Drosophila White Paper-01/12/2001

(aka FlyPaper2001)

Explanatory Note: This document was originally published as the Drosophila White Paper 1999 (15January1999). Revisions to this document have incorporated comments from attendees at the Community Resources Workshop, chaired by Dr. Laurie Tompkins and which took place on Thursday, March 23, 2000 at the Drosophila Research Conference in Pittsburgh, Pennsylvania. As a follow up to this workshop, the Drosophila Community Resources Committee, a standing committee of the North American Drosophila Board, met to re-evaluate the community's needs in the context of the progress of the last two years.

This is a draft prepared by the Drosophila Community Resources Committee. It has already been circulated to the Drosophila Board and revised again. It is now being circulated to the community at large through FlyBase and through directed email. With the input of the community included, the Drosophila Community Resources Committee will submit a final report for approval by the Drosophila Board. This report will then be officially submitted to the Trans-NIH Nonmammalian Model Organism Committee, a committee including representatives of the various NIH institutes that oversees broad resource and infrastructure initiatives for nonmammalian model systems.

Please let us know your view of this document. Input from the Drosophila community is essential! Even if you fully agree with the document and have no scientific or editorial comments, it is important that we know that you agree!

If NIH has a sense that there is broad community support for this report, then it and other funding agencies will be much more sympathetic to funding the various initiatives than if it seems as if they are being proposed by a small group of individuals.

Thus, we ask and beseech you to forward your comments to our committee as soon as possible to the following email address:

FlybaseP@flybase.bio.indiana.edu

Thank you for helping the entire Drosophila community.

Drosophila Community Resources Committee: Lynn Cooley, Claude Desplan, Ulrike Gaul, Pam Geyer, Thom Kaufman, Mark Krasnow, Gerry Rubin, Bill Gelbart (chair).

The Revised White Paper

The advent of complete genomic sequences from many complex organisms has posed an important set of problems. What are all these genes doing, how do their functions interact, and how may we take advantage of the sequences to advance understanding and cure human disease? An important piece of the solution is complete genetic analysis in model organisms where the full array of genetic technologies may be brought to bear on such issues. Drosophila melanogaster is an extraordinarily attractive model organism owing to a combination of its unusually manipulable genetic system, relatively low cost, and biological complexity comparable to that of a mammal. Many organ systems in mammals have well-conserved homologues in Drosophila, and Drosophila research has already led the way in providing new insights into cancer, neurodegenerative diseases, behavior, immunity, aging, multigenic inheritance, and development. Indeed,

analysis of the Celera-BDGP genomic sequence of Drosophila melanogaster provided enormous evidence on the value of the fly as a model for human disease, with about 2/3 of human disease genes having a clear cognate in Drosophila. Together with the information on the Drosophila genome and proteome, the past years of investment in Drosophila research and the anticipated completion of the genomic sequence will catalyze an explosion in outstanding research and insights into normal and disease mechanisms if harnessed properly.

While there is no question that the investigator-initiated RO1 program in Dr osophila is as strong and vital as ever, there are clearly identifiable bottlenecks to more rapid research progress. Thus, our community agrees that to seriously meet the opportunities and challenges in Drosophila genomics and genetics, there must be targeted development of shared genetic resources such as libraries of transposon mutants in all genes, adequate databases, stock centers, complete genomic expression analyses, polymorphism databases, and related resources and enabling technologies. We believe that the successful development of these resources will benefit both the Drosophila and non-Drosophila biomedical research communities alike and catalyze a rapid wave of discovery with significant applications to human biology and disease.

The community gratefully acknowledges the efforts of the BDGP, EDGP and Celera to produce a high quality genomic sequence of Drosophila melanogaster, especially on a more rapid timetable than was originally anticipated. Further, the community is pleased that resources are in place for the BDGP to bring this sequence to finished quality. In order to take advantage of the availability of this sequence and to most fully develop the research potential of the fly in a cost-effective manner, several other resources and initiatives need to be upgraded or established. In general, these recommendations focus on additional resources that will further enrich our understanding of the Drosophila melanogaster genome, and enable more penetrating investigations of the biology of this model system.

1) Support the development of a Drosophila melanogaster cDNA Unigene set where the cDNAs are sequenced and made available in convenient cloning vectors. These cDNAs would have a wide range of uses in functional genomics. The utility of such cDNA collections for all model organisms was recognized in the last NIH/DOE five-year plan for the genome project, and the Mammalian Gene Collection (MGC) project has subsequently been established by the NIH to provide this resource for human and mouse. We ask that a similar project be carried out for Drosophila. We understand that there is already a two-year funding commitment to the BDGP for this Unigene set. We ask for $1,000,000 to fund a third year for the final stages of development of this crucial resource.

2) A significant expansion of stock center capacity for Drosophila melanogaster. This goal will require expansion of the physical space and personnel at the existing U.S. center or establishment of an additional melanogaster center to care for and send out the many genetic strains to the community. We envision that a national capacity in the range of 20,000 different stocks is a necessary minimum to accommodate the anticipated development of mutants in all genes as well as transgenic strains of broad community interest. This goal will cost approximately $500,000 per year beyond current expenditures. Current commitments provide for a collection of 10,000 stocks by 2003, at a total cost of approximately $615,000 per year.

In particular, we endorse the commitment of stock center resources to house another 10,000 unique, characterized P element insertions that will be generated by the BDGP over the next 2.5 years. These lines will go a long way toward the goal of having a genetic disruption in every transcription unit and will be an invaluable resource to the Drosophila research community. It should be noted that at the Community Resources Workshop, as well as among Resource Committee members, the need for increased stock center resources was recognized as one of the key needs of the community.

3) The database capacities available to the community must be significantly expanded. The current torrent of sequences requires annotation, linkage to the genetic maps and phenotypes, and links to databases of diversity. Additional resources to help provide access to functional genomic information on the fly will also be needed. These needs can be met by increasing central Drosophila database budgets by about $500,000

per annum for a total funding commitment of $3,000,000 per year (for the current year), rising to $3,500,000 per year over a 5 year period.

4) Obtain 6X whole genome shotgun coverage of two additional Drosophila species of sufficient evolutionary distance such that nonconserved sequences will display substantial divergence. It is our view that having a 3-way interspecific comparison of the assembled shotgun reads will provide the most robust view of sequence conservation for the experimental annotation of the Drosophila melanogaster genome, including the verification of predicted open reading frames and the identification of conserved regulatory elements and structural domains within the genome.

The first of the species should be sequenced as quickly as possible, hopefully beginning within the next year, and the second one should follow shortly thereafter. The best candidate species should be based upon considerations of genome size, evolutionary distance and available genetic information. From these considerations, the case for making Drosophila pseudoobscura the second species is extremely strong. This species has sufficient evolutionary distance from D. melanogaster so that sequence conservation will be indicative of conserved function, and the genome size is smaller than other candidate species of as great or greater evolutionary distance. Further, D. pseudoobscura has been a model organism in its own right, especially for population genetics, and thus having genome-wide sequence for the species will have some additional benefits.

The choice of the second species will require further investigation and broad community discussion. Experiments and modeling to identify the most appropriate second candidate species should be based upon considerations of genome size, evolutionary distance from both D. melanogaster and D. pseudoobscura, and on the available genetic information for that species.

Based on current costs, we estimate that 6X whole genome shotgun sequencing would cost approximately $4,000,000 per species.

(5) There is a great need for a molecular stock center to house genomic and cDNA clones, and transgenic constructs of various sorts that are in wide demand by the community. Such a resource could be housed separately for Drosophila clones, or could be part of a broader trans-organism facility.

In addition to these major resource needs, we urge NIH to consider program announcements that encourage innovative technological research or resource development in the following areas:

(a) More efficient mutational mapping technologies for gene identification, for example, based on molecular polymorphism markers such as SNPs or microsatellites.

(b) Homologous gene knockout and gene replacement technology.

(c) Improving RNAi technology.

(d) Cryopreservation technology. While "fly freezing" has had a checkered history, it is the only hope for alleviating pressure for more and more national stock center resources, and new avenues of research in this area should be encouraged.

(e) Techniques for isolating primary cells and establishment of continuous cell lines representing identified cell types, including germ cells, and techniques for organ culture.

(f) Capture of spatial expression pattern information on all Drosophila genes, including tissue and subcellular localization.

(g) Improved transgenic technology for inducible spatial or temporal gene expression.