The Anopheles Genome and Comparative Insect Genomics

Thomas C. Kaufman,1 David W. Severson,2 Gene E. Robinson3

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 the Insecta, arguably the dominant group of 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 the Insecta, arguably the dominant group of 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 the Insecta, arguably the dominant group of the planet.
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, which 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 extensive 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. (15) 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 influences 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 Dfd (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.
The Mosquito Genome: Anopheles gambiae

Speciation Within Anopheles gambiae—
the Glass Is Half Full

A. della Torre,1* C. Costantini,1* N. J. Besansky,3 A. Caccone,4 V. Petrarca,2 J. R. Powell,4 M. Coluzzi1

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 (24). 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.

References
22. K. Ishihara et al., Genome Res. 10, 664 (2000).
23. D. A. Papatsenko et al., Genome Res. 12, 470 (2002).
32. A. Alabanzan et al., Development 128, 2803 (2001).