

Meeting report

***Drosophila melanogaster*, two years into the post-genomic era**

Brian Oliver

Address: Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA. E-mail: oliver@helix.nih.gov

Published: 27 June 2002

Genome Biology 2002, **3(7)**:reports4021.1–4021.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/7/reports/4021>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

A report on the 43rd *Drosophila* Research Conference held in San Diego, USA, 10-14 April 2002.

The Berkeley *Drosophila* Genome Project (BDGP) [<http://www.fruitfly.org/>] and Celera Genomics, Inc. released the first version of the sequence of *Drosophila melanogaster* at the 41st *Drosophila* Research Conference [<http://www.drosophila-conf.org/genetics/gsa/dros/dros2000/d-00.htm>] in Pittsburgh two years ago. The genome has now had a chance to percolate through and influence the way fly labs do science. Although there was an abundance of the usual high-quality gene-by-gene analysis at the 43rd *Drosophila* Research Conference [<http://www.faseb.org/cgi-bin/droso2s-cgi/droso2>], this report focuses on the impact of the genome-sequencing landmark on *Drosophila* research and researchers. The take-home lesson is that the sequenced genome facilitates, and perhaps demands, the development of high-throughput biology.

The complete assembly of the *Drosophila* genome has not yet occurred, as is the case with many of the 'completed' genomes of multicellular organisms, but substantial progress has been made to finish the work. Completing the sequencing project beyond the present 'nearly finished' stage is under-appreciated by most, but it is valued by the handful of researchers who have run headlong into one of the previous gaps, and even more people will come to appreciate the importance of a finished product when it is time to perform 'full genome' experiments in their own labs. The gaps in 'version 1' of the sequence are being filled by the BDGP and the Baylor Human Genome Sequencing Center [<http://www.hgsc.bcm.tmc.edu/>] to a stunning degree of accuracy (one error per 100,000 bases). The difficult job of assembling the heterochromatic regions of the genome is also being successfully tackled by using a new Celera assembly, transgene insertions into heterochromatin that serve as bridges for amplifying intervening DNA, and

reverse-transcriptase-coupled PCR, which links the widely dispersed exons buried in the heterochromatin (Sue Celniker, BDGP, Berkeley, USA; Andrew Clark, Penn State University, University Park, USA). The near term prospect for having access to the complete genome is bright.

The version 1 annotation of the *Drosophila* genome relied very heavily on automated gene predictions, which are much better than nothing, but improvements are certainly welcome. Annotation teams at Berkeley and Harvard contributing to the FlyBase database [<http://flybase.bio.indiana.edu:82/>] are progressing well with the third version of this ongoing effort (Gerald Rubin's lab, BDGP, Berkeley, USA; William Gelbart's lab Harvard University, Cambridge, USA). This new annotation relies more extensively on *Drosophila* experts who take *in silico* predictions, biological evidence, and the suggestions of the *Drosophila* community into account. Reannotation occurs in the Apollo software environment, which presents gene predictions and other lines of evidence to the annotators as a visually appealing scaleable map, immediately understandable to anyone who has ever characterized a gene at the molecular level. Ultimately most of what we know and will come to know about fruitflies will need to be mapped onto the sequenced genome. A fully annotated genome is a work in progress and always will be (indeed, the current FlyBase is a logical extension of the print versions of databases that have existed as long as the *Drosophila* community itself).

Drosophila genetics has always been a high-throughput science and several labs are continuing this tradition. Many of the thousand or so genes that were listed in the abstracts for this meeting were isolated as direct results of a few high-throughput screens for embryonic lethality, female sterility, male-sterility, or maternal-effect lethality. The ongoing gene-knockout screen using the transposable element P has already made the lives of countless graduate students and postdocs easier. P elements mapping to about 10% of the predicted genes are available and are easily queried in a P-element

database called P-Screen [<http://flypush.imgen.bcm.tmc.edu/pscreen/>]; they are, most importantly, readily available from the Bloomington *Drosophila* stock center [<http://fly-stocks.bio.indiana.edu/>]. At this meeting we also learned that PiggyBac transposons are useful for insertional mutagenesis, including insertion into genes that are refractory to P-element insertion (Udo Haecker's lab, Lund University, Sweden; Ernst Wimmer's lab, Bayreuth University, Germany; Jon Margolis, Exelixis Inc., San Francisco, USA). Although libraries of PiggyBac insertion alleles are not yet widely available to the community, this situation will certainly be rectified in the near future. Malcolm Fraser (University of Notre Dame, USA), who was one of the original discoverers of PiggyBac as a transposon tool, indicated that the vector would be distributed to academic users through the PiggyBac website [<http://piggyBac.bio.nd.edu>]. Perhaps with a small stable of different transposable element types (with differing insertion biases), an off-the-shelf collection of mutations in every transcription unit is achievable.

Another promising way to map phenotypes onto the genome is systematic RNA interference (RNAi) analysis to inhibit steady-state accumulation of specific mRNAs, or even specific splice forms (Jean Antoine Lepesant's lab, Institut Jacques Monod, Paris, France). A collection of primer pairs to amplify predicted transcripts from genomic DNA that has been converted to a set of dsRNAs for automated RNAi on *Drosophila* tissue culture cells (Norbert Perrimon's lab, Harvard Medical School, Boston, USA). The effect of each of the current known transcripts can be assessed with a suitable assay (such as cell shape, specific antibody, or growth), alone or, in combination by using multiplex RNAi.

There was an explosion of interest in *Drosophila* a couple of decades ago, when molecular biologists began correlating genetics with molecular expression patterns, by northern blotting and *in situ* hybridization. Once again, gene expression is a focus, but this time on a large scale with the use of microarray data. A major push is needed to improve access to arrays and array content, however. Predictably, most of the labs using microarrays are large and well funded, from the microarray incubator at Stanford (pioneered for *Drosophila* by Kevin White, now at Yale University, New Haven, USA), aligned with a commercial collaborator, or both. It is clear that many more labs would like the opportunity to add microarrays to their research toolbox. The community appears to be mobilized at the international level to see this happen. Steve Russell (Cambridge University, UK) announced the creation of the International *Drosophila* Array Consortium [<http://www.indac.net/>], to promote community-wide array distribution and platform development; and Justen Andrews (Indiana University, Bloomington, USA) reported that Indiana University, home of the *Drosophila* stock center, will seek funding for a Genomic Resources Center. Hopefully, momentum evident at the meeting will be maintained.

Finally, and surely this is the way of the future, several groups are beginning to deploy multiple high-throughput methods, such as cataloging quantitative (array) data and the spatial expression patterns (*in situ* hybridization) of all known *Drosophila* genes in a single searchable database (Gerald Rubin's lab). Chris Doe's lab (University of Oregon, Eugene, USA) described a mind-boggling set of interlinked experiments to do the following: mine the fly genome for transcription-factor binding sites; generate array data for that transcription factor; use *in situ* hybridization to determine spatial regulation of targets; and deploy comparative genomics to look for conserved organization of genes and sites in the human genome.

The *Drosophila* community will always have room for 'boutique' work focusing on a very detailed understanding of a single gene. But the *Drosophila* community is also continuing a proud tradition of high-throughput analysis that will be a boon within and beyond the world of the *Drosophila* worker: large datasets made possible by the genome project are helping to annotate the sequence with information and, more importantly, knowledge. Because a large data set is best analyzed with another large data set, look forward to an ever-expanding description of the life and times of the humble fruitfly.

Acknowledgements

Thanks to the members of the *Drosophila* community for discussions, and to Vaijayanti Gupta, Alan Kimmel, and Michael Parisi for comments on the manuscript.