J. Mol. Biol. (1991) **220**, 903–914

Use of High Coverage Reference Libraries of *Drosophila melanogaster* for Relational Data Analysis

A Step Towards Mapping and Sequencing of the Genome

Jörg D. Hoheisel, Gregory G. Lennon, Günther Zehetner and Hans Lehrach

Imperial Cancer Research Fund
P.O. Box 123, Lincoln's Inn Fields
London WC2A 3PX, U.K.

(Received 27 March 1991; accepted 22 April 1991)

Three differently made, primary *Drosophila* cosmid libraries of 16-fold genome coverage have been generated. Also, a jumping library has been created by a new method that takes advantage of methylation differences between genomic DNA and vector. Thirdly, two cDNA libraries have been picked. All these libraries have been arrayed on high-density *in situ* filters, each containing 9216 clones. As a reference system, such filters are distributed and identified clones are provided. Single-copy probes have identified on average 1·4 cosmids per genome equivalent. Together with cytogenetically mapped yeast artificial chromosomes, the libraries are also being used for physically mapping the genome, mainly by oligonucleotide fingerprinting and pool hybridizations. cDNA clones are further examined by a partial sequencing analysis by oligomer hybridization.

Keywords: *Drosophila melanogaster*; genome; reference libraries; hybridization; mapping

1. Introduction

Since the first experimental studies early this century, the fruit fly *Drosophila melanogaster* has become one of the best-analysed organisms. Starting from classical genetic experiments (Morgan, 1910, 1926), many aspects of its biology have been examined, and more than 12,000 manuscripts dealing with *Drosophila* have been published since 1970 alone (MEDLINE (R), National Library of Medicine, Bethesda, U.S.A.). Owing to this very large amount of genetic and molecular information that has already accumulated and to the existence of the polytene chromosomes as an extremely helpful tool, the *Drosophila* genome seems to be an ideal candidate for generating a physical map. Although quite sophisticated methods are available for isolating genes, particularly for *Drosophila* (Pirrotta, 1986), the analysis of large areas of genomic DNA still requires tedious and labour-intensive procedures. Physical maps are extremely helpful in this kind of analysis, and even more so if the different types of mapping data, genetic markers, cytogenetic data, transcriptional and physical maps, can be related to each other. Also, an ordered library would be of immense value for sequencing. D. Hartl and co-workers (Garza *et al.*, 1989) are cytogenetically characterizing yeast artificial chromosomes with the aim of establishing their physical order. In a different approach, Sidén-Kiamos *et al.* (1990) are using restriction fingerprinting of a cosmid library according to the protocol of Coulson *et al.* (1986). All these procedures, however, are limited in their ability to combine the results with other data sets, although Sidén-Kiamos *et al.* take advantage of microamplified divisional material (Saunders *et al.*, 1989) for preselecting cosmid clones. Since all work is carried out with DNA, the obvious way to relate different data is by the intrinsic function of nucleic acids to form a duplex structure. On the basis of hybridization, we use an integrated approach to the analysis of the *Drosophila* genome. High coverage genomic cosmid and jumping libraries, cDNA libraries, and cytogenetically mapped yeast artificial chromosomes have all been used to produce filters containing large ordered arrays of the clones. As a reference system, filters have been distributed and have been used, for example, for chromosome walking within the cosmid library and for gene isolation. They allow other laboratories to take advantage of the existence of the libraries even before the physical mapping of the genome is completed. Data relation can easily be done by all possible means of hybridization. The order of clones is being established by highly efficient hybridization

0022-2836/91/160903–12 $03.00/0

© 1991 Academic Press Limited

fingerprinting protocols, aiming at combining the data to obtain a complete (and not only physical) map of the genome. Hybridization of very short oligomers to the cDNAs, which are assigned to genomic regions by other means of hybridization, produces partial sequence information about gene structure as the next step towards the ultimate mapping analysis, the complete determination of the genome's sequence.

2. Materials and Methods

Unless stated otherwise, all molecular procedures were performed as described by Sambrook *et al.* (1989).

(a) *Cosmid libraries*

DNA of the *Drosophila melanogaster* wild-type strain Canton-S was prepared from adult male and female flies following the protocol of Herrmann & Frischauf (1987). DNA concentrations were determined in a pH 12 buffer by measuring the increased fluorescence of DNA-intercalated ethidium (Morgan *et al.*, 1979; Pohl *et al.*, 1982). Insert DNA for cosmid cloning was generated following three different procedures.

For a partial *Mbo*I digest, 0·2 μg DNA/μl was cut with 0·01 unit *Mbo*I/μl (New England Biolabs (NEB)) at 37 °C for 1 min. The reaction was stopped by the addition of EDTA and heat-inactivation at 68 °C. Subsequent to extractions with phenol and chloroform and a precipitation with ethanol, the DNA was dephosphorylated using calf intestinal phosphatase.

Partial digestion controlled by a mixture of Dam methylase and *Mbo*I was carried out at a DNA concentration of 40 ng/μl with 7·5 units Dam methylase/μl (NEB) and 0·025 unit *Mbo*I/μl (enzyme ratio 300: 1) at 37 °C for 1 h as described in detail (Hoheisel *et al.*, 1989). The enzymes were heat-inactivated at 68 °C for 20 min and the DNA was instantly treated with phosphatase for 30 min (Hoheisel, 1989*b*). Vector arms of the cosmid Lawrist4 were prepared according to a protocol of P. J. de Jong (personal communication). The vector was linearized with *Sca*I and dephosphorylated, followed by cleavage of the *Bam*HI cloning site. A 2 μg sample of partially cut *Drosophila* DNA was ligated to 2 μg vector arms in 25 μl of 25 mM-Tris·HCl (pH 7·5), 10 mM-dithiothreitol, 100 mM-NaCl, 7 mM-$MgCl_2$, 1 mM-spermidine, 1 mM-ATP and 200 units ligase (NEB) at 15 °C overnight. The ligation products were packaged *in vitro* using Gigapack extracts (Stratagene), transfected into the *Escherichia coli* strains 1046 (Murray *et al.*, 1977) or DH5α (Bethesda Research Laboratories (BRL)) and grown on 30 μg kanamycin/ml agar plates. Individual clones were picked in 2 × YT medium supplemented with freezing medium (36 mM-K_2HPO_4, 13·2 mM-KH_2PO_4, 1·7 mM-sodium citrate, 0·4 mM-$MgSO_4$, 6·8 mM-$(NH_4)_2SO_4$, 4·4% (v/v) glycerol) and kanamycin, grown overnight at 37 °C and are stored frozen at −70 °C.

Sheared Canton-S DNA was blunt-ended by a subsequent reaction of S_1 nuclease and Klenow enzyme (Hoheisel & Pohl, 1986). Though in the actual experiment a DNA preparation of 20 to 80 kb† fragments was used, DNA of an appropriate size range can easily be generated from high molecular weight DNA using a vortex stirrer. A 45 μg sample of DNA was incubated in 150 μl 10 mM-sodium acetate (pH 4·6), 250 mM-NaCl, 1 mM-$ZnSO_4$, 5% glycerol and 100 units S_1 nuclease (Pharmacia) at room temperature for 10 min. It was stopped by a pH-shift upon adding 50 μl of 480 mM-Tris·HCl (pH 8·3), 12 mM-EDTA, 48 mM-$MgCl_2$. After incubating the DNA with 3 units Klenow enzyme at 37 °C for 2 min, 1·5 μl of a mixture of all 4 nucleotides (5 mM each) was added and it was incubated for 15 min at 37 °C. The DNA was extracted with phenol and chloroform, precipitated with ethanol and carefully taken up in water. A 1 μg sample of the *Sma*I oligomer d(pCCCGGG) was annealed to a 5-fold molar excess of 8·25 μg d(GATCCCCGGG) by heating it to 75 °C in 100 μl 100 mM-NaCl and slowly cooling it to 4 °C. Since the DNA duplex has a calculated melting temperature of about 16 °C (Rychlik & Rhoads, 1989), it was kept below 5 °C at all times. A 850 ng sample of the linker mixture (46 pmol of d(pCCCGGG)) was ligated to 15·5 μg blunt-end Canton-S DNA (about 0·3 pmol) in 40 μl of 25 mM-Tris·HCl (pH 7·5), 10 mM-dithiothreitol, 100 mM-NaCl, 7 mM-$MgCl_2$, 1 mM-spermidine, 0·5 mM-ATP with 4000 units bacteriophage T4 DNA ligase (NEB) at 4 °C for 8 days. Subsequently, the DNA was size-selected in a 0·5% low-melting-point agarose gel. Fragments larger than about 30 kb were cut out and isolated by agarase digestion (Burmeister & Lehrach, 1989). Ligation into Lawrist4, packaging and transfection were carried out as described above.

(b) *Plasmid jumping library*

Plasmid pHDJ18R was isolated from the *dam*$^-$ strain GM119 and linearized with *Bgl*II. In a volume of 30 μl, 1·7 μg was ligated to 0·7 μg of a Dam/*Mbo*I partial digest of *Drosophila* DNA with 40 units ligase/μl at 14 °C for 24 h under the conditions described above. After extractions with phenol and chloroform and precipitation with ethanol, the DNA was completely cut with *Not*I. Re-ligation was carried out in 870 μl with 10 units ligase/μl. Run in a 0·6% gel of low-melting-point agarose, fragments larger than about 30 kb were isolated by agarase treatment. Subsequent to a complete digest with *Dpn*I, the plasmids were again circularized at a concentration equivalent to 2 ng pBR322/μl. Colonies of transformed DH5α cells were individually picked into microtiter dishes, grown overnight and stored frozen at −70 °C.

Following the identical protocol, but using methylated plasmid isolated from DH5α and *Mbo*I rather than *Dpn*I, a jumping library can be generated from an *Mbo*I partial digest.

(c) *PCR amplification of the cDNA inserts*

Individual clones of the cDNA plasmid library made from 0–4 h embryos (Brown & Kafatos, 1988) were grown to saturation in LB medium containing 100 μg ampicillin/ml. A 1 μl sample of the bacterial suspension was added to 30 μl of 50 mM-KCl, 10 mM-Tris·HCl (pH 8·3), 1·5 mM-$MgCl_2$, 0·01% (w/v) gelatin, 200 μM of each of the 4 nucleotides, 1 μg of the 24mer d(GAATAAACGCTCAACTTTGGGACC), 0·9 μg of the 21mer d(AGACCGGAATTCGGCGGCCGC) and 15 units AmpliTaq DNA polymerase (Perkin-Elmer Cetus). After an initial denaturation at 94 °C for 4 min, 25 cycles with PCR were carried out, each cycle being 1 min at 94 °C, 1 min at 60 °C and 2 min at 73 °C.

† Abbreviations used: kb, 10^3 bases or base-pairs; Mb, 10^6 bases or base-pairs; bp, base-pairs; PCR, polymerase chain reaction.

(d) In situ *filters*

The clones of 96 microtitre dishes were spotted by a robotic device onto Hybond N plus membranes (Amersham) of 22 cm × 22 cm. The filters were transferred onto 2 × YT agar plates and incubated 16 to 24 h for cell growth. Filter processing was carried out as described in detail by Nizetic *et al.* (1991*a*). The protocol consists of 2 steps of denaturation, 4 min at room temperature and 4 min above a steaming waterbath, followed by neutralization. Each filter was digested for 30 min in 600 ml of 50 mM-Tris · HCl (pH 8·5), 50 mM-EDTA, 100 mM-NaCl, 0·9% (w/v) Na-Sarkosyl (BDH/Merck), and 0·28 mg proteinase K/ml (BDH/Merck) and dried overnight blotted on 3MM Whatman paper. The buffer was re-used 5 times. As determined by measuring its absorption at 260 nm and 280 nm, per filter about 270 mg protein was released into the buffer.

YAC filters were grown on plates of 1% yeast extract, 2% Bacto Peptone (Difco) 2% glucose and 1% agar. Prior to denaturation, they were placed on 3MM paper soaked in 1 mM-sodium citrate (pH 5·8), 10 mM-EDTA, 10 mM-dithiothreitol and 4 mg Novozym/ml (Novo Biolabs) and incubated at 37°C overnight. The first denaturation was extended from 4 to 15 min and the proteinase K treatment was carried out in 0·1 M-Tris · HCl (pH 7·6), 0·15 M-NaCl.

(e) *Generation of end-specific insert-probe by primer extension*

For each sample, 50 ng of the 21mer d(TAGGGAGACCGGAAGCTTAGG) or 70 ng of the 30mer d(CATACACATACGATTTAGGTGACACTATAG) was used in separate reactions to prime in the Lawrist4 vector directly adjacent to either end of the insert. The oligomer was 5′ labeled with 25 μCi of [^{32}P]γATP (5000 Ci/mmol) and 5 units of T4 polynucleotide kinase (NEB) for 1 to 2 h at room temperature. After extractions with phenol and chloroform/isoamylalcohol (24 : 1, v/v), 0·5 μl 5 M-NaCl was added and the DNA oligomer was precipitated with 3·5 vol. ethanol at −70°C for 1 h. After centrifugation for 40 min at 14,000 revs/min, the DNA was taken up in 13 μl water; 0·5 μl was used as a control, 12·5 μl was mixed with 10 μl 5 × amplification buffer (250 mM-Tris · HCl (pH 9·0), 75 mM-ammonium sulfate, 35 mM-$MgCl_2$, 250 mM-KCl, 0·85 mg bovine serum albumin/ml; Anglian Biotec Ltd.), 3·3 μl of all 4 nucleotides (5 mM each), 1·5 μl of AmpliTaq DNA polymerase (8 units; Perkin-Elmer Cetus) and 23 μl of cosmid DNA. For each cosmid 100 to 300 ng were used, and up to 48 cosmids per sample. Linear PCR (Saluz & Jost, 1989) was carried out in a BioOven (BioTherm Corp.) with 60 cycles of 2 min 92°C, 2 min 42°C and 2 min 73°C. The efficiency of the primer extension was checked on a 20% acrylamide gel. Prior to hybridization, the probe was competed with 0·5 μg Lawrist4 DNA/μl (sonicated at 18 μm for 10 times, 10 s), 2·5 μg yeast tRNA/μl (Sigma) and 0·2 μg sonicated *E. coli* DNA/μl in 175 μl water, by incubating at 95°C for 5 min, followed by the addition of 25 μl 1 M-sodium phosphate (pH 7·2), and incubation at 65°C for 2 h.

(f) *Hybridizations*

Hybridizations of single-copy or repeated probes as well as the end-probes generated from cosmid pools were hybridized in 0·5 M-sodium phosphate (pH 7·2), 7% SDS, 1 mM-EDTA, 0·1 mg yeast tRNA/ml at 65°C for at least 10 h, usually overnight, the probe concentration being 0·5 to 5 Mcts/min per ml. The filters were briefly rinsed twice at room temperature in 40 mM-sodium phosphate (pH 7·2), 0·1% SDS. A 1·5 l sample of the same buffer of room temperature was then added to up to 4 filters (22 cm × 22 cm) and slowly rocked in a waterbath of 65°C for 45 min. Only filters hybridized with identical probe were washed together. Subsequently, the filters were briefly blotted dry, covered with Saran Wrap (Genetic Research Instrumentation) and the film was exposed overnight at −70°C using intensifying screen. Quantitative analysis of the signal intensities was carried out on a PhosphorImager (Molecular Dynamics). To strip off the probe, 1·5 l of 5 mM-sodium phosphate (pH 7·2), 0·1% SDS, at room temperature, was poured into a box containing up to 10 filters and kept in a waterbath of about 90°C for 30 min, after which the procedure was repeated. The filters were then immediately transferred to prehybridization conditions (0·5 M-sodium phosphate (pH 7·2), 7% SDS, 1 mM-EDTA, 0·1 mg yeast tRNA/ml) and stored between the hybridizations sealed into a plastic bag at 65°C for up to 8 weeks before being used again.

DNA oligomers were labeled at the 5′ end with [^{32}P]γATP and T4 polynucleotide kinase. Hybridizations were carried out in 600 mM-NaCl, 60 mM-sodium citrate, 7·2% (v/v) Na-Sarkosyl, and 20 pM-probe at 8°C overnight (Drmanac *et al.*, 1990*b*; Nizetic & Drmanac, personal communication). For octamers, the probe concentration was 4 nM. The filters were washed in the same buffer at temperatures between 8°C and 28°C over different periods of time. Exposure was done as described above.

3. Results

(a) *Generation of high coverage cosmid libraries*

Insert DNA was prepared in three different ways in an attempt to avoid uneven fragment representation, e.g. that due to preferential cleavage of certain restriction sites. High molecular weight DNA of the wild-type strain Canton-S was either partially cut with *Mbo*I alone, or the digestion was carried out by a combined reaction of Dam methylase and *Mbo*I (Hoheisel *et al.*, 1989). The latter method allows for easy control of the frequency of cleavage by the ratio of Dam/*Mbo*I. Under the conditions used, fragments of an average size of 70 kb were produced. Thirdly, genomic DNA was mechanically broken and enzymically blunt-ended. *Bam*HI adapters were ligated to the ends, and the fragments were subsequently size-selected on a gel. All insert DNA was dephosphorylated before being cloned in cosmid Lawrist4. The clones were transfected into the *E. coli* strains 1046 (Murray *et al.*, 1977) or DH5α (Woodcock *et al.* 1989). Primary clones were manually picked into 96-well microtitre dishes for growth and subsequent long-term storage at −70°C. Altogether, the library consists of 768 microtitre dishes.

A total of 41 cosmids were randomly picked and digested with the enzymes *Hin*dIII and *Pst*I, which in their combined sequence have a G + C content of 50%. The sizes of all 484 fragments larger than 1 kb fit closely to a Poisson distribution based on a G + C content of about 40% for *Drosophila* DNA (Fig. 1). By this means, the total average size of the cosmids

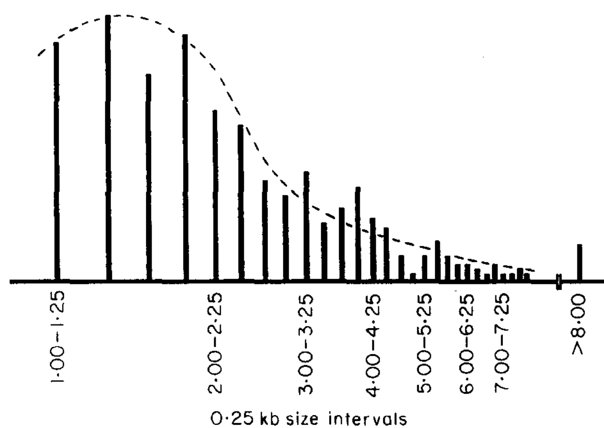

Figure 1. Size distribution of the cosmid libraries. Double digestion with *Hin*dIII and *Pst*I of 41 random cosmids produced 484 fragments larger than 1 kb. Their size was determined by electrophoresis with standards of known length. Ignoring fragments smaller than 1 kb, the apparent average size was 34·6 kb. Using a fitted Poisson distribution, the actual average size of the clones was determined to be 42·5 kb.

was calculated to be 42·5 kb, hence the average insert length is about 37 kb. Additionally, we have characterized by restriction analysis 48 other cosmids found by single-copy probes or chromosome walking. In contrast to human libraries in the same vector (Nizetic *et al.*, 1991*b*), no obviously deleted insert has been detected, irrespective of which *E. coli* host was used.

As determined from control packaging experiments using vector alone as well as from hybridizations with a vector-specific probe to the cosmids of 192 microtitre dishes, 2·7% of the clones contain concatemers of the vector only. Also, material from 4·3% of the wells produced no colony growth after being spotted onto agar plates. Since the *Drosophila* genome is about 160 Mb in size (Rasch *et al.*, 1971), the recombinants represent a 16-fold genome coverage.

In order to present the large clone numbers in a reasonable format that makes them easily accessible for analysis, *in situ* filters of high clone density have been generated in multiple copies. A robotic device permits the highly accurate transfer of bacteria (or phage lysates) from the microtitre dishes to filter membranes (Lehrach *et al.*, 1990). By interleaving patterns, 96 microtitre dishes (9216 clones) are arrayed on a single filter of 22 cm × 22 cm (Fig. 2(a)). Higher densities are anticipated and have been tested successfully (G. Zehetner, unpublished results). Each filter of the genomic cosmid library represents two genome equivalents of *Drosophila*. The whole library fits on eight such filters.

(b) *Complementary jumping library*

In addition to the genomic *Drosophila* library, a jumping library has been generated from the same partial digests used for the cosmid cloning. For the physical mapping of the genomic libraries, it allows one to link areas separated by sequences that are lethal for *E. coli* or unclonable for other reasons. Also, hybridization of potentially overlapping cosmids to this library can confirm (or reject) overlaps suggested by fingerprint data.

As vector, the plasmid pHDJ18R was used which contains eight octanucleotide recognition sites and eight unique hexanucleotide sequences in addition to the 10 unique hexamer sites of the pUC18 polylinker (Fig. 3). The extended multiple cloning site still permits white/blue selection for recombinants due to α-complementation of the β-galactosidase. The plasmid, a derivative of pTZ18R (Mead *et al.*, 1986), was generated as described earlier for other constructs (Hoheisel, 1989*a*; Hoheisel, 1990). The presence of the sites was confirmed by restriction analyses and sequencing.

The method by which the jumping library was made is based on a different methylation status for insert and vector DNA (Fig. 4). With fragments from a partial digestion with Dam methylase and *Mbo*I, the methylated *Drosophila* DNA (b) was ligated to a plasmid vector (a) isolated from an *E. coli dam*$^-$ strain. Concatemers of insert and vector (c) were digested with a rare-cutter enzyme, cutting the plasmid but rarely the genomic DNA (d). The molecules were re-circularized and ligated (e). Subsequent to a size selection by gel electrophoresis, which also removed plasmid DNA not ligated to *Drosophila* fragments, the inserts were then completely digested with *Dpn*I, which specifically cleaves methylated GATC sites, while the vector and the likewise unmethylated cloning sites were not cut. Re-ligation produced clones that contain both terminal *Dpn*I segments of the original genomic fragments. Using *Mbo*I and methylated

Figure 2. Hybridization to high-density filters of the cosmid libraries. (a) Vector-specific probe was hybridized to quantify the amounts of cosmid DNA present at each position. (b) Hybridization with a single-copy probe of the locus *suppressor of forked*, which is located on the X chromosome (division 20E; a gift from K. O'Hare, Imperial College of Science, London). (c) Signal produced by a probe of the transposable element Doc. (d) Hybridization pattern of the oligomer AAAAAAAATAATTT. Washing was at 28°C for 3 h. (e) End-specific probe of the insert-DNA of a pool of 20 randomly picked cosmids was used for hybridization. 69 overlapping clones and a larger number of clones containing repetitive sequences were found. (f) Probe from random hexamer priming of an isolated YAC clone of 275 kb was hybridized, which has been located cytogenetically on chromosome 4 (division 102; D. Hartl, Washington University, St. Louis, personal communication).

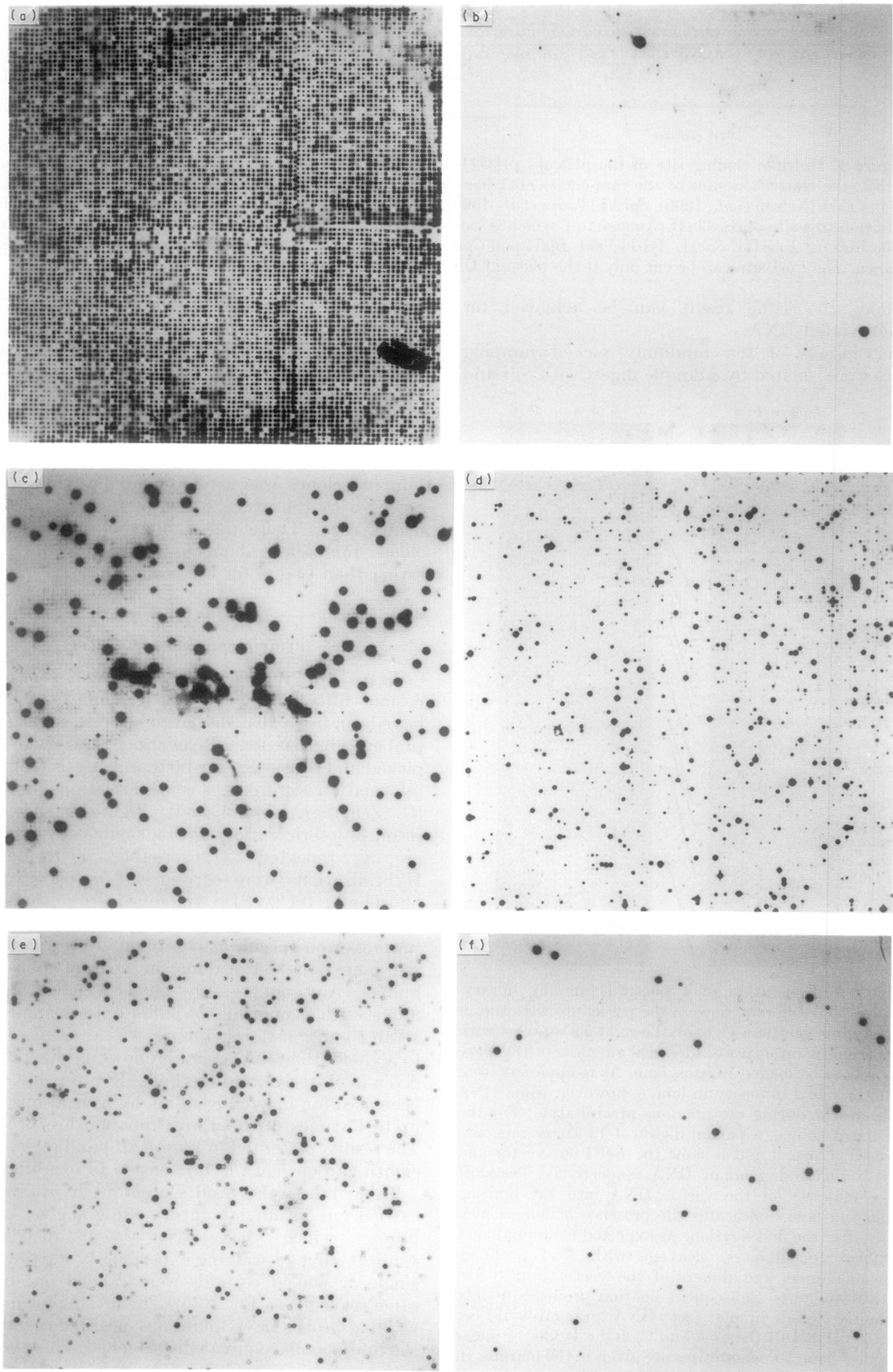

Fig. 2.

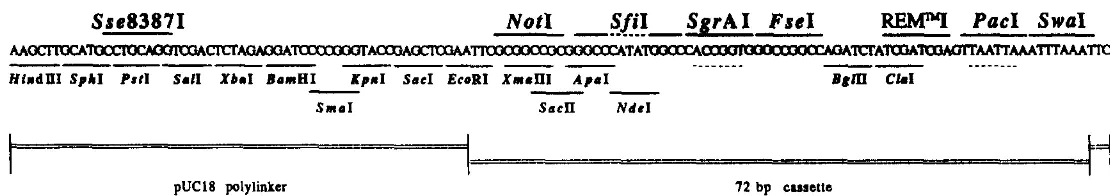

Figure 3. Multiple cloning site of the plasmid pHDJ18R. By insertion of a 72 bp cassette into the *Eco*RI site of pTZ18R, the restriction sites of the rare-cutter enzymes *Swa*I (Boehringer-Mannheim), *Pac*I, REM™I (New England Biolabs) *Fse*I (Nelson *et al.*, 1990), *Sgr*AI (Tautz *et al.*, 1990), *Sfi*I and *Not*I (Qiang & Schildkraut, 1987) were introduced, in addition to a site of *Sse*8387I (Amersham), which is located in the pUC18 polylinker itself. The cassette also contains unique sites for *Xma*III, *Sac*II, *Apa*I, *Nde*I, *Bgl*II and *Cla*I, plus 2 unique hexamers (broken lines), for which no enzyme is known. The *Cla*I site can be cut only if the plasmid DNA is isolated from a dam^- host.

plasmid, the same result can be achieved for unmethylated DNA.

The inserts of ten randomly picked jumping clones were isolated by a double-digest with *Sfi*I and *Pac*I, their average size being 830 bp. The two segments were separated by a *Sau*3A digest, and probe was produced by random hexamer priming (Feinberg & Vogelstein, 1983). One of the cosmids found by each probe was used as starting clone for chromosome walking as described in a later section. From the number of walking steps necessary to link all the cosmids, the average distance bridged by the jumping clones was estimated to be 70 kb, which agrees with the average size of the genomic fragments used. Therefore, a filter of 9216 jumping clones represents about four genome equivalents rather than two as for the cosmid filters.

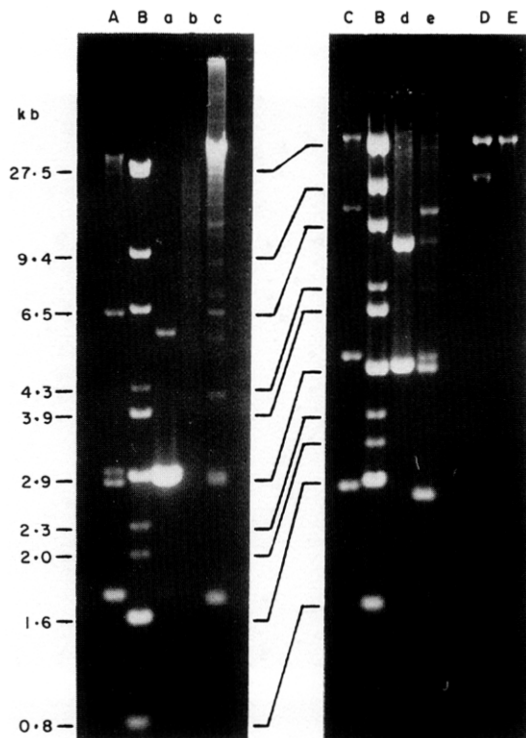

Figure 4. Generation of a plasmid jumping library. Samples from different steps of the procedure were run in 0·6% agarose gels (lanes a to e); the right gel was cast with low-melting-point agarose. Identical amounts of DNA are supposed to be loaded in each lane. As is apparent from the lower signal intensity in lane e, however, some DNA has been lost during the previous precipitation. For the experiment shown, a partial digest of 12 kb average size was used. Lanes a and b show the *Bgl*II-cut vector and partially digested, genomic DNA, respectively. The shift of the majority of the vector DNA into the limiting mobility in lane c indicates the presence of long concatemers after the first ligation. As expected from randomly assembled concatemers, cleavage with *Not*I produces linear monomers and dimers of the vector (lane d) in identical amounts. Subsequent ligation produces mainly covalently closed, circular (ccc) DNA. Standards are ccc DNA of pHDJ18R (lanes A and C) and a ladder of linear molecules (lane B), whose sizes are given in the margin. In lane D linear DNA of 10 and 30 kb is shown; in lane E uncut phage λ DNA.

(c) *Reference library system*

As part of a reference library system, a set of two cosmid filters, carrying four genome equivalents, have been freely distributed to interested parties for probe hybridization. Relevant clones were then picked and forwarded for further analysis. All probe information is stored in a central relational database (G. Zehetner, unpublished). Figure 2 shows the result of hybridizations with a single-copy probe (b) or a repeated transposable element (c). Hybridizations were carried out in 0·5 M-sodium phosphate, 0·1% SDS (modified from Church & Gilbert, 1984) containing 0·1 mg yeast tRNA/ml to suppress non-specific background. Filter wash and stripping were also done in sodium phosphate buffers. Any unnecessary physical stress for the filters, such as pouring hot buffer onto them or using alkali conditions for stripping, was avoided in order to extend their life span. Following the protocol given in Materials and Methods, 18 filters have each been used for an average of ten hybridizations (some up to 17 times) without any apparent loss of signal. The results presented in Figure 2(b) and (c) were the eighth and eleventh hybridization to one filter.

Table 1 lists the results of probe hybridizations carried out in our laboratory using the set of two filters, which has also been distributed. Additionally, 18 walking experiments starting from single or pooled cosmids were carried out on the same set (e.g. see Fig. 2(e)). The 68 single-copy and walking probes have found 394 positive cosmids on the four genome equivalents screened. This averages 1·45 cosmids per probe and genome equivalent. All the probes applied in other laboratories hybridized

Table 1
Probe hybridizations of the cosmid library

Locus	Chromosomal location	No. of positives	Reference
Single copy			
*actin*5C	X 5C	8	Fyrberg *et al.* (1983)
*actin*42A	2 42A	7	Fyrberg *et al.* (1983)
E(spl)	3 96F	12	Preiss *et al.* (1988)
engrailed	2 48A	4	Poole *et al.* (1985)
slit	2 52D	3	Rothberg *et al.* (1988)
pS12·20	X 5E	9	O'Hare & Rubin (1983)
ras	3 64B	11	Mozer *et al.* (1984)
singed	X 7D	4	Roiha *et al.* (1988)
su(f) (pCS7·5)	X 20E	3	A. Michelson, M. Tudor & K. O'Hare (personal communication)
su(f) (pB1·4)	X 20E	4	A. Michelson, M. Tudor & K. O'Hare (personal communication)
white	X 3C	3	Levis *et al.* (1982)
Repeated sequences			
Doc	—	356	Driver *et al.* (1989)
gypsy	—	70	Parkhurst & Corces (1985)
pCS4	—	253	A. Michelson, M. Tudor & K. O'Hare (personal communication)
pogo	—	78	M. Tudor & K. O'Hare (personal communication)
Homologous sequences			
mec-4 (*Caenorhabditis elegans*)	—	2	Driscoll & Chalfie (1991)

The numbers given are the result of hybridizations to 2 filters of the cosmid library, representing 4 genome equivalents. *mec-4*, however, was hybridized to 1 filter only.

to at least one cosmid, the average per genome equivalent being 1·4. As estimated from restriction digests of 48 of the cosmids, about 38% of each insert is not present in any other cosmid of the four genome equivalents used. Taking this into account, each probe finds a stretch of about 20 kb unique DNA per genome equivalent.

In addition to the genomic libraries, two cDNA libraries have been included in the reference system. DH5α *E. coli* was transformed with the amplified cDNA libraries made from 0 to 4 hour and 4 to 8-hour *Drosophila* embryos by Brown & Kafatos (1988; the plasmid DNA was a kind gift from C. Girdham and D. M. Glover, Dundee, U.K.). A total of 19,584 individual colonies were picked into microtitre dishes and high-density filters of 9216 clones each were produced. Such filters were also distributed and have been used for hybridizations with a variety of probes, like, for instance, oligomers representing consensus sequences of a zinc finger motif (Fig. 5(a)) and probe from total mRNA of mouse liver (Fig. 5(b)); T. Gress & H. Lehrach, unpublished results).

(d) *Genome mapping*

To generate a physical map of the entire genome, two different strategies are employed for fingerprinting by hybridization (Lehrach *et al.*, 1990). One is based on the hybridization of short and basically random oligonucleotides. In the other, clone pools are being used to produce similar fingerprint data. The physical order of the clones can then be determined by a statistical analysis of the fingerprints generated (Michiels *et al.*, 1987; Cornette & DeLisi, 1988; Lehrach *et al.*, 1990).

Physical mapping by hybridization fingerprinting has been successfully tested on the genome of *Herpes simplex* type I (Craig *et al.*, 1990) and has also been used extensively in mapping the genome of the fission yeast *Schizosaccharomyces pombe*, with about one probe binding site per 3 to 4 kb of its genome (unpublished results). Actually, the oligomers are unspecifically hybridized at a temperature of 4°C to 8°C, while discrimination between full-match and mismatch hybridization is achieved during the washing (e.g. see Fig. 2(d); and Drmanac *et al.*, 1990*b*). Apart from random sequences, di- and trimer repeats have been used as well as some structure-specific oligonucleotides, which not only produce fingerprint data but in parallel reveal information about DNA structure and organization.

Pool hybridizations have mostly been carried out for chromosome walking, starting from cosmid clones. To that end, two oligonucleotides were hybridized to isolated cosmid DNA, annealing to the vector directly adjacent to either end of the insert. End-specific probe was created by a polymerase extension of the primers. Pools of up to 24

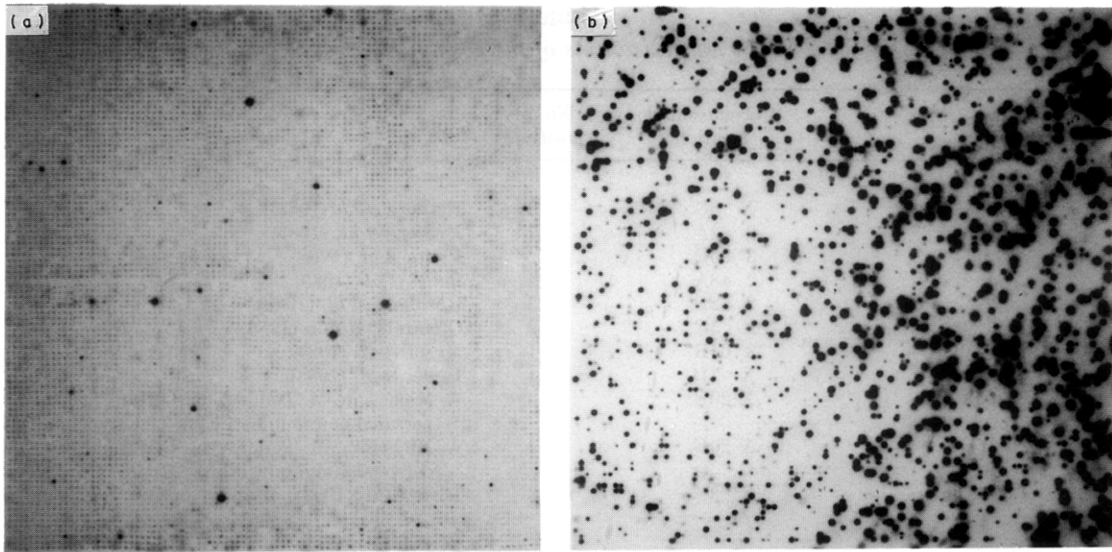

Figure 5. Hybridization to cDNA filters. (a) Probing of a cDNA filter with the zinc finger motif d(GGAGAGAAACC). (b) Probe was generated from total mRNA from mouse liver and hybridized in order to find conserved sequences. One corner of the filter was covered by another filter during the hybridization causing a less intense signal.

clones were used to generate fingerprint data (Fig. 2(e)). Owing to differences in intensity, the weak signal from repetitive sequences could be distinguished from the strong signal from unique overlaps by a quantitative analysis of the hybridization patterns, once a vector probe was hybridized to the filters for normalizing the amounts of cosmid DNA on each position. Pooling the clones in predefined schemes additionally allowed the assignment of positive clones to specific cosmids of the pools by matrix calculations (multiplex walking; Evans & Lewis, 1989). Probe from cosmids was also hybridized to cytogenetically mapped YAC clones (Fig. 6). Compared to the cosmid libraries, far fewer hybridizations will be necessary to produce significant fingerprint data for a physical linkage, due to their larger average insert size of about 220 kb (A. Vellek, J. Ajioka, D. Smoller & D. Hartl, personal communication). However, the yeast DNA in the colonies represents a bigger background problem for oligomer hybridization than the *E. coli* genome in case of the cosmid clones.

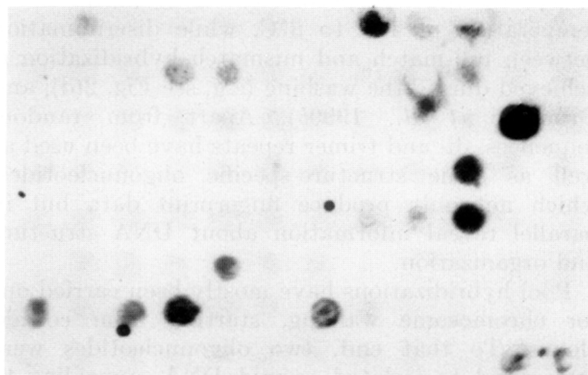

Figure 6. Pool hybridization to YAC clones. Hybridization was carried out with end-specific probe produced from a pool of 12 randomly picked cosmids.

Probe from plasmid pools of the cDNA or jumping libraries was generated by polymerase chain reaction (PCR) amplification of the inserts followed by random priming. In addition to the fingerprint information, the former clones provide information about the presence of coding sequences and assign cDNAs to certain genomic areas, while the latter immediately produce a physical link.

(e) *Gene classification by oligomer hybridization*

Further examination of the cDNA library by the hybridization of very short oligonucleotides (octamers and smaller) can be a very useful method of acquiring at a high rate more detailed information about gene structure and function (Drmanac *et al.*, 1990*a*; Lennon *et al.*, unpublished results). The inserts of cDNA clones, some of which had been hybridized to cosmid filters, were amplified by PCR and run on agarose gels (Fig. 7(a)). The average insert length was 1530 bp. After being transferred to filters (Southern, 1975), the specificity of the bands was checked by hybridization with the two primer molecules. Their signal intensities were in agreement with the amounts of DNA as determined from ethidium staining. A set of octanucleotides was then used for further characterization (e.g. see Fig. 7(b) to (d)). Their sequences were either randomly chosen (b) or based on certain DNA features, like, for example, the zinc finger motifs of the Krüppel type d(GAGAAACC) ("8mf1"; (c)) and d(GGSAARGC) ("8mf2"; (d)) (Bellefroid *et al.*, 1989). The DNA standards of known sequence, pUC18 cut with *Taq*I, pJDH218 linearized with *Bam*HI, and linear DNA of a pUC18 derivative containing an *Mbo*I fragment of λ, provide good

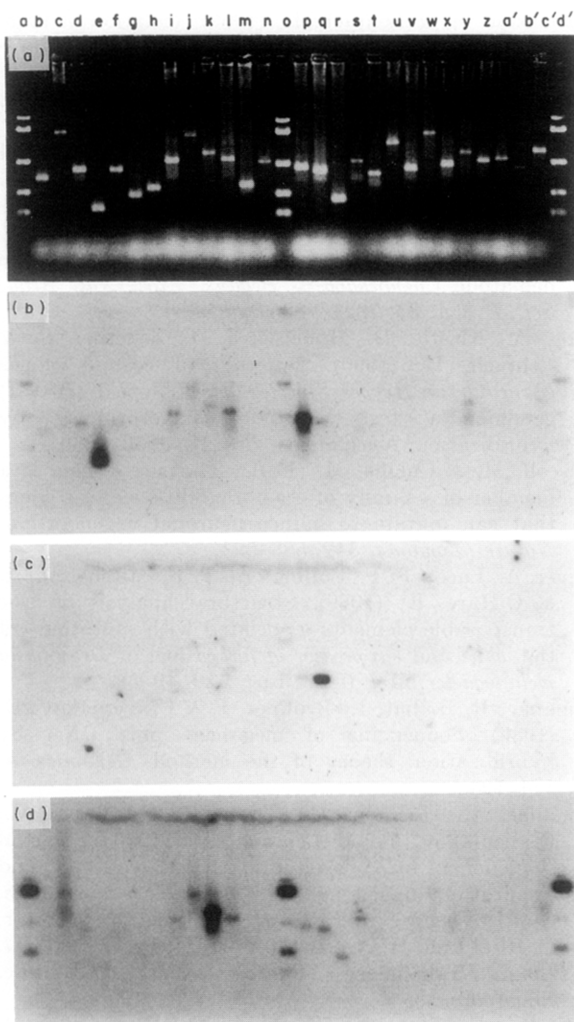

Figure 7. cDNA characterization by oligomer hybridization. (a) PCR products (lanes b to n, p to c′) from cDNA inserts were run in a 1% agarose gel in the presence of ethidium bromide. Standards (lanes a, o, d′) are linear molecules of 3912 bp, 2898 bp, 1444 bp, 736 bp and 476 bp. (b) to (d) A Southern blot of the gel was hybridized with the octamers d(GCTTGGCG) (b), d(GAGAAACC) (c), and d(GGSAARGC) (d).

internal control A complement to the degenerate sequence of 8mf2, for instance, is also present once in the plasmid pJDH218. Also, the 476 bp and 736 bp fragments of pUC18 contain one and three copies, respectively, different in one terminal base, while the 1444 bp band contains no match or a terminal single base mismatch. As calculated from the signal intensities, there is either one full-match (± 3 mismatches) in the PCR product of lane k or 43 (± 3) mismatches of an average stability like the three in the 736 bp pUC fragment. Taking into account the fragment's size of about 1840 bp, however, the latter case is extremely unlikely (about P 0·005). With the signal intensities, the amounts of DNA, and the number of mismatches in the standard molecules known, similar likelihood ratios have been calculated for the other oligonucleotides (Fig. 7(b) and (c)), although no positive control is present. Actually, with the panel of 20 octamers used, no difficulties in discriminate hybridization have been encountered.

4. Discussion

On the basis of hybridization, an integrated approach to the analysis of the genome of the fruit fly *Drosophila melanogaster* is being used with a variety of different libraries. They represent the initial step towards combining genetic, cytogenetic, transcriptional and physical data for a comprehensive molecular understanding of the structure, organization and function (as well as malfunction) of the genome, and by this a great many more aspects of the biology of *Drosophila*.

The high-coverage cosmid library is used as the basic unit of the analysis. The three different protocols by which the inserts were generated ensure an even representation as far as possible, a fact proven by the single-copy hybridizations carried out by us and others. In particular, the sheared library helps in this regard, because regions without appropriate restriction sites (e.g. areas of a high repeat content) are made clonable by this procedure. The supplementary jumping library allows us to span gaps in the genomic libraries due to unclonable sequences. Additionally, contiguous arrays of clones can easily identify related jumping clones to serve as a suitable tool for the generation of sequence tagged sites (Olson *et al.*, 1989) within their two insert segments.

The reference system makes the high-coverage libraries of different types accessible and useful to other laboratories, even before any mapping or other data of their region of interest have been produced. The linkage between the libraries permits an immediate correlation between all levels of DNA analysis. Hybridization with cDNAs or consensus sequences of known gene families, for instance, indicate on the cDNA filters all coding sequences containing conserved and homologous regions, while the genomic DNA can be found in parallel in the cosmid and YAC libraries. If the former are already linked to known genetic markers, or if the latter are cytogenetically mapped, no further determination of the chromosomal location is necessary. In the opposite experiment, cosmids, YACs, large DNA fragments isolated from pulsed-field gels, and microamplified sections of the chromosomes (Saunders *et al.*, 1989) could be used to probe the filters, as to identify cDNAs present in the particular segment, for instance. We wish to emphasize that filters from all the libraries mentioned are available for scientific purposes. Also, we anticipate releasing updates of the database.

The physical mapping of the genomic cosmid library is currently being carried out with the entire library at a rate of about ten oligonucleotide probings per week. Additionally, end-specific probe from cosmid and cDNA pools are being used, after repetitive elements have been excluded from the probe (Hochgeschwender *et al.*, 1989). These hybridizations will be particularly useful in later stages of the

analysis in order to close remaining gaps. The walking procedure of generating end-specific probe by an enzymic extension can be applied to all vector systems, as opposed to the protocol of Evans & Louis (1989), which relies on the presence of promoter sequences for RNA polymerases. We intend to use a similar procedure for a physical linkage of YAC clones.

The examination of cDNAs by hybridizing octanucleotides permits the analysis of coding sequences at a high rate. Obviously, for experiments on a larger scale, high-density filters should be used, like those for the reference libraries. The technique allows us to identify quickly homologies that relate unknown genes to already known gene families (Lennon *et al.*, unpublished results). Also, functions can be predicted from certain features, like, for instance, motifs of DNA binding proteins or homeo box sequences. Potentially, the technique can be extended in order to determine the complete sequence (Drmanac *et al.*, 1989; Khrapko *et al.*, 1989). For this purpose, shorter oligomers would be an advantage, since far fewer hybridizations would have to be carried out. On the basis of a thin-layer chromatographic technique, which mimics the conditions of filter hybridization but permits a much quicker determination of the dissociation temperatures of oligonucleotides (Hoheisel *et al.*, 1990), some hexamer oligonucleotides containing base modification, which improve their duplex stability (Hoheisel *et al.*, 1990; Hoheisel & Lehrach, 1990), have been tested successfully. If all hexamer sequences could be used for discriminate hybridization, 2048 oligonucleotides would need to be hybridized to double-stranded DNA in order to determine the complete sequence. A positive control for such experiments is easily provided by any known DNA, with plasmid pUC19, for instance, containing 2862 different hexamer sequences within its 2686 bp; that is, 70% of the possible 4096.

We thank K. O'Hare for his support throughout the project ranging from the provision of flies and many probes to helpful advice and information. We also thank D. Hartley for probes, I. Goldsmith for the synthesis of oligonucleotides, and D. Hartl and coworkers for providing their libraries and relevant information prior to publication. We are indebted to T. Gress for making his results available to us, and to R. Drmanac and D. Nizetic for helpful discussions. Funding by the Imperial Cancer Research Fund and a long-term EMBO fellowship to J. D. Hoheisel are gratefully acknowledged.

References

Bellefroid, E. J., Lecocq, P. J., Benhida, A., Poncelet, D. A., Belayew, A. & Martial, J. A. (1989). The human genome contains hundreds of genes coding for finger proteins of the Krüppel type. *DNA*, **8**, 377–387.

Brown, N. H. & Kafatos, F. C. (1988). Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol* **203,** 425–437.

Burmeister, M. & Lehrach, H. (1989). Isolation of large DNA fragments from agarose gels using agarase *Trends Genet.* **5**, 41.

Church, G. M. & Gilbert, W. (1984). Genomic sequencing. *Proc. Nat. Acad. Sci., U.S.A.* **81**, 1991–1995.

Cornette, J. L. & DeLisi, C. (1988). Some mathematical aspects of mapping DNA cosmids. *Cell Biophys.* **12**, 271–293.

Coulson, A., Sulston, J., Brenner, S. & Karn, J. (1986). Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 7821–7825.

Craig, A., Nizetic, D., Hoheisel, J. D., Zehetner, G. & Lehrach, H. (1990). Ordering of cosmid clones covering the *Herpes Simplex* Virus Type I (HSV-I) genome: a test case for fingerprinting by hybridization. *Nucl. Acids. Res.* **18**, 2653–2660.

Driscoll, M. & Chalfie, M. (1991). The *mec-4* gene is a member of a family of Caenorhabditis elegans genes that can mutate to induce neuronal degeneration. *Nature (London)*, **349**, 588–593.

Driver, A., Lacey, S. F., Cullingford, T. E., Mitchelson, A. & O'Hare, K. (1989). Structural analysis of Doc transposable elements associated with mutations of the *white* and *suppressor of forked* loci of *Drosophila melanogaster*. *Mol. Gen. Genet.* **220**, 49–52.

Drmanac, R., Labat, I., Brukner, I. & Crkvenjakov, R. (1989). Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics*, **4**, 114–128.

Drmanac, R., Lennon, G., Drmanac, S., Labat, I., Crkvenjakov, R. & Lehrach, H. (1990*a*). Partial sequencing by oligo-hybridization: concept and application in genome analysis. In *Electrophoresis, Supercomputing and the Human Genome* (Cantor, C. R. & Lim, H. A., eds), pp. 60–74, Proc. First *Int. Conf.*, Tallahassee, Florida, World Scientific, Singapore.

Drmanac, R., Strezoska, Z., Labat, I., Drmanac, S. & Crkvenjakov, R. (1990*b*). Reliable hybridization of oligonucleotides as short as six nucleotides. *DNA and Cell Biol.* **9**, 527–534.

Evans, G. A. & Lewis, K. A. (1989). Physical mapping of complex genomes by cosmid multiplex analysis. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 5030–5034.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13.

Fyrberg, E. A., Mahaffey, J. W., Bond, B. J. & Davidson, N. (1983). Transcripts of six *Drosophila* actin genes accumulate in a stage and tissue specific manner. *Cell*, **33**, 115–123.

Garza, D., Ajioka, J. W., Burke, D. T. & Hartl, D. L. (1989). Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science*, **246**, 641–646.

Herrmann, B. G. & Frischauf, A.-M. (1987). Isolation of genomic DNA. *Methods Enzymol.* **152**, 180–183.

Hochgeschwender, U., Sutcliffe, J. G. & Brennan, M. B. (1989). Construction and screening of a genomic library specific for mouse chromosome 16. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 8482–8486.

Hoheisel, J. D. (1989*a*). A cassette with seven unique restriction sites, including octanucleotide sequences: extension of multiple-cloning-site plasmids. *Gene*, **80**, 151–154.

Hoheisel, J. D. (1989*b*). Ultrafast Microprep protocol for large-number preparations of recombinant plasmids and cosmids. *Technique*, **1**, 21–27.

Hoheisel, J. D. (1990). Extension of a pUC18-like polylinker by the octanucleotide recognition sites of *Not*I, *Fse*I, *Sfi*I and *Pac*I. *Trends Genet.* **6**, 346.

Hoheisel, J. D. & Lehrach, H. (1990). Quantitative measurements on the duplex stability of 2,6-diaminopurine and 5-chloro-uracil nucleotides using enzymatically synthesized oligomers. *FEBS Letters*, **274**, 103–106.

Hoheisel, J. D. & Pohl, F. M. (1986). Simplified preparation of unidirectional deletion clones. *Nucl. Acids Res.* **14**, 3605.

Hoheisel, J. D., Nizetic, D. & Lehrach, H. (1989). Control of partial digestion combining the enzymes Dam methylase and *Mbo*I. *Nucl. Acids Res.* **17**, 4571–4582.

Hoheisel, J. D., Craig, A. G. & Lehrach, H. (1990). Effect of 5-bromo and 5-methyl deoxycytosine on duplex stability and discrimination of the *Not*I octadeoxynucleotide; quantitative measurements using thin-layer chromatography. *J. Biol. Chem.* **265**, 16656–16660.

Khrapko, K., Lysov, Y., Khorlyn, A., Shick, V., Florentiev, V. & Mirzabekov, A. (1989). An oligonucleotide hybridization approach to DNA sequencing. *FEBS Letters*, **256**, 118–122.

Lehrach, H., Drmanac, R., Hoheisel, J. D., Larin, Z., Lennon, G., Monaco, A. P., Nizetic, D., Zehetner, G. & Poustka, A. (1990). Hybridization fingerprinting in genome mapping and sequencing. In *Genome Analysis: Genetic and Physical Mapping* (Davies, K. E. & Tilghman, S. eds), pp. 39–81, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Levis, R. L., Bingham, P. M. & Rubin, G. M. (1982). Physical map of the *white* locus of *Drosophila melanogaster*. *Proc. Nat. Acad. Sci., U.S.A.* **79**, 564–568.

Mead, D. A., Szczesna-Skorupa, E. & Kemper, B. (1986). Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Engineering*, **1**, 67–74.

Michiels, F., Craig, A. G., Zehetner, G., Smith, G. P. & Lehrach, H. (1987). Molecular approach to genome analysis: a strategy for the construction of ordered overlapping clone libraries. *Comput. Appl. Biosci.* **3**, 203–210.

Morgan, A. R., Lee, J. S., Pulleyblank, D. E., Murray, N. L. & Evans, D. H. (1979). Review: ethidium fluorescence assay. Part 1. Physicochemical studies. *Nucl. Acids Res.* **7**, 547–569.

Morgan, T. H. (1910). Sex limited inheritance in *Drosophila*. *Science*, **32**, 120–122.

Morgan, T. H. (1926). *The Theory of the Gene*, Yale University Press.

Mozer, B., Marker, R., Parkhurst, S. & Corces, V. (1984). Characterization and developmental expression of a *Drosophila ras* oncogene. *Mol. Cell. Biol.* **5**, 885–889.

Murray, N. E., Brammar, W. J. & Murray, K. (1977). Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Mol. Gen. Genet.* **150**, 53–61.

Nelson, J. M., Miceli, S. M., Lechelavier, M. P. & Roberts, R. J. (1990). *Fse*I, a new type II restriction endonuclease that recognizes the octanucleotide sequence 5′ GGCCGGCC 3′. *Nucl. Acids Res.* **18**, 2061–2064.

Nizetic, D., Drmanac, R. & Lehrach, H. (1991*a*). An improved bacterial colony lysis procedure enables direct DNA hybridization using short (10, 11 bases) oligonucleotides to cosmids. *Nucl. Acids Res.* **19**, 182.

Nizetic, D., Zehetner, G., Monaco, A. P., Gellen, L., Young, B. D. & Lehrach, H. (1991*b*). Construction, arraying, and high-density screening of large insert cosmid libraries of the human chromosomes X and 21: their potential use as reference libraries. *Proc. Nat. Acad. Sci., U.S.A.*, **88**, 3233–3237.

O'Hare, K. & Rubin, G. M. (1983). Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell*, **34**, 25–35.

Olson, M., Hood, L., Cantor, C. & Botstein, D. (1989). A common language for physical mapping of the human genome. *Science*, **245**, 1434–1435.

Parkhurst, S. & Corces, V. (1985). *forked*, gypsys and suppressors in *Drosophila*. *Cell*, **41**, 429–437.

Pirrotta, V. (1986). Cloning *Drosophila* genes. In *Drosophila, A Practical Approach* (D. B. Roberts, ed.), pp. 83–110, IRL Press, Oxford, Washington.

Pohl, F. M., Thomae, R. & Karst, A. (1982). Temperature dependence of the activity of DNA-modifying enzymes: endonucleases and DNA ligase. *Eur. J. Biochem.* **123**, 141–152.

Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. (1985). The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell*, **40**, 37–43.

Preiss, A., Hartley, D. A. & Artavanis-Tsakonas, S. (1988). The molecular genetics of *Enhancer of split*, a gene required for embryonic neural development in *Drosophila*. *EMBO J.* **7**, 3917–3927.

Qiang, B.-Q. & Schildkraut I. (1987). *Not*I and *Sfi*I: restriction endonucleases with octanucleotide recognition sequences. *Methods Enzymol.* **155**, 15–21.

Rasch, E. M., Barr, H. J. & Rasch, R. W. (1971). The DNA content of sperm of *Drosophila melanogaster*. *Chromosoma*, **33**, 1–18.

Roiha, H., Rubin, G. M. & O'Hare, K. (1988). P element insertions and rearrangement at the *singed* locus of *Drosophila melanogaster*. *Genetics*, **119**, 75–83.

Rothberg, M. J., Hartley, D. A., Walther, Z. & Artavanis-Tsakonas, S. (1988). *slit*: an EGF-homologous locus of *D. melanogaster* involved in the development of the embryonic central nervous system. *Cell*, **55**, 1047–1059.

Rychlik, W. & Rhoads, R. E. (1989). A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. *Nucl. Acids Res.* **17**, 8543–8551.

Saluz, H. & Jost, J.-P. (1989). A simple high-resolution procedure to study DNA methylation and *in vivo* DNA–protein interactions on a single-copy gene level in higher eukaryotes. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 2602–2606.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Saunders, R. D. C., Glover, D. M., Ashburner, M., Sidén-Kiamos, I., Louis, C., Monastirioti, M., Savakis, C. & Kafatos, F. (1989). PCR amplification of DNA microdissected from a single polytene chromosome band: a comparison with conventional microcloning. *Nucl. Acids Res.* **17**, 9027–9037.

Sidén-Kiamos, I., Saunders, R. D. C., Spanos, L., Majerus, T., Treanear, J., Savakis, C., Louis, C., Glover, D. M., Ashburner, M. & Kafatos, F. C. (1990). Towards a physical map of the *Drosophila melanogaster* genome: mapping of cosmid clones within defined genomic divisions. *Nucl. Acids Res.* **18**, 6261–6270.

Southern, E. M. (1975). Detection of specific sequences

among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.

Tautz, N., Kaluza, K., Frey, B., Jarsch, M., Schmitz, G. G. & Kessler, C. (1990). *Sgr*AI, a novel class-II restriction endonuclease from Streptomyces griseus recognizing the octanucleotide sequence of 5′-CR/CCGGYG-3′. *Nucl. Acids Res.* **18**, 3087.

Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S. S., Michael, M. Z. & Graham, M. W. (1989). Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucl. Acids Res.* **17**, 3469–3478.

Edited by J.-L. Mandel