Request to sequence the genomes of *Drosophila simulans* and *Drosophila yakuba*
Submitted by: David Begun & Charles Langeley, University of California, Davis
Review date: July 2002

The genomic sequences of two species of Drosophila are currently being determined. The *D. melanogaster* genomic sequence is in the final stages of completion, and a useful draft of the *D. pseudoobscura* genomic sequence should be available by the end of 2002. This white paper requests the sequencing of the genomes of two additional species on the grounds that the data would be valuable for both annotation of the *D. melanogaster* sequence and studies of evolutionary developmental biology and population genetics.

The GRASPP was unconvinced that the sequence of genomes this close to *D. melanogaster* would actually help with computational genome analysis in a significant way. In the Panel's opinion, ~5% neutral divergence is too little to discern the difference between selected and unselected DNA. While more genomes are certainly better than one for comparative analysis, such analysis generally considers more diverged genomes than these. For example, the Cliften et al. study in yeast cited in the white paper used a variety of Saccharomyces species that are about five times more divergent from *S. cerevisiae* than either *D. simulans* or *D. yakuba* is from *D. melanogaster*. The white paper does not provide any hard evidence that would support the authors' claim that the *D. simulans* and *D. yakuba* genomic sequences would contribute comparative information that would significantly improve the annotation of the *D. melanogaster* DNA sequence. The justifications based on molecular evolutionary and population genetics considerations are, on the other hand, quite convincing. *D. simulans* and *D. yakuba* have been used for a wide variety of molecular evolutionary studies by a large and sophisticated research community. Many molecular evolutionary studies (in contrast to functional genome annotation) do generally require closely related genomes. Historically, Drosophila has been the premier model system for molecular evolutionary analysis of animals. Many of the same analyses will eventually be done on a large scale in the much larger primate genomes and such studies will be of enormous import for understanding the evolutionary history of humans. However, adequate tools for such large-scale studies are not yet available and they should be developed first in the *Drosophila* system, not in primates.

However, accepting that the sequences of the genomes of additional Drosophila species are of great biological interest, the GRASPP did not think that the arguments presented for the specific choices of *D. simulans* and *D. yakuba* were convincingly presented. Why does, for instance, the white paper request the sequences of two more species as a "package deal?" How will the combination be more valuable than the sequence of each? Are these the right species? What is the justification for them rather than other Drosophila species? Is there community consensus about these choices?

Recommendation. Defer this request to allow the authors to provide the GRASPP with answers to the questions posed and evidence that there is a broad consensus within the population genetics and molecular evolution communities *that D. simulans* and *D. yakuba* are the right choices.