Insect genomic resources: status, availability and future

Poonam Chilana, Anu Sharma* and Anil Rai

Recent advances in biotechnology have led to the development and evolution in the field of bioinformatics for the analysis and integration of information from genomic, transcriptomic, proteomic, metabolomic and phenomic data. Availability of whole genome sequences, expressed sequence tags, genetic linkage maps and insect transgenesis has opened up new vistas for fundamental research in entomology. This article describes the applications of bioinformatics in applied insect science and pest management. Details of insect genomes sequenced and published, and the available genomic databases along with their features are also discussed. Considering current developments in insect biotechnology and available insect genomic resources, future aspects of bioinformatics applications in insect science have been highlighted.

Keywords: Bioinformatics, bio-rational control, databases, insect genomes, pest management.

**Drosophila melanogaster** serves as a model system for animal and insect genetics. *D. melanogaster* is a species that is extensively studied to understand any particular biological phenomenon because it has characteristics that make it suitable for availability and traceability. A large amount of information is available from *Drosophila* that provides valuable data for the analysis of gene regulation, genetic diseases and evolutionary processes. *Drosophila* research has provided insights into genetics, behaviour, development and disease systems. The *D. melanogaster* genome sequencing project was essentially completed in March 2000. *Drosophila* genome encodes approximately 15,016 genes, fewer than the smaller *Caenorhabditis elegans* genome, but with comparable functional diversity.

Advances in sequencing technologies have provided opportunities in bioinformatics for managing, processing and analysing the sequences. In this genomic era, bioinformatics is used as a bedrock of current and future biotechnology for finding new or better alternatives as designing potential target sites, safer insecticides and developing transgenic insects in applied insect science. The objective of this article is to provide comprehensive information on available insect genomic resources at one place to biotechnologists, molecular biologists, entomologists and physiologists for developing new methods in pest and disease management.

**Sequencing of insect genomes**

Rapid developments in genome sequencing are transforming insect biology with new avenues in the area of insect science. The genome is the set of all genetic...
<table>
<thead>
<tr>
<th>Insect name</th>
<th>Scientific name</th>
<th>Chromosome no</th>
<th>Common name</th>
<th>Scientific name</th>
<th>Chromosome no</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosophila melanogaster</td>
<td>D. melanogaster</td>
<td>4</td>
<td>Fruit Fly</td>
<td>Drosophila simulans</td>
<td>6</td>
</tr>
<tr>
<td>D. pseudoobscura</td>
<td>D. pseudoobscura</td>
<td>6</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. erecta</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. erecta</td>
<td>D. erecta</td>
<td>4</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>D. melanogaster</td>
<td>2</td>
<td></td>
<td>D. melanogaster</td>
<td>4</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>D. yakuba</td>
<td>6</td>
<td></td>
<td>D. yakuba</td>
<td>6</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>D. willistoni</td>
<td>3</td>
<td></td>
<td>D. virilis</td>
<td>4</td>
</tr>
</tbody>
</table>
information of an organism encoded in the deoxyribose
nucleic acid (DNA) of the nucleus and organelles. The
genome contains regions that code for specific proteins,
especially genes and non-coding regions, including some with
structural and regulatory functions. In recent years, the
DNA sequencing boom has made it economically feasible
to obtain the entire sequence of an organism’s genome,
and has opened the door for the establishment of many
publicly or privately funded insect genome projects. In
this section, we describe the status of insect genomes se-
quenced so far in chronological order. A total of 39 insect
genomes have been sequenced to date and numerous
other insect species are in the process of being sequenced.
All the insect genomes which have been sequenced and
published are presented in Table 1.

Drosophila sequencing status

The first insect with published genome is D. melanogaster,
commonly called fruit fly, the most-studied eukaryotic
genome and the most-prominent model organism in mole-
cular biology. The genome of D. melanogaster was
sequenced using a whole genome sequence (WGS) ap-
proach. The first assembly (WGS1) used only plasmid
and bacterial artificial chromosome (BAC) paired-end
sequences, and the second added BAC and P1-based fin-
ished and draft sequences, and then the joint assembly
was submitted to GenBank as Release 1. This sequence
contained many gaps and regions of low sequence qual-
ity. A second release, Release 2, corrected some errors in
the order and orientation of small scaffolds present in Re-
lease 1, and filled a few hundred very small sequence
gaps. Using improved WGS sequence-assembly algo-
rithms, two additional assemblies of the WGS plasmid
and BAC paired-end sequences used in WGS1 were gen-
erated in March 2001 (WGS2) and July 2002 (WGS3),
roughly coinciding with the WGS assemblies of the hu-
man and mouse genomes respectively. Release 3 was
generated to improve Release 2 by closing all the gaps,
improving regions of low sequence quality, and extend-
ing the sequence at the telomeric and centromeric ends of
each chromosome. The Release 3 euchromatic genome
sequence has been reannotated using a new annotation
tool, Apollo, and the complete reannotated improved ge-
nomic sequence of Drosophila with new expressed se-
quence tags (ESTs) and complementary DNA (cDNA)
was then deposited in GenBank.

The improvements made to the genomic sequence in
Release 3 had a large impact on the annotation of transpos-
able elements because of the substantial corrections made
in the assembly of repeated sequences. Release 3 pro-
vided a euchromatic sequence of good quality, gap-free
and of high accuracy. Release 4 and Release 5 are now
available with Flybase and GenBank. They are consid-
ered to be of sufficient accuracy and declared to be sub-
stantially complete and support an initial analysis of
genome structure and preliminary gene annotation and in-
terpretation.

A physical map of a chromosome or a genome shows
the physical locations of genes and other DNA sequences
of interest. The physical map provided a benchmark for
evaluating the accuracy of WGS assemblies. In D. melanogaster, there are five chromosomes (X, 2, 3, 4 and
Y). BAC-based physical maps of chromosomes 2 and 3 of
D. melanogaster constitute 81% of the genome. Sequence
tagged site (STS) content, restriction fingerprinting, and
polypolytene chromosome in situ hybridization approaches
were integrated to produce a map spanning the euchro-
matin. Major computational challenge in the construc-
tion of physical maps is to track the multiple names of
markers. Similarly, multiplicity of names of a single
genome, its associativity with one or more EST clusters,
polymorphisms and STSs are complexities in the con-
struction of physical maps. Further, lack of precise
knowledge of markers, genes and genomic elements
makes mapping more difficult.

Whole genomes of 15 other species of Drosophila
(ananassae, erecta, elegans, grimshawi, mojavensis, per-
similis, pseudoobscura pseudoobscura, sechellia, simu-
lans, virilis, willistoni, takahashii, ficusphiila, kikkawai
and yakuba) have been sequenced and are accessible for
comparative genomics through the internet (http://
insects.eugenens.org/DroSpeGe/). After the sequencing of
Drosophila, much effort was focused on economically
important insects of three categories: agriculturally
important pests (flour beetle, aphid), beneficial insects
(honey bee, parasitic wasp and silkworm) and vectors of
plant and animal diseases (mosquitoes, body louse,
blood-sucking bug, aphids, etc.). Mosquitoes are the sec-
ond most studied insects after fruit flies, as these are vec-
tors of some of the deadliest human diseases. Publication
of the Plasmodium falciparum and Anopheles gambiae
genome sequences has once again paved the way to find a
permanent solution to the malaria problem. The se-
quencing of Culex pipiens, the mosquito vector for the
West Nile virus, shed further light on mosquito biology
and mosquito species-specific gene functions. Next to
mosquito genome, an important lepidopteran silkworm
has joined this group of fully sequenced insects. The
silkworm, Bombyx mori, serves as a central model organ-
ism for the lepidoptera genomics and facilitates studies of
comparative genomics and basic research leading toward
new genome-based approaches for sericulture and pest
control. Within Hymenoptera, an international genom-
ics effort has sequenced the honey bee (Apis mellifera),
an economically important member of this group. The
genome sequences of other medically important body
louse, Pediculus humanus (relapsing fever, trench fever
and epidemic typhus) and Rhodinus prolirus (Chagas dis-
bases) are also available. The genomes of two important
agricultural pests, red flour beetle (Tribolium castaneum)
Many more projects have also developed plans to sequence several key pests and beneficial insect species. In the near future, many more insect genomes will be sequenced as organizations involved with human health and agriculture have also developed plans to sequence several key pests and beneficial insect species.

Gene annotation in insects

Gene annotation is a process during which biological information is attached to the sequence, for example, positions of protein-coding genes, their coding regions and their regulatory elements along with the putative function of each gene. Computer algorithms are used for the identification of structural elements within the genome. Gene annotation helps in comparing the insect genomes with one another and with those of other organisms. These results are helpful in understanding the evolution of insects, the phylogenetic relationships among different orders and the molecular underpinnings of insect-specific processes. The insect genomes sequenced to date differ considerably in terms of size and gene number. The *Drosophila* sequence has been extensively annotated and a wealth of information is available about genomic organization, development, cell biology, neurobiology, behaviour and its evolution. Comparing this with human sequences suggest that the *Drosophila* coding genome is more similar to the human genome than those of yeast and nematode. The annotation summary of *D. melanogaster* is presented in Table 2. The honey bee genome appears to have evolved less rapidly than those of the fruit fly and mosquito, and it displays less similarity, for certain groups of genes, with the latter two insect genomes than with those of vertebrates.

Direct and indirect approaches have been used individually or in combination to identify the functions of insect genes. Most indirect approaches involve the quantification of gene expression at the level of either transcript mRNA or of proteins for finding clues about the role played by a given gene product. In developmental studies, a gene that is expressed only at the onset of metamorphosis such as broad complex may be hypothesized to play a role in that process. Similarly, a gene that is expressed in only one gland or tissue, such as juvenile hormone acid methyl transferase, is likely to have a function that is restricted to the metabolism of that gland or tissue. Various other gene arrays have been generated for insects, including *A. mellifera*, *B. mori*, *Spodoptera frugiperda* and several other gene arrays have been used in these indirect approaches involving the quantification of gene expression at the level of either transcript mRNA or of proteins for finding clues about the role played by a given gene product. In developmental studies, a gene that is expressed only at the onset of metamorphosis such as broad complex may be hypothesized to play a role in that process. Similarly, a gene that is expressed in only one gland or tissue, such as juvenile hormone acid methyl transferase, is likely to have a function that is restricted to the metabolism of that gland or tissue. Various other gene arrays have been generated for insects, including *A. mellifera*, *B. mori*, *Spodoptera frugiperda* and *Choristoneura fumiferana*, typically from cDNAs, and are used to study the differences between bee larvae raised as workers or as queens, the changes in gene expression during silkworm metamorphosis, and the modulation of gene expression following infection with a polydnavirus or with wild type recombinant baculoviruses.

A more direct approach to the study of gene function involves the disruption of individual genes. The genomes of some microorganisms, including viruses, bacteria and yeast, can be manipulated with relative ease to generate mutants that display single-gene defects whose phenotypes allow researchers to infer the function of the disrupted gene. The application of a similar approach to more complex organisms such as insects, however, presents a greater challenge. A detailed comparison between gene content in six insect species (*A. mellifera*, *D. melanogaster*, *A. gambiae*, *T. castaneum*, *B. mori* and the migratory locust, *L. migratoria*) and in three non-insect eukaryotic organisms (yeast, nematode and human) led to the finding that the best represented insect-specific proteins are those associated with stress and stimulus response with cuticle formation and with pheromone or odour perception. Odorant receptors are well represented in the mosquito for host seeking, whereas these type of proteins are altogether absent in the fruit fly. Pheromone or odour receptor proteins are even more
abundant in the honey bee. These proteins are hypothesized to mediate the insect’s perception of pheromone blends, kin recognition signals, and diverse floral odours. The bee genome contains novel genes associated with nectar and pollen utilization. Other genes associated with innate immunity and with the detoxification of xenobiotics are less represented in the honey bee than in the two dipterans (D. melanogaster and A. gambiae). The silk-worm genome contains an estimated 1793 genes for silk production, immunity, development and pheromone production, which were not found in the fruit fly or mosquito. This comparative genomics approach not only helps understand the basic processes involved in insect biology but also leads to the identification of many insect or pest-specific proteins that may be potential targets for insecticide development.

**EST resources in insects**

EST sequencing represents an efficient alternative to whole genome sequencing by providing information of the most expressed parts of the genes at a lower cost. It is also called gene signature, which helps in cloning and characterization of full-length genes. Researchers are putting efforts to construct EST libraries for insect species, which contain collections of cDNA sequences derived from expressed genes only. With the development of ESTs in several insect species, a lot of DNA sequence information has been produced across species and deposited in on-line databases. In the NCBI EST database (dbEST; www.ncbi.nlm.nih.gov/dbEST/), there are up to 214,834, 309,472, 4,448 and 821,005 ESTs available from crop pests, beneficial insects, disease-causing pathogens and Drosophila species respectively, as shown in Figure 1. EST projects on S. frugiperda and C. fumiferana (http://pestgenomics.org) strongly focus on important lepidopteran pests of agriculture and forestry. Many more EST projects are underway in various laboratories.

### Insect genome databases

Traditionally entomologist relies on textbooks and research articles as major resources for ‘omics’ information on insects. With the advancements in molecular biology and genomics technologies in the insect domain, lot of genomic data have been generated in the past decade. This deluge of genomic information has led to an absolute requirement for computerized database to store, organize and index the data along with development of specialized tools to view and analyse the data. Genomic approaches are now becoming an important step for new developments in the biology and pathology of insects. The insect genomic databases contain information of all proteins, biochemical and physiological processes that occur in an insect. A detailed review of databases developed on important insects has been presented in Table 3. Recently, efforts have begun to develop a comprehensive sequence database named Agricultural Pest Genomics Resources (Agripestbase) for storing genomic information from a broad range of pests, including insects, parasites and pathogens. Availability of genomic information on a broad range of agricultural pests will result in comparative genomics and further understanding of these species (http://agripestbase.org/). The information available from these databases will hopefully lead to better management strategies as well as new methods and targets for pest control.

### Bioinformatics in applied insect science and pest management

Bioinformatics is applied in pest management in different ways, which are as follows:
Insect transgenesis

Biotechnology is providing modern improvements and a range of new tools for population control of insects in crop protection. Genetic transformation of insects is another technique that will greatly affect the future role of genomics in applied entomology. Transformations may be used for gene identification and characterization and for creating strains with genes encoding lethality or sterility. Transgenic strains may be created to improve existing biocontrol programmes such as sterile insect management technique, or potentially allow new, highly efficient control strategies. The improved understanding of genome sequencing of the pest insect will stimulate the design of new classes of transgene microorganisms to be used in pest control. Baculoviruses and Bacillus thuringensis, for example, have a narrow host range, and can selectively kill certain lepidopteran larvae, but only do so at very late larval stages (when the animals have already consumed the crop). These microorganisms can be improved by introducing DNA into their genomes coding for lepidopteran proteins that stop larval phytophagy at a very early stage. These larval proteins could be hormones, or signalling chemicals, but many other possibilities may exist that would be revealed only with the upcoming sequenced lepidopteran genome.

Pest genomics and bio-rational target sites

Insect genomics has the greatest potential application in developing novel pest-control products like bio-rational insecticides. These are the chemicals that aim at disrupting a physiological function specific to insects or to a group of insects. Genomics is applied to identify the target sites (proteins) that can be exploited for the developing these bio-rationals. The active ingredients of these insecticides are synthetic compounds; their insect specificity and mode of action make them far more environment-friendly than conventional chemical insecticides. Some bio-rational insecticides are obtained from natural sources or are synthetic analogues of natural compounds. The identification of such compounds can be greatly aided by biotechnology, whereby, the genes encoding insect proteins believed to be suitable targets for enzyme inhibition or hormone receptor antagonistic interactions are cloned and used for the development of in vitro high-throughput screening (HTS) assays\(^2\), where the three-dimensional structure of the target protein can be determined. Computer-assisted design can be used to help identify suitable inhibitors, agonists and antagonists in an approach similar to that currently employed for drug discovery\(^24\).
<table>
<thead>
<tr>
<th>Database</th>
<th>Hosted by</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAIBase</td>
<td>Institute of Applied Molecular Sciences (IAM)</td>
<td><a href="http://www.kaibase.org/index.html">http://www.kaibase.org/index.html</a></td>
</tr>
<tr>
<td>WhISkBase</td>
<td>National Institute of Agrobiological Sciences (NIAS)</td>
<td><a href="http://www.kaibase.org/whiskbase/index.html">http://www.kaibase.org/whiskbase/index.html</a></td>
</tr>
<tr>
<td>SilKBase</td>
<td>China National Tobacco and Vegetable Research Institute (CNTVRI)</td>
<td><a href="http://www.silkbase.org.cn/whiskbase/index.html">http://www.silkbase.org.cn/whiskbase/index.html</a></td>
</tr>
<tr>
<td>The Silkworm Knowledge Base (SilKB)</td>
<td>Beijing Genomics Institute (BGI)</td>
<td><a href="http://www.silkbase.org.cn/whiskbase/index.html">http://www.silkbase.org.cn/whiskbase/index.html</a></td>
</tr>
<tr>
<td>SilKmosh</td>
<td>Center for DNA Fingerprinting and Diagnostics, Hyderabad, India</td>
<td><a href="http://www.silkbase.org.cn/whiskbase/index.html">http://www.silkbase.org.cn/whiskbase/index.html</a></td>
</tr>
<tr>
<td>Microsatellite Database (SilKbase)</td>
<td>Center for DNA Fingerprinting and Diagnostics, Hyderabad, India</td>
<td><a href="http://www.silkbase.org.cn/whiskbase/index.html">http://www.silkbase.org.cn/whiskbase/index.html</a></td>
</tr>
<tr>
<td>FlyExpress</td>
<td>Arizona State University, Tempe, USA</td>
<td><a href="http://flyexpress.net/">http://flyexpress.net/</a></td>
</tr>
<tr>
<td>FlyBase</td>
<td>University of Cambridge, United Kingdom</td>
<td><a href="http://www.flybase.org/">http://www.flybase.org/</a></td>
</tr>
</tbody>
</table>

(Contd.)
<table>
<thead>
<tr>
<th>Databases</th>
<th>Year</th>
<th>Fosted by</th>
<th>URL adds</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berkeley Drosophila Genome Project (BDGP)</td>
<td>2007</td>
<td>Berkeley Drosophila Genome Project, Mailstop Berkeley</td>
<td><a href="http://www.fruitfly.org/">http://www.fruitfly.org/</a></td>
<td>The goal of the Drosophila Genome Centre is to finish the sequence of the <em>D. melanogaster</em> to high quality and to generate and maintain biological annotations of this sequence. In addition to genomic sequencing, the BDGP is producing gene disruptions using P-element-mediated mutagenesis on a scale unprecedented in metazoa, characterizing the sequence and expression of cDNAs and developing informatics tools that support the experimental process, identify features of DNA sequence, and allow us to present up-to-date information about the annotated sequence to the research community. DroSpeGe provides access to new and old <em>Drosophila</em> genomes.</td>
</tr>
<tr>
<td>DroSpeGe</td>
<td>2006</td>
<td>Genome Informatics Lab Biology, Indiana University, Bloomington</td>
<td><a href="http://insect.eugenics.org/-DroSpeGe/">http://insect.eugenics.org/-DroSpeGe/</a></td>
<td>DroSpeGe is a curated collection of known <em>Drosophila</em> transcriptional cis-regulatory modules (CRMs) and transcription factor binding sites (TFBSs). Flynet is a specialized database which focuses on molecular interactions (protein–DNA, protein–RNA and protein–protein) involved in <em>Drosophila</em> development.</td>
</tr>
<tr>
<td>REDfly</td>
<td>2006</td>
<td>Center for Computational Research, State University of New York, USA</td>
<td><a href="http://www.redfly.ccr.buffalo.edu">http://www.redfly.ccr.buffalo.edu</a></td>
<td>REDfly is a curated collection of known <em>Drosophila</em> transcriptional cis-regulatory modules (CRMs) and transcription factor binding sites (TFBSs).</td>
</tr>
<tr>
<td>FlyView – A Drosophila Image Database</td>
<td>1997</td>
<td>Institut fur Neurobiologie Badenstr, Munster</td>
<td><a href="http://pbio07.uai-muenster.de/">http://pbio07.uai-muenster.de/</a></td>
<td>A graphical atlas of expression patterns of genes and enhancer traps at all stages of development.</td>
</tr>
<tr>
<td>Aphidbase</td>
<td>2007</td>
<td>International Aphid Genomics Consortium, France</td>
<td><a href="http://www.aphidbase.com/aphidbase/">http://www.aphidbase.com/aphidbase/</a></td>
<td>It was developed to facilitate community annotation of the pea aphid genome by the International Aphid Genomics Consortium (IAGC).</td>
</tr>
<tr>
<td>BeetleBase</td>
<td>2007</td>
<td>Eioinformatics Center, Kansas State University, Manhattan, USA</td>
<td><a href="http://www.beetlebase.org">http://www.beetlebase.org</a></td>
<td>BeetleBase is a comprehensive sequence database and important community resource for <em>Tribolium</em> genetics, genomics and developmental biology.</td>
</tr>
<tr>
<td>InSatDB</td>
<td>2006</td>
<td>Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India</td>
<td><a href="http://www.cdf.org.in/InSatDB">http://www.cdf.org.in/InSatDB</a></td>
<td>It stores microsatellites from all the five insect genomes separately as well as carries complete annotations of these microsatellites.</td>
</tr>
<tr>
<td>AnoBase</td>
<td>2005</td>
<td>Institute of Molecular Biology and Biotechnology (MBB) Hellas, Greece</td>
<td><a href="http://www.anobase.org">http://www.anobase.org</a></td>
<td>AnoBase is a database containing genomic/biological information on aropheneline mosquitoes with emphasis on <em>Anopheles gambiae</em>.</td>
</tr>
<tr>
<td>Spodoptera</td>
<td>2006</td>
<td>Integrative Biology and Virology of Insects (IBIV), INRA, France</td>
<td><a href="http://spodobase.org">http://spodobase.org</a></td>
<td>An EST database for the lepidopteran crop pest: <em>Spodoptera</em>.</td>
</tr>
<tr>
<td>LocustDB</td>
<td>2006</td>
<td>Fetiing Genomics Institute (BGI), China</td>
<td><a href="http://locustdb.genomics.org.cn/jsp/about.jsp">http://locustdb.genomics.org.cn/jsp/about.jsp</a></td>
<td>Migratory ocus EST database, including homologous/orthologous sequences, functional annotations, pathway analysis, and codon usage based on conserved orthologous groups (COG), gene ontology (GO), protein domain (InterPro), and functional pathways (KEGG). VectorBase contains genome information for three mosquito species: <em>Aedes aegypti</em>, <em>Anopheles gambiae</em> and <em>Culex quinquefasciatus</em>, a body louse, <em>Pediculus humanus</em> and a tick species, <em>Ixodes scapularis</em>.</td>
</tr>
</tbody>
</table>
Comparative genomics approaches

Comparative genomics approaches have identified genes that encode proteins unique to insects or to specific insect taxa. Genome-wide in vivo RNAi screens, such as the one now possible for D. melanogaster, would allow the selection of those insect-specific genes for which transcriptional inhibition induces lethality. Homologues of these genes could then be cloned from pest species and submitted to a similar RNAi analytical approach, thus allowing the identification of genes that are promising bio-rational target sites.

In this way, insect genomics, biotechnology and insect cell lines have begun to provide powerful tools for the identification of new lead compounds. Insect cell lines together with HTS procedures (HTS cell-based assays), can enable the discovery of new modes of action for insecticide candidates. This is new emerging field in entomology and little information is available with us, but sooner or later it will enter a new era.

Future aspects

Drosophila genomic sequence is a major milestone for genomics, as it vindicates a new strategy for sequencing large eukaryotic genomes and as a model system to understand biological functions. In Drosophila post-genomic age, applications of genomics technology to entomology has added volume and quality to the data. At present, only two agricultural pest insect genomes, that of the red flour beetle T. castaneum and pea aphid, A. pism have been fully sequenced, but within the next few years several lepidopteran pest sequences will be available. This information will enormously increase our knowledge for understanding the biology of insects and insecticide resistance, which poses an increasing problem for pest control. In the future, where many genomes will be sequenced, a major application of bioinformatics will be the modelling of genetic and metabolic networks, and then comparative genomics will be an increasingly useful approach for pinpointing common and different genes across species.

Genome comparisons between different organisms will be informative on several levels, and information on genomic sequence and organization will be useful to explore gene functions. Functional genomics is being applied more and more in every aspect of life sciences research, including ecology and evolution. Thus, in future, there is a growing tendency for insect molecular scientists to reach out to the broader molecular biology community with all the benefits that such interactions can have for the application of molecular tools in insect science.

Conclusion

The insect genomic databases are goldmines with information on all the proteins, biochemical and physiological processes of an insect. The newly sequenced insect genomes may harbour many surprises for biochemists, molecular biologists and insect physiologists. Insect pest control will soon enter the genomic era with all its surprises and discoveries, as pest and parasitoids genomes are now available. Thus, genomic advances during the last 10 years will revolutionize insect research.

References


Received 19 August 2011; revised accepted 23 December 2011