Marcus Frohme[1]
Jens Hanke[1]
Lena Åslund[2]
Ulf Pettersson[2]
Jörg D. Hoheisel[1]

[1]Molecular-Genetic Genome Analysis, Deutsches Krebsforschungszentrum, Heidelberg, Germany
[2]Department of Medical Genetics, University of Uppsala, Biomedical Center, Uppsala, Sweden

Selective generation of chromosomal cosmid libraries within the *Trypanosoma cruzi* genome project

From a total genomic cosmid library of the pathogen *Trypanosoma cruzi*, specific sublibraries of the smallest four chromosomes were isolated by hybridization of the respective chromosomal bands obtained from pulsed-field gels. These libraries form the basis for initial mapping analyses that should provide information useful for both the ongoing physical mapping of the entire genome and eventual sequence analyses. Selectivity of the procedure was high with 75% to 92%, although cross-hybridization had to be expected from ubiquitous DNA features, such as centromeric and telomeric sequences, and other regions homologous between individual chromosomes. Overall, the number of identified clones was slightly higher than expected but well within the intrinsic experimental variation considering the uncertainty about the exact genome size, the variability in clonability and the higher frequency of repeat sequences in larger chromosomes. Chromosome III- and IV-specific cosmids were analyzed on Southern blots of chromosomal separations. For strain CL Brener, all clones tested exhibited cross-hybridization to a homologous chromosome larger than 1 Mbp, supporting the assumption of the respective chromosome couple being diploid pairs.

1 Introduction

The members of the genus *Trypanosomatidae* are interesting in two aspects. From the medical point of view, they are important human parasites, causing severe diseases, such as sleeping sickness evoked by *Trypanosoma brucei* in Africa, and Chagas' disease by *Trypanosoma cruzi* in Latin America. Biologists view them foremost as evolutionarily very old eukaryotes with numerous special features [1], such as: mitochondria resembling kinetoplast with its maxicircle and minicircle DNA, the RNA editing process and trans-splicing, the unusual cap and nucleotide structures, and the highly effective parasite strategies for an evasion of the host's immune response and for cell invasion based upon the recruitment of lysosomes to achieve intracellular persistence.

Within the WHO-sponsored *T. cruzi* genome project [2], one major goal is the generation of a physical clone map of the parasite's genome as a resource for subsequent sequence analysis. Of special interest are regions that might be of critical importance to harmful activities of the organism, thus providing potential targets for diagnosis and treatment. The *T. cruzi* genome exhibits a high genetic variability of chromosome sizes. Moreover, poor condensation of the chromosomes hampers evaluation of the karyotype. For the isolate CL Brener, which was chosen as reference strain [3], the chromosome number could only be estimated at around 65, corresponding to only some 42 apparent bands resolved in pulsed field gel electrophoresis (PFGE) in a size range of 0.45–4 Mbp [4]. The haploid nuclear DNA content is currently estimated at approximately 47 Mbp, if the entire genome is diploid, which is not yet known. In the present, still early stages of the genome project, the exact determination of the chromosome number and, moreover, the genome's exact ploidy status is still an important objective besides the creation of an adequate coverage in library clones.

A total genomic cosmid library of some 37000 clones has been created and is being distributed gridded on nylon filters [5]. From both practical experience [6] and theoretical calculations [7] it can be predicted that mapping the entire genome of about 47 Mbp would require some 4500 hybridization experiments if individual and unique probes were to be used. Thus, a significant amount of work has to be performed before any studies can be carried out on a genome-wide scale. For a more immediate analysis of the genomic structure, in what could be crucial information for the proceedings of the global mapping project, the smallest four chromosomes – in sizes ranging from 450 kbp to 680 kbp – were selected for a chromosome-specific pilot phase, mainly because of their good spatial separation in PFGE. To this end, chromosome-specific libraries had to be constructed for efficient progress of the mapping. Here, the procedure for the isolation of such libraries and their evaluation are presented, with emphasis on chromosomes III and IV; subsequent mapping results are described in detail elsewhere in this issue [8].

2 Materials and methods

2.1 PFGE

Agarose plugs containing epimastigotes of three clone stocks (CL Brener, CA-1/72 and Sylvio X10/7) were prepared as previously described [9] and kindly provided by Juan José Cazzulo (Buenos Aires, Argentina). PFGE separation was carried out on a Bio-Rad CHEF Mapper in 0.5 × TBE (44.5 mM Tris, 44.5 mM borate, 1 mM EDTA, pH 8.3) buffer at 14°C. For Southern blotting, a standard separation in the size range of 350–1800 kbp was carried

Correspondence: Marcus Frohme, Molecular-Genetic Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany (Tel: +49-6221-424683; Fax: +49-6221-424682; E-mail: m.frohme@dkfz-heidelberg.de)

Keywords: *Trypanosoma cruzi* / Cosmid libraries / Selective hybridization / Clone mapping / *Trypanosoma cruzi* genome project

0173-0835/98/0404-0478 $17.50+.50/0

out in 1% agarose, using linear ramping with pulse times of 58–169 s at 6 V/cm. DNA corresponding to about 5–10 million epimastigotes was used in each well. The DNA was transferred to nylon membranes by standard techniques [9]; cosmid DNA was labeled and hybridized as described in detail elsewhere [8]. Preparative PFGE was done in 1% SeaPlaque GTG LMP agarose using a pulse time of 55 s at 6 V/cm for 40 h. On these gels, DNA equivalent to about 15–30 million epimastigotes was used per well.

2.2 Sublibrary selection

The construction and characterization of a total genome cosmid library was described in detail elsewhere [5]. Chromosome-specific bands were made visible in PFGE separations and labeled directly in the agarose block by random hexamer priming [10]. Probe hybridization to filter arrays of the total cosmid library was carried out overnight in 0.5 M sodium phoshate, pH 7.2, 7% SDS, 1 mM EDTA, 0.1 mg/mL yeast tRNA at 65°C; filters were washed in 40 mM sodium phosphate, pH 7.2, 0.1% SDS at 65°C, briefly blotted to remove excess liquid and exposed to X-ray films [5]. Positive clones were identified manually or by using a CCD camera system and an appropriate software package (BioImage HDG Analyzer software, version 2.1). Identified clones were picked from the total genome library into fresh 384-well microtiter dishes, grown overnight and stored at −70°C. Sublibrary filters were made by spotting the clones on nylon mem-

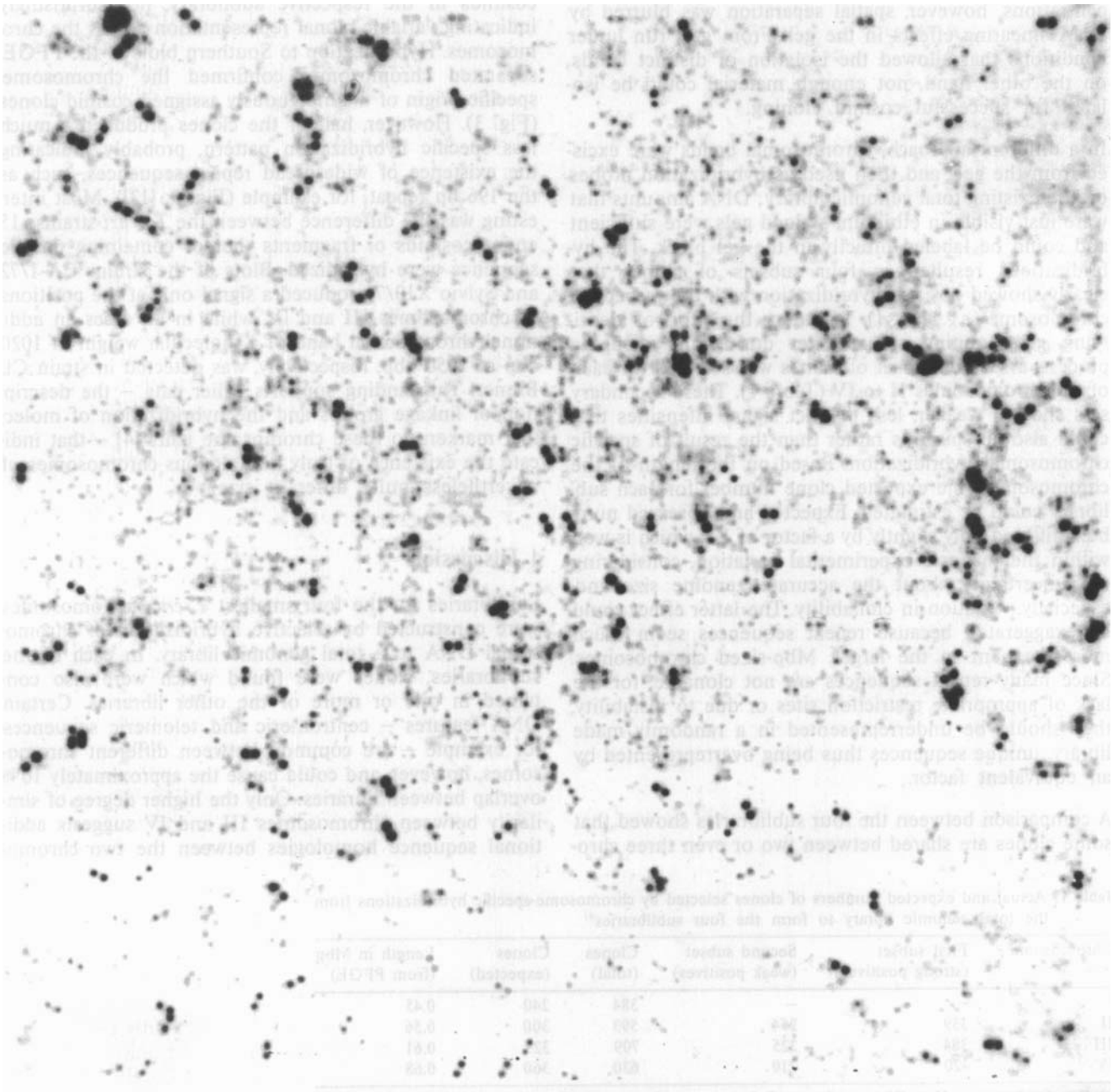

Figure 1. Selective chromosomal hybridization to total genomic library. Chromosome IV was isolated from PFGE and hybridized to a high density filter of some 18000 cosmid clones spotted in a 4 × 4 pattern with double offset.

branes (Hybond N$^+$, Amersham, Braunschweig, Germany) using a robotic device (BioRobotics, Cambridge, UK). Colony growth and filter processing was as reported [5].

3 Results

Initially, cloning of the four smallest *T. cruzi* chromosomes was attempted by directly using material of the respective chromosomal band isolated by PFGE, a method that had yielded good results in several projects, such as the generation of chromosome-specific libraries for the yeast sequencing project [11, 12]. However, excessively large amounts of DNA had to be loaded onto each lane in order to obtain DNA amounts of the small chromosomes sufficient for such cloning. At these high concentrations, however, spatial separation was blurred by DNA smearing effects in the gel. From gels run under conditions that allowed the isolation of distinct bands, on the other hand, not enough material could be isolated for successful cosmid cloning.

In a different approach, chromosomal bands were excised from the gels and then used as hybridization probes on the existing total genomic library; DNA amounts that were just visible in ethidium-stained gels were sufficient and could be labeled directly in the gel block. The hybridizations resulted in four subsets of clones that clearly showed positive hybridization with the respective chromosome (*e.g.*, Fig. 1). To reduce the effort of identifying gap-spanning clones later during the mapping process, a second subset of clones was selected for each of the chromosomes II to IV (Table 1). These secondary sets showed weaker, less distinct signal intensities that could also be spurious rather than the result of specific chromosomal hybridization. Based on the length of the chromosomes, the expected clone number for each sublibrary could be estimated. Expected and observed numbers differed only slightly by a factor of 1.2, which is well within the intrinsic experimental variation, considering the uncertainty about the accurate genome size and, especially, variation in clonability. The latter effect could be exaggerated because repeat sequences seem much more frequent in the larger, Mbp-sized chromosomes. Since many repeat sequences are not clonable, for the lack of appropriate restriction sites or due to instability, they should be underrepresented in a randomly made library, unique sequences thus being overrepresented by an equivalent factor.

A comparison between the four sublibraries showed that some clones are shared between two or even three chromosomal sublibraries (Fig. 2). However, this phenomenon was not caused by highly repetitive sequences present in the *T. cruzi* genome. None of the cosmids that were positive in a hybridization of total genomic DNA to the complete cosmid library was found in the four chromosomal sublibraries. Thus, cross-hybridization was most likely due to ubiquitous DNA features, such as specific centromeric sequences, other regions homologous between individual chromosomes, such as repeat sequences of low frequency for example, and contaminating material present in the PFGE bands used for hybridization. The percentage of clones that were unique to one of the sublibraries ranged between 75% and 92%.

Randomly chosen cosmids unique to the chromosome III and IV sublibraries showed hybridization to 5–50 cosmids in the respective sublibrary, not surprisingly indicating variable clonal representation across the chromosomes. Hybridization to Southern blots of the PFGE-separated chromosomes confirmed the chromosome-specific origin of unambiguously assigned cosmid clones (Fig. 3). However, half of the clones produced a much less specific hybridization pattern, probably indicating the existence of widespread repeat sequences, such as the 196 bp repeat, for example (Fig. 3; [13]). Most interesting was the difference between the *T. cruzi* strains: 15 entire cosmids or fragments thereof containing unique sequence were hybridized. Blots of the strains CA-1/72 and Sylvio X10/7 produced a signal only at the positions of chromosomes III and IV, while in all cases an additional chromosomal band of a molecular weight of 1020 kbp or 1050 kbp, respectively, was detected in strain CL Brener. This finding supports ealier data – the description of linkage groups and the hybridization of molecular markers to these chromosome pairs [4] – that indicate the existence of truly homologous chromosomes of nevertheless quite different size.

4 Discussion

Sublibraries for the four smallest *T. cruzi* chromosomes were constructed by selective hybridization of chromosomal DNA to a total genomic library. In each of the sublibraries, clones were found which were also contained in one or more of the other libraries. Certain DNA features – centromeric and telomeric sequences for example – are common between different chromosomes, however, and could cause the approximately 10% overlap between libraries. Only the higher degree of similarity between chromosomes III and IV suggests additional sequence homologies between the two chromo-

Table 1. Actual and expected numbers of clones selected by chromosome-specific hybridizations from the total genomic library to form the four sublibraries[a)]

Chromosome	First subset (strong positives)	Second subset (weak positives)	Clones (total)	Clones (expected)	Length in Mbp (from PFGE)
I	–	–	384	240	0.45
II	339	254	593	300	0.56
III	384	325	709	325	0.61
IV	420	210	630	360	0.68

a) For chromosomes II to IV two subsets of clones were chosen: the first subset represents the clearly positive clones, the second one contains clones with weak hybridization signals in order to provide candidates for gap closure during the subsequent mapping process.

somes, a result that could be validated by the appearance of some cosmid clones in the physical clone maps of both chromosomes III and IV [8], partly caused by the presence of repeats, for instance. Several dispersed repetitive DNA sequences are known to be present in *T. cruzi*, from the approximately 200 copies of the DGF element to the 10 000 copies of the E13 sequence [14], and could account for part of the overlaps. The latter, highly frequent repeats, however, should have caused a much larger degree of overlap than detected.

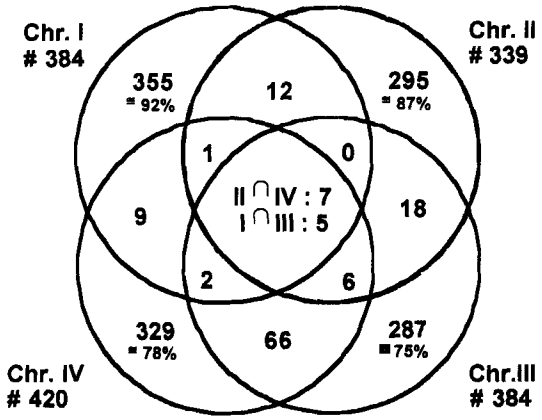

Figure 2. Clone distribution between the sublibraries made from chromosomes I to IV. Each circle resembles the clones of the first subset of each chromosome-specific sublibrary, except for chromosome I, for which no difference was made between the first and a second subset. Total clone numbers are presented as are the percentages of clones not contained in any of the other sublibraries.

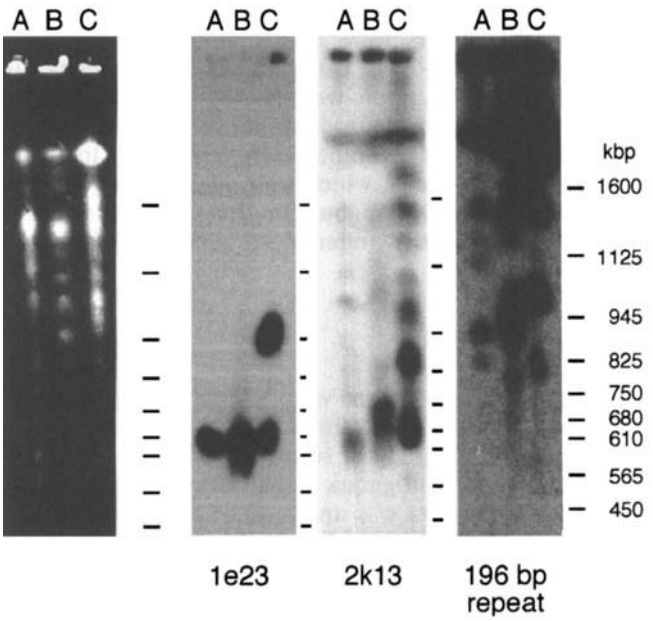

Figure 3. Typical Southern blot hybridization results obtained with chromosome IV-specific cosmids on PFGE-separated chromosomes of different *T. cruzi* strains. (A) Sylvio X10/7, (B) CA-1/72, (C) CL Brener. Clone *1e23*: all three strains present a signal at the position of chromosome IV; strain CL Brener exhibits an additional band at a molecular weight of about 1050 kbp. Similar observations were made for chromosome III. Clone *2k13*: the high overall signal seems to indicate the presence of repeat sequences. Right panel: hybridization of the 196 bp repeat sequence [13], which showed no apparent binding to chromosomes I to IV but was nevertheless found present in the respective chromosomal libraries.

The cross-hybridization of cosmids to a small and a large chromosome of strain CL Brener suggests a high degree of homology between the respective chromosomal pairs. Intriguingly, the number of clones assigned by the selection procedure to the primary subsets is about as high as predicted on such a basis. Also on Southern blots, the 196 bp repeat [13] seemed not to hybridize to the small chromosomes, but it was nevertheless found to be present in the isolated sublibraries. Consequently, it is probable that clones identified by hybridization of the chromosome-specific DNA to the total library could originate from the larger chromosome homologs and be an integral part of the sublibraries. This is supported by the fact that, in all cases tested, cosmids of unique sequence hybridized back to both chromosomes. This data along with earlier evidence [4] strongly suggests that the 1020 Mbp and 1050 Mbp chromosomes are the diploid partners to chromosomes III and IV, different in size mostly, if not exclusively, because of the amount of repeat sequence.

We thank Maria Rydåker for her skillful assistance. We have received financial support from the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (TDR) and the Swedish Agency for Research Co-operation with Developing Countries (SAREC).

Received August 27, 1997

5 References

[1] Boothroyd, J. C., Komoniecke R., *Molecular Approaches to Parasitology*, Wiley-Liss Inc., NY 1995.
[2] The *Trypanosoma cruzi* genome consortium, *Parasitol. Today* 1997, *13*, 16–21.
[3] Cano, M. I., Gruber, A., Vazquez, M., Cortéz, A., Levin, M. J., González, A., Degrave, W., Rondinelli, E., Zingales, B., Ramirez, J. L., Alonso, C., Requena, J. M., da Silveira, J. F., *Mol. Biochem. Parasitol.* 1995, *71*, 273–278.
[4] 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.
[5] Hanke, J., Sánchez, D. O., Henriksson, J., Åslund, L., Pettersson, U., Frasch, A. C. C., Hoheisel, J. D., *BioTechniques* 1996, *21*, 686–693.
[6] Hoheisel, J. D., Maier, E., Mott, R., Lehrach, H., in: Birren, B., Lai, E. (Eds.), *Analysis of Non-Mammalian Genomes – A Practical Guide*. Academic Press, New York 1996, pp. 319–346.
[7] Grigoriev, A. V., *Genomics* 1993, *15*, 311–316.
[8] Hanke, J., Frohme, M., Laurent, J.-P., Swindle, J., Hoheisel, J. D., *Electrophoresis* 1998, *19*, 482–485.
[9] Henriksson, J., Åslund, L., Macina, R. A., Cazzulo, B. M., Cazzulo, J. J., Frasch, A. C. C., Pettersson, U., *Mol. Biochem. Parasitol.* 1990, *42*, 213–223.
[10] Feinberg, A. P., Vogelstein, B., *Anal. Biochem.* 1983, *132*, 6–13.
[11] Scholler, P., Schwarz, S., Hoheisel, J. D., *Yeast* 1995, *11*, 659–666.
[12] Johnston, M., Hillier, L., Riles, L. *et al.*, *Nature* 1997, *387* (Supplement), 87–90.
[13] Sloof, P., Bos, J. L., Konings, A. F., Menke, H. H., Borst, P., Gutteridge, W. E., Leon, W., *J. Mol. Biol.* 1983, *167*, 1–21.
[14] Requena, J. M., López, M. C., Alonso, C., *Parasitol. Today*, 1996, *12*, 279–283.