The Drosophila Genome Sequence: Implications for Biology and Medicine
Thomas B. Kornberg, *et al.*
Science **287**, 2218 (2000);
DOI: 10.1126/science.287.5461.2218

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 G. M. Rubin and D. S. Hogness, unpublished data.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 D. J. Finnegan, G. M. Rubin, D. S. Hogness, unpublished data.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.

mosomes. These allow creation of genetically defined, stable lines with regulated transgenes and efficient production of genetic mosaics, techniques not available in other animals, including *Caenorhabditis elegans* (*5*). These tools are invaluable for genetic and developmental studies. Transposon-based technologies have also been used to screen for lethal mutants with tissue- and cell-specific phenotypes and to screen for gene expression patterns ["enhancer trap" screens (*6*)] in live animals at all stages of development. With this technical arsenal for manipulating gene content, many different ways to identify and analyze genes and genetic interactions in a developmentally and behaviorally complex animal can be exploited.

The molecular cloning and functional analysis of *Drosophila* genes has made it possible to assemble a molecular outline of many cellular and developmental processes. Moreover, these advances have provided entries into studies of the corresponding processes in mammals. Cloned *Drosophila* genes have led to the identification of mammalian cognates, and to an extent no one predicted, many of these cognates have closely related functions in mammals. This includes transcription factors and their regulatory targets, structural proteins, chromosomal proteins, ion channels, and signaling proteins. Because so many basic cellular functions are conserved, attention is drawn to the ones that differ significantly, such as the absence of cytoplasmic intermediate filaments (*7*). Conservation also extends to higher order processes, such as development, behavior, sleep, and physiological response to drugs such as alcohol (*8*), as well as to neurodegeneration (*9*). Homeobox genes were one of the earliest examples of genetic conservation between the fly and mammals. More recently, the pathways controlling development of limbs, nervous system, eyes, and the heart, as well as complex systemic interactions such as circadian rhythms and innate immunity, have also been found to be conserved (*10*). These discoveries of conservation in the underlying control pathways cast doubt on long-held views that eyes, body segmentation, and circadian rhythms arose independently in flies and mammals and are products of convergent evolution. Instead, there appears to be a fundamental unity that makes the fly an important model for many aspects of mammalian biology.

Implications of *Drosophila* Genome Sequence for *Drosophila* Workers

Fly genetics started in 1910 with the discovery in T. H. Morgan's laboratory of a spontaneous mutant with white eye color. Progress accelerated following the discovery of radiation, chemical, and insertional mutagenesis. The current era has witnessed a broad array of systematic screens for phenotypes in development, fertility, behavior, longevity, learning, and drug susceptibility. With the advent of molecular cloning techniques, roughly 2500 *Drosophila* genes have been molecularly defined, and an even larger number have been genetically characterized. With the full complement of ~14,000 *Drosophila* genes now revealed, practitioners can begin to develop methods to ascertain the functions of the genes whose phenotypes are unknown or have not yet been linked to a particular complementation group. Remarkably, an organism as complex as the fly has only twice the number of genes as the unicellular yeast *Saccharomyces cerevisiae*. The relatively small gene complement in *Drosophila*, with few duplicated genes, makes it a streamlined animal genome ideal for genetic analysis.

The analysis of *Drosophila* genes currently underway in many labs will be facilitated by the availability of the full genome sequence. Transposon-induced mutations can now be mapped simply by obtaining a short stretch of genomic sequence flanking the insertion site; sequence polymorphisms between strains can be readily identified for recombinational mapping of chemically induced mutations. Whereas a complete set of loss-of-function mutants like the one being constructed for *S. cerevesiae* is clearly needed, efficient methods for targeted mutagenesis in *Drosophila* have yet to be found. The current strategy is to generate and map 10^5 P-element insertions (*11*). This is the best approach available, and it is expected to generate a collection of mutants representing 80% of all genes within 3 years. Nevertheless, transposon mutagenesis lacks the efficiency and precision of targeted gene replacement, and the full benefit of the *Drosophila* genome sequence will not be realized until efficient methods for directed mutagenesis are found. In addition, an easy way to freeze mutant and transgenic strains is needed in order to preserve the genetic bounty and to relieve the current burden of maintaining thousands of fly stocks in continuous culture.

It is critical to begin linking the new *Drosophila* genes with biochemical pathways. The availability of the full sequence will speed this effort by adding genomic approaches to the time-honored genetic and molecular ones. In *Drosophila*, mRNAs can be readily localized in whole mounts by in situ hybridization, as can proteins by immunohistochemical methods. With the full sequence, the temporal and spatial expression patterns of all genes can now be defined: in situ hybridizations can be undertaken with a full set of gene probes and DNA microarrays containing all predicted *Drosophila* genes can be constructed to allow massive parallel analysis of gene expression (*12*). Genes with related patterns of expression will be revealed by such studies, and the interdependence of their expression can be determined. The sequence can also be used to make reagents to localize any *Drosophila* protein and to define protein-protein interactions. Novel approaches to epistasis studies, interaction networks, and protein functions will likely be developed as well.

Implications of *Drosophila* Genome Sequence for Biology and Medicine

The conservation of biological processes from flies to mammals extends the influence of *Drosophila* to human health. When a *Drosophila* homolog of an important but poorly understood mammalian gene is isolated, the arsenal of genetic techniques in the *Drosophila* system can be applied to its characterization. The *Drosophila* gene's developmental expression pattern, loss-of-function phenotype, and overexpression phenotype can be analyzed to elucidate gene function. Other genes that function in the same pathway can be identified among genes with similar mutant phenotypes or expression patterns, or among mutant genes that enhance or suppress its phenotype. Pathways can be proposed on the basis of genetic epistasis studies. Isolating mammalian homologs of newly identified genes in the *Drosophila* pathway can elucidate the corresponding mammalian pathway. In this way, the power of *Drosophila* genetics has been leveraged to elucidate mammalian pathways involved in cancer biology, the cell cycle, and receptor tyrosine kinase (RTK/RAS) signaling.

The full genome sequence identifies many new candidates for such approaches. Chief among these are *Drosophila* cognates of human disease genes, especially those whose functions are not understood. Finding the *Drosophila* homolog will no longer take months or years of uncertain searching by molecular methods. Homologs of tau and Parkin involved in Parkinsonism, the *p53* tumor suppressor gene, the menin gene in multiple endocrine neoplasia type 1, and many other disease genes are now revealed by the genome sequence. In this vein, the long-sought insulin homologs have finally been found, as have those for receptors for somatostatin, vasopressin, leutotropin, thyroid stimulating hormone, and other hormones.

Looking to the Future

There are many aspects of *Drosophila* biology and physiology waiting to be explored by genetics and the new genomic approaches (*13*). We expect the future of *Drosophila* research to turn increasingly to questions beyond the cellular level, to questions of physiology, maintenance, and regeneration of

whole organs, and to systemic processes. For example, some human renal disorders are associated with defects in genes involved in fluid and electrolyte transport. *Drosophila* orthologs of these genes found in the genomic sequence should spur studies of the physiology and function of Malpighian tubules, which serve as the *Drosophila* "kidney." We anticipate that new collaborations between vertebrate and fly researchers will come about to study behavior, neurodegeneration, aging, and drugs and that important new biological principles and pathways will emerge.

Research is subject to strong selective pressures. Experimental systems that offer the most efficient and direct access to important questions attract the most attention and effort, and as techniques evolve and interests mature, the landscape will change rapidly. The *Drosophila* sequence is a critical resource that ensures that this tiny dew-lover will continue to lead the way to new biological pathways and principles. If *Drosophila* has been difficult for workers in other fields because of an arcane nomenclature and idiosyncratic husbandry, the sequence now provides access through a universal language—the DNA sequence.

References

1. The *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).
2. M. D. Adams *et al.*, *Science* **287**, 2185 (2000); R. A. Hoskins *et al.*, *Science* **287**, 2271 (2000).
3. E. W. Myers, *Science* **287**, 2196 (2000).
4. G. M. Rubin and E. B. Lewis, *Science* **287**, 2216 (2000).
5. T. Xu and G. M. Rubin, *Development* **117**, 1223 (1993).
6. H. J. Bellen *et al.*, *Genes Dev.* **3**, 1288 (1989).
7. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
8. R. D. Riddle and C. Tabin, *Sci. Am.* **280**, 74 (February 1999); J. C. Hendricks *et al.*, *Neuron* **25**, 1299 (2000); M. S. Moore *et al.*, *Cell* **93**, 997 (1998).
9. G. R. Jackson *et al.*, *Neuron* **21**, 633 (1998); J. M. Warrick *et al.*, *Cell* **93**, 939 (1998); K.-T. Min and S. Benzer, *Science* **284**, 1985 (1999).
10. M. Tessier-Lavigne and C. S. Goodman, *Science* **274**, 1123 (1996); J. A. Hoffmann *et al.*, *Science* **284**, 1313 (1999); R. Bodmer and M. Frasch, in *Heart Development*, R. P. Harvey and N. Rosenthal, Eds. (Academic Press, San Diego, CA 1999), pp. 65–90; W. J. Gehring and K. Ikeo, *Trends Genet.* **15**, 371 (1999); J. C. Dunlap, *Cell*, **96**, 271 (1999); A. P. McMahon, *Cell*, **100**, 185 (2000).
11. A. C. Spradling *et al.*, *Genetics* **153**, 135 (1999).
12. G. M. Rubin *et al.*, *Science* **287**, 2222 (2000); K. P. White *et al.*, *Science* **286**, 2179 (1999).
13. G. A. Kerkut and L. I. Gilbert, *Comprehensive Insect Physiology, Biochemistry, and Pharmacology* (Pergamon, New York, 1985); V. B. Wigglesworth, *The Principles of Insect Physiology* (Chapman & Hall, London, 1939).

VIEWPOINT

From Sequence to Chromosome: The Tip of the *X* Chromosome of *D. melanogaster*

Panayiotis V. Benos,[1] Melanie K. Gatt,[2,11] Michael Ashburner,[1,2] Lee Murphy,[3] David Harris,[3] Bart Barrell,[3] Concepcion Ferraz,[4] Sophie Vidal,[4] Christine Brun,[4] Jacques Demailles,[4] Edouard Cadieu,[5] Stephane Dreano,[5] Stéphanie Gloux,[5] Valerie Lelaure,[5] Stephanie Mottier,[5] Francis Galibert,[5] Dana Borkova,[6] Belen Minana,[6] Fotis C. Kafatos,[6] Christos Louis,[7,8] Inga Sidén-Kiamos,[7] Slava Bolshakov,[6,7] George Papagiannakis,[7] Lefteris Spanos,[7] Sarah Cox,[7] Encarnación Madueño,[9] Beatriz de Pablos,[9] Juan Modolell,[9] Annette Peter,[10] Petra Schöttler,[10] Meike Werner,[10] Foteini Mourkioti,[10] Nicole Beinert,[10] Gordon Dowe,[10] Ulrich Schäfer,[10] Herbert Jäckle,[10] Alain Bucheton,[4] Deborah M. Callister,[11] Lorna A. Campbell,[11] Areti Darlamitsou,[11] Nadine S. Henderson,[11] Paul J. McMillan,[11] Cathy Salles,[11] Evelyn A. Tait,[11] Phillipe Valenti,[11] Robert D. C. Saunders,[11,12] David M. Glover[2,11]

One of the rewards of having a *Drosophila melanogaster* whole-genome sequence will be the potential to understand the molecular bases for structural features of chromosomes that have been a long-standing puzzle. Analysis of 2.6 megabases of sequence from the tip of the *X* chromosome of *Drosophila* identifies 273 genes. Cloned DNAs from the characteristic bulbous structure at the tip of the *X* chromosome in the region of the *broad* complex display an unusual pattern of in situ hybridization. Sequence analysis revealed that this region comprises 154 kilobases of DNA flanked by 1.2-kilobases of inverted repeats, each composed of a 350–base pair satellite related element. Thus, some aspects of chromosome structure appear to be revealed directly within the DNA sequence itself.

Fewer than 90 years have elapsed since Alfred H. Sturtevant presented the world with the first-ever genetic map of six visible markers on the *X* chromosome of *Drosophila* (*1*). Now that the sequence of almost the entire euchromatic genome of *Drosophila* has been determined (*2*), we have the opportunity to study the function of each gene. In addition, we can study the relation between DNA sequence and chromosome structure.

The European *Drosophila* Genome Project (EDGP) (*3*) has determined the sequence of a contiguous segment of DNA extending some 2.6 Mb from the tip of the *X* chromosome, that is, from subdivision 1A to subdivision 3C on the standard polytene chromosome map (*4*). We predict the existence of 273 protein-coding genes in this region (one gene every 9.6 kb), which is of some sentimental, as well as much scientific, interest to geneticists. It extends from a position 120-kb distal to the *yellow* locus to 150-kb proximal to the *white* locus, whose mutation was the first clearly visible mutation found in *Drosophila* and whose study led to the discovery of sex-linked inheritance and, hence, to the proof of the chromosome theory of heredity (*5*).

In the region sequenced, we have identified 17 transposon insertions. The most common element was *roo* (six copies), but six other retroviral-like elements, two LINE-like elements, an element with inverted repeat ends (*S*-element), and a foldback (FB) element were also found. The overall density of insertions (one insertion every 155 kb) is similar to that in the *Adh* region [one every 170 kb (*6*)]. This is of some interest because the tip of the *X* is a region of low genetic recombination and, on theoretical grounds, might have been expected to accumulate transposable elements. However, it has been suspected that transposon insertion might reduce fitness. Because the *X* chromosome is hemizygous in the male fly, it is subject to stronger selection pressures. This would lead to the prediction of a lower frequency of insertions on the *X* (*7*). Our finding that the overall transposable element densities in the *X* tip and region 35 to 36 are comparable argues against the maintenance of element copy number by negative selection.

The first physical map of any genome was the description of the polytene chromosomes of *Drosophila* (*8*). These chromosomes arise from endoreduplication resulting in a large number of parallel fibers with each fiber rep-