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.

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: "a" or heterochromatin forming the larger block near the centromere, and "b" 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 a 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 renaturing 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).

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 polytene, 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 k 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).

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

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. Germine transposition 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 cosmids 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 cosmids 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 polyteneized genome has been cloned, counting YACs, cosmids, and walks (49).

To maximize the physical map of the Drosophila
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 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: Hohcsell 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 in formation.

REFERENCES AND NOTES

2. The term "conjug" is defined by A. Coulson et al. [in (1)], is widely used in genome projects. Cysts are clusters of linking clones; the first goal of genome projects is to assemble a conjug from end to end for each chromosome.


22. T. A. Painter, Science 78, 885 (1933); Genetics 19, 179 (1934); J. Hered. 25, 465 (1934).
