

## **Proposal for the sequencing of *Drosophila yakuba* and *D. simulans***

David Begun and Charles Langley, University of California – Davis

February 2003

This is the resubmission of a white paper that the GRASPP deferred in an earlier round, proposing to sequence *Drosophila simulans* and *D. yakuba*. In the decision to defer, the Panel asked that the submitter provide a) a better justification for two species rather than one; and b) a better justification for these species rather than others, including demonstration of community consensus. Both are answered convincingly.

To answer the first request, the authors make a convincing argument that if one is studying genetic differences between *Drosophila melanogaster*, whose genome has been sequenced, and another *Drosophila* species, a third species is needed as an outgroup to determine which version of the sequence is ancestral, and in which lineage the change has occurred. The principal justification for these fly genomes is to enable the population/evolution community to study mechanisms of molecular evolutionary change. It will be essential to have a close outgroup for *simulans/melanogaster* comparisons.

The answer to the second request is not so simple, and can be a frustrating question to have to answer. Any species at the right evolutionary distance is a reasonable choice. From the standpoint of distance alone, given the phylogeny of known species the authors show in Figure 1, for the "close" fly, any of *simulans*, *mauritiana*, or *sechellia* would do; for the "outgroup" fly, any of *teissieri*, *yakuba*, *santomea*, *erecta*, or *orena* would do. The fly literature (and community) clearly favors *simulans* as the close fly; there is already an extensive literature on *simulans/melanogaster* comparison. The choice of *yakuba* is less clear. The authors point out that *D. yakuba* and *D. santomea* are a partially interfertile species pair, which may enable certain follow-on studies that would not be possible in *teissieri* or *erecta*. It is clear from the letters that the great majority of the community supports the choice of *yakuba*.

The white paper reiterates the main justification for sequencing these closely related species: *Drosophila melanogaster* comparisons with *D. simulans* and *yakuba* will be roughly comparable to human comparisons with chimp and rhesus macaque. At these close distances, one can address a large number of questions about evolutionary mechanisms. *D. simulans* and *D. yakuba* genome sequences will help draw the powerful fly evolution/population genetics community into genome-scale analysis, and by developing new tools and analyses in the more accessible *Drosophila* model system, should magnify the power of primate comparative genomics.

In the GRASPP's opinion the white paper still overstates the utility of these genomes for feature-detection purposes. Marty Kreitman states the case far more clearly in his letter of support than the authors do: the fact is that one

wantsBoth the ability to detect important features, and to study how these features evolve. For detection purposes, we have *D. pseudoobscura*; the next logical choice for feature-detection would be *D. willistoni* or *D. littoralis*. For evolution purposes, we currently have no useful genome data, and we need *simulans* and *yakuba* to fill that gap and get the fly evolution/population genetics community engaged as soon as possible.

High priority.