Ordering of cosmid clones covering the Herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation

A.G.Craig+, D.Nizetic, J.D.Hoheisel, G.Zehetner and H.Lehrach
Genome Analysis Laboratory, Imperial Cancer Research Fund Laboratory, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

Received January 25, 1990; Revised and Accepted March 27, 1990

ABSTRACT

To allow the efficient construction of ordered clone libraries, we have been investigating the use of 'oligonucleotide fingerprinting' as an approach to identify overlapping clones, and ultimately restore the linear order of the clone set. To test the effectiveness of the procedure, we have constructed a cosmid library from the genome of the human DNA virus HSV-I and used hybridisation to multiple oligonucleotides selected from the nucleotide sequence to reconstruct the order of clones and oligonucleotides on the genome.

INTRODUCTION

With the rapid increase in interest in cloning human genetic disease loci, the production of a high density molecular map based on ordered libraries of overlapping cosmid or yeast artificial chromosome (YAC) clones has become a high priority. The main approaches discussed to construct such ordered libraries fall into two classes, based either on fingerprinting strategies or on the use of hybridisation techniques analogous to chromosome walking.

Fingerprinting techniques, commonly based on the use of gel electrophoresis to generate restriction digest patterns from each clone, and their use to identify clone overlaps, has been applied successfully to produce ordered clone libraries of different organisms (1−3). This type of technique has two main advantages, the small increase in the mapping effort per clone with the size of the project (proportional to the logarithm of the number of clones) and the insensitivity of the protocol to interspersed repetitive sequences, expected to be a special problem in the analysis of mammalian genomes. Since individual clones have to be analysed in protocols involving a number of steps (growth, DNA isolation, enzymic reactions, gel electrophoresis, scoring of bands), rates of clone analysis are expected to be relatively low, though the rate of analysis can be increased by the use of automated systems using multiple fluorescent tags (4)

As an alternative, mapping approaches based on hybridisation are able to analyse large numbers of clones in parallel and therefore offer the theoretical possibility of much higher analysis rates. Approaches based on hybridisation of probes (derived from subgroups of clones) to clone filters have been proposed (5; Miles Brennan, personal communication), and partly used in the in the end stage of the analysis of the *E. coli* genome (3). Since the number of probes which can be used in each experiment is limited, the analysis of larger genomes requires linear increases in both the number of clones and the number of probe pools. The total effort in the analysis is therefore expected to increase with the square of the genome analysed, making this type of approach less favourable for larger genomes. In addition, hybridisation with pooled probes can be expected to be quite sensitive to repeat sequences in the probes, and especially to low copy number repeats in mammalian genomes.

To be able to combine the advantages of both types of protocols we have proposed a fingerprinting protocol (6−8) based on the use of oligonucleotides as hybridisation probes, which should combine the high data rates achieved by the parallel analysis of many clones in hybridisation experiments with the favourable scaling behaviour (N*log(N)) and repeat insensitivity of the gel fingerprinting approaches. Clones which show a similar pattern of hybridisation to the probes used ('hybridisation fingerprint') can be considered to be overlapping. This information on pairs of overlapping clones can be used to derive the linear order of clones and the oligonucleotide hybridisation sites on the genome.

We describe here an experimental test of the procedure. A cosmid library from HSV-I DNA was picked into 96-well microtitre dishes for storage and reference. Using a robotic device the cosmids were spotted in a high density pattern onto nylon membranes, grown in-situ, and converted into DNA by a colony lysis protocol. Filters were then hybridised to radioactively labelled oligonucleotides, selected from the sequence data (9), to assign oligonucleotide hybridisation signature to all cosmids.

Analysis of the pattern of hybridisation of oligonucleotides to the clones allowed the identification of the linear order of the cosmids within the HSV-I genome and identified cosmids originating from each of the four isomeric forms of HSV-I (10).

+ Present address: Molecular Parasitology Group, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

MATERIALS AND METHODS

Library Construction

The library was constructed in the cosmid vector Lawrist4 (a derivative of LoristB (11); Pieter de Jong, pers. comm.). Vector arms were prepared by digestion with ScaI, phosphatase treatment, and digestion with *Bam*HI. HSV-I DNA (the kind gift of Dr. D. McGeoch) was partially digested with *Mbo*I under conditions maximising the production of fragments in the size range 30 to 50 kb (12), followed by phosphatase treatment (to prevent ligation of non-contiguous fragments during library construction). Ligations were performed and the library was packaged as described (13). The library was plated on *E. coli* strain ED8767 (14).

After plating, single colonies were picked into 2×YT supplemented with Hogness freezing medium and 50 μg/ml kanamycin, contained in 96-well microtitre dishes. Picked colonies were incubated for 15 to 30 hours at 37°C until a majority of the wells contained saturated cultures.

Filter Preparation

Cosmid cultures in the microtitre dishes were spotted onto nylon membranes. This was carried out using a robotic device which has a 96-tip head which can be moved accurately over a working area to spot 96 cosmids in an 8×12 array onto a nylon membrane (GeneScreen Plus−Dupont). By interleaving 16 patterns of the 96-well microtitre dishes it was possible to spot 1,536 cosmids over an area of 8 cm×12 cm. Since this number is in excess of that required to ensure complete coverage of the HSV-I genome of 153 kb, we used 384 cosmids (4 microtitre dishes) representing approximately 100-fold genome coverage (50-fold considering the isomeric forms of the viral DNA) in this experiment, still far in excess of the 20 to 30 fold coverage expected to be used in the analysis of larger genomes (8). A complete high density array of clones was prepared by spotting each cosmid four times (in a row), serving as an internal replica.

After spotting the filters were placed colony side upwards on 2×YT agar plates supplemented with 50 μg/ml kanamycin and grown until discrete colonies could be seen (6 to 7 hours at 37°C). The filters were then treated as described (13) with an additional treatment, after the first denaturing step, of steaming the filters on a pad soaked in denaturing solution above a boiling waterbath for 5 minutes (modified from ref. 15).

Oligonucleotide Hybridisations

Oligonucleotides (12mers—see Fig. 1) were labelled with γ^{32}P-ATP using polynucleotide kinase (16). Filters were prehybridised in 0.5 M sodium phosphate, pH 7.2, 1 mM EDTA, 7% SDS for 30 minutes at room temperature. Hybridisations were carried out in the same solution with the addition of labelled oligonucleotide (at 2×10^6 cpm/ml) at 30°C overnight. After hybridisation the filters were rinsed in 8×SSC (1.2 M NaCl, 0.12 M sodium citrate) three times at room temperature followed by rinsing in 3 M tetramethylammonium chloride, 50 mM Tris-HCl, pH 7.6, 2 mM EDTA, 0.1% SDS (17) (TMAC wash solution). Stringent washes were carried out at 42°C for 2 times 2 minutes in TMAC solution, the filters blotted dry and autoradiographed.

After autoradiography, the filters were stripped of the radioactive oligonucleotide by washing in 2 mM Tris-HCl, pH 8.0, 0.2 mM EDTA, 0.1% SDS at 95°C for 2 times 20 minutes, followed by rinsing in several changes of distilled water.

probe	sequence (5'-3')	position(s)
HSV1	CACGCAACACTG	13 557
HSV2	GTATTCGTCAAA	13 817
HSV3	TTGGCCCGGTAC	17 784 + 96 709
HSV4	GCCCGAGGCGTG	25 572
HSV5	CACGCCCAGGTC	37 216
HSV6	TGTTGTGTCTCT	49 519
HSV7	GTGGATATGTCA	50 270
HSV8	CCGAGCCCCGTC	61 823
HSV9	CCGAGGACGCTC	73 773
HSV10	ATGGAAGGAACA	86 095
HSV11	TCTCGTCTCCTC	98 571
HSV12	CGAATGGCTATG	107 211
HSV13	CCTGGCTCTCAC	109 523
HSV14	AAGAAAATTTCA	114 770
HSV15	GCATGGCGCCAC	116 279
HSV16	GCCCGCCCTGAC	122 758 + 3 600
HSV17	CGGGTCGTTCAG	122 822 + 3 536
HSV18	CAGCCTTGGAGT	132 783
HSV19	ATTCATCTCAGC	133 525
HSV20	GTCGGTGTATCG	136 301
HSV21	CGGGTCGTGCAT	140 042
HSV22	CCAACAGTCGGT	143 461

Figure 1. Oligonucleotides used in the analysis. The sequence and position of the 22 oligonucleotides that were hybridised to the cosmid library are shown. Sequences were chosen mainly for their position in the HSV-I genome but were also screened for their absence in the cosmid vector.

RESULTS

Collation of Hybridisation Signals

The autoradiographs of the oligonucleotide hybridisations (Fig. 2) were read manually to produce binary records for each cosmid (Fig. 3). Different oligonucleotide probes reproducibly produced different signal intensities due to the different stability during the hybridisation step, as well as variable loss during the stringent washing step. Clone to clone variation is most likely due to variations in both colony size and cosmid copy number, detectable by using oligonucleotide probes which recognise the vector DNA (data not shown).

Data Analysis

Hybridisation patterns were used to deduce the order of the clones by either manual analysis or computer programs. One program used (Hoheisel, unpublished) starts from the most frequent pattern and gives preference to patterns present more than once. Copying the manual procedure, it looks for the least different patterns, optimally those with a single change, after the elimination of identical ones. By this means, all possible branches are extended until a clone is found which is already linked to another branch. Also, a program based on the Branching Bound algorithm was successfully used. Clone orders were then checked for consistency with the available DNA sequence.

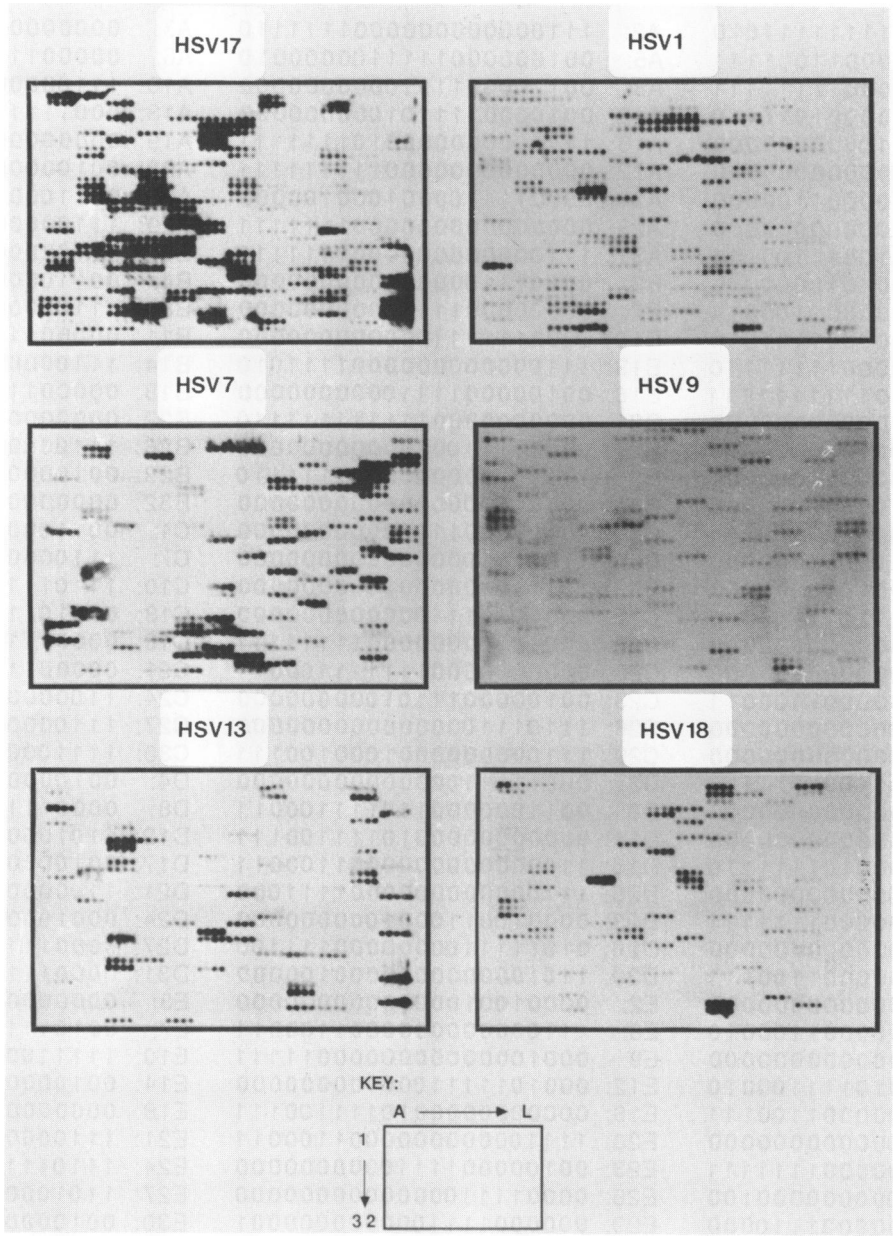


Figure 2. Autoradiographs of oligonucleotide hybridisations to the HSV-I cosmid library. Replica filters were hybridised as described in the MATERIALS AND METHODS section. Positive signals are seen as a row of four dots as the same cosmid was spotted four times per filter. The cosmids were labelled in rows (1−32) and in columns (A−L). For the positions in the HSV-I genome represented by the probes shown here see Fig. 1).

The analysis was complicated by the fact, that HSV-I exists as a mixture of four isomeric forms, which can be considered as subsegments of a permuted circular dimer sequence (Fig. 4). Accordingly, probes located on one end of the published sequence are found in association with probes from the other end. The different combinations of positive hybridisations with oligonucleotides flanking the repeat regions are shown in Fig. 5 along with the relevant oligonucleotide hybridisation data.

Figure 6(a) shows a subset of cosmids covering the entire HSV-I genome. The hybridisation patterns for these cosmids (Fig. 6(b)) can then be be used to derive an oligonucleotide map (Fig.6(c)—only one isomeric form shown). An overall map has been (re-)produced from both the gain(+) and loss(−) of probes. A probe density averaging 1 per 10 kb, in the long unique region, and 1 per 5 kb, in the region around the internal repeats, was sufficient to detect four isomeric forms (Fig. 5). The density of probes used for the analysis of complex genomes would however be likely to be higher than that used here, giving more information and thus helping the resolution of analogous 'difficult' regions.

The predicted order of the positions of twelve cosmids was confirmed by preparing DNA and digesting it with *Eco*RI and *Sca*I, and analysing the restriction digest patterns by agarose gel electrophoresis (data not shown). The fragment sizes produced from the DNA digests were in complete agreement with the predictions from the HSV-I DNA sequence.

While most probes fitted well into the expected pattern, unexpected hybridisation results were observed in a few cases. One such inconsistency was observed with the probe HSV3 which in addition to its expected location was found in combination with other oligonucleotide probes which could not possibly be present

A1: 000100000011111111010	A2: 111000000000000111110	A3: 000000011100100000000
A4: 000100000000001101111	A5: 001000000111110000010	A6: 000001111100000000000
A7: 111000000000001111111	A8: 001100001111000000000	A10: 111000000000000111111
A11: 111000000000000101110	A12: 001000011110100000000	A13: 000111111000000000010
A14: 001000001111000000000	A18: 111000000000010111111	A19: 000000000001011111101 0
A20: 001000011100000000000	A22: 000000000000011111111	A23: 001000000011001110000 0
A24: 111010000000000100000	A25: 000011110000100010000 0	A26: 001100011110100000000
A27: 111111100000000000000	A28: 000000000000000111111	A29: 111000000000000110011 1
A30: 000100000001001110011 1	A31: 111000000000000111111 0	B1: 111000000000000111101 0
B2: 111011100000001000000	B3: 001011100000000000000	B4: 001100001110000000000
B6: 111000000000000110001 1	B7: 001000001111000000000	B8: 111000000000000110001 0
B9: 111000000000000110101 0	B10: 000011111100000000000	B11: 000001110000100000000
B12: 000000000000001111111 0	B13: 111000000000000111101 0	B14: 111000000000000110001 0
B15: 000000000000111111111 1	B16: 001000001111000000000	B18: 000001111100000000000
B19: 000001111000000000000	B21: 000000000011111111110	B22: 000000000011011111111 1
B23: 000011110000000000000	B24: 000011110000000000000	B26: 111000000000000111101 0
B27: 000000011100000000000	B28: 111000000000000111111 0	B29: 001000001111000000000
B30: 001000001110000100000 0	B31: 000011100000000000000	B32: 000000000001011110101 0
C2: 111000000000000110011 1	C3: 001000001110000000000	C4: 001100000011011110001 0
C5: 001000001110100000000	C6: 111111100000000000000	C7: 111000000000100110001 0
C8: 001001111000100000000	C9: 111110000000000000000	C10: 111011100000100000000
C11: 001000000011101110001 0	C12: 000101111100000000000	C13: 000101111100000000000
C14: 111110000000000000000	C16: 000000000000001111111 1	C18: 000011110000000000000
C19: 000111110000000000000	C20: 000000000011111110001 1	C21: 000001111100010000000 0
C22: 110000001000000110011 1	C23: 001000001110100000000	C24: 110000000000100110011 1
C25: 000100011100000000000	C26: 111011100000000000000	C27: 111000000000000111001 0
C28: 111011100000000000000	C29: 111000000000100010011 1	C30: 111100000000000111111 0
C32: 111000000000100111111 0	D3: 000111110000000000000	D4: 001000000111010000000 0
D5: 000011110000000000000	D6: 001100000011011110001 1	D8: 000011111000000000000
D9: 111010000000000000000	D11: 000000000001011110011 1	D12: 110100000000000110001 1
D14: 111011011000010111111 0	D15: 110000000000000110001 1	D17: 001000001110000000000
D18: 000100011100000000000	D20: 111000000000000111100 0	D21: 110000000000000110010 1
D22: 000100000000000111111 1	D23: 000010011000100000000	D24: 000100000011011110001 1
D25: 000100011100000000000	D26: 010111110000000011110 0	D27: 000011110000000000000
D29: 111100000000000110011 1	D30: 110100000000000100000	D31: 000011110000000000000
D32: 111110000000000000000	E2: 000010010000000000000	E3: 000000011100000000000
E4: 111000000000000110001 0	E6: 111000000000000110001 1	E7: 001011110000000000000
E8: 000111100000000000000	E9: 000100000000000011111	E10: 111110000000000110000 0
E11: 000000000001011110001 0	E12: 000101111100000000000	E14: 001000001110000000000
E15: 111100000000000110011 1	E16: 000000000001011110011 1	E18: 000000000000100110000 0
E19: 001000011100000000000	E20: 111100000000000110001 1	E21: 111000000000000111111 1
E22: 111100000000000111111 1	E23: 001000001111000000000	E24: 111011100000000100000 0
E25: 000100111100000000100	E26: 000011110000000000000	E27: 110100000000000100011
E28: 110000000000000111000 0	E29: 000000111100000000001	E30: 001000001111000000000
E31: 000011100000000000010	E32: 001000000101111000001 0	F2: 000001111000000000000
F3: 001100001111000000000	F4: 110100000000000111101 0	F5: 111000000000000111111 0
F6: 111100000000000110001 1	F8: 111100000000001011000 11	F10: 000100000000011111111 1
F11: 111010000000000000000	F13: 001001001111110000000 00	F14: 111010100000000000000
F15: 000000000011011110001 1	F16: 000000011100000000000	F18: 111000000000000110001 1
F19: 000000000001111111111 0	F20: 001011110000100000000	F21: 000111110000000000000
F23: 100000000000001111111 1	F24: 000000111100100000000	F26: 000011110000000000000
F27: 001100000011011111101 0	F28: 011111100000000000000	F29: 000000111100000000000
F30: 111110000000000110001 0	F31: 111010000000000110001 0	F32: 000011010000000000000

Figure 3. Binary signatures of half of the cosmids in the HSV-I library. After scoring, a binary pattern was generated for each cosmid. Some entries are not listed due to the absence of a cosmid at these positions. The order of the oligo probes in each pattern is HSV1 to HSV22.

within the distance limited by the maximal size of a cosmid insert. The sequence recognised by HSV3 lies at position 17784 on the HSV-I genome but the results indicated an alternative association with probes 90 kb away. Inspection of the sequence identified a possible incomplete match at position 96709 of the HSV-I sequence due to an A-A mismatch at the 5′ end of the oligonucleotide.

Similarly inconsistent results were observed with probe HSV4 (mismatch at position 56708). Additionally, it did not hybridise to a subset of cosmids expected to span the region containing this oligonucleotide sequence, identified by the flanking probes HSV3 and HSV5. No differences in restriction pattern could be identified, ruling out large scale deletions as an explanation for this phenomenon. DNA sequence analysis of this region (data

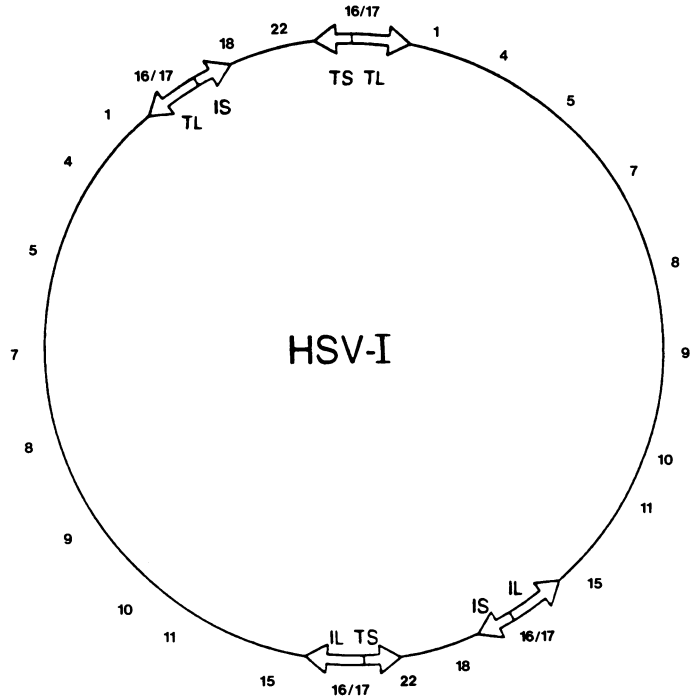


Figure 4. The appearance of the HSV-I genome derived from the oligonucleotide hybridisation data (as a result of the presence of four isomers). The approximate positions of a subset of the oligonucleotide probes are displayed to show the orientation of the molecules.

not shown) did however show a single base deletion, deleting the base opposite to position 10 from the 5′ end of HSV4. It is highly likely that the difference in hybridisation pattern is due to the detection of a polymorphic sequence in the HSV-I DNA used as starting material.

An overall estimate of the potential error rate was based on a thorough analysis of the results for the first 27 cosmids (series A), showing a total of 10 scoring errors not accounted for by stable mismatches during hybridisation in a total of 594 determinations (22 probes scored on 27 cosmids).

Two types of error were observed. As mentioned above, the absence of hybridisation of a probe to a cosmid which contains both flanking sequences could be explained by polymorphisms in the HSV-I DNA population or in the resulting cosmids. Alternatively, in some cases this may be due to the combination of a small colony with a weakly hybridising oligonucleotide probe, verified by hybridisation of the oligomer probe to purified DNA from these cosmids. The complementary error, scoring a colony as positive which was not expected to hybridise to an oligonucleotide probe, conversely might be due to an increased background caused by a larger than usual amount of DNA in a colony (which has been tested by hybridisation to a vector sequence oligomer probe). The error rate of 1.7% observed here should however be well within the expected tolerance of the procedure (6, 8).

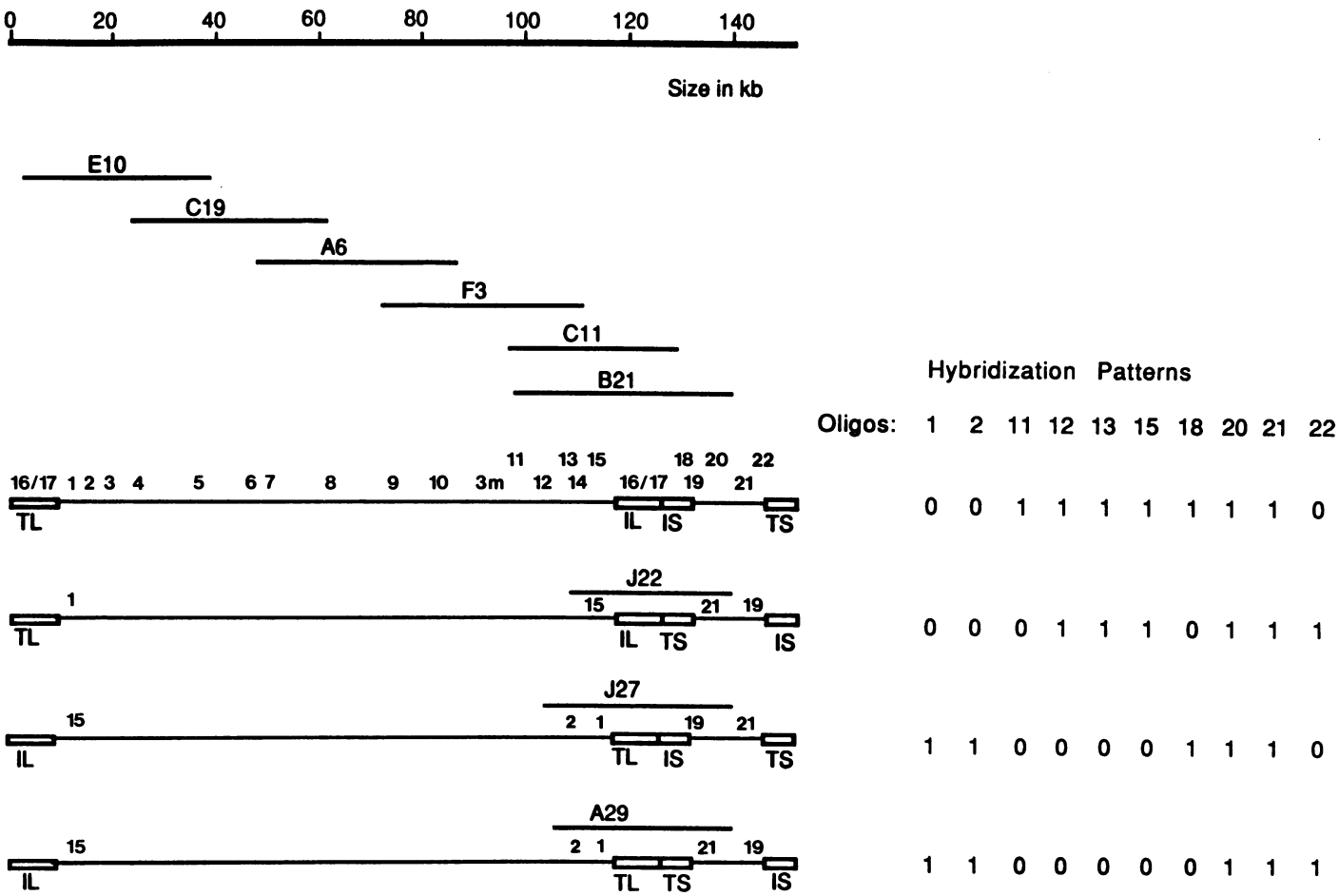
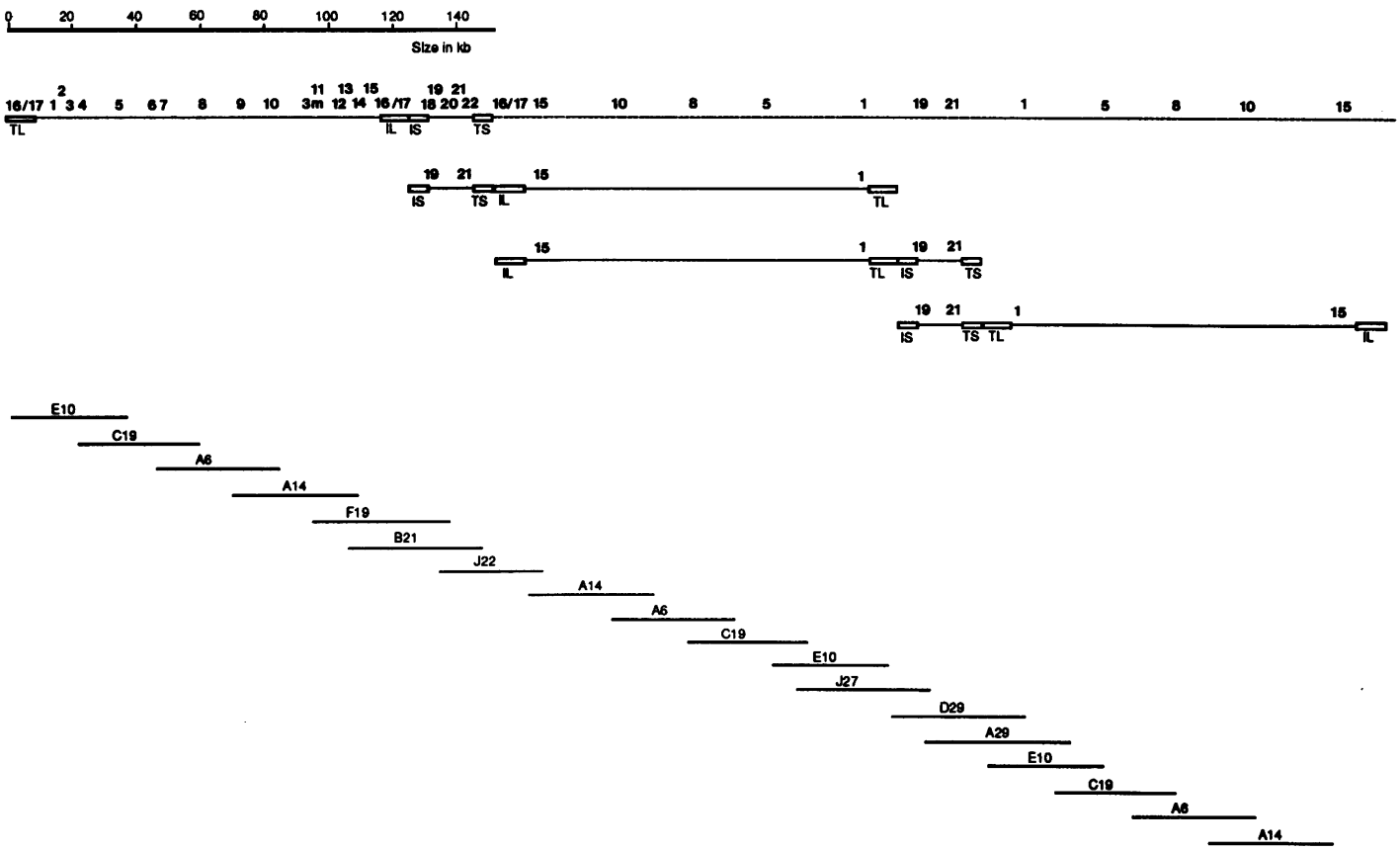


Figure 5. A set of cosmids representing the four isomers of the HSV-I genome. Analysis of the binary signatures allows the differentiation of four classes of cosmid representing the four isomers of the HSV-I genome. One example of each class is shown plus the relevant oligomer hybridisation data which permits the assignment of each cosmid. The labels for the cosmids refer to their positions on the filter.

a



b

Oligo:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	LOSS	GAIN
Cosmid:																								
E10	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0		
C9	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16, 17	
A27	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		7, 6
C19	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1, 3, 2	8
J18	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	9
A6	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	5	10
L27	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	6	
A20	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	7	3
C3	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	8	11
A14	0	0	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0		12
K12	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0		13
G20	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	9	14, 15
L28	0	0	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	10	16, 17, (6)
F19	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	3, 11, (6)	21, 19, 18, 20
G14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1		22
G15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	13, 12	
H3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	14, 15	
J22	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	1	19, 18	13, 14, 12, 15
C20	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	1	20	11
G20	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	21, 16, 17, 22	10, 3
|																								
E10	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0		
F4	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	5, 3	19, 18
J27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	3, 20	
E22	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	22	21
E21	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	4	
D29	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	19, 18	4
E20	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	20	
E10	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	21, 22	5

C

LOSS:	16	1	4	5	6	7	8	9	10	3m		13		14		19	20			21	
	17	3								(6)		12		15		18				16	
		2								11										17	
																				22	
GAIN:					7		8	9	10	3m	11	12	13	14	16			21	22		
					6									15	17			19			
															(6)			18			
																		20			
FINAL:	16	1	4	5	6	7	8	9	10	3m	11	12	13	14	16	19	20	21	22	16	
	17	3												15	17	18				17	
		2																			

Figure 6. **(a)** A reduced set of cosmids covering the HSV-I genome. Due to the isomeric nature of the HSV-I genome the picture seen on re-assortment of the cosmids is of a large circular genome (shown here as linear for clarity). Cosmids E10, C19, A6, and A14 are represented three times as their binary pattern is unaffected by isomerisation. The approximate positions of the oligonucleotides is shown on the top line. **(b)** Binary patterns for cosmids covering the HSV-I genome. The patterns, a subset of which are shown in Fig. 6(a), are detailed, with minimal changes between each step. The sequential loss and gain of oligo signals is shown on the right-hand side. The designations for the cosmids refer to their positions on the filter. The transient occurrence of probe HSV6 in cosmid L28 is probably an error, due to the high copy number of this cosmid. **(c)** Inferred oligonucleotide map of the HSV-I genome. The numbers shown represent the oligonucleotide probes ('3m' indicates the mismatch position of probe HSV3). From the sequential loss and gain of the oligonucleotide sequences from cosmids mapping along the HSV-I genome it is possible to construct oligo maps for both situations. By combining these a final map is produced which is in agreement with the predicted map (from the sequence data). The map for only one isomer is shown.

DISCUSSION

With the increase in the study of complex genetic diseases for which no direct methods of cloning exists, the need for high density molecular maps has become more apparent. As one possible route to the efficient construction of such ordered clone libraries we have proposed a new type of fingerprinting technique, which is based on scoring large numbers of clone DNAs spotted in a high density grid on nylon filters for hybridisation with short oligonucleotides expected to hybridise to multiple positions in the genome. The theoretical feasibility of such an approach has been demonstrated before (6, 8).

The comparison of the hybridisation patterns, analogous to other fingerprinting procedures, should lead to the identification of the linear order of the originally unordered clones, as well as the derivation of a linear map of oligonucleotide hybridisation sites in the genome (Fig. 6(c)). Once such oligonucleotide maps of specific chromosomes or genome exist, a small number of hybridisations should allow the localisation of of any new clone on the pre-existing map, as well as the rapid and efficient ordering of new clone libraries constructed e.g. in more advanced vector systems. In this the oligonucleotide fingerprinting shares essential features with the recently proposed use of 'sequence tagged sites' (STS, 18), allowing the efficient rederivation of both specific clones and ordered clone libraries from information stored in a data base. Such oligonucleotide maps (essentially a form of very partial sequence information, which in theory can be extended to approach the complete sequence (19)) can be used to compare genomes of closely related organisms, identify the position of mutations caused by deletions or translocations, or follow changes in the genomes of cancer cells during tumorogenesis and tumor progression.

Though for practical reasons the experiments described here have been carried out using dodecanucleotides known to occur in the HSV-I sequence (although sequence variants and additional hybridisation sites were identified), this should represent a fair test of the situation encountered in the analysis of unknown genomes. Different types of probes can be considered.

Completely random dodecanucleotides are expected to occur in double stranded DNA approximately once every 8 megabases (16 megabases single stranded DNA), corresponding to an expected frequency of 0.5% in cosmid libraries. If the sequence CpG is avoided in the selection of oligonucleotide sequences, an approximately two fold increase (to 1%) in hybridisation frequency is expected (reduced by the fraction of repeat sequences in the clones). Further improvements should be possible, if higher order Markov predictions are used, or if oligonucleotides hybridising more commonly than average are selected empirically. In addition, using modified hybridisation and washing conditions, a fraction (30 to 50%) of random undecanucleotides has been used successfully as hybridisation probes, leading to a four fold increase the predicted hybridisation frequency. Also, a large number of di, tri, and tetranucleotide repeats (e.g. GTGTGTGTGTGT) have been found to have hybridisation frequencies of up to 10% in cosmid libraries from mammalian genomes (unpublished data).

Though less than the optimal 30% to 50% (8), we expect even hybridisation frequencies of a few percent, easily achievable by the use of (random) eleven or twelve-mers, to be sufficient to allow an efficient fingerprinting of libraries covering mammalian chromosomes.

Probes hybridising to 2% of the clones (assuming binary scoring, and neglecting the effect of errors or polymorphisms) will on the average give 0.14 bits of information per clone and hybridisation cycle (approximately 5.6 bits for each hybridising colony, 0.003 bit for each colony, which does not hybridise with the probe), a total of 1300 bits for each hybridisation of a filter

containing approximately 10,000 colonies. Since in analogy to the experiments described for the multiplex sequencing approach (20) hybridisation of 80 to 100 filters per hybridisation cycle should be feasible, the conditions tested here should allow data rates of minimally 100,000 bits per hybridisation cycle, far in excess of the at most few thousand bits per gel generated with considerable more effort in each fingerprinting experiment. For probes hybridising to 10% of all clones, approximately 0.5 bits per clone would be generated, resulting in potential data rates of close to 5,000 bits per filter, or approximately 400,000 bits for each hybridisation cycle of 80 filters.

This pilot study has tested many of the experimental variables of the oligonucleotide fingerprinting technique. Errors (or polymorphisms), unavoidable in any system set up to handle large numbers of clones, have been shown to be occurring at a level well within the tolerances of this approach and have not prevented the ordering of cosmids covering the entire HSV-I genome. In addition, the unusual structure of the HSV-I genome proved a stringent test for this mapping approach due to the presence of four isomeric forms of the virus. Using a density of probes less than that anticipated for the analysis of mammalian chromosomes (8), cosmids specific for each of the four isomers could be identified (Fig. 5).

We expect this or similar approaches to allow the efficient construction of overlapping clone libraries from mammalian chromosomes and genomes, simplifying the identification of genes responsible for human mutations, as well as offering a possible analytical tool to identify changes in a genome occurring either in mutations or in cancer formation.

ACKNOWLEDGEMENTS

The authors wish to thank Duncan McGeoch for providing the HSV-I DNA, Iain Goldsmith for preparing the oligonucleotides, and Joe Kitau for modifying the robot device. We would also like to thank Janet Roden and Lisa Gellen for technical assistance, Michele Price for help in preparing the manuscript, and Annemarie Frischauf for comments on the manuscript. This work was funded by the Imperial Cancer Research Fund and an EMBO fellowship to Jörg Hoheisel.

REFERENCES

1. Coulson,A., Sulston,J., Brenner,S. and Karn,J. (1986) Proc. Natl. Acad. Sci. USA **83**, 7821–7825
2. Olson,M.V., Dutchk,J.E., Graham,M.Y., Brodeur,G.M., Helms,C., Frank,M., MacCollin,M., Scheinman,R. and Frank,T. (1986) Proc. Natl. Acad. Sci. USA **83**, 7826–7830
3. Kohara,Y., Akiyotama,K. and Isono,K. (1987) Cell 50, 495–508
4. Carrano,A.V., Lamerdin,J., Ashworth,L.K., Watkins,B., Branscomb,E., Slezak,T., Raff,M., De Jong,P.J., McBride,L., Meister,S. and Kronick,M. (1989) Genomics **4**, 129–136
5. Evans, G.A. and Lewis, K.A. (1989) Proc. Natl. Acad. Sci. USA **86**, 5030–5034
6. Poustka,A., Pohl,T., Barlow,D.P., Zehetner,G., Craig,A., Michiels,F., Ehrich,E., Frischauf,A-M. and Lehrach,H. (1986) Cold Spring Harbor Symposia on Quant. Biol. **51**, 131–139
7. Craig,A.G., Michiels,F., Zehetner,G., Sproat,B., Burmeister,M., Bucan,M., Poustka,A., Pohl,T., Frischauf,A-M. and Lehrach,H. (1986) Proc. 7th Int. Congr. Hum. Genet., Berlin 1986. Springer Verlag.
8. Michiels,F., Craig,A.G., Zehetner,G., Smith,G.P. and Lehrach,H. (1987) CABIOS **3**, 203–210
9. McGeoch,D.J., Dalrymple,M.A., Davison,A.J., Dolan,A., Frame,M.C., McNab,D., Perry,L.J., Scott,J.E. and Taylor,P. (1988) J. Gen Virol. **69**, 1531–1574.
10. Delius,H. and Clements,J.B. (1976) J. Gen. Virol. **33**, 125–133.
11. Cross,S.H. and Little,P.F.R. (1986) Gene **49**, 9–22.
12. Hoheisel,J.D., Nizetic,D. and Lehrach,H. (1989) Nucleic Acids Res. 17, 9571–9582.
13. Herrmann,B.G., Barlow,D.P. and Lehrach,H. (1987) Cell **48**, 813–825
14. Murray,N.E., Brammar,W.J. and Murray,K. (1977) Mol. Gen. Genet. **150**, 53–61
15. Maas,R., Silva,R.M., Gomes,T., Trabulsi,L. and Maas,W. (1985) Infect. and Immun. **49**, 46–51
16. Maniatis,T., Fritsch,E. and Sambrook,J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, New York.
17. Wood,W.I., Gitscher,L.A., Lasky,L.A. and Lawn,R.M. (1985) Proc. Natl. Acad. Sci. USA **82**, 1585–1588
18. Olson,M., Hood,L., Cantor,C. and Botstein,D. (1989) Science **245**, 1435–1436.
19. Drmanac,R., Labat.I., Brukner,I. and Crkvenjakov,R. (1989) Genomics **4**, 114–128.
20. Church,G.M. and Kiefer-Higgins,S. (1988) Science **240**, 185–188.