ELSEVIER

Journal of Biotechnology 35 (1994) 121–134

journal of biotechnology

Relational genome analysis using reference libraries and hybridisation fingerprinting

Jörg D. Hoheisel [a,*], Mark T. Ross [b], Günther Zehetner [b], Hans Lehrach [b]

[a] *Molecular-Genetic Genome Analysis, German Cancer Research Centre, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany,*
[b] *Genome Analysis Laboratory, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK*

Abstract

The genomes of eukaryotic organisms are studied by an integrated approach based on hybridisation techniques. For this purpose, a reference library system has been set up, with a wide range of clone libraries made accessible to probe hybridisation as high density filter grids. Many different library types made from a variety of organisms can thus be analysed in a highly parallel process; hence, the amount of work per individual clone is minimised. In addition, information produced on one analysis level instantly assists in the characterisation process on another level. Genetic, physical and transcriptional mapping information and partial sequencing data are obtained for the individual library clones and are cross-referenced toward a comprehensive molecular understanding of genome structure and organisation, of encoded functions and their regulation. The order of genomic clones is established by hybridisation fingerprinting procedures. On these physical maps, the location of transcripts is determined. Complementary, partial sequence information is produced from corresponding cDNAs by hybridising short oligonucleotides, which will lead to the identification of regions of sequence conservation and the constitution of a gene inventory. The hybridisation analysis of the cDNA clones, and the genomic clones as well, could potentially be expanded toward a determination of (nearly) the complete sequence. The accumulated data set will provide the means to direct large-scale sequencing of the DNA, or might even make the sequence analysis of large genomic regions a redundant undertaking due to the already collected information.

Key words: Mapping; Clone library; Hybridisation

1. Introduction

Since the publication of early work on physical mapping of representative genomic DNA libraries (Olson et al., 1986; Coulson et al., 1986; Kohara et al., 1987), major advances have taken place in large-scale mapping projects that aim at the generation of ordered clone libraries covering chromosomes or genomes of multicellular organisms in their entirety. This is most clearly demonstrated by the establishment of a physical clone map of most of the human genome in yeast artificial chromosomes (YACs; Cohen et al., 1993). The availability of molecular maps of higher resolution, such as the cosmid and bacte-

* Corresponding author.

SSDI 0168-1656(94)00035-B

riophage P1 maps spanning the genome of the fission yeast *Schizosaccharomyces pombe* (Hoheisel et al., 1993), is a prerequisite for more detailed analyses of genomic areas, for example, the exact determination and characterisation of gene locations, for they permit immediate access to the DNA. Also, the subdivision of a genome in relatively small ordered pieces facilitates large-scale sequencing – the ultimate objective of the Genome Projects – as demonstrated by the sequence analysis of chromosome III of *Saccharomyces cerevisiae* (Oliver et al., 1992).

The more information can be extracted during the initial phase of mapping in addition to the data that are merely required for the ordering process, the simpler subsequent studies will become. We have therefore embarked on an experimental route that aims at an integrated genome analysis, producing and relating information from all levels of DNA manipulation, from cell radiation hybrids and genetic markers down to the level of partial sequence determination (Fig. 1; Lehrach et al., 1990). Since all experiments deal with nucleic acids, a relation can most conveniently be achieved by their intrinsic function of duplex formation. Hybridisation procedures permit a simple and fast way of identifying homologous DNA and RNA stretches, and large numbers of clones can be analysed concurrently. Any piece of nucleic acid, from a YAC clone of 1 Mbp down to an oligonucleotide of 6 nucleotides, can be used both as a probe and as a probant. Due to this very flexible characteristic, hybridisation allows the optimisation of the gain in information

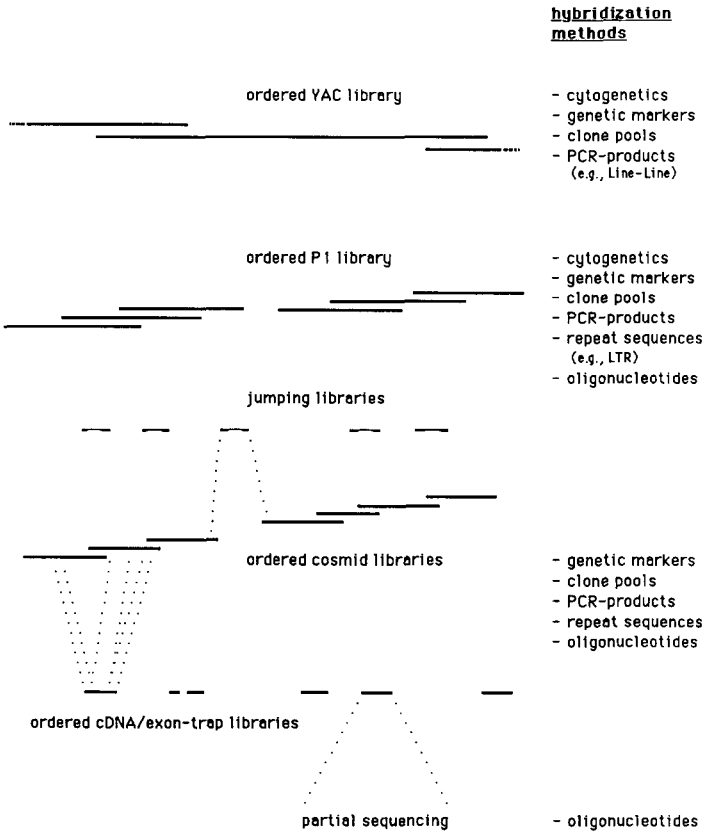

Fig. 1. The different levels of genome analysis on clone libraries by hybridisation fingerprinting. Some exemplary probe types are listed.

from each experiment, an important factor considering the size of the genomes dealt with. Using various hybridisation techniques, we are involved in the process of mapping and further analysing the genomes of several eukaryotic (model) organisms: *S. pombe*, *Drosophila melanogaster*, mouse, pig and human. Additional information is assigned to the clones by distributing clone filters to other laboratories and collection of the locally obtained results in a central reference library system. The successful mapping of the 14 Mbp genome of *S. pombe* (Maier et al., 1992; Hoheisel et al., 1993) demonstrates the efficiency and effectiveness of the existing experimental and analytical procedures and forms the basis from which the technology is advanced further towards the analysis of mammalian genomes.

2. Methodology and results

2.1. Reference libraries

Backbone of our strategy is the existence of clone libraries of high genome coverage. A wide range of cloning vehicles – YAC, bacteriophage P1, cosmid and λ phage as well as plasmid jumping and linking libraries – is being used. Individual and, usually, primary clones are picked and stored in microtiter dishes. The libraries are spotted as ordered arrays of very high density of up to 36 864 clones onto filters of 22 × 22 cm. The filters are incubated for colony growth on agar plates, and the DNA is subsequently attached to the filter in situ. The whole process has been automated to a large extent by the development of robotic devices for picking and spotting which are described in more detail by Maier et al. in this issue.

The filters represent the basic tool for hybridisation analyses. Since many filter replicas can be produced, it is possible to carry out a large number of parallel experiments. Apart from being used in our laboratories, the DNA filter arrays are also distributed to interested parties as part of a reference system. Table 1 shows the reference libraries that are currently available on filters. Usually only a part of a library is distributed; for example, all the cosmid libraries of *S. pombe* represent about 38 genome equivalents, while the filter generally sent contains an 8-fold coverage. Clones identified by the user are provided for further analysis, in exchange for the probe information, which is stored in databases (Lehrach et al., 1990; Hoheisel et al., 1991). The information can be kept confidential for a period of maximal one year or until publication. Apart from linking their probes to the developing physical map, and vice versa, other laboratories can work on a more detailed (e.g., sequence) analysis of the genomic region of their interest, even before a physical map is completed. Their markers will eventually be positioned in the mapped genome without any additional effort on their part. Since one particular genomic library is being used as a reference by a large number of groups, the results generated are easily comparable. Also, the exchange of data is assisted and promoted. If certain clones are found by different probes, for instance, the respective laboratories are informed about this fact.

To date (December 1992), some 200 clones from the libraries of *S. pombe* have been sent to 15 laboratories. Several thousand clones of the *Drosophila* libraries have been identified by probes from 33 laboratories. From these, 1303 clones have been distributed, of which about half (633) are cDNA clones. Of the other libraries, 8842 clones have been provided to 211 laboratories, mainly human and mouse YACs and human cosmid clones. Libraries additional to those listed in Table 1 have been generated or are in the process of being made and will become available, in particular more chromosome specific human cosmids and various cDNA libraries. We are also involved in experiments aiming at the creation of exon-trap libraries (Buckler et al., 1991). Since genomic DNA is the source material, an exon-trap library represents a quasi-normalised library of the transcribed DNA of the relevant genomic region, and intrinsically it is related to its genomic origin.

2.2. Genome analysis

A broad range of hybridisation procedures has been established to allow the analysis of the

reference libraries (e.g., Fig. 2; Craig et al., 1990; Hoheisel et al., 1991, 1993; Monaco et al., 1991; Gress et al., 1992; Ross et al., 1992b). Although the different methods of analysis are listed separately below, they are interconnected and carried out in parallel, so that results from one level improve the analysis of the others. Many probes are actually hybridised in a single experiment to many libraries for instant correlation.

2.2.1. Clone hybridisations

Hybridisation of single clones is technically and analytically the simplest form of ordering genomic libraries by hybridisation fingerprinting.

Table 1
Libraries available on reference filters

Library type	DNA source	Partial digestion	Vector DNA	Host strain	Number of clones	Genome equivalents	Ref.
Schizosaccharomyces pombe (fission yeast); genome size 14 Mbp:							
YACs, *ordered*:							
pomYAC	972h$^-$	*Eco* methyl./ *Eco*RI	pYAC4	AB1380	1248	47	Maier et al., 1992 Hoheisel et al., 1993
P1s, *ordered*:							
pomP1	972h$^-$	*Mbo*I	pAd10sacB11	NS3145	3456	17	Hoheisel et al., 1993
Cosmids, *ordered*:							
pomL4	972h$^-$	*Mbo*I	Lawrist4	ED8767	3072	8	Hoheisel et al., 1993
Drosophila melanogaster (fruit fly); genome size 165 Mbp:							
YACs:							
YAC:DY	Oregon-RC	sheared	pYACP-1	AB1380	768		
YAC:N	Oregon-RC	*Not*I	pYAC5	AB1380	2688	10 [a]	Ajioka et al., 1991
YAC:E	Canton-S	*Eco*RI	pYAC4	AB1380	3744		
P1s:							
DroP1	isogenic strain	*Sau*3A	pNS582 tet14Ad10	NS3145	3840	2 [a]	Smoller et al., 1991
jumping library:							
JHDj70	Canton-S	*Mbo*I	pHDJ18R	JM83	4608	2	Hoheisel et al., 1991
Cosmids:							
JHD1	Canton-S	sheared	Lawrist4	ED8767	13440		
JHD2	Canton-S	Dam/*Mbo*I	Lawrist4	ED8767	17184		
JHD3	Canton-S	*Mbo*I	Lawrist4	ED8767	23136	16	Hoheisel et al., 1991
JHD4	Canton-S	Dam/*Mbo*I	Lawrist4	DH5α	10176		
JHD5	Canton-S	*Mbo*I	Lawrist4	DH5α	9984		
cDNAs:							
cDNA1	0–4 h embryos	–	pNB40	DH5α	9792	1.5 [b]	Hoheisel et al., 1991
cDNA2	0–8 h embryos	–	pNB40	DH5α	10176		Brown and Kafatos, 1988
Sus scrofa (pig); genome size ~ 3000 Mbp:							
YACs:							
pigYAC	female	*Eco* meth./ *Eco*RI	pCGS966	AB1380	31104	1.2	Meier-Ewert, unpubl.
Mus musculus (mouse); genome size ~ 3000 Mbp:							
YACs:							
MYAC	C3H/C57b6	*Eco* methyl./ *Eco*RI	pYAC4	AB1380	19200	4	Larin et al., 1991 Ross et al., 1992a
cDNAs:							
MB1	brain	–	pSPORT	XL1-Blue	20736	0.2 [c]	Meier-Ewert, unpubl.

Table 1 (continued)

Library type	DNA source	Partial digestion	Vector DNA	Host strain	Number of clones	Genome equivalents	Ref.
Homo sapiens (human); genome size ~ 3000 Mbp:							
YACs:							
HYAC	GM1416B OXEN, HD1	*Eco* methyl./ *Eco*RI	pYAC4	AB1380	20544	4	Ross et al., 1992a Larin et al., 1991
P1s:							
HP1	GM1416B	*Mbo*I	pAd10sacB11	NS3145	41472	1.2	Francis, unpubl.
Cosmids, chromosome specific:							
FS6	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	19200	5	
FS11	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	18432	6	
FS13	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	18432	6	
FS17	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	18432	7	Nizetic et al., 1991
FS18	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	18432	7	Lehrach et al., 1990
FS21	GM1416B	Dam/*Mbo*I	Lawrist4	ED8767	9216	3	
FS22	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	9216	4.5	
FSX	GM1416B	Dam/*Mbo*I	Lawrist4	DH5α	18432	3.5	
cDNAs:							
HFB	foetal brain	–	pSPORT	XL1-Blue	31104	0.3 [c]	Meier-Ewert, unpubl.
PATU	pancreas carcinoma	–	pSPORT	XL1-Blue	20736	0.2 [c]	Gress, unpubl.

[a] The libraries were kindly provided by Dr. Daniel Hartl, Washington University, St. Louis, USA
[b] based on the assumption that there are 12000 genes in the genome.
[c] based on the assumption that there are 100000 genes in the genome.

In the mapping of *S. pombe*, a mere 153 hybridisations were sufficient to complete the YAC map of the 14 Mbp genome (Maier et al., 1992), and 764 probes were necessary for spanning it with cosmid and P1 clones (Hoheisel et al., 1993). Following a strategy of 'sampling without replacement', only clones not hit in any of the earlier hybridisations were picked as probes for subsequent experiments until all clones had been hit at least once. To increase the efficiency of probe selection, the unhit clones were sorted on the basis of data from larger fragment hybridisations (YACs and rare cutter DNA fragments) onto cosmid and P1 clones. The clones were not only sub-divided into groups this way but also positioned within the genome and oriented; such information is usually more readily available for large fragment probes. Due to the coarse ordering, evenly spaced clones could be picked as probes for the next round of single clone hybridisations, thus avoiding unnecessary redundancy.

Mapping information was obtained from hybridisation frequencies. The probes were ordered by first calculating a distance between each possible pair of probes, which varied between zero – the probes are identical in terms of hybridisation – and one – the probes are one (or more) average clone length apart. Based on the distances, the probe order was established by a calculation of their shortest possible linear succession using a simulated annealing algorithm (Mott et al., 1993). Thus, an ordered set of probe-tagged sites was created (Hoheisel et al., 1993), named in analogy to the STS concept (Olson et al., 1989), and the clones were fitted to this probe map. The advantage of the approach is that entirely anonymous clones can be used as probes. The procedure is more efficient and accurate than ordering the clones directly. Alignment of the physical map to genetic data was done by hybridising genetically localised markers to the libraries, with some of these results communicated from users of the reference system.

Although the *S. pombe* genome is more than

200-fold smaller than the human genome, single clone hybridisations of a number similar to those carried out for the cosmid/P1 mapping would be sufficient to cover DNA the size of an average human chromosome with YAC clones (Table 2). The efficiency is even further improved by the

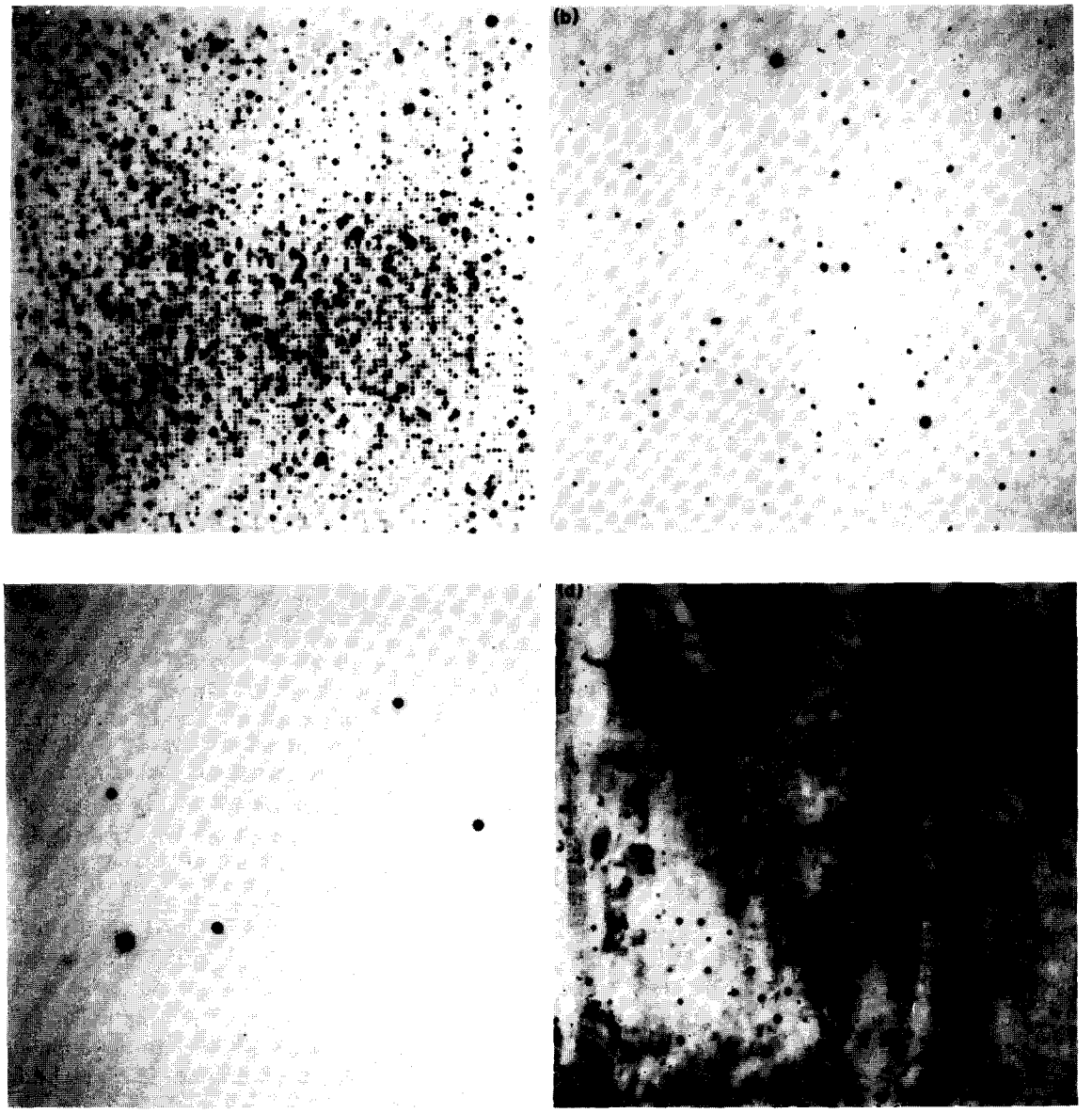

Fig. 2. Use of a variety of probes on high density cosmid filters. (a) Hybridisation pattern generated by the oligomer ATTATAAAAAAAAA. (b) Signal produced by a probe of the repetitive element pCS4 (Michelson, A., Tudor, M. and O'Hare, K., personal communication) on *Drosophila* cosmids. (c) Hybridisation on the same filter with a single-copy probe of the locus *actin42A*. (d) Probe from the entire non-polyA portion of an RNA-preparation from mouse liver was hybridised. The patchy filter background is due to the relatively large amount of radioactivity used in the experiments. (e and f) Inter-species comparison by oligonucleotide hybridisation: filters made from cosmid libraries of human chromosome 21 (e) and *Drosophila* (f) were hybridised with the oligomer ATACTTTAAGTTC, a sequence from the 3′ end of a human Line element.

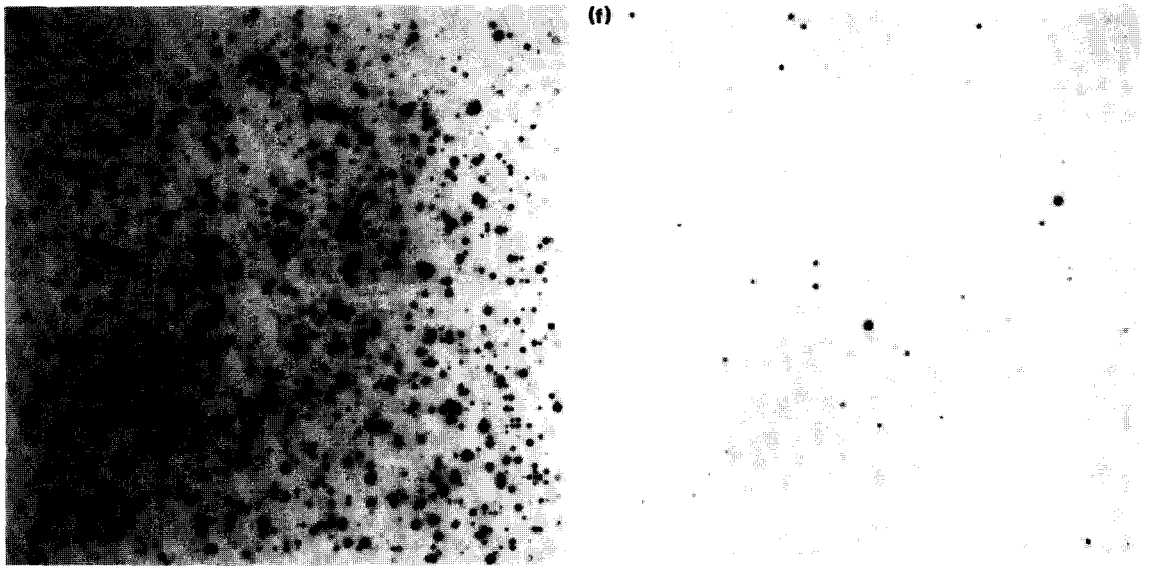

Fig. 2 (continued).

application of clone pools, although hybridisation results of a higher quality are required (Hoheisel et al., 1993). The use of pooling schemes (e.g., Evans and Lewis, 1989) allows the assignment of positive clones to a specific probe by matrix calculations. However, repeat sequences in the probe pools make discerning the relation between the observed signals and the individual clones responsible more complicated or even impossible, thus transforming these and all intersecting pools into multilocus probes rather than a group of unique probes. This complicates the exploitation of genomic clones for ordering mammalian libraries. However, cDNA or exon-trap clones could be used instead. Apart from the fact that they contain far fewer repetitive sequences, they will themselves be positioned onto the map, thus creating a map of transcripts. Work along this line is under way with *Drosophila* libraries and has started with human libraries. For example, pools of up to 74 cDNA clones have been hybridised to *Drosophila* YACs. Since they are also hybridised back to the entire cDNA bank, a variation of the 'sampling without replacement' strategy can be pursued and sequence homologies between the cDNA clones are defined simultaneously.

2.2.2. Complex probes

Although a complex probe will generally not attach specific information to an individual clone, the gain in information about entire libraries can be large. Hybridisations such as shown in Fig. 2, for example, indicate the position of transcription units without a polyA tail. In related experiments, probe from total polyA$^+$ RNA made from different mouse and human tissues was hybridised to filters containing both genomic clones and cDNAs made from either human or *Drosophila* (Gress et al., 1992). Experiments of this kind produce information about species- and tissue-specific expression of the transcribed DNA of the probe and the target library, and might provide essential clues to the function of the mammalian genes from the positioning of cross-hybridising human or mouse sequences on the genetically exceedingly well analysed *Drosophila* genome. As another example for the significant contribution of a complex probe, the subdivision of total human YAC libraries into specific chromosome clone collections was achieved by using a composite probe made from a flow-sorted human chromosomal material (Ross et al., 1992b) or *Alu*-PCR products generated from radiation hybrids as a first step towards the mapping analysis of YAC

libraries. Although complex probes on their own produce more global information about a library, they add considerably to the knowledge on individual clones when analysed in connection with more specific results obtained from other experiments.

2.2.3. Oligomer fingerprinting

Mapping by oligonucleotide hybridisation, which was demonstrated to work very effectively in a test-case on the Herpes simplex virus genome (Craig et al., 1990), combines the high data rates achievable by hybridisation techniques with the advantageous scaling behaviour that the amount of information gained per experiment is independent of the genome size (Lehrach et al., 1990). Furthermore, in comparison to clone hybridisations, there are far smaller effects from repeat sequences in the DNA. In order to achieve reasonable binding frequencies of 5% to 30%, oligomers of 10 to 12 nucleotides are used. The potential power of the method is clearly illustrated by the comparison of the number of hybridisations necessary to obtain sufficient information to map either with clones or oligomers an average sized human chromosome, such as the 160 Mbp X chromosome (Table 2), which is the same size as the *Drosophila* genome. The numbers shown are based on the assumption that the probes (clones and oligomers) are evenly distributed along the chromosome, which is certainly not the case in reality. Also, there can be deviations in the actual numbers as a result of the chosen oligomer sequences, for instance, or the application of non-random clone picking strategies, as mentioned above. Nevertheless, these factors should not affect the basic ratios shown in Table 2. Since the number of clone hybridisations increases exponentially with the size of the DNA analysed, the gap between clone and oligomer hybridisation widens with the size of the examined genome.

Experimentally, oligomers are unspecifically annealed at a temperature of 4°C to 8°C, while discrimination between full-match and mismatch hybridisation is achieved during the washing (Drmanac et al., 1990). The physical order of the library clones is determined by a statistical analysis of the fingerprint data (e.g., Michiels et al., 1987; Lehrach et al., 1990). Due to this fact, no gradually increasing physical map is produced until a significant number of hybridisations have been carried out and the project seems to be stationary for some time. However, due to the use of specific oligomer sequences (e.g., $(CA)_6$), additional information about the genome structure is already revealed during this phase.

Apart from 'random' sequences (e.g., Fig. 2a), oligomers representing di-, tri- and tetramer simple repeats or repeats of higher complexity (e.g., line elements, Fig. 2e and f) have been used, as well as a variety of oligonucleotides comprising consensus sequences. Such sequence-motif containing oligomers are also of major interest in the

Table 2
Calculated number of hybridisations for the mapping of 160 Mbp

Vector system	Average insert size (kbp)	Clone number for 10 × genome coverage	Hits per clone [a]	Single clone hybridisations	12 Clone pool hybridisations [b]	5% Oligomer hybridisations [c]
YAC	700	2300	4	920	150	–
P1	70	23000	4	9200	1500	60 (⇒ 6 hits per clone [d])
Cosmids	37	43200	4	17300	2900	120 (⇒ 6 hits per clone [d])

[a] Based on data from the mapping of *S. pombe* (Maier et al., 1992; Hoheisel et al., 1993).
[b] Two-dimensional pooling scheme that requires 24 experiments for the hybridisation of 144 clones.
[c] Hybridisations with oligomers which yield on average a 5% frequency of positive clones.
[d] Based on data from Craig et al. (1990) and Hoheisel et al. (1993).

analysis of cDNA grids (Fig. 4). In the actual experiments, the cosmid libraries of *Drosophila* (Hoheisel et al., 1991) and the human chromosome 21 (Nizetic et al., 1991), for example, are hybridised together to the same set of probes (Fig. 2e and f). This allows comparisons between the structure and organisation of the two genomes, which will become the more valuable the further the mapping progresses. The more data that accumulates, the more accurately the degree of homology between different libraries can be examined, because the growing oligomer map is in principle a very partial sequence determination of the genomes.

The mapping of *S. pombe* by different hybridisation techniques allowed an assessment of the oligomer data (Hoheisel et al., 1993). Missed positives were the most common discrepancy, with an overall false negative frequency of 26% for this data set. The potentially more troublesome false positives were much less frequent with about 8%. However, these relatively high percentages were mainly caused by the inadequate method of data gathering by digitising. The signal strength of a bound oligomer strongly depends on the amount of target DNA, which varies widely between cosmid clones because of differences in copy number in the cells and the cell number itself. Normalisation by a vector hybridisation is therefore a prerequisite for reliable data. It also

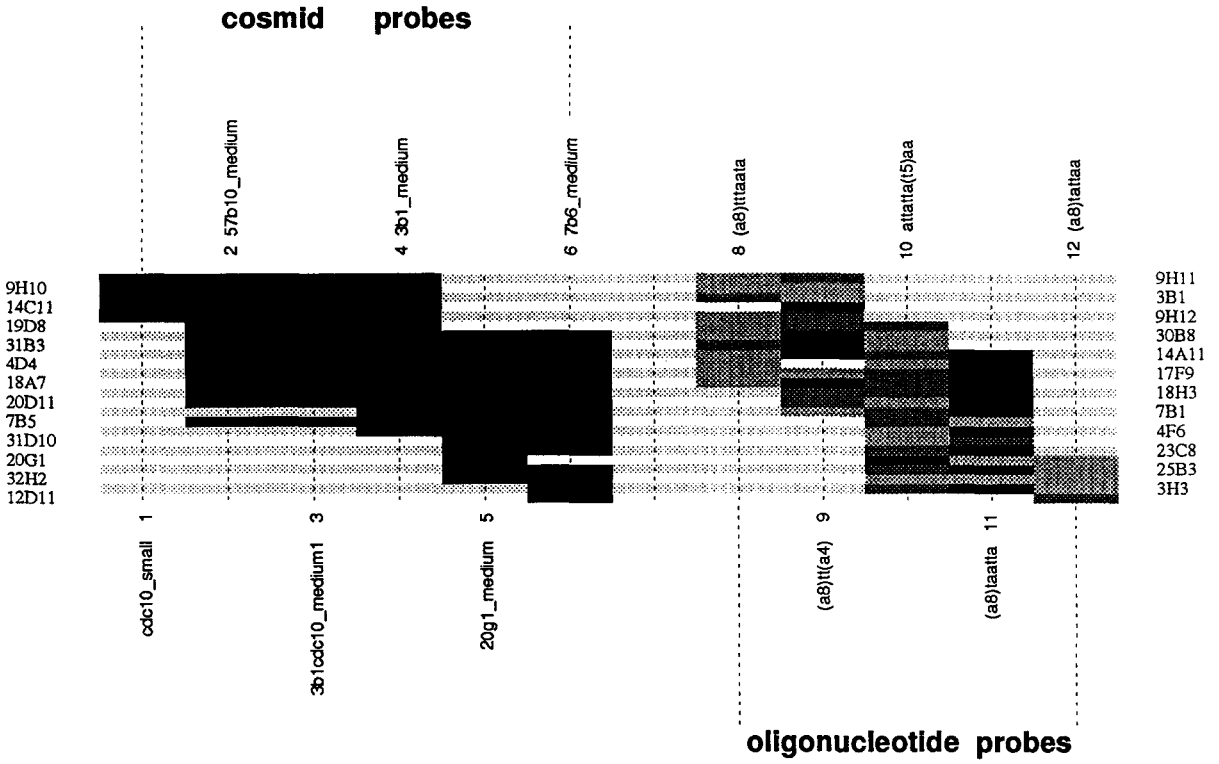

Fig. 3. Comparison of a cosmid map as established by clone and oligonucleotide hybridisation; a small part of the *S. pombe* map is shown. Probe names are given at the top and bottom, clone names at the right and left margin. The cosmid-probe signals were scored by digitising autoradiograms. The intensities of the oligomer hybridisations were quantitatively analysed using a PhosphorImager (Molecular Dynamics). The detection thresholds were adjusted according to empirical data for each oligomer. The grey-levels represent three classes of signal intensity.

improves the efficiency of the procedure, because a large number of less obvious positives is missed without background elimination. When the signal intensities were scored quantitatively using a PhosphorImager (Molecular Dynamics), the error rates were much reduced compared to manual analyses (unpublished results) and enabled the creation of contigs based on the oligomer data that concur with the clone order as determined from clone hybridisation data (e.g., Fig. 3). With the oligonucleotides a better resolution is achieved, since they detect a very short sequence within the cosmid targets, while cosmid probes span an area as wide as the target DNAs. Progress in automated image processing is imperative for the analysis of oligomer data, and is thus a focus of current efforts on the process of data capture.

2.2.4. Partial sequencing by hybridisation

cDNA and exon-trap libraries are subjected to oligomer hybridisations in order to generate partial sequence information on each clone (Lehrach et al., 1990; Drmanac et al., 1991; Meier-Ewert et al., 1993). This clone signature will allow the recognition of identical sequences, subdivide the libraries into clone categories, and, by comparison with known gene sequences, relate these to specific gene families and functions. Especially from sequence-motif oligonucleotides, the type of coding information contained in each clone is indicated. Paired with the information of their location within the genome, this will give each coding sequence of the library a specific label. It will prevent unnecessary redundancy in a complete sequence analysis, because similar sequences will already be known from their oligomer signature, and sequencing efforts can be concentrated on unknown genes.

Fig. 4 shows two such hybridisations to a *Drosophila* cDNA library. Since the filter spots contain genomic *E. coli* DNA besides the cDNA clone, only oligomer sequences that are rare in the *E. coli* genome can be hybridised this way. The *E. coli* contamination also prevents the use

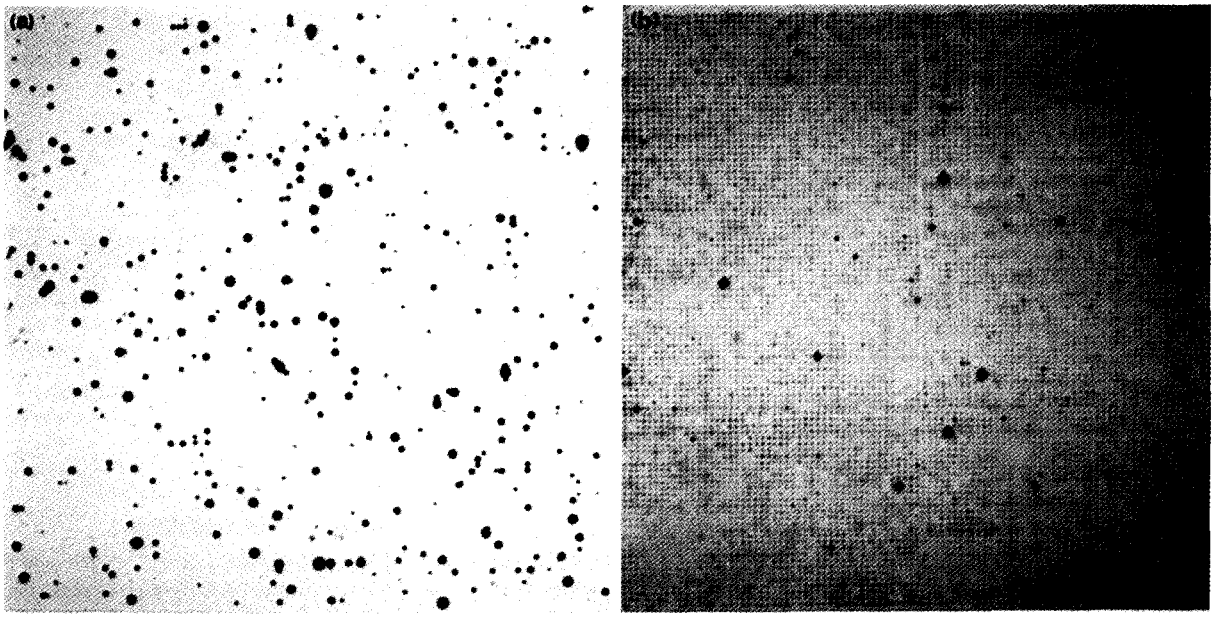

Fig. 4. Analysis of *Drosophila* cDNA clones by oligomer hybridisation. (a) Hybridisation of the oligonucleotide $(ATT)_4$. (b) Probing of cDNA clones with the zinc finger motif d(GGAGAGAAACC).

of very short oligomers (hexamers to octamers), which have the high frequency of clone recognition required for an effective analysis. Spotting of PCR-amplified material circumvents this problem. As described in this issue by Maier et al., we have developed the means to produce and analyse the PCR products of many tens of thousands of cDNAs and are in the process of further developing the degree of automation.

To achieve stable duplex formation with hexamer or even pentamer oligonucleotides, advantage is taken of base modifications (Hoheisel et al., 1990; Hoheisel and Lehrach, 1990) that improve the stability and specificity of oligomer binding. The improvement allowed the discriminative hybridisation of hexamer sequences to PCR products in a TLC system that mimics the conditions of filter hybridisation and also in actual filter experiments (Hoheisel, unpublished results).

2.3. Status of the work

The concept of a relational genome analysis by hybridisation has been developed into a working system. A wide variety of genomic and cDNA libraries have been generated and are distributed via the reference system. The robotics necessary for the handling of large clone numbers has been set up (see Maier et al., this issue) and the analysis procedures were tested successfully on the *S. pombe* genome. As another test-case for the refinement of the techniques, the analysis of the *Drosophila* genome is progressing well. Human and mouse libraries from all levels are being studied, work that will be complemented by data from other mammals (e.g., pig). Technically, non-radioactive detection methods and the achievement of much better hybridisation rates are being pursued as well as the development of improved tools and programs for image detection and analyses.

3. Discussion

A system has been set up to use high coverage reference libraries paired with various hybridisation techniques for a relational mapping analysis, which will lead to a better molecular understanding of many aspects of genome structure, organisation and functions (as well as malfunctions). The major advantages of the method are that there is easy and highly parallel access to large clone numbers; data acquisition by hybridisation is very effective, especially with oligonucleotide probes; probes can be hybridised to many libraries in a single experiment for instant correlation, due to shared experimental conditions; the experimental effort is markedly reduced, since clones from different levels are related to each other from the outset, and different types of data are generated simultaneously.

An alternative to the mapping approach taken by us is the use of sequence-tagged sites (STSs) for the development of long-range physical maps and their communication (Olson et al., 1989). For the creation of STSs, relatively short DNA stretches are isolated from the entire chromosome or genome and subsequently sequenced. 'Tag' primers are designed from areas near the segments' termini that allow their PCR-amplification. By hybridisation or direct PCR on clone libraries, the clone order can be established. Rather than the distribution of the clone libraries that is also necessary for STS mapping, however, the strategy proposes the communication of the primer sequences, so that a DNA region can be re-isolated from genomic libraries by using as probes the PCR products of STSs from that region. However, the advantages of the combination of the reference library system and hybridisation fingerprinting outweigh the disadvantage of having to store and distribute actual clones rather than mere sequence information. As opposed to the STS mapping information, the reference libraries are available to other laboratories from the very beginning of a project and can already be used while mapping is still in progress. Eventually, the results will be assigned to a physical position in the genome. Also, with the number of anticipated STSs, only YAC-sized fragments will readily be connected. More detailed maps will need the generation of very many probes. Furthermore, the retrieval of a certain genomic region requires a new cloning effort. For each new library clone quality must be checked and, most

possibly, a different DNA source is being used. For the reference libraries, any information about, for instance, potential cloning artefacts accumulates during their use and is passed on to all users. The central, relational reference database pools the results from many laboratories on many organisms and makes them easily available and comparable. The feasibility and attraction of central resource centres have been demonstrated by organisations like the Human Polymorphism Study Centre (CEPH) in France and the American Type Culture Collection (ATCC) as well as by the demand for library filters and identified clones from the ICRF reference system.

Apart from the analysis of large genomic areas by hybridisation fingerprinting, gel fingerprinting methods have been used extensively, as for the physical mapping of the genomes of *Caenorhabditis elegans* and *S. cerevisiae* (Coulson et al., 1986, 1991; Olson et al., 1986), but also for the analysis of human chromosomes (e.g., Carrano et al., 1989; Harrison-Lavoie et al., 1989). The fragment patterns generated by restriction-enzyme digestion of clones is used to detect clone overlaps and thus to construct a clone map. Gel and hybridisation fingerprinting differ in a number of respects, although in principle the mapping data generated are very similar in their analysis. An obvious difference is the amount of work involved per clone, due to the far fewer handling requirements of hybridisation techniques and their capacity to analyse almost unlimited clone numbers in parallel. Furthermore, hybridisation permits instant relation between different levels of analysis. Gel fingerprinting, on the other hand, is less sensitive to repeat sequences and exhibits a more favourable scaling behaviour than clone hybridisations (Lehrach et al., 1990). Relation between different analysis levels, however, is difficult to achieve and its range is limited. YACs larger than 120 kbp, for instance, produce too many restriction fragments to allow a regular comparison with cosmid clones (Coulson et al., 1991); in addition, it is technically difficult to obtain sufficient amounts of DNA for such an analysis. Oligomer hybridisation combines the advantageous features of both approaches. Its major drawback is the technical complexity of the experiments, such as the requirement of a quantification of the signal intensities. Many problems have been solved, though, as demonstrated by the results obtained on the *S. pombe* genome.

Eventually, a mixture of the wide variety of hybridisation techniques will provide the most effective way towards a comprehensive genome analysis. While, for example, oligomer hybridisations initially produce information more efficiently than other methods, in later stages of the analysis the gain in information from each oligomer hybridisation will be reduced in a situation analogous to the efficiency performance of random sequencing approaches. The selection as probes of clones from the end of contiguous clone arrays and candidate overlaps, or of as yet unlinked clones, would allow the targeting of regions not covered by oligonucleotide mapping data. Hybridisation of cDNAs towards genomic clones, on the other hand, will link not only the genomic clones, but also position transcripts on the map, while simultaneously information on the functions encoded will be revealed by oligomer hybridisation. The emergence of further automation based on recent technical advances, such as the replacement of filters with glass matrices (Khrapko et al., 1991), for example, and improved signal detection using non-radioactive labelling techniques, will place a comprehensive and highly resolving analysis of large genomes within the reach of our experimental capability.

Acknowledgement

We are grateful to Andrei V. Grigoriev for the kind provision of Fig. 3.

References

Ajioka, J.W., Smoller, D.A., Jones, R.W., Carulli, J.P., Vellek, A.E.C., Garza, D., Link, A.J., Duncan, I.W. and Hartl, D.L. (1991) *Drosophila* genome project: one-hit coverage in yeast artificial chromosomes. Chromosoma 100, 495–509.

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

Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber,

D.A., Sharp, P.A. and Housman, D.E. (1991) Exon-amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA 88, 4005–4009.

Carrano, A.V., Lamerdin, J., Ashworth, L.K., Watkins, B., Branscomb, E., Slezak, T., Raff, M., de Jong, P.J., Keith, D., McBride, L., Meister, S. and Kronick, M. (1989) A high-resolution, fluorescence-based, semi-automated method for DNA fingerprinting. Genomics 4, 129–136.

Cohen, D., Chumakov, I. and Weissenbach, J. (1993) A first generation physical map of the human genome. Nature 366, 698–701.

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

Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J. and Waterston, R. (1991) YACs and the *C. elegans* genome. BioEssays 13, 413–417.

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

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

Drmanac, R., Lennon, G., Drmanac, S., Labat, I., Crkvenjakov, R. and Lehrach, H. (1991) Partial sequencing by oligo-hybridisation: concept and application in genome analysis. In: Cantor, C.R. and Lim, H.A. (Eds.), Proc. First Int. Conf. on Electrophoresis, Supercomputing, and the Human Genome, World Scientific, Singapore, pp. 60–74.

Evans, G.A. and Lewis, K.A. (1989) Physical mapping of complex genomes by cosmid multiplex analysis. Proc. Natl. Acad. Sci. USA 86, 5030–5034.

Gress, T.M., Hoheisel, J.D., Lennon, G.G., Zehetner, G. and Lehrach, H. (1992) Hybridisation fingerprinting of high density cDNA-library arrays with cDNA pools derived from whole tissues. Mammalian Genomes 3, 609–619.

Harrison-Lavoie, K.J., John, R.M., Porteous, D.J. and Little, P.F.R. (1989) A cosmid clone map derived from a small region of human chromosome 11. Genomics 5, 501–509.

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

Hoheisel, J.D., Craig, A.G. and Lehrach, H. (1990) Effect of 5-bromo and 5-methyl deoxycytosine on duplex stability and discrimination of the NotI octadeoxynucleotide; quantitative measurements using thin-layer chromatography. J. Biol. Chem. 265, 16656–16660.

Hoheisel, J.D., Lennon, G.G., Zehetner, G. and Lehrach, H. (1991) Use of high coverage reference libraries of *Drosophila melanogaster* for relational data analysis; a step towards mapping and sequencing of the genome. J. Mol. Biol. 220, 903–914.

Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F. and Lehrach, H. (1993) High-resolution cosmid and P1 maps spanning the 14-Mbp genome of the fission yeast *Schizosaccharomyces pombe*. Cell 73, 109–120.

Khrapko, K.R., Lysov, Y.P., Khorlin, A.A., Ivanov, I.B., Yershov, G.M., Vasilenko, S.K., Florentiev, V.L. and Mirzabekov, A.D. (1991) A method for DNA sequencing by hybridisation with oligonucleotide matrix. DNA Sequence 1, 375–388.

Kohara, Y., Akiyama, K. and Isono, K. (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50, 495–508.

Larin, Z., Monaco, A.P. and Lehrach, H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA 88, 4123–4127.

Lehrach, H., Drmanac, R., Hoheisel, J.D., Larin, Z., Lennon, G., Monaco, A.P., Nizetic, D., Zehetner, G. and Poustka, A. (1990) Hybridisation fingerprinting in genome mapping and sequencing. In: Davies, K.E. and Tilghman, S. (Eds.), Genome Analysis Volume 1: Genetic and Physical Mapping, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 39–81.

Maier, E., Hoheisel, J.D., McCarthy, L., Mott, R., Grigoriev, A.V., Monaco, A.P., Larin, Z. and Lehrach, H. (1992) Yeast artificial chromosome clones completely spanning the genome of *Schizosaccharomyces pombe*. Nat. Genet. 1, 273–277.

Meier-Ewert, S., Maier, E., Ahmadi, A., Curtis, J. and Lehrach, H. (1993) An automated approach to generating expressed sequence catalogues. Nature 361, 375–376.

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

Monaco, A.P., Lam, V.M.S., Zehetner, G., Lennon, G.G., Douglas, C., Nizetic, D., Goodfellow, P.N. and Lehrach, H. (1991) Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridisation of *Alu*-PCR products. Nucleic Acids Res. 19, 3315–3318.

Mott, R., Grigoriev, A., Maier, E., Hoheisel, J.D. and Lehrach, H. (1993) Algorithms and software tools for ordering clone libraries; application to the mapping of the genome of *Schizosaccharomyces pombe*. Nucleic Acids Res. 21, 1965–1974.

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

Oliver, S.G., van der Aart, Q.J.M., Agostoni-Carbone, M.L., Aigle, M., et al. (1992) The complete DNA sequence of yeast chromosome III. Nature 357, 38–46.

Olson, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinmann, R. and

Frank, T. (1986) Random-clone strategy for genomic restriction mapping in yeast. Proc. Natl. Acad. Sci. USA 83, 7826–7830.

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

Ross, M.T., Hoheisel, J.D., Monaco, A.P., Larin, Z., Zehetner, G. and Lehrach, H. (1992a) High-density gridded YAC filters: their potential as genome mapping tools. In: Anand, R. (Ed.), Techniques for the Analysis of Complex Genomes, Academic Press, New York, pp. 137–154.

Ross, M.T., Nizetic, D., Nguyen, C., Knights, C., Vatcheva, R., Burden, N., Douglas, C., Zehetner, G., Ward, D.C., Baldini, A. and Lehrach, H. (1992b) Selection of human chromosome 21 enriched YAC sub-libraries using a chromosome-specific composite probe. Nat. Genet. 1, 284–290.

Smoller, D.A., Petrov, D. and Hartl, D.L. (1991) Characterisation of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75–100 kilobase pairs. Chromosoma 100, 487–494.