

Perspective

The promise of insect genomics

Cornelis JP Grimmelikhuijzen,* Giuseppe Cazzamali, Michael Williamson and Frank Hauser

Centre for Functional and Comparative Insect Genomics, Department of Cell Biology and Comparative Zoology, Institute of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark

Abstract: Insects are the largest animal group in the world and are ecologically and economically extremely important. This importance of insects is reflected by the existence of currently 24 insect genome projects. Our perspective discusses the state-of-the-art of these genome projects and the impacts that they have on basic insect science and pest control.

© 2007 Society of Chemical Industry

1 INTRODUCTION

Insects are the largest animal group in the world (75% of all species are insects) and are ecologically and economically extremely important, because most flowering plants depend on insects for their pollination. Honey bees alone, for example, pollinate about 15 billion dollars worth of crops yearly in the United States.¹ However, insects can also be severe agricultural pests, destroying up to 30% of our potential annual harvest, and can be vectors (intermediate pathogen carriers) for plant diseases and major human diseases such as malaria, elephantiasis, sleeping sickness, dengue fever, and yellow fever. There are an estimated 300–500 million cases of malaria and up to 2.7 million deaths (mainly children) from malaria each year. Other diseases are equally serious: elephantiasis, for example, disables 130 million people worldwide and 1.1 billion people, 17% of the world's population, are at risk of infection. From this short summary it will be clear that insects can be both very beneficial and harmful (pests), and that a few insect species are hampering the welfare of many hundreds of millions of people, especially in developing countries. Vast social benefits would be gained, therefore, if the populations of these pest insects could be selectively reduced.

2 INSECT GENOME PROJECTS

In spite of the importance of insects, our knowledge of their biology is still far from complete. At present, however, there are highly exciting developments occurring within the field of insect research, because the genomes from 24 insect species have recently been sequenced, or are in the process of being sequenced. Among these insects are 12 fruitfly species belonging to the genus *Drosophila*, the best-known being *Drosophila melanogaster* Meigen, which was the first insect with

a sequenced genome.^{2,3} The other *Drosophila* species are given in Fig. 1 (top). All the genomes from the insect species given in this figure have been or will be sequenced with at least an 8–9 times coverage (i.e. the standard for an established sequenced genome). This number implies that only very few sequencing errors are left in the genome draft, and also that there are only minor sequencing gaps. As one can see in Fig. 1, seven of the 12 *Drosophila* species have already reached the standard. In addition to the genome from *D. melanogaster*,² only that from *D. pseudoobscura* has been published,⁴ although information on the other *Drosophila* species is available on the Internet.⁵

Although much sequencing effort has been focused on *Drosophila*, other insect genomes have also been sequenced during the last 6 years. These are the genomes from the malaria mosquito *Anopheles gambiae* Giles (sequenced and published);^{6,7} the silkworm *Bombyx mori* L. (sequenced and published);^{8–10} the honey bee *Apis mellifera* L. (sequenced and published);^{11,12} the red flour beetle *Tribolium castaneum* Herbst (sequenced but unpublished);¹³ the yellow fever mosquito *Aedes aegypti* L. (sequenced but unpublished);¹⁴ the mosquito *Culex pipiens* L. (sequenced but unpublished);¹⁵ the blood-sucking bug *Rhodnius prolixus* Stal (in progress);¹⁶ three parasitic wasp species belonging to the genus *Nasonia* (in progress);^{17,18} the pea aphid *Acyrtosiphon pisum* Harris (in progress);¹⁹ and the louse *Pediculus humanus* L. (in progress).²⁰ These species represent eleven different insect genera, six different insect orders and also the two major evolutionary lineages of insects: Holometabola (insects with a complete metamorphosis) and Hemimetabola (insects having an incomplete metamorphosis). Figure 1 gives the evolutionary relationships between these insect species.

* Correspondence to: Cornelis JP Grimmelikhuijzen, Centre for Functional and Comparative Insect Genomics, Department of Cell Biology and Comparative Zoology, Institute of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark
E-mail: cgrimmelikhuijzen@bi.ku.dk
(Received 15 July 2006; revised version received 18 November 2006; accepted 18 November 2006)
Published online 21 March 2007; DOI: 10.1002/ps.1352

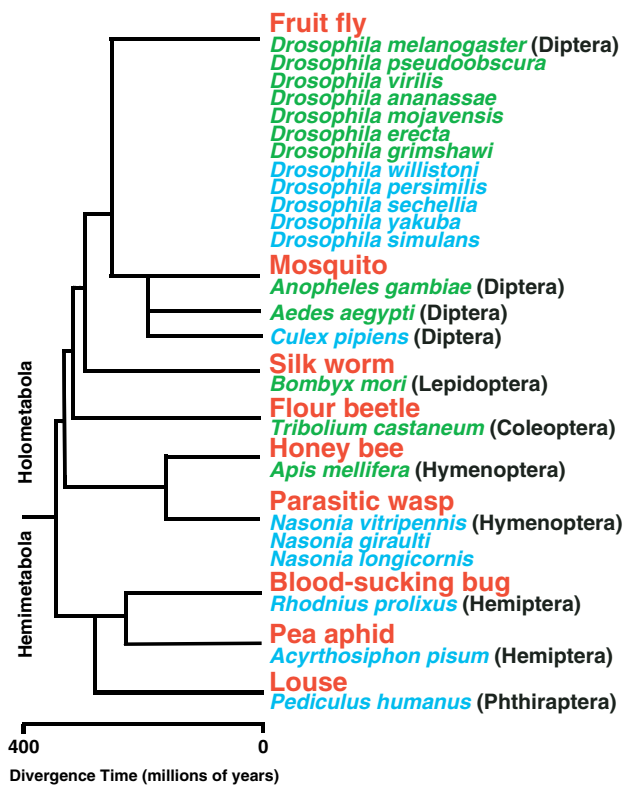


Figure 1. The evolutionary relationships of insects for which genome projects exist. Insects whose genome has been completely sequenced (to at least 8–9 times coverage) are highlighted in green. Insects for which the genome sequence is still incomplete (less than 8–9 times coverage, as of February 2007) are highlighted in blue. The divergence times (the evolutionary distances in time between each species and the common ancestor with its neighbour in this figure) are given as horizontal bars and are from data presented in references 11 and 22.

The sequenced genomes from these insects will provide invaluable information on the basal insects that were the common ancestors of holo- and hemimetabolous insects and on 340 million years of insect evolution. This is an evolutionary distance comparable with that between fishes and humans.

In addition to the finished insect genome projects and the ones that are in progress (Fig. 1), there are several other insect genome projects that are under consideration or that are not publicly accessible. For example, in addition to the silkworm *B. mori*, several other lepidopterans are being discussed for a sequencing project. Since numerous lepidopteran larvae are serious agricultural pests, the sequencing of their genomes is now being encouraged.²¹ For the lepidopteran pest *Spodoptera frugiperda* Smith (fall armyworm), there exists an expressed sequence tag (EST) project (containing cDNA sequences corresponding to mRNAs), but no genuine genome project.²¹ Interestingly, the genome from the lepidopteran pest *Heliothis virescens* Fabricius (tobacco budworm) has already been fully sequenced, but these data are in commercial hands (Bayer AG, Germany, in collaboration with Exelixis Inc., USA) and are not publicly accessible.

3 GOLDMINES

Already now, before the publication of the *Tribolium castaneum* genome, insect genome researchers have made surprising discoveries. For example, by comparing 185 honey bee genes with the corresponding (orthologous) genes from *Tribolium* and the other holometabolous insects with a sequenced genome, they found that the honey bee, which belongs to the insect order Hymenoptera, is evolutionarily clearly more basal than *Tribolium*, which belongs to the order Coleoptera (or beetles).²² Thus, Hymenoptera (bees, wasps and ants) is the most basal order of the holometabolous lineage, followed by Coleoptera (beetles), Lepidoptera (moths and butterflies) and Diptera (flies) (Fig. 1). This is counter to current textbook knowledge and disturbing news for insect systematists, who used to base their phylogenetic trees (and viewpoints) on morphological characteristics and ribosomal RNA or mitochondrial DNA comparisons. This illustrates the power of comparative genomics and suggests that many more insect paradigms will be questioned in the near future.

The insect genomic databases are goldmines, because they contain the information of all proteins and, thereby, of all biochemical and physiological processes that occur in an insect. Therefore, also for biochemists, molecular biologists and insect physiologists, the newly sequenced insect genomes may harbour many surprises, and the main promising outcome is likely to be in their disciplines.

4 WHY HAVE THESE INSECTS BEEN SELECTED FOR A GENOME PROJECT?

The 24 insect species from Fig. 1 have been selected for a genome project for different reasons. First, some insects are model organisms, such as the fruitfly *D. melanogaster*, which is a well-established model for geneticists and developmental and molecular biologists. The other *Drosophila* species have been sequenced to help with the interpretation of the *D. melanogaster* genome.⁴ A second class are the medically important insects that are vectors for serious diseases, such as malaria (the mosquito *A. gambiae*), yellow fever (the mosquito *A. aegypti*), elephantiasis (the mosquito *C. pipiens*), Chagas disease (the blood-sucking bug *R. prolixus*), and typhus (the louse *P. humanus*). A third class are the insects that are agriculturally important. This group can be subdivided into agriculturally beneficial insects, such as the honey bee, *A. mellifera*, which is a major pollinator of food plants and producer of honey, and the silkworm *B. mori*, which produces silk. The red flour beetle *T. castaneum* (which destroys stored grain and many other dried and stored commodities for human consumption) and the pea aphid *A. pisum* (which causes severe damage to green food plants) are serious agricultural pests. The parasitic wasp *Nasonia vitripennis* Walker and the other two *Nasonia* species have been selected because of their potential to control agricultural pest insects.

The divisions between these four groups, however, are not very clear, and several of the insects mentioned in Fig. 1 can be regarded as model organisms. The honey bee, for example, is also used as a model to study sociality, and the red flour beetle is a model for insect embryonic development. The availability of a sequenced genome for each of the mentioned insects will, of course, strongly improve their status as model insects.

5 IMPACT ON CHEMICAL PEST MANAGEMENT

At present, only one agricultural pest insect genome, that of the red flour beetle *T. castaneum*, has been fully sequenced, but within the next few years the pea aphid *A. pisum* and several lepidopteran pests will follow. These achievements will enormously increase our knowledge of the biology of these insects. It will also provide a much better understanding of practical matters such as insecticide resistance, which poses an increasing problem for pest control.²³ Furthermore, many more insecticide targets will be identified than the few, such as acetylcholinesterase, sodium channels and GABA receptors, that are currently being used. Also, the olfactory signalling cascades that are responsible for host-seeking behaviour (plant/insect interactions) will be unravelled. In *T. castaneum*, for example, a high number of olfactory receptors have recently been identified (Walden K and Robertson H, private communication). These cell membrane proteins are new targets that might be used for pest control.

Among the plethora of new drug targets to be discovered, we would like to point to G protein-coupled receptors (GPCRs) as a group of targets that are potentially promising. Insect GPCRs are cell membrane receptors for neurohormones such as biogenic amines, neuropeptides and protein hormones. When activated by their natural extracellular ligands, they initiate an intracellular second messenger cascade, leading to biological effects. Insects have 50–80 neurohormone GPCRs, which, together with their ligands, occupy a high hierarchic position in their physiology and steer central processes, such as reproduction, development, homeostasis and feeding.^{24–27} A blockade or overstimulation of these receptors in pest insects, therefore, will kill the animals, or reduce their fitness and thereby specifically decrease their populations.²⁶ In humans, GPCRs are well-established drug targets, and more than 30% of all currently prescribed human medications act on these membrane receptors.²⁸ High-throughput screening of GPCRs for the presence of non-peptide agonists or antagonists in large chemical libraries is a routine job in the pharmaceutical industry. Thus, now that specific insect GPCRs have been identified,^{24,25} it will be feasible to develop high-affinity non-peptide/non-protein agonists or antagonists. Because many of these new synthetic insect GPCR ligands would act specifically (and would not interfere with vertebrate signalling

pathways, in contrast, for example, to the currently used acetylcholinesterase inhibitors) they would be environmentally safe.

6 IMPACT ON BIOLOGICAL PEST CONTROL

The improved understanding of pest insects made possible by sequencing of their genomes will stimulate the design of new classes of transgene microorganisms to be used in pest control. Baculoviruses, 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).²⁹ The baculoviruses can be improved by introducing DNA into their genomes coding for lepidopteran proteins that stop larval phytophagy at a very early stage.²⁹ These larval proteins could be hormones, signalling satiety, but many other possibilities may exist that are currently 'hidden' in the unsequenced lepidopteran genome. Because these proteins are 'natural' insect proteins (and not toxins), the genetically modified baculoviruses (and the infected insects) would not pose a risk for the environment.

The complete genomic sequences from the three parasitic wasps (Fig. 1) will also break new ground. Parasitic wasps are a diverse group of hymenopterans (over 170 000 species) that are the natural enemies of a broad range of insect pests. Adult parasitic wasps are free living and lay their eggs into various developmental stages (eggs, larvae, pupae) of other insects, thereby killing their host. A major host preference gene has recently been mapped in the genome from *Nasonia* (Werren J, unpublished).³⁰ The newly sequenced genomes from the three *Nasonia* species (Fig. 1), therefore, will help to better understand and eventually manipulate host finding and host preference, and make parasitic wasps more efficient for the biological control of agricultural pests.³⁰ It might perhaps even be possible to create a palette of transgenic parasitic wasps, each having a certain host range, thereby covering a broad spectrum of agricultural pests.³⁰ All these developments will certainly lead to a significant reduction in worldwide pesticide use, which will be beneficial for human health and the environment.

7 CONCLUSIONS

The emergence of a large number of insect genome projects during the last 8 years (Fig. 1) will revolutionize insect research, as large amounts of new data will rapidly become available. Among those insects with a sequenced genome are two agricultural pests (*T. castaneum* and *A. pisum*) and three parasitic wasps that can be used for biological pest control. This means that insect pest control soon will enter the genomic era with all its surprises and discoveries.

REFERENCES

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

- 2 Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, *et al*, The genomic sequence of *Drosophila melanogaster*. *Science (Washington)* **287**:2185–2195 (2000).
- 3 *Drosophila* Genome Project web page. [Online]. Available: <http://www.flybase.org> [10 February 2007].
- 4 Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, *et al*, Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* **15**:1–18 (2005).
- 5 Update on the sequenced genomes from twelve *Drosophila* species. [Online]. Available: <http://rana.lbl.gov/drosophila/> [8 March 2007].
- 6 Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, *et al*, The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science (Washington)* **298**:129–149 (2002).
- 7 *Anopheles gambiae* Genome Project web page. [Online]. Available: <http://www.anobase.org/> [10 February 2007].
- 8 Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, *et al*, A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science (Washington)* **306**:1937–1940 (2004).
- 9 Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanomori H, *et al*, The genome sequence of silkworm, *Bombyx mori*. *DNA Res* **11**:27–35 (2004).
- 10 *Bombyx mori* Genome Project web page. [Online]. Available: <http://sgp.dna.affrc.go.jp> [10 February 2007].
- 11 Honey Bee Genome Sequencing Consortium, Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature (London)* **443**:931–949 (2006).
- 12 Honey Bee Genome Project web page. [Online]. Available: <http://www.hgsc.bcm.tmc.edu/projects/honeybee/> [10 February 2007].
- 13 *Tribolium castaneum* Genome Project web page. [Online]. Available: <http://www.bioinformatics.ksu.edu/BeetleBase/> [10 February 2007].
- 14 *Aedes aegypti* Genome Project web page. [Online]. Available: <http://msc.tigr.org/aedes/aedes.shtml> [10 February 2007].
- 15 *Culex pipiens* Genome Project web page (Online). Available: <http://cpipiens.vectorbase.org/index.php> [10 February 2007].
- 16 *Rhodnius prolixus* Genome Project web page. [Online]. Available: <http://genome.wustl.edu/genome.cgi?GENOME=Rhodnius%20prolixus> [10 February 2007].
- 17 *Nasonia vitripennis* Genome Project web page. [Online]. Available: <http://www.hgsc.bcm.tmc.edu/projects/nasonia> [10 February 2007].
- 18 Information on the three *Nasonia* genome projects. [Online]. Available: <http://www.hgsc.bcm.tmc.edu/projects/> [10 February 2007].
- 19 Pea Aphid Genome Project web page. [Online]. Available: <http://www.hgsc.bcm.tmc.edu/projects/aphid> [10 February 2007].
- 20 *Peduculus humanus* Genome Project web page (Online). Available: <http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/BodyLouseSeq.pdf> [10 February 2007].
- 21 Negre V, Hotelier T, Volkoff AN, Gimenez S, Cousserans F, Mita K, *et al*, SPODOBASE: an EST database for the lepidopteran crop pest *Spodoptera*. *BMC Bioinformatics* **7**:322 (2006).
- 22 Savard J, Tautz D, Richards S, Weinstock GM, Gibbs RA, Werren JH, *et al*, Phylogenomic analysis reveals bees and wasps (*Hymenoptera*) at the base of the radiation of holometabolous insects. *Genome Res* **16**:1334–1338 (2006).
- 23 Hemingway J, Field L and Vontas J, An overview of insecticide resistance. *Science (Washington)* **298**:96–97 (2002).
- 24 Hauser F, Williamson M, Cazzamali G and Grimmelikhuijzen CJP, Identifying neuropeptide and protein hormone receptors in *Drosophila melanogaster* by exploiting genomic data. *Brief Funct Genomics Proteomics* **4**:321–330 (2006).
- 25 Hauser F, Cazzamali G, Williamson M, Blenau W and Grimmelikhuijzen CJP, A review of neurohormone GPCRs present in the fruitfly *Drosophila melanogaster* and the honey bee *Apis mellifera*. *Prog Neurobiol* **80**:1–19 (2006).
- 26 Gäde G and Goldsworthy GJ, Insect peptide hormones: a selective review of their physiology and potential application for pest control. *Pest Manag Sci* **59**:1063–1075 (2003).
- 27 Gäde G, Regulation of intermediary metabolism and water balance of insects by neuropeptides. *Annu Rev Entomol* **49**:93–113 (2004).
- 28 Liebmann C, G protein-coupled receptors and their signalling pathways: classical therapeutic targets susceptible to novel therapeutic concepts. *Curr Pharm Des* **10**:1937–1958 (2004).
- 29 Szewczyk B, Hoyos-Carvajal L, Paluszek M, Skrzecz I and Lobo de Souza M, Baculoviruses – re-emerging biopesticides. *Biotechnol Adv* **24**:143–160 (2006).
- 30 The White Paper of the *Nasonia* Genome Project. [Online]. Available: <http://www.rochester.edu/College/BIO/labs/WerrenLab/nasonia/LettersOfSupport/NasWhitePaper2004.pdf> [10 February 2007].