily transformed with mobile elements such as piggyback, hermes or hobo (Atkinson et al., 1993; Hediger et al., 2001; O'Brachta and Atkinson, 1997; O'Brachta et al., 1996; O'Brochta and Atkinson, 1996; O'Brochta et al., 1994; Sarkar et al., 1997; Warren et al., 1994), and genes can be silenced using RNAi (Burghardt et al., 2005; McGregor et al., 2001).

Readiness of the research community. Currently there are approximately 40 laboratories worldwide whose primary research focus is the house fly. About half of these are engaged in studies of molecular biology that would immediately benefit from a complete genome sequence. Most of the others are studying aspects of toxicology and pest control, and immediate access to design of primers for PCR analysis would open the door to simple but powerful molecular approaches to this group. *Drosophila* researchers should be counted among the community that

would benefit from the sequencing of the house fly genome. The white papers that resulted in funding to sequence an additional 11 genomes of *Drosophila* species failed to include an outgroup to the set of *Drosophila* species, and *Anopheles gambiae* is just too distantly related for optimal analysis (in most cases).

The Danish Pest Infestation Laboratory has been a center of excellence in house fly research since the founding in 1948. House fly biology and especially insecticide resistance has been a primary research topic at DPIL. Numerous strains of house flies are bred and made available for the international research community. House fly research at DPIL has recently adopted molecular approaches, and has also expanded beyond chemical control to include the interaction with biological control agents such as entomopathogenic fungi. DPIL recently merged with the Danish Institute of Agricultural Research, which has a high level of activity in genomics of domestic animals (i.e. identification and application of DNA variation as markers for health and production traits, mapping of hereditary defects, genome scanning and comparative genome analysis) and is currently working on a joint venture to sequence the porcine genome and developing SNPs (Li, 2000). Michael Kristensen (one of the authors of this white paper) works at DPIL.

Bioinformatics. Initial assembly of the house fly genome sequence will be performed at the sequencing center, but it is our goal to make the sequence available to the greater research community as rapidly as possible. Sequence data will be rapidly and regularly released into public databases, prior to assembly. It is likely that the bioinformatics community will perform more than one assembly, judging from their interest in the *Drosophila* genomes, and the *Musca* genome will be of immediate interest to see how it has accomplished the approximately 2-fold expansion relative to *D. melanogaster*. Once there is a shaking out of the assemblies, the most arduous bioinformatics task that will likely not be handled by the genome center is annotation. We would propose to host a genome annotation "jamboree", much in the fashion of the jamboree that was held at Celera Genomics upon its first assembly of the *Drosophila melanogaster* genome sequence. Michael Ashburner and others from the *Drosophila* community have already volunteered to assist in the annotation process, after the automatic annotators have made their first pass. Michael Eisen's group at Berkeley has been one of the most active in pulling together the data from the *Drosophila* genomes sequenced to date, and he is eager to run the pipelines that they have developed for *Drosophila* on the *Musca* genome as well. The plan would be to begin by overlaying as much annotation onto *Musca* by homology to *Drosophila*, and to then explore the remainder armed with BLAST pre-computes and detailed synteny information. By gathering investigators in one place in one intensive effort, there would be an economy of scale in producing an expert-reviewed annotation.

As the annotation information is generated, it needs to be made available in a form that best supports the research efforts by the broader community. This includes features like a genome browser, selectable ftp site, BLAST server, and a SQL interface. FlyBase serves as the model for a tool that provides rapid access to a heterogeneous set of attributes for genomic content. We approached Dr. William Gelbart with the suggestion that *Musca*, as the best outgroup to the *Drosophila* genomes, would be an excellent addition to the FlyBase database. He was enthusiastic about the house fly genome white paper and replied that *Musca* could either be served by FlyBase or InsectBase (pending the outcome of his proposal on the latter). The power of the *Musca* genome informatics comes from comparative genomics, so being imbedded in these multi-species databases is ideal.

References

Adams, M. D., et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185-2195.

Aksoy, S., Merriman, M., Hall, N., Masahira, H., Hide, W., Lehane, M. J., 2005. A case for a Glossina genome project. Trends in Parasitology 21, 107-111.

Atkinson, P. W., Warren, W. D., O'Brochta, D. A., 1993. The hobo transposable element of Drosophila can be cross-mobilized in houseflies and excises like the Ac element of maize. Proc. Natl. Acad. Sci. USA 90, 9693-9697.

Beverley, S. M., Wilson, A., 1984. Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. J. Molec. Evol. 21, 1-13.

Bier, K., Müller, W., 1969. DNA-Messungen bei Insekten und eine Hypothese über retardierte Evolution und besonderen DNA-Reichtum in Tierreich. Biologisches Zentralblatt 88, 425-449.

Bull, J. J. 1983. "Evolution of Sex Determining Mechanisms." Benjamin/Cummings, Menlo Park, CA.

Bull, J. J., Charnov, E. L., 1977. Changes in the heterogametic mechanism of sex determination. Heredity 39, 1-14.

Burghardt, G., Hediger, M., Siegenthaler, C., Moser, M., Dubendorfer, A., Bopp, D., 2005. The *transformer2* gene in *Musca domestica* is required for selecting and maintaining the female pathway of development. Dev. Genes Evol 215, 165-176.

Casida, J. E., Quistad, G. B., 2004. Why insecticides are more toxic to insects than people: the unique toxicology of insects. J. Pestic. Sci. 29, 81-96.

Christophides, G. K., et al., 2002. Immunity-related genes and gene families in *Anopheles gambiae*. Science 298, 159-165.

Consortium, C. e., 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282, 2012-2018.

Consortium, I. H. G. S., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Darbro, J. M., Mullens, B. A., 2004. Assessing insecticide resistance and aversion to methomyl-treated toxic baits in *Musca domestica* L (Diptera: Muscidae) populations in southern California. Pest Manag Sci 60, 901-908.

de Jesus, A. J., Olsen, A. R., Bryce, J. R., Whiting, R. C., 2004. Quantitative contamination and transfer of *Escherichia coli* from foods by houseflies, *Musca domestica* L. (Diptera: Muscidae). Intern. J. Food Microbiol. 93, 259-262.

Denholm, I., Franco, M. G., Rubini, P. G., Vecchi, M., 1983. Identification of a male determinant on the X chromosome of housefly (*Musca domestica* L.) populations in South-East England. Genet. Res. Camb. 42, 311-322.

Denholm, I., Rubini, P. G., Rovati, C., Vecchi, M., 1990. Genetic basis of sex determination in two South African strains of house fly. S. African J. Sci. 86, 41-43.

Dübendorfer, A., Hediger, M., Burghardt, G., Bopp, D., 2002. *Musca domestica*, a window on the evolution of sex-determining mechanisms in insects. Int. J. Dev. Biol. 46, 75-79.

Emerson, P. M., Bailey, R. L., Mahdi, O. S., 2000. Transmission ecology of the fly Musca sorbens, a putative vector of trachoma. Trans. Royal Soc. Trop. Med. Hygiene 94, 28-32.

Emerson, P. M., Lindsay, S. W., Walraven, G. E. L., Faal, H., Bogh, C., Lowe, K., Bailey, R. L., 1999. Effect of fly control on trachoma and diarrhea. Lancet 353, 1401-1403.

Focke, M., Hemmer, W., S., W., 2003. Specific sensitization to the common housefly (Musca domestica) not related to insect panallergy. Allergy 58.
Franco, M. G., Rubini, P. G., Vecchi, M., 1982. Sex-determinants and their distribution in various populations of *Musca domestica* L. of Western Europe. Genet. Res. Camb. 40, 279-293.
1994. Economic Significance of Poultry. *In* "Research and Extension needs for Integrated Pest Management for Arthropods of Veterinary Importance" (C. J. Geden, J. A. Hogsette, Eds.), Gainesville, FL.
Graczyk, T. K., Knight, R., Gilman, R. H., Cranfield, M. R., 2001. The role of non-biting flies in the epidemiology of human infectious diseases. Microbes and Infection 3, 231-235.
Greenberg, B., 1965. Flies and disease. Sci. Amer. 213, 92-99.
Hamm, R., Shono, T., Scott, J. G., 2005. A cline in frequency of autosomal males is not associated with insecticide resistance in house fly, (Diptera: Muscidae). J. Econ. Entomol. 98, 171-176.
Hanley, M. E., Dunn, D. W., Abolins, S. R., Goulson, D., 2004. Evaluation of (*Z*)-9-tricosene baited targets for control of the housefly (*Musca domestica*) in outdoor situations. JEN 128, 478-482.
Hediger, M., Burghardt, G., Siegenthaler, C., Buser, N., Hilfiker-Kleiner, D., Dubendorfer, A., Bopp, D., 2004. Sex determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal regulator *doublesex*. Dev. Genes Evol. 214, 29-42.
Hediger, M., Niessen, M., Wimmer, E. A., Dübendorfer, A., Bopp, D., 2001. Genetic transformation of the housefly *Musca domestica* with the lepidopteran derived transposon piggyBac. Insect Molec. Biol. 10, 113-119.
Hiroyashi, T., 1960. Some new mutants and linkage groups of the house fly. J. Econ. Entomol. 53, 985-.
Hiroyoshi, T., 1964. Sex-limited inheritance and abnormal sex ratio in strains of the housefly. Genetics 50, 373-385.
Hiroyoshi, T., 1977. Some new mutants and revised linkage maps of the housefly, *Musca domestica* L. Japan J. Genetics 52, 275-288.
Hiroyoshi, T., 1980. Formal genetics of the housefly in relation to insecticide resistance. Proceedings of the 16th International Congress of Entomology 1, 397.
Iniative, T. A. G., 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796-815.
Inoue, H., Fukumori, Y., Hiroyoshi, T., 1983. Mapping of autosomal male-determining factors of the housefly, *Musca domestica* L., by means of sex reversal. Jpn. J. Genet. 58, 451-461.
Jensen, A. B., Thomsen, L., Eilenberg, J., 2001. Intraspecific variation and host specificity of *Entomophthora muscae sensu stricto* isolates revealed by random amplified polymorphic DNA, universal primed PCR, PCR-restriction fragment length polymorphism, and conidial morphology. J. Invert. Pathol.
Kalsbeek, V., Mullens, B. A., Jespersen, J. B., 2001. Field studies of Entomophthora (Zygomycetes: Entomophthorales) - induced behavioral fever in *Musca domestica* (Diptera: Muscidae) in Denmark. Biol. Cont. 21, 264-273.
Keiding, J., 1986. The house fly: biology and control. WHO Vector Control Series, 63.
Kelling, F. J., Biancaniello, G., den Otter, C. J., 2002. Electrophysiological characterization of olfactory cell types in the antennae and palps of the housefly. J. Insect Physiol. 48, 997-1008.
Kelling, F. J., Biancaniello, G., J., d. O. C., 2003. Effect of age and sex on the sensitivity of

antennal and palpal olfactory cells of houseflies. Ent.Exp.Appl. 106, 45-51.
Kence, M., Kence, A., 1992. Genetic consequences of linkage between malathion resistance and an autosomal male-determining factor in house fly (Diptera: Muscidae). J. Econ. Entomol. 85, 1566-1570.
Kokoza, V., Ahmed, A., Long, C. W., Jasinskiene, N., James, A., Raikhel, A., 2000. Engineering blood-meal activated systemic immunity in the yellow fever mosquito, *Aedes aegypti*. Proc. Nat. Acad. Sci. 97, 9144-9149.
Kortschak, R. D., Samuel, G., Saint, R., Miller, D. J., 2003. EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. Current Biol. 13, 2190-2195.
Lazzaro, B. P., Clark, A. G., 2003. Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Molec. Biol. Evol. 20, 914-923.
Lazzaro, B. P., Sceurman, B. K., Clark, A. G., 2004. Genetic basis of natural variation in *D. melanogaster* antibacterial immunity. Science 303, 1873-1876.
Lerch, K., Linde, H. J., Lehn, N., 2003. Bacteria ingestion by blowfly larvae: An *in vitro* study. Dermatology 207, 362-366.
Li, H., 2000. Gene sequencing. China, Denmark team up to tackle the pig. Science, 913-914.
Li, S. E., Stutzenberger, F. J., 2000. The housefly (*Musca domestica*) as a possible vector for *Helicobacter pylori* at agricultural sites. Intern. J. Environ. Health Res. 10, 141-152.
Maisnier-Patin, S., Andersson, D. I., 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res. Microbiol. 155, 360-369.
Matsuda, K., Buckingham, S. D., Kleier, D., Rauh, J. J., Grauso, M., Sattelle, D. B., 2001. Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors. Trends in Pharmacological Sciences 22, 573-580.
McGregor, A. P., Shaw, P. J., Hancock, J. M., Bopp, D., Hediger, M., Wratten, N. S., Dover, G. A., 2001. Rapid restructuring of bicoid-dependent hunchback promoters within and between Dipteran species: implications for molecular coevolution. Evol. Devol. 3, 397-407.
Metlitskaia, L., Cabralda, J. E., Suleman, D., Kerry, C., Brinkman, J., Bartfeld, D., Guarna, M. M., 2004. Recombinant antimicrobial peptides efficiently produced using novel cloning and purification processes. Biotech. Appl. Biochem. 39, 339-345.
Milani, R., Rubini, P. G., Franco, M. G., 1967. Sex-determination in the housefly. Genetica Agaria 21, 385-411.
Mongin, E., Louis, C., Holt, R. A., Birney, E., Collins, F. H., 2004. The *Anopheles gambiae* genome: An update. Trends-in-Parasitology, 49-52.
Moriya, K., Fujibayashi, T., Yoshihara, T., Matsuda, A., Sumi, N., Umezaki, N., Kurahashi, H., Agui, N., Wada, A., Watanabe, H., 1999. Verotoxin-producing *Escherichia coli* 0157: H7 carried by the housefly in Japan. Med. Vet. Ent. 13, 214-216.
1994. Housefly, an Everyday Monster (K. Mündi, Ed.). Fox Lorber, New York, NY.
Nauen, T., Ebbinghaus-Kintscher, U., Salgado, V., Kaussmann, M., 2003. Thiamethoxam is a neonicotinoid precursor converted to clothianindin in insects and plants. Pestic. Biochem. Physiol. 76, 55-69.
Nickel, C. A., Wagoner, D. E., 1974. Mutants on Linkage Gropus 3 and 4 of the House Fly. Ann. Entomol. Soc. Am. 67, 775-776.
O'Brachta, D. A., Atkinson, P. W., 1997. Recent Developments in Transgenic Insect Technology. Parasitology Today 13, 99-104.
O'Brachta, D. A., Warren, W. D., Saville, K. J., Atkinson, P. W., 1996. *Hermes,* a Functional

Non-Drosophilid Insect Gene Vector From *Musca domestica*. Genetics 142, 907-914.
O'Brochta, D., Atkinson, P., 1996. Transposable elements and gene transformation in non-Drosophilid insects. Insect Biochem. Molec. Biol. 26, 739-753.
O'Brochta, D. A., Warren, W. D., Saville, K. J., Atkinson, P. W., 1994. Interplasmid transposition of Drosophila hobo elements in non-drosophilid insects. Mol. Gen. Genet. 244, 9-14.
2004. Trachoma. http://www.who.int/pbd/blindness/trachoma/en/.
Osta, M. A., Christophides, G. K., Vlachou, -. D., Kafatos, F. C., 2004. Innate immunity in the malaria vector Anopheles gambiae: comparative and functional genomics. J. Exper. Biol. 207, 2551-2563.
Reed, D. H., Bryant, E. H., 2004. Phenotypic correlations among fitness and its components in a population of the housefly. J. Evol. Biol. 14, 919-923.
Sarkar, A., Coates, C. J., Whyard, S., Willhoeft, U., Atkinson, P. W., O'Brachta, D. A., 1997. The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. Genetica 99, 15-29.
Sasaki, T., Kobayashi, M., Agui, N., 2000. Epidemiological potential of excretion and regurgitation by *Musca domestica* (Diptera: muscidae) in the dissemination of *Escherichia coli* O157: H7 to food. J. Med. Entomol. 37, 945-949.
Schlenke, T. A., Begun, D. J., 2003. Natural selection drives *Drosophila* immune system evolution. Genetics 164, 1471-1480.
Schmidt, R., Hediger, M., Roth, S., Nothiger, R., Dübendorfer, A., 1997. The *Y*-chromosomal and autosomal male-determining *M* factors of *Musca domestica* are equivalent. Genetics 147, 271-280.
Scott, H. G., Lettig, K. S. 1962. "Flies of Public Health Importance and their Control." U.S. Government Printing Office, Washington, D.C.
Scott, J. G., 1999. Molecular basis of insecticide resistance: cytochromes P450. Insect Biochem. Molec. Biol. 29, 757-777.
Shaw, P. J., Salameh, A., McGregor, A. P., Bala, S., Dover, G. A., 2001. Divergent structure and function of the bicoid gene in Muscoidea fly species. Evol. Devol. 3, 251-259.
Srinivasan, P., Abraham, E. G., Ghosh, A., K., Valenzuela, J., Ribeiro, J. M. C., Dimopoulos, G., Kafatos, F. C., Adams, J. H., Fujioka, H., Jacobs-Lorena, M., 2004. Analysis of the *Plasmodium* and *Anopheles* transcriptomes during oocyst differentiation. J. Biol. Chem. 279, 5581-5587.
Sundin, G. W., 1996. Evolution and Selection of Antibiotic and Pesticide Resistance: A Molecular Genetic Perspective. In: (T. M. Brown, Ed.), Molecular Genetics and Evolution of Pesticide Resistance, pp. 97-105. American Chemical Society, Washington, D.C.
Tan, J., Z. Liu, Z., Nomura, Y., Goldin, A. L., Dong, K., 2002. Alternative splicing of an insect sodium channel gene generates pharmacologically distinct sodium channels. J. Neurosci. 22, 5300-5309.
Tomita, T., Wada, Y., 1989. Multifactorial sex determination in natural populations of the housefly (*Musca domestica*) in Japan. Jpn. J. Genet. 64, 373-382.
Tomizawa, M., Casida, J. E., 2003. Selective toxicity of neonicotinoids attributable to specificity of insect and mammalian nicotinic receptors. Annu. Rev. Entomol. 48, 339-364.
Tsukamoto, M., Baba, Y., Hiraga, S., 1961. Mutations and linkage groups in Japanese strains of the housefly. Japan J. Genetics 36, 168-174.
van der Biezen, E. A., 2001. Quest for antimicrobial genes to engineer disease-resistant crops.

Trends Plant Sci. 6, 89-91.
Vandahl, B. B. S., Birkelund, S., Christiansen, G., 2004. Genomic and proteome analysis of *Chlamydia*. Proteomics 4, 2831-2842.
Vecchi, M. L., Rubini, P. G., 1973. Polytene chromosomes of Trichogen-cells in *Musca domestica* L. Ricevuto, 456-463.
Wagoner, D. E., 1969. Presence of male determining factors found on three autosomes in the house fly, *Musca domestica*. Nature 223, 187-188.
Wahl, R., Fraedrich, J., 1997. Occupational allergy to the housefly (*Musca domestica*). Allergy 52, 236-238.
Wakita, T., Kinoshita, K., Yamada, E., Yasui, N., Kawahara, N., Naoi, A., Nakaya, M., Ebihara, K., Matsuno, H., Kodaka, K., 2003. The discovery of dinotefuran: a novel neonicotinoid. Pest. Manag. Sci. 59, 1016-1022.
Warren, W. D., Atkinson, P. W., O'Brochta, D. A., 1994. The Hermes transposable element from the house fly, Musca domestica, is a short inverted repeat-type element of the hobo, Ac, Tam3 (hAT) element family. Genet. Res. Camb. 64, 87-97.
Whitfield, C. W., Band, M. R., Bonaldo, M. F., Kumar, C. G., Liu, L., Pardinas, J. R., Robertson, H. M., Soares, M. B., Robinson, G. E., 2002. Annotated expressed sequence Tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Res. 12, 555-566.
Wiegmann, B. M., Yeates, D. K., Thorne, J. L., Kishino, H., 745-756., S. B., 2003. Time flies, a new molecular time-scale for brachyceran fly evolution without a clock. Systematic Biol. 52, 745-756.
Wilhelm, J., Pingoud, A., Hahn, M., 2003. Real-time PCR-based method for the estimation of genome sizes. Nucleic Acids Research 31, e56.
Wollina, U., Karte, K., Herold, C., 2000. Biosurgery in wound healing - The renaissance of maggot therapy. Journal of the European academy of dermatology and venereology 14, 285-289.
Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., Kinzler, K. W., 2002. Allelic variation in human gene expression. Science, 1143.
Yeates, D. K., Wiegmann, B. M., 1999. Congruence and controversy: toward a higher-level phylogeny of Diptera. Annu. Rev. Entomol. 44, 397-428.
Zdobnov, E. M., et al., 2002. Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science 298, 149-159.
Zurek, L., Watson, D. W., Krasnoff, S. B., Schal, C., 2002. Effect of the entomopathogenic fungus, *Entomophthora muscae* (Zygomycetes: Entomophthoraceae), on sex pheromone and other cuticular hydrocarbons of the house fly, *Musca domestica*. J. Invert. Pathol. 80, 171-176.

APPENDIX

Letters of support for sequencing of the house fly genome

Ashburner, Michael (Cambridge)

Bloomquist, Gary & Claus Tittiger (University of Reno)

Bopp, Daniel & Monica Hediger (University of Zurich)

Broce, Alberto (Kansas State Univ.)

Çağlar, Selim (Hacettepe University)

Casida, John (UC Berkeley)

Collins, Frank (Notre Dame)

Denholm, Ian & Martin Williamson (Rothamstead)

Drosophila Board (Mark A. Krasnow, Stanford University)

Eisen, Michael (UC Berkeley)

Giangaspero, Annunziata & Domenico Otranto (Univ. degli Studi di Foggia)

Kafatos, Fotis (Imperial College)

Lazzaro, Brian (Cornell University)

Mullens, Bradley (University of California, Riverside)

Ozoe, Yoshihisa (Shimane University)

Simpson, Pat (University of Cambridge)

Sparks, Thomas C. (Dow AgroSciences)

Tomita, Takashi, Shinji Kasai, Osamu Komagata & Toshio Shono (National Institute of Infectious Diseases, Japan)

Werren, John H. (University of Rochester)

Wiegmann, Brian (NC State University)

Zurek, Ludek (Kansas State University)

To: jgs5@cornell.edu
Subject: Re: Musca domestica whitepaper and Letters of Support

Dear Jeff,

Thank you for sending me the draft White Paper making the case for a genomic sequence of Musca.

Even as a drosophilist I have long been fascinated by Musca. Like Drosophila melanogaster it is a cosmopolitan synanthropic species. Unlike Drosophila, which only does good in this World (think of how bad it would be if we did not have Drosophila to transmit yeast between grapes !), Musca is a serious disease vector. Many years ago I purchased and read Bernard Greenberg's wonderful two volumes 'Flies and Disease' (Princeton 1971) which includes 67 pages of a table listing the associations then known between Musca (as a genus, but mostly domestica) and other organisms, mostly viruses, bacteria and fungi. It is interesting that this seems to have been funded by the US Army. That would make sense, since Musca can transmit anthrax. However, so much emphasis has been put on the hematophagous vectors in recent years, that the importance of the passive vectors such as Musca seems to have been rather overshadowed. As your White Paper clearly summarizes the cost to human health of Musca is very considerable.

Musca is very interesting from a population genetics point of view. It has, again as reviewd in your document, an extraordinarily complex variety of chromosomal sex determining mechanisms, with, for example populations that are male heterogametic, are female heterogametic, that have the male determining factor on different chromosomes
and so forth. Moreover Musca was one of the first organisms subjected to massive control by insecticides, especially DDT. In some regions, eg. the Po Valley in northern Italy, these were almost military campaigns, as recounted to me many years ago by Milani in Pavia. It is not, therefore, surprising that Musca is an ideal model for the study of the mechanisms, evolution and spread of insecticide resistance.

Above all, however, I am excited by this project because of of what it promises to comparative genomics. The case that Musca is, evolutionarily, 'midway' between Anopheles and Drosophila is well made in your White Paper. It could form a great 'bridge' for the comparative analysis of both genomes. Its relatively large genome size, in comparison to Drosophila, is also of great interest. Thus I see a Musca genome as being of great value not only for the 'run of the mill' comparative genomic purposes (improving annotations of genes and nod-coding sequences), great as that will be, but also for its contribution to our understanding genome evolution at the whole genome level. As you know from the work of the Pavia group there is evidence (pre-molecular) for syntenic conservation of chromosome arms between Musca and Drosophila.

Should you wish me to help in any way in the annotation of the Musca sequence then please let me know. I would be delighted to be involved. We must clearly plan for a close collaboration between that effort and FlyBase, though it is probably premature to be too specific about that right know.

Let me know if I can be of further help.

Best wishes,

Michael Ashburner

December 28, 2004

Jeffrey G. Scott
Daljit S. and Elaine Sarkaria Professor of Insect Physiology and Toxicology
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

Dear Jeff,

This letter is to assure you of our enthusiastic support for your proposal to sequence the genome of the housefly, *Musca domestica*.

Our research in *M. domestica* focuses on the endocrine regulation and biochemistry of cuticular hydrocarbon biosynthesis. Since all insects rely on cuticular hydrocarbons as a defence against dessication, results from this work may be used to develop new pest control strategies. Furthermore, some cuticular hydrocarbons serve as contact sex pheromones. Female housefly pheromone components include various long-chain compounds derived from fatty acids, and their synthesis is regulated by ecdysone. We are now in the process of identifying and purifying cDNA clones for the enzymes involved. These include various cytochromes P450 and fatty acyl elongases, among others.

An annotated *M. domestica* genome will be an invaluable resource that will greatly help with our work, particularly in understanding how genes are regulated during pheromone biosynthesis and/or hydrocarbon production. Comparisons with other known genomes will provide insight into how these processes have evolved in insects.

With best wishes for a successful proposal,

Sincerely,

Gary J. Blomquist, Professor and Chair

Claus Tittiger, Associate Professor

Universität Zürich
Zoologisches Institut · Entwicklungsbiologie

Winterthurerstr. 190
CH-8057 Zürich
Tel. +41 1 635 4869
Fax +41 1 635 6823
dbopp@zool.unizh.ch

www.zool.unizh.ch

Dr. Daniel Bopp

Dr. J.G. Scott
Department of Entomology
Cornell University
Ithaca, NY 14853
USA

Zürich, January, 4th 2005

Dear Jeff

We are writing to express our strong and enthusiastic support for sequencing of the house fly genome. Our research group has a longstanding and successful record of genetic and molecular studies with *Musca domestica*. We are exploiting this system to improve our understanding of how sex-determining pathways have evolved in insects.

How the sexual fate of an individual is determined poses a fascinating problem in biology and has attracted the attention of many researchers since the dawn of genetic studies. Though sex determination is based on a simple binary decision between two alternative developmental programs, male or female, we have only begun to understand the underlying genetic control in a few model systems. In insects, this pathway has been most extensively studied in *Drosophila melanogaster*. This work has led to a detailed description of the underlying genetic architecture and sex determination has become one of the best-studied developmental pathways in this model organism. However, from comparative studies in other dipteran insects it was evident that the instructive part of the Drosophila pathway does not represent a conserved mode of operation. On the contrary, an astounding diversity of sex-determining mechanisms seems to exist at this level in the insect world. For instance, some species make use of dominant Mendelian cues, commonly referred to as *M* when determining male development or *F* when determining the female fate. In other species it is the genetic make-up of the mother that decides whether her progeny will be male (arrhenogenic) or female (thelygenic). The spectrum of sex-determining cues is broad and extends from quantitative chromosomal signals (e.g. differences in ploidy) to environmental cues, such as temperature or population

density. What has specifically attracted our attention to *Musca domestica* is the existence of seemingly different strategies in one and the same species. The spectrum ranges from the use of dominant male or female determiners to a system in which sex is solely determined by the maternal genotype. This special feature makes the housefly a perfect model system for studying evolutionary changes in the sex determination pathway. The different mechanisms seem to have evolved in a relatively short period suggesting that the observed variations are subtle changes in an otherwise well conserved pathway. By identifying the components and their molecular functions in the Musca pathway and comparing them to those found in other fly species we may be able to define a basic principle that is operational in many species. If such a conserved core mechanism does exist it will provide an ideal target for developing effective sexing strategies which can be applied to Sterile Insect Technique based programs to control the spreading of pest fly species of medical and agricultural relevance.

With the recent isolation of a *Musca transformer* homologue (*Mdtra*) we made an important move forward towards understanding the basic principle by which sex is determined in the housefly. *Mdtra* corresponds to the genetically identified *F* gene, the master ON/OFF switch in the pathway (Dübendorfer et al., 2002). Molecular analysis of this locus supports previous genetic evidence that female development relies on an autocatalytic activity of *F* to select and maintain the female fate (Dübendorfer and Hediger, 1998). *Mdtra* functionally corresponds to the *Drosophila Sxl* gene in that it not only executes but also memorizes the female choice through a self-sustaining feedback loop. In both cases, male development follows when the formation of this loop is prevented. In Musca, the loop breaker is likely to be the male dominant factor *M*. The use of dominant male determining factors is widespread in insects. However the molecular nature of these *M* factors has not yet been elucidated. We are currently attempting to isolate the *M* factor from the Musca genome using subtractive hybridization screens of mRNAs isolated from unisexual embryos. These approaches would greatly benefit from the availability of molecular tools, such as BAC libraries and EST libraries. Access to the complete gene content of Musca would also facilitate and accelerate the characterization and validation of candidate *M* genes. To improve our understanding of how *Mdtra* executes the female program we are also interested in identifying

Mdtra targets. With your proposal to sequence the Musca genome it would be feasible to take a functional genomic approach to this end. For instance, candidates can be identified in genome-wide Blast searches by virtue of a set of conserved cis-regulatory sequences and by using the microarray chip technology to produce genome-wide expression profiles for the detection of early sex-specific activities.

Daniel Bopp
PhD

Monika Hediger
PhD

Institute of Zoology, University of Zürich, Switzerland

January 4, 2005

Dr. Jeffrey Scott
Department of Entomology
Cornell University

Dear Dr. Scott:

It is with pleasure that I write this letter as indication of my full support for the request to NIH for funds to sequence the genome of the house fly, *Musca domestica*. The house fly can claim the fame to be the most ubiquitous and cosmopolitan of all the millions described and undescribed species of insects. But more significant yet, it is its role as the most important insect vector of enteric pathogens of humans and domesticated animals. The recent report of *E. coli* multiplying in house flies, and not just being carried by these flies, has done nothing but to reinforce the public health pest status of this fly. As the news media brings us constantly the horrors caused by the devastating effects of the tsunami in Southeast Asia, we just expect house flies to add to the misery of the surviving victims.

Confined livestock feeding operations offer a plethora of habitats for house flies in which to develop in great numbers. And as the urban areas keep alarmingly encroaching more and more on rural areas, the frequency of conflicts between urbanites and livestock producers over the ownership of these pesky flies will just only increase. These flies do not just affect humans by their annoying presence, but as carriers of enteric pathogens they disseminate these within the animal production confines, but also disperse them to wherever the wind and their wings will carry these flies.

For an insect species of such economic and human health importance, we are certainly ill equipped to manage, and least control, their populations. As an example of this poor situation, we do not have an adequate trap for monitoring their populations; in addition, the house fly has demonstrated to possess a formidable capacity to overcome almost every chemical toxicant we have thrown at them. For those of us that have dedicated our professional lives to study filth flies, with the ultimate goal of devising novel and effective control measures, sequencing the genome of undoubtedly the most significance member of this clan is certainly welcoming news as we see promising results of these efforts. These efforts should result in the opening of new promising avenues of attack to control their populations, or at least, the knowledge on how to manage the shrinking chemical arsenal we still have at hands.

Again, I express my full support for the funding of sequencing the genome of the house fly.

Sincerely,

Alberto B. Broce
Professor of Medical and Veterinary Entomology
Department of Entomology
Kansas state University
Manhattan, KS 66506

e-mail (while on leave at the USDA-CMAVE, Gainesville, FL:
abroce@oznet.ksu.edu

01.05.2005

Dear Dr. Jeffrey G. Scott

I and my partners of the "Hacettepe University Ecological Science Research Laboratories (ESRL)" have been working on the field of vector ecology and control for over 20 years. Our research topics span a wide array of fields like population biology of vector organisms, phenotypic variation, resistance development, genetics of resistance and establishment of vector management programs.

The housefly *Musca domestica* L. is a pest well known by many. Its association with household garbage and other organic disposals also makes it a potential vector for many diseases which can affect public health. Therefore efforts to control this organism have been and still are an important research subject. The housefly's natural ability to adapt to various habitats has made it a cosmopolite insect well distributed in most parts of the world. The housefly's high reproduction rate and its ability to adapt to various environments makes control efforts very difficult therefore information of its capacity to increase under different environmental conditions is of high importance.

In order to design an effective and persistent management program it is essential to follow both demographic and genetic changes taking place in target populations as a result of management programs. Resistance is the major problem encountered in nearly all management programs. In order to track the development of resistance and counter attack it is vital to have information exact genetic and enzymatic mechanisms of resistance and their phenotypic expressions. With information obtained from these studies, it will be possible to evaluate the behavior of resistance whether it will settle into a stable state or whether it will oscillate between certain levels. It will also be determine the various resistance mechanisms acting on the population and their relative frequencies.

A full knowledge of the housefly genome would be an invaluable asset to these studies. Full information of the genome will also be helpful in mapping QTL traits underlying the genetic structure of the phenotypic variation taking place in the population. Therefore it will be possible to fully understand and track the selective pressures and evolutionary changes taking place in the population. In addition such research conducted on the housefly can easily be used as a model for other vector organisms such as blowflies, mosquitoes or sandflies.

We give full support for this effort and believe that sequencing the genome of the housefly and other vector organisms is a very important goal and will be a major contribution to the scientific community.

On behalf of the ESRL

Assoc. Prof. Selim Sualp Çağlar
Hacettepe University Department of Biology
Ecological Science Research Laboratory,
06800 Beytepe Ankara/TURKEY
Tel: +90 312 297 80 63 Fax: +90 312 299 20 28
e-mail: sualp@hacettepe.edu.tr

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

Dr. John E. Casida, Director

Environmental Chemistry and Toxicology Laboratory
Department of Environmental Science, Policy and Management
College of Natural Resources, 102 Wellman Hall, Berkeley, California 94720-3112
Phone: (510) 642-5424
Fax: (510) 642-6497
E-mail: ectl@nature.berkeley.edu

January 4, 2005

Professor Jeffrey G. Scott
Department of Entomology
Cornell University
Comstock Hall
Ithaca, NY 14853-0901
Tel.: 607-255-7340
Fax: 607-255-0939
E-mail: jgs5@cornell.edu

Dear Professor Scott:

I strongly support placing high priority on sequencing the house fly, *Musca domestica*, genome. The housefly is the most extensively studied model for insect toxicology and biochemical mechanisms of control. Genetic manipulation of its genes provides insight into mechanisms of toxicity and resistance. Knowledge of its entire genome would identify insect-unique candidate targets. Knockouts of individual genes would provide new possibilities for insect control and opportunities for synthesis of target-specific chemicals affecting insects rather than mammals.

Selective toxicity involving low risk for mammals and high potency to insect pests is an essential requirement for safe and effective pesticides. About 90% of insecticides act on the insect nervous system including the voltage-dependent sodium ion channel, the chloride ion channel of the γ-aminobutyric acid receptor, and the cholinergic enzyme and receptor. The mechanisms for selective toxicity of insecticides are conferred by structural differences in the target sites and balance between metabolic activation and detoxification. The insect genome analysis, particularly in genes for nervous targets and xenobiotic-metabolizing enzymes, is therefore important in defining unique toxicology mechanisms. Also, the genome data will lead to recognition of novel insecticidal targets in insect pests.

Sincerely yours,

John E. Casida, Director
Environmental Chemistry and
Toxicology Laboratory
Professor of Entomology

UNIVERSITY OF
NOTRE DAME

DEPARTMENT OF BIOLOGICAL SCIENCES
107 GALVIN LIFE SCIENCE BUILDING

P.O. Box 369
Notre Dame, Indiana
46556-0369 USA

Telephone (574) 631-6552
Facsimile (574) 631-7413
Web site http://www.bio.nd.edu/biology

July 15, 2005

Jeffrey G. Scott
Sarkaria Professor of Insect Physiology and Toxicology
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

Phone: 607-255-7340
Email: jgs5@cornell.edu

Dear Jeff,

I have read the *Musca domestica* genome project white paper that you have submitted to NHGRI, and I want to let you and your potential reviewers know that I strongly endorse this project. I believe that the genome of this dipteran species will be a particularly useful outgroup – complementing the available *Drosophila* species – in helping us use comparative genomics to better understand the genomes of *Anopheles gambiae* and other mosquitoes. (A draft of the *Aedes aegypti* genome is almost complete and work on *Culex pipiens quinquefasciatus* is underway.)

The house fly is also a widely studied model organism, especially in the area of toxicology and insecticide resistance, and this and other areas of research will benefit significantly from a house fly genome project.

Finally, *Musca* is a very important vector of both human and domestic animal pathogens, and the very large community of scientists who study this species as a human and veterinary pest and vector will benefit enormously.

Again, I strongly endorse this project.

Sincerely,.

Frank H. Collins
Clark Professor of Biological Sciences

4 January 2005

Professor Jeffrey G. Scott
Department of Entomology
Comstock Hall
Cornell University
Ithaca
NY 14853
UNITED STATES OF AMERICA

Dear Jeff

I am delighted to add my support to your proposal to prioritise the housefly (*Musca domestica*) for genome sequencing. The draft of the proposal by yourself, Nannan Liu and Michael Kristensen does an admirable job in highlighting the many areas of research that would benefit from access to a full sequence. First and foremost, I would cite the extraordinary way that *M. domestica* has adapted to a wide range of environments, becoming commensal with man on every continent and one of the primary insect vectors of human diseases, in the developing world especially. As you rightly point out, the range of diseases transmitted has made this species a major target of control programmes worldwide, though often with relatively little success due to its very catholic environmental requirements, its high reproductive potential, and its proven capacity to respond to and withstand exposure to control agents. *M. domestica* remains one of the model organisms for research on mechanisms of insecticide resistance, and information gleaned from this work over the years has fuelled parallel studies on a wide range of medical, veterinary and agricultural pests. Indeed, for many years much of the information available on the genetics and biochemical nature of detoxification mechanisms and target-site modifications was derived directly from painstaking work on *M. domestica*. This exploited the availability of morphological markers for linkage analyses and for dissecting individual mechanisms from strains showing resistance to virtually all known classes of insecticide. Houseflies continue to lead the way in developing resistance to newer classes of chemistry and this, combined with the ease of laboratory maintenance and crossing, render them ideally suited for isolating and characterising the genes responsible. Work in this area is not just of value for improved understanding of resistance *per se*, but assists with identifying molecular mechanisms underpinning micro-evolutionary adaptations to a broad range of xenobiotics and other environmental challenges.

Resistance is just one area in which *M. domestica* has contributed, and can continue to contribute to our broad understanding of the native of adaptations and microevolutionary processes. However, I believe it has the potential to contribute even more substantially to broader, fundamental studies of attributes that enable an organism like this to become such a ubiquitous threat to the welfare of humans and domesticated livestock.

…continued

What enables a species like *M. domestica* to exert such dominance while others, with seemingly rather similar demographic and ecological characteristics, remain only of localised or specialised importance? Is it a question of exceptional phenotypic plasticity (which must in itself have a genetic explanation), or is it equipped with mechanisms that enhance mutability and hence the prospects of recovery of favourable genotypes? Work I was involved in some years ago demonstrated houseflies to exhibit dramatic polymorphism in mechanisms of sex determination, ranging from a 'classical' X-Y system with males being the heterogametic sex, to ones in which sex was established by the presence or absence of a dominant female determinant located on one or more of the five autosomes. The mechanics of what must be a highly complex transition from male to female heterogamety was left rather up in the air due to the (then) lack of incisive tools for taking the research further. It would be a fascinating topic to revisit as it seemed that this polymorphism was of recent origin, associated perhaps with areas where insecticide use was most prevalent, and where a number of independent resistance mutations were appearing and rapidly becoming commonplace. There is certainly very exciting scope for continuing such work armed with the tools and information that comprehensive data on genomic composition and structure would provide.

I wish you success with the proposal to NIH, and am happy to provide further information as required.

Yours sincerely

DR IAN DENHOLM
Head of the Plant and Invertebrate Ecology Division

DR MARTIN WILLIAMSON
Insect Molecular Biology Group, Biological Chemistry Division

STANFORD
UNIVERSITY
School of Medicine

Mark A. Krasnow, M.D., Ph.D.
Professor & Associate Chairman of Biochemistry
Investigator, Howard Hughes Medical Institute

July 11, 2005

Dr. Jeffrey G. Scott
Daljit S. and Elaine Sarkaria Professor of Insect Physiology and Toxicology
Department of Entomology, Comstock Hall
Cornell University
Ithaca, NY 14853

Dear Jeff,

I am writing in regard to the White Paper entitled "Rationale for sequencing the genome of the house fly, Musca domestica" that you are submitting to NIH. I do so in my capacity as the 2005-2006 President of the Drosophila Board of Directors. The Drosophila Board is an elected body of officers and regional representatives. One of the functions of the Board is to oversee community resources and to represent the best interests of the Drosophila research community. In carrying out these roles, the Board does not endorse specific grant applications or advocate for a specific group of researchers, but instead, it defines critical needs of the community-at-large and identifies possible solutions.

With extensive input from the Drosophila community, the Drosophila Board assembles and publishes the Drosophila Board White Paper. We are in the final stages of producing White Paper 2005, which includes a review of recent progress, identification of bottlenecks to more rapid research progress, and defining the most pressing needs of the Drosophila research community over the next several years. A draft version has been prepared and circulated for comments from the Board, the Drosophila research community, and officials at NIH. The final version should be available later this summer.

There are many reasons for embarking on the Musca domestica genome project, including the human health implications of controlling this disease vector, as well as the well-articulated arguments for its importance in genome annotation, evolutionary biology, and understanding the evolution of sex determination mechanisms. Further, a highly inbred line is available and many other reagents are in place, as well as a plan for hands-on annotation involving the community.

One of the twelve most pressing needs of the Drosophila research community identified in the current draft of White Paper 2005 is annotation of genome sequence from additional Drosophila species. The sequencing of 11 additional species of Drosophila is well underway and assemblies should be available soon. These new data present an unparalleled opportunity for rapid progress in a range of areas including (1) using comparative sequence analysis to improve the annotations of D. melanogaster, (2) understanding genome evolution including the functional evolution of genetic pathways, (3) describing variation at a genome scale, and (4) identifying non-coding genes and regulatory elements. To fully realize the potential of this unique resource, continuing support is needed for assembling, aligning and annotating these genomes. In addition to the biological and medical reasons for sequencing the Musca domestica genome, Musca is well situated between Drosophila and Anopheles to serve as an outgroup for the Drosophila genomes. It would also serve as a useful bridge for annotation of the Anopheles and other insect genomes.

Department of Biochemistry, B400 Beckman Center
Stanford University School of Medicine
Stanford, California 94305-5307
(650) 723-7191 • Fax (650) 723-6783 • krasnow@cmgm.stanford.edu

Although we did not receive the Musca White Paper in time to consider the project for inclusion and prioritization in our White Paper 2005, there is support for the project among the nine Board members that were available to review the Musca White Paper.

I would be happy to answer any question you might have about the Drosophila Board and its role in assessing the needs and priorities of the Drosophila research community.

On behalf of the Drosophila Board,

Mark Krasnow, M.D., Ph.D.
President, Drosophila Board of Directors

Michael B. Eisen, Ph.D.

Life Sciences Division
Lawrence Berkeley National Lab

Department of Molecular and Cell Biology
University of California Berkeley

January 2, 2005

Jeffrey G. Scott
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

Dear Jeff,

I am writing to strongly support your proposal to obtain a genome sequence for the housefly *Musca domestica*.

My lab studies the evolution of transcriptional regulation, with a focus on *Drosophila* and other Dipterans. While we have rich resources for the comparative and evolutionary studies within the genus *Drosophila*, many aspects of our work are limited by the absence of a genome sequence of an outgroup. The closest available species – the mosquito *Anopheles gamibiae* – diverged over 200 million years ago from *Drosophila*, and there is extensive sequence divergence between these two groups that makes it nearly impossible to compare the non-coding sequences of these groups. Furthermore, key aspects of development and other functions are radically different between *Drosophila* and *Anopheles*, further complicating comparative analyses. In particular, many of the key regulators of development found in *Drosophila* are absent or have markedly distinct functions in *Anopheles*.

In contrast, *Musca domestica* is far less diverged from Drosophila both at the sequence level and in terms of many functions. In particular, development in *Musca* proceeds very similarly to *Drosophila*, and the two taxa share an almost identical developmental toolkit. Our group, and many others, are poised to use this genome in our research. For these reasons we strongly support obtaining the *Musca* genome sequence.

Sincerely,

Michael B. Eisen

Faculty Scientist
Department of Genome Sciences
Life Sciences Division
Lawrence Orlando Berkeley National Lab

Assistant Professor of Genetics and Development
Center for Integrative Genomics
Department of Molecular and Cell Biology
University of California Berkeley

Dear Dr Scott,

among the ectoparasites affecting worldwide animals and humans, *Musca domestica* (Diptera: Muscidae), commonly known as “house fly”, represent a major concern for the human and animal health. This is mainly due to the fact that it is diffused and can live in several environments, the town as well as the countryside.

Musca domestica act as pathogen both at the larval (as causative agent of facoltative myiasis) and at the adult stage. At adult stage *M. domestica* may also cause nuisance to the livestock by flying or feeding on animal secretions or wounds. Meanwhile face flies are very active mechanic and/or biological vectors of pathogens. In particular *M. domestica* is mechanical vector of viruses (enterovirus, rotavirus), bacterial (salmonellae, shigellae, streptococci, stafilococci, *Escherichia coli),* protozoa *(Giardia*, amoeba, *Cryptosporidium*, etc.), nematodes (tricurids, ancylostomatids, ascarids). In fact more than 100 pathogens have been isolated from house flies and, among them, about 60 are transmitted by flies. *Musca domestica* is also biological vector of nematodes (e.g. *Thelazia, Parafilaria, Stephanofilaria, Habronema*) (Giangaspero, 1997).

Genomic studies on some *M. domestica* might be of interest to elucidate the direct and indirect pathogenic role this species plays in different geographical areas and on different hosts.

For example, the presence of different size or shape of prestomal teeth may be a character linked to the different pathogenic activity *M. domestica* have under the evolutionary pressure of different habitats and/or feeding habits. Similarly, genetic studies could be interesting to achieve insights on the indirect role of face flies as vectors of pathogen (see above). An example on this respect is represented by *M. domestica* acting as vector of nematodes belonging to the genus *Thelazia*.

Thelazia spp. commonly named “eyeworms”, comprises about 16 species of nematodes of animal and human concern. In fact *Thelazia* includes the more common species of spirurids infecting a wide range of domestic animals all over the world that are responsible for subclinical to clinical diseases, with variable symptoms (from mild conjunctivitis, epiphora and photophobia to keratitis and ulcers).

The life cycle of this nematode requires the presence of secretophagous flies (Diptera: Muscidae) as intermediate hosts. Thirteen species of *Musca* have been incriminated in the transmission of eyeworms, (*Musca autumnalis*, *Musca larvipara*, *Musca osiris* and *Musca domestica*) but mainly face flies (i.e. *M. autumnalis* and *M. larvipara*) have been demonstrated, both under experimental and natural conditions, to act as vectors in a few countries (Stoffolano, 1970; Otranto et al.,2003).

The role of *M. domestica* as vector of thelaziae is still controversial since, while in some countries (i.e. India and Cecoslovacchia) this species was demonstrated to be a good vector of eyeworms, experimental trials carried out in US demonstrated that it is not suitable to transmit bovine eyeworms. The latter evidences were based on epidemiological (only occasionally this species eat on the conjunctive) and biological evidences (*M. domestica* larvae develop in the haematocele in an anomalous manner) (Broce and Elzinga, 1984; Geden and Stoffolano, 1981; 1982).

The evidence that the house fly may act as intermediate host of *Thelazia gulosa* and *Thelazia rhodesi* –which infect cattle- (Vilagiova, 1967; Gupta, 1970) indicates that where *M. domestica* and eyeworms have shared a common habitat for long periods of time, selection could have favoured a successful vector-nematode relationship (Geden & Stoffolano, 1981).

A recent molecular epidemiological survey demonstrated that *M. domestica* may act as vector of *T. gulosa* in southern Italy (Otranto et al., 2003).

By the means of genomic studies on *M. domestica* the role of this species in the aetiology, epidemiology and pathogenesis of thelaziosis in different geographical areas might be of interest.

Genome projects for several medically important parasites have been initiated and are in progress. We are firmly convinced that the NIH sequencing project of house fly genome might add important information on a topic which is often little considered, as the epidemiology of animal and human parasitic diseases.

A.Giangaspero & D. Otranto

Dipartimento PR.I.M.E.
Via Napoli, 25
71100 Foggia, Italia
tel.+39.881.589227
fax.+39.881.740211
e-mail:a.giangaspero@unifg.it

Domenico Otranto
DVM, PhD, Dip. E.V.P.C.
Associate Professor of Animal Parasitic Diseases
Dipartimento di Sanità e Benessere Animale.
Str. prov. per Casamassima km 3
70010 Valenzano (Bari)
ITALY
tel/fax +39 080 4679839
e.mail: d.otranto@veterinaria.uniba.it

General References

Branch G J, Stoffolano JG (1974). Face fly: invertebrate vector and host of a mammalian eyeworm in Massachussetts. J Econ Entomol 67: 304-305.

Broce AB, Elzinga RJ (1984). Comparison of prestomal teeth in the face fly *Musca autumnalis* and the house fly Musca domestica (Diptera: Muscidae). J. Med. Entomol. 21: 82-85.

Chirico J (1994a). A comparison of sampling methods with respect to cattle-visiting Muscidae and their nematode infections. Med Vet Entomol 8: 214-218.

D'Esposito L (1949). Di alcuni casi di infezione del sacco congiuntivale dei bovini da *Thelazia*. Clinica Veterinaria 72: 108-118.

Garippa G, Polinas I, Suelzu P, Merella P, Sini T, Scala A (2002). Epidemiology of bovine thelaziosis in Northeast Sardinia. Parassitologia 44 (Suppl.1): 79.

Geden CJ, Stoffolano JG (1981). Geographic range and temporal patterns of parasitization of *Musca autumnalis* De Geer by *Thelazia* spp. in Massachussets with observation on *Musca domestica* as an unsuitable host. J Med Entomol 18: 449-456.

Geden CJ, Stoffolano JG (1982). Development of the bovine eyeworm, *Thelazia gulosa* Railliet and Henry 1910 in experimentally infected female *Musca autumnalis* De Geer. J Parasitol 68: 287-292.

Giangaspero A (1997). Le mosche di interesse veterinario. I Muscidae. Guida alla conoscenza e al riconoscimento. Edagricole, Bologna, Italy.

Giangaspero A, Broce AB (1993). Micromorphology of the prestomal teeth and feeding behaviour of *Musca autumnalis*, *M. larvipara* and *M. osiris* (Diptera: Muscidae). Med Vet Entomol 7: 398-400.

Giangaspero A, Otranto D, Vovlas N, Puccini V (2000). *Thelazia gulosa* Railliet & Henry, 1910 and *T. skrjabini* Erschow, 1928 infection in southern Europe (Italy). Parasite 7, 327-329.

Krafsur ES, Church CJ (1985). Bovine thelaziasis in Iowa. J Parasitol 71: 279-286.

Krastin NI (1950). A study of the developmental cycle of the nematode *Thelazia gulosa* (Railliet & Henry, 1910) a parasite of the eyes of cattle. Doklady Akademy Nauk SSSR 70: 549-551.

Lyons ET, Drudge JH (1975). Two eyeworms, *Thelazia gulosa* and *Thelazia skrjabini* in cattle in Kentucky. J Parasitol 61: 1119-1122.

Moonlenbeek WS, Surgeoner GA (1980). Southern Ontario survey of eyeworms *Thelazia gulosa* and *Thelazia lacrymalis* in cattle and larvae of *Thelazia* spp. in the face fly *Musca autumnalis*. Can Vet J 21: 50-52.

O'Hara JE, Kennedy MJ (1991). Development of the nematode *Thelazia skrjabini* (Nematoda: Thelazioidea) in experimentally infected face flies *Musca autumnalis* (Diptera: Muscidae). J Parasitol 77: 417-425.

Otranto D, Tarsitano E, Traversa D, Giangaspero A., De Luca F, Puccini V (2001). Differentiation among three species of bovine *Thelazia* (Nematoda: Thelaziidae) by PCR- RFLP of the ITS-1 (rDNA). Int J Parasitol 31: 1693-1698.

Otranto D, Ferroglio E, Lia R, Traversa D, Rossi L (2003*a*). Current status and epidemiological observations of *Thelazia callipaeda* (Spirurida, Thelaziidae) in dogs, cats and foxes in Italy: a "coincidence" or a parasitic disease of the Old Continent? Vet Parasitol 116: 315-325.

Otranto D, Tarsitano E, Traversa D, De Luca F, Giangaspero, A. (2003*b*). Molecular epidemiological survey on the vectors of *Thelazia gulosa, Thelazia rhodesi* and *Thelazia skrjabini* (Spirurida: Thelaziidae). Parasitology 127: 365-373.

Otranto D., Traversa D. 2004. Molecular charachterization of the first internal transcribed spacer (ITS1) of ribosomal DNA (rDNA) of the most common species of eyeworms (Thelazoidea: Thelazia). J. Parasitol. 90: 185-188.

Puccini V., Giangaspero A., Bisceglia D. 1986. Telaziosi dei bovini del Gargano. Atti SISVet 40: 1001-1002.

Puccini V, Giangaspero A, Bisceglia D (1988). La telaziosi dei bovini. OD&V 9: 17-19.

Vilagiova I (1967). Results of experimental studies on the development of pre-invasive stage of worms of the genus *Thelazia* Bosc 1819 (Nematoda: Spirurata) parasitic in the eye in the cattle. Folia Parasitologica 74, 275-280.

Vilagiova I. (1968). Ecological specificity of intermediate hosts of Thelazia in cattle and the relation of their infective larvae to some unusual hosts. Helminthologia 9, 563-566.

Traversa D, Giangaspero A, Otranto D (2004). Characterization of the ribosomal ITS1 and ITS2 (rDNA) of *Thelazia lacrymalis* (Spirurida, Thelaziidae): genetic markers for its identification in vectors. XXIII Congresso Nazionale SoIPA, Vietri sul mare, 9-12 giugno 2004.

Imperial College London

Division of Cell and Molecular Biology
Imperial College London

Room 612, Sir Alexander Fleming Building
South Kensington Campus
SW7 2AZ
Tel: +44 (0)20 7594 1267 Fax: +44 (0) 20 7594 1759

f.kafatos@imperial.ac.uk
www.imperial.ac.uk

Fotis C. Kafatos, PhD FRS
Professor
Chair of Insect Immunogenomics

22 November 2005

To Jeffrey G. Scott

Dear Jeff,

I enjoyed visiting you recently, discussing the prospects of *Musca* genome sequencing and reading your white paper and follow-up arguments.

I am absolutely convinced that sequencing *Musca* is very sensible at this stage. Its control really is important to human wellbeing; its sequence will be extremely valuable as an intermediate outgroup, for improving annotation of both the mosquito genomes and the drosophilid genomes; the new estimate of genome size and the available inbred lines make sequencing straightforward; and its biology promises significant enrichment of our understanding of key processes such as insect immunity, sex determination, olfactory physiology and insecticide resistance mechanisms.

The size of this fly, ease of laboratory maintenance, genetic tractability and availability of transgenesis, will permit rapid growth of the *Musca* molecular genetics community, complementing the existing strength of the toxicology and physiology communities. I would dearly love to have the *Musca* genome available as we work to improve the annotation of vector genomes (currently Anopheles and soon Aedes, followed by other vectors in the pipeline). As our community expands further, we will need to lean on reasonably close relatives, to increasingly understand the thus far neglected genome compartments (such as the rapidly evolving gene families). The biological tractability of the house-fly coupled with its special features will be invaluable.

Best wishes,

Fotis C. Kafatos

CORNELL

Department of Entomology Fax: 607 255-0939
Cornell University
Comstock Hall
Ithaca, NY 14853-2601

College of Agriculture and Life Sciences

January 4, 2005

Dear Jeff,

I am writing to express strong support for your proposal to have the genome of *Musca domestica* sequenced. I see a number of reasons why housefly should be sequenced, including the practicality of "doing genetics" in the system and the phylogenetic placement of the housefly. As you know, my particular research interest is in the population genetics and evolutionary biology of insect immune systems. The ability to conduct research in this area was greatly expanded by the complete sequencing of the *Drosophila melanogaster* and *Anopheles gambiae* genomes. I envision a similar explosion in research productivity with the data and technological resources that would become available with a *Musca* sequence.

Because of the housefly's evolutionary intermediacy between *Drosophila* and *Anopheles*, availability of a *Musca* genome sequence would have immediate impact on several research questions related to my interest. The high divergence between *Drosophila* and *Anopheles* impairs comparison of these species for inference of evolutionary pressures on many proteins. This problem is compounded by the fact that the most interesting proteins are frequently those most quickly evolving. The housefly, being more closely related to *Drosophila melanogaster* than *Anopheles*, but less closely related than the recently sequenced *Drosophila pseuodobscura*, is a useful outgroup for evolutionary genomic comparisons. Furthermore, the availability of genome sequences at multiple phylogenetic distances confers optimal power for testing long-term evolutionary hypotheses. For one example, when the complement of immune response genes in the *Anopheles* genome were compared to those of *Drosophila*, it was found that the mosquito genome had much larger gene families related to suppression of eukaryotic parasites (presumably due to pressure from malaria) while the fruit fly had many more genes relevant for suppression of microbial infection (presumably reflecting the septic environment in which it lives). Because the environmental pathogens afflicting *Musca* differ from those affecting either *Anopheles* or *Drosophila*, further comparison with the housefly will extend our ability to address the degree to which gene complement is determined by the ecology of the organism.

I am quite enthusiastic about the prospect of a *Musca* genome sequence both for my own research and for the advancement of insect evolutionary and functional genetics. Please do not hesitate to call on me if I can help advance your effort.

Sincerely,

Brian Lazzaro
Assistant Professor
Department of Entomology
Cornell University

UNIVERSITY OF CALIFORNIA, RIVERSIDE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

COLLEGE OF NATURAL AND
AGRICULTURAL SCIENCES
DEPARTMENT OF ENTOMOLOGY - 041
FAX: (909) 787-3086

RIVERSIDE, CALIFORNIA 92521-0314

Dr. Jeffrey G. Scott
Daljit S. and Elaine Sarkaria Professor of Insect Physiology and Toxicology
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

7 Jan. 2005

Dear Jeff,

I am pleased to be able to write in support of your NIH proposal to sequence the genome of the house fly, *Musca domestica*. This species is a superior laboratory model for studying many aspects of basic biological processes. Therefore the genome would no doubt be of immense utility for developmental or evolutionary biologists, geneticists and neuroscientists.

My own area is biology and management of pests affecting human and animal health. In this regard, in my opinion, **the house fly truly deserves to be ranked as a high priority target**. It absolutely should be up there with the primary obligate vectors of biologically-transmitted human disease agents. Genome description already is either complete or underway for several significant vectors- the house fly deserves no less attention.

The house fly is one of mankind's most abundant, widespread, and unwelcome associates. Unlike a species like *Anopheles gambiae* (a principal African malaria vector and one of the first targets of genome description), house flies now exist almost everywhere humans do. Most of the pathogens they transmit also can get around by other means, so it has been easy to underestimate, or at least not fully appreciate, the actual role of flies in disease agent transmission.

Especially over the past decade, however, it has become steadily clearer just how important house flies are in linking contaminated substrates with humans or domestic animals. Enteric (and rather frequently antibiotic-resistant) pathogens such as *Salmonella* or *Campylobacter* cause immense misery, or sometimes death, around the world. Sequencing the house fly genome could have tremendous applications for novel control efforts, from development of new toxicants to manipulating behavior. I enthusiastically support your efforts.

Sincerely,

Bradley A. Mullens

Bradley A. Mullens
Professor

January 4, 2005

Letter of Support for Housefly Genome Project

Neurotransmitter receptors, including ligand-gated ion channels and G protein-coupled receptors, are targets for medicines and insecticides. I have been working on neurotransmitter receptors as sites for insecticidal action, using the housefly, *Musca domestica* L. The advantage of using the housefly in such studies includes its short life cycle, the large size of the body (compared to *Drosophila*), and easiness of rearing. The housefly has served as an excellent organism to study the biochemistry and physiology of hygienic insect pests. Studies with the housefly have led to an accumulation of data regarding insect neurochemistry and toxicology.

Although the low density of receptors in insect tissue hampered detailed analysis of the mode of action of ligands on isolated receptors, recent molecular biological techniques have made it possible to use heterologously-expressed receptors in such studies. For example, the γ-aminobutyric acid (GABA)-gated channel, a target of the phenylpyrazole insecticide, was considered to be an only chloride channel in the insect nervous system several years ago, but glutamate has recently been shown to open a distinct chloride channel. It is necessary to promptly examine whether insecticides acting at the GABA receptor act on the glutamate receptor as well. We cloned genes that encode subunits of GABA- and glutamate-gated chloride channels from the housefly (Accession No.: AB177547 and AB177546), and are investigating molecular and pharmacological characteristics of each channel. Sequencing the genome will provide information to isolate other subunit(s) or protein(s) that interact(s) with the above subunits to form functional, native GABA and glutamate receptors. I strongly support the proposal for sequencing housefly genome, which will facilitate the molecular toxicological studies of insect pests.

Yoshihisa Ozoe
Professor of Bioorganic Chemistry
Department of Life Science and Biotechnology
Shimane University
Matsue, Shimane 690-8504, Japan

January 21, 2005

Dr. Jeffery G. Scott
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

Dear Dr. Scott,

This letter is in support of your request to NIH to sequence the genome of the house fly, *Musca domestica*. The house fly has long history in the scientific community as a tool for the study of a host of basic and applied problems including population genetics and target site investigations that are central to the understanding and development of new insect control technologies and molecules. Equally important, the house fly is itself an insect pest of world wide importance.

Because of it's pest status, a dizzying array of insect control agents have been used against the house fly over the past 60 years leading to numerous cases of resistance. According to the Michigan State University Resistance Database, the house fly ranks in the top 10 most resistant insect pests, having developed resistance to more than 35 different insecticides. Because it has developed resistance to such a wide spectrum of insect control agents, the house fly has been the subject of a wide array of studies address to behavioral, biological, biochemical and molecular basis of resistance. Likewise, the numerous resistant strains provide a doorway to understand the population genetics of resistance and resistance management. In addition, these same strains also provide a tool to investigate and identify new targets and approaches for the control of pest insects. The availability of the genome sequence for the house fly would be of enormous value to the scientific community as well as industry by providing the means more fully study and understand this important insect pest.

I applaud the efforts of you and your colleagues to induce NIH to sequence the house fly genome.

Best regards,

Thomas C. Sparks
Advisor
Discovery Research, Insect Management Biology
Dow AgroSciences
9330 Zionsville Road
Indianapolis, IN 46268

Support letter for sequencing the genome of the housefly, *Musca domestica*

Takashi Tomita, Shinji Kasai, Osamu Komagata and Toshio Shono
Department of Medical Entomology, National Institute of Infectious Diseases,
1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, JAPAN

The housefly is recognized as the significant medically important insect worldwide. Houseflies breed in dumping sites, livestock sheds and organic fertilizer in greenhouses. The housefly is not only an unpleasant pest for humans and livestock, but also the vector of emerging and re-emerging diseases. A total case number of food poisoning due to enterohemorrhagic *E. coli* O-157 shows no signs of decreasing in last several years in Japan (1445 patients were confirmed diagnosis by O157:H7 serotypes in 2001) and one of the factors for this is failure is lack of appropriate control of houseflies. Furthermore, flies (including houseflies) have recently been strongly implicated to be the transmitters of bird-influenza among poultry farms in Japan. Accordingly, control of houseflies is becoming more and more important in this era to secure human health from such emerging diseases.

Historically houseflies have been the targets of chemical control, and that resulted in the development of high levels of resistance to most of the insecticide families including organochlorines, organophosphates, pyrethroids and insect growth regulators (such as chitin synthesis inhibitors and juvenile hormone mimics). Therefore, the natural population of the housefly is like a department store of insecticide resistance, and thus it is one of the most difficult species for us to control. We believe that the genome-sequencing project of the housefly will bring us vast amounts of valuable information to overcome a number of problems regarding to this insect and other important pest insects. Furthermore, there are some unique advantages for the housefly genome project compared to the projects conducted for other insect species. In this letter, we would like to support the housefly genome project from the following three aspects.

(I) Elucidation of the mechanisms of insecticide resistance
As mentioned above, a number of housefly local population have been already confirmed to develop insecticide resistance in livestock barns and that causes fear of spreading pathogens especially in Asian country where livestock farms are situated close to

residential areas. There are two major mechanisms in insecticide resistance; insensitivity of insecticidal targets, such as acetylcholinesterase and sodium channels, and the increased activity of insecticide-detoxifying enzymes, such as cytochrome P450s, carboxylesterases and glutathione *S*-transferases. Housefly is the representative and ideal model insect in the research field of insecticide resistance. Abundant studies of the mechanisms of insecticide resistance have been conducted using housefly, compared to *Drosophila melanogaster*, *Bombix mori* and *Anopheles gambiae*, although the function of each enzyme is still not yet fully studied. One of the highest hurdles for identification of resistance genes involves the P450s and glutathione *S*-transferases, because they are large multigene families. For example, *Drosophila* and *Anopheles* genome project clarified that they have 91 and 100 P450 isoforms, respectively. Each isoform has very similar protein structure and molecular mass so that it had been very difficult to isolate each isoform from others for the function analysis. Since most P450s have high substrate specificities, multiple numbers of P450s theoretically have roles in insecticide metabolism. However, CYP6D1 identified in Scott's laboratory is the only P450 that has been proven to be a significant factor in pyrethroid-resistance. Sequencing the whole genome of housefly will be conductive in order to identify unknown P450s involved in resistance. A goal of the research will be application of these accomplishments to molecular monitoring of insecticide-resistance that will help prompt selection of efficacious compounds in the agricultural and medical fields.

(ii) Exploration of novel insecticidal targets

Some cytochrome P450s are well studied as target sites for the development of medicines (anti-tuberculosis), fungicides, and herbicides. Thus, the P450s of insects also have potential for use as insecticidal targets. However, due to multiplicity of isoforms, the study of insect P450s has not made progress in this area. P450s metabolize endogenous hormones (juvenile and ecdysteroid) and pheromones as well as exogenous chemicals such as pesticides and environmental pollutants. Four P450 genes involved in ecdysteroid synthetic pathway have recently identified. Since this pathway is restricted to invertebrates, but essential for them to maintain their lives, P450s participating in this pathway will be good candidates as target of new insect regulators. The *Drosophila* genome project contributed (in part) to the discovery of these molecules. Using information from the Drosophila genome our group conducted transcriptional analysis of *Drosophila* P450s. We spotted all *Drosophila* P450 genes on slide glass and then the expression of all P450 genes were comprehensively analyzed by microarray method.

Several sex-specific unique isoforms were discovered (Kasai and Tomita, 2003, *Biochem. Biophys. Res. Commun*., **300**, 894-900) and have potential as candidates for the target of new insecticides. We could have never identified these proteins without the information obtained from the genome project. Because the P450s of each insect species are unique, our results from Drosophila cannot be readily utilized in other insects. Therefore, we are interested in using similar approaches in housefly, in order to address the important health issues of this species (mentioned above).

(III) Elucidation of epistatic sex-determination system

Epistatic sex-determination is a major system in insects, however, the substance of sex-determinant, the genes involved in downstream gene cascade, and their interactions are not known. The housefly provides multiply located, but functionally equivalent male determinants that helps to look for a great genetic switch. A model insect, *D. melanogaster*, takes rather a unique sex-determining system: balance of autosome and sex chromosome doses. Mosquitoes, and most fly species (including the housefly), as well as lepidopteran species have epistatic sex-determining system. In the housefly, the presence or absence of a male-determining gene (M) in the genome will destine male or female. Y chromosome is basically a carrier of M in the housefly. However, the existence of M in each autosomal linkage and furthermore a mutant female-determining factor (epistatic to M; on chromosome 4) are segregating in the natural population. These sex-determinants are already genetically mapped in the respective linkage groups. Identification of M, the wild-type gene of F, and how they regulate downstream cascades, including absence of male crossing-over are quite interesting to be elucidated.

By the reasons mentioned above, we strongly support the proposal to sequence the housefly genome.

U N I V E R S I T Y O F
ROCHESTER
THE COLLEGE
ARTS • SCIENCES • ENGINEERING

COLLEGE OF ARTS AND SCIENCE
DEPARTMENT OF BIOLOGY

Jeffrey G. Scott
Daljit S. and Elaine Sarkaria Professor of Insect Physiology and Toxicology
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

12 December 2005

Dear Jeff,

This letter is written to strongly endorse full genome sequencing for the housefly. *Musca domestica* is an obvious choice. It is well positioned phylogenetically, providing an important intermediary between the drosophilid flies and more basal mosquitoes. It is a well-studied tractable genetic organism, which means that the genome can be exploited to investigate the functional basis of different genes. And finally, *Musca* has very interesting biology. Among the obvious features of interest is its sex determination. The sex determining system of the housefly is quite different from *Drosophila*, as involves dominant sex determining factors rather than X:Autosomal balance sex determination. Considerable progress has been made in unraveling the basis of sex determination in this system, and the genome sequence will greatly accelerate these studies.

A housefly genome will also present the unique opportunity to study the interactions between a pest insect and its parasitoid(s) at the whole genome level. Parasitoid wasps lay their eggs in or on other insects, and the developing parasitoids eventually kill the host. This process involves a number of interactions, which can range from suppression of the host immune system, to manipulation of host physiology and development by parasitoid venoms. Parasitoids have a rich pharmacopia of venoms that have yet to be tapped for medically relevant compounds.

The parasitioid wasp *Nasonia vitripennis* has been selected for full genome sequencing, and the project is already underway (over 400,000 sequence reads by last week). *Nasonia* and its relatives are parasitoids of houseflies. Once the genome of the housefly is underway, the information can be used for constructions of microarrays for analysis of the effects of parasitoid venoms of host gene expression and the effects of the host on parasitoid gene expression during development. Similarly, studies can also be used to enhance the effectiveness of parasitoids for biological control of houseflies and other disease transmitting fly relatives. This would be the only system where the genome sequence of a host and its parasitoid are available.

If I can provide any additional information to support your effort, please let me know.

Sincerely

John H. Werren
Professor of Biology
Biology Department
University of Rochester

North Carolina State University is a land-grant university and a constituent institution of The University of North Carolina

Department of Entomology

NC STATE UNIVERSITY

Brian M. Wiegmann
Entomology

College of Agriculture and Life Sciences
Gardner Hall
Campus Box 7613
Raleigh, NC USA 27695-7613

919.515.2703 (phone)
919.515-7746 (fax)

3 January 2005

Jeffrey G. Scott
Daljit S. and Elaine Sarkaria Professor of Insect Physiology and Toxicology
Department of Entomology
Comstock Hall
Cornell University
Ithaca, NY 14853

Dear Dr. Scott,

I am writing to express my enthusiasm and support for the proposed *Musca domestica* genome project. *Musca* is a key model organism critical to diverse fields of study, including genetics, physiology, toxicology, neurobiology, and medical/veterinary entomology. Obtaining a full genome sequence for this fly will be critically important for framing detailed evolutionary comparisons both among and within genomes. *Musca* provides a thoroughly studied, phenotypically well-characterized, fly species useful for detailed experimental and descriptive comparisons with the completed *Drosophila* and mosquito genomes. This project will have an immediate impact on my own phylogenetic research program and on a broad array of research programs that use flies as a model system.

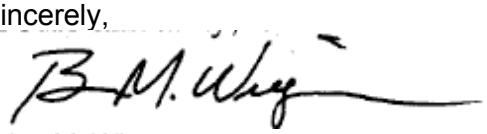
Musca is a member of the cyclorrhaphan fly lineage -- Calyptratae. This is highly diverse fly group that shares a common ancestor with *Drosophila* (30-80 mya) and *Anopheles* (230-250mya) spanning a long history of fly diversification and genetic evolution. This closer, but still quite old relationship with *Drosophila* provides an important intermediate-aged comparison allowing finer scale establishment of orthology relationships, tracking of genome rearrangements, and calibration of evolutionary rate assessments.

A fully detailed genome for *Musca* will be extremely useful in providing genetic markers useful within our NSF-funded, Assembling the Tree of Life (ATOL) project on Diptera phylogeny, called FLYTREE (http://www.inhs.uiuc.edu/cee/FLYTREE/). A major component of this large collaborative project involves mining the established, and newly emerging, fly genomes for genes that are informative of fly evolutionary relationships. Within our project, we are also currently completing full sequencing and annotation of the *Musca domestica* mitochondrial genome - this will be compared to other mitochondrial genomes for dipteran model organisms.

I would be pleased to help in any way with the Musca project, especially in providing evolutionary phylogenetic insights in the context of comparative genomic analysis as the results of our own phylogenetic projects emerge.

I look forward to this important genomic initiative for *Musca*, and please feel free to contact me for further information.

Sincerely,



Brian M. Wiegmann
Associate Professor, Department of Entomology

Jeffrey G. Scott July 6, 2005
Department of Entomology
Cornell University
Ithaca, NY 14853

Dear Dr. Scott,

I would like to express my strong support for your proposal to sequence the house fly genome. The house fly is undoubtedly one of the most successful insect species. Sequencing the house fly genome is interesting and important from many perspectives, including biology, phylogeny, insecticide resistance, microbial ecology and public health. The last two perspectives are of great interest of mine.

Due to the house fly developmental habitats (decaying organic materials - primarily animal manure with numerous and diverse microbial communities), unrestricted movement, attraction to the urban environment, and mode of feeding (regurgitation), house flies greatly amplify the risk of human exposure to pathogens and antibiotic resistant strains. Considering the heavy use of antibiotics in the U.S. livestock industry and convergence of agricultural and urban/sub-urban environments over the past two decades, the public health importance of house flies has increased tremendously. Research results from my and other labs. show that this insect plays an important role in the ecology of antibiotic resistance strains/resistance genes and to some degree in dissemination of human pathogens, such as *Escherichia coli* O157:H7 and *Campylobacter jejuni*.

Moreover, it has become clear that house fly biology is closely linked to microbes. Development of house fly larvae is strictly dependant on a live and active microbial community in a natural developmental habitat. Larvae cannot develop beyond the first instar in sterilized a natural or artificial substrate/medium. The principle of this symbiosis is unknown although it has been shown that different bacteria support the house fly development to different degrees.

Sequencing of the house fly genome would offer new tools in the research on house fly microbial ecology. For example, the spotted microarray approach would allow to investigate how different bacterial strains effect expression of different house fly genes and result in better understanding of the house fly-bacterial symbiosis that can lead to new approaches in house fly management based on transgenesis and paratransegensis.

Again, I strongly support the proposal to sequence the house fly genome.

Best regards,

Ludek Zurek
Assistant Professor
Department of Entomology
Kansas State University
Manhattan, KS 66506