VIEWPOINT

# A Brief History of *Drosophila*'s Contributions to Genome Research

**Gerald M. Rubin[1] and Edward B. Lewis[2]**

**The sequence of the *Drosophila melanogaster* genome presented in this issue of *Science* is the latest milestone in nine decades of research on this organism. Genetic and physical mapping, whole-genome mutational screens, and functional alteration of the genome by gene transfer were pioneered in metazoans with the use of this small fruit fly. Here we look at some of the instances in which work on *Drosophila* has led to major conceptual or technical breakthroughs in our understanding of animal genomes.**

In 1910, T. H. Morgan, having chosen *Drosophila* for his studies of heredity, was rewarded with the first of many mutants, a white-eyed fly. Morgan was soon joined in the famous Fly Room at Columbia University by three principal students, A. H. Sturtevant, C. B. Bridges (see Fig. 1), and H. J. Muller. Within the space of 5 years, they formulated a revolutionary chromosome theory of heredity (*1*). Their accomplishments, which led to Morgan winning the Nobel Prize in 1933, are all the more remarkable because their sole experimental method was to do controlled crosses with these mutants and count progeny.

In 1913, Sturtevant constructed the first genetic map and showed that genes are arranged in a linear order (*2*). In two papers published in 1914 and 1916, Bridges, exploiting chromosome nondisjunction in XXY females, provided the elegant first proof that chromosomes must contain genes; this ruled out the alternative possibility, assumed by some at the time, that chromosomes and genes were separate hereditary elements (*3*). In 1918, Muller introduced the use of balancers, chromosomes bearing inversions that allow the stable maintenance of lethal mutations as heterozygotes in a manner that does not require selection (*4*); this is only now becoming possible in the nematode and is still not possible in mice.

The physical mapping of genes has its roots in the discovery by Heitz and Bauer in 1933 of salivary gland polytene chromosomes in the fly *Bibio hortulanus* (*5*). Polytene chromosomes could easily be seen in the microscope because, after numerous rounds of replication, the chromosomes remained aligned and were patterned in cytological bands. T. S. Painter at the University of Texas promptly realized their importance and in 1934 published the first drawings of *Drosophila melanogaster* polytene chromosomes, which included the chromosomal localization of several genes (*6*). In 1935 and 1938, Bridges published polytene maps of such accuracy that they are still used today (*7*). Making extensive use of chromosomal rearrangements, Bridges also constructed cytogenetic maps that assigned genes to specific sections and even specific bands (see Fig. 2A). We now know that these maps are often accurate enough to place genes within intervals of less than 100 kb.

In 1927, Muller showed that ionizing radiation causes genetic damage and that mutations, including chromosomal rearrangements, be induced with x-rays, a finding for which he received a Nobel Prize in 1946 (*8*). In the late 1930s, two groups demonstrated the feasibility of generating deficiencies and duplications by combining x-ray–induced chromosomal aberrations with closely spaced break points (*9*). This method was systematically exploited by D. L. Lindsley, L. Sandler, and 14 co-workers in 1970 to generate an ordered set of duplications and deletions spanning the major autosomes in ~500-kb segments (*10*). This work initiated the concept of whole-genome scanning in metazoans for phenotypic perturbations; such a resource has never been duplicated for any other metazoan. It is interesting to note that this genome-wide effort occupied a higher percentage of the total *Drosophila* research community at the time than has the current genome project.

The foundation for modern genome research can be traced to a grant application (*11*) written in 1972 by D. S. Hogness of Stanford University. Anticipating the first successful cloning of eukaryotic DNA a year later (*12*), Hogness proposed using large insert clones to construct physical maps of whole chromosomes to facilitate the detailed study of chromosome structure (Fig. 3). The first random clones of any organism were generated in the Hogness laboratory in early 1974, and a cloned DNA segment was mapped to a specific chromosomal location a few months later (*13*) (see Fig. 2B). By early 1975, clone libraries representing the entire genome had been generated (*14*) and screened for clones carrying specific sequences (*15*) with the newly developed method of colony hybridization (*16*). Overlapping segments of chromosomal DNA cloned in bacteriophage lambda (*17*) covering more than 200 kb were constructed by "chromosome walking" by the end of 1978 (*18*, *19*). An inversion that linked this region to the Bithorax complex of homeobox genes was used to achieve the first positional cloning of a gene, Ultrabithorax, in early 1979 (*18*, *20*). By late 1980, many mutant alleles had been located on the restriction map of the complex and shown to be the result of chromosomal breakage or transposable element insertion (*21*) (see Fig. 4).

In 1980, C. Nusslein-Volhard and E. Wie-


**Fig. 1.** (**A**) Bridges (left) and Sturtevant in 1920. (**B**) Morgan in 1917. The photo of Morgan, who was camera shy, was taken by Sturtevant using a camera hidden in an incubator and operated remotely by means of a string. The books and microscope in the background were at Sturtevant's desk (*1*). Both photos courtesy of the Archives, California Institute of Technology.

[1]Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720–3200, USA. [2]Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

schaus extended to animals the use of a systematic genome-wide mutational screen to attempt to identify all genes involved in a fundamental process (*22*), a feat that had previously been attempted only in microorganisms. Their work on embryonic development soon led to the discovery of the components of most major signaling pathways, as a new generation of fly workers were eager to isolate and sequence the genes defined in their screen using the techniques of positional cloning and transposon tagging (*23*). In recognition of this work, Nusslein-Volhard and Wieschaus shared the 1995 Nobel Prize.

An important breakthrough for manipulating the genome was made in late 1981 when methods for making transgenic flies with the use of transposable element vectors were developed and used to achieve the first rescue of a mutant phenotype in an animal by gene transfer (*24*). The availability of stable, single-copy, integrative transgenesis enabled a range of powerful techniques to be developed in *Drosophila*, many of which have since been adapted to other metazoans. These methods include the use of enhancer traps to screen for genes based on their pattern of expression, developed in 1987 (*25*), large-scale insertional mutagenesis with engineered transposable elements, developed in 1988 (*26*), site-specific recombination for generating chromosomal rearrangements, developed in 1989 (*27*), and two-component systems for controlling ectopic gene expression, developed in 1993 (*28*).

Ironically, the success in cloning and studying individual genes dampened enthusiasm for an organized genome project, which was seen as unnecessary. Over 1300 genetically characterized genes—nearly 10% of all the genes in *Drosophila*—have been cloned and sequenced by individual labs (*29*). This is over twice the percentage of genes in any other animal for which both the loss-of-function phenotype and sequence have been determined. Nevertheless, for flies (*30*) as well as other animals (*31*), less than a third of genes have obvious phenotypes when mutated, emphasizing the critical importance of genome sequencing as a gene discovery method.

The annotated sequence of the *Drosophila* genome reported in this issue (*32*) is the product of both publicly and privately funded efforts and is the first application of the whole-genome shotgun approach (*33*) to the sequencing of an animal genome. It provides a model for the large-scale annotation of a genomic sequence, which was accomplished through the concerted efforts of 40 experimental and computational biologists from 20 institutions in five countries. These sequencing and annotation efforts follow the collaborative tradition of *Drosophila* research established over 80 years ago; as observed by J. Schultz, "it derives from Morgan, and paradoxically has not so much to do with cooperation as with the paramount importance

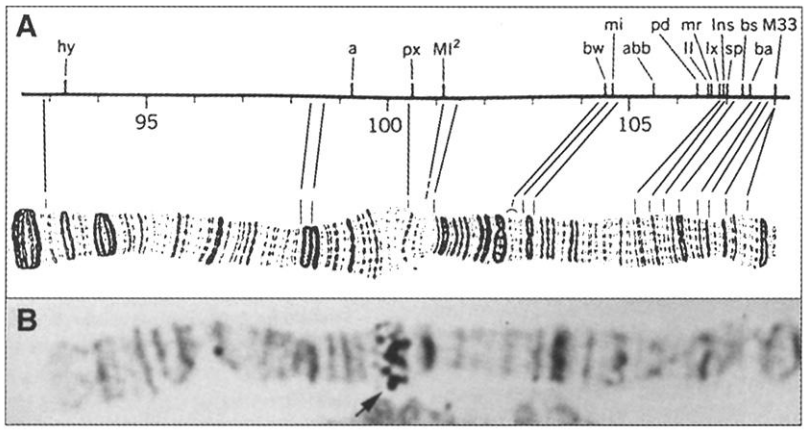


**Fig. 2.** (**A**) Corresponding points in the polytene chromosome map and the linkage map for the tip of the second chromosome [modified from (*35*)]. The region shown covers about 5 Mb of DNA. (**B**) In situ hybridization (*36*) of a cloned segment of *Drosophila* DNA to polytene chromosomes, demonstrating the first mapping of a cloned gene to its chromosomal location [modified from (*13*)].

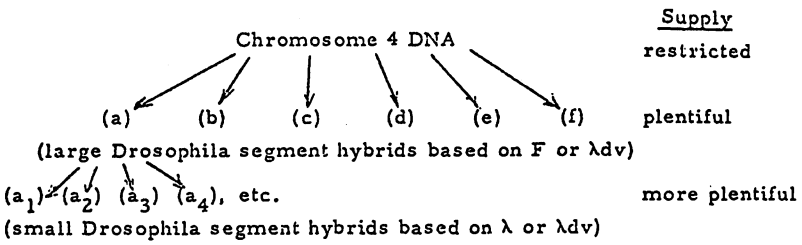


**Fig. 3.** Diagram taken from D. Hogness's 1972 grant application (*11*) showing his proposed strategy for making a physical map of a whole chromosome, starting with ordering large insert clones based on the F factor [now known as bacterial artificial chromosomes (BACs)] and then subcloning each of these into bacteriophage lambda or plasmid vectors. "One could then obtain a set of overlapping segments covering all the DNA in the chromosome, and the overlaps between segments could be detected and mapped.... In this way, many of the sophisticated physical techniques can be applied in an ordered manner to specific segments of a *Drosophila* chromosome" (*11*).

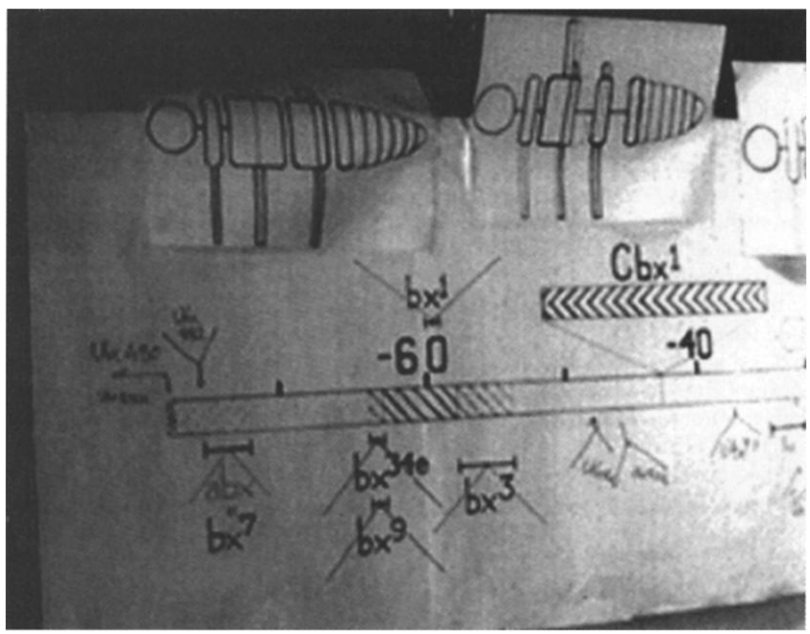


**Fig. 4.** Poster displaying a partial map of the Bithorax complex displayed at the Stanford University Biochemistry Department retreat at Asilomar, California, in late 1980. Note the molecular mapping of various mutant alleles relative to the scale in kilobases derived from the restriction map of the cloned region.

attached to getting on with the work. I cannot recall any instance of explicit discussion of the value of cooperation; it was always taken for granted, and taught by example" (*34*).

**References and Notes**


1. A. H. Sturtevant, *Am. Sci.* **53**, 303 (1965); A. H. Sturtevant, *A History of Genetics* (Harper and Row, New York, 1965), pp. 1–167; H. L. K. Whitehouse, *Towards an Understanding of the Mechanism of Heredity* (Arnold, London, ed. 2, 1969), pp. 1–447.
2. A. H. Sturtevant, *J. Exp. Zool.* **14**, 43 (1913).
3. C. B. Bridges, *Science* **40**, 107 (1914); *Genetics* **1**, 1 (1916).
4. H. J. Muller, *Genetics* **3**, 442 (1918).
5. E. Heitz and H. Bauer, *Z. Zellforsch.* **17**, 67 (1933).
6. T. S. Painter, *J. Hered.* **25**, 464 (1934).
7. C. B. Bridges, *J. Hered.* **26**, 60 (1935); *Genetics* **23**, 142 (1938).
8. H. J. Muller, *Science* **66**, 84 (1927); *J. Genet.* **22**, 299 (1930); _____ and W. S. Stone, *Anat. Rec.* **47**, 393 (1930).
9. N. V. Dubovsky and L. V. Kelstein, *Bull. Biol. Med. Exp. URSS* **6**, 733 (1938); J. T. Patterson, M. S. Brown, W. Stone, *Univ. Texas Publ.* **4032**, 167 (1940).
10. D. L. Lindsley *et al.*, *Genetics* **71**, 157 (1972).
11. D. S. Hogness, grant application GB-25769-1 submitted to the National Science Foundation on 24 May 1972 and awarded 25 August 1972.
12. J. F. Morrow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1743 (1974).
13. P. C. Wensink, D. J. Finnegan, J. E. Donaldson, D. S. Hogness, *Cell* **3**, 315 (1974).
14. G. M. Rubin and D. S. Hogness, unpublished data.
15. D. J. Finnegan, G. M. Rubin, D. S. Hogness, unpublished data.
16. M. Grunstein and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961 (1975).
17. T. Maniatis *et al.*, *Cell* **15**, 687 (1978).
18. W. Bender, personal communication.
19. See also W. Bender, P. Spierer, D. Hogness, *J. Supramol. Struct.* **8**, 32 (1979); P. Spierer, A. Spierer, W. Bender, D. S. Hogness, *J. Mol. Biol.* **168**, 35 (1983).
20. See also W. Bender, P. Spierer, D. S. Hogness, E. B. Lewis, *Annu. Rep. Biol. Calif. Inst. Technol.* **1980**, 165 (1980).
21. D. S. Hogness, W. W. Bender, M. E. Akam, R. Saint, P. Spierer, *J. Supramol. Struct. Cell. Biochem.* **5**, 385 (1981); W. Bender *et al.*, *Science* **221**, 23 (1983).
22. C. Nusslein-Volhard and E. Wieschaus, *Nature* **287**, 795 (1980).
23. P. M. Bingham, R. Levis, G. M. Rubin, *Cell* **25**, 693 (1981).
24. A. C. Spradling and G. M. Rubin, *Science* **218**, 341 (1982); G. M. Rubin and A. C. Spradling, *Science* **218**, 348 (1982).
25. C. O'Kane and W. J. Gehring, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9123 (1987).
26. L. Cooley, R. Kelley, A. C. Spradling, *Science* **239**, 1121 (1988).
27. K. G. Golic and S. Lindquist, *Cell* **59**, 499 (1989).
28. A. H. Brand and N. Perrimon, *Development* **118**, 401 (1993).
29. FlyBase, *Nucleic Acids Res.* **27**, 85 (1999) (flybase.bio.indiana.edu/).
30. M. Ashburner *et al.*, *Genetics* **153**, 179 (1999).
31. The *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).
32. G. M. Rubin *et al.*, *Science* **287**, (2000); M. D. Adams *et al.*, *Science* **287**, 2185 (2000); E. W. Myers, *Science* **287**, 2196 (2000).
33. J. C. Venter *et al.*, *Science* **280**, 1540 (1998).
34. From a letter written by Jack Schultz to George Beadle in 1970, recalling his days as a student in the Columbia fly lab (Jack Schultz papers, American Philosophical Society Library, Philadelphia, PA).
35. C. B. Bridges, *Cytologia* (Fujii Jubilee Volume) (1937), p. 745.
36. M. L. Pardue, D. D. Brown, M. L. Birnsteil, *Chromosoma* **42**, 191 (1973).


37. We thank our many colleagues for sharing their recollections, for checking facts in their old lab notebooks, and for helpful comments on the manuscript. Supported by grant HD06331 to E.L. from NIH.

VIEWPOINT

# The *Drosophila* Genome Sequence: Implications for Biology and Medicine

**Thomas B. Kornberg[1] and Mark A. Krasnow[2]**

**The 120-megabase euchromatic portion of the *Drosophila melanogaster* genome has been sequenced. Because the genome is compact and many genetic tools are available, and because fly cell biology and development have much in common with mammals, this sequence may be the Rosetta stone for deciphering the human genome.**

The genome sequence of the fruit fly *Drosophila melanogaster* reported in this issue is a landmark achievement that marks the end of a century of gene hunting and heralds a new era of exploration and analysis. It is the second and largest animal genome sequenced (*1*), containing ~180 million base pairs (Mbp), of which most of the 120-Mbp euchromatic, gene-rich portion has now been determined (*2*). The importance of this accomplishment stems in part from the monumental technical feat it represents and the swiftness with which it was completed as a combined academic and industry effort. The foundation was laid by the Berkeley, European, and Canadian *Drosophila* Genome Projects, which contributed a detailed chromosomal map and 28 Mbp of sequence. The remaining 75% of the sequence was obtained this past year in a collaboration between Celera Genomics Group and the Berkeley *Drosophila* Genome Project.

Three million short (~500 bp) sequence reads were made from the ends of random genomic fragments, and overlaps between the obtained sequences were used to assemble the nearly complete sequence of the four *Drosophila* chromosomes. This random ("shotgun") strategy had not previously been attempted for genomes so large and complex, because repeated sequences hundreds to thousands of base pairs long scattered throughout the genome cause ambiguities in assembly. The solution for this was to obtain sequences from both ends of fragments that were ~2, 10, and 150 kb in length (*3*). These oriented bits of sequence were assembled into increasingly dense and interlinked scaffolds that ultimately generated long continuous stretches of chromosome sequence with few gaps or ambiguities. An estimated 2% of euchromatin remains unfinished; it is thought to be mostly repeat-dense regions that border heterochromatin and are difficult to assemble. The success of this strategy with *Drosophila* is encouraging for a similar combination of directed and shotgun sequencing to elucidate larger and more complex genomes, including the human genome, which is nearly 30 times larger than *Drosophila*.

Beyond the technical achievement, the importance of the *Drosophila* sequence rests partly on the role this fly has played in the history of experimental biology. Even more significant is the accelerated rate of discovery it will catalyze in new areas of *Drosophila* biology important for human biology and medicine.

## *Drosophila* as a Model Animal

Throughout the last century, the fly has been the workhorse for genetic studies in eukaryotes. These studies provide the basis of much of our conceptual understanding of fundamental aspects of eukaryotic genetics, including the chromosomal basis of sex determination, genetic linkage, and chromosomal mechanics and behavior (*4*). *Drosophila* now has a wealth of mutants, and many special chromosomes that have been endowed with visible and molecular markers and other properties that facilitate genetic manipulations. These tools enable saturating genome screens directed to the isolation of a broad spectrum of visible and lethal phenotypes, even ones that are manifested in the $F_2$ or $F_3$ generations of mutagenized individuals. Transposon-based methods for manipulating genes have also been developed, all made possible because the P transposon can be modified and stably integrated into the chro-

[1]Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA. [2]Howard Hughes Medical Institute and Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA.