Jens Hanke[1]
Marcus Frohme[1]
Jean-Pierre Laurent[2]
John Swindle[2]
Jörg D. Hoheisel[1]

[1]Molecular-Genetic Genome Analysis, Deutsches Krebsforschungszentrum, Heidelberg, Germany
[2]Seattle Biomedical Research Institute, Seattle WA, USA

Hybridization mapping of *Trypanosoma cruzi* chromosomes III and IV

As part of the *Trypanosoma* Genome Initiative launched by the World Health Organization (WHO), a physical clone map of *Trypanosoma cruzi* chromosomes III and IV was generated to facilitate both DNA sequence analysis of the parasite's genome and the investigation of chromosome organization. Apart from a few genetic markers, anonymous cosmids were taken from chromosomal sublibraries and individually hybridized to filter arrays of the relevant cosmid library. The probe order was determined from the hybridization fingerprint results and used to define a fitting clone order, with few gaps remaining. The results were independently verified by hybridizations to a bacterial artificial chromosome (BAC) library and, in case of chromosome III, restriction mapping. For gap closure, additional experiments on a total cosmid library were carried out. The possible tiling paths consist of 26 clones for chromosome III (610 kbp) and 28 clones for chromosome IV (680 kbp).

1 Introduction

Mapping by hybridization has proved to be a powerful tool in genome analysis [1]. Hybridization allows a simple and fast identification of homologous stretches of nucleic acid and permits parallel examination of large clone numbers, as the degree of handling per individual clone is small. Since any piece of nucleic acid can be used both as a probe and a target, the experiments can be arranged to yield the maximum amount of information. The degree and thus the effect of parallelism is increased by the use of robotic devices for the gridding of large clone numbers on small surfaces. Using a total genomic cosmid library, this methodology is currently applied to the generation of a map of the entire *T. cruzi* genome (J. Bua *et al.*, unpublished results). Because of its relevance to human disease, *T. cruzi* has been chosen as one target organism within the Parasite Genome Initiative of the World Health Organization (WHO), with the initial aims being the generation of a clone map and the sequence analysis of genes and genomic regions [2]. Within a collaborative effort, it was decided to map independently the four smallest chromosomes of reference strain CL Brener. The resulting maps should not only serve as starting material for DNA sequencing but also provide data for the identification and analysis of experimental problems that may be encountered during the whole genome approach.

2 Material and methods

2.1 Hybridization

Filter arrays of chromosome-specific sublibraries were generated as described elsewhere in this issue [3]. DNA of individual clones was isolated by alkaline lysis and checked on agarose gels subsequent to *Eco*RI cleavage. Labeling of complete cosmid DNA or *Eco*RI fragments was by random hexamer priming [4]. Cross-hybridization of the vector portion in the former probe was reduced by a competition reaction with sonicated vector DNA prior to hybridization [5]. Both hybridization and the washing procedure were carried out in glass tubes. Hybridization was in 0.5 M sodium phosphate, pH 7.2, 7% SDS, 1 mM EDTA, 0.1 mg/mL yeast tRNA at 65°C overnight, the probe concentration being 0.5–1 Mcpm/mL. Each filter was washed twice at 65°C in 40 mL of 40 mM sodium phosphate, pH 7.2, 0.1% SDS, for 30 min. Subsequently, filters were briefly blotted dry and film was exposed at −70°C using intensifying screens. To strip off the probe, filters were heated to about 90°C for 30 min in 0.5 L of 5 mM sodium phosphate, pH 7.2, 0.1% SDS. They were then immediately transferred to prehybridization conditions, which were identical to those of the actual hybridization but for the probe, and stored between experiments under these conditions for up to eight weeks.

2.2 Contiguous assembly

The hybridization results were scored manually or by using a CCD camera system and image analysis software (BioImage HDG Analyzer software, version 2.1). For the construction of contiguous clonal arrays, the hybridization fingerprint data was analyzed by a special software package [6], which can be freely downloaded from http://www.mpimg-berlin-dahlem.mpg.de/~andy/welcome.html. In short, the procedure orders the probes rather than the clones, since there are far fewer probes than library clones. Moreover, a comparison of two probes is based on the information from the relatively large number of positive clones, while a clone-clone comparison relies on less data. Between each possible pair of probes a distance was calculated from the percentage of clones hybridizing to only one probe but not to both. From such distances, the shortest possible linear succession of all probes was calculated by a simulated annealing algorithm, and the probes were ordered in contigs. Only subsequently were the clones fitted to the given probe order [1].

Correspondence: Dr. Jens Hanke, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg (Tel: +49-6221-424681; Fax: +49-6221-424682; E-mail: j.hanke@dkfz-heidelberg.de)

Abbreviation: **BAC**, bacterial artificial chromosome

Keywords: *Trypanosoma cruzi* / Hybridization mapping / Chromosomal maps

© WILEY-VCH Verlag GmbH, 69451 Weinheim, 1998

0173-0835/98/0404-0482 $17.50+.50/0

2.3 Genetic marker and repeat probes

A plasmid clone of the 3′-portion of the gene coding for the major cystein protease cruzipain (cysprot) [7] was obtained from Lena Åslund (Uppsala, Sweden); the insert was PCR-amplified and labeled by random priming. During sequence analysis of one cosmid clone (chromosome III, 1c9) an abundant repeat sequence had been identified (Björn Andersson, personal communication) and subcloned as a *Pst*I fragment, which was also kindly provided by Lena Åslund. A PCR product of the repetitive RIME sequence [8] was a gift from Sara Melville (Cambridge, UK). An oligonucleotide resembling the telomeric repeat sequence $(CCCTAA)_{3.5}$ of the closely related parasite *T. brucei* [9] was hybridized after radioactive labeling of the 5′-end with T4-polynucleotide kinase according to the manufacturer's instructions.

2.4 Restriction mapping

For each digest, about 2 µg cosmid DNA was cut completely with *Eco*RI. Band patterns were examined in 1% agarose gels after ethidium staining. Fragment sizes were calculated with a software tool (Gel 1.01), written by Jean-Michel Lacrois (Gainesville, FL, USA). This program is available at ftp.embl-heidelberg.de.

3 Results

Since for chromosomes III and IV of *T. cruzi* only a few markers were known, mapping was based on the use of probes of anonymous clones taken at random from the library (Fig. 1). To increase the efficiency of the procedure, chromosome-specific cosmid sublibraries [3] were used in this study; each consisted of two clone subsets, one of high, the other of low specificity to the respective chromosomal probe. Probes were selected by an iterative process of 'sampling without replacement', being picked from the ever-decreasing number of library clones that had not been positive in earlier experiments until all clones were hit at least once. The probes, although anonymous, should be relatively evenly spaced by this strategy. Also, redundant analysis of existing contigs was avoided. In a second phase, contig end clones were used as probes in order to define undetected overlaps. Several contig breaks remained, however, after the sublibraries had been analyzed exhaustively. Since it could not be excluded that some chromosome-specific clones had been missed during the sublibrary selection process, clones from contiguous ends were hybridized to the total genomic cosmid library [10] and a bacterial artifical chromosome (BAC) library of about 2-fold genome coverage (J. P. Laurent and J. Swindle, unpublished results) in order to identify connecting clones. In addition, the existing contigs were independently verified by also hybridizing relevant probes to the BAC library. Altogether, some 140 and 160 probe hybridizations – which is about three times the initial number – were carried out for the analyses of chromosomes III and IV, respectively, in an extensive attempt to improve the map quality. However, mostly redundant data was produced, confirming the existing clone order.

In Fig. 2 the resulting cosmid maps are presented; probes which yielded redundant data are omitted. Only

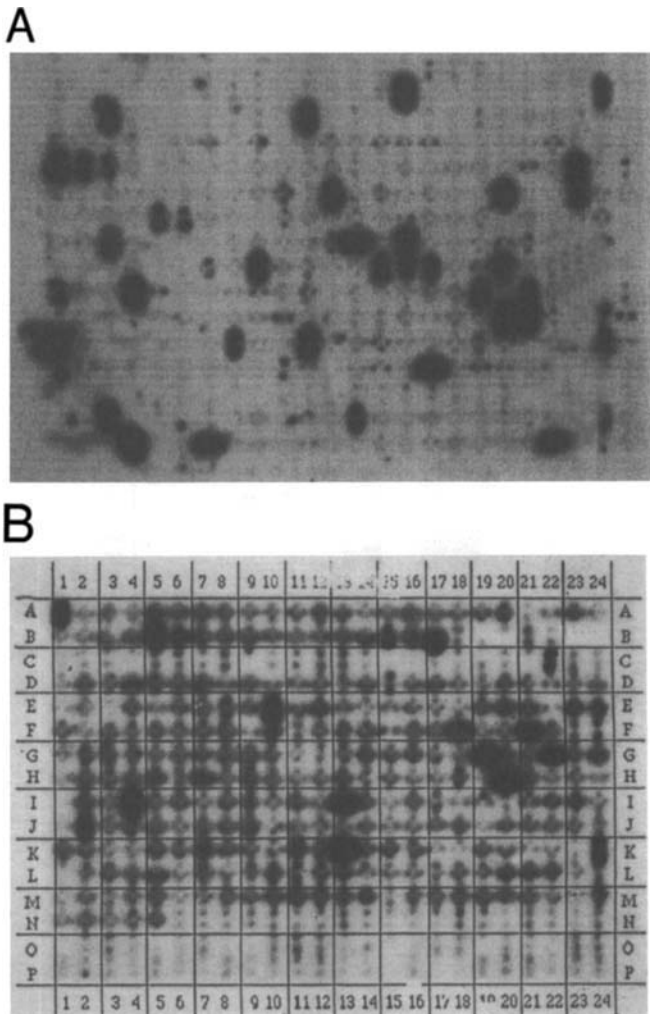

Figure 1. Typical hybridization results on chromosome IV-specific cosmid filters. Each library clone had been spotted onto the filters in duplicate [6]: subset 1 was spotted in a vertical, and clones of subset 2 in a horizontal arrangement. (A) Hybridization of cosmid 1f2 taken from the sublibrary; (B) hybridization of probe $(CCCTAA)_{3.5}$, which represents a telomeric repeat sequence of *Trypanosoma brucei*.

one gap (probes *5* to *6*) in the chromosome III cosmid coverage could be closed by using the BAC library. A second one was eliminated by hybridization of the genetic marker cysprot (probe *16*) [7]. Two remaining gaps (probes *17–18; 19–20*) could not be closed by any of the described measures. In a different approach, mapping was repeated, based on restriction fragment analyses [11] but was found no more successful (data not presented); again, only the existing contigs were verified. The map of chromosome IV was of overall better quality and only a single gap between probes *10* and *11* could not be bridged. The connection from probe *16* to *18* seems to be weak. However, analyses suggest that probe *17* is a chimeric clone; while its insert overlaps with the clone used as probe *16*, the remaining portion that overlaps with probe *18* is missing and replaced by an unrelated stretch of DNA. Removal of probe *17* would notably increase the significance of overlap *16* to *18*.

The tiling paths selected by the span option of the mapping program package [6] consist of 26 cosmids for chro-

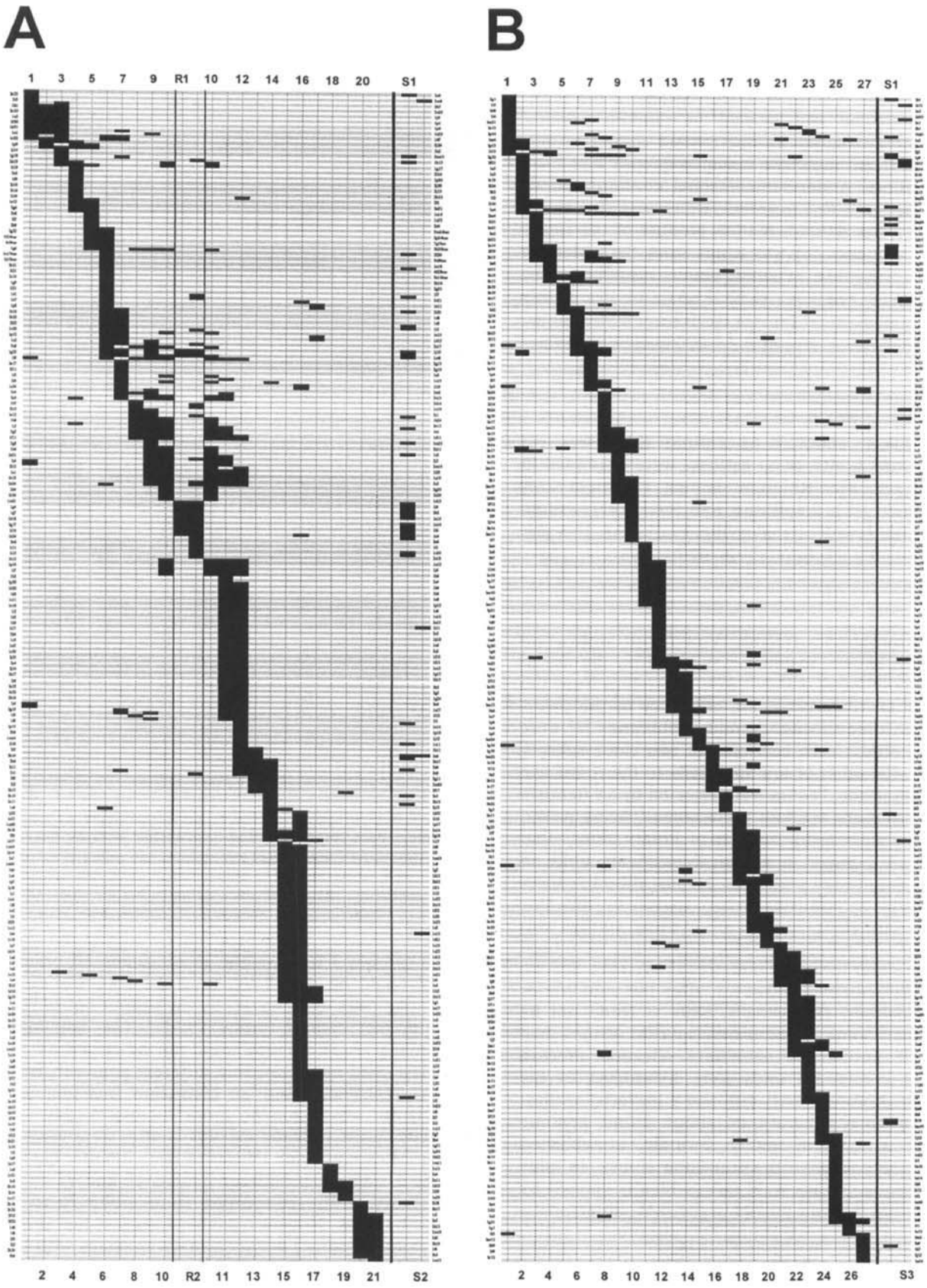

Figure 2. Cosmid map of chromosomes III (A) and IV (B). The hybridization results are presented in a two-dimensional matrix, probes correspond to columns, clones to rows. A black bar at an intersection represents a positive hybridization event. Probes *R1* and *R2* contained the RIME sequence and thus produced false overlaps if taken into account for map assembly. Here they were inserted only to illustrate their disruptive potential and their large degree of similarity to the results obtained with the actual RIME repeat; the map proceeds from probe *9 via 10* to *11*. Probes *S1* to *S3* were the RIME sequence, *Pst*I repeat and the *T. brucei* telomeric sequence $(CCCTAA)_{3.5}$; they were also not considered for map assembly but added to the given clone order.

mosome III (610 kbp) and 28 clones for chromosome IV (680 kbp); the average clone overlap of this spanning set, which is by no means minimal, was thus one third of each insert, equivalent to about 12 kbp. This is analogous to results obtained with this algorithm for other genomic maps based on such hybridization data [12].

The results obtained with the telomeric oligonucleotide – used in an attempt to identify terminal clones *via* a potential sequence homology between *T. brucei* and *T. cruzi* telomeric sequences – and the *Pst*I repeat are not clearly interpretable. The hybridizations produced distinct signals (Fig. 1). Surprisingly, however, these clones are distributed across the map; even in well-covered regions of the map, rarely are even two positive clones located next to each other (Fig. 2). The most probable explanation for this is the occurrence of an insertion in the genomic DNA of individual parasites of the cultured population causing polymorphisms unique in position and, possibly, copy number.

In an initial version of the chromosome III map, a loop structure seemed to be the best interpretation of some of the data, connecting two map positions by two possible routes. Further analyses, however, revealed that this result was caused by the overlapping occurrence of repeated sequences. Clones that were located in different regions of the chromosome were thus artificially clustered. The repeat mainly responsible for this phenomenon was the well-described RIME sequence (probe *S1*) [8]. Few of the clones that had been used as probes contained this repeat. To illustrate their effect, the relevant hybridization results of two probes (*R1* and *R2*) are presented in Fig. 2A, although they do not contribute to the overall clone order. The hybridization patterns were similar to the one obtained with the RIME probe itself.

4 Discussion

The mapping of the *T. cruzi* chromosomes III and IV indicates the complexity of the genomic DNA in this organism. Apart from the gaps and repetitive sequences discussed above, there were many minor complications that could not be completely solved even by extensive probings. One factor could be the use of sublibraries picked from a total library by the hybridization of chromosomal fragments [3]. By this means, some DNA from other chromosomes will be selected because of the existence of repetitive and homologous sequences. For chromosomes III and IV, larger, homologous chromosomes exist [3, 13] which could well be their diploid partners, differing only in the amount of repetitive sequence. It is therefore likely that in either case the maps represent a hybrid of both respective homologs. Only sequence analyses, however, will give a clear answer to this.

The very large number of hybridizations and the restriction analyses – carried out in the unsuccessful attempt to close the remaining gaps – suggest that some genomic regions are actually absent from the sublibrary and also the primary cosmid and BAC libraries. Alternatively, the unconnected contigs of the chromosome III map might not belong to the chromosome at all but be part of the chromosome III-specific sublibrary only because of sequences of low homology; any strong homology to chromosome III sequences would have linked them to the main contig during the mapping process. This hypothesis is supported by the fact that these contigs nearly exclusively consist of clones of the second subset of the chromosome III sublibrary, which had exhibited only weak interaction with the selective, chromosomal probe [3]; also, some could actually be found in the second subset of one of the other chromosomal sublibraries.

In conclusion, it can be said that *T. cruzi* is a difficult organism for molecular genomic analyses; the occurrence of many repeat sequences, homologous chromosomes and high genetic variability complicates any mapping effort, irrespective of the technique applied. For the global mapping project, the use of cDNAs as hybridization probes should reduce the problem of repeat sequences, apart from the intrinsic advantage of simultaneously positioning the genes on the physical clone map. The creation of a clone map prior to DNA sequencing seems to be an essential preparative step for the identification of rearrangements and homologous areas. Investigations on the exact degree of sequence polymorphism in chromosomes III and IV and their homologs await the results of ongoing sequence analyses.

The technical assistance by Sandra Clauder is gratefully acknowledged. We thank Lena Åslund, Björn Andersson and Sara Melville for providing material and information prior to publication. This work was partly funded by the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (TDR 940538).

Received July 27, 1997

5 References

[1] Hoheisel, J. D., Maier, E., Mott, R., Lehrach, H., in: Birren, B., Lai, E. (Eds.), *Integrated Genome Mapping by Hybridisation Techniques in Nonmammalian Genomic Analysis,* Academic Press, London 1996, pp. 319–346.

[2] The *Trypanosoma cruzi* Genome Consortium, *Parasitol. Today* 1997, *13*, 16–22.

[3] Frohme, M., Hanke, J.. Åslund, L., Pettersson, U., Hoheisel, J. D., *Electrophoresis* 1998, *19*, 478–481.

[4] Feinberg, A. P., Vogelstein, B., *Anal. Biochem.* 1983, *132*, 6–13.

[5] Hanke, J., Hoheisel, J. D. in: Rapley, R. (Ed.), *Construction and Screening of Cosmid Libraries in the Nucleic Acids Protocols Handbook,* Human Press, Totowa, New York, 1998 in press.

[6] Mott, R., Grigoriev, A., Maier, E., Hoheisel, J. D., Lehrach, H., *Nucleic Acids Res.* 1993, *21*, 1965–1974.

[7] Åslund, L., Henriksson, J., Campetella, O., Frasch, A. A. C., Pettersson, U., Cazzulo, J. J., *Mol. Biochem. Parasitol.* 1991, *45*, 345–347.

[8] Hasan, G., Turner, M. J., Cordingley, J. S., *Cell* 1984, *37*, 333–341.

[9] van Leeuwen, F., Wijsman, E. R., Kuyl-Yeheskiely, E., van der Marel, G. A., van Boom, J. H., Borst, P., *Nucleic Acids Res.* 1996, *24*, 2476–2482.

[10] Hanke, J., Sanchez, D. O., Henriksson, J., Åslund, L., Pettersson, U., Frasch, A. C., Hoheisel, J. D., *BioTechniques* 1996, *21*, 686–693.

[11] Olson, M. V., Dutchik, J. E., Graham, M. Y., Brodeur, G. M., Helms, C., Frank, M., MacCollin, M., Scheinman, R., Frank, T., *Proc. Natl. Acad. Sci. USA* 1986, *83*, 7826–7830.

[12] Hoheisel, J. D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A. V., Schalkwyk, D., Nizetic, F., Francis, F., Lehrach, H., *Cell* 1993, *73*, 109–120.

[13] Henriksson, J., Porcel, B., Rydåker, M., Ruiz, A., Sabaj, V., Galanti, N., Cazzulo, J. J., Frasch, A. C. C., Pettersson, U., *Mol. Biochem. Parasitol.* 1995, *73*, 63–74.