

Proceedings of the Schistosome Genome Project

Manami Tanaka/⁺, Tomoo Tanaka*, Johji Inazawa**, Shigeo Nagafuchi***,
Youji Mutsui, Anne Kaukas****, David A Johnston****,
David Rollinson****

The National Institute of Bioscience and Human-Technology, Institute of Industrial Science and Technology, Tsukuba Science City, Ibaraki 305, Japan *Division of Host-Defence Mechanism, Department of Infectious Diseases, Tokai University School of Medicine, Isehara, Kanagawa 259, Japan **Department of Hygiene, Kyoto Prefectural University of Medicine, Kamigyo-ku, Kyoto 602, Japan ***Department of Pediatrics, Chigasaki City Hospital, Chigasaki, Kanagawa 253, Japan ****Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5 BD, United Kingdom

*"The host-parasite relationship" is a vast and diverse research field which, despite huge human and financial input over many years, remains largely shrouded in mystery. Clearly, the adaptation of parasites to their different host species, and to the different environmental stresses that they represent, depends on interactions with, and responses to, various molecules of host and/or parasite origin. The schistosome genome project is a primary strategy to reach the goal; this systematic research project has successfully developed novel technologies for qualitative and quantitative characterization of schistosome genes and genome organization by extensive international collaboration between top quality laboratories. Schistosomes are a family of parasitic blood flukes (Phylum Platyhelminthes), which have seven pairs of autosomal chromosomes and one pair of sex chromosomes (ZZ for a male worm and ZW for a female), of a haploid genome size of 2.7×10^8 base pairs (Simpson et al. 1982). Schistosomes are ideal model organisms for the development of genome mapping strategies since they have a small genome size comparable to that of well-characterized model organisms such as *Caenorhabditis elegans* (100 Mb) and *Drosophila* (165 Mb), and contain functional genes with a high level of homology to the host mammalian genes. Here we summarize the current progress in the schistosome genome project, the information of 3,047 transcribed genes (Expressed Sequence Tags; EST), complete sets of cDNA and genomic DNA libraries (including YAC and cosmid libraries) with a mapping technique to the well defined schistosome chromosomes. The schistosome genome project will further identify and characterize the key molecules that are responsible for host-parasite adaptation, i.e., successful growth, development, maturation and reproduction of the parasite within its host in the near future.*

Key words: schistosome - genome project - expressed sequence tags (EST) - yeast artificial chromosome (YAC) library - cosmid library - chromosomal *in situ* hybridization (CISS)

PROJECT OBJECTIVE

The purpose of this project is to initiate the physical mapping of the schistosome genome in a world-wide network, and molecular characterization of strain and species diversity.

The work was supported by grants from Ministry of Education, Science, and Culture, Japan (Nos. 06044032 and 06670253); Agency of Industrial Science and Technology Human Science Frontier Program; Nissan Science Foundation, Japan; Yamada Science Foundation, Japan; Ichiro Kanahara Foundation, Japan; the UNDP/ WORLD BANK/WHO Special Programme for Research and Training in Tropical Diseases Nos. 940332 and 940325.

⁺Corresponding author. Fax: +81-298-54-6503. E-mail: tmanami@nibh.go.jp

Received 20 August 1997

Accepted 10 September 1997

The specific aims and scheme of work are:

Phase I *Place known gene markers on the chromosome*

1) Construction of a Yeast Artificial Chromosome (YAC) library

2) Establishment of the standard technique for chromosome domain identification of YACs (chromosome *in situ* suppression hybridization; CISS).

Phase II *Retrieval of new transcribed sequences*
Screening of cDNA pools with YACs and cosmids.

Phase III *Characterization of YAC specific mini-libraries*

1) Construction of a cosmid library

2) Screening of high density grids (YAC and cosmid libraries)

3) Assemble chromosome specific YACs into contigs and link cosmids, together with visual links using CISS technique

4) Screening of YAC and cosmid inserts for new markers using repetitive sequences and known gene markers.

Phase IV *Characterization of the newly-identified transcribed segments*

1) DNA sequencing to cover whole region (Mb size regions) of the schistosome genome

2) Comparison of the sequences among stains and species

3) Assembly into transcription units (RNA analysis)

4) Development of *in vitro* and *in vivo* assay systems for transcribed genes and gene systems.

These studies cover the complete range of genome mapping strategies with world-wide collaborators, and have been already providing up to date information concerning genome structure, organization, and transcribed sequences in *Schistosoma*. Extensive contig assembly covering the entire chromosome sets and the construction of a reasonably detailed gene map should be feasible in the project.

SCIENTIFIC PROGRESS

Scientific progress is as follows (according to the original scheme of the experimental design):

Phase I *Place known gene markers on the chromosomes*

Phase I is the basic and essential process to start the project. It represents the preparation of total genome pool, and represents the standard technique for mapping of genes. We have established a YAC library and developed CISS technique (Tanaka et al. 1995a, b). These materials and techniques were distributed for common use among the whole of the schistosome genome mapping group to facilitate the mapping process.

1) Construction of YAC library - The numbers of YAC clones and the coverage of the schistosome genome by the library at the beginning of the project were 1283 clones, this represented a 1.7 fold coverage of the haploid genome. Good progress in the period under review has increased the number clones to more than 2283, and the library now represents a 2.6 fold coverage of the genome. This estimated total size of the library is enough as a genome pool to proceed towards the characterization and further analysis of the insert DNA (Tanaka et al. 1995a).

High density filter grids of the YAC libraries have been produced with the collaboration of Dr Hans Lehrach, Imperial Cancer Research Fund, London, and have been widely distributed among the WHO genome project collaborators (Drs Simpson, Pena, Rollinson, Pierce, and etc.) for hybridization screening.

2) Establish a standard technique for chromosome domain identification of YACs (chromosome

in situ suppression hybridization; CISS). The necessity for a standard technique such as CISS is strongly emphasized since the chromosome domain identification is the most labour intensive part of physical mapping and requires critical evaluation of the data by different individuals. In addition, different strains of *S. mansoni* are currently maintained in each laboratory involved in the genome mapping group, and little is known about the general genome organization and chromosome structure.

We have established the standard CISS technique using schistosome chromosomes. The CISS method is a simplified and improved extension of FISH, now well applied to human genome mapping (Inazawa et al. 1993) The advantage of the method is a wide application to any kind of DNA templates (especially cosmid DNAs), and this technique is appropriate for the visualization of cosmid links using DNA strands from chromosome materials. The technique has been distributed to colleagues on request, and the details are now in press (Tanaka et al. 1995b).

A noteworthy progress has been made after the establishment of the CISS technique, and the most recent schistosome genome map is shown as Fig. 1. These observations were reconstructed for confirmation in three different Labs (Tsukuba Univ., Tokai Univ., and Kyoto Prefectural Medical Univ.) and discussed for critical evaluation. YAC or equivalent large insert DNA library (BAC or cosmid libraries) is required for efficient access to the construction of a high-resolution physical map. We have succeeded in constructing a cosmid library using the same source of DNA as the YAC library (cercariae DNA). However, phage and plasmid clones could not yield a distinguishable signal onto the schistosome chromosome unless multi-copy genes, located on the same loci of the chromosome, were used. Even phage clones containing an 8 Kb region encoding cytosolic Cu/Zn superoxidase dismutase gene (CT-SOD) did not yield any signals (Mei et al. 1995).

In addition, the CISS technique requires species-specific repetitive sequences as a masking agent. We had two choices in the schistosome genome, one of which is Sau3A-digested schistosome genomic DNA, and the other is the schistosome specific repetitive sequence, Sm α (Spotila et al. 1989). These DNAs were tested to demonstrate their masking ability by hybridization with schistosome chromosomes. Unlike a former publication (Spotila et al. 1989), the Sm α sequence (approximately 1 copy per 50 Kb) did not demonstrate any signals to any of the schistosome chromosomes, only faint backgrounds signals were obtained according to the wash condition. This result, together

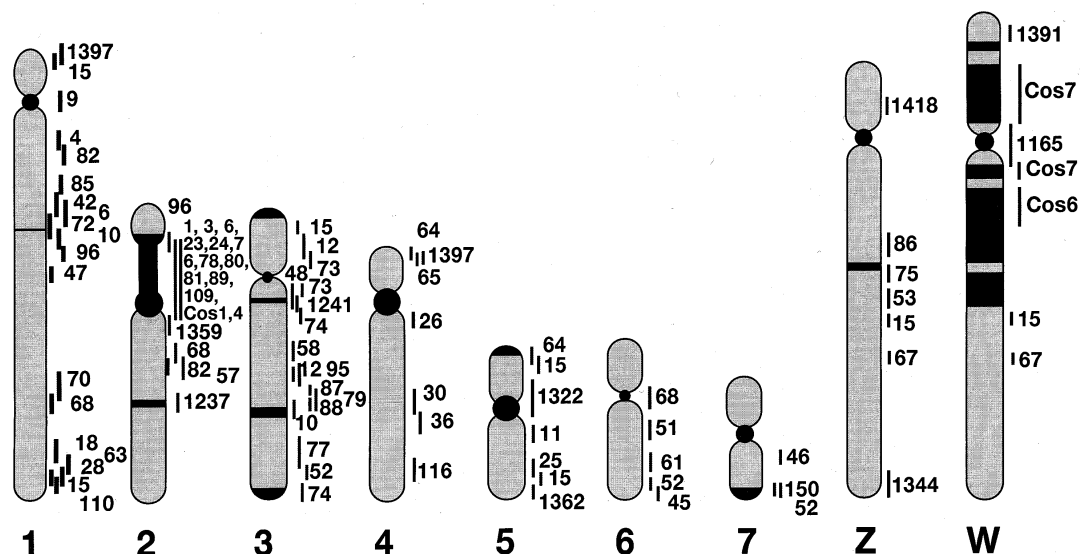


Fig. 1: genome map of *Schistosoma mansoni*. Numbers represent the YAC clones, and Cos number represents the mapping results using cosmid clones.

with the critical evaluation of the data using other schistosome genes, compelled us to re-examine and re-evaluate the previous results. Again, it is emphasized that the critical evaluation of the data by different individuals is necessary to be confident in proceeding to the chromosome domain identification.

3) Homogeneous structure of *S. mansoni* chromosomes - The hybridization of schistosome chromosomes with total genomic DNA provided valuable data and general information about the genome organization. This examination should be commonly performed to know the basic chromosome band patterns, and is useful to know the background repetitive sequence band patterns before hybridization with specific probes.

Repetitive sequences which represent approximately 30% of the schistosome genome did not distribute equally among the chromosomes, instead they tended to accumulate in centromeric and telomeric regions, and confired to the limited number of bands in each chromosome. The patterns of band formation and genome organization judged by this repetitive sequence domains revealed that schistosome strains NMR, Kenya, and Brazil have homogeneous chromosome structures. The homogeneity has been confirmed by pulse-field gradient gel electrophoresis and RAPD using various primer sets, and none of these experiments could demonstrate significant polymorphism and variance among those strains. Interestingly, *S. japonicum* showed significant polymorphisms in band formation between strains. These results pro-

vide valuable information for the comparison of genome organization of the different schistosome species; *S. mansoni* showed gene rearrangements in small fragments (1-2 Kb at largest) during development, with a very homogeneous genome structure among geographic locations, *S. japonicum* has very distinguishable polymorphisms in chromosome structure among strains obtained from different geographic locations, however, without any evidence of gene rearrangement.

4) The CISS technique is useful to distinguish host DNA contaminants from schistosome DNA. In addition, it has been observed that mouse DNA did hybridize to schistosome chromosome by the CISS technique. It is well known that the possibility of contamination with host DNA molecules can not be excluded in the preparation of schistosome adult DNAs. Conserved small tandem repeat DNA molecules of host origin may remain in the schistosome adult and egg DNAs. The CISS technique provides a very good procedure to distinguish accidental host contaminants from schistosome gene markers, simply by detecting positive signals onto schistosome chromosomes.

Thus it appears acceptable to use different strains of *S. mansoni*, at least those commonly maintained in the genome mapping group as a source of equal quality for physical mapping.

Phase II *Retrieval of new transcribed sequences*
Screening of cDNA pools with YACs and cosmids. Our recent effort to establish the positional cloning methods provided us the novel molecular understanding of abundant host-related

molecules and receptors in the schistosome genome. High density grids of each library (generated ESTs, YAC and cosmid library) have been constructed and distributed among all the laboratories of WHO/TDR genome project members. More EST grids are now being processed from Dr Simpson's project, the YAC grids are ready, and the cosmid library has been successfully constructed.

Phase III *Characterization of YAC specific mini-libraries*

1) Construction of a cosmid library. A cosmid library has been constructed with partial digested DNA from cercariae in the cosmid vector pWE15 and SuperCos provided by STRATAGENE (Tanaka et al. 1995b). This library will be screened with the gene markers, and designated to high density grids as with the YAC library.

2) Screening of high density grids (YAC and cosmid libraries) - High density grids were screened with the known gene markers described above. Preliminary data for detailed mapping of the YAC insert will be obtained in the laboratories.

3) Assemble chromosome specific YACs into contigs, link cosmids together with visual linking using the CISS technique. Together with the contigging of the cosmid library, the application of the CISS technique to make visual links of cosmid to chromosome strands is being performed.

4) Screening of YAC and cosmid inserts for new markers using repetitive sequences and known gene markers - The CISS technique using YAC clones provides a valuable tool to demonstrate the repetitive sequences existing in the gene markers. For example, a CT-SOD gene marker was shown to include a chromosome 3 specific repetitive sequence by signal formation of YAC clones which contain the gene. As described above, the repetitive sequences in the schistosome genome did not disperse equally among the chromosomes, and seems to accumulate together in very limited regions. The CT-SOD gene contains one of these repetitive sequences, which provides a suitable point for genome walking of chromosome 3 (Tanaka et al. 1995b).

Thus the progress obtained, includes the development and application of CISS and the development of the YAC libraries which are available for the mapping project in all schistosome research laboratories. Research is also providing basic information concerning chromosome structure and genome organization in species and strains of *Schistosoma*.

In addition, a schistosoma genome project gopher server is being constructed by Dr David A Johnston, of the Natural History Museum. Such

Internet resources will assist in the provision, deposition, and exchange of information.

FUTURE SCHEME

As the genome pools and standard techniques have been established and distributed among WHO/TDR collaborators (Drs Simpson, Pena, Rollinson, Pierce, Brindley, and McManus), rapid progress can be obtained toward constructing a reasonable quality physical map. The scheme in detail is:

Phase II *Retrieval of new transcribed sequences*

1) Screening of cDNA pools with YACs and cosmids - Fig. 2 summarizes the screening methods and the high-resolution maps which will be obtained through the process. YAC insert DNAs are being used for rapid discovery and characterization of the new transcribed genes. YAC and cosmid clones are directly random-labeled with radio isotope, and used to screen the generated EST high density grids. These insert DNAs can be applied to the GeneTrapper system (Gibco-BRL) to obtain the positive clones from pooled cDNA libraries. These procedures enable us to pool and analyze the chromosome specific transcribed genes.

Phase III *Characterization of YAC specific mini-libraries*

1) Construction of cosmid library - Cosmid libraries and cosmid high density grids are currently under construction.

2) Screening of high density grids prepared from YAC and cosmid libraries - YAC and cosmid high density grids will be screened with known genetic markers and the generated EST markers. PCR products from degenerative primers, and characterized gene fragments of interest will also be used to screen YAC/cosmid grids for further mapping and analysis (Fig. 2).

3) Assemble chromosome specific YACs into contigs and link cosmids, together with visual links using the CISS technique. Chromosome specific YACs will be used for the screening of cosmid grids to complete contig assembly and to further macromapping of the YAC inserts (Fig. 3).

4) Screening of YAC and cosmid inserts for new markers using repetitive sequences and known gene markers. Chromosome-specific, roughly macro-mapped YAC and cosmid will be subcloned into plasmids and further screened with ESTs and the known gene markers to provide a detailed physical and genetic map.

Phase IV *Characterization of the newly-identified transcribed segments*

1) DNA sequencing to cover whole regions (Mb size regions) of schistosome genome 50 Kb

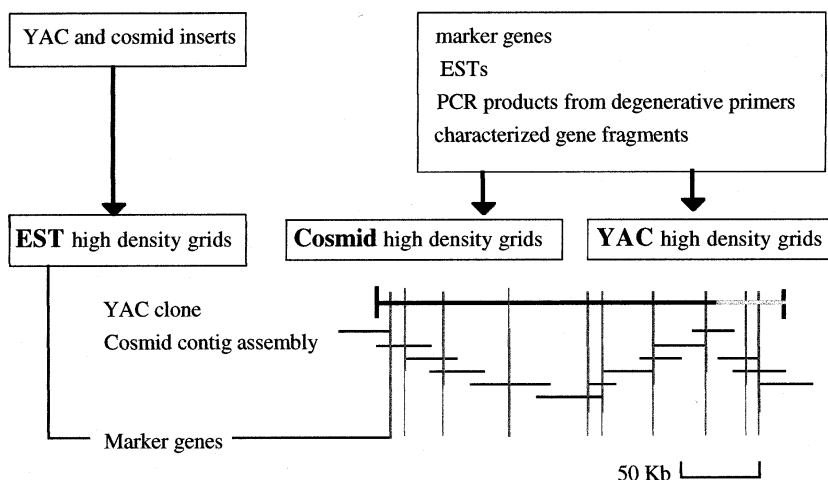


Fig. 2: schematic strategy for a high-resolution map.

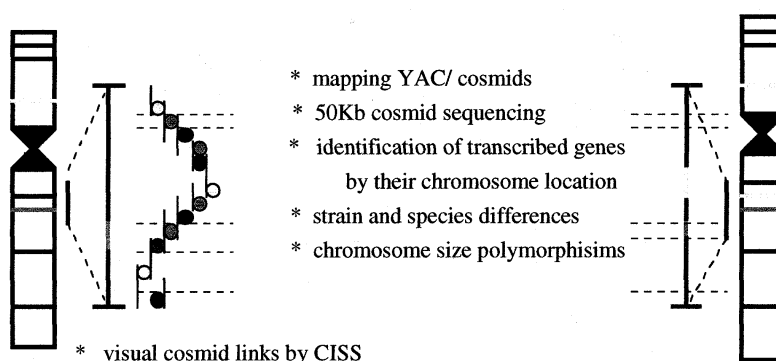


Fig. 3: future scheme of the schistosome genome project.

cosmid inserts will be pursued and sequences will be aligned to cover Mb regions of the chromosome. The sequences will be characterized by homology searches against databases.

2) Comparison of the sequences among stains and species. Species and strain-specific differences in chromosome structure, related to sequence variation will be analyzed. The nucleotide and amino acid sequences, chromosomal localizations, and allelic polymorphisms are being determined using laboratory maintained schistosome species and strains. Chromosomal *in situ* suppression hybridization using the chromosomal preparations from the different stages of the parasite will demonstrate the possible DNA rearrangement of the genes. The chromosomal assignments of the genes would also be helpful in defining its *in vivo* functions in the different stages of the parasites. Schistosome counterpart of human disease-related gene homologues will be characterized with phylogenetic considerations.

3) Assembly into transcription units (RNA analysis). The developmental control of the gene expression of the new genes will be analyzed by RNA blot, and an expression map will be constructed.

4) Development of *in vitro* and *in vivo* assay systems for transcribed genes and gene systems. The information described above will finally be used to investigate which molecular systems are essential for parasitism. Gene transfer systems (incorporation and "knock-outs") should be tested in both parasite and host. The YAC cosmid libraries are therefore multi-functional at this point, and their production constitutes integral and economic components of the project.

5) Sequence, hybridization and mapping data will be described in relevant Internet databases [eg. GenBank, EMBL, schistosome genome project server under construction (see over)].

Thus, the proceedings of schistosome genome project promises not only the generation of a high-

resolution map of the schistosome genome within a few years, but also the beginning of the new era of Frontier Science in Parasitology research toward the real understanding and control of "The host-parasite relationship". Recently many questions and discussions have broken out asking what is the purpose of the gigantic project like human genome project. My answer is: "to find exact gene and gene systems in limited time and limited budget.

The author was the first to publish evidence that the schistosome genome contained host retrovirus-related molecules. This work was based on the hypothesis that, in order to evade immunological attack, schistosome has, during the course of evolution, acquired host genes and that retroviruses provided a likely means of transfer. This data provoked a great deal of controversy and it took six years for independent research to prove that the schistosome genome does contain retroviral sequences, in the long list of 3,047 ESTs.

The advent of technology and extensive international collaboration will enable us to solve many questions within a reasonable amount of time, human resources, and, more importantly, within a limited budget.

ACKNOWLEDGEMENTS

To Mr Nobuhiko Shirato, Mr Kouji Saegusa, Ms Toshie Sawada, and Mr Kozo Kobarashi for technical

assistance; to Drs Takashi Imai, Schizhen Qin, and Masao Yamada for much value advice.

REFERENCES

- Inazawa J, Saito H, Ariyama T, Abe T, Nakamura Y 1993. High-resolution cytogenetic mapping of 342 new cosmid markers including 43 RFLP markers on human chromosome 17 by fluorescence *in situ* hybridization. *Genomics* 17: 153-162.
- Mei H, Hirai H, Tanaka M, Hong Z, Rekosh D, Loverde PT 1995. *Schistosoma mansoni*: Cloning and characterization of a gene encoding cytosolic Cu/Zn superoxidase dismutase. *Exp Parasitol* 80: 250-259.
- Simpson AJG, Sher A, McCachan TF 1982. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol Biochem Parasitol* 6: 125-137.
- Spotila LD, Hirai H, Rekosh D, Loverde PT 1989. A retroposon like short repetitive DNA element in the genome of the human blood fluke, *Schistosoma mansoni*. *Chromosoma (Berl)* 97: 421-428.
- Tanaka M, Hirai H, Nagafuchi S, Loverde PT, Franco GR, Simpson AJG, Pena SD 1995a. Yeast Artificial Chromosome (YAC) based genome mapping of schistosomes. *Mol Biochem Parasitol* 69: 41-51.
- Tanaka M, Tanaka T, Inazawa J, Nagafuchi S, Dias-Neto E, Simpson AJG, Kaukas A, Johnston DA, Rollinson D 1995b. Chromosomal *in situ* suppression hybridization exhibits homogeneous chromosome structure and genome organization of *Schistosoma mansoni* strains. *Mol Biochem Parasitol* (in press).