# *Culex pipiens quinquefaciatus* Genome Project

**Contacts:** **Peter W. Atkinson <peter.atkinson@ucr.edu>, 909-787-4782**

**Frank H. Collins <frank@nd.edu>, 574-631-9245**

## 1. Justification for a *Culex pipiens quinquefasciatus* Genome Project.

*Culex* species are important vectors of human pathogens in the United States and world-wide, including the aetiologic agents of West Nile encephalitis, Eastern equine encephalitis, Venezuelan equine encephalitis, Japanese encephalitis, St. Louis encephalitis, Ross River encephalitis, Murray Valley encephalitis, Rift valley fever, and lymphatic filariases. The most important of the *Culex* vectors are members of the *Culex pipiens* complex, a very closely related group of species (or incipient species - the taxonomy remains unclear) that originated in Africa but has spread by human activity to tropical and temperate climate zones on all continents but Antarctica. More than 100 million people are infected worldwide with the *Wuchereria bancrofti* form of lymphatic filariasis that is transmitted primarily by *Cx. pipiens* complex mosquitoes in urban and suburban settings, and approximately 43 million filariasis cases are seriously disabled (UNDP/World Bank/WHO/TDR and WHO/UNICEF "Research on Rapid Geographical Assessment of Bancroftian Filariasis" July, 1997). It has been estimated that 750 million people are exposed to filariasis every year (WHO Expert Committee on Filariasis, 1992).

Recently, both *Cx. pipiens pipiens* (Northern house mosquito) and *Cx. pipiens quinquefasciatus* (Southern house mosquito), together with *Cx. tarsalis* and several other *Culex* species, have attracted much attention within the United States and Canada as significant vectors of West Nile Virus. When one considers the number of human pathogenic viruses that *Culex* species transmit combined with the numbers of people afflicted, in some form, with lymphatic filariasis, the overall disease burden created by members of the genus *Culex* is very large. This burden caused by *Culex* species exceeds those of dengue and yellow fever transmitted by *Aedes aegypti* (estimated at ~ 20 million) and the number of people who are HIV-positive ( ~ 42 million). The combination of both the global distribution of the *Cx. pipiens* complex and its distribution within the United States as both a rural and urban mosquito species, also makes this mosquito a potential vector for mosquito-borne pathogens released for the purposes of bioterrorism.

The genome sequence of a member of the *Cx. pipiens* complex would greatly simplify the identification of mosquito genes required for pathogen transmission, potentially enabling the development of new strategies for combating and controlling these diseases. Furthermore, initiation of a *Cx. pipiens* complex species genome project would complement ongoing work with *Anopheles gambiae* (draft genome completed) and *Aedes aegypti* (extensive EST and BAC end sequencing underway) and lead to the completion of draft genomes of the three most important mosquito vectors of human pathogens. Comprehensive comparisons between representatives of these three mosquito genera will also greatly inform the evolutionary relationships among these species and perhaps lead to advances in our understanding of mosquito genes involved in important phenomena like vectorial capacity and insecticide resistance. The genome of a *Cx. pipiens* complex species will also be invaluable in helping illuminate mechanism whereby the endosymbiotic organism *Wolbachia* (first discovered in *Cx. pipiens*) is responsible for phenomena like cytoplasmic incompatibility in mosquitoes and other insects.

## 2. Relationship of Proposed *Culex* Project with Other Culex Research .

The work outlined in this proposal does not duplicate any ongoing or planned *Culex* research. However, this project does build on and complement *Culex* research currently underway in laboratories of many of this project's participants. The specific projects and their relationship with this proposal are as follows:

(a) Dr. A.J. Cornel of the University of California, Davis is supported by NIAID in a project that has the goal of establishing polytene chromosome maps for members of the *Cx. pipiens* species complex. One aspect of his work involves the development of a *Culex* polytene chromosome physical mapping procedure and use of this procedure to map sequence tagged sites to the polytene chromosome complement. Dr. Cornel has been working with other participants in this project to identify a *Cx. pipiens* complex colony with good quality polytene chromosomes that would be used in this genome project. Furthermore, it is expected that physical mapping of sequence tagged clones like BAC genomic clones or cDNAs will eventually prove important in assigning the assembled scaffolds from a full genome sequencing project to ordered and oriented locations along the mosquito chromosomes.

(b) Professor J. Hemingway, Director of the Liverpool School of Tropical Medicine, is funded by the Welcome Trust to develop a normalized cDNA library from mixed *Cx. pipiens* tissues and sequence clones from this library in the first phase of a gene discovery project that will also provide reagents that will contribute to a first generation *Cx. pipiens* microarray. Professor Hemingway has agreed to reallocate the funds originally designated for cDNA sequencing to the production of additional normalized *Culex* libraries that will then be sequenced in this Microbial Sequencing Center project (see section **7. Goals and Rationale for Proposed Culex Project**).

(c) Professors B.M. Christensen of the University of Wisconsin, Madison and S. Higgs of the University of Texas Medical Branch at Galveston are both funded to work on *Cx. pipiens* complex mosquitoes. Professor Christensen's laboratory is currently investigating the response of *Cx. pipiens* to bacterial challenge with an EST project, and Professor Higgs routinely infects *Cx. pipiens* colonies with West Nile virus and occasionally with other flaviviruses. Professor Christensen, in a project funded by the Department of State for technical exchange in science between the United States and Egypt, has been working with Dr. Reda Ramzy of Ain Shams University, Cairo. They have produced a limited EST dataset derived from two cDNA libraries, one constructed from midgut tissues and one from bacteria-inoculated whole bodies of *Cx. p. pipiens*. This project, to date, has produced 1,204 midgut sequences that were compiled into 386 contigs, and 1169 sequences from the immune-activated library produced 585 contigs (Bartholomay *et al*. 2003, Mol. Biochem. Parasitol. 130:43-50). Professors Christensen and Higgs will provide purified mRNA from bacterial challenged and WNV infected *Cx. pipiens* mosquitoes to the laboratory of J. Hemingway, where this RNA will be used to produce normalized cDNA libraries (see **7** below).

(d) Professor F.H. Collins and J. Romero-Severson of the University of Notre Dame are funded by the State of Indiana through the Indiana Center for Insect Genomics (F.H. Collins is Director, J. Romero-Severson is Assistant Director) to produce genomic reagents for a wide variety of insects. Professors Collins and Severson have been working with Dr. J. Tomkins of the Clemson University Genomics Institute on the production of a number of insect genomic DNA BAC libraries. Through this collaboration between Notre Dame and Clemson and with Indiana Center funds, a 10X coverage BAC genomic DNA library will be produced for the *Culex p. quinquefasciatus* JHB colony (see **8** below).

(e) Professor F.H. Collins is the Principal Investigator of a pending NIAID contract proposal to develop an integrated relational bioinformatics resource center for vectors of human pathogens. [This proposal has undergone the

preliminary scientific and business review and has been judged in the competitive range.] This center, provisionally called ***VectorBase***, will pull together resources from a number of existing bioinformatics projects and institutions, including the European Bioinformatics Institute/EMBL at Hinxton, UK, the bioinformatics resources at EMBL, Heidelberg, Germany, the FlyBase project at Harvard, and several others. If the ***VectorBase*** contract is funded, it will have the resources for managing, analyzing, and displaying all genomic data that emerge from the *Culex* genome project. ***VectorBase*** staff will work with the relevant Microbial Sequencing Center to determine whether ***VectorBase*** resources will also contribute to annotation of the *Culex* genome.

The *Cx. p. quinquefasciatus* BAC genomic DNA library and normalized cDNA libraries will form the basic reagents to be sequenced in this MSC project, described in detail in **7** below. Some smaller *Culex* EST projects are currently being funded, including work already done or underway in the laboratories of Professors Hemingway and Christensen. Obviously, these data and mapping data that will be obtained by Dr. Cornel at Davis, will contribute materially to the eventual assembly and annotation of a full *Cx. p. quinquefasciatus* genome. All participants in this project will work carefully to insure that *Culex* sequence and mapping efforts in their laboratories and in the laboratories of other Culex investigators are carefully coordinated with this project so as to avoid any duplication of effort.

## 3. Interest of the Scientific Community in the *Culex* Genome

As described above, several species of the genus *Culex* are vectors of important human diseases. Consequently, they have been the subject of many decades of research by the biomedical and entomological community, with many ongoing research projects based in developed and developing countries throughout the world. The constitution of the *Culex* genome sequence committee (see following section) reflects the interest of the *Culex* research community in the completion of a *Culex* genome project. Members of the committee have discussed this project over the past year by conference calls and by two small meetings held at the 4$^{th}$ International Workshop on Transgenesis and Genomics of Invertebrates held at Asilomar, CA in May, 2003 and at the EMBO Mosquito Biology Meeting held in Crete, Greece in August, 2003. This committee at present includes laboratories currently engaged in research with *Culex* mosquitoes. Clearly, this committee now has the responsibility for expanding its membership to include investigators who work directly on programs to control *Culex* transmission of arbovirus and filarial worm pathogens. It will be particularly important to build stronger bridges with and expand committee membership to include investigators who work directly on filariasis transmission in less developed countries and on WNV transmission in the US and Canada.

## 4. Management of the *Culex* project

Management of the project will be conducted by a *Culex* genome committee consisting of 12 members (listed below), together with at least one member from the chosen sequencing center. Members of this committee are:
**Peter W. Atkinson**, University of California, Riverside, CA; **Frank H. Collins**, University of Notre Dame, Notre Dame, IN; **David W. Severson**, University of Notre Dame, Notre Dame, IN; **Greg C**. **Lanzaro**, University of California, Davis, CA; **Brendan Loftus**, The Institute for Genome Research, MD; **Bruce M. Christensen**, University of Wisconsin, Madison, WI; **Alexander S. Raikhel**, University of California, Riverside, CA; **Janet Hemingway**, Liverpool School of Tropical Medicine, Liverpool, U.K.; **Stephen Higgs**, University of Texas Medical School, Galveston, TX; **Anthony Cornel**, University of

California, Davis, CA (stationed in at the Kearney Agricultural Field station, Parlier, CA); **Michel Raymond**, Universite Montpellier II, France; and **Rob A. Holt**, Michael Smith Genome Science Centre, Vancouver, BC, Canada.

P. Atkinson will chair this committee and will be responsible for preparing and delivering genomic DNA of the strain selected for sequencing to other members of the consortium for genomic library constructions and to the sequencing center. He has the responsibility for coordinating the project. F.H. Collins and R.A. Holt have extensive experience with overseeing the completion of the *An. gambiae* genome project. R.A. Holt has extensive experience directing genome projects both in his former position with Celera Genomics and in his current position with the Genome Science Center in Canada. D. Severson is PI of the ongoing, preliminary stage *Ae. aegypti* genome project. Collins, Severson, and Holt therefore bring significant expertise and experience to this project. As mentioned above, B. Christensen, S. Higgs and J. Hemingway are currently undertaking construction and analysis of EST libraries from West Nile virus and bacteria-infected and insecticide-resistant, *Cx. p. pipens*. F.H. Collins is also PI of ***VectorBase***, a relational database bioinformatics resource center contract proposal currently under review by NIAID. The *Cx. p. quinquefasciatus* genome has already been identified in the ***VectorBase*** contract proposal as one of the vector genomes that this resource will serve in the event that the contract is funded.

**5. Utility of *Culex* Project Results.**

The scientific community is already prepared to utilize sequence data arising from this project. Groups working on insecticide resistance, pathogen transmission, sensory perception, endosymbiotic relationships, endocrinology, metamorphosis, interactions with microbial pesticides, innate immunity, and tissue specific expression of genes in salivary glands, fat body and the midgut will immediately benefit. *Culex* species have been the target of genetic control strategies. Transposable elements, genes involved in meiotic drive, and genes involved in interactions with the endosymbiont *Wolbachia* will be identified as a result of a *Culex* genome project and so should lead to the initiation of new approaches to the genetic control of *Culex* species.

**6. Suitability of *Culex* as an Experimental Organism.**

In addition to biochemical and epidemiological studies on *Culex* as vectors of human pathogens, *Culex* species have also been the subject of the following sample of research foci:

***(a) Systematics of the* Cx. pipiens *Complex*.** The *Cx. pipiens* complex comprises the subspecies *Cx. pipiens pipiens* and *Cx. p. quinquefasciatus*, biotypes “molestus” and “pallens” and in Australia two sibling species *Culex australicus* and *Culex globocoxite*. The question of genetic and systematic relatedness of the complex members is still debatable. These subspecies and biotypes overlap geographically in many locations in North America, South America, Australia, Europe, Africa and in the Middle and Far East. According to Barr (Barr, A. R. “The *Culex pipiens* complex” in “Recent Developments in the Genetics of Insect Disease Vectors” eds, W. W. M. Steiner, W. J Tabachnick, K. S. Rai and S. Narang, Stipes Publishing Co., Champaign, IL, pp551-572, 1982), the only definitive morphological difference between the two subspecies is the male genitalia (nature of the tips of the dorsal arms of the phallosome - blunt in *Cx. p.pipiens*, pointed in *Cx. p. quinquefasciatus*., the alignment of the dorsal arms – divergent in *Cx. p. pipiens*., parallel in *Cx. p. quinquefasciatus*., and the nature of the ventral arms of the phallosome – narrow in *Cx. p. pipiens*, broad in *Cx. p.quinquefasciatus* (from Barr, 1982)).

The degree of autogeny (ability to lay the first egg batch without the need of a blood meal), blood host seeking propensities, breeding site preferences and geographic distribution are used to distinguish the biotypes. The sequence of second internal transcribed spacer (ITS2) of the rDNA is divergent between *Cx. p. quinquefasciatus* and *Cx. p. pipiens* (Severini *et al*., Insect Mol. Biol, 5, 181-186, 1996). The acetylcholinesterase gene has been utilized as a diagnostic tool to distinguish between these two subspecies (Bourguet *et al*., J. Am. Mosq. Control Assoc., 1998, 14: 390-396). For this project we aim to use a colony identified as *Cx. p. quinquefasciatus*.

***(b) Genetic mapping in* Culex p. pipiens.** Narang and Seawright identified 110 mutants of the *Cx. pipiens* complex ("Linkage relationships and genetic mapping in *Culex* and *Anopheles*", in "Recent Developments in the Genetics of Insect Disease Vectors" eds, W. W. M. Steiner, W. J Tabachnick, K. S. Rai and S. Narang, Stipes Publishing Co., Champaign, IL, pp231-289, 1982). Twenty six mutants were located on linkage group 1 (chromosome 1), the shortest of the 3 chromosomes, and 12 were mapped along this chromosome. The sex factor is also located on chromosome 1. Large variations in linkage distances between these loci are observed between *Culex p. pipiens* from different geographical locations and between *Cx. p. pipiens* and *Cx. p. quinquefasciatus.* Twenty one mutants were assigned to linkage group 2 (chromosome 3) of which 15 were mapped. Twenty three mutants were assigned to linkage group 3 (chromosome 2) of which 13 were mapped. Thus, of 110 identified mutants, 70 have been assigned to chromosomes of which 40 were able to be used to develop linkage groups along each of the 3 chromosomes. Some insecticide resistance genes (esterase and acetylcholinesterase) were also assigned to chromosome 2, and the correspondence between linkage groups and chromosomes was settled by *in situ* hybridization (Heyse *et al*., J. Amer. Mosq. Ctr. Assoc., 12, 199-205, 1996). The first photographic map of the polytene chromosomes was published in 1998 (Zambetaki *et al*., Genome, 41, 751-755, 1998).

Mori, Severson and Christiansen (J. Heredity, 90: 160-164, 1999) utilized cDNAs previously isolated from, and mapped in, *Aedes aegypti* to examine the question of synteny between *Ae. aegypti* and *Cx. p. pipiens*. Of 71 cDNAs from *Aedes*, 52 were found to hybridize to *Culex* DNA under stringent conditions and 34 of these displayed strain specific polymorphisms. Twenty one were then used for recombinational mapping. The results showed that the order of all 4 cDNAs along chromosome 1 was preserved between the 2 species but that significant translocations had taken place between chromosomes 2 and 3 such that the left arm of *Aedes* chromosome 2 was homologous to the left arm of *Culex* chromosome 3 but the right arm of *Culex* chromosome 3 was homologous to the left arm of *Aedes* chromosome 3. The left arm of *Culex* chromosome 2 was homologous to the right arm of *Aedes* chromosome 2 and the right arm of *Culex* chromosome 2 was homologous to the right arm of *Aedes* chromosome 3. The total map distance in *Cx. pipiens* is 165.8cM (7.1 + 80.4 + 78.3 cM) vs. 150.8 cM for *Aedes aegpti* (44.0 + 59.8 + 54.3 cM). The basis for the large discrepancy in the total map distances between chromosomes 1 of these species remains unknown. In addition, the larger map distance for *Culex* is in contrast to the estimates of the physical sizes of these genomes: 0.54 pg for *Culex pipiens* and 0.83 pg for *Aedes aegypti* (*Anopheles gambiae* = 0.27 pg).

***(c) Estimates of genome size of* Culex.** The following estimates of mosquito genome size are based on re-association kinetics (from Knudson *et al*., "Genome organization of Vectors: in "The Biology of Disease Vectors" eds. B. J. Beaty and W. C. Marquardt, University Press of Colorado, 1996, pp. 175-214) are consistent with independent estimates obtained by nuclear staining methods (D.W. Severson, personal communication).

| Species | Genome Size (pg) | %FB | %High | %Mod. | %Unique |
|---|---|---|---|---|---|
| *Cx. pipiens* | 0.54 | 0.11 | 0.38 | 0.29 | 0.22 |
| *Ae. aegypti* | 0.83 | 0 | 0.2 | 0.2 | 0.6 |
| *An. gambiae* | 0.27 | 0.06 | 0.33 | 0 | 0.6 |

***(d)* Cx. p. quinquefasciatus *can be genetically transformed using transposons as gene vectors.*** This provides the opportunity to use this technology to examine gene function in *Culex* and to introduce modified genes into *Culex* for the purposes of altering the vectoral capacity of this mosquito (Allen et al.,(2001) J. Med. Entomol.38: 701-710).

***(e)* *Cytoplasmic incompatability*.** The phenomenon of cytoplasmic incompatibility (CI) caused by the presence of the endosymbiotic bacteria *Wolbachia* was first demonstrated in *Cx. pipiens* (see review in Rousset and Raymond, Trends Ecol. Evol. 6**:**54-57, 1991), and this insect remains a model for how effective CI can be in driving a novel genotype through a mosquito population. Elucidating the molecular basis of this should positively influence the development of genetic drive systems in mosquitoes and other insects using CI or related biological phenomena.

***(f)* *Studies on insecticide resistance*.** *Cx. pipiens* has been the target of insecticide control programs since the early sixties. Mechanisms of resistance have been identified and genes coding for targets of DDT and pyrethroids, organophosphates (OP) and carbamates, dieldrin, and *Bacillus sphaericus* binary toxin have been cloned and sequenced (see review by H. Ranson & Hemingway, Ann. Rev. Ent. 45, 371-391, 2000). Gene amplification of dexoxifying esterases conferring resistance to OP has been extensively studied, at the molecular and population level, with the demonstration of the spread of amplified esterases and insensitive acetylcholinesterase across continents (M. Raymond et al., Nature 350,151-153, 1991 and M. Weill et al. Nature, 423,136-137, 2003). Besides occurring in the *C. pipiens* complex, esterase gene amplification was also observed in *C. tarsalis* and *C. tritaeniorynchus*, but never in *Anopheles* or *Aedes*. Comparing genome sequences may reveal why gene amplification is easily possible in *Culex* species but not in other mosquitoes.

***(g)* *Genetic control strategies*.** *Cx. pipiens, Cx. tritaeniorhynchus*, and *Cx. tarsalis* have been the subject of efforts to use genetic manipulations to develop novel control strategies for these species.

**7. Goals and Rationale for the Proposed *Culex* Project**

We are proposing that the complete *Cx. p. quinquefasciatus* genome of the JHB strain be sequenced, assembled and annotated in a two stage process. The first stage would include the work outlined in this proposal (below). The second stage would be a full genome shotgun sequencing project that would use the results of this first project in scaffold assembly, assignment of scaffolds to chromosomes, and annotation of the genome. The completed draft of the *Cx. p. quinquefasciatus* genome project will directly benefit from the completed *An. gambiae* and *D. melanogaster* genomes as well as from the *Ae. aegypti* genome project now underway. For example, based on previous work, we know that the map of the *Ae. aegypti* genome will be an important guide to organizing the *Culex* map, at least at a low resolution level (Mori *et al*., J. Heredity, 90: 160-164, 1999). Likewise the high percentage of *Ae. aegypti* cDNAs that displayed strong hybridization to the *Culex* genome indicates that many of the *Aedes* cDNAs, and perhaps anopheline cDNAs as well, will be of use in defining the gene structures and in characterizing coding sequences obtained from the *Culex* project and in establishing early maps of the genome. Furthermore, the ability to perform polytene mapping on *Culex* salivary gland chromosomes will provide an important tool for organizing and orienting scaffolds along the length of each of the 3 chromosomes.

The two stages of this project are:

(1). The sequencing of clones from normalized cDNA libraries, the sequencing of ends of all clones in a 10X coverage genomic DNA BAC library (approximately 45,000 clones of average insert size of 120 Mb), and the sequencing of approximately 1 Mb of BAC clones chosen based on potential homology with the four *Ae. aegypti* BAC clones that have been sequenced recently at TIGR. Three hundred of these cDNA and BAC clones will also be physically mapped to the polytene chromosomes of *Culex* to construct a medium resolution physical map of *Culex*.

(2). The whole genome sequencing of *Culex* using information obtained in stage 1 (and from the *Anopheles* and *Aedes* projects) to assemble the genome. This would most likely be achieved through a shotgun sequencing project involving plasmid genomic DNA libraries of multiple insert sizes produced and sequenced at the MSC.

At this point, funding is sought for stage 1 only. As described below, cDNA and BAC libraries will be constructed within the consortium which will be sequenced by the selected sequencing center. Selected clones will then be used in polytene chromsosome *in situ* hybridizations and once again this will be accomplished within the consortium. The outcome of stage 1 will be a large amount of sequence data obtained from the cDNA and BAC libraries and the ordering of many of these clones along the polytene chromosomes of *Culex*.

On demonstration of good progress in stage 1, funding will then be sought for the whole genome sequencing of *Culex p. quinquefasciatus*.

**Details of Stage 1 Project.**

***1 Construction and sequencing of normalized cDNA libraries for gene discovery and preliminary stage microarrays for* Culex p. quinquefasiatus.**

**Justification***:* Pathogens vectored by *Culex* mosquitoes pass through various insect tissues before being transmitted to human hosts. For viruses, such as West Nile, invaded tissues are the mid gut epithilium, the hemocoel (and fat body), and the salivary glands. For filarial worms the tissues are the mid gut epithilum, the hemocoel (and fat body), the indirect flight muscles, and the salivary glands. Determining the genes that are transcriptionally up or down regulated in each of these tissues in response to infection by viruses or filarial worms will be an important milestone in tracing the genetic and biochemical consequences of pathogen transmission in *Culex*. Identification of these genes should lead to new strategies for the control of pathogen transmission through *Culex* mosquitoes. In addition, other important mosquito-specific physiological phenomena, including blood-feeding and host-choice, will be investigated by sequencing clones from normalized cDNA libraries.

**Research Plan**: Professor Hemingway has already produced a normalized *Cx. p. quinquefasciatus* cDNA library from mixed tissues (embryo, larvae, pupae, adult males and females) from mixed insecticide resistant and insecticide susceptible strains with support from the Welcome Trust. We propose that 3' ends of 20,000 clones from this library be sequenced at MSC. [All cDNA sequencing will be from the 3' ends to maximize sequence overlap of independent clones and to provide a resource of more value than 5' reads in cDNA microarrays.] This mix of tissue has been selected to maximize gene discovery. In addition, four more normalized libraries will be constructed in the Hemingway laboratory using Welcome Trust support, and 10,000 clones from each will be sequenced as part of this stage 1 project. These libraries will be based on (1) RNA from mosquito heads (including males during night time swarming and day time inactivity and females during

blood meal seeking and feeding as well as day time inactivity), from blood fed females (at a series of time points over the 48 hours after blood feeding when digestion and egg development occur), from West Nile virus infected mosquitoes over a series of time points post infection (RNA supplied by S. Higgs), and from bacteria inoculated mosquitoes over a series of time points post infection (RNA supplied by B. Christensen). Similar numbers of clones from comparable normalized cDNA libraries prepared from *An. gambaie* RNA were sequenced and these numbers yielded excellent gene discovery with little redundancy. This will result in 60,000 EST sequences generated by this component of the project.

**Proposed work sites**: RNA for these libraries will be prepared at Galveston (WNV-infected), Madison (bacteria-infected), Liverpool (mixed stage and tissue), and Riverside (blood fed and heads) for RNA synthesis from infected and non-infected mosquitoes. The mixed stage and tissue normalized library has already been prepared at Liverpool in Hemingway laboratory from a *Cx. p. quinquefasciatus* strain other than JHB. The other four normalized libraries will also be prepared at Liverpool but the strain used will be JHB.

**2. *BAC library construction, BAC end- sequencing, and sequencing of a select set of BAC clones*.**

***Justification**:* End sequencing of clones in a genomic DNA BAC library acts as an important resource for providing architectural information of use in scaffold assembly and assignment to chromosomes. For a genome the size of *Culex* (~540 Mb), a 10-fold coverage BAC library of clones with 120 kb average insert size would include approximately 45,000 clones. The proportion of BAC clones that would be end-sequenced should be determined by the MSC to complement a full genome sequencing plan. Because the end-sequenced BACs will provide both important architectural information as well as sequence-tagged clones that can be used to orient and order scaffolds produced during assembly, we would recommend that most BAC clones be end-sequenced if the cost is not markedly greater than the cost of sequencing genomic DNA containing plasmids produced for a shotgun sequencing project in phase 2.

***Research Plan**:* F.H. Collins and J. Romero-Severson of Notre Dame are currently working with J. Tomkins of the Clemson University Genomics Institute to produce this BAC library from *Cx. p. quinquefasciatus* JHB strain genomic DNA. Both ends of these BAC clones will be sequenced at the chosen sequencing center as well the entirety of several BAC clones shown to overlap with four *Ae. aegypti* BAC clones that have been completely sequenced at TIGR (D.Severson and B. Loftus, personal communication). The purpose of the BAC clone sequencings will be to obtain both a preliminary view of the architecture of the *Cx. p. quinquefasciatus* genome and to compare the gene order and orientation in the sequenced *Ae. aegypti* BAC clones with the homologous regions from *Cx. p. quinquefasciatus*. It is also hoped that by sequencing a relatively large region, a representative sample of the *Culex* repeats will be present within the assembly. This will help gain valuable insight into the size and complexity of these repeats within the genome and allow us to anticipate the difficulties in assembling sequence data coming from a future whole genome approach at a variety of sequence coverage levels. To maximize the possibility of sequence comparisons between genomic sequences of *Aedes* and Culex, 2-3 overlapping BAC-clone tiling paths will be selected to match each of the four *Aedes* BAC clones. We expect that each of these BAC contigs will be about 200-250 kb long for a total of approximately 1 megabase of *Culex* genomic DNA sequence (including the overlaps).

***Proposed Work Sites**:* BAC library construction will be done collaboratively by Notre Dame and Clemson. Sequencing of BAC ends to be conducted at the chosen sequencing center. Identification of BAC contigs for sequencing will be done at Notre Dame, and subcloning and sequencing of these BACs will be done at the MSC.

***(3) __in situ__ __polytene chromosome mapping of select__ Culex p. quinquefasciatus __BAC and cDNA clones__.***

**Justification**: Following the examples of *D. melanogaster* and *An. gambiae*, *in situ* chromosome hybridizations provide an excellent means for positioning and ordering scaffolds along the lengths of insect chromosomes. This was a crucial component of the *An. gambiae* genome project. *Cx. p. quinquefasciatus* has a genome approximately twice the size of *An. gambiae*, and this difference is likely due almost entirely to repetitive DNA, which could complicate assembly of shotgun reads into large scaffolds. Physical mapping information could help compensate for any compromise in the size of assembled scaffolds due to repetitive DNA content. We expect that extensive physical mapping of genomic DNA clones (primarily BAC clones) will be a necessary prerequisite to assignment of assembled contigs to correct chromosome locations and correct orientation. Our experience with the *An. gambiae* genome suggests that approximately 2,000 randomly mapped BAC clones were necessary to reliably assign approximately 85% of the assembled scaffolds to the genome. Had clone choice been based on the location of BAC clones in assembled scaffolds as inferred from BAC end sequences, a smaller number of clones would have provided the same information. Because physical mapping is a very slow procedure, however, chromosome assignment of scaffolds assembled from a shotgun seqencing project should not be entirely on clones that are physically mapping after sequence assembly. Our plan is to do some preliminary mapping in conjunction with the phase 1 sequencing of cDNAs and BAC ends. More extensive clone mapping, will then be done in conjunction with the shotgun sequencing effort, with clones selected for mapping based on their inferred location within assembled scaffolds. The first steps in mapping work will be done largely in the laboratories of A.J. Cornel, F.H. Collins, and J. Hemingway, with resources that are not part of this MSC project. More extensive physical mapping will be an important prerequisite for the phase 2, full genome shotgun sequencing project, and the phase 2 proposal will not be submitted until a clear plan for this more extensive mapping effort is in place.

**Research Plan**: A. Cornel is currently developing a cytogenetic map for *Cx. p. quinquefasciatus* with support from NIAID/NIH. Cytogenetic mapping will occur in conjunction with the phase 1 proposal but will not use MSC resources. This mapping will include A. Cornel's NIAID/NIH-supported work as well as additional mapping in the laboratories of F.H. Collins and J. Hemingway covered by other resources (Indiana Center for Insect Genomics funds in the Collins laboratory and Liverpool School of Tropical Medicine Director's funds in the Hemingway laboratory). We estimate that approximately 20-30 BAC clones will need to be mapped in order to validate assembly of the four BAC contigs that will be sequenced as part of the phase 1 project. Furthermore, we will determine the map location of a set of 300 randomly selected BAC clones. Assuming this set is uniformly distributed (we recognize this is an unrealistic assumption), we should have a clone spacing of about one BAC every ~2 mb. This should enable us to physically locate and orient most scaffolds that are ~5 Mb or larger. For the *An. gambiae* genome, this represents ~50 % of the scaffolds in the 278 Mb assembly, but it will probably be a smaller percent of the *Culex* genome because of this mosquito's larger genome size and greater amount of repetitive DNA. Nonetheless, this preliminary set of mapped BACs would enable a useful chromosome assignment of much of the first assembly from a shotgun sequencing project. Assignment and orientation of additional scaffolds to genome locations could then be done by mapping selected BAC or cDNA clones.

***Proposed Work Sites**:* Clone mapping will be done at the University of California at Davis, Kearney agricultural field station (CA), the Liverpool School of Tropical Medicine, U.K., and the University of Notre Dame (IN).

**8. Strain choice and availability of DNA for the project.**

As described above, we have elected to sequence a strain of *Culex p. quinquefasciatus* that consistently yields high quality chromosome spreads suitable for *in situ* hybridization studies. This strain is currently maintained at the Kearney Agricultural Field Station (Parlier, CA), is insecticide susceptible and so can be transported and maintained in other laboratories world-wide. The Johannesburg (JHB) strain was established from offspring reared from 20 egg rafts all collected in a single unused fish pond in March 2001. Population genetic studies revealed that the Johannesburg *Cx. p. quinquefasciatus* population was genetically isolated from a sympatric *Cx. p. pipiens* population (Cornel AJ, McAbee RD, Rasgon J, Stanich MA, Scott TW and Coetzee M. 2003. *J. Med. Entomol*. 40: 36-51). Consequently, the JHB strain is expected to possess a high degree of genetic divergence from *Cx. p. pipiens.* The JHB colony is currently in its 30$^{th}$ generation since colonization. Genomic DNA preparations suitable for cloning and other molecular uses can be readily obtained from this strain using protocols developed for Drosophila and other mosquito species. Based on the many generations of inbreeding at relatively small population size (often no more than about 50 females per generation), we anticipate that the extent of genetic variation in this colony will be much smaller than that present in the *An. gambiae* PEST colony that was used in the *An. gambiae* genome project. In addition to having excellent polytene chromosomes in larval salivary gland tissues, this colony has also been found to be free of any visible polytene chromosome inversions.

**9. Availability of other sources of funding.**

Other funding that will directly or indirectly supports the goals of this project is being contributed by the Wellcome Trust, the Indiana Center for Insect Genomics, the Cornel NIAID R01, and the Liverpool School of Tropical Medicine. When application for phase 2 funding is initiated, the Michael Smith Genome Science Centre in Canada will be asked to support sequencing of full length *C. p. quinquefasciatus* cDNAs.