Toward Cloning and Mapping the Genome of *Drosophila*

J. MERRIAM, M. ASHBURNER, D. L. HARTL, F. C. KAFATOS

An ultimate goal of *Drosophila* genetics is to identify and define the functions of all the genes in the organism. Traditional approaches based on the isolation of mutant genes have been extraordinary fruitful. Recent advances in the manipulation and analysis of large DNA fragments have made it possible to develop detailed molecular maps of the *Drosophila* genome as the initial steps in determining the complete DNA sequence.

THE WALL CHART INCLUDED IN THIS ISSUE OF *SCIENCE* describes the current status of the clone map in *Drosophila melanogaster*. Many overlapping molecular clones of specific chromosome sections have been isolated. When this map is complete the DNA of each chromosome will be available as a set of defined clones stretching from one chromosomal end to the other. Coordinating individual clones to rearrangements and marker genes (that is, creating a physical map) will be a major advance toward making full use of the cloned genome to detect and analyze additional genes.

Genome projects are under way for many organisms (*1, 2*). Their aim is to provide detailed molecular genetic information about fundamental biological processes. It is at this level that *Drosophila* plays a major, if not unique, role. It has already been thoroughly analyzed genetically; mutations defining thousands of genes have been isolated (*3, 4*) and the phenotypes of these mutations are usually accessible in detail at the level of single cells (*5*). While many of the tools for functional analysis are specific for *Drosophila*, such as polytene chromosomes, the results of the analysis will have profound value for all biology because of the universality of cellular and developmental mechanisms.

Many genetic systems, such as the cellular responses to heat shock and to steroid hormones were originally discovered in flies and are now recognized as important in the vertebrates. At the level of genes, the homeobox coding sequences represent a class of fundamentally important genes that was discovered by work on *Drosophila* (*6*). The genes *boss* and *sev*, which control the differentiation of one of the photoreceptor cells in the eye, form one of the best examples of ligand-receptor interactions in developmental biology (7). By discovering all the genes of a complex but genetically tractable organism such as *Drosophila*, we achieve access to a "functional language" common to all biology.

J. Merriam is a professor in the Department of Biology, University of California, Los Angeles, CA 90024. M. Ashburner is a professor in the Department of Genetics, Cambridge University, Cambridge, U.K. CB2 3EH. D. L. Hartl is a professor in the Department of Genetics, Washington University, St. Louis, MO 63110. F. C. Kafatos is a professor in the Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138 and Institute of Molecular Biology and Biotechnology, Forth, Heraklion Crete, Greece.

Drosophila mutations are central to functional analyses of genes for which clones, but not mutations, are available in vertebrates. The usual approach is to use the vertebrate sequence as a probe to obtain a similar sequence from flies. The goal is then to map the *Drosophila* sequence and either correlate the gene with an existing mutation or select new mutations in the sequence to demonstrate its function. A recent example is the effort to map *Drosophila* genes coding molecules similar to kinesin (which was originally identified in the vertebrates) and to correlate these with *Drosophila* cell division mutations (*8*). Genes that are redundant or have overlapping function may not show an altered phenotype in the vertebrates when only one of the genes is mutant, and may require simultaneous mutations in two or more genes to be detected. It is much easier to make flies homozygous for mutations in two genes simultaneously than it is with other organisms (*9*).

Because of the power of *Drosophila* genetics, researchers have been able to devise screens capable of recognizing all the mutations that affect a biological process (*10*) or are located within a chromosome segment (*11*). A bottleneck for further analysis has been the time required to clone the genes for which mutations have been induced chemically or with ionizing radiation. The immediate goal of a physical map of the *Drosophila* genome is to provide strategies for rapid cloning through knowledge of gene location. This will be enhanced by the detailed mapping of molecular and genetic markers including rearrangement breakpoints (junction fragments), at both the chromosomal and molecular scales.

Description of the Genome

The size of the *D. melanogaster* genome is estimated to be 165 Mb (or 165,000 kb) (*12*) compared with 3000 Mb for the human genome. Estimates for the number of genes range from about 5,000 (on the basis of the number of lethal mutations) to more than 15,000 (on the basis of the number of transcription units). The actual number of genes will not be determined until the entire genome has been sequenced. The clones generated to cover the genome, as described here, are reasonable starting points for such a project.

Mitotic chromosomes, DNA families, euchromatin, and heterochromatin. One advantage of *Drosophila* as an organism for molecular cloning is its relatively small genome size and low chromosome number. The largest *Drosophila* chromosome is about the size of the smallest human chromosome (*13*). The diagrammatic appearance of the haploid set of four chromosomes is presented in Fig. 1. Each chromosome arm (except chromosome 4) has a length of about 1.5 μm in mitotic preparations, and consists of terminal euchromatin and a pericentric block of heterochromatin (the latter accounts for about 25% of each chromosome). These distinctions are important because most of the genes are located in the euchromatin (*14*). Moreover, because of the problem of cloning the heterochromatin,

which consists of repetitive DNA, a complete contig of overlapping clones will only span one chromosome arm from telomere to the boundary of the repetitive DNA. The complete molecular map should consist of six contigs; five of similar sizes and one much smaller.

The heterochromatin blocks are further divided into two distinct regions: "α" or heterochromatin forming the larger block near the centromere, and "β" heterochromatin forming a somewhat diffuse boundary with euchromatin (*15*). Such distinctions in structure are the consequence of the different families of DNA found in each location. According to the renaturation kinetics, 18 to 21% of the DNA consists of α heterochromatin—highly repeated, simple sequences found in satellite DNA (*16*). About 9 to 12% of the DNA is moderately repeated, formed by members of repeated gene families such as the histone genes and the ribosomal RNA genes, as well as copies of different sequences termed mobile elements because neither their location nor number of copies are fixed in the genome (*17*). β heterochromatin is formed from moderately repeated sequences although most of the moderately repeated sequences are located in the euchromatin.

Drosophila is unusual in that the copy number of most repeated gene families is low, and a high proportion of moderately repeated sequences (more than half) consist of mobile elements. There are at least 50 kinds of mobile elements with an average of 50 copies each in the genome (*18*). Unlike mammals, there are no highly repetitive small interspersed nuclear element (SINE) like sequences such as Alu, in the euchromatin. The distribution of repeated sequences is also different in *Drosophila*, which follows a "long period" interspersion pattern with an average distance of >5 kb between repeated elements (*19*). About 64 to 67% of the genome is single-copy according to renaturation kinetics; this forms the bulk of the euchromatin. Less well described is the proportion of the genome present as very rapidly reannealing sequences ("snap-back"). These may form up to 6% of the genome, and are present in euchromatin in part as the "fold-back" class of mobile elements (*20*).

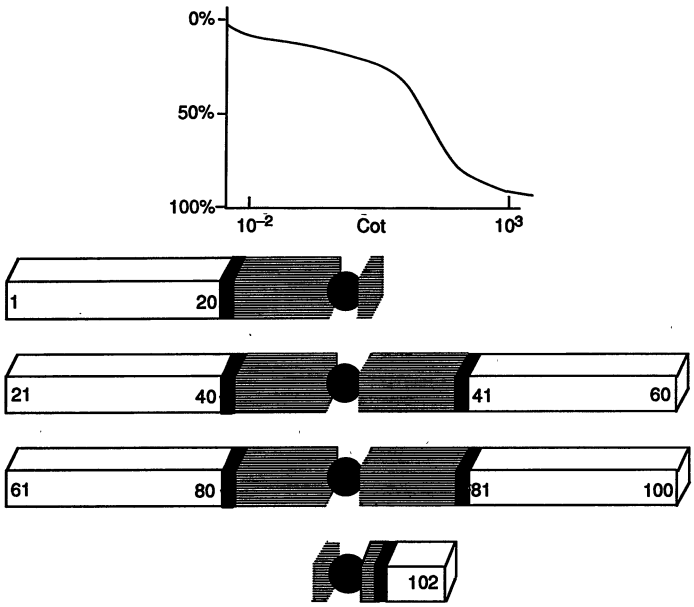

Fig. 1. Typical renaturation kinetics of single-stranded DNA (top) and the structure of a haploid set of *Drosophila melanogaster* chromosomes (bottom) highly repeated DNA, striped; euchromatin (single-copy and some moderately repeated DNA), white; transition border between euchromatin and α heterochromatin, stippled. In mitotic preparations, the chromosomes may be told apart on the basis of slight size differences and constrictions not indicated here. The X and 4th chromosomes appear to be acrocentric (subtelocentric) or single armed. The 4th chromosome, labeled 102, is shown here disproportionally larger than its "dot" size. Although the centromeres (filled circles) appear as distinct entities here they form part of the single continuous DNA molecule making up each chromosome (*13*). Both of the major autosomes are metacentric, with arms about the same size as the X. With five arms of approximately the same size it is customary to use arm designation (such as 2R) for location as if they were separate elements. Each arm is labeled to indicate centromere and telomere positions on the wall chart although the divisions are not distinct on mitotic chromosomes. X (1, 20); 2L (21, 40); 2R (41, 60); 3L (61, 80); 3R (81, 100); 4 (102).

Polytene chromosomes. "Giant," or polytenic chromosome arms as long as 400 μm are found in certain dipteran tissues (for example, salivary glands, Malpighian tubules, fat body cells, and nurse cells). These tissues undergo terminal differentiation accompanied by up to ten rounds of DNA replication without mitosis. The interphase sister chromatids are held together and precisely aligned, like many parallel strands to long ropes. Only the euchromatin and β heterochromatin polytenize, that is, replicate to form the long rope-like arms. The α heterochromatin (including the Y chromosome) does not replicate, instead melting into a diffuse "chromocenter" from which the polytenized arms project. Each polytene arm is "banded" with local concentrations of chromatid coiling.

These chromosomes offer both cytogenetic and molecular resolution unsurpassed in any organism. The banding patterns are unique and provide a reliable high resolution map of the chromosomes. Bridges (*21*) and Painter (*22*) provided the first physical map of the genome when they demonstrated from chromosome rearrangements that the polytene banding map is co-linear with the genetic map obtained by recombination data. Bridges (*21*) established a coordinate system to indicate location on the polytenized map, which is still used. The entire genome is divided into 102 sections called divisions; each division is further subdivided into six lettered subdivisions, within which distinct bands are numbered sequentially. The total number of bands is about 5100 (*23*). The size of the polytenized sections is estimated to be about 110,000 kb (*23*).

Salivary gland chromosomes are ideal substrates for in situ hybridization with labeled nucleic acid probes (*24*). The resolution has been improving; initially localizations to the level of the lettered subdivision (average size 200 kb) were common. With biotin-labeled probes, resolution to a band (average size 20 kb) is frequently possible (*25*).

Cloning Strategies

Cloning genes. How can a euchromatic genome of more than 100,000 kb be cloned and the clones arranged in an ordered array? Even for flies this would not be feasible as a single walking project at the usual walk rate of 20 kb per month in phage clones (*26*). Nevertheless, almost 1300 clones isolated from λ or plasmid libraries have been recorded on the basis of their salivary chromosome locations (*27*). Most of these clones describe genes, although some are of anonymous regions of the DNA. More than a third are genes defined by their product rather than by mutation. As listed in Table 1, more clones have been isolated on the basis of sequence similarity to other clones, or through oligonucleotides, or through some form of transcription than by their location.

Because of the resolution of in situ hybridization, the option of cloning *Drosophila* genes on the basis of their location near a previously available clone or a chromosome rearrangement is now becoming more common. Examples include the "walk" to *Ubx* (*28*) and the isolation of *per* (*29*). Numerous strategies have been implemented to speed up the walk process. These include "jumping" between breakpoints of a chromosome rearrangement (*26*), microdissection of the desired region (*30*), and "transposon tagging" from nearby mobile elements (*31*).

Table 1. The cloning techniques used to isolate a sample of DNA clones (*27*).

Method of isolating clones	Number
Expression screen/differential transcript screen	181
Sequence similarity probes	218
Oligonucleotide probes	30
Transposon tag	97
Walk or jump	125
Micro clone	36
Direct RNA isolation, chance, other	595

From "transposon tagging" to the "enhancer trap". Control over P element mobility has been the basis for transposon mutagenesis screens (*32*). Adding markers to the P elements has resulted in a new approach for identifying genes through activity. Germline transformation by P mobile element vectors is reviewed by Spradling (*33*). O'Kane and Gehring (*34*) started the procedure to introduce P elements tagged with the *Escherichia coli lacZ* gene into the genome. Active transcript units are identified through the influence of enhancers near the insertion site that drive lacZ transcription. P elements are usually marked with other genes to allow the individual elements to be followed in crosses (*35*). Many thousands of inserts have been examined by this "enhancer trap" technique, uncovering both known and previously unknown genes (*36*). An important aspect of the technique is that it can detect genes that are not otherwise detected by mutagenesis, for example when the gene is redundant or mutations have no phenotype. In light of the threefold difference between the number of lethals and the number of transcription units, this problem of genes that mutate without phenotypic effect is an important one that otherwise limits *Drosophila* genetics.

Large DNA technology. An early advance was the development of cosmid vectors, capable of carrying inserts of 33 to 47 kb (*37*). Electrophoretic methods for the separation of even larger DNA molecules were pioneered with pulsed field gradient gel electrophoresis (*38*). The subsequent development of methods for cloning large DNA molecules in yeast artificial chromosomes (YACs) (*39*) or in bacteriophage P1 (pacmids) (*40*) has made it feasible to undertake the molecular analysis of complex genomes as large as that of *Drosophila* or of chromosomes the size of those in humans or the mouse.

The physical mapping of large DNA molecules initially focused on bacteria, yeast, and nematodes (*1, 2, 41*). A "bottom-up" strategy, with λ or cosmid clones was used, in which large numbers of clones were analyzed by restriction mapping or fingerprinting in order to arrange overlapping clones into contigs through the recognition of shared DNA fragments. The ability to clone large DNA fragments allows a "top-down" strategy, which relies on the large DNA inserts that can be cloned in YACs to obtain comprehensive coverage of the target genome with fewer clones (*39, 42*). The approaches have offsetting strengths and weaknesses. Bottom-up mapping provides ease of library construction, recombinant clone purification, manipulation, and detailed characterization of the clone; the main drawback is that the relatively small size of the cloned DNA fragments puts a practical limit on the size of the initial contigs (usually two to three times the size of the insert). Top-down mapping gives greater genome coverage with fewer clones, but at the cost of greater difficulty in library construction, lower yields of target DNA from the recombinant clones, and the necessity of subcloning before detailed characterization (*43*). The pacmid cloning system yields inserts intermediate in size between cosmids and YACs (*40, 44*), and in theory it combines many of the technical advantages of alternative cloning systems; its advantages and disadvantages remain to be established in practice.

Distribution of Clones, Genes, and Rearrangement Breakpoints

The distribution of clones is presented on the wall chart and listed by division in Table 2. A detailed picture of sections 1 and 2 is presented in Fig. 2. These data combine the results of both bottom-up and top-down mapping strategies. The former is represented by the large number of clones in chromosome walks and the cosmids (individual or in contigs) that have been mapped to the polytene chromosomes (*45*); the latter is represented by the YAC clones (*42, 46*). The cosmid map is being assembled in sections, by means of probes that are microdissected from polytene chromosomes and amplified by the polymerase chain reaction (PCR) (*47*). Thus far, nearly complete coverage has been achieved within several divisions of the X, approximately 50% in the X as a whole, and approximately 10% in the autosomes. The *Drosophila* DNA in the YAC clones that have been mapped to the polytene chromosomes is equivalent to about 2.5 times the euchromatic genome. Assuming that all euchromatic sequences have an equal chance of representation in the YACs, then about 92% of the euchromatic sequences should be included at least once among the mapped clones. Although their overlaps have not been established directly, theoretical calculations (*48*) suggest that the number of YAC contigs is about 100, the average contig size is greater than 900 kb, and the longest contig is about 4 Mb. The data displayed on the wall chart suggest that about 100,133 kb, or 91% of the polytenized genome has been cloned, counting YACs, cosmids, and walks (*49*).

To maximize the value of the physical map of the *Drosophila*

Table 2. Genetic and molecular information for each numbered chromosome division. Kb indicates DNA sizes (*23*); % indicates the estimated proportion cloned to date (see Fig. 2 for discussion); and RNA indicates the lengths (in kilobases) known to be transcribed. Genes summarizes the updated data of Ashburner (*4, 50*). Insert lists the number of localized, marked P or H mobile elements (*32*) or transposable elements of Ising (*52*). Break lists the number of chromosome rearrangement breakpoints (*51*).

Div	Kb	%	RNA	Genes	Insert	Break	Div	Kb	%	RNA	Genes	Insert	Break
1	1125	100	46.2	100	15	478	52	980	80	24.6	17	14	90
2	1255	99	51.4	99	12	260	53	1290	87	7.4	15	14	64
3	1400	100	49.7	96	7	867	54	1000	74	18.7	25	11	56
4	1340	97	30.1	52	10	194	55	1310	80	13	27	8	76
5	1240	86	6.1	57	4	124	56	1520	99	13	34	12	114
6	1140	57	12.1	45	5	77	57	1560	93	20.1	84	11	203
7	1670	95	54.7	160	7	231	58	1050	84	4.7	11	12	109
8	1230	100	9.9	60	4	118	59	1290	93	8.8	29	8	194
9	1460	95	0	48	13	129	60	1820	97	44.2	39	25	242
10	1500	63	37.6	102	8	144	61	1140	89	18.7	40	30	155
11	1240	92	13.6	93	2	154	62	1390	89	14.7	26	16	126
12	1160	100	9.4	83	21	152	63	1110	94	19.4	25	15	76
13	1260	90	30.5	38	7	65	64	1370	95	28.4	53	15	152
14	900	98	16.5	51	3	140	65	1300	83	21.7	22	23	89
15	640	87	13.6	37	3	120	66	1640	96	16.5	25	25	149
16	990	50	14	46	5	117	67	1330	96	18.9	58	13	136
17	980	91	15.5	35	9	187	68	1150	81	7.7	63	28	118
18	965	94	4.7	38	11	106	69	900	79	4.5	28	5	66
19	855	100	16.7	58	12	275	70	1260	97	17.4	30	22	155
20	320	100	10.1	28	9	837	71	920	100	9	17	16	106
21	1140	91	22.6	40	12	307	72	810	100	5.5	22	4	104
22	990	100	6.1	31	9	331	73	920	98	23.3	29	14	103
23	760	89	6.9	9	10	113	74	690	100	12.8	7	4	33
24	800	96	14.2	30	10	177	75	1140	100	12.2	9	24	110
25	1280	85	19.3	64	15	242	76	1040	90	4.5	9	11	56
26	890	100	9.2	20	6	154	77	910	87	17.1	23	13	53
27	890	96	8.7	13	5	96	78	990	76	3.1	8	12	45
28	870	98	15.6	32	10	102	79	850	96	15.7	9	16	64
29	690	49	9.3	21	6	113	80	320	98	3.5	13	3	492
30	970	89	12.9	17	18	114	81	190	21	0.8	12	5	189
31	910	44	8.7	47	4	82	82	910	96	0.4	4	21	66
32	760	94	0.5	12	5	95	83	1000	86	11.2	6	16	114
33	970	100	9.1	10	5	98	84	1290	100	45.5	100	26	365
34	1140	100	1	35	13	297	85	1850	81	45.1	47	30	212
35	1040	93	12.7	58	11	568	86	1320	86	6.5	32	27	159
36	1150	96	35.9	54	15	173	87	1690	89	38.7	90	23	384
37	1000	92	16.1	72	7	128	88	1330	100	46.3	28	29	171
38	950	100	2.6	5	16	122	89	1460	100	26.7	40	31	389
39	750	86	3.1	27	14	107	90	810	100	14.1	34	15	111
40	370	81	11.4	29	8	606	91	1010	95	10.1	24	18	134
41	740	100	0	16	4	481	92	1180	97	13.2	24	30	177
42	1140	100	17.7	59	20	141	93	1260	95	46	48	29	128
43	1150	74	7.7	33	12	172	94	1310	90	6	18	21	129
44	1010	100	18.2	19	10	121	95	1210	84	24.7	25	20	83
45	760	97	2.2	2	18	98	96	1660	98	22.1	48	17	180
46	860	86	5.8	14	13	48	97	1150	73	15.2	21	17	162
47	1280	97	9.2	25	21	72	98	1090	100	26.8	25	12	170
48	900	88	11.6	27	14	81	99	1200	100	58.6	49	21	140
49	1010	92	30	64	16	197	100	1210	100	42.6	24	13	114
50	1110	98	8.5	6	12	127	101	260	23	0	6	1	342
51	1020	100	1.6	33	5	68	102	1290	32	5.4	6	2	109
							Sum	110650	90.5	1715.9	3758	1359	17970

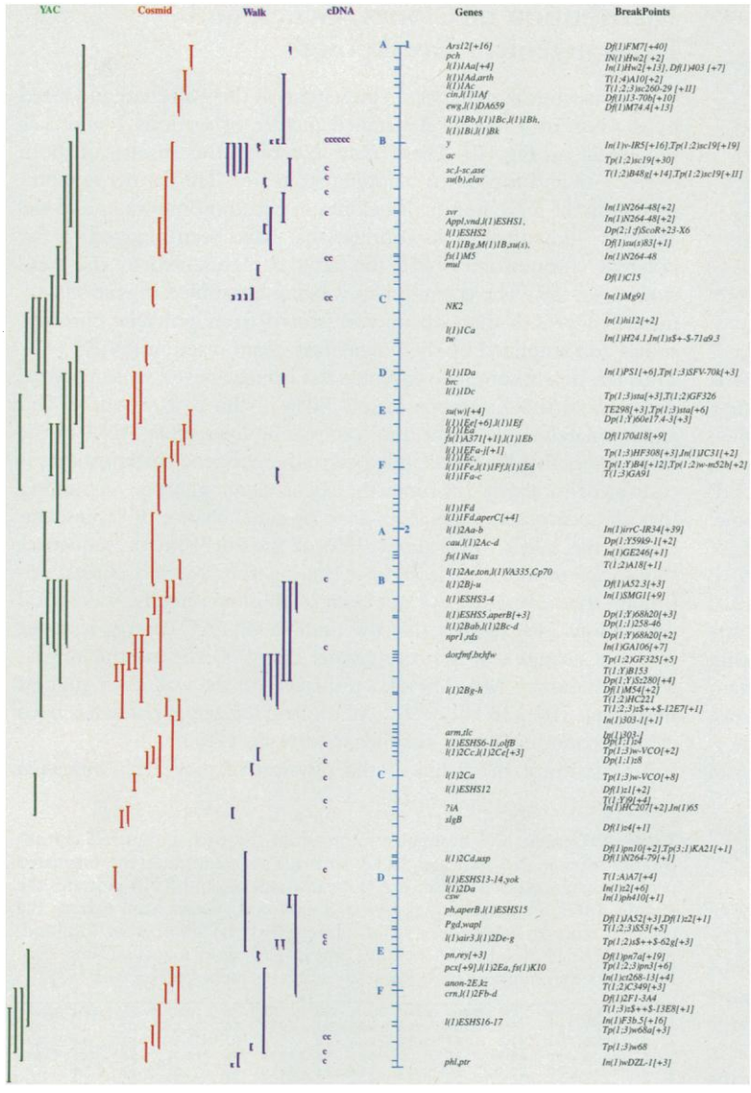

Fig. 2. Detail of DNA clone, gene and rearrangement breakpoint locations for the 2380 kb forming the X chromosome terminal sections 1 and 2. The light blue vertical axis in the middle shows the relative sizes of each lettered subdivision and of the bands (separated by horizontal tics within each lettered subdivision). DNA values were as described (*23*). The column to the immediate right of the DNA map lists genes (not necessarily sequenced) by their locations; at the far right are rearrangement breakpoints. The (+*n*) designation indicates the number (*n*) of additional genes or breakpoints that could not be listed at that location because of space limits. Genes listed are updated from (*50*); the list of rearrangements was compiled from (*51*). The columns on the left of the DNA map show the approximate sizes and locations of YACs, cosmids, and walks. These data are the same as the wall chart except that 23 walks each less than 40 kb and 27 cDNA clones not included on the wall chart are also shown here (*27*). Apparent overlaps between clones are mostly tentative as only a few have been located relative to each other. Clones localized only to the subdivision will appear to be "stacked" with common endpoints at the upper, or left side of the subdivisions (*49*). A tic mark at the upper (or left) end of a clone line indicates the left end was specified precisely; on the lower (or right) end a tic indicates the right end was specified; a thin space indicates that neither end was specified.

euchromatin as a resource, the map should be richly annotated with the locations of genes, coding sequences, chromosome breakpoints, transposon insertion sites, and all other kinds of genetic markers. Table 2 lists the numbers of genes, insertions, and chromosome rearrangement breakpoints that are located in each division. Approximately 3700 genes have already been recognized and mapped cytologically or by recombination (*50*). Most have not yet been identified in clones. The number of insertions localized to each division refers to marked P or H mobile elements (*32*) or the TE elements of Ising (*52*). Chromosome rearrangements are especially valuable for mapping because they link three different maps: the genetic recombination map, the cytogenetic chromosome map through their breakpoints, and the molecular map through their junction fragments. Almost 18,000 breakpoints have been reported throughout the genome. An additional 1500 breakpoints in the Y chromosome were not listed in Table 2 (*51*). The number of strains of flies bearing rearrangements is lower because many rearrangements are broken at more than one euchromatic location.

Included also in Table 2 are current (as of 1 August 1991) summaries of how much of the DNA length in each division is known to be transcribed. These minimal estimates were obtained from the UCLA database (*27*) by adding the longest lengths of the mature transcripts for different cloned genes in the same division. A total of 537 genes (not counting transfer RNA genes) produce an aggregate 1714 kb in transcripts; the average size per mature transcript is about 3.2 kb, which is larger than expected. Mapping transcripts and complementary DNA probes will surely be increasingly used as a method of annotating the molecular map.

Evolutionary Considerations

Evolutionary considerations have always been important in *Drosophila* genetics, in part because the heyday of *Drosophila* genetics in the 1930s coincided with development of the modern synthesis of evolutionary theory that combined Darwinian natural selection with Mendelian heredity (*53*). This synthesis was made possible in large part by the analysis of polymorphisms in the banding patterns of the salivary gland chromosomes within species and by comparisons of the banding patterns between species (*54*). Virtually every important concept in population genetics and evolution has been influenced to some extent by studies of natural or laboratory populations of *Drosophila* (*55*), and the importance of *Drosophila* in evolutionary studies shows no sign of decreasing emphasis shifts to the analysis of DNA sequences.

The technology developed for genome projects has applications in

evolutionary studies. Among the aspects of genome evolution that have not yielded to conventional molecular biology are those involving the organization of large tracts of DNA, including centromeric regions, telomeric regions, heterochromatin, or other levels of chromosome structure exceeding a few hundred kilobases. In some cases the limitation results from difficulty in cloning the sequences, but in other cases the sequences of interest are simply too long to be isolated and manipulated in conventional cloning systems. Although more is known about the sequence structure of the centromeric heterochromatin in *Drosophila* than about any other species (*14, 16, 17*), the limits of past technology have been a problem. It is a challenge to extend these limits with new large DNA molecule technology to the megabase level to uncover more information about the genome organization and evolution of these important and interesting regions.

Note added in proof: Hoheisel *et al.* described high-density filters made from three genomic libraries, a jumping library, and two cDNA libraries. The order of clones is established by hybridization fingerprinting protocols; oligomers produce partial sequence information.

REFERENCES AND NOTES

1. Including *Escherichia coli* [Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987)], yeast [M. V. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986)], the nematode *Caenorhabditis elegans* [A. Coulson, J. Sulston, S. Brenner, J. Karn, *ibid.*, p. 7821, A. Coulson, R. Waterston, J. Kiff, J. Sulston, Y. Kohara, *Nature* **335**, 184 (1988)], the mustard weed *Arabidopsis thaliana* [report of the NSF *Arabidopsis* genome research project, NSF bulletin 91-60, Washington, DC (1991)], and mice [reviewed by N. G. Copeland and N. A. Jenkins, *Trends Genet.* **7**, 113 (1991)].
2. The term "contig," defined by A. Coulson *et al.* [in (*1*)], is widely used in genome projects. Contigs are clusters of linking clones; the first goal of genome projects is to assemble a contig from end to end for each chromosome.
3. C. B. Bridges and K. S. Brehme, *The Mutants of* Drosophila melanogaster (Publ. 552, Carnegie Institution of Washington, Washington, DC, 1942); D. L. Lindsley and E. H. Grell, *Genetic Variations of* Drosophila melanogaster (Publ. 627, Carnegie Institution of Washington, Washington, DC, 1968); D. L. Lindsley and G. Zimm, *The Genome of* Drosophila melanogaster (Academic Press, San Diego, CA, in press).
4. M. Ashburner, Drosophila *Inf. Serv.* **69** (1991).
5. G. M. Rubin, *Science* **240**, 1453 (1988); D. B. Roberts, Ed., Drosophila: *A Practical Approach* (IRL Press, Oxford, 1986); M. Ashburner, Drosophila, *A Laboratory Handbook* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
6. See recent reviews by E. M. De Robertis, G. Oliver, and C. V. E. Wright, *Sci. Am.* **263**, 46 (July 1990); M. Affolter, A. Schier, W. J. Gehring, *Curr. Biol.* **2**, 485 (1990); M. D. Goulding and P. Gruss, *Curr. Sci.* **1**, 1088 (1989); M. Akam, *Nature* **349**, 282 (1991); and C. Wright and B. Hogan, *ibid.* **350**, 458 (1991).
7. H. Kramer, R. L. Cagan, S. L. Zipursky, *Nature* **352**, 207 (1991).
8. S. A. Endow and M. Hatsumi, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4424 (1991); H. McDonald and L. B. S. Goldstein, *Cell* **61**, 991 (1990).
9. For examples, see T. Elkins, M. Hortsch, A. J. Bieber, P. M. Snow, C. S. Goodman, *J. Cell Biol.* **110**, 1825 (1990); F. B. Gertler, R. L. B. Bennett, M. J. Clark, F. M. Hoffmann, *Cell* **58**, 103 (1989); F. B. Gertler, J. S. Doctor, F. M. Hoffmann, *Science* **248**, 857 (1990).
10. C. Nusslein-Volhard and E. Wieschaus, *Nature* **287**, 795 (1980).
11. B. H. Judd, M. W. Shen, T. C. Kaufman, *Genetics* **71**, 139 (1972).
12. E. M. Rasch, H. J. Barr, R. W. Rasch, *Chromosoma* **33**, 1 (1971).
13. R. Kavenoff and B. H. Zimm, *ibid.* **41**, 1 (1973).
14. Reviewed by M. L. Pardue and W. Hennig, *Chromosoma* **100**, 3 (1990); see also R. H. Devlin, B. Bingham, B. T. Wakimoto, *Genetics* **125**, 129 (1990); M. J. Healy, R. J. Russell, G. L. Miklos, *Mol. Gen. Genet.* **213**, 63 (1988); G. L. Miklos, M.-T. Yamamoto, J. Davies, V. Pirrotta, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2051 (1988); J. Tomkiel, S. Pimpinelli, L. Sandler, *Genetics* **128**, 583 (1991); P. Dimitri, *ibid.* **127**, 553 (1991); M. G. Hearn, A. Hedrick, T. A. Grigliatti, B. T. Wakimoto, *ibid.* **128**, 785 (1991), and references therein.
15. E. Heitz, *Biol. Zentralbl.* **54**, 588 (1934).
16. A. R. Lohe and D. L. Brutlag, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 696 (1986).
17. A. C. Spradling and G. M. Rubin, *Annu. Rev. Genet.* **15**, 219 (1981).
18. D. J. Finnegan and D. H. Fawcett, in *Oxford Surveys on Eukaryotic Genes*, N. MacLean, Ed. (Oxford Univ. Press, Oxford, 1986), vol. 3, pp. 1–62.
19. J. E. Manning, C. W. Schmid, N. Davidson, *Cell* **4**, 141 (1975).
20. R. Lovering, N. Harden, M. Ashburner, *Genetics* **128**, 357 (1991).
21. C. B. Bridges, *J. Hered.* **26**, 60 (1935); *ibid.* **29**, 11 (1938).
22. T. S. Painter, *Science* **78**, 585 (1933); *Genetics* **19**, 175 (1934); *J. Hered.* **25**, 465 (1934).
23. V. Sorsa, *Chromosome Maps of* Drosophila (CRC Press, Boca Raton, FL, 1988), vols. I and II.
24. M. L. Pardue, in D. B. Roberts (*5*), p. 111.
25. Under exceptional circumstances, resolution of fluorescent probes has been reported as less than 10 kb [M. C. Rykowski, S. J. Parmelee, D. A. Agard, J. W. Sedat, *Cell* **54**, 461 (1988)].
26. W. Bender, P. Spierer, D. S. Hogness, *J. Mol. Biol.* **168**, 17 (1983).
27. J. Merriam, S. Smalley, A. Merriam, B. Dawson, Drosophila *Inf. Serv.* **63**, 173 (1986); J. Merriam, S. Adams, G. Lee, D. Krieger, in *Genetic Maps*, S. J. O'Brien, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 5, 1990); electronic access to periodically updated tables by remote FTP from IUBIO.BIO-.INDIANA.EDU in the FLY subdirectory—since 1983 a database on cloned *Drosophila* genes has been maintained at the University of California, Los Angeles.
28. W. Bender *et al.*, *Science* **221**, 23 (1985).
29. M. Roshbash and J. C. Hall, *Cell* **43**, 3 (1985).
30. F. Scalanghe, E. Tuvco, J. E. Edstrom, V. Pirotta, M. Melli, *Chromosoma* **82**, 805 (1981).
31. P. M. Bingham, M. G. Kidwell, G. M. Rubin, *Cell* **29**, 995 (1982).
32. Over 1300 inserts have been mapped by in situ hybridization (Table 2); most are available from the Bloomington Stock Center. Although the distribution of marked P elements is broad, there are some limits in expecting transposon mutagenesis to identify every transcript unit. The P element tends to insert near an apparent consensus site that is usually located at the 5′ region of some, but not all, genes [W. R. Engels, in *Mobile DNA*, D. Berg and M. Howe, Eds. (American Society for Microbiology, Washington, DC, 1989), pp. 437–484]. Several hot spots of insertion have been noted (several labs isolating the same gene from independent insertions). Only about 10% of insertions are associated with mutant phenotypes (*35*), but excision of an insert is sometimes imprecise and results in local deletions, which frequently are associated with a mutant phenotype. H (for Hobo) mobile elements can also be used for germline transformation and transposon tagging, and they may insert near different chromosome sites [R. K. Blackman, M. M. Koehler, R. Grimaila, W. M. Gelbart, *EMBO J.* **8**, 211 (1989). It may become a useful complement to P elements for a more complete coverage of the genome].
33. A. C. Spradling, in D. B. Roberts (*5*), p. 175.
34. C. O'Kane and W. J. Gehring, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9123 (1987).
35. L. Cooley, R. Kelley, A. Spradling, *Science* **239**, 1121 (1988); reviewed by L. Cooley, C. Berg, and A. C. Spradling [*Trends Genet.* **4**, 254 (1988)].
36. U. Grossniklaus, H. J. Bellen, C. Wilson, W. J. Gehring, *Development* **107**, 189 (1989); C. Wilson *et al.*, *Genes Dev.* **3**, 1301 (1989); H. J. Bellen *et al.*, *ibid.*, p. 1288; H. J. Bellen *et al.*, *J. Physiol. (London)* **84**, 33 (1990); E. Bier, L. V. Jan, Y. N. Jan, *Genes Dev.* **4**, 190 (1990); L. Fasano *et al.*, *Cell* **64**, 63 (1991).
37. J. Collins and B. Hohn, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4242 (1978).
38. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984); G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984); D. Vollrath and R. W. Davis, *ibid.* **15**, 7865 (1987).
39. D. T. Burke, G. F. Carle, M. V. Olson, *Science* **236**, 806 (1987).
40. N. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 103 (1990).
41. J. P. Boucle *et al.*, *J. Mol. Biol.* **154**, 21 (1982).
42. D. Garza, J. W. Ajioka, D. T. Burke, D. L. Hartl, *Science* **246**, 641 (1989).
43. *Drosophila* chimeric YACs account for about 10% of the clones; most of these can be detected by hybridizing to more than one location on salivary gland chromosomes.
44. D. A. Smoller, D. Petrov, D. L. Hartl, *Chromosoma* **100**, 487 (1991).
45. I. Siden-Kiamos *et al.*, *Nucleic Acids Res.* **18**, 6261 (1990); reviewed by F. Kafatos *et al.*, *Trends Genet.* **7**, 155 (1991).
46. J. W. Ajioka *et al.*, *Chromosoma* **100**, 495 (1991).
47. R. D. C. Saunders *et al.*, *Nucleic Acids Res.* **17**, 9027 (1989).
48. M. J. Palazzolo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
49. We prepared Fig. 2 and the positioning of the clones on the wall chart on a Macintosh II computer using a program, "GENEDRAW," written in Think C by G. Lee (unpublished). This program uses data on clone sizes and chromosome locations available from the UCLA database (*27*) to position the clones along the projected molecular map. It was designed to estimate what fraction of the genome has been cloned after allowing for potential overlaps.
50. The data on genes are from the 1 August 1991 update of (*4*) and are available electronically as in (*27*) and from the European Molecular Biology Laboratory Netserver and SEQNET.
51. Included are deficiencies (4605 breakpoints listed), inversions (4436 breakpoints), translocations (5618 breakpoints), duplications (1979 breakpoints), and transpositions (1332 breakpoints). These data are compiled by M. Ashburner (unpublished) and are available electronically as in (*50*).
52. The data on inserts were compiled by the UCLA database and by the Bloomington Stock Center at Indiana University. TEs are described by G. Ising and K. Block, *Mol. Gen. Genet.* **196**, 6 (1984).
53. A. H. Sturtevant, *A History of Genetics* (Harper & Row, New York, 1965).
54. T. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
55. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, New York, 1974); D. L. Hartl and A. G. Clark, *Principles of Population Genetics* (Sinauer Associates, Sunderland, MA, 1989).
56. J. D. Hoheisel, G. C. Lennon, G. Zehetner, H. Lehrach, *J. Mol. Biol.* **220**, 903 (1991).

57. Major financial support provided for the YAC project by the NIH; for the cosmid project by the Stimulation Action Programme of the European Community, with initial support from the Fondation Schlumberger, France; for the clone walk database by the National Library of Medicine. J.M. thanks G. Lee, E. Castro, M. Valdivia, and S.-Y. Huh for preparing the figures and tables. M.A. and F.C.K. wish to thank the other participants in the cosmid mapping project, I. Sidén-Kiamos, R. Saunders, B. Savakis, K. Louis, D. Glover, and J. Modolell; as well as B. Jasny and reviewers for helpful comments.