

ment—then the benefits of this genome sequence become obvious. The genome sequence will enable access to the major regulatory genes involved in resistance, particularly if orthologous regulators control metabolically based resistance in insects generally. For example, management of resistance in practice currently involves basic rotations and mixtures or mosaics of different insecticides. Access to insect-specific metabolic enzyme regulators will provide a target for “add-ons” to current insecticides, which should expand their natural life-span by blocking common resistance pathways while leaving mammalian toxicity unaffected.

Insecticide resistance can result from direct changes to the proteins that normally bind to the insecticides. For example, mutations in sodium channels (the target of DDT and pyrethroids) and in acetylcholinesterase (AChE; the target of organophosphates and carbamates) have been well documented in many insect species including, in the case of sodium channels, mosquitoes (3, 8). The *Anopheles* genome will provide information on the target site genes, facilitating cloning and mutagenesis studies to determine the precise nature of the mutations and to aid in predicting interactions between insect proteins and insecticides. In the longer term, this could lead to new insecticidal molecular targets. This approach may be especially important for AChE as there is increasing evidence for multiple AChE genes from the *Anopheles* and *Drosophila* genome databases (9). Two AChE genes are apparent in the *A. gambiae* genome as

detailed by Ranson *et al.* in this issue (10), and to date no resistance-linked mutations have been identified in mosquitoes predominantly in studies on the sex-linked AChE gene. The *Anopheles* genome in conjunction with that of *Drosophila* also provides sequences of nicotinic acetylcholine receptor subunits, which will facilitate their cloning from other insect species. This receptor is the target of an important new group of agrochemicals, but until now studies of insect receptors have relied on coexpression of insect genes with a vertebrate subunit (11).

Many instances of resistance, resulting from a change in a single regulator gene, may trigger complex cascades of expression of unrelated genes. The presence of the full genome sequence will prompt a move away from the reductionist approach that has dominated the last two decades of resistance research. Such an approach has tended to result in a lack of appreciation for how the large physiological changes that often accompany resistance can influence other characteristics of the insect vector. For example, many of the large scale-up regulations of enzyme families that accompany insecticide resistance result in profound changes in oxidative stress levels in the cells where these enzymes are expressed. These are often the identical tissues in which parasites or viruses reside during transmission by insect vectors from one human to another.

Microarray and cell biology approaches are being used to define mechanisms in *A. gambiae* that enable this insect vector to be refractory to infection by malaria and other parasites. Avail-

able data already strongly indicate that an enhanced ability of the refractory insects to tolerate oxidative stress is integral to their ability to resist parasite infection (12). The convergence of these two lines of resistance research will be greatly facilitated by our ability to look at regulation patterns in different phenotype combinations across the whole genome. There are already indications that these are not mutually exclusive systems; for example, filarial parasites fail to develop in highly insecticide-resistant *Culex* mosquitoes (13).

Overall, the *A. gambiae* genome provides us with exciting opportunities to move from knowledge and understanding of how resistance genes work to the practical application of that knowledge in the field. This will fundamentally improve the control of malaria and other important vector-borne diseases and will contribute to the wider studies of resistance in agricultural pests.

#### References

1. C. F. Curtis *et al.*, *Philos. Tran. R. Soc London Ser. B.* **353**, 1769 (1998).
2. H. Ranson *et al.*, *Insect Mol. Biol.* **9**, 499 (2000).
3. F. Chandre *et al.*, *Bull. WHO* **77**, 230 (1999).
4. H. Ranson *et al.*, *Insect Mol. Biol.* **11**, 409 (2002).
5. R. L. Blackman *et al.*, *Heredity* **82**, 80 (1999).
6. M. G. Paton *et al.*, *Biochem. J.* **346**, 17 (2000).
7. L. M. Field, *Biochem. J.* **349**, 863 (2000).
8. H. Ranson *et al.*, *Insect Mol. Biol.* **9**, 491 (2000).
9. M. Grauso *et al.*, *FEBS Lett.* **424**, 279 (1998).
10. H. Ranson *et al.*, *Science* **298**, 179 (2002).
11. Y. Huang *et al.*, *Neurosci. Lett.* **284**, 116 (2000).
12. G. Dimopoulos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8814 (2002).
13. L. McCarroll *et al.*, *Nature* **407**, 961 (2000).

#### VIEWPOINT

## The *Anopheles* Genome and Comparative Insect Genomics

Thomas C. Kaufman,<sup>1</sup> David W. Severson,<sup>2</sup> Gene E. Robinson<sup>3</sup>

The *Anopheles gambiae* genome sequence, coupled with the *Drosophila melanogaster* genome sequence, provides a better understanding of the insects, a group that contains our friends, foes, and competitors.

The Phylum *Arthropoda* is the most species-rich and morphologically diverse animal group on the planet. Since their appearance in the Early Cambrian and their subsequent radiation, arthropods have come to inhabit and dominate the vast majority of ecological habitats. From the many different arthropod groups that existed in the Early Cambrian,

only four have survived to the present: the Chelicerata, Myriapoda, Crustacea, and Insecta. Members of these four groups plague us, transmit diseases, benefit us, and feed us. The genome sequence of the African malaria vector, the mosquito *Anopheles gambiae*, reported on page 129 of this issue (1), coupled with the *Drosophila melanogaster* genome sequence (2), provides us with new insights into the genetic makeup of two members of the Insecta, arguably the dominant group of arthropods.

The genome sequences of *A. gambiae* and *Plasmodium falciparum*, the malaria parasite it transmits (3, 4), will yield fresh insights

into parasite and vector biology that will lead to more efficient disease control strategies. A new approach to vector-borne disease control based on the genetic manipulation of the mosquito has already received considerable attention (5, 6). The *A. gambiae* genome sequence will accelerate efforts to identify molecules that can inhibit parasite development in the vector and subsequently prevent transmission to humans. Stable germline transformation has been demonstrated for several vector mosquitoes (7–9). This is encouraging news given that transgenic anopheline mosquitoes engineered to express an anti-*Plasmodium* molecule turn out to be inefficient vectors for disease transmission in the laboratory (10).

Of the ~3500 mosquito species, molecular information exists for only a small num-

<sup>1</sup>Department of Biology, Indiana University, Bloomington, IN 47405, USA. <sup>2</sup>Center for Tropical Disease Research and Training, Department of Biological Sciences, University of Notre Dame, IN 46556, USA.

<sup>3</sup>Department of Entomology and Neuroscience Program, University of Illinois, 505 South Goodwin Avenue, Urbana, IL 61801, USA.

ber, and even this is limited (11). The *Anopheles* sequence will facilitate elucidation of biological processes unique to mosquitoes, including genes and pathways associated with blood feeding, host-seeking behavior, and immune responses to pathogens. Comparison of orthologous genes should help to illuminate the crucial and vexing issue of interspecific variability in vector competence. Why is one species of mosquito a fully competent vector for a given pathogen, whereas another is completely refractory to infection?

The *Anopheles* genome sequence forms the foundation for comparative genomic analyses across mosquito species. *A. gambiae* represents the subfamily *Anophelinae*, which contains the primary vectors of malaria parasites. But it is the subfamily *Culicinae* that contains the majority of mosquito species, including the primary vectors of several emerging or reemerging arbovirus diseases (yellow fever, dengue fever, and West Nile encephalitis) and also of lymphatic filariasis. These two mosquito subfamilies appear to differ significantly in genomic structure (11, 12)—gene order conservation between *A. gambiae* and the culicine mosquito *Aedes aegypti* (the primary vector of yellow and dengue fever viruses) is characterized by ex-

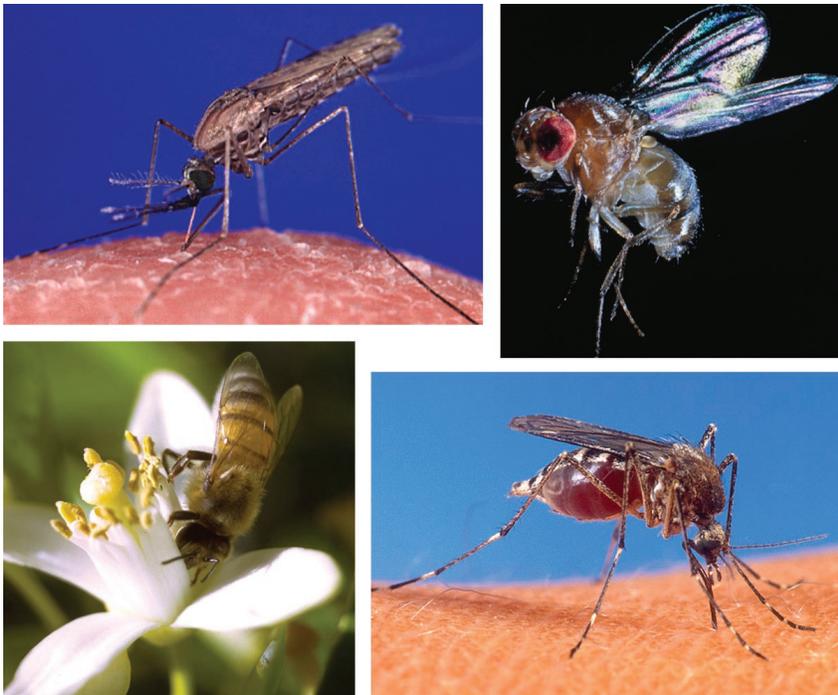
tensive local rearrangements within chromosomal arms (13). This is similar for the *Drosophila* and *Anopheles* genomes, which show conservation of whole chromosome arms but considerable local rearrangement within arms, as reported by Zdobnov *et al.* (14) and Sharakhov *et al.* (13) in this issue. Conversely, comparisons within the *Culicinae* indicate conservation of linear gene order (15). Given the diversity within the mosquito lineages, the availability of the *A. gambiae* genome sequence should fuel interest in the study of additional mosquito genomes. Indeed, a National Institutes of Health/National Institute of Allergy and Infectious Diseases (NIH/NIAID)-sponsored genome-sequencing project for *Ae. aegypti* has recently been initiated (16). Also, the NIH's National Human Genome Research Institute has just selected the honeybee *Apis mellifera* to be among the next group of organisms for genome sequencing (17). Bees belong to the Order Hymenoptera, and so the honeybee will be the first insect chosen for sequencing that is not a member of the Order Diptera. Sequencing the genome of the highly social honeybee will have a strong impact on sociogenomics (18, 19), which seeks to develop a comprehensive understanding in molecular terms of social life in all creatures: how it evolved, how it is governed, and how it influ-

ences all aspects of genome structure, gene expression, and organismal development, physiology, and behavior. Comparative analyses of the genomes of *Anopheles*, *Drosophila*, and the honeybee will be valuable for identifying bee genes that are lacking in the two dipteran genomes, some of which may be of importance for understanding sociality. More likely, the bee's vaunted behavioral complexity and social skills will be found to be due largely to differences in gene regulation. Comparative genomics and new algorithms doubtless will identify conserved regulatory sequences and regulatory networks (20, 21) or new candidate cis-regulatory sequences (22, 23).

Of course, insects differ not only in behavior, but in form as well. In fact, one key to the success of all arthropods has been their ability to evolve increasingly complex body plans and specialized appendages for locomotion and food acquisition (24, 25). Much of the genetic variability that underlies this spectacular divergence is likely to reside in regulatory differences, but the situation is complex. Often we tacitly assume that if we understand the developmental pathway of a morphological trait in one species, then it is likely to proceed in the same way in a related species with the same morphology. That this assumption is incorrect is demonstrated by recent studies on the development of insect wings and mouthparts.

Abouheif *et al.* have shown that the signaling pathways for wing development are different among distinctive castes of ants, some winged and some not (26). By analyzing the *Drosophila* wing specification pathway, these authors cloned several orthologous genes from different ant species and determined their expression patterns in the embryonic anlagen of the wings for the different ant castes. The ant orthologs were conserved in their expression in the wing primordia of the winged castes but were apparently not expressed correctly in wingless siblings. Moreover, the point at which the block in wing specification occurred differed among the four ant species. Given that winglessness is thought to have evolved only once in ants, there seems to be regulatory lability in wing specification among wingless castes, but a conserved signaling cascade among winged forms. Clearly, we need to elucidate the precise regulation of wing-specification genes, which will require knowledge of the regulatory sequences of these loci.

Understanding the regulatory plasticity of mouthpart development directed by the Hox gene *proboscipedia* (*pb*) will also require elucidation of regulatory sequences. In most insects, *pb* expression overlaps that of *Deformed* (*Dfd*) in the maxillary segment and of *Sex combs reduced* (*Scr*) in the adjacent labial segment. In the *Drosophila* embryo, *Dfd* and *Scr* activate the expression of *pb* in the maxillary and labial segments, respectively



**Fig. 1.** Depicted are the insect species with genomes completely sequenced or where genome sequencing has been initiated. Currently, complete genome sequences are available for two Dipterans, *Anopheles gambiae* (top left) and *Drosophila melanogaster* (top right); sequencing of genomes from a Hymenopteran, *Apis mellifera* (bottom left), and an additional Dipteran, *Aedes aegypti* (bottom right), are under way. Thus, 2 of the extant 33 orders of the Insecta have been or will be sampled—a rather small and specialized representation of a large, morphologically and behaviorally diverse group of organisms. [Photos: J. Gathany/U.S. Centers for Disease Control and Prevention (*A. gambiae*); D. Dale/Photo Researchers, Inc. (*D. melanogaster*); A. J. Ferreira/California Academy of Sciences (*Apis mellifera*); U.S. Department of Agriculture (*Ae. aegypti*)].

(27). In the beetle *Tribolium castaneum*, the three genes have a similar expression pattern but *pb* expression is not diminished in a *Dfd* mutant, suggesting that *Dfd* is not necessary to activate *pb* in this species (28). This probably reflects a difference in *pb* regulatory sequences, as the *Tribolium* Deformed protein can activate *Drosophila pb* when it is expressed in *Drosophila* (29). In the bug *Oncopeltus fasciatus*, *Dfd* cannot activate *pb* because *pb* is not even present in the maxillae (30). Additionally, although *pb* and *Scr* are coexpressed in the labial appendages, RNA interference analysis suggests that *Scr* does not activate *pb* (31). Lastly, even in the *Drosophila* adult the regulatory hierarchy appears to be different from that in the *Drosophila* embryo; *Scr* does not activate *pb* in adults, but rather *pb* is necessary to activate *Scr* (32). Thus, we have three insect species and four different regulatory systems to control the expression of *proboscipedia*. Considering the millions of different insect species, these results suggest enormous diversity in the regulation of this, and other, developmental genes.

The completed sequences of the *Dro-*

*sophila* and *Anopheles* genomes and the prospective sequencing of the *Apis* and *Aedes* genomes will provide significant insights into the insects and their development, behavior, and evolution (Fig. 1). But these four species represent only the beginning of an analysis of the Insecta, much less of the whole of the Arthropoda. Next we might consider sequencing genomes of representatives from the Coleoptera and Lepidoptera. These two insect orders contain many of our most serious agricultural pests and, together with the Diptera and Hymenoptera, comprise the “Big Four” insect orders that have evolved “complete” (holometabolous) development. The scientific community is now blessed with a wealth of sequencing capacity. Given the obvious importance of insects to our well-being and existence, it is important that some of it be used to build a strong empirical foundation for comparative insect genomics.

#### References

1. R. A. Holt *et al.*, *Science* **298**, 129 (2002).
2. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
3. S. L. Hoffman, G. M. Subramanian, F. H. Collins, J. C. Venter, *Nature* **415**, 702 (2002).
4. World Health Organization, *The World Health Report*

- 2000, *Health Systems: Improving Performance* (World Health Organization, Geneva, 2000).
5. B. J. Beaty, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10295 (2000).
6. K. S. Aultman *et al.*, *Trends Parasitol.* **17**, 507 (2001).
7. F. Catteruccia *et al.*, *Nature* **405**, 959 (2000).
8. M. L. Allen *et al.*, *J. Med. Entomol.* **38**, 701 (2001).
9. Z. N. Adelman *et al.*, *Mol. Biochem. Parasitol.* **121**, 1 (2002).
10. J. Ito *et al.*, *Nature* **417**, 452 (2002).
11. D. W. Severson *et al.*, *Annu. Rev. Entomol.* **46**, 183 (2001).
12. K. S. Rai *et al.*, *Adv. Genet.* **41**, 1 (1999).
13. I. V. Sharakhov *et al.*, *Science* **298**, 182 (2002).
14. E. M. Zdobnov *et al.*, *Science* **298**, 149 (2002).
15. J. R. Anderson *et al.*, *Mol. Phylogenet. Evol.* **20**, 316 (2001).
16. D. W. Severson, personal communication.
17. E. Pennisi, *Science* **296**, 1589 (2002).
18. G. E. Robinson, *Trends Ecol. Evol.* **14**, 202 (1999).
19. ———, *Science* **297**, 204 (2002).
20. B. P. Bertran *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 757 (2002).
21. G. G. Loots *et al.*, *Science* **288**, 136 (2000).
22. K. Ishihara *et al.*, *Genome Res.* **10**, 664 (2000).
23. D. A. Papatsenko *et al.*, *Genome Res.* **12**, 470 (2002).
24. R. C. Brusca, G. J. Brusca, *Invertebrates* (Sinauer Associates, Sunderland, MA, 1990).
25. W. C. Wheeler *et al.*, *Cladistics* **17**, 113 (2001).
26. E. Abouheif, G. A. Wray, *Science* **297**, 249 (2002).
27. D. B. Rusch, T. C. Kaufman, *Genetics* **156**, 183 (2000).
28. M. DeCamillis *et al.*, *Genetics* **159**, 1643 (2001).
29. S. J. Brown *et al.*, *Dev. Genes Evol.* **209**, 389 (1999).
30. B. T. Rogers *et al.*, *Evol. Dev.* **4**, 1 (2002).
31. C. L. Hughes *et al.*, *Development* **127**, 3683 (2000).
32. A. Abzhanov *et al.*, *Development* **128**, 2803 (2001).

#### VIEWPOINT

## Speciation Within *Anopheles gambiae*—the Glass Is Half Full

A. della Torre,<sup>1\*</sup> C. Costantini,<sup>1\*</sup> N. J. Besansky,<sup>3</sup> A. Caccone,<sup>4</sup> V. Petrarca,<sup>2</sup> J. R. Powell,<sup>4</sup> M. Coluzzi<sup>1</sup>

Restrictions to gene flow among molecular forms of the mosquito *Anopheles gambiae sensu stricto* reveal an ongoing speciation process affecting the epidemiology of malaria in sub-Saharan Africa.

The most important vector of the malaria parasite in sub-Saharan Africa is the mosquito *Anopheles gambiae sensu stricto* (*s.s.*). It belongs to a group of sibling species—known as the *A. gambiae* complex—that are morphologically indistinguishable but exhibit distinct genetic and eco-ethological differences reflected in their ability to transmit malaria. *Anopheles gambiae s.s.* shows extreme genetic heterogeneity, revealed not only by the traditional study of chromosomal inversions (*I*) but also by recent studies of

molecular markers such as X-linked ribosomal DNA (rDNA). So far, extensive molecular analyses have attempted to distinguish the number of isolated or semi-isolated genetic units of *A. gambiae s.s.* that exist and whether these are evolving into separate species (speciation). Elucidating the genetic population structure of the *A. gambiae s.s.* complex is a prerequisite for determining which genetic units of the complex are the vectors of malaria, and unraveling the ecological and ethological differences that are relevant to disease transmission. Such knowledge will improve our understanding of malaria epidemiology and will help in implementing appropriate vector control strategies.

Genotyping X-linked rDNA of *A. gambiae s.s.* has led to the characterization of two molecular forms (M and S) that differ in both the transcribed and nontranscribed spacers in the rDNA repeat unit (2–4). The relationship between the M and S molecular forms and the

chromosomal forms—defined according to nonrandom associations of inversions in chromosome 2 (*I*)—varies according to their ecological and geographic distribution (Fig. 1). In some areas of West Africa (for example, Mali and Burkina Faso), there is a one-to-one correspondence between the M molecular form and the Mopti chromosomal form. Similarly, the S molecular form always corresponds to the Savanna or Bamako chromosomal form (5). In other areas of West Africa, this clear correspondence breaks down (2). For example, in populations inhabiting forests or humid savannas, both molecular forms are characterized by high frequencies of the standard arrangement in chromosome 2 indicative of the Forest chromosomal form. Within the S form, a small proportion show ambiguous cytological configurations, indicating the presence of chromosome 2 arrangements typical of chromosomal forms other than Savanna and Bamako. Outside Mali and Burkina Faso, the M form may exhibit chromosomal arrangements typical of the Bissau, Savanna, or Forest forms.

<sup>1</sup>Parasitology Unit, Department of Public Health, <sup>2</sup>Department of Genetics and Molecular Biology, University of Rome “La Sapienza,” P.le Aldo Moro 5, 00185 Rome, Italy. <sup>3</sup>Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. <sup>4</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA.

\*To whom correspondence should be addressed. E-mail: ale.dellatorre@uniroma1.it; carlo.costantini@uniroma1.it