Grant proposal for the construction of four lepidopteran BAC libraries

Collaborative proposal from:

Antónia Monteiro (contact person)
Department of Biological Sciences
University at Buffalo
Buffalo, NY 14260
Tel: 716 6452363 ext: 135
e-mail monteiro@buffalo.edu

Paul Brakefield
Institute of Biology
Leiden University
2300 RA Leiden , The Netherlands

Bas Zwaan
Institute of Biology
Leiden University
2300 RA Leiden , The Netherlands

Patricia Beldade
Dept. of Ecology and Evolutionary Biology
University of California Irvine
Irvine, CA 92697-2525

Ilik Saccheri
School of Biological Sciences
University of Liverpool
Liverpool L69 7ZB, UK

Richard ffrench-Constant
Department of Biology and Biochemistry
University of Bath, UK

Marian Goldsmith
Biological Sciences Department
University of Rhode Island
Kingston, RI 02881-0816

Fred Gould,
Department of Entomology
North Carolina State University
Raleigh, NC 27695-7634

Chengcang Wu
The Texas A&M BAC Center
Department of Soil & Crop Scien
Texas A&M University
College Station, Texas 77843-212

J. Mark Scriber
Dept. of Entomology
Michigan State University
E. Lansing, MI 48824

Richard Harrison
Dept. of Ecology and Evol. Biolo
Cornell University
Ithaca, NY 14853

June 10, 2003

Project Goals

The goal of this proposal is to build BAC libraries for a set of well-established lepidopteran insect models in the study of the genetic and developmental basis of adaptation and diversification. Readily available BAC libraries for these species will accelerate research into the genetic mechanisms underlying lepidopteran development, morphology, behavior, and evolution.

Our specific objectives are to produce 8 high-quality BAC libraries for each of four different experimentally important lepidopteran species, a Satyrid butterfly, *Bicyclus anynana* (a model for polyphenism/phenotypic plasticity, wing pattern "evo-devo" and life history evolution, including longevity) a Papilionid butterfly, the Eastern tiger swallowtail *Papilio glaucus* (a classic example of Batesian mimicry, melanism, and the biochemical basis of pigment production, and an important model for studying the molecular and evolutionary mechanisms underlying hostplant preference and detoxification of plant allelochemicals and xenobiotics), a Noctuid moth, *Heliothis subflexa* (a model for the evolution of resistance to chemical and biological pesticides, and for polyphagy *vs.* specialized hostplant preference), and a Pyralid moth, the European corn borer *Ostrinia nubilalis*, (a model for insect communication via pheromones).

Background and significance: Why BAC libraries for Lepidoptera?

Project rationale. The real challenge of the genomic and post-genomic era of scientific investigation is the integration of sequence information with functional genomics within a well-understood behavioral, evolutionary and physiological context. In other words, DNA sequences will be most valuable when they can be readily related to function, and understood in the context of the developmental, physiological, and biochemical processes they underlie.

Although completion of the *Drosophila melanogaster* and *Anopheles gambiae* genome projects has accelerated work on Lepidoptera molecular genetics, many genes and regulatory networks, as well as overall genome organization have likely undergone radical evolutionary change in the 230-240 million year that separate the Diptera and the Lepidoptera. In addition, many characteristics, such as the development of pigmented scale cells and an elaborate patterning mechanism for coloring the wing, are novel features of Lepidoptera (Nijhout, 1991, Galant et al., 1998). Furthermore, even though the Lepidoptera is represented by more than 170,000 species (one of the most diverse orders among animals), including major agricultural pests (*e.g.* Noctuids and Pyralids), plant pollinators, the only truly domesticated insect (the mulberry silkworm, *Bombyx mori*), and numerous species that are part of tropical conservation efforts, species of Lepidoptera are clearly under-represented in terms of genomic resources relative to their biological and economic status. The present proposal will contribute to filling this gap and provide the necessary material for comparative genomic studies across the Lepidoptera (*e.g.* evolution of genome organization, synteny).

A growing Expressed Sequence Tag (EST) database (*SilkBase*: http://www.ab.a.u-tokyo.ac.jp/silkbase; currently 36,000 sequences), BAC libraries (Wu et al., 1999; K. Mita and P. deJong, personal communication), physical maps, and a 2.5X whole genome shotgun draft (K. Mita and T. Shimada, personal communication) are now available or underway for *B. mori*; additionally, a BAC library has been made for the fall armyworm, *Spodoptera frugiperda* (P. Fournier, personal communication). Other BAC libraries and limited EST projects are planned or in progress (http://www.ab.a.u-tokyo.ac.jp/lep-genome/), and GenBank (http://www.ncbi.nlm.nih.gov:80/entrez/) lists nearly 40,000 entries for Lepidoptera, but the majority of these (~30,000) are silkworm-derived. Although the silkworm genomic resources can provide a partial foundation for lepidopteran studies, *B. mori* does not exhibit many of the features that make moths and butterflies interesting model systems. For example, it lacks derived wing patterns, complex feeding and mating behavior, or variation to study the genetic and developmental basis of adaptive change (including life history evolution). Lepidoptera are also a high profile group for assessing biodiversity and educating the general public

about the importance of conservation. Additional genomic resources such as the BAC libraries proposed here are urgently needed to enable investigation of a much broader set of phenomena and to enable the rigorous comparative approach needed to unravel the underlying genetic and epigenetic mechanisms responsible for diversification in this important taxonomic group.

Lepidopteran evolution and targeted species. The construction of a set of lepidopteran BAC libraries will provide critically important tools for every aspect of insect science, especially for comparative genomics and future sequencing of model genomes. Comparative studies on genome sequences between Lepidoptera, Diptera, nematodes, and vertebrates has already revealed lepidopteran-specific mechanisms of sex determination (Suzuki et al. 2001) and lepidopteran-specific genes (K. Mita and T. Shimada, personal communication). As well as illustrating at the molecular level what makes Lepidoptera different from other insects, newly identified sequences and structural features made possible by large-fragment genomic libraries will provide tools for understanding the evolution of the Lepidoptera themselves. The paucity of the fossil record and lack of informative taxonomic characters leaves much early lepidopteran evolution poorly resolved. Whereas current efforts in phylogenetics are aimed at analyzing coding regions of conserved genes with sufficient variation to sort out relationships deeper in the lepidopteran phylogeny (Friedlander et al., 2000), the knowledge of gene organization on a broader scale, including details of exon-intron structure, regulatory sequences, changes in the number and arrangement of closely linked genes and clustered multigene families, and the degree of conservation of specific chromosome segments, promises to provide an important new set of tools for unraveling evolutionary pathways. Only with a global picture of the overall structure and plasticity of the genome can we understand and appreciate how, for example, the detailed mechanisms giving rise to diverse feeding behavior of closely related species or the startling and rapid changes in butterfly wing patterns have arisen and been maintained.

All of the species under consideration in this proposal belong to the Macrolepidoptera, for which the first known fossils date from the Paleocene, about 50 mya (Walley 1986). The origin of the group is likely to be much older. The clade subsequently underwent an explosive radiation, and today includes more than 86,000 extant species (Kristensen 1999; Wagner 2001). In all, members of 3 superfamilies are included in this proposal: the butterflies, *Bicyclus anynana* and *Papilio glaucus,* are Papilionoidea, which encompass 13,600 species. *Heliothis subflexa* belongs to the Noctuoidea, the largest lepidopteran superfamily, with 41,300species. *Ostrinia nubilalis* belongs to the Pyraloidea, which comprises 16,000 named species and probably at least as many more remaining to be described. In addition to the existing BAC libraries for *B.* mori, BAC libraries are presently under construction for *Heliconius erato*, another papilionid, *H. virescens*, a noctuid, and *Manduca sexta*, which together with *B. mori* represents a fourth superfamily, the Bombycoidea, a medium-sized clade with 3550 species. Macrolepidopteran superfamilies not included in this or in previous proposals include the Geometroidea with around 20,000 species and four other superfamilies (Axioidea, Calliduloidea, Hedyloidea, and Drepanoidea) with less than 1000 described species each.

1-The importance of the organisms to biological research:

Bicyclus anynana* and *Papilio glaucus: Butterfly wing patterns, the genetic basis of pattern formation, mimicry and polyphenism, and life history evolution (A. Monteiro, P. Beldade, P.M. Brakefield, B. Zwaan, I. Saccheri, R. ffrench-Constant, M. Berenbaum, M. Scriber, P. Adolfatto, S. Carroll, D. Stern). Research on butterfly wing patterns spans over a hundred and fifty years and is expanding rapidly, fueled by innovative studies into the developmental and genetic basis of pattern formation (Nijhout 1991, Brakefield et al.1996, Brakefield and French 1999, McMillan et al. 2002, Beldade and Brakefield 2002, Monteiro et al. 2003) and the ecological and evolutionary significance of these patterns (Scriber et al 1995, 1996, Brakefield 1996; Brakefield and French 1999, McMillan et al. 2001; Jiggins et al.2001; Kapan 2001, Deering and Scriber 2002, Scriber 2002a, Scriber et al. 2003, Andolfatto et al 2003). As a result, butterfly wing patterns are exceptional model systems with which to link the developmental and genetic processes that generate morphological variation with the ecological and evolutionary processes that mold this variation in natural populations (reviewed in

McMillan et al. 2002). Although there are still wide gaps in our understanding, there is perhaps no better animal group where the transitions from genes, through developmental pathways, to phenotype, function, and fitness can be clearly illuminated (Brakefield 1998).

The generation of BAC libraries for two additional model butterfly species, *Bicyclus anynana* and *Papilio glaucus*, will accelerate research on butterfly wing patterns and allow a far more comprehensive picture of the interplay between genes, development and evolution. The two targeted species are representatives of taxonomically divergent lineages within butterflies and were chosen because they are already important models for research addressing the ecology, evolution, and development of butterfly wing patterns. *B. anynana* (Nymphalidae, Satyrinae) is one of the best examples of seasonal polyphenism and phenotypic plasticity in animals, in wing patterns as well as in life history, and is emerging as a model organism to study the molecular developmental events of pattern formation (reviews in Brakefield and French 1999, Beldade and Brakefield 2002). The Eastern Tiger Swallowtail, *P. glaucus* (Papilionidae, Papilioninae), is a striking and well-studied example of batesian mimicry and sex-linked melanism (Clarke & Sheppard 1959, 1962; Clarke & Clarke 1983; Scriber et al. 1996). In addition to falling on diverse evolutionary lineages within the butterflies, the wing patterns of *B. anynana and P. glaucus* are subjected to different selective pressures (*i.e.*, warning versus cryptic coloration, developmentally plastic versus canalized patterns; seasonal versus geographic variation). Thus, understanding the genetic and developmental basis of pattern formation in these species will help determine the relative importance of phylogenetic versus selective history in the generation of novel morphological features. There are presently more than 20 different laboratories working on wing patterns on these or closely related species that would directly benefit from BAC libraries. Moreover, *B. anynana* is used in several research projects on the evolutionary genetics of life histories in seasonal environments. Among these projects is a large multidisciplinary research program on the genetics of longevity involving humans and model organisms. A complete set of molecular genetic tools is essential for the exploitation of the full potential of this and similar projects.

Both model species are poised to rapidly exploit the genomic resources provided by BAC libraries. Our understanding of the molecular basis of the early events of pattern formation was greatly expanded with the discovery that homologues of many conserved regulatory genes, and even whole regulatory pathways, are co-opted to play a role in eyespot formation (Carroll et al. 1994; Brakefield et al. 1996; Weatherbee et al. 1998; Weatherbee et al. 1999; Keys et al. 1999, Brunetti et al. 2001). Recently, genetic variation surrounding the butterfly homologue of the transcription factor *Distal-less* (*Dll*), was directly implicated in the control of eyespot size (Beldade et al. 2002). The construction of a BAC library will aid in the detailed study of cis-regulatory regions of this and other known patterning genes in *Bicyclus*. Transgenic manipulations are also underway in this species, where the *piggyBac* transposable element carrying the marker gene GFP under the control of an eye specific promoter has successfully integrated into the genome of this butterfly (J. Marcus and A. Monteiro, in prep.). These experiments represent the first transgenic manipulations in a butterfly. This technique will allow researchers to test the functional role of specific genes or cis-regulatory sequences in pattern formation and to identify novel patterning loci via transposable element mediated mutagenesis. A BAC library will prove invaluable for further identification of the interrupted genes, facilitating the retrieval and identification of the genomic sequences flanking insertion sites. Furthermore, genetic linkage maps are being constructed in *Bicyclus* (B. Zwaan, P. M. Brakefield, I.Saccheri, personal communication) to dissect the genetic architecture of important life histories and wing pattern traits. *Bicyclus* is easily and rapidly reared (12 generations per yr) and mapping strategies can be developed to pinpoint novel Quantitative Trait Loci (QTL), or single gene mutations associated with pattern change.

In another system, recent research by Koch and co-workers on melanization in *P. glaucus* is providing insights into the latter stages of pattern development. Within *P. glaucus* the change from the nonmelanic to melanic forms is controlled by a dominant allele at a sex- or W-linked locus, plus a modifying Z-linked locus that can suppress the melanic phenotype (Scriber et al. 1996). This switch is achieved by finely coordinating the expression of two genes that cause concordant changes in both

pigment synthesis and scale maturation (Koch et al. 1998; Koch et al. 2000a; Koch et al. 2000b). Understanding the coordinated regulation of these genes is the essential next step and is critical for determining how a single locus can produce such dramatic morphological changes.

The development of scale-covered wings, pigmentation, and an elaborate patterning system are key evolutionary innovations of the Lepidoptera which might have been at the basis of the extraordinary diversification of this group (Nijhout 1991, Galant et al. 1998). To date, we know something about early gene expression in one pattern element, the eyespot, but there is much to be discovered (McMillan et al. 2002, Beldade and Brakefield 2002). Access to genomic resources, such as arrayed BAC libraries, will have direct and synergistic effects on an active research community. Over the short term, the accessibility of BAC libraries in *Bicyclus* and *Papilio* will increase our understanding of the regulation of known patterning genes and facilitate the discovery and characterization of novel patterning loci. As color-pattern loci and their regulatory regions are located and sequenced, it will be possible to probe directly the history of the genetic modifications that led to adaptive change and that underlie species differences. Ultimately, an understanding of the molecular and developmental basis of wing pattern formation in these species will allow researchers to address some of the more challenging questions in ecological and evolutionary research. Not only will we know how mimicry and polyphenism are achieved, but we will have a remarkable comparative data set to explore

1) how development constrains or biases evolutionary change,

2) the extent to which patterning loci and their cis-regulatory regions are reused in evolution, and

3) the kinds of developmental changes that take place during evolution and how they vary across diverse butterfly lineages and different selective landscapes.

Although we have focused on the importance of these species for understanding adaptive morphological change, the papilionids present remarkable examples of herbivorous specialists and generalists that are metabolically adapted to detoxify and sequester a suite of noxious compounds from their host plants, an ability that underlies their roles in mimicry. For example, understanding the plasticity of structure and function of the large multigene family of cytochrome p450 enzymes that enable these insects to metabolize plant allelochemicals and xenobiotics is a subject of intense study, and more than a dozen genes have been characterized in the polyphagous *P. glaucus* and its specialist congener, *P. canadensis* (Hung et al., 1997; Li et al., 2001, 2002). A complete inventory of the relevant genes in *P. glaucus* would go a long way toward understanding host plant patterns in this group.

Heliothis subflexa: Effects of evolutionary adaptations on genomic structure. (N. Vickers, C. Schal, S. Gill, M. Adang, F. Gould)**.** Recent comparative analysis of the genomes of *Drosophila melanogaster* and *Anopheles gambiae* has demonstrated that major gene families such as those coding for the P450 metabolic enzymes (Ranson et al. 2002) and those coding for odorant receptors (Hill et al. 2002) are evolutionarily labile. Gene phylogenies indicate that most odorant binding genes in *A. gambiae* are derived from other *A. gambiae* genes and not from genes of a common ancestor of *D. melanogaster* and *A. gambiae* that may have had a similar function. These species are ecologically quite distinct so the differences may have evolved due to selection for altered olfactory systems. However, these species are distantly related, so it is possible that the change resulted from non-Darwinian evolutionary forces. When the genome of *D. pseudobscura* is sequenced it will be possible to assess the degree of genomic differentiation between two closely related species, but in this case little is known about the natural ecology of either of these two species. To determine if evolution of divergent ecological niches could result in rapid alteration of the genome it would be useful to compare two closely related species that are ecologically distinct. The species pair, *Heliothis virescens* and *H. subflexa,* offer this opportunity. *H. virescens* and *H. subflexa* both feed on plants as larvae. However, *H. virescens* is a generalist that feeds on over 14 families of plants while *H. subflexa* only feeds on some species within a single plant genus. The two species are so closely related that they can be hybridized in the laboratory. (They do not hybridize naturally because they have 4 distinct chemical components to their female pheromone blends.)

It has been hypothesized that generalist herbivores must produce a broader array of detoxifying enzymes than specialists because their varied diet will include a more diverse set of toxins. Some evidence in support of this hypothesis is available (Gould et al. 1982, Li et al. 2001). It has been demonstrated that P450 enzymes are important in detoxification of plant toxins. As with genes coding for odorant receptors, Ranson et al. (2002) demonstrated that the P450-coding genes of *A. gambiae* were very distinct from most of the P450-coding genes of *D. melanogaster*. A genomic comparison of *H. virescens* and *H. subflexa* would be very helpful in gaining a broad assessment of the origin and diversity of a number of detoxification enzyme-coding gene families (e.g. P450s, esterases, transferases) in these species (see Ranson et al. 2002).

Because *H. subflexa* females are only attracted to *Physalis* plant species, their antennae are expected to be highly enriched with receptors for odors from this single plant genus. In contrast, *H. virescens* is expected to have receptors for a few odors that are common to many plants, and/or to have a large set of receptors that bind specific odorants from their host plants. As in the case of the detoxification-related genes, evolutionary selective pressures could cause modification of existing alleles or result in a proliferation of new gene families. A genomic analysis would be valuable in distinguishing between these two extreme possibilities.

Work on developing a BAC library for *H. virescens* is almost completed. A similar BAC library for *H. subflexa* would enable researchers to test the above hypotheses.

Because of its status as one of the most damaging agricultural pests in the US, *H. virescens* has been exposed to an array of insecticides over the past 60 years. It has developed resistance to many classes of pesticides (e.g. organophosphates, pyrethroids). In contrast, *H. subflexa* has not been exposed to insecticides. A genomic analysis could reveal changes in recently collected *H. virescens* specimens when compared to *H. virescens* museum specimens from the 1960s, and to *H. subflexa*. Molecular genetic analyses have already revealed much about *H. virescens* adaptation to agricultural control practices. Resistance mechanisms in *H. virescens* that have evolved in response to pyrethroids show point mutations in a sodium channel. This work has helped define the role of these sites in sodium channel function (Zhao et al., 2000). By using a straightforward genetic mapping strategy, it was recently shown that a laboratory selected strain of *H. virescens*, YHD2, which shows a complete loss of binding to a *Bacillus thuringiensis* (Bt) Cry1Ac toxin (Jurat-Fuentes et al., 2000), developed resistance based on disruption of a cadherin-super family gene due to a retrotransposon-mediated insertion (Gahan et al., 2001). This example reveals different classes of mutations in response to different insecticides, although knowledge of the genes involved in additional instances of Bt resistance in *H. virescens* is needed to determine whether disruption of cadherin or retrotransposon-mediated insertions are widespread genomic responses. A genomic analysis would be critical in such an assessment.

Even though *H. virescens* and *H. subflexa* differ substantially in their ecology and are known not to be sister species, they can be hybridized in the laboratory. The hybrid female offspring are fertile and can be backcrossed to either parent. After two generations of backcrossing the males regain fertility in most types of crosses. Because Lepidoptera have many small chromosomes (31 in *Heliothis*) and have no recombination in females, it is possible to create specific crosses that keep all of the genes of entire chromosomes together for ease of initial mapping of QTL, or break up these linkage groups when finer scale mapping is desired.

In most comparative genomic studies, differential gene expression between species is used as a means to identify candidate genes possibly responsible for phenotypic differences. From these candidate genes it is also possible to infer gene function. However, because so many genes differ between species it is difficult to determine, without doubt, if the differences in single genes are responsible for phenotypic differences between the species. One great advantage of the *H.subflexa/ virescens* system comes from the ability to hybridize and backcross. This utility was exploited successfully by Sheck

and Gould (1996) to move genes for ability to feed on soybean from the *H. virescens* genome into the *H. subflexa* genome, allowing for a comparison of the genetics of host range in these two species. Such comparisons cannot be readily made between species that cannot be crossed. Such hybridizations were also used to first map and then identify the major *B. thuringiensis* resistance gene in *H. virescens* (Heckel et al. 1999; Gahan et al. 2001).

A number of studies are in progress that use AFLP analysis of backcross progeny to discover QTL that are responsible for differences between *H.subflexa* and *H. virescens* in host range, mate communication and virus resistance. Two independent QTL found can partially explain variation in each of two pheromone components, and fine scale mapping of pheromone components has begun (F. Gould, unpublished observations). Four QTL have been found that are related to host range, but each of them explains less than 10% of the phenotypic variation, which means that additional factors remain to be identified. Studies on the genetics of virus resistance are just beginning. However, because the difference between the two species in virus resistance is one thousand fold, it seems likely that QTL will be found.

In order to move beyond AFLP mapping of QTL to the actual identification of ecologically relevant genes it will be necessary to complete fine scale mapping. BAC libraries will then become essential to identify and sequence candidate genes via positional cloning strategies. Direct comparison of the corresponding chromosomal regions of these two congeners carrying ecologically relevant traits will begin to answer the question of what kinds of evolutionary forces have shaped lepidopteran genomes.

***Ostrinia nubilalis*:** Pheromone communication (R. Harrison, W. Roelofs, C. Linn, D. Heckel). In many groups of Lepidoptera, mate finding and mate recognition depend on pheromone communication systems in which females produce and males respond to specific mixtures of volatile organic molecules (Lofstedt 1993; Roelofs 1995). Variation in pheromone production and response can result in pre-zygotic barriers to gene exchange and thus contribute directly to speciation and diversification. Although studies of natural populations have revealed a remarkable diversity of moth pheromones (H. Arn, The Pherolist, www.nysaes.cornell.edu/pheronet/), and the chemical pathways for pheromone production are well described (Roelofs and Wolf 1988; Roleofs 1995), little is known about genetic variation underlying differences in pheromone production and male response, and how such variation accounts for the remarkable diversity and species specificity of these communication systems. Initial efforts to clone and express genes encoding enzymes involved in moth pheromone production have been successful (e.g., Rosenfield et al. 2001; Liu et al. 2002), and hypotheses to explain the origin of novel pheromone blends (in the face of strong selection) have been proposed (Roelofs et al. 2002). Nonetheless, the full array of genes and proteins responsible for conferring specificity in moth pheromone communication systems remain to be elucidated (e.g., see Willett and Harrison 1999). Lepidoptera provide unique opportunities to dissect the genetic and molecular basis of chemical communication; the data obtained will complement work being done in model organisms (e.g., *Drosophila* and mouse; see Dulac 1997, Takahashi et al. 2001, Stowers et al. 2002).

The European Corn Borer (ECB; *Ostrinia nubilalis*) is perhaps the most important lepidopteran model for studying the genetics of pheromone communication systems. This moth was introduced into the United States early in the 20th century and has had a significant economic impact as one the most important pests of maize. Female corn borers produce and males respond to a mixture of the Z and E isomers of 11-tetradecenyl acetate. In most populations in Europe and North America, the major component appears to be the Z isomer. Trapping data reveal that males respond to a 3:97 blend of the (E)/(Z)- 11-tetradecenyl acetates (Klun and Cooperators 1975; Anglade et al. 1984). However, males from some populations in both Europe and the United States respond to a 99:1 (E)/(Z) sex pheromone blend of the two components. Klun and Maini (1979) first showed that female pheromone production in the ECB is controlled by a single autosomal factor with two alleles. Subsequent data (Roelofs et al. 1987) have confirmed this result. Further studies have characterized the genetic basis of male electrophysiological and behavioral responses to pheromone. Recordings from single olfactory sensilla on the male antennae differ between the two strains and these differences appear to be determined by a

single autosomal factor (Roelofs et al. 1987; Hansson and Lofstedt 1987). Male behavioral response to a range of sex pheromone stimuli can be assayed in a flight tunnel (Glover et al. 1990, 1991a), and comparisons of the responses of parental, F1, F2 and backcross individuals revealed that behavioral response is controlled by a single sex-linked factor (Roelofs et al. 1987; Glover et al. 1990) and therefore is clearly not linked to the locus controlling pheromone production in females.

The pheromone production "gene" and the male behavioral response "gene" have recently been placed in the context of an AFLP/microsatellite genetic linkage map for ECB (Dopman, Bogdanowicz and Harrison, unpublished). Thus, chromosomal regions in which the relevant genes occur can now be defined and (with a dense map and a BAC library) candidate genes can be identified and sequenced. For pheromone production, Roelofs (personal communication) has suggested one possible candidate gene (a reductase), variation in which may account for the difference in specificity between the ECB pheromone strains. Furthermore, major shifts in pheromone biosynthetic pathways between different species within the genus *Ostrinia* appear to result from changes in the activity level of sex-pheromone desaturases, which occur as a small gene family in all species examined (Roelofs et al. 2002). A detailed understanding of the evolution of pheromone communication systems and of the mechanistic basis for species specificity depends on characterizing the structure and function of the molecules (including the pheromone receptor) that are involved. Given the substantial background information on pheromone chemistry and biosynthesis, the simple genetic determination of both production and response, the availability of a genetic linkage map, and the ease with which the system can be manipulated (e.g., the moth is easily reared with about 12 generations/year), ECB appears to be an ideal system in which to pursue this understanding. A BAC library for ECB will obviously facilitate proposed work at both the mechanistic and the comparative levels.

2-Uses to which the BAC library would be put, in addition to genomic sequencing

All lepidopteran BAC libraries: The first fruits of research undertaken using the BAC libraries proposed here, together with those already made or under construction (Wu et al., 1999; Mita et al., unpublished; Wu and Zhang, unpublished, Fournier and d'Alencon, unpublished), will be to provide a set of conserved anchor loci that will serve as a framework to unify genetic studies in the Lepidoptera by enabling a rigorous test for synteny, and to investigate phylogenetic relationships, which are still largely unresolved for most of the clade.

Collinearity of chromosomal gene arrangement or “synteny” is widespread in vertebrate and plant genomes. It enormously facilitates the identification, map-based cloning, and annotation of orthologous genes, especially for poorly characterized organisms for which a reference species is available. Thus, synteny would be highly advantageous for tracking the rise of insecticide resistance in agricultural pests like the noctuids and pyralids as well as in non-target species; it would also serve as an important tool for investigating similarities and differences in other evolutionarily adaptive traits as described here. No study has rigorously and systematically tested for synteny in Lepidoptera, although several investigators are now undertaking this work using a variety of species (e.g., M. Goldsmith, D. Heckel, P. Andolfatto, C. Wheat, K. Sahara, Y. Yasukochi, personal communication). Sex (Z)-linkage of a few apparently similar morphological traits and allozymes (Sperling, 1994; Raijmann et al., 1997) suggest that synteny may exist within this order of insects. Nevertheless, a key evolutionary feature of the Lepidoptera is the presence of “holocentric” chromosomes, which bear many, longitudinally distributed microtubule attachment points or “diffuse kinetochores” (Goldsmith, 1995; Wolf et al., 1997). These have been postulated to help maintain chromosome fragments through many rounds of cell division (Marec et al., 2001), possibly accelerating the potential for chromosome rearrangement and evolution. Among invertebrates (protostomes), nematodes, which also have holocentric chromosomes (Dernburg 2001), and Diptera, whose chromosomes are monocentric, exhibit some broad-scale chromosome conservation, but extensive internal reshuffling has occurred, so that conserved gene order is generally limited to microsynteny (Guiliano et al., 2002; Ranz et al., 2001;

Roethele et al., 2001; Sharakov et al., 2002; Zdbodnov et al., 2002). Given that the evolutionary time scale for the lepidopteran species under consideration here is of the same order as for several of the better-studied dipterans (Ranz et al., 2001; Sharakov et al., 2002), the level of chromosomal synteny is of interest, not only for its potential practical application in genetic and genomic research, but also to provide a more comprehensive understanding of the modes of evolution of insect genomes.

***Bicyclus anynana*:** We will use the BAC libraries to **A)** understand the molecular basis of gene co-option, e.g., compare color pattern genes and their cis-regulatory regions across different species of Lepidoptera (from more basal to more derived lineages), that originated before and after the appearance of the respective gene expression pattern and color pattern. **B)** Identify interrupted genes that will result from a mutagenesis/enhancer trap screen using *piggyBac* transposable elements. **C)** Assist current QTL mapping approaches for traits such as egg size, developmental time, starvation resistance and ventral eyespot size, where fine scale mapping and (candidate) gene cloning is planned. **D)** Map genes affecting inbreeding depression, and controlling the major differences in fertility load (i.e. deleterious mutations) among the sexes.
***P. glaucus*:** We will use BACs to **A)** search for Z- or female linked genes involved in the melanic pattern polymorphism, and for a W-linked color suppressor of the mimetic black female form. Individual BACs can be probed to genomic DNA of males and females to look for Z and W-linked polymorphisms. Z, or W-linked BACs can be sequenced in their entirety to look for candidate melanic genes like the homolog of *Drosophila* ebony (Koch et al. 2000a). **B)** To map Z-linked genes involved in regulating diapause/obligate diapause in *P. glaucus/P. canadensis.* These ecologically very significant genes are important in hybrid zone dynamics and speciation (Hagen 1989; Hagen and Scriber 1995). The temporal isolation of introgressed hybrid zone individuals relative to the parental genotypes of *P. canadensis* and *P. glaucus* may have given rise to a cryptic (putative) new mountain species (*P. appalachiensis*); and we believe this is achieved by post-diapause delay in spring emergences regulated by the od+ locus of the Z chromosome (and the closely linked ldh-100 allozyme locus; Scriber 2002b; Scriber and Ording unpubl.). **C)** To characterize the large multigene family of cytochrome p450 enzymes that enable these insects to metabolize plant allelochemicals and xenobiotics.
***H. subflexa*:** We will use BACs to identify genes **A)** involved in metabolism of plant toxins. **B)** involved in host odor recognition **C)** coding for distinct pheromone blends **D)** involved in male response to specific blends. **E)** involved in insecticide resistance.
O. nubilalis: We will use BACs to **A)** identify and characterize the gene(s) responsible for differences in pheromone production and male behavioral response. Once identified, these gene regions are obvious targets of study in other lepidopteran genera in which species utilize different pheromone blends. **B)** Characterize the desaturase gene family, including analysis of cis-regulatory regions. Comparative data from this region will provide insights into the history of gene duplication events and, therefore, into the genetic basis for the origin of novelty. **C)** Ultimately, provide a resource to other researchers - including those interested in the evolution of insecticide (e.g. Bt) resistance.

3-Size of the research community that could potentially use the BAC libraries.

There are already a number of labs interested in these species as model systems to address fundamental questions in development, evolution, ecology, behavior, life history, genetics, neurobiology, cell biology, physiology, and systematics, who can immediately benefit from the proposed BAC resources. In addition, there is a large, international scientific community investigating basic biological/physiological processes in pestiferous species (e.g., close relatives to *H. subflexa* and *O. nubilalis*) whose research will be accelerated by the ability to more quickly isolate and study genes with potential as species-selective targets for control, with consequent reduction in the negative impact on human health and ecosystems of conventional chemical insecticides. The BAC resources will also allow for an expansion in the community interested in comparative genomics of Lepidoptera, of insects in general, of arthropods, and of animals. As an indicator of the impact of this research on

international science, there are 13,000+ citations in PubMed for the keyword “lepidoptera,” of which 3500 are listed under “lepidoptera AND gene,” more than 600 under “lepidoptera AND evolution,” and nearly 500 for “lepidoptera AND pheromone” (May, 2003).

The infrastructure for networking among lepidopteran scientists has steadily expanded in recent years. This has included establishment of an International Lepidopteran Genome Project Consortium in August, 2001, specifically “to promote international cooperation to sequence the genome of *Bombyx mori* and undertake comparative genomics of other economically and scientifically important Lepidoptera” (http://www.ab.a.u-tokyo.ac.jp/lep-genome/). This consortium formally elected a Steering Committee of 10 members from 8 countries at the First International Workshop on Lepidopteran Genomics (attended by nearly 160 persons) in Tsukuba, Japan in November, 2002 (http://www.nias.affrc.go.jp/i-genomics/). A listserv for information exchange which now has more than 120 members (Lepgen-L) has also been set-up as a result of the growing interest in networking among Lepidopteran scientists. The International Butterfly Ecology and Evolution Symposium, established in 1982, has recently held its fourth meeting, and The International Workshop on Molecular Biology and Genetics of the Lepidoptera, established in 1988, is scheduled for its sixth meeting this summer, together with the Lepidopteran Genomics Workshop.

***B. anynana*:** A. Monteiro at SUNY Buffalo, A. Long at UC Irvine, D. Stern at Princeton, P. Brakefield & B. Zwaan at Leiden, S. Carroll at UW Madison, I. Saccheri at Liverpool, K. Fisher at Bayreuth, J. Marcus at Western Kentucky (see Appendix with two supporting letters).
***P. glaucus*:** M. Scriber at Michigan, R. ffrench-Constant at Bath, M. Berenbaum at Illinois, P Adolfatto at Toronto, F Sperling, at. Univ. Alberta, J. Bossart at Univ. NJ; F. Marec at Czeck Academy at Branisovska, W. Traut at Lubeck Germany, N. Wahlberg at Univ Stockholm, A. Porter at Univ. Massachusetts, T. Emmel at Univ. Florida.
***H. subflexa*:**:F. Gould at North Carolina, S. Gill at UC Riverside, M. Adang at U. Georgia, D. Heckel at Max Plank Institute, L. Gahan at Clemson University, K. Hopper at USDA, N. Vickers at U of Utah, K. Hoover at Penn State, R. Feyereisen at INRA-Antibes.
O. nubilalis: R. Harrison at Cornell, W. Roelofs and C. Linn at New York Agricultural Experiment Station, D. Heckel at University of Melbourne.

4-Whether the organisms will be, or have been, proposed to NHGRI or another publicly funded agency for BAC-based genomic sequencing and the status of that request.

Three of the four species in the current proposal (*B. anynana, P. glaucus, and H. subflexa*) were submitted together with three others (*Manduca sexta, Heliothis virescens,* and *Heliconius erato*) to an NSF call for proposals for BAC library construction (M. Goldsmith was the PI). The grant received excellent reviews and no major criticism. Despite this, the grant was only funded in part due to a lack of funds, and three of the species were removed in a rather arbitrary manner. According to the NSF program director, Dr. Judith Plesset, similar cuts were made across all proposals. The funded BAC libraries are now being produced by the Laboratory of Drs. Zhang and Wu at Texas A&M University. These researchers have helped pioneer BAC techniques, and have produced four of the five publicly available BAC libraries for *Bombyx mori*. They have recently successfully completed four additional lepidopteran libraries (*M. sexta* and *H. virescens*) and are responsible for primary storage and maintenance of the libraries, and their public distribution.
To our knowledge no requests for BAC library construction have been made for *Ostrinia nubilalis* (to NHGRI or any other agency).

5-Other genomic resources that are available that will complement this resource.

B. anynana **-** A genetic linkage map, including both AFLP and microsatellite markers, is nearing completion (Zwaan, personal communication). A10,000 EST database is presently under construction (P. Beldade and A.D. Long, personal communication) and is due to be completed in 2006. These ESTs, all derived from the pre-adult "wing primordia" where wing patterns are being determined, will serve to: i) to identify novel "wing patterning genes", and ii) identify DNA sequence polymorphisms in these genes, which can be used to map the QTLs responsible for variation in wing pattern in *B. anynana*. A BAC library, because it represents *B. anynana* genomic sequences, will enable us to identify gene regulatory regions for the mapped QTLs that are not in their corresponding cDNAs. The different types of libraries will complement each other and provide the resources necessary to the fully explore the great potential of *B. anynana* for an integrated study of adaptive evolution and development.
P. glaucus **-** Small EST and cDNA collections containing a few thousand sequences have been made from specific tissues such as midgut, pheromone glands, antenna and endocrine organs for *P. glaucus* (Li et al., 2001). A cDNA library of *P. glaucus* yellow and dark morphs) was recently started with P. Andolfatto.
H. subflexa **-** An AFLP linkage map for the cross between *H. virescens* and *H. subflexa* is available. Over 500 informative bands have been scored. A BAC library is available for *H. virescens*. A number of mapping families have already been created and phenotyped for host range and pheromone production. Specimens are being held in a -80 freezer. Funds from a USDA grant and an NSF grant will be used for the genotyping of these specimens. A number of strains of *H. virescens* and *H. subflexa* are being maintained for use in other crosses.
O. nubilalis **-** A genetic linkage map has been constructed based on about 200 AFLPs and 40 microsatellite loci. The backcross progeny are phenotyped for phermone production and response, and additional markers are being added to the map. Genomic sequences (from population samples of both Z and E strain ECB, as well as from Asian Corn Borer (*O. furnacalis*) are available from multiple genes (mtDNA, *Tpi, Ldh, kettin, Pheromone binding protein, _11-desaturase* , etc.).

6-The strain of the organism proposed and rationale for its selection

B. anynana **-** We will supply animals from a wild type strain of this species from the live colony maintained in the Monteiro laboratory, at the University of Buffalo, NY. This colony was established in 2001, from a couple of thousand eggs taken from a laboratory colony kept in Leiden, The Netherlands by Prof. Paul Brakefield, and always maintained at high adult numbers (>400 animals). The original founder animals were 80 gravid females collected in Malawi, Africa, in 1988.
P. glaucus **-** The Scriber lab has several populations of *P. glaucus* available every summer (and spring from Florida collections). These individuals are obtained each year from geographically-variable locations (FL, GA, MI, PA, and sometimes MO,IN,OH,VA, etc.). In addition, the Scriber lab can maintain a culture of these butterflies (hybrids/backcrosses) for an additional 2-3 generations.
H. subflexa **-** We will provide animals from a strain of *H. subflexa* that has already been used in a number of genetic and phenotypic studies at NCSU, University of Utah, Penn State Univ., and the USDA, Delaware.
O. nubilalis - We will provide pupae from a Z strain colony maintained by the Roelofs group in the Department of Entomology at the New York Agricultural Experiment Station in Geneva, NY. These moths have been used in the genetic crosses referred to above, and as one of the parental strains in the mapping data that have been generated in the Harrison lab. This colony has been established from Z borers collected around Geneva, NY and has periodically been "refreshed" with field caught material.

7-The size of the genome

The genome of *B. anynana* has 490 Mb, P. glaucus has 440 Mb (T. R. Gregory, personal communication), and *H. subflexa* has 400 Mb (J.S. Johnston, personal communication). Genome size of *O. nubilalis* is not known but will be characterized prior to BAC library construction. Based on fragment numbers produced in AFLP analyses, we would estimate genome size to be of the order of

500 MB. All four species are at the lower end of the range of the few other lepidopteran genomes for which data are available (0.5-1 Mb; T. R. Gregory, http://www.genomesize.com). The karyotype of *B. anynana* is N=28, (T. R. Gregory, http://www.genomesize.com), *P. glaucus* N=30 (Scriber and Procunier, unpublished), *H. subflexa* N=31 (Robinson, 1971), and *O. nubilalis* N= 31 (Guthrie et al. 1965).

8-The availability of a source of DNA for construction of the BAC library (evidence of its quality for this purpose).

The most critical step in BAC library construction is to obtain high quality DNA from a genetically homogeneous source in order to minimize genetic polymorphisms. Polymorphisms will make eventual sequencing of the BAC library ambiguous. To obtain high quality megabase DNA we have to optimize the DNA preparation procedure. The Wu/Zhang laboratory has considerable experience in making insect BAC libraries (Wu et al., 1999; Zhang, 2000; Wu and Zhang, unpublished). Currently we have successfully optimized a procedure for megabase DNA preparation from several lepidopteran species by testing different buffers and insect developmental stages. As a result, we found that day-10 pupae of *M. sexta*, day-4 pupae of *H. virescens*, and day-6 pupae of *H. erato* are most suited for megabase DNA preparation from these insects, using a buffer system named STE (0.1M NaCl, 10 mM Tris-HCl, 10 mM EDTA, pH 9.4, 0.15% β-mercaptoethanol). We used 50 pupae of *M. sexta* and 200 pupae of the smaller *H. virescens*. DNA fragments isolated with this method are not only large (> 600 kb), but also readily digestible, thus being well-suited for the proposed BAC library constructed. A similar procedure should be also suited for the megabase DNA preparation of the proposed species here.

In order to minimize genetic polymorphism, each lab will provide material as offspring from single pair crosses. Successive generations of brother-sister matings will be employed, to the extent possible, without reducing offspring number as a result of inbreeding depression (Saccheri et al. 1996. 1999).

B. anynana - Live pupae will be provided as needed from the Monteiro and/or Brakefield Lab, at any time during the year. An inbreeding program using single pair matings is currently being established. Single pair crosses can produce up to 250 offspring.

P. glaucus - *P. glaucus* from Levy Co. Florida will be used to assure no introgression from *P. canadensis*, as recently observed in the Great Lakes/New England region of the USA. Sibling pairings will be made in the lab to enhance inbreeding in the next generation. Single pair crosses can produce up to 250 offspring.

H. subflexa - The Gould lab will provide 250 or more pupae from a single pair mating that will be produced for this work.

O. nubilalis - The Harrison Lab will provide, whenever needed, pupae from single pair matings or from inbred lines derived from the laboratory culture of the Z strain. Single females can produce 150-200 offspring.

9-Specifications for the library (e.g., library depth, BAC insert size) and supporting scientific rationale for these specifications

a) Library depth:

In order to achieve complementary overlaps and deep genome coverage, we request that two BAC libraries with an average insert size of at least 150 kb, 30,000~35,000 clones in total, equivalent to 10X genome coverage, will be made for each species by partial digestion of high molecular weight DNA using two enzymes (Bam HI and Eco RI) that recognize different base composition restriction sites.

Almost all large-insert BAC libraries developed to date were generated from partial digests of high molecular weight (HMW) DNA with restriction enzymes. However, the distribution of the sites of a restriction enzyme throughout a genome is uneven. Consequently, the genome regions with a high, or low density of the restriction sites are difficult to clone because small (<40 kb) or large (>350 kb)

DNA fragments generated by partial digestion are removed by PFGE-based size selection during BAC cloning. Therefore, it is desirable to develop complementary large-insert BAC libraries with different individual restriction enzymes having different nucleotide contents in their restriction sites. Such large-insert BAC libraries are similar to shotgun libraries in genome coverage - the cloned DNA fragments are randomly derived from the different regions of a genome, and thus, would have a truly high genome coverage. These have been demonstrated by the successful development of the whole-genome BAC-based physical maps of Arabidopsis (Chang et al. 2001) and rice (Tao et al. 2001) using BAC libraries developed with different enzymes in Drs. Zhang and Wu's laboratory.

In addition, lepidopteran genomes carry a wide variety of transposable elements (Ogura et al., 1994; Eickbush, 1995; Robertson and Asplund, 1996; Prasad et al., 2002) which may comprise 40-50% of the DNA based on reassociation kinetics (Gage, 1974; Efstratiadis et al., 1976). To minimize cloning failure from these kinds of sequences, we propose the construction of two BAC libraries for each species using restriction enzymes with different target base composition. Each library should be produced using either Bam HI or Eco RI which are complementary in restriction site nucleotide composition [Bam HI (66% GC), Eco RI (66% AT)].

b) BAC insert size: Average insert size should be 150 kb or larger. This standard has been achieved for the most recent Lep BAC libraries using pECBAC1 as vector producing a range of insert sizes from 100-380 kb (Figure 1).

c) Percentage of insert-empty clones: should be below <5% (Figure 1).

10-The time frame in which the libraries are needed.

B. anynana **-** Library needed as soon as possible. Library will facilitate ongoing work in the cloning of important wing patterning genes and their cis-regulatory regions. Library will also facilitate the cloning and identification of interrupted genes resulting from a mutagenesis enhancer-trap screen. Transgenic lines are presently being created for this screen and expected to be completed by the end of 2004. As mentioned, QTL mapping experiments have been conducted for a suite of traits. Fine scale mapping and (candidate) gene cloning approaches would hugely benefit immediately from a BAC library.

P. glaucus - The library will facilitate ongoing work with the diagnostic Z-linked Ldh allozyme loci believed to be involved directly, or via close linkage with a locus or loci under very strong selection pressure as on the diapause regulation, creating high frequencies of recombinant interspecific hybrids across the historical hybrid zone. P. Andolfatto has cloned the *P. glaucus* Ldh and would use the library immediately to retrieve the relevant BACs and analyze hundreds of *P. glaucus* recombinants (currently stored in –80 freezers) with interspecifically mixed traits obtained recently from the natural hybrid zone, as well as lab-paired backcrosses.

H. subflexa - Availability of this library within one year would fit into our schedule for completion of fine scale mapping of pheromone and host range genes.

O. nubilalis **-** The library is needed as soon as possible - to be used in coordination with the linkage map being developed in the Harrison lab. Initial focus will be on detailed analysis of the Z chromosome and identification of chromosome regions in which the two pheromone strains are reciprocally monophyletic. Such regions are likely candidates for recent "selective sweeps" and also must represent regions in which introgression is suppressed. It is in these regions that we would initiate our search for genes that contribute to post-diapause development and male pheromone response (both known to be controlled by Z-linked genes).

11-Other support that is available or has been requested for the construction of the desired library.

We have requested the support of T. R. Gregory, who has extensive experience with insect material (see his website at http://www.genomesize.com) to make a precise determination of the genome size of ***O. nubilalis*** .

12- Other relevant information.

We propose that this project be performed at Texas A&M University (TAMU) by Dr. Wu , who helped pioneer BAC techniques, and has made eight of the ten publicly available BAC libraries for Lepidoptera. He would also be responsible for characterizing the libraries for percentage of insert-empty clones, average insert size, and insert size distribution. In addition, Dr. Wu would be responsible for primary storage and maintenance of the libraries, and their public distribution.

Wu's Laboratory. Together with his long time collaborator, Dr. Zhang, Wu has helped pioneer BAC techniques, with more than ten years of research experience in the development of techniques, production of high-quality BAC and Binary BAC (BIBAC) libraries from animal, plant and insect tissues, and library maintenance, archiving and dissemination. The lab's contributions to the field include pioneering the preparation of high-quality megabase DNA from nuclei (Zhang et al. 1994, 1995a, 1996a; Wu et al. 1999; Wu and Zhang, unpublished), which is essential for large-insert BAC cloning, as well as pioneering BAC library making techniques (Wing et al., 1995; Zhang et al. 1995b, 1996a, b; Zhang and Wing 1997; Zhang 1999; Zhang 2000; Zhang and Wu 2001). Using these techniques and their modified procedures Wu and Zhang have constructed more than 80 high quality libraries with average insert sizes of 97-250 kb from a wide variety of microbe, plant, animal and insect species (Zhang 1997, 1998; Scheuring and Zhang 1999;Zhang and Scheuring 2000; McCuine and Zhang 2001; see http://hbz.tamu.edu BAC library-Library List). High-quality BAC libraries have been constructed from the mosquito, fire ant, silkmoth, *H. virescens*, and *M. sexta* (Figure 1). Work is presently underway on a third lepidopteran species, *H. erato*. Utilizing present facilities and equipment, to the lab can produce two million BACs with an average insert size of 150 kb or larger within an academic year.

Because of rapid development and wide applications of the BAC libraries, Wu and Zhang established the Texas A&M BAC Center-A Public Facility for Accelerated Genomics Research in 1996, now renamed as GENE*finder* Genomic Resources (http://hbz.tamu.edu). It is a dedicated operation, with all facilities and instruments (see Resources) required for BAC library production, characterization, long-term maintenance, archiving and dissemination. The GENE*finder* Genomic Resources unit now has two full-time and three half-time research scientists. It is directed by Drs. Zhang and Wu. The major activities of the GENE*finder* are library maintenance, archiving and dissemination.

A systematic mechanism and procedure for library distribution has been established. BAC libraries are available to the public in forms of duplicated libraries, high-density clone filters (8 x 12 cm with 1,536 or 3,072 doubled-spotted clones per filter and 22 cm x 22 cm with 9,216 or 18,432 double-spotted clones per filter), and individual clones on a cost recovery basis. The lab has eight years of experience in BAC library distribution. At the present time, it is estimated that thousands of the laboratories world wide are using BAC libraries in their research supplied by GENE*finder* Genomic Resources. To facilitate access of users to the BAC libraries at GENE*finder*, a web site at http://hbz.tamu.edu was created. Users can readily check, request and use the libraries online. In the past 1-1/2 years more than 6,000 people have visited the BAC Library page of the web site http://hbz.tamu.edu.

Zhang and colleagues have published a manual of BAC library production in three editions (Wing et al. 1995; Zhang 1999; Zhang 2000), organized five international public workshops in BAC and BIBAC technologies, and hosted 34 visiting scientists in the past five years. A total of 116 scientists were trained in BAC technology at these workshops and in the Zhang/Wu laboratory. Additionally, about 300 copies of the third edition of our BAC manual (Zhang 2000) have been requested and distributed to genome scientists world wide.

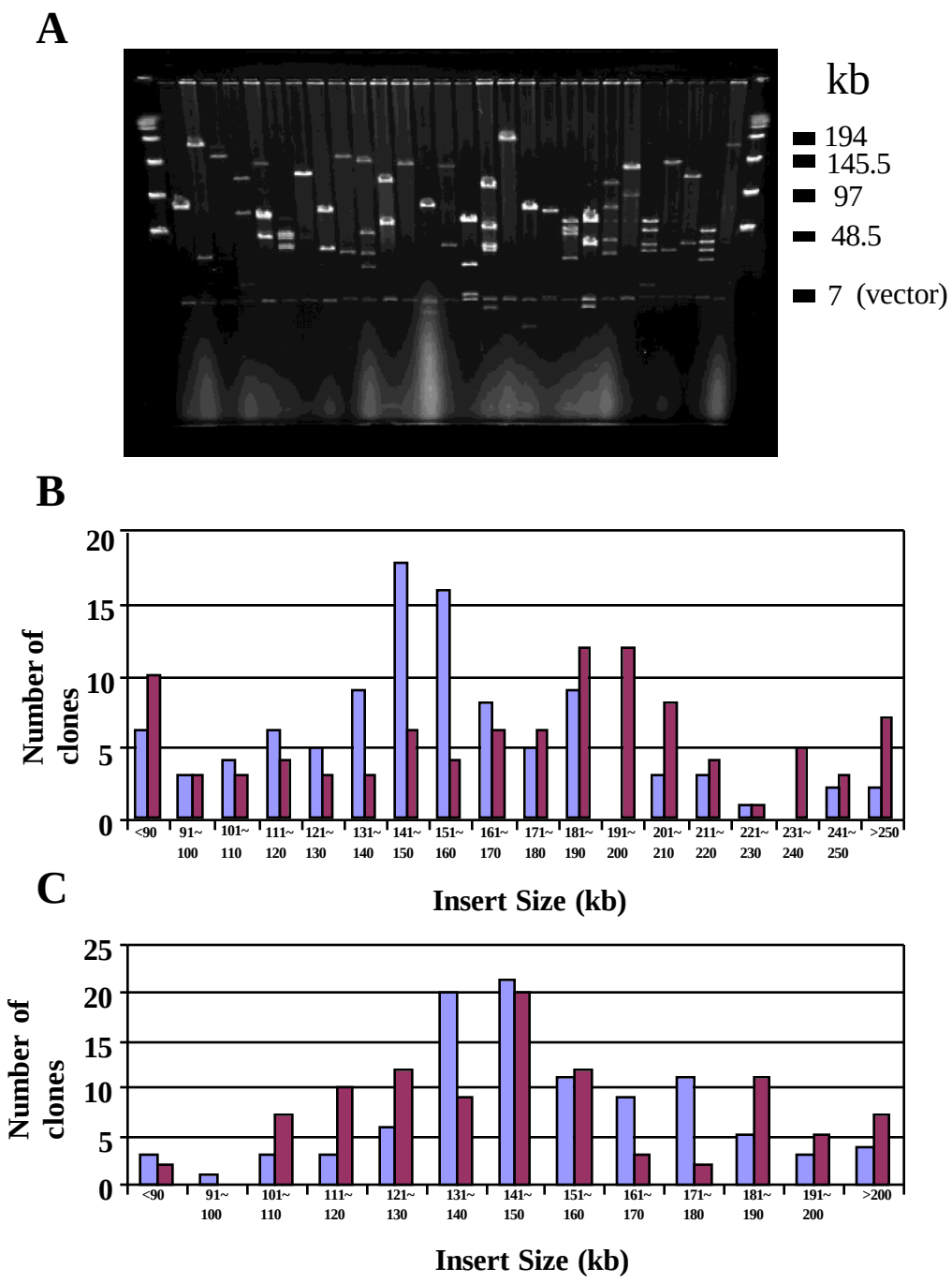

Figure 1. Insert DNA analysis of random BAC clones of the BAC libraries of two lepidopteran species: *Manduca sexta* and *Heliothis virescens*. **A)**. BAC clones randomly selected from the Manduca *Eco* RI BAC library and analysed by pulsed-field gel electrophoresis. BAC DNA was isolated, digested with *Not* I to release the insert DNA, separated on a 1% agarose gel and stained with ethidium bromide. Outer lanes contain the Lambda Ladder PFG Marker (New England BioLabs). The common band (~7.4 kb) is the pECBAC1 cloning vector. **B)** The insert size distributions of 100 BACs randomly selected from both *Manduca Eco* RI (red bars, mean=170 kb) and *Bam* HI (gray bars, mean=152 kb) libraries. C. The insert size distributions of 100 BACs randomly selected from both *Heliothis Eco* RI (red bars, mean=153 kb) and *Bam* HI (gray bars, mean=150 kb) libraries. No empty clone was detected in all randomly selected samples of the four lepidopteran BAC libraries.

References

Andolfatto, P., J.M. Scriber, and B. Charlesworth. 2003. No association between mitochondrial DNA haplotypes and a female-limited mimicry phenotype in *Papilio glaucus*. Evolution 57: 305- 316.

Anglade, P., J. Stockel, and I. W. G. O. Cooperators. 1984. Intraspecific sex-pheromone variability in the European corn borer, *Ostrinia nubilalis* Hbn. (*Lepidoptera, Pyralidae*). Agronomie 4:183-187.

Beldade, P., and P. M. Brakefield. 2002. The genetics and evo devo of butterfly wing patterns. Nat. Rev. Genet. 3:442-452.

Beldade, P., P. M. Brakefield, and A. D. Long. 2002. Contribution of *Distal-less* to quantitative variation in butterfly eyespots. Nature 415:315-317.

Brakefield PM. 1996. Seasonal polyphenism in butterflies and natural selection. TREE 11:275-277.

Brakefield, P. M. 1998. The evolution-development interface and advances with the eyespot patterns of *Bicyclus* butterflies. Heredity 80:265-272.

Brakefield, P. M., and V. French. 1999. Butterfly wings: the evolution of development of colour patterns. BioEssays 21:391-401.

Brakefield, P. M., J. Gates, D. Keys, F. Kesbeke, P. J. Wijngaarden, A. Monteiro, V. French, and S. B. Carrol. 1996. Development, plasticity, and evolution of butterfly eyespot patterns. Nature 384:236-242.

Brunetti, C. R., J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield, and S. B. Carroll. 2001. The generation and diversification of butterfly eyespot color patterns. Current Biology 11:1578-1585.

Carroll, S. B., J. Gates, D. N. Keys, S. W. Paddock, G. E. F. Panganiban, J. E. Seleque, and J. A. Williams. 1994. Pattern formation and eyespot determination in butterfly wings. Science 265:109-114.

Chang Y-L, Tao Q, Scheuring C, Ding K, Meksem K, Zhang HB 2001. An integrated map of Arabidopsis thaliana for functional analysis of its genome sequence. Genetics 159:1231-42

Clarke, C.A. and Clarke F.M.M. 1983. System. Entomol. 8: 25.

Clarke, C.A. and Sheppard P.M. 1959. Linn. J. Genet. 56: 237.

Clarke, C.A. and Sheppard P.M. 1962. Ecology. 43: 159.

Deering, M.D. and J.M. Scriber. 2002. Field bioassays show heterospecific mating preference asymmetry between hybridizing North American *Papilio* butterfly species (Lepidoptera: Papilionidae). J. Ethology 20: 25-33.

Dernburg, A. F. 2001. Here, there, and cverywhere: Kinetochore function on holocentric chromosomes. J. Cell Biol. 153: F33–F38.

Dulac, C. 1997. Molecular biology of pheromone perception in mammals. Seminars in Cell and Developmental Biology. 8:197-205.

Efstratiadis, A., Crain W.R., Britten R.J., Davidson E.H., Kafatos F.C. 1976. DNA sequence organization in the lepidopteran Antheraea pernyi. Proc. Natl. Acad. Sci. USA 73: 2289–93.

Eickbush, T. 1995. Mobile elements of lepidopteran genomes. In Molecular Model Systems in the Lepidoptera, M.R. Goldsmith, A.S. Wilkins, ed's; NY: Cambridge U. Press, pp. 77–105.

Friedlander, T.P., J.C. Regier, C. Mitter, D.L. Wagner and Q.Q. Fang. 2000. Evolution of heteroneuran Lepidoptera (Insecta) and the utility of dopadecarboxylase for Cretaceous age phylogenetics. Zoological Journal of the Linnean Society, 130: 213-234.

Gage L.P. 1974. The *Bombyx mori* genome: analysis by reassociation kinetics. Chromosoma 45:27–42.

Gahan, L.J., Gould, F. and Heckel, D.G. 2001. Identification of a gene associated with Bt resistance in Heliothis virescens. Science 293: 857-60.

Galant, R., J. B. Skeath, S. Paddock, D. L. Lewis, and S. B. Carroll. 1998. Expression pattern of a butterfly *achaete-scute* homolog reveals the homology of butterfly wing scales and insect sensory bristles. Curr. Biol. 8:807-813.

Glover, T., M. Campbell, P. Robbins, and W. Roelofs. 1990. Sex-linked control of sex pheromone behavioral responses in European corn-borer moths (*Ostrinia nubilalis*) confirmed with TPI marker gene. Archives Insect Biochem. Phys. 15:67-77.

Glover, T. J., M. G. Campbell, C. E. Linn, Jr., and W. L. Roelofs. 1991a. Unique sex chromosome mediated behavioral response specificity of hybrid male European corn borer moths. Experientia 47:980-984.

Goldsmith M.R. 1995. Genetics of the silkworm: revisiting an ancient model system. In Molecular Model Systems in the Lepidoptera, M.R. Goldsmith, A.S. Wilkins, ed's; NY: Cambridge U. Press, pp. 76.

Gould F, Carroll C. R, Futuyma, D.J. 1982. Cross-resistance to pesticides and plant defenses - a study of the 2-spotted spider mite.. Entomol. Exp. Appl. 31 (2): 175-180

Guiliano D. B., N. Hall, S. J. Jones, L. N. Clark, C. H. Corton, B. G. Barrell, B& M. L. Blaxter. 2002. Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. Genome Biol 2002 Sep 26;3(10):RESEARCH0057

Guthrie, W. D., E. J. Dollinger, and J. F. Stetson. 1965. Chromosome studies of the European corn borer, smartweed borer, and lotus borer (Pyralidae). Ann. Entomol. Soc. Am. 58(1):100–105.

Hagen, R.C. and J.M. Scriber, 1989. Sex linked diapause, color, and allozyme loci in *Papilio glaucus*: Linkage analysis and significance in a hybrid zone. Heredity 80: 179-185.

Hagen, R. and J.M. Scriber. 1995. Sex chromosomes and speciation in the *Papilio glaucus* group. Chapter 22 In (J.M. Scriber, Y. Tsubaki, and R.C. Lederhouse, eds.) The Swallowtail Butterflies: Their Ecology and Evolutionary Biology. Scientific Publishers, Inc., Gainesville, FL. (pp. 211-228).

Hansson, B. S. and C. Löfstedt. 1987. Inheritance of olfactory response to sex pheromone components in *Ostrinia nubilalis*. Naturwissenschaften 74:497-499.

Heckel, D.G., Gahan LJ, Liu YB, Tabashnik BE. 1999. Genetic mapping of resistance to Bacillus thuringiensis toxins in diamondback moth using biphasic linkage analysis. Proc Natl Acad Sci U S A 96 :8373-7.

Hill, C.A., Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ. 2002 G protein coupled receptors in *Anopheles gambiae*. Science 298 (5591): 176-178

Hung, C.F., Berenbaum M.R., Schuler M.A. 1997. Isolation and characterization of CYP6B4, a furanocoumarin-inducible cytochrome P450 from a Polyphagous caterpillar (Lepidoptera: Papilionidae). Insect Biochem. Molec. 27 (5): 377-385

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302-305.

Jurat-Fuentes, J. L., F. L. Gould, et al. 2000. High levels of resistance and cross-resistance to *Bacillus thuringiensis* Cry1 toxins in *Heliothis virescens* are due to reduced toxin binding and pore formation. Resistant Pest Management 11(1): 23-24.

Kapan, D. D. 2001. Three-butterfly system provides a field test of müllerian mimicry. Nature 409:338-340.

Keys, D. N., D. L. Lewis, J. E. Selegue, B. J. Pearson, L. V. Goodrich, R. J. Johnson, J. Gates, M. P. Scott, and S. B. Carroll. 1999. Recruitment of a *hedgehog* regulatory circuit in butterfly eyespot evolution. Science 283:532-534.

Klun, J. A. and Cooperators. 1975. Insect sex pheromones: Intraspecific pheromonal variability of *Ostrinia nubilalis* in North America and Europe. Environ. Entomol. 4:891-894.

Klun, J. A. and S. Maini. 1979. Genetic basis of an insect chemical communication system: The European corn borer. Environ. Entomol. 8:423-426.

Koch, P. B., B. Behnecke, and R. H. ffrench-Constant. 2000a. The molecular basis of melanism and mimicry in a swallowtail butterfly. Current Biology 10:591-594.

Koch, P. B., B. Behnecke, M. Wigmann-Lenz, and R. ffrench-Constant. 2000b. Insect pigmentation: activities of Beta-alanyldopamine syntase in wing color patterns of wild-type and melanic mutant swallowtail butterfly Papilio glaucus. Pigment Cell Research 13 (Supplement 8):54-58.

Koch, P. B., D. N. Keys, T. Rocheleau, K. Aronstein, M. Blackburn, S. B. Carroll, and R. H. ffrench-Constant. 1998. Regulation of dopa decarboxylase expression during colour pattern formation in wild-type and melanic tiger swallowtail butterflies. Development 125:2303-2313.

Kristensen, N.P. and A.W. Skalski. 1999. Phylogeny and palaeontology. In N.P. Kristensen (Ed.), Lepidoptera, Moths and Butterflies, Volume 1: Evolution, Systematics, and Biogeography, Chapter 2: pp. 7-25. Hawthorne, NY: Walter de Gruyter, Inc.

Li W., Berenbaum M.R., Schuler, M.A. 2001. Molecular analysis of multiple CYP6B genes from polyphagous *Papilio* species. Insect Molec. Biol. Biochem. 31: 999–1011.

Li , W., Petersen , R. A., Schuler, M. A., & Berenbaum, M. R. 2002. CYP6B cytochrome p450 monooxygenases from *Papilio canadensis* and *Papilio glaucus*: Potential contributions of sequence divergence to host plant associations. Insect Mol. Biol. 11: 543–51.

Liu, W., H. Jiao, N. C. Murray, M. O'Connor, and W. L. Roelofs. 2002. Gene characterized for membrane desaturases that produces (E)-11 isomers of mono- and diunsaturated fatty acids. Proc. Natl. Acad. Sci. USA 99:620-624.

Löfstedt, C. 1993. Moth pheromone genetics and evolution. Phil. Trans. R. Soc. Lond. B 340:167-177.

Marec, F., A. Tothova, K. Sahara, and W. Traut. 2001. Meiotic pairing of sex-chromosome fragments and its relation to atypical transmission of a sex-linked marker in *Ephestia kuehniella* (Insecta: Lepidoptera). Heredity 87: 659-671

McCuine S. A. and H. B. Zhang. 2001. Seventy large-insert, ordered and high-quality BAC and BIBAC libraries for accelerated plant and animal genomics research. Plant & Animal Genome IX. P78

McMillan, W. O., A. Monteiro, and D. D. Kapan. 2002. Development and evolution on the wing. TREE 17:125-133.

Monteiro, A., J. Prijs, M. Bax, T. Hakkaart, and P. M. Brakefield. 2003. Mutants highlight the modular control of butterfly eyespot patterns. Evol. Dev. 5:180-187.

Nijhout, H. F. 1991. The development and evolution of butterfly wing patterns. Smithsonian Institution Press, Washington.

Ogura T., Okano K., Tsuchida K., Miyajima N., Tanaka H., Takada N., Izumi S., Tomino S., Maekawa H. 1994. A defective non-LTR retrotransposon is dispersed throughout the genome of the silkworm, *Bombyx mori*. Chromosoma 103: 311–323.

Prasad, M. D., D. L. Nurminsky, & J. Nagaraju. 2002. Characterization and molecular phylogenetic analysis of mariner elements from wild and domesticated species of silkmoths. Mol. Phylogenet. Evol. 25: 210–17.

Raijmann, L. E. L., W. E. vanGinkel, D. G. Heckel & S. B. J. Menken. 1997. Inheritance and linkage of isozymes in *Yponomeuta padellus* (Lepidoptera; Yponomeutidae). Heredity 78: 645–54.

Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R 2002. Evolution of supergene families associated with insecticide resistance. Science 298 (5591): 179-181

Ranz, J. M., F. Casals & A. Ruiz. 1991. How malleable is the eukaryotic genome? Extreme fate of chromosomal rearrangement in the genus *Drosophila.* Genome Research 11:230–239.

Ranz, J. M., F. Casals, & A. Ruiz. 2001. How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the Genus *Drosophila*. Genome Res. 11: 230–239.

Robertson H.M., Asplund M.L. 1996. Bmmar1: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem Mol Biol. 26:945–54

Robinson R. 1971. Lepidoptera Genetics. Oxford: Pergamon Press Ltd.

Roelofs, W. L. 1995. Chemistry of sex attraction. Proc. Natl. Acad. Sci. USA 92:44-49.

Roelofs, W., T. Glover, X.-H. Tang, I. Sreng, P. Robbins, C. Eckenrode, C. Löfstedt, B. S. Hansson, and B. O. Bengtsson. 1987. Sex pheromone production and perception in European corn borer moths is determined by both autosomal and sex-linked genes. Proc. Natl. Acad. Sci. 84:7585-7589.

Roelofs, W. L., W. Lie, G. Hao, H. Jiao, A. P. Rooney, and C. E. Linn 2002. Evolution of moth pheromones via ancestral genes. Proc. Natl. Acad. Sci. USA 99:13621-13626.

Roelofs, W. L. and W. A. Wolf. 1988. Pheromone biosynthesis in Lepidoptera. J. of Chemical Ecology. 14:2019-2031.

Roethele, J. B., J.Romero-Severson and J. Feder. 2001. Evidence for broad-scale conservation of linkage map relationships between *Rhagoletis pomonella* (Diptera: Tephritidae) and *Drosophila melanogaster* (Diptera: Drosophilidae). Ann. Entomol. Soc. Am. 94: 936–947.

Rosenfield, C.-L., K. M. You, P. Marsella-Herrick, W. L. Roelofs, and D. C. Knipple. 2001. Structural and functional conservation and divergence among acyl-CoA desaturases of two noctuid species, the corn earworm, *Helicoverpa zea*, and the cabbage looper, *Tricoplusia ni*. Insect Biochem. and Mol. Biol. 31:949-964.

Saccheri, I. J., P. M. Brakefield, and R. A. Nichols. 1996. Severe inbreeding depression and rapid fitness rebound in the butterfly *Bicyclus anynana* (Satyridae). Evolution 50:2000-2013.

Saccheri, I. J., I. J. Wilson, R. A. Nichols, M. W. Bruford, and P. M. Brakefield. 1999. Inbreeding of bottlenecked butterfly populations: estimation using the likelihood of changes in marker allele frequencies. Genetics 151:1053-1063.

Scheuring, C. and H. B. Zhang. 1999. The Texas A&M BAC Center - A Public Facility for Accelerated Plant Genomics Research. Plant & Animal Genome VII. P114

Scriber, J.M., R.H. Hagen and R.C. Lederhouse. 1996. Genetics of mimicry in the tiger swallowtail butterflies, *Papilio glaucus* and *P. canadensis*. (Lepidoptera: Papilionidae). Evolution 50: 222-236.

Scriber, J. M. 2002a. The evolution of insect-plant relationships; Chemical constraints, coadaptation and concordance of insect/plant traits. (from keynote address). Entomologia Expt. Appl. 104: 217-235

Scriber, J. M. 2002b. Latitudinal and local geographic mosaics in host plant preferences as shaped by thermal units and voltinism in *Papilio* spp. (Lepidoptera) .
European J. Entomol. 99: 225-239.

Scriber, J.M., A. Stump and M. Deering. 2003. Hybrid zone ecology and tiger swallowtail trait clines in North America. Pp 367-391 In (C. Boggs, W. Watt and P. Ehrlich, eds.) . Ecology and evolution taking flight: Butterflies as model study systems. University of Chicago Press, Chicago, IL.

Scriber, J.M., Y. Tsubaki, and R.C. Lederhouse (eds). 1995. Swallowtail Butterflies: Their Ecology and Evolutionary Biology. Scientific Publ., Gainesville, FL (35 Chapters, 460 pages with color plates).

Sharakhov, I V., A. C. Serazin, O. G. Grushko, A. Dana, N. Lobo, M. E. Hillenmeyer, R. Westerman, J. Romero-Severson, C. Costantini, N'F. Sagnon, F. H. Collins, and N. J. Besansky. 2002. Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus.* Science 298: 182–185.

Sheck A.L., Gould F. 1996. the genetic basis of differences in growth and behavior of specialist and generalist herbivore species: selection on hybrids *of Heliothis virescens* and *Heliothis subflexa* (Lepidoptera). Evolution 50: 831–841.

Sperling, F.A.H.. 1994. Sex-linked genes and species differences in Lepidoptera. Can. Entomol. 126: 807-818.

Stowers, L., T. E. Holy, M. Meister, C. Dulac, and G. Koentges. 2002. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. Science 295:1493-1500.

Suzuki M.G., Ohbayashi F., Mita K., Shimada T. 2001. The mechanism of sex-specific splicing at the doublesex gene is different between *Drosophila melanogaster* and *Bombyx mori.* Insect Biochem and Mol Biol. 31: 1202-1211.

Takahashi, A., S-C. Tsaur, J. A. Coyne and C.-I. Wu. 2001. The nucelotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 98:3920-3925.

Tao Q, Y. L. Chang, J. Wang, H. Chen, M. N. Islam-Faridi, C. Scheuring, B. Wang, D. M. Stelly, and H. B. Zhang. 2001. Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158:1711-24

Wagner, D.L. 2001. Moths. In S.A. Levin (Ed.), Encyclopedia of Biodiversity, Volume 4: pp. 249-270. San Diego, CA: Academic Press.

Weatherbee, S. D., G. Halder, J. Kim, A. Hudson, and S. B. Carroll. 1998. Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the *Drosophila* haltere. Genes Dev. 12:1474-1482.

Weatherbee, S. D., H. F. Nijhout, L. W. Grunert, G. Halder, R. Galant, J. Selegue, and S. Carroll. 1999. Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. Current Biology 9:109-115.

Whalley, P.E.S. 1986. A review of the current fossil evidence of Lepidoptera in the Mesozoic. Biol. J. Linn. Soc., 28: 253-271.

Willett, C. S. and R. G. Harrison. 1999. Pheromone binding proteins in the European and Asian corn borers: No protein change associated with pheromone differences. Insect Biochem. Mol. Biol. 29:277-284.

Wing R. A., H-B. Zhang, and S-S. Woo. 1995. BAC Workshop Manual. Southern Crop Improvement Facility, Texas A&M University, College Station, Texas.

Wolf, K. W., K. Novak & K. Marec. 1997. Kinetic organization of metaphase I bivalents in spermatogenesis of Lepidoptera and Trichoptera species with small chromosome numbers. Heredity 97:135–143.

Wu, C., S. Asakawa., N. Shimizu., S. Kawasaki, and Y. Yasukochi. 1999. Construction and characterization of bacterial artificial chromosome libraries from the silkworm, *Bombyx mori*. Mol. Gen. Genet. 261, 698-706

Wu C., Z. Xu, H-B. Zhang.2003. DNA Libraries. In: Meyers RA (ed) Encyclopedia of Molecular Cell Biology and Molecular Medicine. Wiley-VCH Verlag GmbH, Weinheim, Germany (in press)

Zdobnov E. M, C. von Mering, I. Letunic, D. Torrent, M. Suyama, R. R. Copley et al. 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298:149–59.

Zhang H-B 1997. Toward development of a BAC Center for genome research. Plant & Animal Genome V. P36.

Zhang H-B. 1998. The Texas A&M BAC Center: Ten new BAC libraries constructed in 1997. Plant & Animal Genome VI. P15

Zhang H-B. 1999. BAC Cloning and Library Construction Workshop – Manual. CSIRO Tropical Agriculture, Brisbane, Australia.

Zhang H-B. 2000. Construction and Manipulation of Large-insert Bacterial Clone Libraries – Manual. Texas A&M University, Texas, USA.

Zhang H-B and Wing RA. 1997. Physical Mapping of the rice genome with BACs. Pl. Mol. Biol.35:115-127.

Zhang H-B, X. Zhao, R. A. Wing, and A. H. Paterson. 1994. A simple method for cotton megabase DNA isolation. Plant Mol. Biol. Rep. 12:126-131.

Zhang H-B, X. Zhao, X. Ding, A. H. Paterson, and R. A.Wing. 1995a. Preparation of megabase-size DNA from plant nuclei. Plant J. 7:175-184.

Zhang H-B, S. Choi, S-S. Woo, Z. Li, and R. A. Wing. 1995b. Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Plant Genome III. P276.

Zhang H-B, S-S. Woo, and R. A. Wing. 1996a. BAC, YAC and Cosmid Library Construction, in: Foster G., Twell D., (Eds.), Plant Gene Isolation: Principles and Practice: John Wiley & Sons, Ltd., England. pp. 75-99.

Zhang H-B, S. Choi, S-S. Woo, Z. Li, and R. A. Wing. 1996b. Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Molecular Breed. 2:11-24.

Zhang H-B and C. Scheuring. 2000. Fifty large-insert, ordered and high-quality BAC and BIBAC libraries for accelerated plant and animal genomics research. Plant & Animal Genome VIII. P112.

Zhang H-B and C. C. Wu. 2001. BACs as tools for genome sequencing. Plant Physiology and Biochemistry 39:195-209.

Zhao, Y; Y. Park,and M. E. Adams. 2000. Functional and evolutionary consequences of pyrethroid resistance mutations in S6 transmembrane segments of a voltage-gated sodium channel. Biochemical And Biophysical Research Communications 278:516-521.