

# Flying to New Heights

Barbara R. Jasny and Floyd E. Bloom

Even the most cynical spectator of the genome races will be inspired by the accomplishments presented in this special issue. *Drosophila melanogaster* has served biology for nearly a century, since Thomas Hunt Morgan and his students selected the fly as the subject for their studies of heredity. Not only did they establish the relationships between chromosomes and genes and exploit the natural mutations they observed for early chromosomal mapping, but their followers have also repeatedly developed innovative technology for inferring gene function based on genetic analysis. Studies of *Drosophila* genetics became such an attractive subject for analysis that the organized genome project was once regarded as probably unnecessary, given the progress that had been achieved.

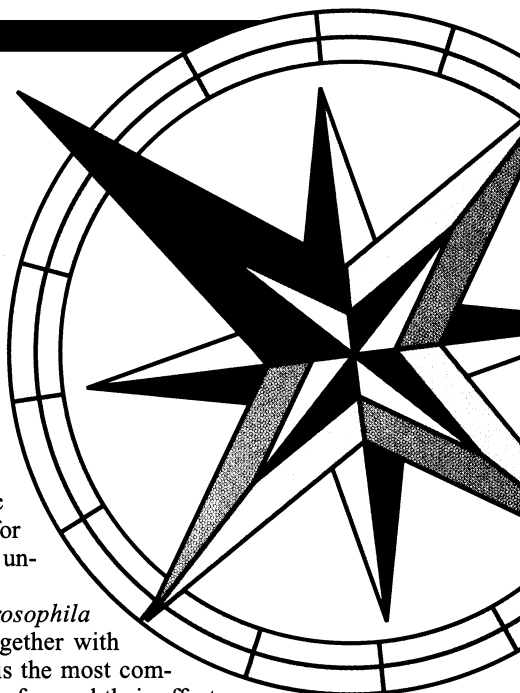
Now with the joint announcements of essential completion of the *Drosophila* sequence by the Berkeley and European *Drosophila* Genome projects, together with Celera Genomics Inc., a new phase of biological insight has begun. This is the most complex organism whose genome has yet been analyzed at this level. The teams focused their efforts on the 120 megabases of the genome sequence known to contain the most genes (the so-called euchromatic region). The early results of their analysis bring even more surprises than the most enthusiastic supporters of the effort had expected, ensuring that *Drosophila* genetics will continue to be critical to our understanding of biological processes. The similarities between *Drosophila* genes and genes involved in human physiological processes and disease are staggering. Yet *Drosophila* has only twice the number of genes as yeast and somewhat fewer than were previously noted in *Caenorhabditis elegans*. Although many have high similarity, and presumably similar function, to mammalian genes and their proteins, some two-thirds of *Drosophila* genes resist functional classification by prior knowledge.

Outside of the enormous scientific potential that will derive from having the complete inventory of genes, numerous technological and collaborative issues deserve readers' attention. The vast majority of the *Drosophila* genome sequencing was achieved by application of a refined version of the whole genome shotgun approach on a random array of 2- and 10-kilobase DNA fragments, previously applied to bacteria and viruses but never before attempted on this scale. When this challenge was first placed before the traditional sequencing community in 1998, with the added motive that *Drosophila* might serve as a pilot project for the human genome, the whole genome shotgun strategy was met with serious skepticism and indeed was declared unfeasible given the large genome size and the number of repetitive regions, which would preclude an accurate reassembly. Although there are certain to be debates over the definition of completeness, it is clear that the approach will be viable for the mammalian genomic efforts that are now well along.

In order to complement the depth of genetic inventory with an organized first guess at their functional assignments, Celera and the Berkeley *Drosophila* Genome Project instituted a ground-breaking "annotation jamboree." Forty scientists, representing some of the sharpest thinkers in *Drosophila* research and bioinformatics, were brought together for a roughly 2-week period to begin the process of extracting meaning from the nucleotide sequence. This process will undoubtedly

continue for many years to come, and new and better functional assignment tools are still needed.

Finally, the collaboration between academia and industry in this complex project sets an example that should become a new standard for the community. The combination of the long-standing intellectual fervor and knowledge base of academic research scientists with the resources, drive, and expertise of the Celera scientists created a product of higher quality and in less time (less than a year) than either side could have accomplished alone. Their joint efforts now provide the scientific community a first glimpse at some fundamental comparative and evolutionary molecular processes and pathways. The unfettered public availability of the current "Release 1" sequence data and its future refinements means that all researchers will be free to derive their own insights, applications, and functional validations. Within that cooperative framework, Celera's business model permitted advance looks at the data by its partners. This spirit of mutual cooperation was how T. H. Morgan created the *Drosophila* community and is how the process should work throughout science. For this achievement, we salute all participants.



EDITORIAL

LETTERS

ESSAYS ON  
SCIENCE AND  
SOCIETY

POLICY  
FORUMS

BOOK AND  
NEW MEDIA  
REVIEWS

PERSPECTIVES

TECH. SIGHT

REVIEWS

**dEbates!**

Respond online.  
All of Compass is now  
open to dEbates.

"[W]ith...essential  
completion of  
the *Drosophila*  
sequence...a new  
phase of  
biological insight  
has begun."