GENOME ANALYSIS
OF PROTOZOAN PARASITES

PROCEEDINGS OF A WORKSHOP HELD AT ILRAD
NAIROBI, KENYA
11–13 NOVEMBER 1992

Edited by
S.P. Morzaria

THE INTERNATIONAL LABORATORY FOR RESEARCH ON ANIMAL DISEASES
BOX 30709 • NAIROBI • KENYA

The International Laboratory for Research on Animal Diseases (ILRAD) was established in 1973 with a global mandate to develop effective control measures for livestock diseases that seriously limit world food production. ILRAD's research program focuses on animal trypanosomiasis and tick-borne diseases, particularly theileriosis (East Coast fever).

ILRAD is one of 18 centres in a worldwide agricultural research network sponsored by the Consultative Group on International Agricultural Research. In 1992 ILRAD received funding from the United Nations Development Programme, the World Bank and the governments of Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom and the United States of America.

Production Editor: Peter Werehire

This publication was typeset on a microcomputer and the final pages produced on a laser printer at ILRAD, P.O. Box 30709, Nairobi, Kenya, telephone 254 (2) 632311, fax 254 (2) 631499, e-mail BT GOLD TYMNET CGI017. Colour separations for cover were done by PrePress Productions, P.O. Box 41921, Nairobi, Kenya. Printed by English Press Ltd., P.O. Box 30127, Nairobi, Kenya.

ISBN 92-9055-296-4

The correct citation for this book is *Genome Analysis of Protozoan Parasites: Proceedings of a Workshop Held at ILRAD, Nairobi, Kenya, 11–13 November 1992*, ed. S.P. Morzaria. Nairobi: The International Laboratory for Research on Animal Diseases, 1993.

Contents

List of authors

Foreword

Opening address
A.R. Gray

MAJOR MAPPING PROJECTS

Physical strategies for the molecular dissection of genomes
C.L. Smith, R. Oliva, D. Wang, D. Grothues and S. Lawrance

Bovine genome research
A. Teale

Establishing a laboratory for parasite genome analysis: focus on the parasitic protozoa *Toxoplasma, Trypanosoma* and *Leishmania*
J.M. Blackwell, J.W. Ajioka and S.E. Melville

Summary of discussion

GENETIC ANALYSIS OF *THEILERIA* AND *PLASMODIUM*

Genetic recombination in field populations of *Plasmodium falciparum*
D. Walliker, L.C. Ranford-Cartwright and H.A. Babiker

Genome mapping and linkage analysis in *Plasmodium falciparum:* toward identification of the chloroquine resistance determinant
T.E. Wellems

Theileria parva: current status
V. Nene, R. Bishop, T.T. Dolan, D. McKeever, O.K. ole-MoiYoi, S.P. Morzaria, A.J. Musoke, M.K. Shaw, P. Toye and A.S. Young

Genomic polymorphisms in *Theileria parva*
R. Bishop, H. Baylis, B. Allsopp, P. Toye, V. Nene, T. Dolan, P. Spooner, M. Kibe and S. Morzaria

Theileria parva: a restriction map and genetic recombination
S.P. Morzaria, J.R. Young, P.R. Spooner, T.T. Dolan and R.P. Bishop

Summary of discussion

GENETIC ANALYSIS OF TRYPANOSOMES

An introductory review of the important trypanosomes of livestock
P.R. Gardiner

Variability of *Trypanosoma congolense*
P. Majiwa

Genetic recombination and karyotype inheritance
in *Trypanosoma brucei* species
A. Tait, N. Buchanan, G. Hide and M. Turner

The identification of genes of importance in *Trypanosoma brucei brucei* through the use of transfection techniques
N.B. Murphy, A. Muthiani, P. Pandit, S. Kemei, S.K. Moloo, K. Hirumi, H. Hirumi and A.S. Peregrine

Chromosome polymorphism in *Trypanosoma brucei brucei*
and the selection of chromosome-specific markers
S.E. Melville, M.P. Barrett, J.P. Sweetman, J.W. Ajioka and R.W.F. Le Page

Summary of discussion

TECHNIQUES FOR PHYSICAL AND FUNCTIONAL ANALYSIS

Markers for mapping trypanotolerance genes
S.J. Kemp

Derivation of large insert clones from human chromosome 22
using the bacteriophage P1 vector system
N.S. Shepherd, J.N. Coulby and S.L. Ackerman

Yeast artificial chromosomes: cloning and analysis of complex genomes
G.A. Silverman, K.M. Lia and S.S. Schneider

Yeast artificial chromosome cloning and contig
construction in protozoan parasites
J.W. Ajioka, J. Espinoza and J.T. Swindle

Summary of discussion

ROUND TABLE DISCUSSION

RECOMMENDATIONS

LIST OF PARTICIPANTS

List of authors

Workshop Organizing Committee
P. Majiwa
S. Morzaria
N. Murphy

Ackerman, S.L.
Research and Development
The Du Pont Merck Pharmaceutical Company
Wilmington, Delaware 19880-0328, USA

Ajioka, J.W.
Microbiology and Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP, UK

Allsopp, B.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Babiker, H.A.
Division of Biological Sciences
Institute of Cell, Animal and Population Biology
University of Edinburgh
West Mains Road, Edinburgh EH9 3JN, UK

Barrett, M.P.
Microbiology and Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP, UK

Baylis, H.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Bishop, R.P.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Blackwell, J.M.
Microbiology and Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP, UK

Buchanan, N.
Wellcome Unit of Molecular Parasitology
University of Glasgow
Glasgow, G61 1QH, UK

Coulby, J.N.
Research and Development
The Du Pont Merck Pharmaceutical Company
Wilmington, Delaware 19880-0328, USA

Dolan, T.T.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Espinoza, J.
Department of Pathology
University of Cambridge
Tennis Court Road
Cambridge CB2 1QP, UK

Gardiner, P.R.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Grothues, D.
Department of Molecular and Cell Biology
Division of Structural Biology
University of California
Lawrence Berkeley Laboratory
Berkeley, California 94720, USA

Hide, G.
Wellcome Unit of Molecular Parasitology
University of Glasgow
Glasgow, G61 1QH, UK

Hirumi, H.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Hirumi, K.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Kemei, S.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Kemp, S.J.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Kibe, M.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Lawrance, S.
Life Sciences Department
Otterbein College
Westerville, Ohio 43081, USA

Le Page, R.W.F.
Microbiology and Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP, UK

Lia, K.M.
Joint Program in Neonatology
Department of Paediatrics
Harvard Medical School
Children's Hospital
300 Longwood Avenue, Enders 970
Boston, Massachusetts 02115, USA

Majiwa, P.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

McKeever, D.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Melville, S.E.
Microbiology and Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP, UK

ole-MoiYoi, O.K.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Moloo, S.K.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Morzaria, S.P.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Murphy, N.B.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Musoke, A.J.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Muthiani, A.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Nene, V.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Oliva, R.
Department of Molecular and Cell Biology
Division of Structural Biology
University of California
Lawrence Berkeley Laboratory
Berkeley, California 94720, USA

Pandit, P.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Peregrine, A.S.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Ranford-Cartwright, L.C.
Division of Biological Sciences
Institute of Cell, Animal and Population Biology
University of Edinburgh
West Mains Road
Edinburgh EH9 3JN, UK

Schneider, S.S.
Joint Program in Neonatology
Department of Paediatrics
Harvard Medical School
Children's Hospital
300 Longwood Avenue, Enders 970
Boston, Massachusetts 02115, USA

Shaw, M.K.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Shepherd, N.S.
Research and Development
The Du Pont Merck Pharmaceutical Company
Wilmington, Delaware 19880-0328, USA

Silverman, G.A.
Joint Program in Neonatology
Department of Paediatrics
Harvard Medical School
Children's Hospital
300 Longwood Avenue, Enders 970
Boston, Massachusetts 02115, USA

Smith, C.L.
Centre for Advanced Biotechnology
Boston University
36 Cummington Street
Boston, Massachusetss 02215, USA

Spooner, P.R.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Sweetman, J.P.
Microbiology and Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP, UK

Swindle, J.T.
Department of Microbiology and Immunology
University of Tennessee, Memphis
858 Madison Avenue
Memphis, Tennessee 38163, USA

Tait, A.
Wellcome Unit of Molecular Parasitology
University of Glasgow
Glasgow, G61 1QH, UK

Teale, A.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Toye, P.
International Laboratory for Research
on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Turner, M.
Department of Zoology
Laboratory for Biochemical Parasitology
University of Glasgow
Glasgow, G12 8QQ, UK

Walliker, D.
Division of Biological Sciences
Institute of Cell, Animal and Population Biology
University of Edinburgh
West Mains Road
Edinburgh EH9 3JN, UK

Wang, D.
Department of Molecular and Cell Biology
Division of Structural Biology
University of California
Lawrence Berkeley Laboratory
Berkeley, California 94720, USA

Wellems, T.E.
Genetics and Pharmacology Section
Laboratory of Malaria Research
National Institute of Allergy and Infectious Diseases
Bethesda, Maryland 20892, USA

Young, A.S.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Young, J.R.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi, Kenya

Foreword

This book provides the record of a workshop on `Genome Analysis of Protozoan Parasites' held at the International Laboratory for Research on Animal Diseases (ILRAD), Nairobi, Kenya. This workshop was one of a series of meetings that ILRAD holds on a regular basis to provide guidance, focus and planning for various research activities within the Institute's three program areas, namely Tick-Borne Diseases, Trypanosomiasis and Socioeconomic and Epidemiology.

ILRAD's mandate is to provide improved control methods for the economically important livestock diseases in developing countries, particularly Africa. Two main disease syndromes, theileriosis and trypanosomiasis, caused by the parasitic protozoa of the genus *Theileria* and *Trypanosoma*, respectively, are the major causes of morbidity and mortality in cattle and other small ruminants. Thus, they contribute significantly towards restricting agricultural development in Africa and other parts of the tropics. ILRAD initially undertook to improve the control of these diseases through gaining a better understanding of their epidemiology and by developing novel subunit vaccines. It aims to accomplish this primarily by focusing research on the immunological and molecular aspects of theileriosis and trypanosomiasis. Research on parasite genome analysis, which includes a range of activities from gene cloning to physical mapping, forms an integral part of the basic research conducted at ILRAD and this has made significant contribution in the fields of theileriosis and trypanosomiasis. For example, in trypanosomiasis research, the molecular karyotypes of the important *Trypanosoma* species have been generated, DNA probes have been produced for accurate differentiation of species and transfection techniques have been exploited to investigate the genetic basis of drug resistance. In theileriosis research, a complete macro-restriction map of the *T. parva* genome has been constructed, important genes of the parasite have been located on various chromosomes, sexual recombination has been identified as one cause of diversity in parasite populations and a large number of markers has been generated to identify and differentiate various *Theileria* species. These studies have enhanced our knowledge of the different epidemiological patterns of the disease and contributed towards better understanding of parasite genome composition, parasite behaviour and the interaction of parasites with their hosts. Similar advances have been made in other laboratories studying parasitic protozoa such as *Plasmodium, Toxoplasma* and *Leishmania.* Given ILRAD's goal-oriented research, the financial constraints it faces and a change in its research program to project-based research, it was considered timely to hold this meeting with a view to focusing the laboratory's research towards generating outputs that will have practical applications in controlling these important livestock diseases. Thus, ILRAD invited nine scientists from Europe and the USA with expertise in various aspects of eukaryotic genome analysis to participate in this workshop. The workshop aimed to do the following.

- Review the current state of some of the major genome programs in the world.
- Review recent developments in the genome analysis of the important protozoan parasites such as *Leishmania, Toxoplasma, Trypanosoma* and *Theileria.*

- Review major technological advancements contributing towards eukaryotic genome analysis.
- Identify major applications of the knowledge gained from the genome analysis projects.
- Identify ILRAD's needs for *Theileria* and *Trypanosoma* genome analysis.
- Explore possibilities of collaboration with an international group of scientists involved in genome analysis projects.

This book contains all the formal presentations and summaries of discussions in five sessions held over a three-day period and a set of recommendations made by the group. The meeting achieved all its goals. The recommendations made by the group will provide guidelines for setting ILRAD's research priorities in genome analysis.

I would like to thank the members of the organizing committee for this workshop, Drs. P. Majiwa and N. Murphy, for their help in formulating the program; Dr. A. Teale for his support in keeping ILRAD's needs sharply in focus during the meeting; the invited participants for keeping the discussions open and interactive; and ILRAD staff members who contributed as presenters and rapporteurs.

The workshop would not have been possible without the support of the Cooperative Programs, Training and Information Department. My particular thanks go to Dr. Rob Eley and Mr. Kepher Nguli for their help and assistance in organizing the workshop. I am also grateful to Ms. Susan MacMillan for editing the summaries of discussion, Mr. Dave Elsworth for redrawing the illustrations in Prof. C.L. Smith's paper and Mr. Peter Werehire for proofreading the text and typsetting the book.

Subhash Morzaria
International Laboratory
for Research on Animal Diseases
Nairobi, July 1993

Opening address

A.R. Gray
Director General, ILRAD, Nairobi, Kenya

Good morning ladies and gentlemen. I warmly welcome you all to the International Laboratory for Research on Animal Diseases (ILRAD) as we convene this workshop on Genome Analysis of Protozoan Parasites. I note that most of the invited speakers are visiting the institute for the first time. I assure you that we will do our best to make your stay pleasant and comfortable. I hope that while you are here you will take the opportunity to meet as many ILRAD scientists as possible, visit our research facilities and learn a bit about the various research activities conducted here.

I would like to say a few words about the institute and its goals for the benefit of those visiting us for the first time. ILRAD was founded in 1973 and is one of several institutes supported by the Consultative Group on International Agricultural Research. The institute is charged with the responsibility of providing improved methods for the control of livestock diseases adversely affecting food production in developing countries. For the last 15 years ILRAD has concentrated its research on various forms of animal African trypanosomiasis and East Coast fever of cattle caused by the tick-transmitted protozoan parasite *Theileria parva*. These diseases, together with other tick-borne diseases, such as anaplasmosis, babesiosis and cowdriosis, still constitute a major threat to livestock development in Africa.

ILRAD recently introduced a project-based research management structure. One of the ways the Laboratory ensures the relevance of projects within the research programs is by holding small scientific workshops to discuss specific areas of research. These workshops enable ILRAD scientists to consult with scientists internationally recognized for their expertise in specialized areas. Discussions resulting from such interactions help us to formulate long- and short-term objectives that keep the institute's research focused on important problems related to improving control of animal diseases.

ILRAD is involved in various aspects of genome analysis of *Trypanosoma* and *Theileria*. In the future, we might also be interested in analyzing the genomes of other tick-borne protozoa of veterinary importance. ILRAD needs to understand the global perspective on genome analysis in higher and lower eukaryotes and become up-to-date in the latest developments in the field of parasite genome analysis. Because some confusion exists about what constitutes genome analysis, there is a need to define this broad field more precisely, to break down the various aspects of genome analysis and to pick out the specific areas in which ILRAD must become involved in order to achieve its goals. We have made a major commitment in the bovine genome program to identify the genetic factors that underly the trypanotolerance trait which might eventually allow us to exploit the knowledge to produce cattle resistant to the severe effects of trypanosome infection. ILRAD is also producing genetic crosses from defined breeds of cattle as a resource for several

collaborating international laboratories, and is also generating DNA-based markers for the trypanotolerance trait. A question we hope discussions at this workshop will help to answer is whether ILRAD should use similar approaches in its analyses of parasite and bovine genomes, given the differences in complexities of the parasites and host. I hope you will enlighten us on your research goals in work in related areas since we need to identify research groups with common interests to initiate collaborative research in the present climate of dwindling resources.

I note from the program that there is ample opportunity and scope to discuss the specific objectives outlined for the workshop. I urge the session chairpersons and rapporteurs to lead structured discussions and to produce concise summaries of all the discussions. I expect that the workshop, with help from the invited speakers, will produce a set of recommendations that may form the basis for program development and further research in this area. I am pleased to open this workshop. Before I hand over to the chairman of the first session, I thank the visitors for finding time from their busy schedules to attend this meeting.

MAJOR MAPPING PROJECTS

Physical strategies for the molecular dissection of genomes

C.L. Smith, R. Oliva, D. Wang, D. Grothues and S. Lawrance

Molecular studies were, in the past, confined to organisms with well developed genetic systems. Recently a number of physical methods have been developed that allow the dissection of essentially all genomes. These methods focus on characterizing the genome using a top down approach. A top down approach involves 1) the identification of number and size of chromosomes, 2) the creation of low resolution genomic restriction maps, 3) the creation of high resolution restriction maps and overlapping genomic libraries and eventually 4) the sequencing of the entire genome. This approach allows the organization of genes to be analyzed as they are placed on the physical maps. The physical characterization of new genomes should aid the development of other necessary techniques for genome manipulation *in vivo.*

INTRODUCTION

Traditional studies in molecular biology, anchored in genetics, have focused on a small number of organisms (e.g. *Escherichia coli, Mus musculus)* because of the synergy generated by their well developed genetics, extensive phenotypic characterization, and the availability of effective analytical tools. The development of techniques for physical analysis of DNA, including recombinant DNA techniques, extended the utility of many of the techniques developed in model systems to other organisms lacking genetic systems. In particular, the development of pulsed-field gel electrophoresis (PFGE) (Schwartz *et al.*, 1983; Schwartz and Cantor, 1984) and ancillary physical techniques for manipulating large DNA molecules (e.g. Smith and Cantor, 1987; Smith *et al.*, 1986, 1987a, 1988, 1992) irrevocably changed the focus of genome studies from model systems to a wide variety of organisms. For instance, before the development of these techniques, for many organisms, there was no simple and reliable method for determining the size and number of chromosomes nor any fractionation technique that allowed their purification. In one of the first applications of these techniques, physical characterization of the *Trypanosome* chromosomes led to an understanding of how the variable surface glycoprotein was controlled (van der Ploeg *et al.*, 1984). The power of PFGE combined with DNA sequencing, the polymerase chain reaction (PCR) (Saiki *et al.*, 1985) and restriction fragment length polymorphism analysis (Botstein *et al.*, 1980) prompted the realization that physical studies would even allow the dissection of complex genomes, including that of *Homo sapiens.*

The first step in a physical approach to genome analysis requires the construction of genomic restriction maps which provide an overview of the genome. Subsequent construction of libraries of overlapping clones allows immediate access to any part of genome and sets the stage for eventual complete DNA sequencing. Here, we focus on the first steps in the physical analysis of genomes.

The first complete, low-resolution genomic restriction map of *Escherichia coli* (Smith *et al.,* 1987a) was soon followed by complete genomic restriction maps for several other small genomes ranging in size from 1 to 15 Mb and a number of large regions of complex genomes (e.g. Lawrance *et al.*, 1987; Kohara *et al.*, 1987). Although the first overlapping genomic library for *E. coli* was completed at about the same time as the first genomic map, only a small number of complete libraries has subsequently been reported. This reflects the relative ease of constructing genomic restriction maps as compared to the difficulty in constructing overlapping genomic libraries. Recently, sets of overlapping clones spanning the euchromatic region of the Y chromosome (Foote *et al.*, 1992) and the short arm of chromosome 21 (Chumakov *et al.*, 1992) have been reported.

Strategies for generating genomic restriction maps can be divided into top down, mostly directed approaches, and bottom up, mostly random approaches (Figure 1). Top down approaches are hard to automate but lead to complete maps. Although the first end product of a top down approach is a complete map on paper, no cloned DNA is generated. However, such maps can be of great value in constructing libraries of chromosomes or specific chromosomal regions. Alternatively, bottom up strategies use clone fingerprints to detect overlaps between clones and to reconstruct the original genomic structure. Bottom up, random strategies are easy to automate, but difficult, if not impossible, to use by themselves to complete a map.

TOP DOWN APPROACHES

Top down approaches divide the genome into segments that are analyzed independently. For instance, the natural division of a genome into a number of chromosomes can provide the first step in a top down approach. The conventional construction of chromosomal, low-resolution, restriction maps (Figure 1A) involves specific fragmentation of chromosomes (or regions of chromosomes) into large restriction fragments which are fractionated by PFGE, and ordered by hybridization experiments using single copy probes that have been regionally assigned and/or partially ordered using one of a number of techniques.

DNA SOURCE

The first decision in any mapping experiment is the source of DNA to be used. Many experimental organisms have a readily identifiable progenitor or a genetically or molecularly well-characterized isolate that is widely used or known. Sometimes choices are made on the basis of convenience or some unusual aspect of a particular species. For instance, a particular cell line or isolate may be easy to grow and readily available.

Cell lines from higher eukaryotes may differ from *in situ* tissue cells as a result of somatic or epigenetic changes, or as a result of immortalization. Frequently, cell lines in culture have a variable ploidy. DNA methylation is reported to drift in tissue culture, especially with lymphoblastoid lines (Shmookler and Goldstein, 1982; Silva and White, 1988). Thus, tissue samples (e.g. blood and sperm) are a valuable source of DNA for map construction. Sperm samples are particularly useful because they represent germ line

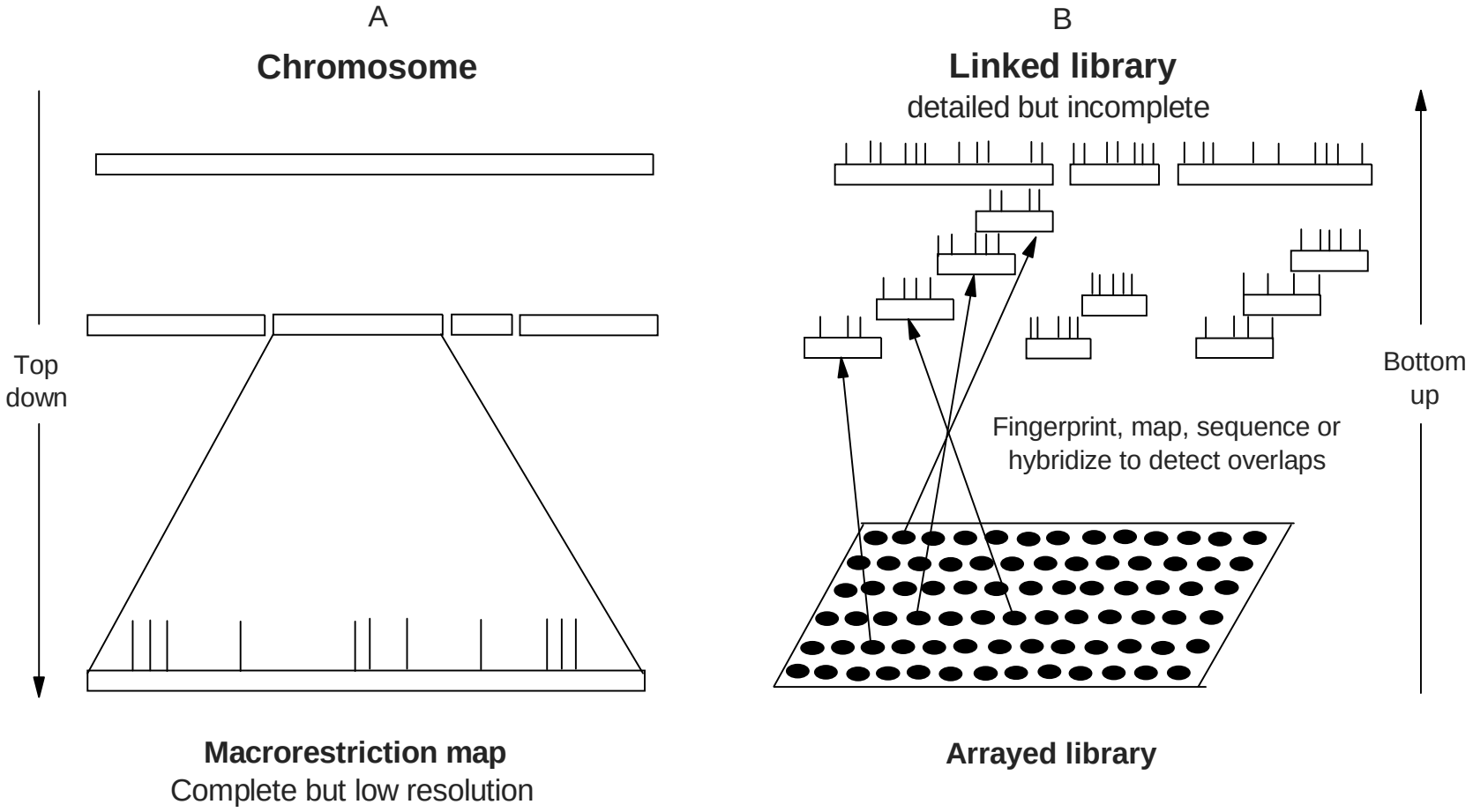

Figure 1. Two basic genomic mapping strategies. Top down approaches (A) produce complete low resolution restriction maps. Bottom up approaches (B) produce high resolution restriction maps and overlapping libraries.

chromosomes whose essential features must be conserved. All other samples have the potential to contain developmental changes involving chromosomal DNA rearrangements.

Interpretation of hybridization experiments using DNA from diploid cells may be difficult. Detection of a single band in a complete restriction enzyme digest of DNA from a diploid cell with a single copy genomic probe is simple to interpret. Sometimes single copy probes will detect different bands in different cell lines or multiple bands in the same cell line. Multiple bands may be indicative of methylation at a nearby restriction enzyme site, or may represent true megabase polymorphisms, e.g. a restriction enzyme site polymorphism or a megabase fragment size variation. If two bands of equal intensity are detected, then it is likely the two homologous chromosomes are different. In some cases, such differences can seriously complicate the interpretation of experimental results. These problems may be avoided in several ways. For instance, human-rodent hybrid cell lines containing one or a few human chromosomes can be used as sources of haploid DNA as it is unlikely that such cells would have two different copies of any one human chromosome. Another way around this problem is to use human cell lines that are homozygous or hemizygous for a particular chromosome as was done to analyze the HLA region on chromosome 6 (Lawrance *et al.*, 1986, 1987).

PURIFICATION AND MANIPULATION OF DNA

Purification of genomic DNA in agarose is necessary to prevent shear damage. The procedures are simple and have been applied to a large number of organisms. A recent summary of the protocols and their applications can be found in Smith *et al.* (1993) and Burmeister and Ulanovsky (1992).

Purified DNA must be cleaved with an appropriate restriction enzyme. In a top down mapping approach, it is most convenient to start mapping with enzymes that cut DNA infrequently, so that few fragments need to be ordered. Generally, GC content, or more correctly dinucleotide content of a genome, are good but not totally reliable predictors of fragment sizes produced with a particular enzyme. This is because genomic DNA sequences are not distributed randomly in nature. For instance, the most infrequently occurring restriction sites in *E. coli* are those that contain the stop codon UAG (Smith and Condemine, 1990). In higher eukaryotes the dinucleotide sequence, CpG, occurs 80% less than it would in a random sequence of the same GC content and usually the cytosine is modified to 5-methyl cytosine (5-MC). The most common spontaneous damage to DNA, deamination, converts 5-MC to thymine and cytosine to uracil. DNA repair systems remove uracil from DNA (thus restoring the original GC base pair) but restore the original base pair at T-G mismatches only part of the time (Brown and Jiricny, 1987). This leads to genomes rich in TpG dinucleotide, but deficient in CpG sequences. The CpG dinucleotide is clustered in hypomethylated CpG-rich islands present at the 5' region of many genes (Bird, 1986, 1987; Antequera and Bird, 1988). Nearly all restriction enzymes which generate large fragments from the human genome have at least one CpG dinucleotide in their recognition sequence. Cleavage by many of these restriction enzymes is inhibited by cytosine methylation. This means that these enzymes will tend to cut chromosomal DNA at the 5' region of the genes; thus genes will be frequently located at the ends of large

restriction fragments. Consequently the low-resolution restriction maps will often assist in locating genes.

Potentially useful enzymes are summarized in Table 1. Surprisingly, the restriction enzymes which appear to produce the largest (~1 Mb) DNA fragments from the human genome are *Not*I, *Mlu*I and *SgrA*I (Table 1). When none of the obvious choices produce a

Table 1. Restriction enzymes that produce megabase fragments.

Recognition	Enzyme site (5⌈-3⌈)
Enzymes with >6 bp recognition site	
*Pac*I	TTAAT/TAA
*Pme*I	GTTT/AAAC
*Swa*I	ATTT/AAAT
*Sse838887*I	CCTGCA/GG
Enzymes >6 bp recognition and cut in CpG islands	
*Rsr*II (Csp*)	CG/GWCCG
*SgrA*I	CR/CCGGYG
*Not*I	GC/GGCCGC
*Srf*I	GCCC/GGGC
*Fse*I	GGCCGGCC
*Sfi*I	GGCCNNNN/NGGCC
*Asc*I	GG/CGCGCC
Enzymes that cut in CpG islands	
*Mlu*I	A/CGCGT
*Sal*I	G/TCGAC
*Nru*I	TCG/CGA
*BssH*II	G/CGCGC
*Sac*II	CCGC/GG
*Eag*I (*Ecl*XI*, *Xma*III)	C/GGCCG
*Nar*I	GG/CGCC
*Sma*I	CCC/GGG
*Xho*I	C/TCGAG
*Pvu*I	CGAT/CG
*Apa*I	GGGCC/C
Enzymes with TAG in their recognition sequence	
*Avr*II (*Bln* I*)	C/CTAGG
*Nhe*I	G/CTAGC
*Xba*I	T/CTAGA
*Spe*I	A/CTAGT
*Nhe*I	G/CTAGC
Other useful enzymes	
*Dra*I	TTT/AAA
*Ssp*I	AAT/ATT

*Indicates preferred isoschizomer

manageable number of fragments, testing a battery of `ordinary' enzymes can reveal the usefulness of a particular enzyme.

PULSED-FIELD GEL ELECTROPHORESIS

Once genomic DNA purified in agarose has been cleaved with a restriction enzyme the results need to be analyzed by PFGE. Pulsed-field gel electrophoresis separates DNA molecules on the basis of size (for recent reviews see Cantor *et al.*, 1988; Burmeister and Ulanovsky, 1992). The largest known molecule fractionated with this technique is the largest *Schizosaccharomyces pombe* chromosome, which is 5.7 Mb in size (Fan *et al.*, 1988, 1991). Although DNA molecules greater than 5.7 Mb in size can be fractionated by PFGE (Orbach *et al.*, 1988; Cole and Williams, 1988; Cox *et al.*, 1990), it is most convenient to analyze DNA molecules that can be properly compared to size standards and hence are less than 5.7 Mb in size.

In conventional electrophoresis using agarose or acrylamide, DNA molecules smaller than 20 kb are fractionated by size because they are sieved as they move. Molecules larger than 20 kb cannot be fractionated because they move through agarose at the same constant speed. Pulsed-field gel electrophoresis separates DNA by size because the ability of these molecules to change direction in the gel is size-dependent; larger molecules take longer to change direction than smaller molecules. DNA molecules are forced to change direction by exposure to alternating electrical fields. The time each field is on is called the pulse time. The pulse time is the major variable that must be adjusted in a PFGE experiment. A pulse time must be chosen such that the size class of interest spends most of the pulse time changing direction, and very little of the pulse time actually moving through the gel. In this way, the majority of the pulse time is spent in a size-dependent movement rather than a size-independent movement. Hence, PFGE fractionation is based on reorientation and retardation, which explains the requirement for long run times. Different PFGE conditions will open different fractionation windows. Factors other than pulse time that influence the speed at which DNA molecules change direction and move can significantly affect the fractionation window. These include agarose concentration, temperature, buffer composition, field strength and the angle through which molecules are forced to change direction. These variables should be carefully controlled in order to ensure reproducibility.

There have been a large number of PFGE apparatus that have been developed. The most popular instruments appear to be the Chef Mapper (Biorad) and the Pulsphor (Pharmacia-LKB). All of the PFGE instruments work on the same principle. The basic difference in these instruments is the angle between the fields and whether the fields are homogeneous or not. Theoretically the same experiment done on different instruments should lead to the same results. This may not be obvious. Experiments carried out using different PFGE instruments may have hidden variations in other parameters (outlined above) that influence DNA movement.

DNA fractionation by PFGE is usually carried out using 1% agarose gels, 15 °C and high field strengths of 10 V/cm. Pulse times of 25 or 100 sec fractionate molecules up to 0.5 or 1 Mb, respectively. The best separations of molecules above 1 Mb in size occur at lower field strengths (e.g. 3 V/cm) and a small number of directional changes. For

instance, the *S. pombe* chromosomes are slowly fractionated in six days using 3600 sec pulse times at 5 V/cm and 8 °C. This means that these DNA molecules have only changed direction about 150 times, compared to approximately 500 directional changes used to fractionate molecules less than 1 Mb in size. This slow and limited movement is needed to prevent shear damage to these very large DNA molecules during electrophoresis. In fact, at 10 V/cm, molecules even smaller than 1 Mb experience some, but acceptable, shear damage.

Annealed λ DNA provides a convenient and accurate ladder of multimers which range between the monomer (48.5 kb) and 30 mer (1.4 Mb) (Mathew *et al.*, 1988). For studies of DNA <1 Mb in size, it is convenient to load *Saccharomyces cerevisiae* chromosomal DNAs next to an annealed λ size standard to ensure correct counting of the phage ladder. The size of molecules >1 Mb in size may be determined by comparison with the three *S. pombe* chromosomal DNAs (5.7, 4.7 and 3.5 Mb, respectively) as well as the ease in preparing chromosomal DNAs of a number of fungi (T. Zhang, S. Ringquist, J. Fan, C. Cantor and C. Smith, unpublished observations).

The final step in mapping experiments involves the characterization of the PFGE fractionated restriction fragments by hybridization experiments using various probes. It is enormously useful to know the number and size of restriction fragments that will make up the map. For small genomes this may be determined by simple ethidium bromide. For human chromosomes, this information has been obtained by identifying human DNA restriction fragments from hybrid cell lines using human specific interspersed repetitive sequences, e.g. *Alu* (Singer, 1982) as hybridization probes (Sainz *et al.*, 1992).

The use of ordered genetic markers as hybridization probes is a powerful aid in constructing genomic restriction maps. These probes serve as anchor points between genetic and physical maps. A physical map between two closely linked genetic markers is constructed by hybridization analysis of complete and partial restriction enzyme digests. Hybridization of probes to complete digests will reveal the size of the fragment containing the probe but will not reveal the position of a particular probe on a fragment, nor the order of different probes within the same fragment, nor the neighbouring fragment. Analysis of partial digests using closely linked probes allows the physical order and distance to be established. For example, in a complete digest of a haploid cell line, two probes (A and B) may detect different fragments of 1.5 and 0.7 Mb, respectively (Figure 2). The detection of a band of common size (2.2 Mb) equal to the combined size of the two fragments in the partial digest suggests that the probes are located on adjacent fragments. Usually putatively linked probes are successively tested on the same blot. Other confirmatory evidence must be used to prove linkage. In some cases partial digest analysis allows the determination of several flanking restriction fragments. Usually, interpretation of partial mapping data requires combining information derived from two or more probes located on nearby fragments. It is particularly useful to employ linking clone fragments (Smith *et al.*, 1986, 1987b). Each half of a linking clone recognizes a different subset of partial digest products (Figure 2A). The true map must be consistent with both sets of data.

A complete linking clone library would allow the identification of all adjacent fragments. Clones that identify adjacent restriction sites will identify a common fragment and establish linkage between the two restriction sites. This means that only a small number

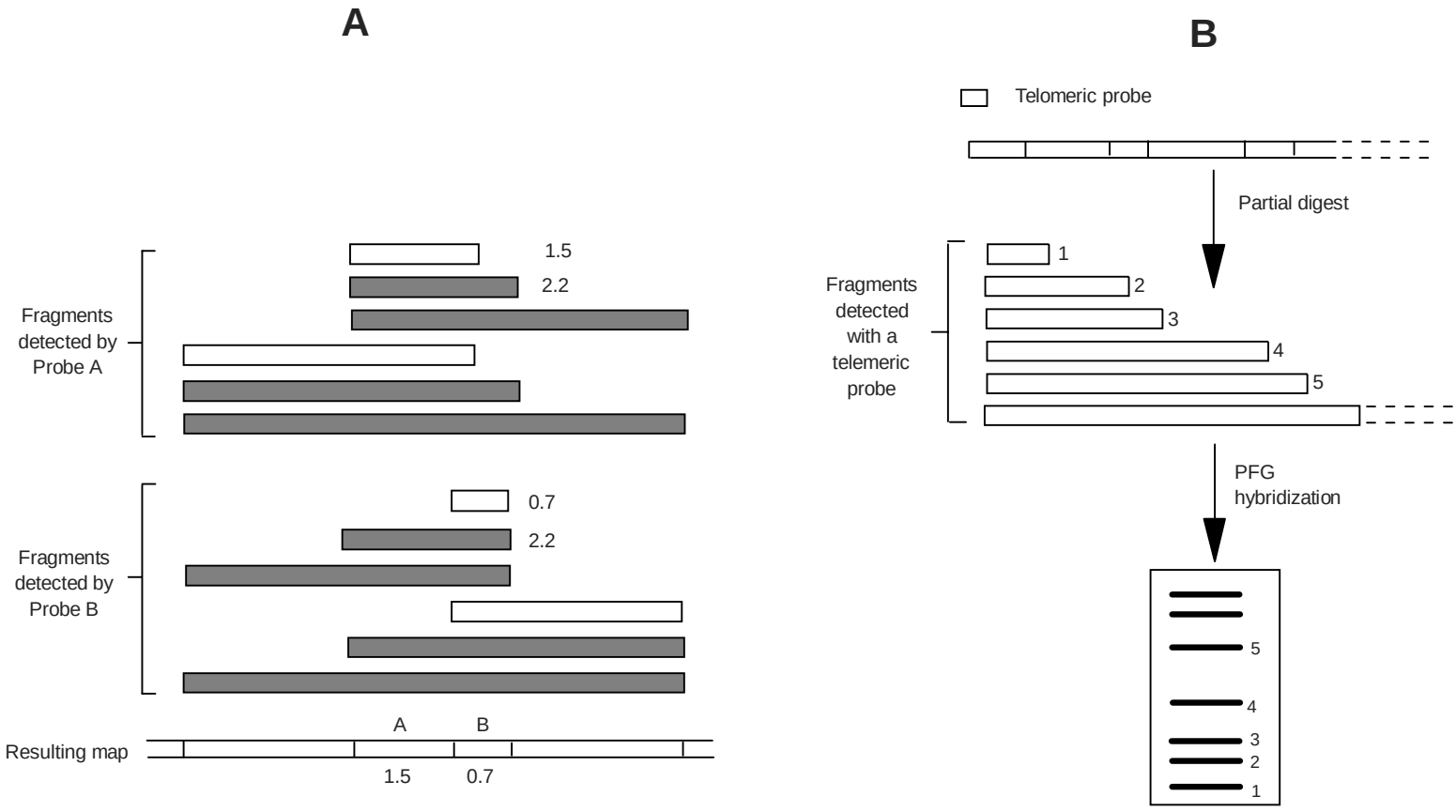

Figure 2. Use of partial restriction enzyme digestion to construct physical maps (A). Single probes (A or B) used in hybridization experiments to partially digested DNA detect bidirectional partial digestion products. Probes on adjacent fragments detect common (shaded) and unique (unshaded) bands. (B) Indirect-end labelling experiment with telomeric sequences reveals the restriction sites close to the end of the chromosome. Partial digestion experiments are easy to interpret with telomeric probes because fragments extend in only one direction.

of linking clones are needed to map an entire chromosome (e.g. ~50 for chromosome 21) or even the entire human genome (~3000 clones) using the restriction enzyme *Not*I. In practice, the proof of a common fragment involves fingerprinting the neighbourhood of the fragment or the fragment itself using partial digest.

Polymorphism link-up can also establish linkage in much the same way as a partial digest (Smith *et al.*, 1992, 1993). Polymorphism link-up takes advantage of existing polymorphic variation between different DNAs (Figure 3). For instance, hybridization of two different probes to complete digests of different DNAs reveals the same megabase polymorphism pattern of polymorphisms although different size fragments are detected. Furthermore, the restriction site between the probes may not be cleaved in different cell lines, leading to the detection of the same size fragment by both probes. In other cell lines, the probes detect different fragments because cleavage has occurred at the intervening site. Thus, the sometimes complex and undesired presence of multiple bands in gels can be turned into a methodological advantage.

Bacterial chromosomes are circular, thus map closure should produce a circular map. Linear eukaryotic chromosomes have special telomeric structures at the ends of chromosomes designed to preserve their structure. The ends of a linear chromosomal restriction map may be identified using telomeric probes (for review see Cheng and Smith, 1990). These clones are especially powerful using partial digestion strategies because the partial digestion products extend from only a single direction and the results are easy to interpret (Figure 2B).

Starting a map is a relatively easy task. The difficulty lies in finishing it. End game strategies require focused efforts. Each gap represents unique problems and solutions. Thus, no global method guarantees closure. Rather, the most efficient solution for a particular gap will depend on what is known about the surrounding region and what samples are available for analysis. For instance, it has been particularly useful to generate probes directly from unassigned megabase restriction fragment to finish the 50 Mb restriction map of the long arm of chromosome 21 (D. Wang and C. Smith, unpublished observations).

BOTTOM UP APPROACHES

The construction of overlapping libraries for chromosomes or chromosomal regions requires a representative clone library (Figure 1B). In this approach, random clones from an arrayed library are fingerprinted to detect overlaps between clones, or are used as hybridization probes to fish overlapping clones from a library. The overlapping clones are then formed into contigs with the eventual goal of linking up all the contigs from a region or a chromosome. Coulson *et al.* (1986, 1988) have fingerprinted cosmid clones by identifying fragment patterns generated using two restriction enzymes with 4 bp recognition sequences. While this approach assigns fragments to clones, it does not order restriction sites. Hence no restriction map is created. A protocol using this approach has been adapted to a commercial DNA sequencing device, and it is being used to make an overlapping library of human chromosome 19 (Carrano *et al.*, 1989).

Kohara *et al.* (1987) fingerprinted clones in an overlapping *E. coli* λ library using partial digestion experiments with eight restriction enzymes having 6 bp recognition sites. Here

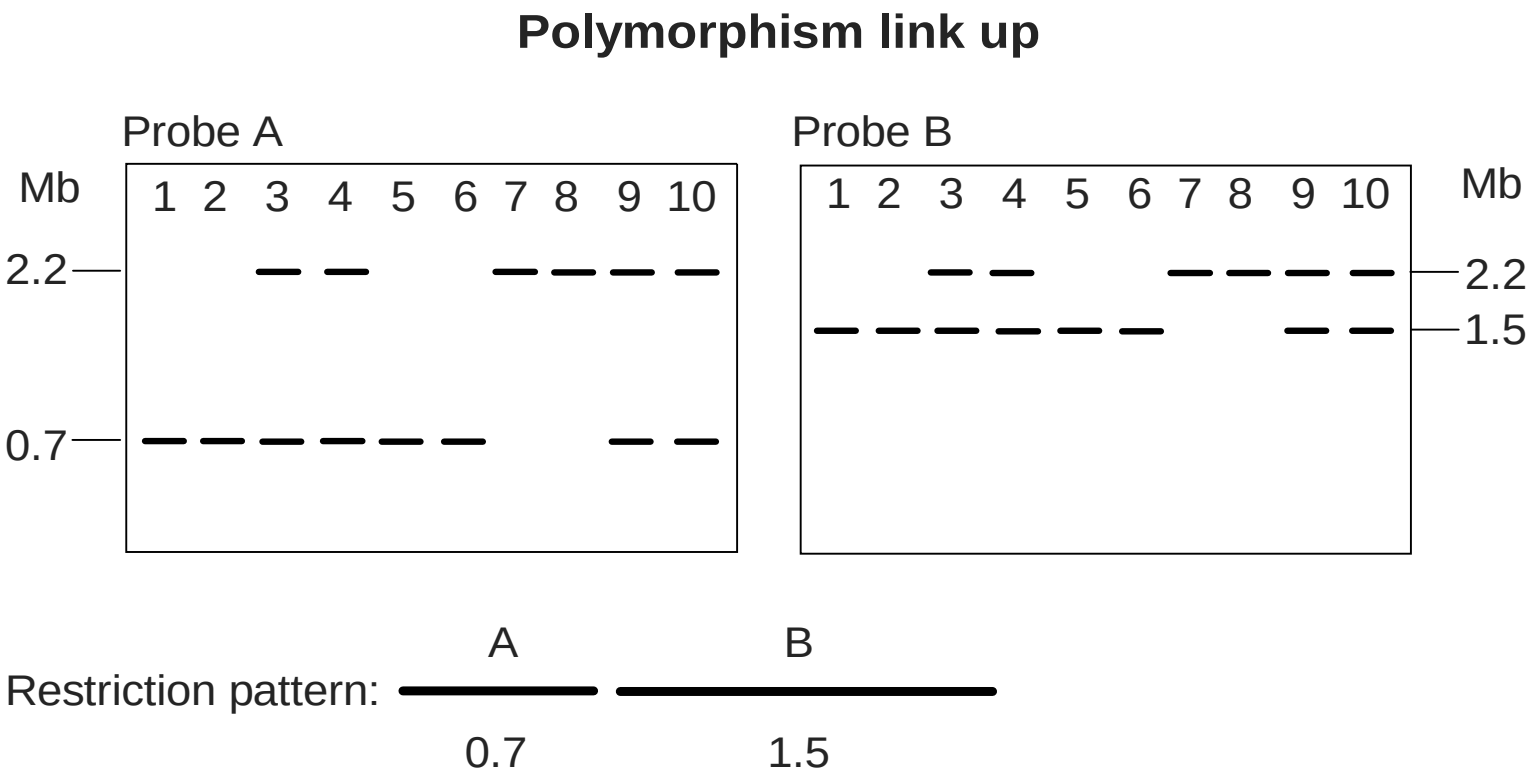

Figure 3. Physical map construction using polymorphism link-up. Hybridization of DNA from different cell lines (lanes 1–10) with two putatively linked probes (A and B) detects identical polymorphism patterns, but common and different restriction fragments.

indirect end-labelling hybridization experiments using probes specific for the ends of λ allowed the creation of a high resolution restriction map for the *E. coli* chromosome. Others have proposed, and are using, different fingerprinting approaches which include direct sequencing of clones (Church and Kieffer-Higgins, 1988) or portions of clones around rare cutting sites (Oliva *et al.*, 1991), hybridization to oligonucleotides (Rackwitz *et al.*, 1984; Michiels *et al.*, 1987) and hybridization with pooled groups of clones (Evans and Lewis, 1989), or genomic restriction fragments (D. Grothues, C. Cantor and C. Smith, unpublished observations). The latter method combining advantages of both top down and bottom up approaches is being tested now on an *S. pombe* cosmid library.

The common problem in all the random approaches is that, as the map nears completion, new information is gained more and more slowly since most clones have been seen before (Olson *et al.*, 1986; Lander and Waterman, 1988). Gaps may contain DNA sequences that are unclonable or unstable. Thus far, completion of overlapping library construction requires an approach directed at gap filling (Coulson *et al.*, 1988; Kohara *et al.*, 1987).

A direct way to expand the number of overlapping clones in a contig is to use classical `chromosome walking' approaches. Here, the end clones of a contig are used to screen a library to find adjacent clones. These new overlapping clones are then themselves used to identify other overlapping clones (Figure 2). Classical chromosome walking requires the screening of many clones in sequential `walking' (or `crawling') steps in order to walk a relatively small distance since the amount of DNA contained in the clones identified in a single walk greatly exceeds the net distance covered. This method can be speeded up by using pools of clones for the screening procedure and by identifying overlapping clones for the ends of all contigs at the same time.

In principle, the methods described above can be applied to any type of clone (plasmid, phage, cosmid or yeast artificial chromosome (YAC)). In practice it is more efficient to use the largest possible clone to minimize the total number of clones that have to be linked up. Although cosmid clones (containing about 40 kb of DNA) have been the preferred vectors for chromosome walking, recent progress with YAC (Burke *et al.*, 1987) and P1 phage (Sternberg, 1990) large insert cloning systems should ease the construction of overlapping libraries.

SUMMARY

The physical map of a chromosome is the structure of its DNA. Hence the most detailed physical map would contain the complete primary DNA sequence of that chromosome. Until that goal is reached, the various maps generated along the way can be used to identify and study functional domains on chromosomes. For instance, combinations of various maps and strategies have allowed the identification of an increasing number of human disease genes.

ACKNOWLEDGEMENT

This work was supported by contract number DE-AC03-76SF00098, USA.

REFERENCES

ANTEQUERA, F. and BIRD, A.P. 1988. Unmethylated CpG islands associated with genes in higher plant DNA. *EMBO Journal* 8: 2295–2299.

BIRD, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* 321: 209–213.

BIRD, A.P. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics* 3: 342–347.

BOTSTEIN, D., WHITE, R.L., SKOLNICK, M. and DAVIS, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314.

BROWN, T.C. and JIRICNY, J. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methyl cytosine. *Cell* 50: 945–950.

BURKE, D.T., CARLE, G.F. and OLSON, M.V. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.

BURMEISTER, M. and ULANOVSKY, L., eds. 1992. *Pulsed-Field Gel Electrophoresis: Protocols, Methods and Theories.* Totowa, New Jersey: The Humana Press, Inc., 481 pp.

CANTOR, C.R., SMITH, C.L. and MATHEW, M.K. 1988. Pulsed-field gel electrophoresis of very large DNA molecules. *Annual Review of Biophysics and Biophysical Chemistry,* Vol. 17. Palo Alto, California: A. Perlis Annual Reviews, Inc., pp. 287–304.

CARRANO, A.V., LAMERDIN, J., ASHWORTH, L.K., WATKINS, B., BRANSCOMB, E., SLEZAK, T., RAFF, M., de JONG, P.J., KEITH, D., McBRIDE, L., MEISTER, S. and KRONICK, M. 1989. A high-resolution, fluorescence-based, semiautomated method for DNA fingerprinting. *Genomics* 4: 129–136.

CHENG, J.F. and SMITH, C.L. 1990. YAC cloning of telomeres. *Genetic Analysis: Techniques and Applications* 7: 119–125.

CHUMAKOV, I., RIGAULT, P., GUILLOU, S. and 33 others. 1992. Continuum of overlapping clones spanning the entire human chromosome 21q [see comments]. *Nature* 359 (6394): 380–387.

CHURCH, G. and KIEFFER-HIGGINS, S. 1988. Multiplex DNA sequencing. *Science* 249: 185–188.

COLE, R.A. and WILLIAMS, K.L. 1988. Insertion of transformation vector DNA into different chromosomal sites of *Dictyostelium discoideum* as determined by pulse field electrophoresis. *Nucleic Acids Research* 16 (11): 4891–4902.

COULSON, A., SULSTON, J., BRENNER, S. and KARN, J. 1986. Towards a physical map of the genome of the nematode *Caenorahabditis elegans. Proceedings of the National Academy of Sciences of the USA* 83: 7821–7825.

COULSON, A., WATERSON, R., KIFF, J. and KOHARA, Y. 1988. Genome linking with yeast artificial chromosomes. *Nature* 335: 184–186.

COX, E.C., VOCKE, C.D., WALTER, S., GREEG, K.Y. and BAIN, E.S. 1990. Electrophoretic karyotype for *Dictyostelium discoideum. Proceedings of the National Academy of Sciences of the USA* 87: 8247–8251.

EVANS, G.A. and LEWIS, K.A. 1989. Physical mapping of complex genomes by cosmid multiplex analysis. *Proceedings of the National Academy of Sciences of the USA* 86: 5030–5034.

FAN, J.B., CHIKASHIGE, Y., SMITH, C.L., NIWA, O., YANAGIDA, M. and CANTOR, C.R. 1988. Construction of a *Not*I restriction map of the fission yeast *Schizosaccharomyces pombe* genome. *Nucleic Acids Research* 17: 2801–2817.

FAN, J.B., GROTHUES, D. and SMITH, C.L. 1991. Alignment of *Sfi*I sites with the *S. pombe Not*I restriction map. *Nucleic Acids Research* 19: 6289–6294.

FOOTE, S., VOLLRATH, D., HILTON, A. and PAGE, D.C. 1992. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* 258 (5079): 60–66.

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

LANDER, E.S. and WATERMAN, M.S. 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 2: 231–239.

LAWRANCE, S.K., SMITH, C.L., WEISSMAN, S.M. and CANTOR, C.R. 1987. Megabase-scale mapping of the HLA gene complex by PFG electrophoresis. *Science* 235: 1387–1390.

LAWRANCE, S.K., SRIVASTAVA, R., CHORNEY, M.J., RIGAS, B., VASAVADA, H., GILLESPIE, G.A., SMITH, C.L., CANTOR, C.R., COLLINS, F.S. and WEISSMAN, S.M. 1986. Molecular approaches to characterization of megabase regions of DNA: studies of the human histocompatibility complex. *Cold Spring Harbor Symposium on Quantitative Biology* 51: 123–130.

MATHEW, M.K., SMITH, C.L. and CANTOR, C.R. 1988. High-resolution separation and accurate size determination in pulsed-field gel electrophoresis of DNA. I. DNA size standards and the effect of agarose and temperature. *Biochemistry* 27: 9204–9210.

MICHIELS, F., CRAIG, A.G., ZEHETNER, G., SMITH, G.P. and LEHRACH, H. 1987. Molecular approaches to genome analysis: a strategy for the construction of ordered overlapping clone libraries. *CABIOS* 3: 203–210.

OLIVA, R., LAWRANCE, S.K., WUE, Y. and SMITH, C.L. 1991. Chromosomes: molecular studies. In: Dulbecco, R., ed. *Encyclopedia of Human Biology.* San Diego: Academic Press, pp. 475–488.

OLSON, M.V., DUTCHIK, J.E., GRAHAM, M.Y., BRODEUR, G.M., HELMS, C., FRANK, M., MacCOLLIN, M., SCHEINMAN, R. and FRANK, T. 1986. Random clone strategy for genomic restriction mapping in yeast. *Proceedings of the National Academy of Sciences of the USA* 83: 7826–7830.

ORBACH, M.J., VOLLRATH, D., DAVIS, R.W. and YANOFSKY, C. 1988. An electrophoretic karyotype of *Neurospora crassa. Molecular and Cellular Biology* 8: 1469–1473.

RACKWITZ, H.R., ZEHETNER, G., FRISCHAUF, A.M. and LEHRACH, H. 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. *Gene* 30: 195–200.

SAIKI, R.K., SCHARF, S., FALOONA, F., MULLIS, K.B., HORN, G.T., ERLICH, H.A. and ARNHEIM, N. 1985. Enzymatic amplification of the b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230: 1350–1354.

SAINZ, J., PEVNY, L., WU, Y., CANTOR, C.R. and SMITH, C.L. 1992. Distribution of interspersed repeats *(Alu* and Kpn) on *Not*I restriction fragments of human chromosome 21. *Proceedings of the National Academy of Sciences of the USA* 89: 4539–4543.

SCHWARTZ, D.C. and CANTOR, C.R. 1984. Separation of yeast chromosome-sized DNAs by pulsed-field gradient gel electrophoresis. *Cell* 37: 67–75.

SCHWARTZ, D.C., SAFFRAN, W., WELSH, J., HAAS, R., GOLDENBERG, M. and CANTOR, C.R. 1983. New techniques for purifying large DNA's and studying their properties and packaging. *Cold Spring Harbor Symposium on Quantitative Biology* 47: 189–195.

SHMOOKLER REIS, R.J. and GOLDSTEIN, S. 1982. Variability of DNA methylation patterns during serial passage of human diploid fibroblasts. *Proceedings of the National Academy of Sciences of the USA* 79: 3949–3953.

SILVA, A.J and WHITE, R. 1988. Inheritance of allelic blueprints for methylation patterns. *Cell* 54: 145–152.

SINGER, M.F. 1982. Highly repeated sequences in mammalian genomes. *International Review of Cytology* 76: 67–112.

SMITH, C.L. and CANTOR, C.R. 1987. Purification, specific fragmentation and separation of large DNA molecules. In: Wu, R., ed. *Methods in Enzymology, `Recombinant DNA'*, Vol. 155. New York: Academic Press, pp. 449–467.

SMITH, C.L. and CONDEMINE, G. 1990. New approaches for physical mapping in small genomes. *Journal of Bacteriology* 172: 1167–1172.

SMITH, C.L., ECONOME, J., SCHUTT, A., KLCO, S. and CANTOR, C.R. 1987a. A physical map of the *E. coli* K12 genome. *Science* 236: 1448–1453.

SMITH, C.L., KLCO, S., ZHANG, T., FANG, H., OLIVA, R., WANG, D., BREMER, M. and LAWRENCE, S. 1992. Analysis of megabase DNA using pulsed-field gel electrophoresis techniques. In: Adolph, K.E., ed. *Methods in Molecular Genetics.* San Diego: Academic Press, pp. 155–193.

SMITH, C.L., KLCO, S.R. and CANTOR, C.R. 1988. Pulsed-field gel electrophoresis and the technology of large DNA molecules. In: Davies, K., ed. *Genome Analysis: A Practical Approach.* Oxford: IRL Press, pp. 41–72.

SMITH, C.L., LAWRANCE, S.K., GILLESPIE, G.A., CANTOR, C.R., WEISSMAN, S.M. and COLLINS, F.S. 1987b. Strategies for mapping and cloning macro-regions of mammalian genomes. In: Gottesman, M., ed. *Methods in Enzymology, `Molecular Genetics of Mammalian Cells'*, Vol. 151. New York: Academic Press, pp. 461–489.

SMITH, C.L., WANG, D. and GROTHUES, D. 1993. Construction of physical maps of chromosomes. In: *The Encyclopedia of Molecular Biology and Biotechnology.* New York: VCH Publishers, in press.

SMITH, C.L., WARBURTON, P., GAAL, A. and CANTOR, C.R. 1986. Analysis of genome organization and rearrangements by pulsed-field gradient gel electrophoresis. In: Setlow, J.K. and Hollaender, A., eds. *Genetic Engineering,* Vol. 8. New York: Plenum Press, pp. 45–70.

STERNBERG, N. 1990. Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobases pairs. *Proceedings of the National Academy of Sciences of the USA* 87: 1103–1105.

Van der PLOEG, L.H.T., SCHWARTZ, D.C., CANTOR, C.R. and BORST, P. 1984. Antigenic variation in *Trypanosoma brucei* analyzed by electrophoretic separation of chromosome-sized DNA molecules. *Cell* 37: 77–84.

Bovine genome research

A. Teale

Progress is being made towards effective coordination of bovine genome research internationally. Such research is primarily stimulated by the prospect of rapidly identifying potentially valuable genes controlling aspects of food productivity and quality. The approach being adopted is linkage analysis in populations segregating traits of interest. It is considered that the role of maps and mapping in bovine genome research is equivocal for four reasons. First, it has been argued that maps have a role in ensuring even genome coverage as markers are developed and in avoiding marker oversaturation of specific areas. Oversaturation is an unwarranted concern. Second, it is likely that individual research groups will develop only modest numbers of markers and will wish to apply all in-house markers in resource populations. Third, the costs of marker application are decreasing. Consequently, resource requirements for application of large numbers of markers, rather than a subset, will not be the limiting factor in the search for interesting genes by linkage analysis. Fourth, the limited need for maps in bovine genetics research will to a large extent be met by the advanced maps of the genomes of other mammalian species which are currently under development. Bovine molecular genetics research is more likely to develop around resource populations, which will be limiting, rather than around maps, and ILRAD is a typical example of this process. The Laboratory is seeking markers linked to genes controlling trypanotolerance—the ability of some West African *Bos taurus* cattle breeds to resist the effects of trypanosome infection. To this end a population of N'Dama × Boran cattle segregating trypanotolerance is under construction and marker development is in progress.

A coordinated `bovine genome project' at international level is still developing, with a small number of laboratories taking a lead and many others considering how, and to what extent, they can contribute. The project is not driven by the prospect of a complete sequence of the bovine genome. Rather, the principal driving force is the prospect of being able to move directly and rapidly to small parts which could have useful practical applications.

Cattle species are arguably the most important livestock species on a worldwide basis. Genome research is therefore stimulated by the possibility that it will lead to identification of genes controlling, either directly or indirectly, aspects of productivity in a valuable resource. Indeed, it can be contended that genetic approaches to improvements in livestock productivity have greater potential to impact on immediate human health and welfare needs than research on the human genome.

A second, although less powerful, stimulus to bovine genome research is a basic scientific interest in comparing the genomes of a diverse range of organisms, each being representative of significant parts of the animal and plant kingdoms. In this context, cattle are representative of the ruminants, and are the most studied.

As far as a bovine genome map is concerned, at the present time several hundred genes have been assigned to chromosomes and syntenic groups in cattle, and the first reports of the relative positions, in genetic terms, of genes within syntenic groups are beginning to appear. The map is nonetheless a rudimentary one, and for reasons to be discussed, is quite likely to remain so for some time. Currently, by far the greatest effort on the part of the

bovine molecular genetics community is being placed in development of a set of highly polymorphic markers capable of giving `reasonable coverage' of the genome. The application of such a marker set will be in linkage studies in populations segregating important traits. The definition of `reasonable coverage' is that coverage providing an acceptable chance of any one or more markers in the set being linked to a gene of interest in the types of resource populations which can be foreseen. As in other species, the emphasis as far as markers are concerned is on microsatellites. The number of markers of this type available worldwide is difficult to estimate, but certainly numbers are in the hundreds and are increasing rapidly.

As the primary objective of most bovine genome research groups is identification of important genes in segregating populations, the role of mapping is somewhat equivocal. It is clear that if markers are mapped as they are developed, this will help to avoid oversaturation of certain areas of the genome and undersaturation of others. In these circumstances, inherently involving a high degree of collaboration and extensive information exchange between research groups, a set of markers giving the required `reasonable coverage' would be developed in an efficient manner, without unnecessary effort and with minimal resources. However, it can be argued that the prospect of any possible oversaturation should not generate concern as every marker has the potential to `have its day'—there cannot be too many markers. More importantly, perhaps, there are few signs at present that there will be effective collaboration in development of a reasonable coverage set of mapped markers. It is more likely that as each genome group reaches the stage of applying markers to the DNA of a segregating population, the group will apply all of the markers available to it, which in most cases will not be an inordinately large number. There will hopefully be a tendency for groups to supply DNA from their resource populations to other groups to do the same. This tendency will be a consequence of the fact that the limiting factor in the search for important livestock genes will be good populations in which to search for them. A third factor which reduces the need for limiting the number of markers to be developed, and therefore also reduces the need to map them as they are produced, is that it is becoming increasingly easy with advancing technologies to apply markers to the DNA of study populations, and the costs on a per marker basis are decreasing. The costs of developing useful populations are, on the contrary, escalating, and the phenotyping processes are essentially unchanging. If maps will not be particularly useful in marker development, will they be of use in moving from marker to gene of interest? The answer in this case is that maps will be of some help, although with rapid advances in physical analysis technologies, the value of high resolution maps in identifying a gene of interest linked to any given marker will possibly decrease over time. Moreover, for this purpose, the maps of the genomes of other mammalian species, which are already at an advanced stage, will provide much of the information needed by bovine genome researchers. However, these considerations aside, for reasons already outlined, it is entirely possible that bovine molecular genetics research will develop around traits of interest rather than around maps, with the emphasis on identification of markers *per se* being a primary and important objective because of their immediate value for marker-assisted selection. This has been the case at ILRAD where research is being undertaken on the ability of some West African *Bos taurus* breeds of cattle to resist the effects of trypanosome infections and to remain productive in the face of disease challenge. This resistance trait

is commonly referred to as `trypanotolerance' (Trail *et al.*, 1989). In the zebu *Bos indicus* breeds, trypanosomiasis causes marked losses in productivity and is often fatal. The objectives of the research are the identification of genetic markers linked to trypanotolerance genes and ultimately isolation of the genes themselves. Markers once identified will have application in livestock breeding programs aimed at increasing the numbers and productivity of trypanotolerant cattle. Isolation of the genes will provide the opportunity to develop transgenic animals particularly fitted to demanding tropical environments in which trypanosomiasis is a factor. With the genes may also come an understanding of how the trait operates at the molecular level.

The Laboratory is therefore developing a resource population of cattle segregating trypanotolerance based on a cross between N'Dama (tolerant) and Boran (susceptible) founder animals (Teale, 1990, 1992). An F_2 generation will be challenged with a pathogenic *Trypanosoma congolense* clone and immunological, physiological and pathological indicators of trypanotolerance will be scored.

With respect to marker development, the Laboratory is engaged in production of microsatellites for use in the trypanotolerance program and to add to the global pool (Brezinsky *et al.*, 1992). Effort is also being put into the identification of polymorphisms in `candidate' genes—that is, genes which, because of their known functions, could be associated with important effects (Kemp *et al.*, 1992) . A third type of marker which has potential application in large segregating populations is provided by random priming with short oligonucleotides—so-called random amplified polymorphic DNA (RAPD) markers (Welsh and McClelland, 1990; Williams *et al.*, 1990). The potential of this marker class for studies of trypanotolerance is also being explored at the present time.

Finally, concern has been expressed that the identification of quantitative trait loci (QTL) by linkage analysis may prove to be very difficult. However, the reported success in locating chromosome regions containing QTL controlling apparently complex traits in plants is encouraging (Paterson *et al.*, 1991). Certainly, in view of the potential benefits of improved food productivity achieved through sustainable approaches such as genetic improvement, there is every reason at present to proceed with all speed.

REFERENCES

BREZINSKY, L., KEMP, S.J. and TEALE, A.J. 1992. ILSTS001—a polymorphic bovine microsatellite. *Animal Genetics* 23: 81.

KEMP, S.J., MAILLARD, J.C. and TEALE, A.J. 1992. A polymorphism in the bovine gamma-S-crystallin gene revealed by allele-specific amplification. *Animal Genetics* 24: 125–127.

PATERSON, A.H., DAMON, S., HEWITT, J.D., ZAMIR, D., RABINOWITCH, H.D., LINCOLN, S.E., LANDER, E.S. and TANKSLEY, S.D. 1991. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. *Genetics* 127: 181–197.

TEALE, A.J. 1990. Genes controlling disease resistance as targets for bovine genome research, with particular reference to trypanotolerance. In: Womack, J., ed. *Mapping the Genomes of Agriculturally Important Animals.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 65–72.

TEALE, A.J. 1992. Mapping and trypanotolerance. In: Teale, A.J., ed. *Bovine Genome Mapping and Trypanotolerance: Proceedings of a Workshop Held at ILRAD, Nairobi, Kenya, 9–11 April 1991.* Nairobi: International Laboratory for Research on Animal Diseases, pp. 5–7.

TRAIL, J.C.M., D'IETEREN, G. and TEALE, A.J. 1989. Trypanotolerance and the value of conserving livestock genetic resources. *Genome* 31: 805–812.

WELSH, J. and McCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 7213–7218.

WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSKI, J.A. and TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531–6535.

Establishing a laboratory for parasite genome analysis: focus on the parasitic protozoa *Toxoplasma*, *Trypanosoma* and *Leishmania*

J.M. Blackwell, J.W. Ajioka and S.E. Melville

Over the past 12 months we have been working towards the establishment of a new laboratory for parasite genome analysis at the University of Cambridge. This laboratory will build on, and provide new tools for, research already being undertaken in our respective laboratories, but taking on board the philosophy of other major (e.g. human, *Drosophila*, *Caenorhabditis elegans)* genome projects which provide a major resource for the international research community. The major aims of the initiative are:

- To establish a laboratory for parasite genome analysis in Cambridge concentrating initially on production of yeast artificial chromosome (YAC) and dual purpose shuttle vector/mapping pacmid P1 and cosmid libraries for the protozoan parasites *Toxoplasma*, *Trypanosoma* and *Leishmania*. High density filters will be produced for mapping purposes and made available to the international research community. A central mapping database will be established in Cambridge.
- To develop a series of polymerase chain reaction (PCR)-based sequence tagged site (STS), polymorphic STS (pSTS) and expression tagged site (ETS) markers to provide genetic and physical mapping tools which can be used to analyze genetic/phenotypic/karyotypic variation in natural populations of these parasites.
- To use, and make generally available, the mapping and dual purpose shuttle vector/mapping libraries as a major resource in addressing specific biological problems by reverse genetics.
- To exploit these parasite genomes, where the chromosomes are resolvable by pulsed-field gel electrophoresis and the genome sizes are two orders of magnitude smaller than mammalian genomes, as models to address broader questions relating to the evolution and organization of eukaryote genomes.

This paper describes the rationale behind the venture, some of the tools developed in conjunction with other genome projects which we plan to exploit, and the three specific chromosome mapping projects we are undertaking related to the biology of the three different parasitic protozoa we are studying.

RATIONALE BEHIND THE CHOICE OF PARASITES

Toxoplasmosis, African trypanosomiasis and leishmaniasis are important protozoal diseases for which no effective or sustained control strategies have been developed. *Toxoplasma gondii* is a prevalent opportunistic pathogen which infects 25–90% of adults worldwide. Along with its more common association with congenital birth defects (Desmonts and Couvreur, 1974; Luft and Remington, 1988) it has emerged as a serious cause of mortality in immuno-compromised patients such as those suffering from AIDS (Mills, 1986). It is also an important agricultural problem causing high abortion rates in the intermediate (sheep) host (Dubey, 1977). The sexual cycle occurs in enterocytes in the gut of cats (Cornelissen and Overdulve,

1985). Ability to experimentally manipulate infection through the cat has permitted classical genetic crosses to be performed and a genetic map to be developed, mainly through the heroic efforts of Boothroyd and Sibley (Sibley, *et al.,* 1991; Sibley, *et al.,* 1992; Sibley and Boothroyd, 1992) working with crosses generated by Pfefferkorn (Pfefferkorn, *et al.,* 1977; Pfefferkorn and Pfefferkorn, 1980; Pfefferkorn and Kasper, 1983).

Trypanosomiasis remains a serious problem across approximately 10 million square kilometres of Africa, causing long-term debilitation and death in humans and restricting the use of cattle and small ruminants in agriculture. There are several species and subspecies of African trypanosomes: *Trypanosoma brucei gambiense* and *T. brucei rhodesiense* cause chronic and acute sleeping sickness in man while *T. brucei brucei, T. congolense* and *T. vivax* cause trypanosomiasis in livestock. In addition to its importance as a major pathogen affecting the lives of many people, the trypanosome has long been a favoured organism of study for biochemists and molecular biologists (reviewed in Clayton, 1988). Its study has led to many novel biological discoveries, such as kinetoplast networks, antigen gene switching, RNA editing and guide RNAs, tandemly repeated genes and polycistronic transcription. The existence of a sexual stage in the life cycle of the African trypanosome had been the subject of speculation for many years before an international collaboration, including Dr Richard Le Page's laboratory in Cambridge, demonstrated unequivocally that genetic recombination between genetically distinct trypanosome isolates could take place in the tsetse fly host (Jenni *et al.,* 1986). The precise mechanism by which recombination/meiosis occurs remains unclear (Turner *et al.,* 1990), but the feasibility of performing laboratory crosses clearly increases the scope of genetic analysis possible in African trypanosomes compared to, for example, the related parasite *Leishmania.*

Leishmaniasis is a protozoal disease which affects 12 million people in Old and New World tropical countries, with an estimated 300 million people at risk of infection (Ashford *et al.,* 1992). The disease takes many different forms including fatal visceral infection of liver, spleen and bone marrow, as well as disfiguring mucocutaneous and diffuse cutaneous disease. Although field and laboratory studies suggest that genetic exchange may occur in *Leishmania* (Pages *et al.,* 1989; Kelly *et al.,* 1991), the existence of a sexual cycle has not been formally demonstrated. Hence, physical mapping is the only way forward in analyzing *Leishmania* genomes. We believe that much can be gained by establishing a comparative approach to genetic and physical mapping in these three groups of parasitic organisms.

CHOICE OF `BIG DNA' LIBRARIES

In developing a new program of genome research, the overall rationale and intention is to produce a series of `big DNA' libraries for genome mapping, utilizing YAC (Burke *et al.,* 1987; Ajioka *et al.,* 1991), P1 (Sternberg, 1990), cosmid (Ish-Horowicz and Burke, 1981) and YAC telomere (Blackburn 1984; Blackburn and Szostak 1984; Chang *et al.,* 1989) cloning systems according to the size of genome, size of insert and scale of contig mapping required, which will provide a resource not only for our own laboratory but for the international parasitology research community. Initially, three libraries are being constructed for each organism (see Table 1) in two different cloning vectors: a pacmid P1 genomic library, insert sizes ranging from 75–110 kb, to map chromosomes ≤2 Mb (Ajioka *et al.,*

Table 1. The number of YAC, P1 and YAC/telomere clones required for an ordered library in different species.

Species	Genome size	Chromosome number	Chromosome range	YAC	P1	YAC/ telomere
Homo sapiens	3×10^9 bp	23	70 to 300 Mb	60000	150000	230
Leishmania	5×10^7 bp	22 to 28	200 kb to 2 Mb	1000	2500	260
T. b. brucei	4×10^7 bp	12*	1 to >5 Mb	800	2000	120*
T. gondii	8×10^7 bp	11	2 to 5 Mb	1600	4000	110

The calculations for the number of YAC (250 kb average), P1 (100 kb average) and YAC/telomere clones needed for an ordered library are based on five genome equivalents assuming random (Poisson) cloning. In the case of the YAC/telomere, a `genome equivalent' is the number of chromosome ends. The estimations used in this table for *Leishmania* (Scholler *et al.*, 1986; Samaras and Spithill 1987; Galindo and Ramirez, 1989; Pages *et al.*, 1989; Bishop 1990; Bastien *et al.*, 1990), *T. b. brucei* (*for the megabase chromosome class only; see S.E. Melville *et al.*, this volume), and *Toxoplasma gondii* (Sibley and Boothroyd, 1992) may vary between related species or from strain to strain, but serve as useful figures for comparative purposes.

1991); a YAC genomic library with average insert size 250–400 kb to map chromosomes 2–5 Mb (Ashford *et al.*, 1992); and a YAC telomere library to ensure coverage of chromosome ends which will not clone into conventional cloning vectors (Belle *et al.*, 1992). YACs will clearly allow construction of contigs for each chromosome very rapidly, P1s and cosmids facilitate physical mapping of small chromosomes and allow finer mapping within regions of larger chromosomes. New P1 and cosmid vector constructs (collaboration with Dr Steve Beverley, Harvard University) are also being engineered to mediate transfection of cosmids and P1s back into *Leishmania* and trypanosomes. These are based on transfection systems already developed for both groups of parasites (Coburn *et al.*, 1991; Cruz and Beverley, 1990) and will facilitate a reverse genetic/complementation approach to solving biological questions. The new shuttle vectors will allow production of dual purpose libraries which can be used for mapping and transfection genetics.

A similar strategy will be adopted for *Toxoplasma* when transfection constructs become available. In addition to the obvious academic advantages of a comparative approach to studying the three groups of protozoan parasites, *Toxoplasma,* African trypanosomes and *Leishmania* spp., establishing an overall program base at one site will maximize efficient use of the specialized laboratory (robotics, automated sequencing) and computing equipment needed to support genome research (Hultman *et al.*, 1989; Bentley *et al.*, 1992; McCombie *et al.*, 1992). The parasite genome analysis program will also benefit from the experience already gained in *Escherichia coli/Drosophila/C. elegans* and human genome projects (Coulson *et al.*, 1986; Kohara *et al.*, 1987; Larin *et al.*, 1991; Ajioka *et al.*, 1991; Chumakov *et al.*, 1992; Sulston *et al.*, 1992; Waterston *et al.*, 1992). Since parasite genomes are much smaller ($3–8 \times 10^7$ bp compared to 3×10^9 bp in man), the task is not so `monumental'. Moreover, since the chromosomes in parasite genomes are generally resolvable by pulsed-field gel electrophoresis (PFGE) (e.g. Samaras and Spithill, 1987; Galindo and Ramirez, 1989; Holmes-Giannini *et al.*, 1990; Sibley and Boothroyd, 1992), the ability

to make chromosome-specific subsets of clones will greatly facilitate contig construction and physical mapping (see Table 1). The new laboratory in Cambridge will be concerned from the outset with development of the resource. High density filters for screening libraries and the clones identified will be made freely available. Databases will be created to collect and collate the information feeding back into the mapping project, with the overall objective being to create databases compatible with, and available through, the genome databases (e.g. Sulston *et al.*, 1992) currently on-line from the CRC/Human Genome Mapping Program Resource Centre.

DEVELOPMENT OF MAPPING TOOLS FOR PARASITE GENOME ANALYSIS

Although the state-of-the-art genome mapping tools are PCR-based STS (Olson *et al.*, 1989; Green and Olson 1990, 1991; Palazzolo *et al.*, 1991), ETS (Venter *et al.*, 1992; Waterston *et al.*, 1992) and pSTS (Nickerson *et al.*, 1992) markers, a given organism's genome structure will determine the utility, and mode of identification and development of these types of markers. Perhaps this is best illustrated by comparing *Toxoplasma gondii* with *T. b. brucei* and *Leishmania* spp. *Toxoplasma gondii* has a well defined sexual cycle and meiosis (Cornelissen and Overdulve 1985; Pfefferkorn *et al.*, 1977), where the major life form is haploid (Cornelissen *et al.*, 1984). In contrast, *T. b. brucei* has a somewhat more obscure sexual cycle and at least part of its genome appears to be diploid (Turner *et al.*, 1990). *Leishmania* spp. also appear to be diploid (Iovannisci and Beverley, 1989). For these reasons, identifying markers which will be useful both genetically and physically, i.e. pSTSs, will be much simpler in *Toxoplasma gondii*, which appears to have a genome structure and mode of gene expression similar to a `standard' eukaryote (e.g. Burg *et al.*, 1988). Coding sequences are separated by *cis* introns and there is little base composition or codon bias (J. Boothroyd, personal communication). Moreover, since many loci in African trypanosomes and *Leishmania* are tandemly arrayed, ETSs will not be as useful as pSTSs. On the other hand, repetitive sequences are probably less abundant and so mapping conundrums in certain regions may be less frequent. Although it may be more work, it is possible to develop pSTSs and ETSs for *T. b. brucei* and *Leishmania* spp., particularly if each strain carries different alleles but the alleles are homozygous at most loci within a strain. However, the rationale and procedures involved with developing these tools is most easily illustrated using *Toxoplasma gondii* as the focus of discussion.

The current genetic map for *Toxoplasma gondii* consists of 64 restriction fragment length polymorphic (RFLP) markers mapped by hybridization to chromosomes separated on TAFE PFGE (Sibley and Boothroyd, 1992) and by linkage analysis of F_1 clones from a genetic cross (Sibley *et al.*, 1992). Eleven linkage groups with a total of 147 map units (centimorgans; cM) have been defined. The relationship between physical distance and recombination frequency is in the order of 500 kb/cM (Sibley *et al.*, 1992). The current map thus provides (on average) one polymorphic marker every 2 cM or 1 Mb. This means that, even if an interesting phenotype (e.g. drug resistance) is genetically mapped, it would be necessary to search through hundreds of kilobases of DNA to find the locus. The alternative is to continue mapping by analyzing more F_1 progeny. This is only useful if there are sufficient markers distributed across the chromosome. If there were a polymorphic marker every 0.2 cM, and it were feasible to analyze

several hundred F_1 progeny, then the physical region where the candidate locus could reside would be limited to less than 100 kb. Clearly, this approach would be greatly facilitated by development of a larger series of PCR-based pSTSs. However, the ability to identify pSTSs is highly dependent upon the level of nucleotide heterozygosity between strains in a cross, i.e., the frequency of nucleotide polymorphism. Although not every pSTSs would be useful in every cross, a simple calculation from the genetic data estimates that there is one nucleotide difference per 240 bp between the P and C strains (Sibley *et al.*, 1992). If nucleotide polymorphisms are randomly (Poisson) distributed, then better than 60% of a random sample of 240 bp segments would contain a difference between the two parental strains, 95% if the size were increased to 750 bp. These calculations suggest that it is not only reasonable to screen simple STSs and ETSs for polymorphisms, but feasible to generate pSTSs from random fragments of DNA since there are several methods available (e.g. Belle White *et al.*, 1992) which can detect single base changes over this size range. Moreover, there are DNA motifs seen in other eukaryotic organisms, such as VNTRs (Jeffries *et al.*, 1986, Nakamura *et al.*, 1987) and dinucleotide repeats (Weber and May, 1989) which are highly polymorphic. If similar motifs exist in *Toxoplasma gondii*, screening for these motifs improves the odds of generating a pSTS.

Simple STS markers can be derived from two sources. The existing RFLP markers which are in the form of cosmids (Sibley *et al.*, 1992) can be converted to STSs by subcloning into a plasmid vector (collaboration with Dr. J. Boothroyd, Stanford University, USA, and Dr. L.D. Sibley, Washington University, St. Louis, USA). Clones from the subclone library can be selected for small, PCR-favourable insert sizes (300 bp to 1 kb), checked for repetitive sequences and DNA sequenced across the cloning junctions. The sequence generated should provide the information to design oligonucleotide PCR primers which will have compatible T_ms, between 50 and 60⌈C (Saiki, 1989). Although the primers will be designed for an optimal annealing temperature, the PCR reactions must necessarily be reproducible and, without artefacts, each of the STS/PCR reactions will be tested for high specificity using a range of annealing temperatures for the oligonucleotide primers and a range of Mg^{++} concentrations (Saiki, 1989). The only caveat is that the RFLP seen from the cosmid could be quite distant from the STS, if for example, they are separated by several introns. The second source of simple STSs can be generated from DNA sequence derived from the ends of the YAC or P1 cloned inserts. This DNA sequence can be from `end-rescued' PCR products (Riley *et al.*, 1990) or from direct linear amplification DNA sequencing across the cloning junctions (Krishnan *et al.*, 1991). These STSs are specifically for contig construction, although they should be checked for genetically useful polymorphisms (cf. below).

Expression tagged sites (ETS) will be generated from a selected or `normalized' (removal of clones which represent highly expressed genes) set of cDNAs (Sankahavaram *et al.*, 1991). For this purpose, a directionally cloned cDNA library has been constructed from poly(A)+ RNA from RH tachyzoites in a λ-based vector (collaboration with Drs. J. Alexander and W. Harnett, Strathclyde University, UK) which has two vital properties. First, the cloning protocol demands that the 5' and 3' ends of the mRNA always be cloned in the same direction relative to the vector itself, thus it is possible to get DNA sequence information which is assured to be 5' or 3' specific. Second, the clone can be converted into a plasmid by growth with a `helper phage' for routine manipulation. Unfortunately, technical limitations make screening for chromosome specific ETSs difficult. However, for an overall

genome project, DNA sequencing of random clones from a partially normalized cDNA library is a reasonable strategy to generate ETSs, as the estimated number of genes, 13,000, is large enough such that random cDNA sequencing can proceed through several thousand clones with very little redundancy. Although the ETSs are by definition genetic information, they may not contain genetically useful polymorphisms, particularly since protein coding sequence is, on average, much less variable than non-coding sequence. Nonetheless, like the simple STSs, they can be tested for useful polymorphisms.

Since the future utility of these maps is somewhat dependent upon having at least a subset of the genetic markers tightly correlated with the physical marker for fine-scale genetic mapping, simple methods for either generating pSTSs *de novo* or screening simple STSs and ETSs must be developed. Although the calculations above suggest that generating pSTSs from random pieces of DNA is feasible, this directed approach would be secondary to the conversion of simple STSs and ETSs until more evidence for the existence and frequency of highly polymorphic structures is found. Although there are several methods which can detect single base substitutions, the heteroduplex method seems to be the simplest and most effective procedure (Belle White *et al.,* 1992). Using a series of point mutations in the equine infectious anaemia virus (EIAV), eight of nine point mutations over a 372 bp PCR product were detected with this procedure. Each of the point mutants were PCR amplified in the presence of ^{32}P-dCTP. Heteroduplex molecules were formed by subjecting the pair-wise-mixed PCR products to heat denaturation followed by slow cooling to room temperature. The heteroduplex molecules were best resolved by 5% polyacrylamide and 10% urea gel electrophoresis and subsequent autoradiographic exposure of the gel. The procedures for testing STSs and ETSs can be performed exactly as in the EIAV experiment, since the parents and the F_1 progeny in a *Toxoplasma gondii* cross are all haploid. This method should work with diploid *T. b. brucei* if the strains are homozygous at most loci, rendering them effectively haploid. Even if they are not, if the strains do not share common alleles at a locus, the procedure will still effectively mark alleles. This makes all the manipulations and data interpretation very rapid and simple. Moreover, the testing for polymorphism only requires parental DNA for PCR. Given the estimation that a polymorphism will occur about 1 per 240 bp, a random distribution (Poisson) of base changes will result in finding a polymorphism in somewhat better than 60% of the STSs in this size range. Once the STS has been shown to be a pSTS, scoring F_1 progeny from a cross should be limited only by speed of running the gels as the PCR and heteroduplex formations are adaptable to current robotic automation (Bentley *et al.,* 1992). Eventually, even running gels may be eliminated as a ligation-mediated assay for heteroduplex molecules has been adapted to an ELISA assay (Nickerson *et al.,* 1990).

SPECIFIC CHROMOSOME MAPPING PROJECTS

Toxoplasma gondii

The ultimate goal of the *Toxoplasma gondii* genome project will be the construction of a complete genome map, i.e. the combination of physical and genetic maps (Ajioka *et al.,* 1991). Ordered whole genomic libraries will be subdivided into chromosome-specific sets of clones.

Contigs will be constructed using a clone-limited strategy (Palazzolo *et al.*, 1991). The assembly of the genome map will begin by mapping existing RFLPs, and the new STSs, pSTSs, ETSs, onto the contigs, with other laboratories contributing to development of the map through their access to high density filters and clones from the `big DNA' libraries produced. The genome map of *Toxoplasma* will thus become an important tool for future molecular genetic studies. In our own laboratory we will focus initially on constructing a physical map of chromosome III, a 3 Mb chromosome which shows a 15% size decrease in the RH strain relative to the P and C strains used in the genetic cross from which the genetic linkage map was derived. The reasons for this size difference may be uncovered with STS/ETS mapping. Working in parallel, Dr. David Sibley (Washington University) will use the libraries to construct a physical map of chromosome V, which has a marker showing linkage to a locus regulating resistance to adenine ara inoside (ara-1; Pfefferkorn *et al.*, 1977).

Trypanosoma brucei brucei

African trypanosomes can also be crossed in the laboratory, but there is a dearth of genetic markers available to analyze crosses. Work on the *T. b. brucei* genome will build on current research already being undertaken (S.E. Melville *et al.*, this volume). This has included analysis of hybrids derived from a cross, carried out by Prof. A. Tait and Dr. C.M.R. Turner in Glasgow, between TREU 927/4 (originally isolated from a tsetse fly in Kenya in 1970) and STIB 247 (originally isolated from a vertebrate host in Tanzania in 1971). The parental stocks and nine hybrid clonal stocks were made available to Dr. Le Page in Cambridge. When the clones were analyzed by CHEF PFGE, astonishing chromosome polymorphisms were observed, with each clonal hybrid population differing in its karyotype from all others and from the parental lines. Chromosomes differ in size and number, and all hybrids contain megabase-sized chromosomes which do not correspond in electrophoretic mobility to a chromosome from either parent. When Southern blots of karyotype gels are probed with single-copy markers, these mostly detect two chromosomes of very different sizes. Long-range restriction mapping reveals homology over physical distances up to 1 Mb. Since the size differences are too great to attribute to antigenic variation, there is yet no clue as to the source of the extra DNA in these apparent `homologues'. Markers used in these initial studies were not in linkage groups, or even assigned to specific chromosomes. Priority has therefore been given to develop the first set of linked markers along one pair of homologous chromosomes using two complementary approaches: cDNA selection using a single radioactively labelled PFGE-purified chromosomal DNA band to develop chromosome-specific markers (Ajioka *et al.*, 1991) and preparation of a 6 × genomic equivalent cosmid library and screening using the whole chromosome probe to generate a chromosome-specific subgenomic library for chromosome walking (Ashford *et al.*, 1992). The project is now poised to link into the strategies outlined for the *Toxoplasma* genome project, with plans to develop STSs, pSTSs and ETSs for genetic and physical mapping, as well as construction of additional P1, YAC and YAC telomere libraries for mapping over a broader range of chromosome sizes. Construction of physical/genetic maps and analysis of the hybrid clones provides a unique way of analyzing and

understanding the dramatic karyotypic variability observed in trypanosomes and its contribution to the biology and survival of the parasite.

Leishmania Species

Current work in our laboratory has already begun to look at genetic variability in leishmanial parasites, particularly in relation to molecular determinants of virulence (reviewed Belle White *et al.,* 1992, Chang and Chaudhuri, 1990). One aspect of this has again been the demonstration of remarkable genetic heterogeneity (Reiner *et al.,* 1989) and CHEF PFGE-analyzed karyotypic variability (Blackwell, 1992; E. Espinoza, J.W. Ajioka and J.M. Blackwell, unpublished data) in fresh isolates of *L. peruviana.* Southern blot analysis with specific gene probes (e.g. GP63, HSP70) showed hybridization to different sized chromosomes in different isolates (Blackwell, 1992) and clones (E. Espinoza, J.W. Ajioka and J.M. Blackwell, unpublished data). Without good physical/genetic maps of *Leishmania,* it has thus far been impossible to analyze this kind of karyotypic variation or to determine whether isolates and clones with two GP63 or two HSP70 bearing chromosomes of different sizes represent homologous pairs of chromosomes. Work has commenced to produce ordered P1 and YAC genomic libraries of *L. peruviana,* and to modify (in collaboration with Dr. Steve Beverley, Harvard) the P1 vector to allow its use as a dual purpose mapping/shuttle vector library. A YAC telomere library has been made, restriction maps generated for several telomeres, and sequence data obtained to begin to develop STSs to map the telomeres onto specific chromosomes. One project already well established in the laboratory is to study genetic variability at the GP63 locus and to construct a physical map for the GP63-containing chromosome of *L. peruviana* by creating a contig from a subset of selected overlapping clones. Sequence analysis at the sites of overlaps will allow us to design oligonucleotide primer pairs to develop STS and pSTS PCR markers. cDNA libraries have also been made and will be used to develop ETSs. Sequence tagged site and expression tagged site maps of a `gold standard' GP63 chromosome and a few known size variants can then be constructed, and the STS/PCR technology used to analyze large numbers of field isolates and clones from Peru. The *Leishmania* work will also build on current research in the Department of Pathology being undertaken by Dr. Douglas Barker and his colleagues, in particular, the work of Eresh (1992) on karyotypic variability and genome organization in *L. braziliensis* which has concentrated on two small (50 and 350 kb) chromosomes. These chromosomes show karyotypic variability between species and isolates of *L. braziliensis,* the smaller chromosome possibly acting as a transposable element (Eresh, 1992). This work again emphasizes the need for the development of a physical map for *Leishmania.* Since classical transmission genetics is impossible in *Leishmania,* physical mapping techniques provide the only way forward.

APPLICATIONS AND USE OF THE RESOURCE

The resource created will be used in this and other laboratories to address specific biological problems. The dual purpose shuttle vector/mapping libraries will facilitate the

use of reverse genetics/complementation to find genes regulating drug resistance in parasites, and to find missing links in unique parasite biochemical pathways. For example, in our own laboratory complementation will be used to screen for reconstitution of virulence in genetically stable avirulent clones in an attempt to identify genes responsible for major steps in the biosynthetic pathway of *Leishmania* lipophosphoglycan. Collaborative research (A. Skinner and J. Blackwell with Dr. Malcolm McConville at Dundee University) demonstrates that avirulent *L. mexicana* clones make truncated LPG molecules. Differentiation of avirulent clones appears to be blocked such that full-length metacyclic LPG is not made. Work is already in progress to identify genes differentially expressed in virulent and avirulent clones by developing subtractive cDNA libraries. Once a good `normalized' set of cDNA clones has been assembled, a combined approach of sequencing/database searching and screening for cosmid/P1s to determine genomic organization and provide clones for complementation studies can be initiated.

Finally, the parasite genomes to be studied, and the techniques which will be developed, provide excellent model systems in which to address key questions relating to the evolution and organization of eukaryote genomes.

ACKNOWLEDGEMENTS

Core support for the Laboratory for Parasite Genome Analysis in Cambridge, UK, is being provided by Glaxo Group Research Ltd., Greenford, London, UK.

REFERENCES

AJIOKA, J.W., SMOLLER, D.A., JONES, R.W., CARULLI, J.P., VELLEK, A.E.C., GARZA, D., LINK, A.J., DUNCAN, I.W. and HARTL, D.L. 1991. *Drosophila* genome project: one-hit coverage in yeast artificial chromosomes. *Chromosoma* 100: 495–509.

ASHFORD, R.W., DESJEUX, P. and DERAADT, P. 1992. Estimation of population at risk of infection and number of cases of leishmaniasis. *Parasitololgy Today* 8: 104–105.

BELLE WHITE, M., CARVALHO, M., DERSE, D., O'BRIEN, S.J. and DEAN, M. 1992. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 12: 301–306.

BASTIEN, P., BLAINEAU, C., TAMINH, M., RIOUX, J.A., ROIZES, G. and PAGES, M. 1990. Interclonal variations in molecular karyotype in *Leishmania infantum* imply a `mosaic' strain structure. *Molecular and Biochemical Parasitology* 40: 53–62.

BENTLEY, D.R., TODD, C., COLLINS, J., HOLLAND, J., DUNHAM, I., HASSOCK, S., BANKIER, A. and GIANELLI, F. 1992. The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. *Genomics* 12: 534–541.

BISHOP, R. 1990. Extensive homologies between *Leishmania donovani* chromosomes of markedly different size. *Molecular and Biochemical Parasitology* 38: 1–12.

BLACKBURN, E.H. 1984. Telomeres: do the ends justify the means? *Cell* 37: 7–8.

BLACKBURN, E.H. and SZOSTAK, J.W. 1984. The molecular structure of centromeres and telomeres. *Annual Review of Biochemistry* 53: 163–94.

BLACKWELL, J.M. 1992. Leishmaniasis epidemiology: all down to the DNA. *Parasitology* 104: S19–S34.

BURG, L.J., PERLEMAN, D., KASPER, L.H., WARE, P.L. and BOOTHROYD, J.C. 1988. Molecular analysis of the gene encoding the major surface antigen of *Toxoplasma gondii. Journal of Immunology* 141: 3584–3591.

BURKE, D.T., CARLE, G.F. and OLSON, M.V. 1987. Cloning large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.

CHANG, J.-F., SMITH, C.L. and CANTOR, C.R. 1989. Isolation and characterization of a human telomere. *Nucleic Acids Research* 17: 6109–6127.

CHANG, K.P. and CHAUDHURI, G. 1990. Molecular determinants of *Leishmania* virulence. *Annual Review of Microbiology* 44: 499–529.

CHUMAKOV, I., RIGAULT, P., GUILLOU, S. and 33 others. 1992. Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* 359: 380–387.

CLAYTON, C.E. 1988. The molecular biology of the Kinetoplastidae. In: Rigby, P.W.J., ed. *Genetic Engineering 7*. New York: Academic Press, 1–56.

COBURN, C.M., OTTENAN, K.M., McNEELY, T., TURCO, S.J. and BEVERLEY, S.M. 1991. Stable transfection of a wide range of trypanosomatids. *Molecular and Biochemical Parasitology* 46: 169–180.

CORNELISSEN, A.W.C.A. and OVERDULVE, J.P. 1985. Sex determination and sex differentiation in coccidia: gametogony and oocyst production after monoclonal infection of cats with free-living and intermediate host stages of *Isopora (Toxoplasma) gondii*. *Parasitology* 90: 35–44.

CORNELISSEN, A.W.C.A., OVERDULVE, J.P. and van der PLOEG, M. 1984. Determination of nuclear DNA of five Eucoccidian parasites *Isopora (Toxoplasma) gondii, Sarcocystis cruzi, Eimeria tenella, E. acervulina* and *Plasmodium berghei* with special reference to gamontogenesis and meiosis in *Isopora (Toxoplasma) gondii*. *Parasitology* 88: 531–553.

COULSON, A., SULSTON, J., BRENNER, S. and KARN, J. 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the USA* 83: 7821–7825.

CRUZ, A. and BEVERLEY, S.M. 1990. Gene replacement in parasitic protozoa. *Nature* 348: 171–173.

DESMONTS, G. and COUVREUR, J. 1974. Congenital toxoplasmosis: a prospective study of 378 pregnancies. *New England Journal for Medicine* 290: 1110–1116.

DUBEY, J.P. 1977. *Toxoplasma, Hammmondia, Besnotitia, Sarcosystis* and other tissue cyst-forming coccidia of man and animals. In: Krier, J.P., ed. *Parasitic Protozoa*. New York: Academic Press, pp. 101–237.

ERESH, S. 1992. Karyotypic analysis of *Leishmania*. PhD thesis. Department of Pathology, University of Cambridge.

GALINDO, I. and RAMIREZ OCHOA, J.L. 1989. *Leishmania mexicana mexicana* electrokaryotype by clamped homogeneous electric field electrophoresis. *Moleluar and Biochemical Parasitololgy* 34: 245–252.

GREEN, E.D. and OLSON, M.V. 1991. Chromosomal regions of the cystic fibrosis gene in yeast artificial chromosomes: a model of human genome mapping. *Science* 250: 94–98.

GREEN, E.D. and OLSON, M.V. 1990. Systematic screening of yeast artificial chromosomes by use of the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the USA* 87: 1213–1217.

HOLMES-GIANNINI, S., CURRY, S.S., TESH, R.B. and van der PLOEG, L.H.T. 1990. Size-conserved chromosomes and stability of molecular karyotype in cloned stocks of *Leishmania major. Molecular and Biochemical Parasitololgy* 39: 9–22.

HULTMAN, T., STAHL, S., HORNES, E. and UHLEN, M. 1989. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Research* 17: 4937–4946.

IOVANNISCI, D.M. and BEVERLEY, S.M. 1989. Structural alterations of chromosome 2 in *Leishmania major* as evidence for diploidy, including spontaneous amplification of the mini-exon arrray. *Molececular and Biochemical Parasitology* 34: 177–188.

ISH-HOROWICZ, D. and BURKE, J.F. 1981. Rapid and effective cosmid cloning. *Nucleic Acids Research* 9: 2989.

KELLY, J.M., LAW, J.M., CHAPMAN, C.J., Van EYS, G.J.J.M. and EVANS, D.A. 1991. Evidence of genetic recombination in *Leishmania. Molecular and Biochemical Parasitololgy* 46: 253–264.

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

JEFFRIES, A.J., WILSON, V., THEIN, S.L., WEATHERALL, D.J. and PONDER, B.A.J. 1986. DNA `fingerprints' and segregation analysis of multiple markers in human pedigrees. *American Journal of Human Genetics* 39: 11–24.

JENNI, L., MARTI, S., SCHEWIZER, J., BETSCHART, B., le PAGE, R.W.S., WELLS, J.M., TAIT, A., PAINDEVOINE, P., PAYS, E. and STEINERT, M. 1986. Hybrid formation between African trypanosomes during cyclical transmission. *Nature* 322: 173–175.

KRISHNAN, B.R., BLAKESLEY, R.W. and BERG, D.E. 1991. Linear amplification DNA sequencing directly from single phage plaques and bacterial colonies. *Nucleic Acids Research* 19: 1153.

LARIN, Z., MONACO, A.P. and LERACH, H. 1991. Yeast artificial chromosome libraries containing inserts from mouse and human DNA. *Proceedings of the National Academy of Sciences of the USA* 88: 4123–4127.

LUFT, B.J. and REMINGTON, J.S. 1988. Toxoplasmic encephalitis. *Journal of Infectious Diseases* 157: 1–6.

McCOMBIE, W.R., ADAMS, M.D., KELLEY, J.M., FITZGERALD, M.G., UTTERBACK, T.R., KHAN, M., DUBNICK, M., KERLAVAGE, A.R., VENTER, J.C. and FIELDS, C.R. 1992. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genetics* 1: 124–131.

MILLS, J. 1986. *Pneumocystis carnii* and *Toxoplasma gondii* infections in patients with AIDS. *Reviews of Infectious Diseases* 8: 1001–1011.

NAKAMURA, Y., LEPPERT, M., O'CONNELL, P., WOLFF, R., HOLM, T., CULVER, H., MARTIN, C., FUJIMOTO, E., HOFF, M., KUMLIN, E. and WHITE, R. 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616–1622.

NICKERSON, D.A., KAISER, R., LAPPIN, S., STEWART, J., HOOD, L. and LANDERGREN, U. 1990. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proceedings of the National Academy of Sciences of the USA* 87: 8923–8927.

NICKERSON, D.A., WHITEHURST, C., BOYSEN, C., CHARMLEY, P., KAISER, R. and HOOD, L. 1992. Identification of clusters of biallelic polymorphic sequence-tagged sites (pSTSs) that generate highly informative and automatable markers for genetic linkage mapping. *Genomics* 12: 377–387.

OLSON, M.V., HOOD, L., CANTOR, C. and BOTSTEIN, D. 1989. A common language for physical mapping of the human genome. *Science* 245: 1434–1440.

PAGES, M., BASTIEN, P., VEAS, F., ROSSI, V., BELLIS, M., WINCKER, P., RIOUX, J.A. and ROIZES, G. 1989. Chromosome size and number polymorphisms in *Leishmania infantum* suggest amplification/deletion and possible genetic exchange. *Molecular and Biochemical Parasitology* 36: 161–168.

PALAZZOLO, M.J., SAWYER, S.A., MARTIN, C.H., SMOLLER, D.A. and HARTL, D.L. 1991. Optimized strategies for sequence-tagged-site selection in genome mapping. *Proceedings of the National Academy of Sciences of the USA* 88: 8034–8038.

PFEFFERKORN, E.R. and KASPER, L.H. 1983. *Toxoplasma gondii:* genetic crosses reveal phenotypic suppression of hyroxyurea resistance by fluorodeoxyuridine resistance. *Experimental Parasitology* 55: 207–218.

PFEFFERKORN, E.R., PFEFFERKORN, L.C. and COLBY, E.D. 1977. Development of gametes and oocysts in cats fed with cysts derived from cloned trophozoites of *Toxoplasma gondii. Journal of Parasitology* 63: 158–159.

PFEFFERKORN, L.C. and PFEFFERKORN, E.R. 1980. *Toxoplasma gondii:* genetic recombination between drug resistant mutants. *Experimental Parasitology* 50: 305–316.

REINER, N.E., LO, R., LLANOS-CUENTAS, A., GUERRA, H., BUTTON, L.L. and McMASTER, W.R. 1989. Genetic heterogeneity in Peruvian *Leishmania* isolates. *American Journal of Tropical Medicine and Hygiene* 41: 416–421.

RILEY, J., BUTLER, R., OGILVIE, D., FINNIEAR, R., JENNER, D., POWELL, S., ANAND, R., SMITH, J.C. and MARKHAM, A.F. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Research* 18: 2887–2890.

SAIKI, R.K. 1989. The design and optimization of the PCR. In: Ehrlich, H., ed. *PCR Technology.* New York: Stockton Press, pp. 7–16.

SAMARAS, N. and SPITHILL, T.W. 1987. Molecular karyotype of five species of *Leishmania* and analysis of gene locations and chromosomal rearrangements. *Molecular and Biochemical Parasitology* 25: 279–291.

SANKAHAVARAM, R.P., PARIMOO, S. and WEISSMAN, S. 1991. Construction of a uniform abundance (normalized) cDNA library. *Proceedings of the National Academy of Sciences of the USA* 88: 1943–1947.

SCHOLLER, J.K., REED, S.G. and STUART, K. 1986. Molecular karyotype of species and subspecies of *Leishmania*. *Molecular and Biochemical Parasitology* 20: 279–293.

SIBLEY, L.D., LEBLANC, A.J., PFEFFERKORN, E.R. and BOOTHROYD, J.C. 1992. Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii*. *Genetics*, in press.

SIBLEY, L.D. and BOOTHROYD, J.C. 1992. Construction of a molecular karyotype for *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* 51: 291–300.

SIBLEY, L.D., PFFEFERKORN, E.R. and BOOTHROYD, J.C. 1991. Proposal for a uniform genetic nomenclature in *Toxoplasma gondii*. *Parasitology Today* 7: 327–328.

STERNBERG, N. 1990. Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobase pairs. *Proceedings of the National Academy of Sciences of the USA* 89: 103–107.

SULSTON, J., DU, Z., THOMAS, K., WILSON, R., HILLIER, L., STADEN, R., HALLORAN, N., GREEN, P., THIERRY-MIEG, J., QIU, L., DEAR, S., COULSON, A., CRAXTON, M., DURBIN, R., BERKS, M., METZSTEIN, M., HAWKINS, T., AINSCOUGH, R. and WATERSTON, R. 1992. The *C. elegans* genome sequencing project: a beginning. *Nature* 356: 37–41.

TURNER, C.M.R., STERNBERG, J., BUCHAN, N., SMITH, E., HIDE, G. and TAIT, A. 1990. Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitology* 101: 377–386.

VENTER, J.C., ADAMS, M.D., MARTIN-GALLARDO, A., McCOMBIE, W.R. and FIELDS, C. 1992. Genome sequence analysis: scientific objectives and practical strategies. *TIBTECH* 10: 8–11.

WATERSTON, R., MARTIN, C., CRAXTON, M., HUYNH, C., COULSON, A., HILLIER, L., DURBIN, R., GREEN, P., SHOWNKEEN, R., HALLORAN, N., METZSTEIN, M., HAWKINS, T., WILSON, R., BERKS, M., DU, Z., THOMAS, K., THIERRY-MIEG, J. and SULSTON, J. 1992. A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genetics* 1: 114–123.

WEBER, J.L and MAY, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44: 388–396.

Summary of discussion

Chairperson: Dr. O. ole-MoiYoi
Rapporteurs: Drs. S. Kemp and S. Morzaria

Three papers were presented in this session describing the rationale and requirements for undertaking major mapping projects. Prof. C. Smith (Boston University) reviewed the aims and objectives of the internationally coordinated Human Genome Project. The ultimate goal of this project is to provide within 15 years the entire sequence of the human genome. The short-term objectives are to generate high-resolution genetic and physical maps and produce overlapping DNA libraries for each of the human chromosomes. Dr. A. Teale (ILRAD) described the bovine genome project, the principal aims of which are to identify genes controlling aspects of food production and quality. ILRAD's role in the project is to identify markers linked to genes controlling trypanotolerance. This is being achieved through the use of markers (RAPDs, microsatellites and specific genes) and the generation of N'Dama and Boran cattle crosses segregating trypanotolerance. Prof. J. Blackwell (Cambridge University) described the initiative taken by her group to establish a new laboratory for parasite genome analysis. The aim of this group is to produce YAC, P1 and cosmid libraries of the three important protozoan parasites: *Toxoplasma, Trypanosoma* and *Leishmania*. They will establish a central database, develop a series of PCR-based markers and make these resources available to the international scientific community for use in solving biological problems induced by these organisms.

The discussion that followed was limited to the bovine and protozoan genome analysis programs. Although it was recognized that the approach to identifying the genes for trypanotolerance was relevant, it was not clear what time period would be required to achieve this goal. It is estimated that the required markers may be available by 1997 and the genes may be identified two to three years after that. The possibility of linkage between trypanotolerance and some undesirable traits and its effect on breeding programs was raised. Current breeding data do not show that this is the case but a definite statement cannot be made until further results are obtained. With regard to the question of whether the trypanotolerance trait is determined by a single gene or by a number of genes, it was pointed out that the preliminary analysis of the F_2 cattle from the N'Dama-Boran crosses indicates that the trait is likely to be multigenic. It is also estimated that with 150 markers and 160 cattle, 10% recombination can be detected. However, the difficulty of differentiating 10% of the phenotype from that background was recognized.

The advantages and disadvantages of investing in parasite genome analysis projects were discussed extensively. Parasite/host interactions are not static. Generation of polymorphism is an important survival strategy for parasitic protozoa and once new control strategies are introduced, parasites will change and adapt themselves for survival. Therefore, this was enough justification for funding genome analysis programs. Studies on genome analysis would allow researchers a better understanding of the

underlying mechanisms of polymorphism and perhaps enable them to stay a step ahead of the parasites. It was pointed out that in mapping parasite genomes, it would be desirable first to take the approach of identifying the traits of interest, as has been done in the bovine genome project, and then to perform linkage analysis, to determine recombination frequencies and to follow the steps in identifying relevant genes. Because parasite genomes are small, it will probably be easier and cheaper to map them than the genomes of higher eukaryotes. However, it is necessary to take into account gene density, which appears to be higher in these organisms than in higher eukaryotes. Additionally, the cost of sequencing is still high (~US$1/base).

In view of the relatively small sizes of protozoan chromosomes compared to higher eukaryotes, the desirability of constructing YAC libraries was questioned. It was suggested that cosmid and P1 libraries may be more appropriate. However, it was recognized that all libraries have some short-comings and a combination of libraries may be required to obtain complete contig maps. It was appreciated that multiple libraries involve a lot of work and increase the complexity of characterization. With regard to the technical problems of fingerprinting clones for alignment, it was suggested that use of techniques such as image analysis and exploitation of relevant computer software programs, developed at the Department of Energy, USA, would greatly expedite this work.

GENETIC ANALYSIS OF *THEILERIA* AND *PLASMODIUM*

Genetic recombination in field populations of *Plasmodium falciparum*

D. Walliker, L.C. Ranford-Cartwright and H.A. Babiker

Malaria parasites undergo a mainly haploid life-cycle. The only diploid stage is the zygote, formed by fusion of gametes in the mosquito stomach. The first division of the zygote is a meiotic one, producing, after further mitotic divisions, haploid sporozoites. Genetic recombination occurs at meiosis, following cross-fertilization of gametes of parasites with different genotypes. This has been shown in laboratory studies by feeding mosquitoes on a mixture of *Plasmodium falciparum* clones and analyzing the resulting progeny for parasites with non-parental combinations of the clone markers. Such recombinants are produced at a higher than expected frequency.

There is considerable genotype diversity in field populations of *P. falciparum.* Evidence that recombination in mosquitoes is the principal cause of this diversity is two-fold. First, parasites isolated from patients in small isolated communities at the same time are genetically very diverse. No two isolates examined for polymorphic markers at some 20 loci have been found to possess identical combinations of the allelic variants of these genes. Second, examination of oocysts in wild-caught mosquitoes by the PCR technique has shown that a high proportion are heterozygotes. There is thus frequent crossing in natural populations of this parasite.

In addition to recombination at meiosis, it is also clear that genetic changes can occur during asexual multiplication of *P. falciparum* blood forms, as shown by deletions of regions of certain chromosomes during *in vitro* culture. The extent to which this occurs in nature is not known.

INTRODUCTION

Genetic recombination is classically defined as the process by which organisms with novel combinations of genes are produced in crosses between two parent organisms. Recombination occurs primarily at meiosis, in two principal ways: (i) independent segregation of variant forms (alleles) of genes on different chromosomes and (ii) crossing-over events between linked genes on the same chromosome. A third aspect that must be considered is intragenic recombination, by which novel alleles of a given gene can be formed by crossing-over events within the gene itself. This process is likely to occur very rarely, but may be significant where selection pressures against existing alleles are high. Recombination at mitosis also occurs, although much more rarely. In the case of haploid organisms such as *Plasmodium,* mitotic recombination is not expected to have important genetic consequences, since it can only involve exchange between identical (sister) chromatids, which contain identical alleles of any given gene.

It is now clear that recombination is of primary importance in generating the genetic diversity seen in natural populations of *Plasmodium falciparum.* We review here briefly current knowledge on the organization of the parasite genome and on ways in which chromosome diversity arises. Data on recombination in this parasite both in laboratory and field studies are then discussed.

ORGANIZATION OF THE *PLASMODIUM FALCIPARUM* GENOME

The nuclear genome of *P. falciparum,* and most probably of other species of *Plasmodium,* consists of 14 chromosomes. They range in size from 0.65 to 3.4 megabases (Mb) (Triglia *et al.,* 1992), and are visible only by pulsed-field gradient gel electrophoresis (PFGE). Two cytoplasmic genetic elements are also present, a linear element of tandemly repeated units of 6 kb which contains mitochondrial genes and a 35 kb circular DNA element which encodes chloroplast-like genes (Wilson *et al.,* 1991).

Genes can be mapped to chromosomes by hybridization of gene probes to blots of pulsed-field gels (Triglia *et al.,* 1992). The linear order of genes is also being elucidated by this approach, using blots of chromosome fragments cut by rare-cutting restriction enzymes (Sinnis and Wellems, 1988; Corcoran *et al.,* 1988) and of yeast artificial chromosome (YAC) constructs from certain chromosomes (Triglia and Kemp, 1991).

GENETIC EVENTS IN THE LIFE-CYCLE

The parasite is haploid for most of its life-cycle. The only diploid form is the zygote, formed by union of gametes in the mosquito midgut. Meiosis occurs within a few hours of zygote formation (Sinden and Hartley, 1985), resulting eventually in the production of haploid sporozoites in each oocyst.

Recombination during meiosis is expected to be responsible for the formation of most novel genotypes. Laboratory crossing work has demonstrated that this process is very efficient in *Plasmodium.* In these studies, deliberate mixtures of parasite clones are fed to mosquitoes in order to allow crossing between gametes of each clone. The resulting sporozoites are used to establish new infections, and the resultant organisms examined for the presence of parasites with non-parental (recombinant) combinations of characters. Numerous crosses done with rodent malaria species and with *P. falciparum* have shown that recombinants are produced in this way at a high frequency (Walliker, 1989).

A recent advance in genetic analysis of malaria parasites is the ability to examine the genetic composition of oocysts in mosquitoes. This has been achieved by using the polymerase chain reaction (PCR) on certain *P. falciparum* antigen genes (Ranford-Cartwright *et al.,* 1991). When oocysts derived from mosquitoes fed on a mixture of two clones, denoted 3D7 and HB3, were examined in this way, approximately 50% were found to be heterozygous for the parent genes (Ranford-Cartwright *et al.,* 1993). This showed directly that random fertilization events had occurred between the gametes of each clone, producing hybrid and parental type zygotes. In the hybrid zygotes, meiotic recombination leads to the production of haploid recombinant forms. The inheritance of the 6 kb cytoplasmic element has recently been studied. In crosses between *P. falciparum* clones, oocysts found to be hybrid for nuclear gene markers proved to have only a single parental form of this element (Creasey *et al.,* 1993). Such uniparental inheritance is typical for mitochondria in many other organisms.

CHROMOSOME POLYMORPHISM

A striking early finding on *Plasmodium* chromosomes was the remarkable variation in the size of homologous chromosomes in different isolates and clones. This could be seen in

parasites taken directly from different infected patients (Corcoran *et al.*, 1986; Babiker *et al.*, 1991b), and so the phenomenon was clearly not an artefact of *in vitro* culture. In the past five years, it has become clear that several mechanisms are involved in generating these size polymorphisms, as follows:

Crossing-over at Meiosis

As might be expected for a eukaryotic organism, meiosis results in frequent crossing-over events. This has been demonstrated in laboratory crosses between *P. falciparum* clones with different sized chromosomes. For example, in a cross between clones 3D7 and HB3 which possessed a different-sized chromosome 4, Sinnis and Wellems (1988) showed that a crossover in the central region of this chromosome accounted for the production of a chromosome intermediate in size between those of the parent clones used in the cross. Extensive crossing-over events in all 14 chromosomes have been demonstrated in another cross by examining the inheritance patterns of numerous chromosome-specific markers (Walker-Jonah *et al.*, 1992).

Chromosome Breakage During Asexual Division

There are several reports of deletions of portions of chromosomes in blood forms of *P. falciparum* maintained in culture. This happens especially near chromosome ends. For example, Scherf *et al.* (1992) showed that a large part of a sub-telomerically located gametocyte-specific gene (Pf11.1) on chromosome 10 became deleted during culture of a cloned parasite line. Parasites with the truncated gene were found among sub-clones of this clone. Similar findings have been made with a knob-associated histidine-rich protein (KAHRP) on chromosome 2 (Pologe and Ravetch, 1988). New telomeres are formed at the breakage points in these instances. The mechanisms by which this occurs are not understood. It seems unlikely that recombination is involved since telomeric repeat sequences are not present at any of the breakpoints which have been studied (Scherf and Mattei, 1992).

Increases in Chromosome Size

There are occasional reports of an increase in the size of certain chromosomes following prolonged asexual passage. This has been studied particularly in the rodent malaria species *P. berghei* (Janse *et al.*, 1989). In some instances, these increases appear to be due to additions of a 2.3 kb sequence commonly found in subtelomeric sites on several chromosomes of this species (Dore *et al.*, 1990; Pace *et al.*, 1990). Janse *et al.* (1992) describes a line of *P. berghei* in which a chromosome of larger than normal size appeared to contain a translocated portion of another chromosome, although it was not clear whether this had occurred during asexual passage of this parasite.

Gene Amplification

Amplification of genes in *Plasmodium* has been studied particularly with respect to the so-called multi-drug resistance (mdr) genes, which have been thought to be involved in resistance to chloroquine. Foote *et al.* (1989) described lines and clones of *P. falciparum* with differing numbers of copies of the Pfmdr1 gene, arranged in tandem arrays on chromosome 5, which, they considered, were responsible for the size polymorphisms seen in this chromosome. There was some evidence of instability of such tandem arrays in one clone (B8) during culture.

In nature, it can be expected that meiotic recombination is the principal mechanism for the generation of chromosomal polymorphisms. While the other phenomena outlined above have been shown to occur frequently in laboratory cultures, the extent to which they occur in nature is not known.

RECOMBINATION IN NATURAL PARASITE POPULATIONS

The extent of crossing, and hence of recombination, in nature is now being addressed. The topic has been the subject of some recent debate (Tibayrenc amd Ayala, 1991; Walliker, 1991; Dye, 1991). If recombination occurs rarely, or never at all, one would expect to find in circulation only a limited number of genetically different parasites, with some forms being over-represented. If it is frequent, then a large diversity of parasite clones would be expected, exhibiting all possible combinations of the alleles of polymorphic genes.

This subject can be examined in two ways. First, the genotypes of parasites in patients can be examined to determine whether their frequencies are in accordance with a randomly interbreeding model. Second, direct evidence of crossing can be sought by determining the frequency of hybrid oocysts in wild-caught mosquitoes.

Diversity of Parasites in the Blood of Patients

Characters Studied

As mentioned above, blood forms of *P. falciparum* from different patients can be examined for gross genetic differences by comparing their chromosomes, using the PFGE technique. More precise genetic differences can be studied by examining allelic variation in specific genes. Characters available for such work include enzymes and other proteins revealed by electrophoretic techniques, antigens using monoclonal antibodies and variations in drug-sensitivity (Creasey *et al.*, 1990). Each of these characters requires the availability of cultured parasite material.

In recent years, the polymerase chain reaction (PCR) technique has been developed to amplify alleles of genes whose DNA sequence is known. Two merozoite surface antigens, MSP-1 and MSP-2, have proved particularly useful in this regard (Snewin *et al.*, 1991). Each contains regions of repetitive sequences, and alleles of each gene vary in both the numbers of repeats present and in their sequence. The PCR-amplified products of these regions can be examined for size variation by electrophoresis, and by blotting and

hybridization with allele-specific oligonucleotides for sequence differences. A particular advantage of this technique is that it can be performed on small quantities of parasite material, for example in fingerprick blood samples, without the need for culturing (Foley et al., 1992).

Results

Numerous studies have now been carried out on genetic diversity in *P. falciparum* and other species in various countries. Those most relevant to answering questions on the frequency of recombination include those on enzymes in *P. chabaudi* (Beale *et al.,* 1978) and *P. falciparum* (Carter and Voller, 1975), and on enzymes, 2D-PAGE proteins and antigens in *P. falciparum* (Creasey *et al.,* 1990; Babiker *et al.,* 1991a, 1991b; Conway and McBride, 1991). The principal findings from all these studies are:

(i) There is considerable allelic diversity of many genes, especially those encoding antigens. For example, in *P. falciparum* Conway and McBride (1991) found 36 alleles of antigen MSP-1, differing by epitopes recognized by monoclonal antibodies, among isolates of *P. falciparum* from a peri-urban region in the Gambia. H.A. Babiker and L.C. Ranford-Cartwright (unpublished data) have used PCR to differentiate 22 alleles of the same gene among only 50 isolates in Tanzania. It is quite probable that this diversity of alleles has come about by extensive intragenic recombination, as suggested by Tanabe *et al.* (1987), although formal proof for this has not been obtained in crossing experiments.

(ii) No two isolates have been found in which parasites possess identical genotypes. This is the case even in a small community in which malaria transmission is highly seasonal (Babiker *et al.,* 1991a, 1991b) and in which a restricted number of genotypes might be expected to be in circulation.

(iii) Mixed infections with more than one genetically distinct clone are common.

(iv) There is geographical variation in the frequencies with which alleles of many genes occur. For example, the frequency of an electrophoretic form of adenosine deaminase denoted ADA-2 is very rare in Thailand and the Gambia, but common in Brazil (Creasey *et al.*, 1990).

Diversity of Parasites in Mosquitoes

Work is now starting to examine the genotypes of oocysts of *P. falciparum* in wild-caught mosquitoes in villages near Ifakara, Tanzania (H.A. Babiker, L.C. Ranford-Cartwright, D. Charlwood and P. Billingsley, unpublished data). This is a region where malaria is highly endemic and some 60% of the inhabitants exhibit *P. falciparum* parasitaemias at any given time. Mixed infections with more than one clone are common. Sixty-eight percent of oocysts examined in *Anopheles gambiae* and *A. funestus,* caught in houses of these villages, have proved to be heterozygous for alleles of MSP-1, MSP-2, or of both genes.

CONCLUSIONS

The genetic diversity seen in natural infections of *P. falciparum* has now been shown to be considerable. As discussed above, chromosomal polymorphisms are extensive in

different parasite isolates, even in those from small communities (Babiker *et al.*, 1991b). There is also remarkable allelic variation of many individual genes, especially those encoding antigens. Perhaps the most significant finding is that no two isolates yet studied have identical genotypes.

It is now quite clear that the numerous combinations of alleles of the genes studied seen among clones in natural infections are evidence of extensive recombination in nature. This presumably happens as a consequence of crossing between clones when mosquitoes take up mixtures of gametocytes. Self-fertilization events between genetically identical gametes also occur in such instances. In addition, selfing is the only possibility if mosquitoes take up gametocytes of a single clone. In Tanzania, some 32% of oocysts are homozygous at two highly polymorphic loci, and these are almost certainly derived from such selfing events.

Thus, evidence from population studies of *P. falciparum* strongly supports the view that gametes of this parasite undergo random mating. When cross-mating occurs, recombination along classical Mendelian lines ensures that recombinants are produced readily. The `clonality' model proposed by Tibayrenc *et al.* (1990) is thus inappropriate when applied to this parasite. It could, at least in theory, apply in regions where clonal infections are the norm, in which mosquito transmission would maintain pure clones. In practice, no regions with infections of this type have yet been found.

ACKNOWLEDGEMENTS

We are happy to acknowledge support from the Medical Research Council of Great Britain, the World Health Organization, and the Commission of the European Communities for some of the work described here.

REFERENCES

BABIKER, H.A., CREASEY, A.M., FENTON, B., BAYOUMI, R.A.L., ARNOT, D.E. and WALLIKER, D. 1991a. Genetic diversity of *Plasmodium falciparum* in a village in eastern Sudan. 1. Diversity of enzymes, 2D-PAGE proteins and antigens. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85: 572–577.

BABIKER, H.A., CREASEY, A.M., BAYOUMI, R.A.L., WALLIKER, D. and ARNOT, D.E. 1991b. Genetic diversity of *Plasmodium falciparum* in a village in eastern Sudan. 2. Drug resistance, molecular karyotypes and the *mdr1* genotype of recent isolates. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85: 578–583.

BEALE, G.H., CARTER, R. and WALLIKER, D. 1978. Genetics. In: Killick-Kendrick, R. and Peters, W., eds. *Rodent Malaria*. London: Academic Press, pp. 213–245.

CARTER, R. and VOLLER, A. 1975. The distribution of enzyme variation in populations of *Plasmodium falciparum* in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 69: 371–376.

CONWAY, D.J and McBRIDE, J.S. 1991. Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology* 103: 7–16.

CORCORAN, L.M., FORSYTH, K.P., BIANCO, A.E., BROWN, G.V. and KEMP, D.J. 1986. Chromosome size polymorphisms in *Plasmodium falciparum* can involve deletions and are frequent in natural parasite populations. *Cell* 44: 87–95.

CORCORAN, L.M., THOMPSON, J.K., WALLIKER, D. and KEMP, D.J. 1988. Homologous recombination within sub-telomeric repeat sequences generates chromosome size polymorphisms in *Plasmodium falciparum. Cell* 53: 807–813.

CREASEY, A., FENTON, B., WALKER, A., THAITHONG, S., OLIVEIRA, S., MUTAMBU, S. and WALLIKER, D. 1990. Genetic diversity of *Plasmodium falciparum* shows geographical variation. *American Journal of Tropical Medicine and Hygiene* 42: 403–413.

CREASEY, A.M., RANFORD-CARTWRIGHT, L.C., MOORE, D.L., WILLIAMSON, D.H., WILSON, R.J.M., WALLIKER, D. and CARTER, R. 1993. Uniparental inheritance of the mitochondrial gene cytochrome b in *Plasmodium falciparum. Current Genetics,* in press.

DORE, E., PACE, T., PONZI, M., PICCI, L. and FRONTALI, C. 1990. Organization of subtelomeric repeats in *Plasmodium berghei. Molecular and Cellular Biology* 10: 2423–2427.

DYE, C.M. 1991. Population genetics of nonclonal, nonrandomly mating malaria parasites. *Parasitology Today* 7: 236–240.

FOLEY, M., RANFORD-CARTWRIGHT, L.C. and BABIKER, H.A. 1992. Rapid and simple method for isolating malaria DNA from fingerprick samples of blood. *Molecular and Biochemical Parasitology* 53: 241–244.

FOOTE, S.J., THOMPSON, J.K., COWMAN, A.F. and KEMP, D.J. 1989. Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of *P. falciparum*. *Cell* 57: 921–930.

JANSE, C.J., BOORSMA, E.G., RAMESAR, J., van VIANEN, P., van der MEER, R., ZENOBI, P., CASAGLIA, O., MONS, B. and van der BERG, F.M. 1989. *Plasmodium berghei:* gametocyte production, DNA content and chromosome-size polymorphisms during asexual multiplication *in vivo. Experimental Parasitology* 68: 274–282.

JANSE, C.J., RAMESAR, J. and MONS, B. 1992. Chromosome translocation in *Plasmodium berghei*. *Nucleic Acids Research* 20: 581–586.

PACE, T., PONZI, M., DORE, E., JANSE, C., MONS, B. and FRONTALI, C. 1990. Long insertions within telomeres contribute to chromosome-size polymorphism in *Plasmodium berghei. Molecular and Cellular Biology* 10: 6759–6764.

POLOGE, L.P. and RAVETCH, J.V. 1988. Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* 55: 869–874.

RANFORD-CARTWRIGHT, L.C., BALFE, P., CARTER, R. and WALLIKER, D. 1991. Genetic hybrids of *Plasmodium falciparum* identified by amplification of genomic DNA from single oocysts. *Molecular and Biochemical Parasitology* 49: 239–244.

RANFORD-CARTWRIGHT, L.C., BALFE, P., CARTER, R. and WALLIKER, D. 1993. Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum. Parasitology,* in press.

SCHERF, A. and MATTEI, D. 1992. Cloning and characterisation of chromosome breakpoints of *Plasmodium falciparum:* breakage and new telomere formation occurs frequently and randomly in subtelomeric genes. *Nucleic Acids Research* 20: 1491–1496.

SCHERF, A., CARTER, R., PETERSEN, C., ALANO, P., NELSON, R., AIKAWA, M., MATTEI, D., PEREIRA da SILVA, L. and LEECH, J. 1992. Gene inactivation of Pf11.1 of *Plasmodium falciparum* by chromosome breakage and healing: identification of a gametocyte-specific protein with a potential role in gametogenesis. *EMBO Journal* 11: 2293–2301.

SINDEN, R.E. and HARTLEY, R.H. 1985. Identification of the meiotic division of malarial parasites. *Journal of Protozoology* 32: 742–744.

SINNIS, P. and WELLEMS, T.E. 1988. Long-range restriction maps of *Plasmodium falciparum* chromosomes: crossing-over and size variation among geographically distant isolates. *Genomics* 3: 287–295.

SNEWIN, V.A., HERRERA, M., SANCHEZ, G., SCHERF, A., LANGSLEY, G. and HERRERA, S. 1991. Polymorphism of the alleles of the merozoite surface antigens MSA1 and MSA2 in *Plasmodium falciparum* wild isolates from Colombia. *Molecular and Biochemical Parasitology* 49: 265–276.

TANABE, K., MacKAY, M., GOMAN, M. and SCAIFE, J.G. 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum. Journal of Molecular Biology* 195: 273–287.

TIBAYRENC, M. and AYALA, F.J. 1991. Towards a population genetics of micro-organisms: the clonal theory of parasitic protozoa. *Parasitology Today* 7: 228–232.

TIBAYRENC, M., KJELLBERG, F. and AYALA, F.J. 1990. A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the USA* 87: 2414–2418.

TRIGLIA, T. and KEMP, D.J. 1991. Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Molecular and Biochemical Parasitology* 44: 207–212.

TRIGLIA, T., WELLEMS, T.E. and KEMP, D.J. 1992. Towards a high resolution map of the *Plasmodium falciparum* genome. *Parasitology Today* 8: 225–229.

WALKER-JONAH, A., DOLAN, S., GWADZ, R.W., PANTON, L.J. and WELLEMS, T.E. 1992. An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favored linkage groups in a genetic cross. *Molecular and Biochemical Parasitology* 51: 313–320.

WALLIKER, D. 1989. Implications of genetic exchange in the study of protozoan infections. *Parasitology* 99: S49–S58.

WALLIKER, D. 1991. Malaria parasites: randomly interbreeding or `clonal' populations? *Parasitology Today* 7: 232–235.

WILSON, R.J.M., GARDNER, M.J., FEAGIN, J.E. and WILLIAMSON, D.H. 1991. Have malaria parasites three genomes? *Parasitology Today* 7: 134–136.

Genome mapping and linkage analysis in *Plasmodium falciparum:* toward identification of the chloroquine resistance determinant

T.E. Wellems

Determinants of important parasite traits in malaria include genes that affect disease transmission and infectivity, drug response, host immunity, erythrocyte invasion and cytoadherence. Detailed knowledge of these determinants will improve our understanding of malaria and lead to new targets for vaccine development and drug design. But these determinants are often difficult to identify and characterize. Biochemical experiments on malaria parasites are limited by the relatively small numbers of parasites available from preparative cultivations and their contamination by host cell components. When appropriate probes have been available, receptors, enzymes, antigens and structural elements have been isolated and studied through heterologous expression in bacterial or eukaryotic cells. It is fair to say, however, that the biochemical bases of most important parasite traits remain unknown.

A non-biochemical strategy to locate determinants of parasite traits is that of positional cloning, sometimes loosely described as reverse genetics. Crosses of cloned parasite lines, linkage analysis and chromosome mapping techniques are used to locate the chromosome segments harbouring genetic loci that govern traits of interest. The chromosome segments are then cloned as large pieces of DNA, mapped in detail and analyzed for transcribed regions and candidate genes. At the foundation of this approach lie genetic and physical maps of the parasite genome. In this review I discuss these maps as developed for *Plasmodium falciparum*, the agent of the most deadly human malaria, and describe how they are being used to identify the determinant of chloroquine resistance, now a serious therapeutic problem across the tropical world.

MAPS OF THE *PLASMODIUM FALCIPARUM* GENOME

There are two types of maps generally distinguished in genomic studies: the linkage map and the physical map. The linkage map refers to recombination frequencies between genes as measured by laboratory crosses. Map distances, assigned statistically from large numbers of progeny, are defined in terms of centimorgans (cM) where 1 cM represents 1% recombination between the parental forms of genes in the progeny. Map units are additive in the linkage map, reflecting the fact that genes lie in a linear arrangement along individual chromosomes.

The physical map is based upon distances along individual chromosome DNAs. Direct measurements of gene separations and chromosome size are typically given in units of nucleotide base pairs (bp). Distances on the map generally correlate with recombination rates in linkage studies, but variations occur because of recombination `hotspots' and variable crossover rates in the chromosomes.

Laboratory crosses of *P. falciparum* clones are expensive in the resources and time they require, so physical mapping has predominated in assigning relative locations of

genetic markers. For this purpose, methods of pulsed-field gel electrophoresis (PFGE) (Schwartz and Cantor, 1984; Carle and Olson, 1984) are used to separate the intact chromosome DNA molecules in agarose gels. Figure 1 shows a photograph of one such separation in which the DNA was stained with ethidium bromide. The bands represent intact chromosome DNA molecules; in some bands the intensity indicates obvious multiplets containing DNA from two or more chromosomes. Optical scanning methods and genetic analysis have shown that there are 14 nuclear chromosomes in *P. falciparum* (Wellems *et al.*, 1987; Kemp *et al.*, 1987; Sinnis and Wellems, 1988; Gu *et al.*, 1990), a number in agreement with that found by electron microscopy of kinetochores (Prensier and Slomianny, 1986).

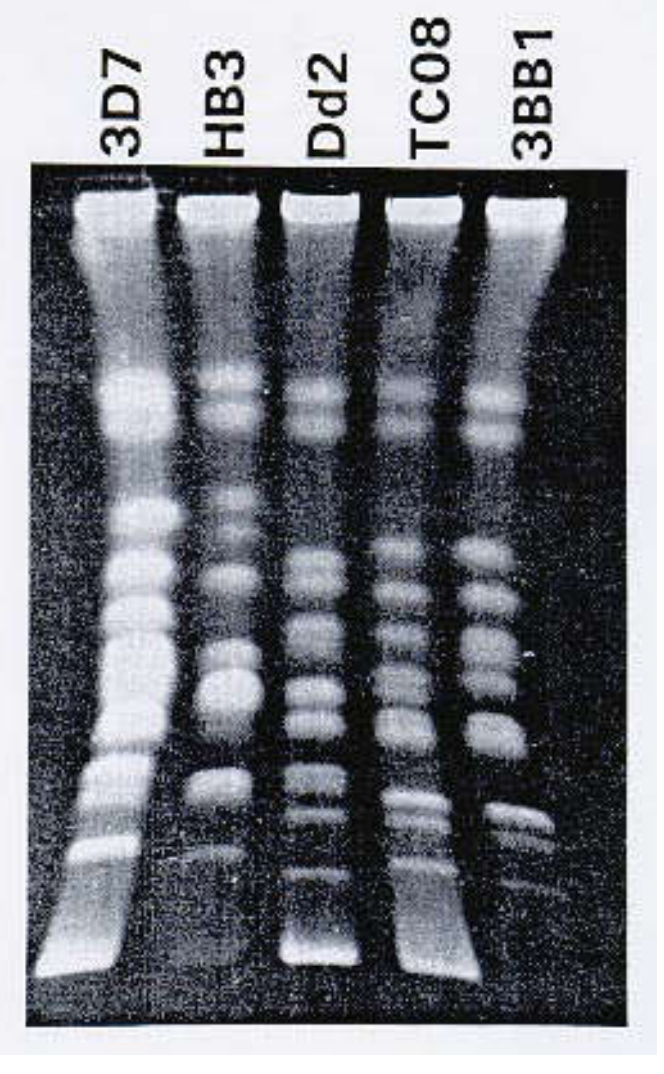

Figure 1. Separation of *Plasmodium falciparum* chromosome DNA molecules in an agarose gel as revealed by ethidium bromide staining. The linear DNA molecules range in size from about 800 kb to 3500 kb. Chromosome size polymorphisms are evident among the four parasite clones.

Full length chromosome maps are generated by excising *P. falciparum* chromosome DNA from pulse field gels, digesting the DNA with rare cutting restriction enzymes and mapping the resulting fragments (Wellems *et al.*, 1987; Sinnis and Wellems, 1988; Dolan *et al.*, 1993). Restriction enzymes with G+C rich recognition sites serve well for this purpose, presumably because of the AT-rich nature of the *P. falciparum* genome (Pollack *et al.*, 1982). Maps of nine *P. falciparum* chromosomes were recently summarized (Triglia *et al.*, 1992). The nuclear chromosomes have sizes of 600 to 3500 kb, constituting a haploid nuclear DNA content of 25,000–30,000 kb (Wellems *et al.*, 1987). DNA content in individual chromosomes may vary considerably among different parasite isolates as size variations of several hundred kilobases are common (Figure 1). The nature of this expendable DNA, still largely unknown, has been shown to contain several genes, some of which may be involved in cytoadherence and gametogenesis (Pologe and Ravetch, 1986; Shirley *et al.*, 1990). Mechanisms of chromosome size polymorphism involve recombination during meiosis (Sinnis and Wellems, 1988), breaking and healing events at telomeres (Pologe and Ravetch, 1988) and copy number changes in regions of internal chromosome structure (Triglia *et al.*, 1991). Homologous recombination among subtelomeric repeat regions has also been proposed as a mechanism of large size changes (Corcoran *et al.*, 1988).

In addition to 14 nuclear chromosomes, two other genetic elements occur in *P. falciparum*, a tandemly repeated 6 kb extrachromosomal sequence and a 35 kb circle (Vaidya *et al.*, 1989; Aldritt *et al.*, 1989; Gardner *et al.*, 1991). The 6 kb tandemly repeated element contains an intriguing set of genes encoding interrupted rRNA fragments, cytochrome b and subunit I of cytochrome oxidase. Analysis of these genes has led to the suggestion that the 6 kb element may be mitochondrial in origin. In contrast the 35 kb circle contains coding regions for uninterrupted rRNA transcripts and RNA polymerase subunits. In its structure the 35 kb circle has been observed to resemble a chloroplast genome. Both the 6 kb repetitive element and the 35 kb circle probably reside in parasite cytoplasmic organelles; whether they are in the same or separate subcellular compartments remains to be determined.

Recombination rates among *P. falciparum* markers have been determined by genetic analysis in the laboratory. These experiments involved two laboratory crosses of parasite clones (3D7 × HB3 and HB3 × Dd2) performed according to the general strategy depicted in Figure 2 (Walliker *et al.*, 1987; Wellems *et al.*, 1990). Sexual stage parasites (gametocytes) are produced *in vitro* and fed artificially to anopheline mosquitoes where they develop into gametes and cross-fertilize. After zygote formation and oocyst development, sporozoites enter the mosquito salivary glands. Sporozoites are then inoculated into a splenectomized chimpanzee by allowing the mosquitoes to blood-feed on the abdomen of the animal. After about two weeks, a patent blood infection develops and individual progeny are cloned from the chimpanzee blood.

In analyzing inheritance data from a genetic cross, it is essential first to confirm that progeny used in the analysis are independent recombinants, i.e. they contain unique combinations of genetic material from the parent clones. This is ascertained through examination of restriction fragment length polymorphisms (RFLPs) on different chromosomes and may be complemented by isozyme profiles, monoclonal antibody reactivities and drug response assays. The process is facilitated by the use of DNA

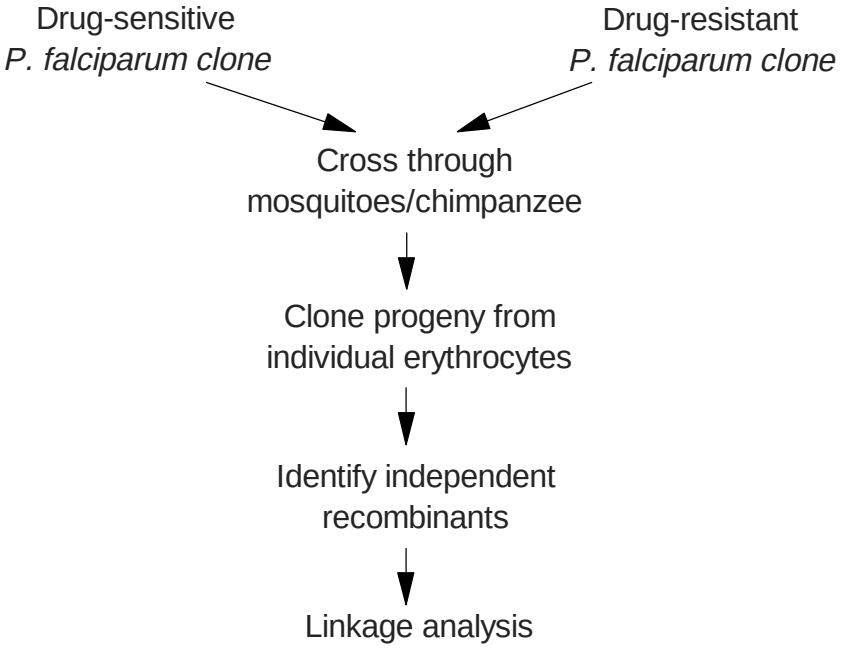

Figure 2. Outline of the strategy for producing a *Plasmodium falciparum* laboratory cross. Progeny used for genetic linkage analysis must be independent recombinants.

`fingerprint probes' that detect polymorphic repetitive elements on the chromosomes. Providing such probes are meiotically and mitotically stable, distinct banding patterns among the progeny reflect recombination and reassortment among chromosome segments during meiosis. One such probe, pC4.H32 (Dolan *et al.*, 1990), used in conjunction with selected single-locus markers to analyze over 100 progeny clones, identified 16 progeny that were unique recombinants from the HB3 × Dd2 *P. falciparum* cross (Wellems *et al.*, 1990). Redundant progeny appeared to result from expansion of select clonal populations *in vivo* and *in vitro.*

Chromosome assignments and inheritance data for nearly 90 RFLP markers have been reported for the 16 unique HB3 × Dd2 progeny (Walker-Jonah *et al.*, 1992). Markers were assigned to polymorphic loci on the 14 nuclear chromosomes. Unique patterns of chromosome recombination and reassortment were detected in each of the progeny, and indicated that progeny from cross-fertilization events were favoured over progeny from self-fertilization of either parent alone. The inheritance data showed that selective factors were absent from most chromosomes although some polymorphisms on chromosomes 2, 3, 12 and 13 suggested that certain parental linkage groups were favoured in the cross. Recombination frequencies measured on five chromosomes indicated an approximate map unit size of 15–30 kb per cM for *P. falciparum*, a value that compares to average rates of 5 kb per cM in yeast, 500 kb per cM in *Drosophila* and 1000 kb per cM in man (Lewin, 1990). Such

rates of course vary over the map and may be affected by recombination hotspots on the chromosomes.

Restriction fragment length polymorphisms that distinguish inheritance of the 6 kb tandemly repeated element and the 35 kb circle in the HB3 ⁄ Dd2 *P. falciparum* cross are under investigation (A.B. Vaidya, unpublished data).

POSITIONAL CLONING

Toward Identification of the Chloroquine Resistance Determinant

A focus of intensive effort has been the use of linkage analysis and physical mapping to identify the determinant of chloroquine resistance in *P. falciparum*. As a result of the worldwide spread of drug resistance (Figure 3) chloroquine, once the first choice as a safe, cheap and effective drug, is no longer reliable against *P. falciparum* malaria. A replacement drug having the favourable characteristics of chloroquine remains to be found, emphasizing the need to understand chloroquine action and resistance at the molecular level.

Experimental evidence indicates that chloroquine acts by inhibiting the parasite's ability to sequester ferriprotoporphyrin-IX, the toxic byproduct of haemoglobin digestion within the red blood cell (Slater and Cerami, 1992). Chloroquine-resistant (CqR) parasites counter the action of the chloroquine by expelling it: efflux rates *in vitro* were found to be 40 ⁄ to 50 ⁄ more rapid for resistant parasites than for drug-sensitive (CqS) parasites (Krogstad *et al.*, 1987). The mechanism of chloroquine rapid efflux is unknown, but drug release data from different parasite strains and the history of chloroquine resistance (Clyde, 1987a, 1987b) suggest that the efflux mechanism is the same in all CqR parasites.

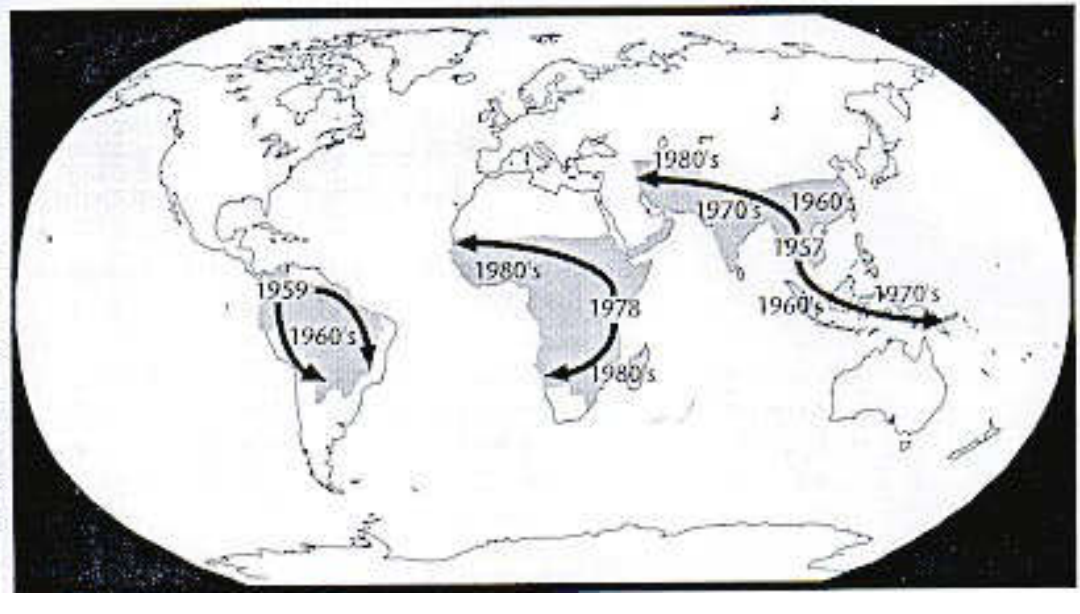

Figure 3. Map of the spread of chloroquine resistance (adapted from Clyde, 1987a, 1987b).

The HB3 × Dd2 *P. falciparum* cross, a cross designed to examine the genetics of chloroquine resistance, was used to determine inheritance of chloroquine response from the drug-resistant and drug-sensitive parents (Wellems *et al.*, 1990). Of the 16 independent recombinants recovered from the cross, eight exhibited the same absolute levels of resistance and rapid chloroquine efflux rates as the CqR parent (Dd2) while the other eight exhibited the chloroquine response and slow efflux rate of the CqS parent (HB3). No recombinant from the cross exhibited a non-parental chloroquine phenotype, indicating that a single genetic locus was responsible for the rapid chloroquine efflux mechanism.

Verapamil inhibits the rapid chloroquine efflux mechanism in resistant parasites, and because of this effect the efflux mechanism was thought to involve a gene analogous to a multiple drug resistance *(mdr)* gene in mammalian tumour cells (Martin *et al.*, 1987). This verapamil reversal phenomenon was confirmed in the HB3 ⟩ Dd2 cross: the CqR Dd2 parent and all eight CqR progeny exhibited inhibition of chloroquine release and decreased resistance when exposed to verapamil *in vitro,* while the CqS HB3 parent and the eight CqS progeny showed no such effects. Two *P. falciparum* genes with homology to mammalian *mdr* genes *(pfmdr*1, *pfmdr*2) were not, however, linked to chloroquine-resistance in the cross (Wellems *et al.*, 1990) and exceptions to an association of resistance with point mutations in *pfmdr*1 have been apparent in population surveys (Foote *et al.*, 1990). A mechanism other than one involving amplification or mutation of *pfmdr*1 or *pfmdr*2 must account for the rapid drug efflux from CqR parasites.

Chromosome inheritance patterns in the HB3 × Dd2 progeny have been examined by RFLP analysis to locate the determinant of chloroquine response (Wellems *et al.*, 1991). The search now has included markers to all 14 nuclear chromosomes, the 6 kb element and the 35 kb element. Results from this search show that a single segment of chromosome 7 is linked to chloroquine response. Long-range restriction mapping has been used to map the boundaries of this DNA segment from meiotic crossover events evident in progeny chromosomes. These boundaries place the chloroquine resistance determinant within a 200 kb segment of chromosome 7 (Figure 4).

While a 200 kb segment of DNA is a small fraction (0.7%) of the *P. falciparum* genome, it is still very large by sequencing standards and harbours many genes. Experiments are now under way to develop a detailed map of this 200 kb segment and clone the genes that lie within it. This has required that large (50 kb and more) overlapping segments of *P. falciparum* DNA be stably cloned for analysis. Recently, libraries of such segments have been generated in *S. cerevisiae* as yeast artificial chromosomes (YACs) (Triglia and Kemp, 1991; Lanzer *et al.*, 1992; de Bruin *et al.*, 1992)*. Screening of one of these libraries (Lanzer *et al.*, 1992; de Bruin *et al.*, 1992) produced a set of overlapping YACs that encompass the 200 kb segment (S. Wertheimer, M. Lanzer, D. de Bruin, T.E. Wellems and J.V. Ravetch, unpublished data). DNA from these YACs provides material to identify transcribed sequences from the chloroquine resistance region. Analysis of these sequences should lead to a candidate chloroquine resistance gene.

---

* Attempts to clone large *Plasmodium falciparum* fragments into *Escherichia coli* using cosmid and λ systems have not been successful, presumably due to the extremely high A+T content of *Plasmodium falciparum* DNA.

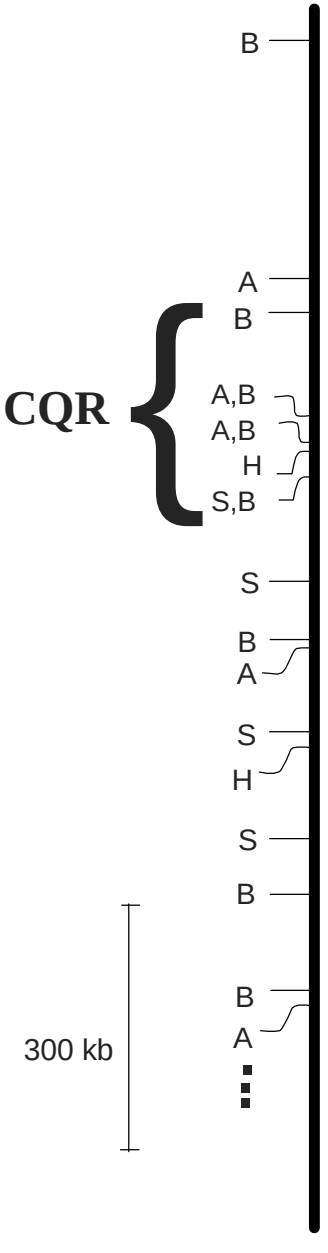

Figure 4. Long range restriction map of chromosome 7 and boundaries of the chloroquine response determinant as determined from the HB3 × Dd2 *Plasmodium falciparum* cross.

PROSPECTS

Advances in laboratory genetics, linkage analysis and physical mapping methods facilitate the identification of genetic loci in *P. falciparum*. The cloning in YACs of large chromosome

segments that span these loci is advancing efforts to identify specific genes. Characterization of genes that affect drug response, infectivity, host immunity and disease severity will lead to new strategies in the understanding and control of malaria.

ACKNOWLEDGMENTS

I thank the many colleagues and collaborators with whom I have had the privilege of working, including Lindsey Panton, Photini Sinnis, Stephen Dolan, Annie Walker-Jonah, Jacqueline Herrfeldt, David Peterson, Akhil Vaidya, Ilya Gluzman, Donald Krogstad, Sam Wertheimer and Jeffrey Ravetch.

REFERENCES

ALDRITT, S.M., JOSEPH, J.T. and WIRTH, D.F. 1989. Sequence identification of cytochrome b in *Plasmodium gallinaceum. Molecular Cell Biology* 9: 3614–3620.

de BRUIN, D., LANZER, M. and RAVETCH, J.V. 1992. Characterization of yeast artificial chromosomes from *Plasmodium falciparum:* construction of a stable, representative library and cloning of telomeric DNA fragments. *Genomics,* in press.

CARLE, G.F. and OLSON, M.V. 1984. Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucleic Acids Research* 12: 5647–5664.

CLYDE, D.F. 1987a. Variations in response of malaria parasites to drugs in Asia and Oceania. *La Medicina Tropicale nella Cooperazione allo Sviluppo* 3: 3–22.

CLYDE, D.F. 1987b. Genesis of chloroquine-resistant *Plasmodium falciparum* in the American region. *La Medicina Tropicale nella Cooperazione allo Sviluppo* 3: 41–44.

CORCORAN, L.M., THOMPSON, J.K., WALLIKER, D. and KEMP, D.J. 1988. Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum. Cell* 53: 807–813.

DOLAN, S.A., ADAM, R.D. and WELLEMS, T.E. 1993. Chromosome mapping methods for parasitic protozoa. In: Hyde, J.E., ed. *Molecular Biology Protocols.* Clifton, New Jersey: Humana Press, Inc., in press.

DOLAN, S.A., MILLER, L.H. and WELLEMS, T.E. 1990. Evidence for a switching mechanism in the invasion of erythrocytes by *Plasmodium falciparum. Journal of Clinical Investigation* 86: 618–624.

FOOTE, S.J., KYLE, D.E., MARTIN, R.K., ODUOLA, A.M., FORSYTH, K., KEMP, D.J. and COWMAN, A.F. 1990. Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in *Plasmodium falciparum* [see comments]. *Nature* 345: 255–258.

GARDNER, M.J., WILLIAMSON, D.H. and WILSON, R.J.M. 1991. A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Molecular and Biochemical Parasitology* 44: 115–124.

GU, H., INSELBURG, J.W., BZIK, D.J. and LI, W.B. 1990. *Plasmodium falciparum:* analysis of chromosomes separated by contour-clamped homogenous electric fields. *Experimental Parasitology* 71: 189–198.

KEMP, D.J., THOMPSON, J.K., WALLIKER, D. and CORCORAN, L.M. 1987. Molecular karyotype of *Plasmodium falciparum:* conserved linkage groups and expendable histidine-rich protein genes. *Proceedings of the National Academy of Sciences of the USA* 84: 7672–7676.

KROGSTAD, D.J., GLUZMAN, I.Y., KYLE, D.E., ODUOLA, A.M., MARTIN, S.K., MILHOUS, W.K. and SCHLESINGER, P.H. 1987. Efflux of chloroquine from *Plasmodium falciparum:* mechanism of chloroquine resistance. *Science* 238: 1283–1285.

LANZER, M., de BRUIN, D. and RAVETCH, J.V. 1992. Transcription mapping of a 100 kb locus of *P. falciparum* identifies an intergenic region in which transcription terminates and reinitiates. *EMBO Journal* 11: 1949–1955.

LEWIN, B. 1990. *Genes* IV. Cambridge: Cell Press, pp. 89–90.

MARTIN, S.K., ODUOLA, A.M. and MILHOUS, W.K. 1987. Reversal of chloroquine resistance in *Plasmodium falciparum* by verapamil. *Science* 235: 899–901.

POLLACK, Y., KATZEN, A.L., SPIRA, D.T. and GOLENSER, J. 1982. The genome of *Plasmodium falciparum.* I. DNA base composition. *Nucleic Acids Research* 10: 539–546.

POLOGE, L.G. and RAVETCH, J.V. 1986. A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. *Nature* 322: 474–477.

POLOGE, L.G. and RAVETCH, J.V. 1988. Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* 55: 869–874.

PRENSIER, G. and SLOMIANNY, C. 1986. The karyotype of *Plasmodium falciparum* determined by ultrastructural serial sectioning and 3D reconstruction. *Journal of Parasitology* 72: 731–736.

SCHWARTZ, D.C. and CANTOR, C.R. 1984. Separation of yeast chromosome-sized DNAs by pulsed-field gradient gel electrophoresis. *Cell* 37: 67–75.

SHIRLEY, M.W., BIGGS, B.A., FORSYTH, K.P., BROWN, H.J., THOMPSON, J.K., BROWN, G.V. and KEMP, D.J. 1990. Chromosome 9 from independent clones and isolates of *Plasmodium falciparum* undergoes subtelomeric deletions with similar breakpoints *in vitro*. *Molecular and Biochemical Parasitology* 40: 137–145.

SINNIS, P. and WELLEMS, T.E. 1988. Long-range restriction maps of *Plasmodium falciparum* chromosomes: crossing over and size variation among geographically distant isolates. *Genomics* 3: 287–295.

SLATER, A.F.G. and CERAMI, A. 1992. Inhibition by chloroquine of a novel haem polymerase enzyme activity in malaria trophozoites. *Nature* 355: 167–169.

TRIGLIA, T., FOOTE, S.J., KEMP, D.J. and COWMAN, A.F. 1991. Amplification of the multi-drug resistance gene *pfmdr*1 in *Plasmodium falciparum* has arisen as multiple independent events. *Molecular Cell Biology* 11: 5244–5250.

TRIGLIA, T. and KEMP, D.J. 1991. Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Molecular and Biochemical Parasitology* 44: 207–212.

TRIGLIA, T., WELLEMS, T.E. and KEMP, D.J. 1992. Towards a high-resolution map of the *Plasmodium falciparum* genome. *Parasitology Today* 8: 225–229.

VAIDYA, A.B., AKELLA, R. and SUPLICK, K. 1989. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Molecular and Biochemical Parasitology* 35: 97–108.

WALKER-JONAH, A., DOLAN, S.A., GWADZ, R.W., PANTON, L.J. and WELLEMS, T.E. 1992. An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favored linkage groups in a genetic cross. *Molecular and Biochemical Parasitology* 51: 313–320.

WALLIKER, D., QUAKYI, I.A., WELLEMS, T.E., McCUTCHAN, T.F., SZARFMAN, A., LONDON, W.T., CORCORAN, L.M., BURKOT, T.R. and CARTER, R. 1987. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* 236: 1661–1666.

WELLEMS, T.E., PANTON, L.J., GLUZMAN, I.Y., do ROSARIO, V.E., GWADZ, R.W., WALKER-JONAH, A. and KROGSTAD, D.J. 1990. Chloroquine resistance not linked to *mdr*-like genes in a *Plasmodium falciparum* cross. *Nature* 345: 253–255.

WELLEMS, T.E., WALKER-JONAH, A. and PANTON, L.J. 1991. Genetic mapping of the chloroquine-resistance locus on *Plasmodium falciparum* chromosome 7. *Proceedings of the National Academy of Sciences of the USA* 88: 3382–3386.

WELLEMS, T.E., WALLIKER, D., SMITH, C.L., do ROSARIO, V.E., MALOY, W.L., HOWARD, R.J., CARTER, R. and McCUTCHAN, T.F. 1987. A histidine-rich protein gene marks a linkage group favored strongly in a genetic cross of *Plasmodium falciparum*. *Cell* 49: 633–642.

Theileria parva: current status

V. Nene, R. Bishop, T.T. Dolan, D. McKeever, O.K. ole-MoiYoi, S.P. Morzaria, A.J. Musoke, M.K. Shaw, P. Toye and A.S. Young

Theileria parva is a tick-transmitted intracellular protozoan parasite of cattle which usually causes an acute and often fatal lymphoproliferative disease called East Coast fever (ECF). Approximately 24 million cattle in 12 countries in eastern, central and southern Africa are at risk. The high rate of morbidity and mortality of cattle introduced into ECF endemic areas severely impedes improvement of the cattle industry. The epidemiology of the disease is confused by the presence of *Theileria* species which are nonpathogenic to cattle and the presence of a *T. parva* reservoir in buffalo complicates the development of ECF control measures. Furthermore there is polymorphism in the clinical aspects of the disease and there is antigenic diversity within and between different parasite stocks.

Cattle that recover from a natural infection or those that are immunized by a simultaneous inoculation of a lethal dose of cryopreserved sporozoites and a long-acting oxytetracycline develop a long-lasting but parasite strain-specific immunity to reinfection. This immunity is thought to be dependent on the generation of class I MHC restricted cytotoxic T cells (CTLs) which are specific for schizont-infected cells. None of the available MAbs or DNA probes correlate with cross-immunity data and the nature of the parasite antigenic diversity remains undefined. A role for an immune response to sporozoites in mediating protection has been demonstrated by immunizing cattle with a recombinant derivative of the major sporozoite surface antigen p67. In contrast to the antigens that induce cell-mediated immune responses, the sequence of the p67 molecule is highly conserved.

The parasite genome is remarkably small, being about 10^7 bp in size with an estimated G+C content of 31%. The genome consists of four Mbp size chromosomes and at least two small extrachromosomal DNA elements. Several parasite genes which code for antigens, enzymes, rRNA and mitochondrial components have been characterized. In addition, telomeric and repetitive DNA sequences have been analyzed. All the genes have been assigned to specific locations on the genome as based on a physical linkage map. The DNA sequence data generated accounts for about 0.6% of the *T. parva* genome and preliminary information regarding gene structures have been deduced from this database.

INTRODUCTION

Bovine theileriosis is a complex disease syndrome caused by the tick-transmitted protozoan parasite *Theileria parva* (reviewed by Norval *et al.*, 1992). The usually acute and often fatal lymphoproliferative disease is characterized by pyrexia, swelling of the lymph glands and a panleucopenia (reviewed by Irvin and Morrison, 1987). The parasite affects approximately 24 million cattle in 12 countries in eastern, central and southern Africa. In 1989, an estimated US$ 168 million was lost as a direct consequence of the disease (Mukhebi *et al.*, 1992). The high rate of morbidity and mortality of cattle introduced into areas where the disease is endemic impedes improvement of the productivity of the cattle industry in a large part of sub-Saharan Africa.

The clinical behaviour of different parasite stocks has in the past led to sub-speciation of the parasite. Under this nomenclature *T. p. parva* is the cause of East Coast fever (ECF)

and *T. p. bovis*, a parasite from Zimbabwe, is the cause of January disease, so called because of its seasonal occurrence. Both sub-species are maintained in cattle whereas *T. p. lawrencei* is a parasite of buffalo which, when transmitted to cattle, causes Corridor disease. There is no zoological justification for sub-species status and the parasites are now referred to as either cattle- or buffalo-derived stocks of *T. parva* (Anon., 1989). However, different stocks of *T. parva* clearly exhibit polymorphism in the clinical disease, parasitological parameters and antigenic type. Factors such as virulence and low or high levels of parasitosis are phenotypic characteristics of parasite stocks, which invariably contain several strains, and a number of these characteristics may depend on host-parasite interaction. Analysis of homogeneous parasite populations is required to unambiguously assign a `trait' to the parasite since both the host and parasite are genotypically polymorphic.

The wild life reservoir of *T. parva* complicates the development of control measures since cattle and buffalo often share grazing and the epidemiology of the disease can be confused by the presence of other *Theileria* species which are non-pathogenic to cattle. The most extensively used disease control measure is the frequent dipping of cattle to limit tick infestation (reviewed by Norval *et al.*, 1992). This has the added advantage of controlling other tick-borne diseases. However, acaricides are toxic compounds and ticks can develop resistance to them. Drugs can be used in chemotherapy (McHardy, 1984), but these are expensive, not always available and usually have to be used fairly early during the course of an infection to be completely effective. It is also possible to immunize cattle against the disease by the simultaneous inoculation of a lethal dose of cryopreserved sporozoites together with a long-acting oxytetracycline (reviewed by Morzaria and Nene, 1990). Cattle undergo a mild disease reaction from which they recover and then exhibit a long-lived immunity. Problems associated with this infection-and-treatment method of immunization include the use of live parasites and drug treatment. A more serious limitation is that immunity is often parasite stock-specific as challenge of cattle with a heterologous parasite stock can result in breakthrough infections. Vaccine potential of the infection-and-treatment method can be increased by immunizing with a cocktail of sporozoites from stocks of different cross-immunity profiles. Since none of the available *in vitro* parasite characterization methods correlate with cross-immunity data the extent of parasite antigenic diversity is difficult to predict and the success of this approach on a wide scale remains to be determined. Hence, alternate control measures are required and subunit vaccines offer a potentially cheap and efficient method of parasite control.

PARASITE LIFE CYCLE

Sporogony takes place in type III acini in the salivary glands of the main tick vector, *Rhipicephalus appendiculatus* (Fawcett *et al.*, 1985). Sporogony initiates when the tick attaches to its host and mature sporozoite release peaks four to five days later. Approximately 40,000–50,000 sporozoites develop within one infected acinus. This represents a minimum parasite challenge and it can be lethal. Sporozoites enter bovine lymphocytes and develop into a multinucleate stage called the schizont. The events of the entry process and the possible involvement of MHC class I molecules in entry has been recently

described in detail (Shaw *et al.,* 1991). Parasite development is highly restricted in the mammalian host range and there is no small animal model for the parasite.

The schizont lies free in the host cytoplasm and it induces the host cell to proliferate. The host cell and parasite divide in synchrony which results in a rapid increase in schizont parasitaemia. Since the proliferative phenotype of infected cells can be reversed by schizonticides, the cells are not immortalized. *In vitro,* it is possible to infect and `transform' B cells, $CD4^+$ and $CD8^+$ T cells and T cells which lack the CD4 and CD8 markers (Baldwin *et al.,* 1988) but which express γδ T cell receptor genes. The molecular mechanisms that cause the host cell to proliferate are unknown but casein kinase II in the host cell and the parasite are candidate enzymes involved in the signalling of cell growth (ole-MoiYoi *et al.,* 1992).

Some schizonts undergo merogony to produce merozoites which are released by host cell rupture. These merozoites invade red blood cells where they develop into piroplasms, the tick-infective stage. There is little if any multiplication of the piroplasm stage of the parasite and the primary cause of pathology is the schizont-infected cell. Cattle that recover from infection become carriers of the parasite but what maintains this status is unclear.

Morphologically different forms of gametes are only seen in the tick gut and these fuse to form a zygote which invades the gut epithelium. The zygote differentiates into a kinete which invades the tick salivary gland to complete the life cycle. Recent experiments have shown that genetic exchange can take place during this event. However, whether antigenic diversity is generated during the sexual cycle remains to be determined. Transmission of the parasite is trans-stadial and, since the tick is a three-host tick, infections are picked up by larvae and nymphs. Only the next instar, namely nymphs and adults, are able to transmit the parasite.

SCHIZONT ANTIGEN-BASED VACCINE

Cattle that recover from a natural infection or those that are immunized develop a long-lasting but parasite strain-specific immunity to reinfection. This immunity is thought to be dependent on the generation of class I MHC restricted cytotoxic T cells (CTLs) which are specific for schizont-infected cells (reviewed by Morrison and Goddeeris, 1990). The specificity of the CTL response is known to be influenced by the bovine MHC and the immunizing parasite stock and these findings could complicate the development of an anti-schizont vaccine. The relative importance of this type of response in very young animals has been recently questioned since preliminary studies have not been able to detect a cytotoxic T cell response in immune calves.

Two strategies to identify schizont epitopes that are the targets of the CTL response are being developed. The first involves the isolation and characterization of parasite peptides associated with bovine class I MHC gene products. The second strategy involves a screen of COS cells, permanently expressing bovine MHC gene(s) and transiently expressing randomly isolated schizont cDNAs in a direct killing assay, using CTLs as probes. The identification of a schizont antigen carrying a CTL epitope will allow a number of important technical and theoretical issues to be investigated.

A role for a humoral response to the sporozoite stage in mediating immunity has been hypothesized based on the finding of sporozoite neutralizing antibodies in sera taken from cattle in endemic areas. Sporozoite antigens that are the targets of these antibodies are, therefore, potential components of subunit vaccines (reviewed by Musoke and Nene, 1990). A significant rationale in the development of an anti-sporozoite vaccine is that the severity of the disease appears to be sporozoite dose-dependent (Jarrett *et al.,* 1969). Hence, a vaccine does not have to induce a sterile immunity.

Monoclonal antibodies (MAbs) that neutralize sporozoite infectivity *in vitro* primarily bind to a 67 kDa stage specific antigen termed p67. Antisera to this antigen will neutralize sporozoite infectivity from a range of different stocks suggesting that p67 is a conserved antigen. The gene coding for p67 of *T. parva* (Muguga) has been cloned and characterized (Nene *et al.,* 1992). The gene is present in a single copy and it is split into two exons by an intron that is 29 bp long. The open reading frame codes for a protein consisting of 709 amino acid residues which has a signal sequence at the N-terminus, a hydrophobic C-terminal region and it does not contain short tandem amino acid sequence repeats.

Recombinant p67 has been expressed in both prokaryotic and eukaryotic cells. Nine cattle immunized with a semi-purified preparation of a bacterial fusion protein with saponin as an adjuvant developed high levels of anti-p67 antisera. When given a syringe challenge of homologous sporozoites, a range of responses was observed. Some cattle did not react, some developed mild disease and recovered while others developed severe disease and were indistinguishable from controls (Musoke *et al.,* 1992).

The protection data has been validated in a larger number of cattle and overall 19 of 32 immunized cattle have been protected against challenge. Interestingly, a similar level of protection was observed with a heterologous sporozoite challenge indicating that p67 may be able to induce a broad spectrum immunity. Unfortunately, antibody titres to p67 were not predictive of protection and *in vitro* correlates with immunity remain to be defined. In a more recent experiment, five cattle that received purified recombinant antigen and a different adjuvant were immune to sporozoite challenge. This result is extremely promising and the immunization regime is being optimized before a tick challenge is undertaken. Preliminary data suggest that sporozoite release from the tick occurs as a trickle rather than as a burst and this may favour an anti-sporozoite vaccine.

As a variation in amino acid sequence in the p67 protein may complicate vaccine development, the genes coding for this protein from a variety of different parasite stocks have been sequenced. The data indicates that there is very little if any sequence variation in the gene from cattle-derived parasites whereas the predicted gene product of buffalo-derived parasites contains several amino acid differences. Hence, it is reasonable to presume that extensive cross-immunity trials with cattle-derived parasites do not have to be undertaken but the capacity of recombinant p67 to induce immunity to buffalo-derived parasites needs to be established.

In considering vaccine development a mild disease reaction in p67-immunized cattle has the advantage that a cell-mediated immune response to the schizont stage of the parasite is developed and such cattle would maintain immunity against both the infective and the pathogenic stage of the parasite. Differences in the local or regional

epidemiology of the disease may dictate that different vaccine strategies be developed. For example, because of the wildlife reservoir, disruption of the endemic stability of the disease by a vaccine could have disastrous consequences if control measures are not sustained.

THE PARASITE GENOME

All life cycle stages of the parasite except the kinete, and presumably the zygote, are haploid. The parasite genome is remarkably small being about 10^7 bp in size with an estimated G+C content of 31% (Allsopp and Allsopp, 1988). The genome consists of four Mbp size chromosomes (Morzaria and Young, 1992) and contains at least two small extrachromosomal DNA elements. Several parasite genes which code for antigens, enzymes, rRNA and mitochondrial components have been characterized. In addition, telomeric and repetitive DNA sequences have been analyzed. All the genes have been assigned to specific locations on the genome as based on a physical linkage map. The DNA sequence data generated accounts for about 0.6% of the *T. parva* genome.

Preliminary information regarding gene structure and *cis* spliced introns has been deduced from a database. The characteristics of the introns are different to higher eukaryotic introns and in one example COS cells were not able to express a protein from a gene containing a 29 bp intron. The majority of the genes that have been analyzed are present as a single copy and protein antigens rarely contain short tandemly repeated amino acid sequences. Regulation of gene expression can take place at the transcriptional and post-transcriptional level and there is no evidence, so far, of co-transcription of genes. One approach being pursued to identify schizont antigens bearing CTL epitopes is to screen randomly isolated cDNA. Analysis of these clones should yield valuable information on the repertoire of genes expressed in the pathogenic stage of the parasite and help to gain an understanding of the cellular and molecular biology of *T. parva*.

REFERENCES

ALLSOPP, B.A. and ALLSOPP, N.T.E.P. 1988. *Theileria parva:* genomic studies reveal intra-specific sequence diversity. *Molecular Biochemical Parasitology* 28: 77–84.

ANON. 1989. Nomenclature in *Theileria.* In: Dolan, T.T., ed. *Theileriosis in Eastern, Central and Southern Africa. Proceedings of a Workshop on East Coast Fever Immunization Held in Lilongwe, Malawi, 20–22 September, 1988.* Nairobi: International Laboratory for Research on Animal Diseases, pp. 182–186.

BALDWIN, C.L., BLACK, S.J., BROWN, W.C., CONRAD, P.A., GODDEERIS, B.M., KINUTHIA, S.W., LALOR, P.A., McHUGH, N.P., MORRISON, W.I., MORZARIA, S.P., NASSENS, J. and NEWSON, J. 1988. Bovine T cells, B cells and null cells are transformed by the protozoan parasite *Theileria parva. Infection and Immunity* 56: 426–467.

FAWCETT, D.W., YOUNG, A.S. and LEITCH, B.L. 1985. Sporogony in *Theileria* (Apicomplexa: Piroplasmida). A comparative ultrastructural study. *Journal of Submicroscopic Cytology* 17: 299–314.

IRVIN, A.D. and MORRISON, W.I. 1987. Immunopathology, immunology and immunoprophylaxis of *Theileria* infections. In: Soulsby, E.J.L., ed. *Immune Responses in Parasitic Infections: Immunology, Immunopathology and Immunoprophylaxis.* Boca Raton, Florida: CRC Press, pp. 223–274.

JARRETT, W.F.H., CRIGHTON, G.W. and PIRIE, H.W. 1969. *Theileria parva:* kinetics of replication. *Experimental Parasitology* 24: 9–25.

McHARDY, N. 1984. Recent advances in the chemotherapy of theileriosis. In: Riemann, H.P. and Burridge, M.J., eds. *Impact of Diseases on Livestock Production in the Tropics*. Amsterdam: Elsevier, pp. 179–192.

ole-MOIYOI, O.K., SUGIMOTO, C., CONRAD, P.A. and MACKLIN, M.D. 1992. Cloning and characterization of the casein kinase II, a subunit gene from the lymphocyte-transforming intracellular protozoan parasite *Theileria parva*. *Biochemistry* 31: 6193–6202.

MORRISON, W.I. and GODDEERIS, B.M. 1990. Cytotoxic T cells and immunity to *Theileria parva* in cattle. *Current Topics in Microbiology and Immunology* 155: 70–93.

MORZARIA, S.P. and NENE, V. 1990. Bovine theileriosis: progress in immunization methods. *International Journal of Animal Science* 5: 1–14.

MORZARIA, S.P. and YOUNG, J.R. 1992. Restriction mapping of the genome of the protozoan parasite *Theileria parva*. *Proceedings of the National Academy of Sciences of the USA* 89: 5241–5245.

MUKHEBI, A.W., PERRY, B.D. and KRUSKA, R. 1992. Estimated economics of theileriosis control in Africa. *Preventive Veterinary Medicine* 12: 73–85.

MUSOKE, A., MORZARIA, S.P., NKONGE, C., JONES, E. and NENE, V. 1992. A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. *Proceedings of the National Academy of Sciences of the USA* 89: 514–518.

MUSOKE, A.J. and NENE, V. 1990. Development of recombinant antigen vaccines for the control of theileriosis. *Parassitologia* 32: 73–85.

NENE, V., IAMS, K.P., GOBRIGHT, E. and Musoke, A.J. 1992. Characterization of the gene encoding a candidate vaccine antigen of *Theileria parva* sporozoites. *Molecular Biochemical Parasitology* 51: 17–28.

NORVAL, R.A.I., PERRY, B.D. and YOUNG, A.S. 1992. *The Epidemiology of Theileriosis in Africa*. London: Academic Press, 481 pp.

SHAW, M.K., TILNEY, L.G. and MUSOKE, A.J. 1991. The entry of *Theileria parva* sporozoites into bovine lymphocytes: evidence for MHC class I involvement. *Journal of Cell Biology* 113: 87–101.

Genomic polymorphisms in *Theileria parva*

R. Bishop, H. Baylis, B. Allsopp, P. Toye, V. Nene, T. Dolan, P. Spooner, M. Kibe and S. Morzaria

A *Theileria parva* repetitive sequence designated *Tpr* has been extensively characterized. Most *T. parva* stocks examined can be distinguished by restriction fragment length polymorphisms (RFLPs) within the *Tpr* locus. DNA sequencing of Tpr clones reveals open reading frames (ORFs) with a novel organization. Synthetic oligonucleotides derived from *Tpr* sequences can distinguish selected stocks in a positive/negative fashion.

Ribosomal RNA genes are polymorphic between *T. parva* stocks particularly in the transcribed spacer region. Telomeric DNA sequences are polymorphic between *T. parva* stocks and most telomeric *Sfi*I fragments are size polymorphic. The majority of telomeres can be distinguished using a cloned *T. parva* telomeric sequence by *Eco*RI RFLPs.

Polymorphism has been examined in the genes encoding for two *T. parva* antigens. The p67 sporozoite antigen gene is highly conserved among cattle-derived *T. parva*, but p67 from buffalo-derived parasites has a 129 bp insertion potentially encoding an additional 43 amino acids. The central section of the gene encoding the polymorphic immunodominant molecule (PIM), which is found on the surface of *T. parva* schizonts and sporozoites, contains complex repeated sequences which are polymorphic between different *T. parva* stocks. Polymerase chain reaction (PCR) amplification of DNA using single arbitrary primers (decamers) has recently been used to detect *T. parva* genomic polymorphisms.

INTRODUCTION

Polymorphic sequences within the *Theileria parva* genome which have been characterized include repetitive DNAs (in particular a tandemly repeated sequence designated *Tpr),* ribosomal RNA genes, telomeric DNA sequences and the genes encoding the p67 and polymorphic immunodominant (PIM) antigens. Attempts have been made to look for polymorphic minisatellites and microsatellites in *T. parva.* Polymerase chain reaction (PCR) amplification of DNA using arbitrary primers (AP-PCR) has recently been shown to be a powerful method for detecting genomic polymorphisms in *T. parva.*

REPETITIVE DNA SEQUENCES

A major repetitive DNA sequence designated *Tpr*1 (Allsopp *et al.,* 1989), is highly polymorphic in the *T. parva* genome and has been extensively used for characterization of *T. parva* stocks using RFLP analysis (Conrad *et al.*, 1987; Allsopp and Allsopp, 1988).

Sequence analysis of an 8.0 kb EMBL 3 phage clone containing *Tpr* DNA from the *T. parva* Muguga stock reveals the existence of long open reading frames (ORFs) which are partially repeated and tandemly arranged (Baylis *et al.,* 1991). The repeats are approximately 90% homologous within the *T. parva* Muguga genome and conserved as ORFs, although many of the ORFs do not have an in-frame ATG codon, for initiation of

translation, close to the 5' end. Hybridization of *Tpr* probes to pulsed-field gel electrophoresis (PFGE) blots shows that the *Tpr* genes are situated in a limited region of the *T. parva* Muguga genome on two adjacent *Sfi*I fragments on chromosome 3 (Morzaria and Young, 1992). When *Tpr* DNA sequences from different *T. parva* stocks are compared, the 5' ends of the ORFs are sometimes not homologous, whereas the 3' ends are 70–80% homologous at the DNA level. Transcripts homologous to *Tpr* sequences have been detected in the tick-infective piroplasm stage of *T. parva*, but not in the schizont stage (Baylis *et al.*, 1991). *Tpr*-homologous piroplasm cDNA clones from a single infected animal are derived from at least four loci and show considerable sequence heterogeneity. The *Tpr* ORFs have the appearance of a system designed to generate diversity and have features reminiscent of vertebrate immunoglobulin or troponin T genes. However, it has so far proved impossible to demonstrate the existence of proteins corresponding to the genomic ORFs.

A practical consequence of *Tpr* sequence diversity is that synthetic oligonucleotides derived from *Tpr* sequences can be used to distinguish selected *T. parva* stocks or groups of stocks in a positive/negative fashion (Allsopp *et al.*, 1989; Bishop *et al.*, 1992). This is an important element of the strategy employed for isolating recombinant *T. parva* parasites. A second multicopy, polymorphic, transcribed sequence has recently been discovered in the *T. parva* genome. This is situated on a single *Sfi*I fragment on chromosome 4. Unlike the *Tpr* sequences, these repeats are transcribed in both schizonts and piroplasms. A third repeated sequence (Conrad *et al.*, 1987), which has not been extensively characterized, is situated primarily on a single *Sfi*I fragment on chromosome 2.

A general property of the repeated sequences so far identified in *T. parva* is that they are confined to specific regions and not dispersed in the genome.

RIBOSOMAL RNA GENES

Hybridization of *Eco*RI-digested *T. parva* DNA with a *T. parva* small subunit ribosomal RNA (SSU, rRNA) gene probe separates all *T. parva* stocks so far examined into two groups, exhibiting either one or two fragments hybridizing to the probe. It has been shown that there are two ribosomal RNA coding units in *T. parva*, situated on chromosomes 1 and 3. There is evidence to suggest that in some *T. parva* stocks the two transcription units are similar to one another, whereas in other stocks they are different. It is not yet known if there are sequence differences in the SSU rRNA coding genes between the two units, as has been observed in *Plasmodium berghei* (Gunderson *et al.*, 1987). There are considerable differences, both between different *T. parva* stocks and within a single parasite genome, in the transcribed spacer sequences between the large and small ribosomal subunits, involving large insertions or deletions.

TELOMERIC DNA SEQUENCES

It is known that telomere-associated DNA sequences are often major regions of genome instability in protozoa, for example deletion in subtelomeric repetitive sequences is responsible for generating chromosome size polymorphisms in *Plasmodium* (Corcoran *et*

al., 1988; Biggs *et al.,* 1989). In *T. parva* initial studies involving hybridization of a synthetic oligonucleotide derived from a published *P. berghei* telomeric sequence (Ponzi *et al.,* 1985) to blots of *Sfi*I-digested *T. parva* DNA run on pulsed-field gels demonstrated that the majority of size-polymorphic *Sfi*I fragments were telomerically located (Morzaria *et al.,* 1990). A *T. parva* telomeric sequence has been isolated using the *P. berghei* oligonucleotide as a probe. Using this cloned sequence most *T. parva* telomeres can be distinguished in *Eco*RI-digested DNA and many *Eco*RI fragments are size-polymorphic between different stocks. The probe can therefore be employed to detect independent assortment of chromosomes in recombinant *T. parva* parasites.

Analysis of telomere-associated DNA sequences situated up to 15 kilobases upstream of the simple telomeric repeats, cloned in the λ vector EMBL 3, does not reveal the presence of large blocks of repetitive DNA sequences. Homology with other telomeres does not extend beyond 4–5 kilobases from the simple repeats. There are however RFLPs between different *T. parva* stocks in the telomere-associated DNA. The cloned *T. parva* subtelomeric DNA sequences are derived from the left hand end of chromosome 1.

ANTIGEN GENES

In view of the importance of the gene encoding the p67 sporozoite surface antigen (Nene *et al.,* 1992) as a candidate vaccine antigen, its complete sequence has been determined for eight *T. parva* stocks. Five cattle-derived *T. parva* stocks did not exhibit significant variation at the DNA or amino acid levels. The p67 genes of the *T. parva* Muguga and Marikebuni stocks, which were analyzed by direct sequencing of PCR products, were found to be identical. The homogeneity of the gene in these two stocks is surprising since they are known to differ in many other properties, including the strain specificity of schizont immunity (Irvin *et al.,* 1983; Mutugi *et al.,* 1989). The sequence of the p67 gene from three Kenyan *T. parva* buffalo-derived parasites exhibits 64 nucleotide differences relative to the gene from the cattle-derived parasites, 37 of which result in amino acid differences. The gene from the buffalo-derived parasites also contains a 129 base pair insertion potentially encoding an additional 43 amino acids inserted in-frame between amino acids 303 and 304 of the cattle-derived p67 protein. The insertion is flanked by a six nucleotide direct repeat, a feature typical of sequences which have been inserted into DNA. The p67 gene is situated on chromosome 3 (Morzaria and Young, 1992).

The polymorphic immunodominant antigen (PIM) is an antigen recognized by all mouse monoclonal antibodies (MAbs) raised against *T. parva* schizont-infected cells, which have been tested on Western blots (Toye *et al.,* 1991). It is also the predominant antigen recognized by sera from infected cattle on Western blots. More recently the PIM antigen has also been shown to be expressed in the sporozoite stage of *T. parva*. On Western blots the antigen is size polymorphic between different parasite stocks or, in the case of *T. parva* Marikebuni, between clones within a stock (Shapiro *et al.,* 1987; Toye *et al.,* 1991). A panel of anti-PIM MAbs can also be used to distinguish different *T. parva* stocks on the basis of +/– reactions in an immunofluorescent antibody test (IFAT) (Minami *et al.,* 1983).

The gene encoding PIM proved refractory to cloning using bacterial expression systems but the *T. parva* Muguga PIM gene has recently been cloned by using one of the MAbs to

screen a library expressed in the eukaryotic COS cell system. The PIM gene has also been isolated from the buffalo-derived 7014 stock using the *T. parva* Muguga PIM gene and the genes from both stocks have been sequenced. The PIM gene is situated on chromosome 1. The sequence data reveals the 5' and 3' ends of the genes to be relatively conserved, approximately 96–97% homologous at the nucleotide level. The central portion of the PIM gene contains a complex series of polymorphic repeated sequences of three different types. The relative size of the polymorphic central region of the gene between the two stocks corresponds to the size difference observed in the PIM protein in that it is larger in buffalo-derived 7014 than in *T. parva* Muguga. Deletion constructs have allowed mapping of the epitopes detected by the MAbs to a region in the 3' end of the gene which is outside the central region of the molecule which contains the polymorphic repeated sequences. This suggests that the size polymorphism in the PIM antigen and the epitope polymorphisms detected using the IFAT test are located in independent regions of the molecule. This result also contrasts with the situation in many *Plasmodium* antigens in which the immunodominant epitopes are usually located in regions of repeated amino acids (reviewed in Nussenzweig and Nussenzweig, 1985).

MINISATELLITES, MICROSATELLITES AND PCR AMPLIFICATION USING ARBITRARY PRIMERS

A probe derived from the region of M13 phage which is known to contain homology to sequences present as variable number tandem repeats or minisatellites in vertebrate genomes (Vassart *et al.*, 1987) hybridizes to *T. parva* DNA. The sequences involved display only limited polymorphism between *T. parva* stocks and are primarily confined to a single *Sfi*I fragment situated on chromosome 2.

Length polymorphisms in simple dinucleotide or trinucleotide repeats, detected using PCR amplification, are useful polymorphic markers for genome analysis in many organisms (Tautz, 1989). Two oligonucleotides $(GT)_{10}$ and $(GA)_{10}$ derived from the most common microsatellites present in vertebrate genomes do not hybridize to *T. parva* DNA. Two short microsatellites (four copies of a dinucleotide repeat) have been observed in cloned *T. parva* DNA sequences but it is not known whether these are polymorphic between *T. parva* stocks.

PCR amplification of DNA using single arbitrary sequence primers (AP-PCR) has recently been shown to be a useful method of detecting polymorphisms between organisms (Williams *et al.*, 1990; Welsh and McLelland, 1990). Selected decamer primers have been found to generate PCR products polymorphic between different *T. parva* stocks and clones. When gel-purified and hybridized to Southern blots of *T. parva* DNA, individual polymorphic PCR products detect restriction fragments which are size-polymorphic between different *T. parva* clones. Since this technique is likely to detect polymorphisms which are dispersed in the *T. parva* genome, and which may sometimes be within protein coding genes, AP-PCR represents a powerful method for detecting and analyzing genomic polymorphisms in *T. parva*.

Extensive polymorphism has been discovered in the *T. parva* genome but there is no evidence that any of the genomic polymorphisms so far characterized correlate with

phenotypic parameters of interest such as the strain-specifity of schizont immunity, parasite virulence or infectivity to the vertebrate and invertebrate hosts. A high proportion of the polymorphic markers so far available sample variation from multiple loci or at multiple sites within a complex locus. The available markers are heavily biased towards chromosomes 1 and 3. For the purposes of further mapping the *T. parva* genome there is a requirement for a series of simple allelic markers which are dispersed throughout the *T. parva* genome. In the context of mapping, isolation of a dispersed repetitive DNA sequence in *T. parva* would also be useful for ordering clones from large fragment libraries constructed in P1 or YAC vectors.

REFERENCES

ALLSOPP, B.A. and ALLSOPP, M.T.E.P. 1988. *Theileria parva:* genomic DNA studies reveal intra-specific sequence diversity. *Molecular and Biochemical Parasitology* 28: 77–84.

ALLSOPP, B.A., CARRINGTON, M., BAYLIS, H., SOHAL, S., DOLAN, T. and IAMS, K. 1989. Improved characterisation of *Theileria parva* isolates using the polymerase chain reaction and oligonucleotide probes. *Molecular and Biochemical Parasitology* 35: 137–148.

BAYLIS, H.A., SOHAL, S.K., CARRINGTON, M., BISHOP, R.P. and ALLSOPP, B.A. 1991. An unusual repetitive gene family in *Theileria parva* which is stage specifically transcribed. *Molecular and Biochemical Parasitology* 49: 133–142.

BIGGS, B.A., KEMP, D.J. and BROWN, D.V. 1989. Subtelomeric deletions in field isolates of *Plasmodium falciparum* and their relationship to loss of cytoadherence *in vitro*. *Proceedings of the National Academy of Sciences of the USA* 86: 2428–2432.

BISHOP, R., SOHANPAL, B., KARIUKI, D.P., YOUNG, A.S., NENE, V., BAYLIS, H., ALLSOPP, B.A., SPOONER, P.R., DOLAN, T.T. and MORZARIA, S.P. 1992. Detection of a carrier state in *Theileria parva*-infected cattle by the polymerase chain reaction. *Parasitology* 104: 215–232.

CONRAD, P.A., IAMS, K., BROWN, W.C., SOHANPAL, B. and ole MOI-YOI, O. 1987. DNA probes detect genomic diversity in *Theileria parva* stocks. *Molecular and Biochemical Parasitology* 25: 213–226.

CORCORAN, L.M., THOMPSON, J.K., WALLIKER, D. and KEMP, D.J. 1988. Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *Plasmodium falciparum*. *Cell* 53: 807–813.

GUNDERSON, J.H., SOGIN, M.L., WOLLETT, G., HOLLINGDALE, M., de la CRUZ, V.F., WATERS, A.P. and McCUTCHAN, T.F. 1987. Structurally distinct stage-specific ribosomes occur in *Plasmodium*. *Science* 238: 933–937.

IRVIN, A.D., DOBBELAERE, D.A.E., MWAMACHI, D.M., MINAMI, M., SPOONER, P.R. and OCAMA, J.G.R. 1983. Immunisation against East Coast fever: Correlation between monoclonal antibody profiles of *Theileria parva* stocks and cross-immunity *in vivo*. *Research in Veterinary Science* 35: 341–346.

MINAMI, T., SPOONER, P.R., IRVIN, A.D., OCAMA, J.G.R., DOBBELAERE, D.A.E. and FUJINAGA, T. 1983. Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. *Research in Veterinary Science* 35: 334–340.

MORZARIA, S.P. and YOUNG, J.R. 1992. Restriction mapping of the genome of the protozoan parasite *Theileria parva*. *Proceedings of the National Academy of Sciences of the USA* 89: 5241–5245.

MORZARIA, S.P., SPOONER, S.P., BISHOP, R.P., MUSOKE, A.J. and YOUNG, J.R. 1990. *Sfi*I and *Not*I polymorphisms in *Theileria* stocks detected by pulsed-field gel electrophoresis. *Molecular and Biochemical Parasitology* 40: 203–212.

MUTUGI, J.J., YOUNG, A.S., MARITIM, A.C., NDUNGU, S.K., MINING, S.K., LINYONYI, A., NGUMI, P.N., LEITCH, B.L., MORZARIA, S.P. and DOLAN, T.T. 1989. Immunization of cattle in

Coast Province, Kenya: laboratory evaluation of a *Theileria parva parva* stabilate for use in infection-and-treatment immunization in the field. *Research in Veterinary Science* 27: 170–177.

NENE, V., IAMS, K.P., GOBRIGHT, E. and MUSOKE, A.J. 1992. Characterisation of the gene encoding a candidate vaccine antigen of *Theileria parva* sporozoites. *Molecular and Biochemical Parasitology* 51: 17–28.

NUSSENZWEIG, V. and NUSSENZWEIG, R.S. 1985. Circumsporozoite proteins of malaria parasites. *Cell* 42: 401–403.

PONZI, M., PACE., T., DORE., E. and FRONTALI, C. 1985. Identification of a telomeric DNA sequence in *Plasmodium berghei. The EMBO Journal* 4: 2991–2995.

SHAPIRO, S.Z., FUJISAKI, K., MORZARIA, S.P., WEBSTER, P., FUJINAGA, T., SPOONER, P.R. and IRVIN, A.D. 1987. A life-cycle stage-specific antigen of *Theileria parva* recognised by anti-schizont monoclonal antibodies. *Parasitology* 94: 29–37.

TAUTZ, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463–6471.

TOYE, P.G., GODDEERIS, B.M., IAMS, K., MUSOKE, A.J. and MORRISON, W.I. 1991. Characterisation of a polymorphic immunodominant molecule in sporozoites and schizonts of *Theileria parva*. *Parasite Immunology* 13: 49–62.

VASSART, G., GEORGES, M., MONSIEUR, R., BROCAS, H., LEQUARRE, A.S. and CHRISTOPHE, D. 1987. A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* 235: 683–684.

WELSH, J. and McCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research.* 18: 7213–7218.

WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSKI, J.A. and TINGEY, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531–6535.

Theileria parva: a restriction map and genetic recombination

S.P. Morzaria, J.R. Young, P.R. Spooner, T.T. Dolan and R.P. Bishop

In order to study the nature of observed polymorphisms in *Theileria parva*, we have constructed a long-range restriction map of the parasite genome and generated several random markers in the genome. Subsequently, we have used various chromosome-specific markers to demonstrate that sexual reproduction occurs during the parasite life cycle.

The map reveals 29 *Sfi*I and four *Not*I sites within the four chromosomes of the 10 million base pair haploid nuclear genome. Several genes, including those for parasite antigens, and the single locus carrying the major repetitive sequence have been located on the map. Of the 29 *Sfi*I sites, 23 were identified using linking clones, and of the remaining eight links, four were established by analysis of partial digests and four were implied because no alternative arrangements of fragments were possible.

In sexual recombination studies, polymorphic markers from 11 different loci on the genome (major repetitive sequences, ribosomal RNA gene, major schizont antigen gene and eight telomeres) were used. Several recombinants were identified and analysis of two of these showed that both were derived by independent assortment of chromosomes. The methodology developed for mapping the *T. parva* genome may be applicable to genomes of other lower eukaryotes. Both the physical mapping of the parasite genome and the ability to produce genetic crosses provide the foundation for the construction of a higher resolution map, the production of genetic crosses for linkage analysis, and the subsequent localization and identification of genes of important biological traits.

INTRODUCTION

Theileria parva, a tick-transmitted protozoan parasite of cattle and buffalo, causes a severe disease syndrome in cattle known variously as East Coast fever, January disease and Corridor disease. The disease is important because it affects cattle in large parts of eastern, central and southern Africa and limits the introduction of more productive breeds in endemic areas (Mukhebi *et al.*, 1992). The International Laboratory for Research on Animal Diseases (ILRAD) is attempting to develop improved methods for control of the disease through a better understanding of its epidemiology and by developing novel vaccines.

Theileria parva stocks and strains exhibit extensive genotypic and phenotypic polymorphisms. Genomic polymorphisms have been characterized using the parasite repetitive sequence, designated TpR, ribosomal RNA gene probes and telomeric sequences (Bishop *et al.*, 1993). Size polymorphisms have been identified in *Sfi*I macro-restriction fragments of different *T. parva* stocks (Morzaria *et al.*, 1990). Important phenotypic differences observed in *T. parva* include the presence of different immunological strains (Radley *et al.*, 1975), variations in the virulence of parasites derived from buffalo and cattle (Grootenhuis and Young, 1981), the presence of strains causing mild disease (Irvin *et al.*, 1989) and the occurrence of parasites causing no carrier state in cattle (Bishop *et al.*, 1992). Some

of these characteristics affect the epidemiology of the disease and may be important in planning strategies for developing novel vaccines.

A systematic approach to the genome analysis was initiated to study the genetic basis of the observed polymorphisms. This approach involved the construction of a restriction map, generation of numerous chromosome-specific polymorphic markers, and the generation and characterization of sexual recombinants following crosses between strains of *T. parva*.

MAPPING THE *THEILERIA PARVA* GENOME

When the mapping project was initiated only rudimentary information on the *T. parva* genome was available with only four genes cloned and sequenced. Preliminary analysis had shown that the nuclear genome size was ~10 million base pairs (Allsopp and Allsopp, 1988). Rare-cutting restriction enzymes such as *Sfi*I and *Not*I produced 33 and 8 fragments, respectively, and these could be easily resolved on pulsed-field gel electrophoresis (PFGE).

Based on this information, three mapping approaches were considered. Simple restriction mapping by double and partial digestion would have been the easiest and quickest approach, but such a map would have provided little information and no markers for easy comparison of stocks. The production of overlapping clones for the whole genome, using cosmids, would provide a physical map with a large number of markers distributed throughout the genome. However, this approach was rejected because of the problems in mapping the *Plasmodium falciparum* genome (Triglia and Kemp, 1991). They encountered difficulties in cloning DNA fragments in *Escherichia coli* because of the parasite's A+T-rich genome and the *T. parva* genome is approximately 70% A+T-rich. Thus a third approach, commonly described as the `top down' approach, was used. This involved the separation and subdivision of the chromosomes into their macro-restriction fragments and localization of various genes on these fragments. Combining this approach with the ordering of the macro-restriction fragments along the genome, using linking clones, provided a complete macro-restriction map of the *T. parva* genome.

Chromosomes, *Sfi*I, *Not*I Fragments and Telomeres

A combination of different PFGE techniques was used to characterize the chromosomes and *Sfi*I and *Not*I fragments of the *T. parva* genome. The results of these studies showed that there were four linear chromosomes in the nuclear genome of *T. parva*, with 29 *Sfi*I and four *Not*I sites. A *Plasmodium berghei* telomeric oligonucleotide probe, which also hybridizes with *T. parva* telomeres, identified eight *Sfi*I fragments. This probe also recognized eight *Eco*RI fragments that were susceptible to *Bal*-31 digestion (R. Bishop and B. Sohanpal, unpublished data) thus confirming the presence of four chromosomes in the genome. The summation of the sizes of all the *Sfi*I fragments confirmed the previously estimated genome size to be 10^7 bp (Allsopp and Allsopp, 1988).

Ordering *Sfi*I Fragments Using Linking Clone Strategy

Linking clones are unique genomic DNA fragments bearing a particular restriction site and when used as hybridizing probes they identify two fragments joined together by the restriction site. The linking clone strategy (Smith *et al.*, 1987) was used with *T. parva* to order all the 33 *Sfi*I fragments along the four separated chromosomes. The mapping involved the construction of a genomic library, selection of clones bearing *Sfi*I restriction sites, characterization of *Sfi*I linking clones and the use of the linking clones as probes in Southern blots of PFGE separated *Sfi*I fragments (Morzaria and Young, 1992).

A library of 4–6 kb *Sau*3AI fragments from the partially digested genomic DNA of *T. parva* was constructed into the Bluescript SKM13+ plasmid vector. In order to identify *Sau*3AI DNA fragments bearing *Sfi*I sites, the total DNA from the library was digested with *Sfi*I to linearize the appropriate fragments. These linear molecules were then physically separated by agarose gel electrophoresis (Young and Morzaria, 1991), gene-cleaned and re-ligated to obtain an enriched population of the plasmid containing *Sau*3AI fragments bearing the *Sfi*I site (Morzaria and Young, 1992). This was then transformed into appropriate bacteria and small-scale plasmid preparations (minipreps) from individual colonies were analyzed in order to select unique linking clones.

The minipreps were analyzed by agarose gel electrophoresis, and those containing inserts of appropriate length were characterized further by electrophoretic analysis of *Sau*3AI and *Sau*3AI + *Sfi*I digests. Agarose gels were used to reveal the larger *Sau*3AI fragments and acrylamide gels were used to reveal the smaller (less than 500 bp) *Sau*3AI fragments (Young and Morzaria, 1991). Of the 355 minipreps examined, 199 contained the expected size inserts of between 4–6 kb. Of the 199, 123 were found to contain *T. parva* genomic fragments bearing *Sfi*I sites and these were sorted into 21 groups, each representing a single *Sfi*I site. The linking clones were used in three kinds of analysis; as hybridization probes on Southern blots of *Sfi*I-digested *T. parva* genomic DNA to detect linkage between *Sfi*I fragments, as specific markers to assign *Sfi*I fragments to separated chromosomes and *Not*I fragments and two clones were hybridized to partial digestion ladders.

Four *Not*I sites were detected by double digestion with *Not*I and *Sfi*I. Their locations within *Sfi*I fragments were determined by hybridization of *Not*I + *Sfi*I digests with linking clones including the ends of the cut *Sfi*I fragments. The complete map of the *Sfi*I and *Not*I sites in the genome of the *T. parva* Muguga stock is shown in Figure 1. The main results of the genome analysis are listed below.

- The nuclear genome of 10^7 bp consists of four chromosomes of 3.2, 2.3, 2.2 and 2.1 mb.
- The genome is subdivided into 29 *Sfi*I and four *Not*I sites.
- The order of *Sfi*I and *Not*I fragments along the chromosomes has been established.
- Twenty-one of the linkages between *Sfi*I fragments have been established by linking clones.
- *Sfi*I fragments bearing telomeres have been identified.
- Several house-keeping and antigen genes, random schizont cDNAs and a major repetitive sequence have been localized in the map.

The type and the distribution of chromosome specific markers available for the *T. parva* genome are summarized in Table 1.

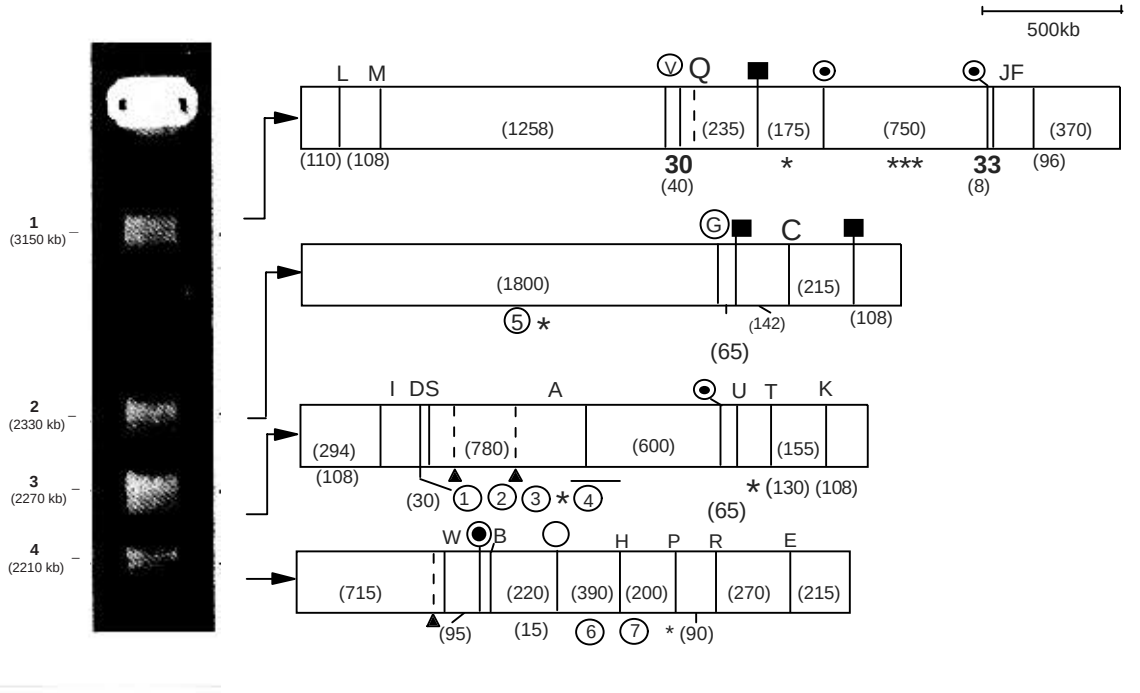

Figure 1. The genome map of *Theileria parva* Muguga, adapted from Morzaria and Young (1992). Four chromosomes separated by field inversion gel electrophoresis are shown in the ethidium bromide-stained gel on the left hand side of the figure. The sizes of the chromosomes are given in kb. The *Sfi*I restriction map of each chromosome is shown on the right hand side of the gel. *Sfi*I fragments are numbered 1–33 and their sizes in kb are given in brackets. Telomere-bearing *Sfi*I fragments are indicated as rounded ends. The linking clones spanning *Sfi*I sites are indicated by circled and uncircled capital letters. Circled numbers 1–7 and asterisks represent the location of various antigen genes and uncharacterized schizont cDNAs, respectively. Approximate sizes of the linking clones and the *Sau*3AI fragments bearing *Sfi*I sites are shown in the table.

Table 1. *Theileria parva* chromosome-specific DNA markers.

Markers	Number of markers	Chromosome number 1	2	3	4
Linking clones	21	6	2	7	6
Antigen genes	5	1	0	3	1
House keeping genes	3	0	1	1	1
Random cDNAs	8	4	1	2	1
TpR major repetitive sequence	1	0	0	1	0
Total	38	11	4	14	9

SEXUAL CYCLE IN *THEILERIA PARVA*

By analogy with related Apicomplexan parasites of the genera *Plasmodium* and *Babesia*, it has been assumed that a sexual cycle occurs in *T. parva* (Irvin and Boarer, 1980; Mehlhorn and Schein, 1984), however no genetic evidence has been presented. As part of our study of polymorphism in *T. parva* it was important to identify if a sexual cycle occurred in *T. parva* and whether this provided a mechanism for generating polymorphism.

The macro-restriction map of the genome formed the backbone of the study of the sexual cycle. In designing experiments to identify the presence of a sexual cycle, we made certain assumptions based on the findings with malaria parasites. It was assumed that the parasite had a predominantly haploid life-cycle, except when diploid zygotes and kinetes were formed, and that the sporozoites were derived from a single meiotic division followed by a multiple asexual division. It has been estimated that the probability of a single acinar cell in a tick salivary gland becoming infected with two or more kinetes is 1 in 10^6 (Fujisaki *et al.*, 1988) and, therefore, a sporozoite population derived from a single-infected acinus was considered a product of a single meiotic division.

Experimental Design to Investigate the Presence of a Sexual Cycle in *Theileria parva*

Two cattle were each infected with both the Muguga and Uganda stocks of *T. parva*, for which a number of phenotypic markers were available. An important feature of the selected stocks was that they developed piroplasms in cattle at about the same rate, providing the opportunity for ticks to become infected with both parasites. Salivary glands from ticks which fed on these cattle were dissected, removed and examined under interference contrast microscopy for identification of glands containing only single infected acini. Sporozoites from these single infected acini were analyzed. It was postulated that a population of sporozoites from a single meiotic product should contain either Muguga, Uganda or a Muguga/Uganda mixed genotype.

Results

Sporozoite DNA from single infected acinar cells was amplified by the polymerase chain reaction (PCR) using conserved primers (Bishop *et al.,* 1992) derived from a *T. parva* repetitive DNA sequence TpR (Allsopp *et al.,* 1989). The primers flanked a region of sequence which was variable between stocks. The amplified DNA was probed using oligonucleotides that distinguish between the Muguga and Uganda parasite stocks on a +/- basis (Bishop *et al.,* 1993). Cloned, schizont-infected lymphoblastoid cell lines derived from infection with sporozoites from single-infected acinar cells, which reacted with both oligonucleotides, were characterized using a panel of schizont-specific monoclonal antibodies (MAbs) as phenotypic markers (Minami *et al.,* 1983) and a range of DNA probes, including the TpR repetitive sequence, a small subunit ribosomal RNA gene, and a *T. parva* telomeric sequence (Bishop *et al.,* 1993) as genotypic markers. The PCR analysis showed that the sporozoite populations from 38% of the single infected acinar cells had genotypes representing mixtures of both stocks. One of the sporozoite populations of mixed genotype was used to infect bovine lymphocytes and amplified *in vitro* as a cloned schizont-infected cell line. Analysis of parasite DNA from the cell-line using the telomeric probe confirmed that the parasite was the product of a sexual cross between the two stocks and that this recombinant was produced by an independent assortment of chromosomes.

CONCLUSIONS

The investigations reported above provide evidence that sexual reproduction occurs in *T. parva.* Analysis of one of the recombinants showed that genetic exchange occurred between two parasite populations and is one mechanism which could account for the polymorphism detected in *T. parva.* Other putative recombinants are currently being characterized. The significance of the sexual cycle in the generation of polymorphism during natural infection and its contribution to antigenic diversity is yet to be determined.

Now that it is possible to produce genetic crosses in *T. parva*, classical genetic studies can be initiated. The combination of genetic and physical maps in the future may provide a powerful tool for the identification of important parasite genes and for the studies on the molecular structure of the genome.

ACKNOWLEDGEMENTS

The technical help of Mr. J. Kiarie, Mr. F. Mwakima, Mr. T. Ndolo and staff in the Tick Unit, ILRAD, is greatly appreciated.

REFERENCES

ALLSOPP, B.A. and ALLSOPP, M.T.E.P. 1988. *Theileria parva:* genomic DNA studies reveal intra-specific sequence diversity. *Molecular and Biochemical Parasitology* 28: 77–84.

ALLSOPP, B.A., CARRINGTON, M., BAYLIS, H., SOHAL, S., DOLAN, T.T. and IAMS, K. 1989. Improved characterisation of *Theileria parva* isolates using the polymerase chain reaction and oligonucleotide probes. *Molecular and Biochemical Parasitology* 35: 137–148.

BISHOP, R., SOHANPAL, B., KARIUKI, D.P., YOUNG, A.S., NENE, V., BAYLIS, H., ALLSOPP, B.A., SPOONER, P.R., DOLAN, T.T. and MORZARIA, S.P. 1992. Detection of a carrier state in *Theileria parva*-infected cattle by the polymerase chain reaction. *Parasitology* 104: 215–232.

BISHOP, R.P., SOHANPAL, B.K., ALLSOPP, B.A., SPOONER, P.R., DOLAN, T.T. and MORZARIA, S.P. 1993. Detection of polymorphisms among *Theileria parva* stocks using repetitive, telomeric and ribosomal DNA probes and anti-schizont monoclonal antibodies. *Parasitology,* in press.

FUJISAKI, K., IRVIN, A.D., VOIGT, W.P., LEITCH, B.L. and MORZARIA, S.P. 1988. The establishment of infection in the salivary glands of *Rhipicephalus appendiculatus* ticks by transplantation of kinetes of *Theileria parva* and the potential use of the method for parasite cloning. *International Journal of Parasitology* 18: 75–78.

GROOTENHUIS, J.G. and YOUNG, A.S. 1981. The importance of wildlife in the epidemiology of theileriosis. In: Karstad, L., Nestel, B. and Graham, M., eds. *Wildlife Disease Research and Economic Development.* Ottawa: IDRC, pp. 33–39.

IRVIN, A.D. and BOARER, C.D.H. 1980. Some implications of a sexual cycle in *Theileria*. *Parasitology* 80: 571–579.

IRVIN, A.D., MORZARIA, S.P., MUNATSWA, F.C. and NORVAL, R.A.I. 1989. Immunisation of cattle with a *Theileria parva bovis* stock from Zimbabwe protects against challenge with virulent *T. p. parva* and *T. p. lawrencei* stocks from Kenya. *Veterinary Parasitology* 32: 271–278.

MEHLHORN, H. and SCHEIN, E. 1984. The piroplasms: life cycle and sexual stages. *Advances in Parasitology* 23: 37–103.

MINAMI, T., SPOONER, P.R., IRVIN, A.D. and FUJINAGA, T. 1983. Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. *Research in Veterinary Science* 35: 334–340.

MORZARIA, S.P. and YOUNG, J.R. 1992. Restriction mapping of the genome of the protozoan parasite *Theileria parva. Proceedings of the National Academy of Sciences of the USA* 89: 5241–5245.

MORZARIA, S.P., SPOONER, P.R., BISHOP, R.P., MUSOKE, A.J. and YOUNG, J.R. 1990. *Sfi*I and *Not*I polymorphisms in *Theileria* stocks detected by pulsed-field gel electrophoresis. *Molecular and Biochemical Parasitology* 40: 203–211.

MUKHEBI, A.W., PERRY, B.D. and KRUSKA, R. 1992. Estimated economics of theileriosis control in Africa. *Preventive Veterinary Medicine* 12: 73–85.

RADLEY, D.E., BROWN, C.G.D., BURRIDGE, M.J., CUNNINGHAM, M.P., KIRIMI, I.M., PURRELL, R.E. and YOUNG, A.S. 1975. East Coast fever. 1. Chemoprophylactic immunization of cattle against *Theileria parva* (Muguga) and five theilerial strains. *Veterinary Parasitology* 1: 35–41.

SMITH, C.L., LAWRENCE, S.K., GILLESPIE, G.A., CANTOR, C.R., WEISSMAN, S.M. and COLLINS, F.S. 1987. Strategies for mapping and cloning macro-regions of mammalian genomes. In: Gottesman, M., ed. *Methods in Enzymology, Molecular Genetics of Mammalian Cells*, Vol. 151. New York. Academic Press, pp. 461–489.

TRIGLIA, T. and KEMP, D.J. 1991. Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Molecular and Biochemical Parasitology* 44: 207–211.

YOUNG, J.R. and MORZARIA, S.P. 1991. Agarose entrapment method for the production of *Sfi*I linking library for *Theileria parva. Genetic Analysis: Techniques Applications* 8: 148–150.

Summary of discussion

Chairperson: *Prof. A. Tait*
Rapporteurs: *Drs. P. Majiwa and D. McKeever*

Most of the discussion on this session focused on use of physical maps and genetic crosses in *Theileria parva* for the detection of genes that determine traits of interest. Virulence was one trait that was discussed in detail.

One of the problems in detecting the gene(s) for virulence in *T. parva* was lack of correlation between *in vitro* characteristics of the parasite and *in vivo* behaviour. It was emphasized that the assessment of the virulence of a *T. parva* stock is rendered complex by the influence of the phenotype of the infected host cell. Indeed, many of the parasite's traits may be determined or strongly influenced by the host, in which case it may be futile to look for gene associations in the parasite. An additional complicating factor is that the parasite turns on some host genes and turns off others. Because of these host influences, it may be difficult to look for parasite genes responsible for such traits. Since there is no information on the genetic basis for virulence in *T. parva*, it was difficult to formulate a mapping strategy for the identification of relevant genes. However, it was pointed out that lack of information on the chromosomal location of a gene or genes has not prevented the search for them. Examples given were the search for genes responsible for schizophrenia and homosexuality, which are believed to be hereditary but are also environmentally influenced.

A question was raised about the possibility of generating *in vitro* phenotypes of *T. parva* and whether such phenotypes had any correlation with *in vivo* traits. It was pointed out that this was attempted in one instance by Dr. C.G.D. Brown, who used attenuated *T. parva*-infected lymphoblastoid cell lines. He also attempted to generate drug-resistant strains without success. One other trait considered for study was merogony and the carrier state. The difficulty with this trait again is the problem of inducing merogony *in vitro*.

It was suggested that mutagenesis be considered as an approach to generating new traits that could be studied in the laboratory, particularly where haploid organisms are concerned. It must be borne in mind, however, that mutagenesis techniques have been applied in malaria and have not always given correct answers and may in fact lead one down the wrong alley. An example of this is the mutations caused in the parasite by nitrogen mustard and drug pressure in the laboratory, which have no relevance to the basis of drug resistance in the field.

There was a lengthy discussion on the potential uses of performing parasite crosses for the identification of gene loci for important parasite traits. Generation of large numbers of recombinants from a single cross in the *T. parva* should not be a problem since one can clone large numbers of sporozoites from a single infected acinus. But it was pointed out that performing crosses in *T. parva* is an expensive undertaking, particularly considering the large number of cattle required for complete analyses of products from each cross. It

would therefore be useful to know the number of crosses that need to be made before genes or traits of significance are mapped in *T. parva.* It was suggested that if the fluorescent-activated cell sorter can be used to sort sporozoites into 96-well plates as has been done to sort single sperms, degenerate oligonucleotides could be used to amplify genetic material for subsequent analysis.

The group felt that despite reservations expressed about the expenses involved in analysing genetic crosses in *T. parva,* it would be a good idea to make one cross that would service all the conceivable mapping needs of the program. As many recombinants as possible should be generated and stored safely for future use. The choice of parental parasite populations for such a cross would be crucial and would be complicated by the fact that phenotype may not actually reflect genotype and that organisms may switch genes on and off.

There was also a discussion on the requirement for markers and the types of markers for characterization of parasites. It was clear that random and specific DNA markers would be of considerable use in characterizing allelic diversity in parasite populations in the field. This would provide better understanding of parasite behaviour and enable better planning of immunization in the field. For the immediate future it was thought that polymorphic markers for monitoring vaccine breakdowns will be useful in the application of novel vaccines. ILRAD will be in a better position to identify relevant markers for vaccine breakthrough populations after at least one of the antigens that provoke CTL responses has been identified. One approach to the identification of such a marker is to use polymorphic markers to identify CTL antigens by generating recombinants from two parasite clones, one of which was seen by a CTL clone and the other of which was not. By screening these recombinants with an array of polymorphic markers, the gene encoding the antigen should be detectable.

Markers can also be used to find genes without having to generate maps. Random DNA markers have been successfully exploited in finding genes, particularly in plant systems. However a map is a resource to which one can always turn for both information and more detailed analyses of specific genetic loci.

In view of the current thrust of the theileriosis program, the group believed that the construction of a *T. parva* genome map of a resolution higher than the one currently available was of low priority.

GENETIC ANALYSIS OF TRYPANOSOMES

An introductory review of the important trypanosomes of livestock

P.R. Gardiner

The pathogenic trypanosomes which contribute to the disease complex affecting livestock productivity in Africa are briefly reviewed. Attention is drawn to the important biological features of trypanosomes and opportunities for genetic research. Differences, as well as similarities, amongst the livestock trypanosomes are highlighted to assist the formulation of research efforts which will effectively address the animal health and economic problems posed by these trypanosomes as a group.

The majority of experimental work conducted with the tsetse-transmitted African trypanosomes has focused on *Trypanosoma brucei brucei.* This has been because of the ease of its propagation in laboratory rodents, the early establishment of culture systems for these haemoflagellate protozoan organisms and its close relationship to two other subspecies, *T. b. rhodesiense* and *T. b. gambiense,* which cause two forms of human trypanosomiasis in Africa. Many important findings, and even biological precedents, have resulted from such work, some of which are outlined below. However, in terms of combating livestock trypanosomiasis, both in Africa and elsewhere, other sub-genera and species of trypanosomes present much more serious problems (Hoare, 1972).

Trypanosoma brucei brucei and its human infective relatives belong to the *Trypanozoon* sub-genus of trypanosomes. These parasites have a relatively complicated cycle of development in the tsetse fly vector during which they exhibit differences in mitochondrial development and energy metabolism in the vector and in mammalian hosts. The parasites display clearly demarcated phenotypes during the cycle in the vector and in bloodstream infections. These parasites tend not to be pathogenic in livestock although, in mixed infections, they may contribute to antigenic load and may interfere with the integrity of the blood/brain barrier, making radical cure difficult (Stephen, 1970).

Although there are host susceptibility differences and variations across the continent, the most pathogenic trypanosomes in Africa belong to the *Nannomonas* and *Duttonella* sub-genera. Amongst the *Nannomonas* parasites, *T. congolense* produces acute and/or chronic fatal disease in cattle and *T. simiae* is an acute pathogen of pigs. These *Nannomonas* parasites have a transmission cycle through tsetse of intermediate complexity, passing through the midgut of the fly and returning to the mouth parts. However, for these parasites the site for development to infective forms for mammals is the proboscis rather than the salivary glands favoured by the *Trypanozoon* species (Hoare, 1970).

Trypanosoma vivax is the principal species of the *Duttonella* group which causes a similar disease in cattle and small ruminants but with, in general, more frequent cases of spontaneous recovery and less chronic depression of bone marrow function. However, some isolates of this organism can produce an acute haemorrhagic disease which is usually fatal. The cycle of development which *T. vivax* undergoes in tsetse is relatively simple and

confined to the proboscis of the fly. This simplicity leads to relatively high infection rates and efficient transmission of this species which is often encountered as the first trypanosome infection following introduction of animals into a tsetse-infested area. *Trypanosoma vivax* has established itself in South America in the absence of the tsetse fly vector and is now transmitted by other biting flies such as tabanids. A more stable epidemiological situation has resulted, although outbreaks of disease when they occur in cattle can still have severe economic consequences (Gardiner, 1989).

Trypanosoma evansi, another member of the *Trypanozoon* subgenus, and closely related to *T. brucei*, has a lesion in its mitochondrial DNA which prevents its establishment in tsetse. Nevertheless, through the agency of mechanical transmission, *T. evansi* has become the most widespread of the pathogenic trypanosomes of livestock with a distribution extending through the tropical parts of South America, North Africa, the Middle East, Southern Europe and Asia. Different host species are affected in different areas: horses and cattle in South America, the camel in North Africa, the Middle East and northern Asia, and the water buffalo in Asia. The disease is particularly severe in horses, which quickly develop secondary nervous system complications and paralysis. In many areas of the world, *T. evansi* infection is the most important disease of camels and in Asia, particularly Indo-China, epidemics of *T. evansi* infections involving thousands of working or milking buffalo still occur (Gardiner and Mahmoud, 1992).

It can be seen from the foregoing that animal trypanosomiasis is not one disease, but a disease complex, produced by organisms which, although similar, are sufficiently different in their life cycles, transmission and pathogenesis to warrant different emphases in programs of research and control.

Trypanosomes, principally *Trypanosoma brucei*, have been studied for nearly a century (c.f. Plimmer and Bradford, 1899; Carruthers and Cross, 1992). Not only have they remained pathogens of human and veterinary importance, but their study has revealed a number of unique biological capabilities and processes which have intrigued cellular biologists and parasitologists alike (recently reviewed in Logan-Henfrey *et al.*, 1992). Principal amongst these unique capabilities is the ability of the salivarian trypanosomes to alter the expression of their surface coat molecules. The surface coat of a trypanosome is an all-enveloping layer of a single glycoprotein arrayed in a molecular palisade. The majority of a population of bloodstream trypanosomes usually express a single type of surface coat antigen. However, minor types will also exist and when the bulk of the population is removed by an antibody response, the minor types will grow up to re-establish parasitaemia in the blood (Van der Ploeg *et al.*, 1982). With *brucei*-group trypanosomes possessing an estimated 300 to 1000 genes for the variable surface glycoproteins (VSGs), the phenomenon of antigenic variation usually overcomes the immune capabilities of the host animal and results in chronic, wasting infections with attendant losses in productivity and through mortality.

The molecular mechanisms governing antigenic variation are known in some detail, but the biological event which triggers expression of new VSGs in a trypanosome population remains elusive. Whilst the reader is referred to any of a number of excellent reviews on the subject (Boothroyd, 1985; Borst, 1986; van der Ploeg, 1987; Pays and Steinert, 1988), from the point of view of genetic analysis it should be noted that the duplication of basic copies of antigen genes and their expression from telomeric sites both indicate the great plasticity of the

trypanosome genome and can lead to alterations in telomere and chromosome lengths. Whilst it has been proved that all the salivarian trypanosomes described above undergo antigenic variation, the number of antigen genes, the mechanisms of expression and the extent of antigen gene diversity are much less well known outside the *Trypanozoon* group. In tsetse-transmitted trypanosomes, VSG expression is shut down on transformation of bloodstream forms to the insect stages in the vector. The trypanosomes then express a different, invariant surface molecule termed procyclin (Roditi and Pearson, 1990) which, however, differs quite markedly between *T. brucei* and *T. congolense.* When the trypanosomes in the vector mature to metacyclic forms (which occurs in the salivary glands for *T. brucei* and in the proboscis for *T. congolense* and *T. vivax),* the trypanosomes once again assume the VSG coat and are the infective stage for mammals. Interestingly, the metacyclic population expresses a limited set of the total VSG repertoire which can be developed by bloodstream forms (Barry *et al.,* 1985). This at one time seemed to hold out the promise of vaccination against a limited number of trypanosome antigenic variants. However, each species of trypanosome possesses non-overlapping variant antigens. Serodemes (antigenic repertoires) of individual species can also differ markedly in antigen expression providing no serological cross-resistance between the majority of parasite isolates. This degree of complexity in antigenic expression has caused research at ILRAD and elsewhere to shift focus to key invariant molecules of the trypanosome in the search for vaccines against these organisms.

The major trypanosome sub-genera are distinguished by characteristic patterns of motility and morphology. There is considerable phenotypic differentiation during the life cycle of these parasites with each group showing distinct morphologies associated with dividing and non-dividing stages in the blood. Further forms characterize passage through different organs of the tsetse vector with multiplication taking place both as procyclic trypanosomes in the midgut (this stage is absent in the life cycle of *T. vivax)* and attached epimastigotes in the mouth parts of the fly. The metacyclic trypanosomes are in the resting, G_0, stage of the life cycle and are triggered to differentiate and divide as bloodstream forms upon inoculation into the mammalian bloodstream as a tsetse feeds. Differential expression of genes during the life cycle of trypanosomes, the genetic and functional linkage of these genes and their control by parasite or host factors are all important topics for research with long-term prospects for revealing mechanisms through which to control the growth of these parasites (Murphy *et al.,* this volume).

The differentiation that *T. brucei* undergoes during its life cycle is accompanied by a remarkable modulation in the form and function of the mitochondrion (Fairlamb and Opperdoes, 1986). This allows the parasite, which has respired relatively inefficiently in the nutrient-rich environment of the mammalian bloodstream, to effectively extend its life cycle into the nutrient poor environment of the insect. Again, there are key differences in this ability between trypanosome species, with the mitochondrion in *T. vivax* apparently being fully activated in both bloodstream and insect stages. In *T. evansi,* the kinetoplast maxicircles—the true mitochondrial DNA which encodes several respiratory enzymes—are lacking and this species is unable to be cyclically transmitted through tsetse (Gardiner and Mahmoud, 1992). The kinetoplast also contains several thousand copies of minicircular loops of DNA linked together in dense networks to form a body which is visible even at the level of the light microscope. Topoisomerases, enzymes which assist the unfolding and replication of minicircle DNA, constitute a rare enzyme type and perhaps a target for

chemotherapy (Shapiro and Englund, 1990). For many years, the function of minicircles was unknown, but recently trypanosomes have set another important biological precedent with the discovery that the minicircles can serve as guides for RNA editing of the transcripts of the mitochondrial genome (Stuart, 1991). This partitioning of form and function of the organism's DNA requires that these parasites separately control kinetoplast and nuclear division (Sherwin and Gull, 1989) and adds further considerations for those wishing to study the genetic complexity and reassortment of important characteristics.

Trypanosomes as bloodstream forms are entirely dependent upon glycolysis for the derivation of their ATP. The glycolytic enzymes of the trypanosomes and related kinetoplastids are sequestered in vesicular organelles termed glycosomes (reviewed in Michels, 1988). These enzymes are cytosolic in other cells. Glycosomes are similar to the microbodies and peroxisomes of other cells in which enzymes performing related or integrated functions are gathered together. It may be that the extremely high rate of glucose utilization by kinetoplastids (Opperdoes, 1987) has provided the evolutionary pressure for the formation of this unique organelle. Targeting mechanisms to ensure transport of the nascent enzymes to the glycosome have required genetic and amino acid sequence differences between the glycolytic enzymes and their cytosolic counterparts.

A further characteristic of the trypanosomes and related kinetoplastid protozoa is their method of RNA splicing in which a common leader sequence contained within a mini-exon is spliced to the gene transcript (Sutton and Boothroyd, 1988). Polycistronic messenger RNAs are common in trypanosomes, even when gene members of a cistron will be differentially regulated subsequently. To what extent the mechanisms, learnt from *T. brucei,* which govern RNA splicing can be extrapolated to the other important trypanosomes is not yet clear. Recent research provides unexpected results with *T. vivax* which, although it has mini-exon sequences which are linked to 5S ribosomal RNA genes (Roditi, 1992), lacks one of the two RNA polymerase 2 genes usually found in the trypanosomes which undergo antigenic variation (Smith *et al.,* 1993).

The genetic organization of trypanosomes will be covered more thoroughly elsewhere in this workshop, but it is interesting to note that at least judged by chromosome size (see Majiwa, Melville *et al.* and Tait *et al.,* this volume) trypanosome isolates can differ markedly even within a species. Karyotype analysis shows *T. brucei* to possess large, medium-sized and mini chromosomes (Van der Ploeg *et al.,* 1989) although our present knowledge of nuclear division in these organisms is insufficient to determine whether each of the three classes are replicated and separated to daughter cells by the same or different mechanisms. Whilst the size classes of chromosomes found in *T. congolense* parasites are roughly similar, the majority of *T. vivax* chromosomes are in the megabase range and have not been separated completely (Dirie *et al.,* 1993a). It has not yet been determined what functional differences, if any, may result from the location of genes in different size classes of chromosomes. Certainly, genetic differences between isolates of the same trypanosome species are sufficient for random primed PCR to be useful in yielding isolate specific patterns (Waitumbi and Murphy, 1993; Dirie *et al.,* 1993b). These fundamental differences and occasional major changes caused by translocation, deletions and the process of antigenic variation which results in alterations in chromosome size make the selection of organisms for detailed genetic studies problematic, as would be the direct extrapolation of results from one species to another (see also Murphy *et al.,* this volume).

As we have seen, trypanosomes possess several unique biochemical pathways and organelles, but rational drug design for control of the trypanosomiases of veterinary importance is still at the developmental stage. The currently available trypanocides are relatively toxic molecules which have been important for the control of this disease in livestock for decades without their modes of action having been properly elucidated (Leach and Roberts, 1981). These shortcomings in knowledge of drug action and drug development have stemmed in large part from the relative lack of interest of commercial, pharmaceutical companies, who have anticipated relatively modest economic returns from the farmers of developing countries compared with modern drug development and licensing costs. However, work continues on the presently available trypanocides because of field evidence that trypanosomes resistant to the existing compounds can arise. How drug resistance arises and how it spreads through trypanosome populations are major questions for genetic analysis (see Tait *et al.*, this volume). Meanwhile, rational drug design against the function of the unique enzymes of the trypanosome glycosome is proceeding as a test case for the feasibility of this approach in scientific and economic terms (Verlinde *et al.*, 1992).

This paper has stressed both the uniqueness of the trypanosomes and the differences amongst those parasite species which are important constraints to livestock productivity in large areas of the tropical world. Control of trypanosomiasis is through vector control, management and chemotherapy of infected cattle and, to a limited extent, the farming of trypanotolerant cattle. The control of parasites and vectors requires continued vigilance and the repeated administration of potentially toxic compounds. Trypanotolerance may provide a sustainable means of control, but the numbers of trypanotolerant cattle are likely to remain too small to effectively combat this disease on a large scale in the medium-term future. ILRAD is pursuing genetic marker production for the trait to enhance the selection and dissemination of trypanotolerance genes in cattle (Teale, this volume). An effective, low cost vaccination procedure remains the most worthwhile and appropriate goal for research on the trypanosomes themselves. We have seen that the variable surface antigens are unlikely to yield appropriate vaccine targets and that the concentration is now on invariant target molecules. A pragmatic view would be that trypanosome somatic antigens or released pathogenic products are equally likely to provide targets for vaccines. It may be sufficient to limit the extent of the pathology associated with the disease rather than eradicate a trypanosome infection to allow animals to remain productive. The problem for research becomes not only the identification of the key molecules, but the degree of correspondence between similar molecules in different trypanosome species (Ziegelbauer *et al.*, 1992). It will therefore be necessary, whilst concentrating on the antigenic and genetic description of particular trypanosomes, to bear in mind the inherent variability amongst the pathogenic salivarian trypanosomes.

REFERENCES

BARRY, J.D., CROW, J.S. and VICKERMAN, K. 1985. Neutralization of individual variable antigen types in metacyclic populations of *Trypanosoma brucei* does not prevent their subsequent expression in mice. *Parasitology* 90: 79–88.

BOOTHROYD, J.C. 1985. Antigenic variation in African trypanosomes. *Annual Review of Microbiology* 39: 475–502.

BORST, P. 1986. Discontinuous transcription and antigenic variation in trypanosomes. *Annual Review of Biochemistry* 55: 701–732.

CARRUTHERS, V.B. and CROSS, G.A.M. 1992. High efficiency clonal growth of bloodstream- and insect-form *Trypanosoma brucei* on agarose plates. *Proceedings of the National Academy of Sciences of the USA.* 89: 8818–8821.

DIRIE, M.F., OTTE, J., THATTHI, R. and GARDINER, P.R. 1993a. Comparative studies of *Trypanosoma (Duttonella) vivax* isolates from Colombia. *Parasitology* 106: 21–29.

DIRIE, M.F., MURPHY, N.B. and GARDINER, P.R. 1993b. DNA fingerprinting of *Trypanosoma vivax* isolates rapidly identifies intraspecific relationships. *Eukaryotic Microbiology* 40: 132–134.

FAIRLAMB, A.H. and OPPERDOES, F.R. 1986. Carbohydrate metabolism in African trypanosomes, with special reference to the glycosome. In: Morgan, M.J., ed. *Carbohydrate Metabolism in Cultured Cells.* New York: Plenum Press, pp. 183–224.

GARDINER, P.R. 1989. Recent studies of the biology of *Trypanosoma vivax. Advances in Parasitology* 28: 229–317.

GARDINER, P.R. and MAHMOUD, M.M. 1992. Salivarian trypanosomes producing disease in livestock outside sub-Saharan Africa. In: Kreier, J.P. and Baker, J.R., eds. *The Parasitic Protozoa,* Second Edition, Volume 2. New York: Academic Press, pp. 277–314.

HOARE, C.A. 1970. The mammalian trypanosomes of Africa. In: Mulligan, H.W., ed. *The African Trypanosomiases.* London: George Allen and Unwin, pp. 3–19.

HOARE, C.A. 1972. *The Trypanosomes of Mammals: A Zoological Monograph.* Oxford and Edinburgh: Blackwell Scientific Publication.

LEACH, T.M. and ROBERTS, C.J. 1981. Present status of chemotherapy and chemoprophylaxis of animal trypanosomiasis in the Eastern Hemisphere. *Pharmacology and Therapeutics* 13: 91–147.

LOGAN-HENFREY, L., GARDINER, P.R. and MAHMOUD, M.M. 1992. Livestock trypanosomiasis in sub-Saharan Africa. In: Kreier, J.P. and Baker, J.R., eds. *The Parasitic Protozoa,* Second Edition, Volume 2. New York: Academic Press, pp. 157–275.

MICHELS, P.A.M. 1988. Compartmentation of glycolysis in trypanosomes: a potential target for new trypanocidal drugs. *Biology of the Cell* 64: 157–164.

OPPERDOES, F.R. 1987. Compartmentation of carbohydrate metabolism in trypanosomes. *Annual Review of Microbiology* 41: 127–151.

PAYS, E. and STEINERT, M. 1988. Control of antigen gene expression in African trypanosomes. *Annual Review of Genetics* 22: 107–126.

PLIMMER, H.G. and BRADFORD, J.R.C. 1899. A preliminary note on the morphology and distribution of the organism found in the tsetse-fly disease. *Proceedings of the Royal Society* B65: 274.

RODITI, I. 1992. *Trypanosoma vivax:* linkage of the mini-exon (spliced leader) and 5S ribosomal RNA genes. *Nucleic Acids Research* 20: 1995.

RODITI, I. and PEARSON, T.W. 1990. The procyclin coat of African trypanosomes (or the not-so-naked trypanosome) *Parasitology Today* 6: 79–82.

SHAPIRO, T.A. and ENGLUND, P.T. 1990. Selective cleavage of kinetoplast DNA minicircles promoted by anti-trypanosomal drugs. *Proceedings of the National Academy of Sciences of the USA* 87: 950–954.

SHERWIN, T. and GULL, K. 1989. The cell division cycle of *Trypanosoma brucei brucei:* timing of event markers and cytoskeletal modulations. *Philosophical Transactions of the Royal Society, London,* B323: 573–588.

SMITH, J.L., CHAPMAN, A.B. and AGABIAN, N. 1993. *Trypanosoma vivax:* Evidence for only one RNA polymerase II largest subunit gene in a trypanosome which undergoes antigenic variation. *Experimental Parasitology* 76: 242–246.

STUART, K. 1991. RNA editing in Trypanosomatid Mitochondria. *Annual Review of Microbiology* 45: 327–344.

STEPHEN, L.E. 1970. Clinical manifestations of the trypanosomiases in livestock and other domestic animals. In: Mulligan, H.W., ed. *The African Trypanosomiases.* London: George Allen and Unwin, pp. 774–794.

SUTTON, R.E. and BOOTHROYD, J.C. 1988. Trypanosome transsplicing utilises 2⌈–5⌈ branches and a corresponding branching activity. *EMBO Journal* 7: 1431–1437.

VERLINDE, C.L.M.J., RUDENKE, G. and HOL, W.G.J. 1992. In search of new lead compounds for trypanosomiasis drug design: a protein structure-based linked-fragment approach. *Journal of Computer-Aided Molecular Design* 6: 131–147.

Van der PLOEG, L.H.T. 1987. Control of variant surface antigens switching in trypanosomes. *Cell* 51: 159–161.

Van der PLOEG, L.H.T., VALERIO, D., De LANGE, T., BERNARDS, A., BORST, P. and GROSVELD, F.G. 1982. An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic Acids Research* 10: 5905–5923.

Van der PLOEG, L.H.T., SMITH, C.L., PALVER, R.I. and GOTTESDIENER, K.M. 1989. Improved separation of chromosome-sized DNA from *Trypanosoma brucei* stocks 427-60. *Nucleic Acids Research* 17: 3217–3225.

WAITUMBI, J.N. and MURPHY, N.B. 1993. Inter- and intra-species differentiation of trypanosomes by genomic fingerprinting with arbitrary primers. *Molecular and Biochemical Parasitology* 58: 181–186.

ZEIGELBAUER, K., MULTHAUP, G. and OVERATH, P. 1992. Molecular characterization of two invariant surface glycoproteins specific for the bloodstream stage of *Trypanosoma brucei*. *The Journal of Biological Chemistry* 267: 10797–10803.

Variability of *Trypanosoma congolense*

P. Majiwa

Parasitic protozoa are organisms with many interesting and variable phenotypes. Trypanosomes, for example, change their antigenic profiles in the process of antigenic variation and also display variability in the severities of the disease they cause in their vertebrate hosts. Among the trypanosomes that cause human sleeping sickness, two pathotypes are recognizable: *Trypanosoma brucei rhodesiense* which causes an acute disease syndrome and *T. b. gambiense* which causes a chronic disease syndrome. Several pathotypes have been described among trypanosomes collectively designated *T. congolense* which are responsible for livestock trypanosomiasis. In these trypanosomes, at least five distinct genotypes have been observed.

The extent of genetic polymorphism and proximity among trypanosomes can be estimated by the analysis of randomly amplified polymorphic DNA (RAPD) markers and restriction enzyme fragment length polymorphisms (RFLPs), and by direct comparison of nucleotide sequences of conserved genes.

It has been established that different species of trypanosomes can infect a single vector and thus a single host. It is therefore possible that genetic recombination could occur resulting in trypanosomes with new genotypes which may display entirely different clinical effects in different hosts. Given that different parasite genotypes can be identified accurately using DNA-based markers, it will be useful to determine recombination frequencies within parasite species by comparative analyses of their respective genomes, and the consequences of such events on parasite genetic repertoires and population types. Global comparison of the genomes of the trypanosomes is likely to yield useful information concerning their evolutionary relationships.

INTRODUCTION

Trypanosomes are unicellular parasitic protozoa that infect man as well as domestic and wild animals in many different parts of the world. The parasites are normally transmitted from one host to another by biting flies. The tsetse fly *(Glossina* spp.) is the primary vector of African trypanosomes. Some trypanosomes have evolved the capacity for mechanical transmission, thus dispensing altogether with the need for cyclical development in the fly vector. Such trypanosomes include *Trypanosoma evansi* and *T. equiperdum.* There is some evidence that *T. vivax* can be transmitted mechanically. Thus, although in Africa trypanosomiasis occurs mainly in the tsetse belt, the disease is found in other areas where the vector is absent.

The disease is caused by various species and subspecies of trypanosomes and results in serious health and economic problems. This brief review, with an emphasis on the livestock infective *T. congolense*, focuses on variabilities that have been observed both at the phenotypic and genotypic levels.

EARLY EVIDENCE

The initial phenotypic evidence which indicated that the trypanosomes classified as *T. congolense* have variable characteristics was presented during the early part of this century,

in the period spanning 1904 to 1906 (Dutton and Todd, 1903; Broden, 1904, 1906; Balfour, 1906). From the very start, much controversy, doubt, confusion and apparent disorder accompanied the nomenclature of this trypanosome. Different names such as *T. dimorphon, T. nanum, T. confusum* and *T. pecorum* were used by different workers to describe the trypanosome. Frustration concerning the exact identity of what would be called *T. congolense* is illustrated by the statement of Montgomery and Kinghorn (1908) to the effect that `. . . the confusion in nomenclature is appalling, and the number of specific or suggested names, based largely upon the country of origin or the first found host rather than upon morphological or biological characteristics, renders absolute diagnosis . . . almost impossible . . . .'

MORPHOMETRY

Evidence concerning phenotypes of *T. congolense* was initially based primarily on morphological variation among different isolates as could be determined from biometrical observations—mainly length of the organism (Hoare, 1970; Stephen, 1986). The question of whether the *T. congolense* parasites of different mensural characteristics could be regarded as one or different species continued well into the early 1960s without much of a resolution. What became widely accepted was that among *T. congolense* there exists a number of morphological and biological strains, types, varieties and demes, each with different biological behaviour, and that in the future they would be properly classified based not upon morphology but on biochemical and immunological properties (Stephen, 1986). Thus, for lack of better methods of parasite characterization, Mackenzie and Boyt (1969) described a strain designated simply as `A', whose identity could not be established unequivocally because (1) it did not cause the fatal disease in pigs as would *T. simiae,* but instead caused only chronic infection and (2) it was morphologically and morphometrically identical to *T. congolense;* however, it proved to be non-infective for all other domestic and laboratory animals. It was later shown that the tsetse species transmitting a trypanosome may influence its virulence to the host (Janssen and Weijers, 1974). Thus, a trypanosome such as the one described by Mackenzie and Boyt (1969) could have been a *T. simiae* or *T. congolense* whose virulence was influenced by the tsetse species used in the study. Godfrey (1982) proposed that *T. congolense* should be regarded as a ` . . . collection of diverse organisms . . . not just a species within the subgenus *Nannomonas* . . . .'

It is clear from these early studies that the classification of organisms based on morphology and morphometry alone has severe limitations, and that such studies would not lead to an understanding of the true nature of these particular parasites. Nonetheless, a consensus could be reached for the convenience of clinical veterinarians that *T. congolense* comprised at least three strains separable by mean lengths and proportion of short and long forms present in an infection, a property which was thought to correlate with parasite virulence (Godfrey, 1961). *Trypanosoma congolense* were then classified as: (a) *congolense*-type or the short-type, (b) intermediate-type and (c) dimorphon-type or the long-type (Godfrey, 1960). These differences in morphology appeared to correlate somewhat with the course of infection in the host (Godfrey, 1960, 1961).

ISOENZYME PROFILES

With the refinement of methods for analysis of micro-organisms using isoenzyme electrophoresis, studies were undertaken in an attempt to clarify the situation. A study performed on numerous isolates from West and Central Africa essentially separated *T. congolense* parasites into two groups distinct from each other by isoenzyme patterns. These were designated the West African riverine/forest-type and the savannah-type (Young and Godfrey, 1983). It was proposed in these studies (i) that the different types of *T. congolense* were confined to certain ecological zones by the types of tsetse vectors that inhabited such zones, (ii) that certain species of the tsetse vector were better than others at transmitting some types of *T. congolense* and (iii) that such tsetse inhabited only certain ecological niches. From other data obtained much later, it appears that this is not always the case since the trypanosomes, at least the Kilifi-type *T. congolense*, have been observed in places other than the ones in which they were originally found (Majiwa and Otieno, 1990; Nyeko *et al.*, 1990).

MOLECULAR KARYOTYPES AND SERODEMES

More concrete data concerning variability of *T. congolense* were obtained from studies employing biochemical/molecular biological techniques of chromosome separation and repetitive DNA hybridization. The technique of pulsed-field gel electrophoresis (PFGE) (Schwatz and Cantor, 1984) was applied to the analysis of *T. brucei* (Van der Ploeg *et al.*, 1984) and later to *T. congolense* (Majiwa *et al.*, 1985) chromosomes. The comparative approach, where the chromosomes of trypanosome clones from different isolates are compared with each other, has proved to be very informative. Using this approach, it was observed that trypanosomes generally classified as *T. congolense* have molecular karyotypes that differ significantly from those of the *Trypanozoon* trypanosomes and that *T. congolense* comprises parasites with at least two significantly different molecular karyotypes. The two were therefore described as belonging to two karyotypic groups, although it was not possible by this method to differentiate the West African riverine/forest and *T. simiae* from the savannah-type. Among members of a karyotypic group, there was chromosomal variation which appeared to correlate with antigenic repertoire. From these observations, a proposition was made that *T. congolense* clones of a single antigenic repertoire have essentially similar molecular karyotypes (Majiwa *et al.*, 1986). This proposition was verified experimentally by combining the PFGE separation of chromosomes of *T. congolense* with the classical serological methods of serodeme analysis (Masake *et al.*, 1988), which led to the observation that recovery sera from an animal infected with a *T. congolense* clone of a particular molecular karyotype completely neutralizes other *T. congolense* clones or stocks with identical molecular karyotypes. This observation does not hold true for members of the *Trypanozoon* subgenus.

HIGHLY REPETITIVE DNA SEQUENCES

The two karyotypic groups of *T. congolense* were shown to be genetically different from each other by the possession of different repetitive DNA sequences, and different sequences

of kinetoplast minicircle DNA sequences (Majiwa *et al.,* 1985). The analyses of the predominant repetitive DNA sequences was extended to the West African riverine/forest type *T. congolense* and indeed it was shown to have repetitive DNA of a different nucleotide sequence (Gibson *et al.,* 1988).

By 1988, *T. congolense* could be viewed as comprising three distinct genotypic groups which probably correlate with what were designated strains in earlier studies: (1) the savannah-type, (2) the West African riverine/forest-type, and (3) the Kilifi-type. The availability of cloned repetitive DNA sequences, more sensitive and rapid assays employing specific oligonucleotide primers and the polymerase chain reaction (PCR) (Saiki *et al.,* 1988) for the detection of trypanosomes in both tsetse vectors and mammalian hosts (Moser *et al.,* 1989), and improved methods for the *in vitro* cultivation of trypanosomes directly from either the tsetse vector (McNamara and Snow, 1991) or the vertebrate hosts have facilitated investigations on natural infections among tsetse vectors and the mammalian hosts. Such studies have indicated the presence of other trypanosomes belonging to the *Nannomonas* subgenus but which are neither *T. simiae* nor any of the known *T. congolense* types. Since there are classically only two accepted species within this subgenus, such trypanosomes have been called one type or another of *T. congolense.* At least five different genotypic groups of *T. congolense* based upon (1) isoenzyme polymorphisms and (2) repetitive DNA sequences have been described thus far. To this list can be added two new types of *T. congolense:* the Tsavo-type (Majiwa *et al.,* 1993) found in East Africa, and a `new' *Nannomonas* (McNamara *et al.,* 1991) found in a relic forest in the Gambia.

CONCLUSION: THE EMERGING TECHNOLOGIES AND THE PHYLOGENETIC RELATIONSHIPS AMONG PROTOZOAN PARASITES

It would appear that there may be more types of *T. congolense* and that these will be discovered during the application of the sensitive and specific reagents in epidemiological or field-based investigations. These differences must have a basis in the parasite genome. It will be necessary to understand the phylogenetic relationships among these morphologically identical but genotypically variable trypanosomes and to determine how they relate to each other at the genetic and population levels in order to fully appreciate the genetic population structure of *T. congolense,* as an essential part of integrated control of diseases caused by livestock-infective trypanosomes. It remains to be seen what solutions the emerging technologies will bring to this problem. One of the best and possibly most effective ways in which these problems can be approached is by the global comparative analysis of the genomes of the parasites.

REFERENCES

BALFOUR, A. 1906. Trypanosomiasis in the Anglo-Egyptian Sudan. *Report of Wellcome Tropical Research Laboratory, Khartoum* 2: 113.

BRODEN, A. 1904. Les infections a trypanosomes au Congo chez l'homme et les animaux (communication preliminaire). *Bulletin de la Société d'Etudes Coloniales,* pp. 116–139.

BRODEN, A. 1906. Rapport sur les travaux du laboratoire medicale de Leopoldville. *Bulletin de l'Academie Royale de Medécine de Belgique* 20: 387.

DUTTON, J.E. and TODD, J.L. 1903. Gland puncture in trypanosomiasis compared with other methods of demonstrating the presence of parasites. *Fourth Interim Report, Expedition of the Liverpool School of Tropical Medicine to the Congo.*

GIBSON, W.C., DUKES, P. and GASHUMBA, J.K. 1988. Species-specific DNA probes for the identification of African trypanosomes in tsetse flies. *Parasitology* 97: 63–73.

GODFREY, D.G. 1960. Types of *T. congolense.* I. Morphological differences. *Annals of Tropical Medicine and Parasitology* 54: 428.

GODFREY, D.G. 1961. Types of *T. congolense.* II. Differences in the course of infection. *Annals of Tropical Medicine and Parasitology* 55: 154.

GODFREY, D.G. 1982. Diversity within *Trypanosoma congolense.* In: Baker, J.R., ed. *Perspectives in Trypanosomiasis Research.* London: John Wiley, pp. 37–46.

HOARE, C.A. 1970. Systematic description of the mammalian trypanosomes in Africa. In: Mulligan, H.W., ed. *The African Trypanosomiases.* London: Ministry of Overseas Development, pp. 24–59.

JANSSEN, J.A.H.A. and WEIJERS, D.J.B. 1974. *Trypanosoma simiae* at the Kenya coast. A correlation between virulence and the transmitting species of *Glossina. Annals of Tropical Medicine and Parasitology* 68: 5–19.

MacKENZIE, P.K.I. and BOYT, W.P. 1969. Notes upon a trypanosome strain resembling *T. congolense* apparently adapted to the porcine species. *British Veterinary Journal* 125: 414–421.

MAJIWA, P.A.O. and OTIENO, L.H. 1990. Recombinant DNA probes reveal simultaneous infection of tsetse flies with different trypanosome species. *Molecular and Biochemical Parasitology* 40: 245–254.

MAJIWA, P.A.O., HAMERS, R., van MEIRVENNE, N. and MATTHYSSENS, G. 1986. Evidence for genetic diversity in *Trypanosoma (Nannomonas) congolense. Parasitology* 93: 291–304.

MAJIWA, P.A.O., MAINA, M., WAITUMBI, J.N., MIHOK, S. and ZWEYGARTH, E. 1993. *Trypanosoma (Nannomonas) congolense:* molecular characterization of a new genotype from Tsavo, Kenya. *Parasitology* 106: 151–162.

MAJIWA, P.A.O., MASAKE, R.A., NANTULYA, V.M., HAMERS, R. and MATTHYSSENS, G. 1985. *Trypanosoma (Nannomonas) congolense:* identification of two karyotypic groups. *European Molecular Biology Organization Journal* 4: 3307–3313.

MASAKE, R.A., NYAMBATI, V.M., NANTULYA, V.M., MAJIWA, P.A.O., MOLOO, S.K. and MUSOKE, A.J. 1988. The chromosome profiles of *Trypanosoma congolense* isolates from Kilifi and their relationship to serodeme identity. *Molecular and Biochemical Parasitology* 30: 105–112.

McNAMARA, J.J and SNOW, W.F. 1991. Improved identification of *Nannomonas* infections in tsetse-flies from the Gambia. *Acta Tropica* 48: 127–136.

McNAMARA, J.J., GIBSON, W.C., MOHAMED, G., SNOW, W.F. and GODFREY, D.G. 1991. New technology, old protozoology and a new species of *Nannomonas. Transactions of the Royal Society for Tropical Medicine and Hygiene* 85: 695.

MONTGOMERY, R.E. and KINGHORN, A. 1908. A report on the trypanosomiasis of domestic stock in North-West Rhodesia. *Annals of Tropical Medicine and Parasitology* 2: 97.

MOSER, D.R., COOK, G.A., OCHS, D.E., BAILEY, C.P., McKANE, M.R. and DONELSON, J.E. 1989. Detection of *Trypanosoma congolense* and *Trypanosoma brucei* by DNA amplification using the polymerase chain reaction. *Parasitology* 99: 57–66.

NYEKO, J.H.P., ole-MOIYOI, O.K., MAJIWA, P.A.O., OTIENO, L.H. and OCIBA. P.M. 1990. Characterization of trypanosome isolates from Uganda using species-specific DNA probes reveals predominance of mixed infections. *Insect Science and its Application* 11: 271–280.

SAIKI, R.K., GELFAND, D.H., STOFFEL, S., SCHARF, S.J., HIGUCHI, R., HORN, G.T., MULLIS, K.B. and ERLICH, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239: 487–491.

SCHWATZ, D.C. and CANTOR, C.R. 1984. Separation of yeast chromosome-sized DNAs by pulsed-field gradient gel electrophoresis. *Cell* 37: 67–75.

STEPHEN, C.A. 1986. *Trypanosomiasis: A Veterinary Perspective.* Oxford: Pergamon Press, 551 pp.

Van der PLOEG, L.T.H., SCHWATZ, D.C., CANTOR, C.R. and BORST, P. 1984. Antigenic variation in trypanosomes analysed by electrophoretic separation of chromosome-sized DNA molecules. *Cell* 37: 77–84.

YOUNG, C.J. and GODFREY, D.G. 1983. Enzyme polymorphism and the distribution of *Trypanosoma congolense* isolates. *Annals of Tropical Medicine and Parasitology* 77: 467–481.

Genetic recombination and karyotype inheritance in *Trypanosoma brucei* species

A. Tait, N. Buchanan, G. Hide and M. Turner

This paper reviews recent research on the genetic analysis of *Trypanosoma brucei* spp. Crosses between seven different stocks have been undertaken generating a total of 109 hybrid progeny clones allowing some of the main features of this genetic system to be defined. Five main aspects are considered, namely: the frequency and stage of mating, the inheritance of phenotypic and genotypic markers, the inheritance of karyotype, the genetics of the kinetoplast and mating types and self-fertilization.

The available data show that mating is non-obligatory and occurs during the life cycle stages in the tsetse fly. Analysis of the progeny produced by infecting tsetse flies with two cloned parental stocks using phenotypic and genotypic markers shows that they are equivalent to an F_1 in a typical diploid mendelian system. Evidence that such F_1 progeny are produced by the fusion of haploid nuclei arising from meiosis of the parental diploid nuclei is provided by the segregation and recombination of alleles at several loci for which one of the parental clones is heterozygous. Although the majority of F_1 progeny clones conform to this conventional mendelian model, a proportion have a raised DNA content which has been shown to be due to trisomy for several chromosomes (presumably arising from non-disjunction at meiosis in one of the parental clones).

Analysis of the karyotype of the progeny clones by pulsed-field gel electrophoresis and the identification of homologous chromosomes using chromosome specific gene probes shows that each pair of homologues segregates into the progeny clones, producing recombinant karyotypes, and that some chromosomes change in size by as much as 500 kb. Thus recombination at a karyotype level occurs both by independent assortment of non-homologous chromosomes and by the generation of novel size chromosomes possibly derived by homologous recombination. The question of whether all stocks of *T. brucei* can mate or whether mating barriers exist is discussed and recent results demonstrating that crossing can occur within a cloned stock (selfing) are presented. These results suggest that the mating system of trypanosomes may be complex although the question of whether mating types exist remains open.

INTRODUCTION

The question of whether a process of genetic exchange exists in African trypanosomes has long been the subject of controversy (Tait, 1983) and evidence based on enzyme variation in natural populations (Tait, 1980; Gibson *et al.*, 1980) has subsequently been questioned on the basis of high levels of linkage disequilibrium and other criteria (Cibulskis, 1988; Tibayrenc *et al.*, 1990). However, the direct demonstration of genetic exchange between trypanosome stocks (Jenni *et al.*, 1986) of *Trypanosoma brucei* puts this question beyond doubt, although the contribution of mating to variation in natural populations of *T. brucei*, *T. b. rhodesiense* and *T. b. gambiense* is still a matter of debate which will not be considered here.

The ability to cross stocks of *T. brucei* in the laboratory has allowed the investigation of some of the basic features of the trypanosome genetic system. To date a total of six crosses have been made and 109 progeny clones have been analyzed (Table 1). Based on

the origins of the stocks in terms of host and geographical region, it is clear that genetic exchange can occur between *T. b. rhodesiense* and *T. b. brucei,* between stocks from East and West Africa and between stocks within East Africa. In this paper, we will review the current state of knowledge about this system based on the published work and some of the more recent work from our own laboratories.

FREQUENCY AND STAGE OF MATING

All crosses have been made by feeding tsetse flies on mixtures of two cloned stocks at the bloodstream stage then allowing the infection to develop through the stages in the fly followed by sampling the metacyclic stage trypanosomes from the salivary glands (Tait and Turner, 1990). Either by cloning the extruded metacyclics directly or by cloning trypanosomes from the bloodstream infections resulting from tsetse feeding, it has been shown by marker analysis that some of the trypanosome clones are hybrid and therefore the products of mating. The frequency of mating can be analyzed by determining the phenotype or genotype of either the population of metacyclics extruded by each of a series of mixed infected flies or a series of clones derived from a single fly. In the first approach, flies can be scored as producing either parental, mixtures of parental or

Table 1. Crosses between cloned stocks of *T. brucei.*

Parental Clones				Progeny Clones		
P1	Origin	P2	Origin	Total	Hybrid	Parental
STIB 247	Hartebeest Tanzania	STIB 386	Man Ivory Coast*	38	17	21
STIB 247	Hartebeest Tanzania	TREU 927/4	Tsetse Kenya†	21	10	11
STIB 386	Man Ivory Coast	TREU 927/4	Tsetse Kenya†	18	9	9
STIB 247	Hartebeest Tanzania	STIB 777	Tsetse Uganda‡	4	4	0
J10	Hyena Zambia	196	Pig Ivory Coast§	14	9	5
058	Man Zambia	196	Pig Ivory Coast¶	14	12	2

P1 and P2 refer to the two parental cloned stocks used in each cross and the origin refers to the host and geographical location from which the original stock was isolated. Data from *Jenni *et al.*, 1986; *Sternberg *et al.*, 1989; †Turner *et al.*, 1990; ‡Schwizer *et al.*, 1993; ¶Gibson *et al.*, 1991;*Gibson, 1989).

hybrid trypanosomes; the results obtained are summarized in Table 2 and show that mating is non-obligatory as some flies only produce one or other parental type while others produce hybrid trypanosomes. Furthermore, analysis of the phenotype of the metacyclic population from individual tsetse flies with time shows that initially both parental types are produced with no mating and hybrids are subsequently produced, thus further emphasizing the non-obligatory nature of mating. The results of all published work using the second approach are summarized in Table 1. Here, individual infected flies have been analyzed (although the data presented are a summation of the results from several flies) and it has been shown that each fly produces both hybrid and parental trypanosomes. Thus from both types of analysis it is clear that mating is non-obligatory in the life cycle of the trypanosome.

From these experiments, it is presumed that mating takes place at some stage during the trypanosome life cycle in the insect vector i.e. at the procyclic, epimastigote or metacyclic stages, although mating between the bloodstream stages as they enter the tsetse fly cannot be formally excluded. The fact that metacyclic clones are clearly hybrid when analyzed by phenotypic or genotypic markers suggests, but does not prove, that this stage is the product of mating and implies that mating occurs at an earlier stage. Analysis of the electrophoretic enzyme phenotype of procyclics derived from the midgut of mixed infected flies which are producing hybrid metacyclics has shown iso-enzyme patterns which suggest the presence of hybrid trypanosomes (Schweizer *et al.*, 1991). Similar experiments using *in vitro* mixtures of procyclic stage trypanosomes have also shown iso-enzyme patterns consistent with the occurrence of mating at this stage of the life cycle (Schweizer *et al.*, 1991). However, no cloned hybrid trypanosomes have been derived from the procyclic stage and therefore the question of the stage at which mating takes place remains open and requires further investigation.

Table 2. Metacyclic population phenotype of tsetse flies infected with two stocks.

Phenotype	No. of tsetse
P1	23
P2	8
P 1 + P2	5
H	12

These data were obtained by screening the trypanosome population using an electrophoretic enzyme marker for which both parental stocks were homozygous but different; hybrids are detected by their heterozygous enzyme electrophoretic pattern. P1 indicates parental type 1; P2 indicates parental type 2; P1 + P2 indicates a mixture of both parental types; H indicates F_1 hybrids progeny (Turner et al., 1990).

INHERITANCE OF PHENOTYPIC AND GENOTYPIC MARKERS

As the stage at which mating takes place has not been established and reagents are not available for defining the genotype or phenotype of individual cells, it is not possible to examine the mating events morphologically. Our knowledge of the mechanism of genetic exchange has relied on the analysis of markers (genotypic and phenotypic) which distinguish the two stocks used to infect flies. Analysis of the segregation and recombination of such markers in cloned trypanosomes derived from mixed infected flies has allowed the elucidation of some of the features of the system.

Considering markers for which each parental stock is homozygous but different, the progeny clones show heterozygous patterns consistent with such clones being the F_1 progeny of mating between the parental stocks. An example of the data using an isoenzyme (malic enzyme-B) and a gene probe (phosphoglycerate kinase-PGK) for the two parental stocks, STIB 247 and TREU 927/4, and the progeny clones derived from them is illustrated in Figure 1. Such heterozygous patterns using a number of markers have been shown in 61 progeny clones out of the 109 derived from all crosses screened for markers for which both parents are homozygous but different, the remaining 48 clones are of one or other parental type.

These hybrid progeny clones thus show the phenotype and genotype predicted for the F_1 of a classical diploid mendelian system. In order to provide evidence that meiosis occurs, a backcross or F_2 should be undertaken to demonstrate segregation of alleles and recombination between loci; to date such crosses have not been undertaken. An equivalent analysis (in genetic terms) of a backcross has been undertaken by typing the F_1 hybrid progeny clones for markers for which one parental stock is heterozygous and the other is homozygous (Tait and Turner, 1990; Sternberg *et al.*, 1989; Turner *et al.*, 1990; Sternberg and Tait, 1990; Gibson *et al.*, 1991). The data obtained from such an analysis is illustrated, in general terms, in Table 3 and show that there is segregation of alleles at each locus and recombination between loci leading to recombinant non-parental genotypes (1–2, ++ or 2, +–). Such analysis has been extended to several loci and in almost all cases segregation and recombination have been observed. Overall, marker analysis of the progeny clones supports the existence of a mendelian diploid genetic system involving meiosis leading to allele segregation and recombination.

Three hypotheses have been put forward to explain these results, firstly the fusion of the two parental stocks to yield tetraploid progeny (Paindavoine, 1986; Wells *et al.*, 1987), secondly meiosis of the parental nuclei followed by fusion of the resulting haploid products to yield diploid progeny (Sternberg *et al.*, 1989; Turner *et al.*, 1990) and thirdly fusion of the parental nuclei followed by meiosis to yield diploid progeny (Gibson, 1989; Tait and

Table 3. Phenotype and genotype of hybrid progeny clones at loci where one parental stock is heterozygous.

Clone		Enzyme	Gene probe
Parental	P1	1–2	+–
	P2	2	++
Progeny	Class 1	1–2	+–
	Class 2	1–2	++
	Class 3	2	+–
	Class 4	2	++

The enzyme phenotype is indicated by numbers where 2 indicates a single electrophoretic band of activity (homozygous) and 1–2 indicates a multiple banded activity (heterozygous) derived from two alleles 1 and 2. The polymorphism detected by a gene probe is indicated as + for the presence of a restriction site and – for its absence, thus a heterozygous stock is indicated as +– and a homozygous stock as ++. These patterns and the nomenclature are illustrated in Figure 1.

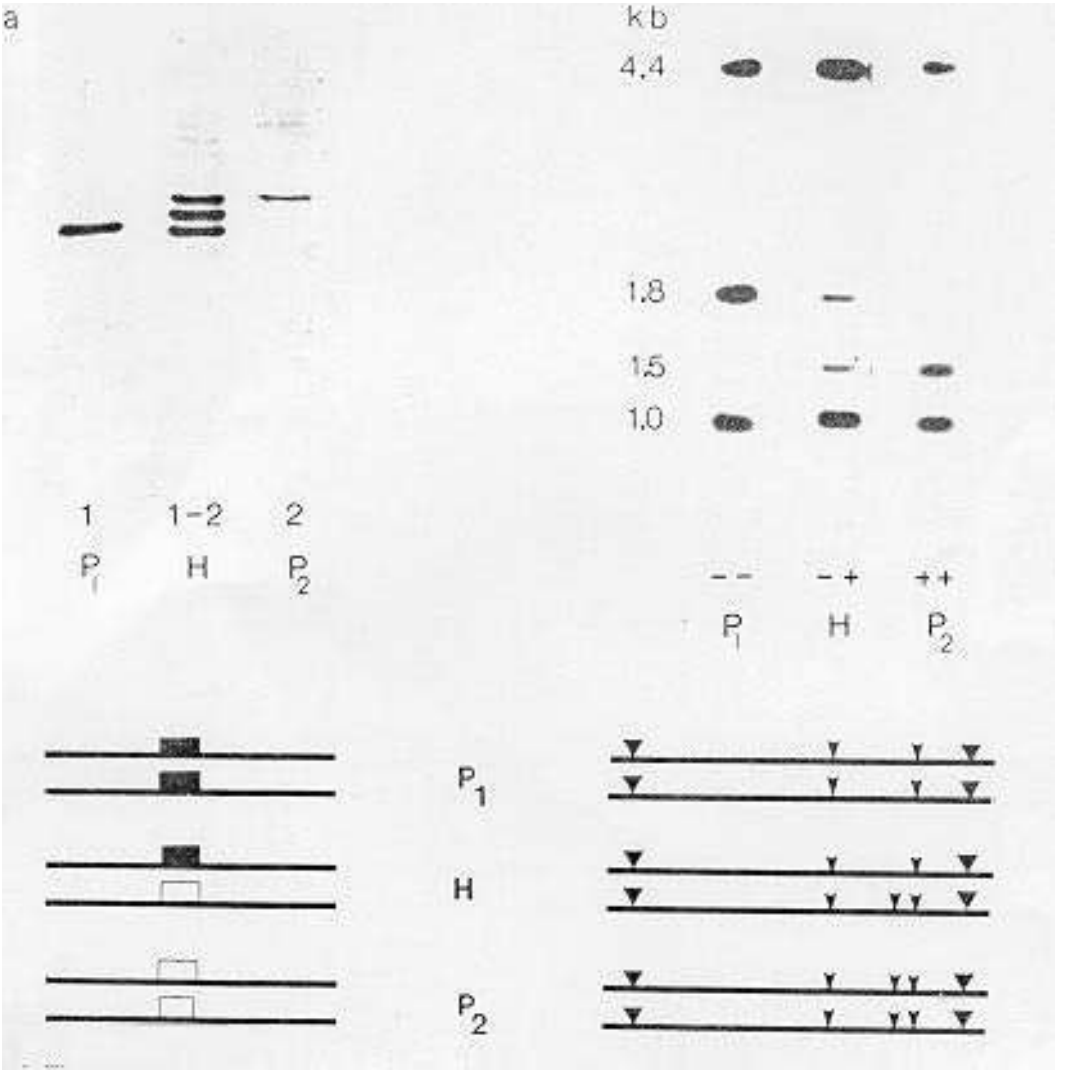

Figure 1. The genotype and phenotype of parental and hybrid trypanosomes. The left panel shows the enzyme electrophoretic pattern stained for malic enzyme (ME_B); the two parental stocks (P1—TREU 927/4; P2-STIB 247) give single-banded patterns which differ in mobility while the hybrid, F_1 progeny clones (H) give a three-banded pattern, typical of a heterozygote. The right hand panel shows a Southern blot of agarose gel separated total genomic DNA digested with *Pst*I and probed with a phosphoglycerate kinase cDNA probe (PGK) using genomic DNA from the same stocks. The size of the detected fragments is given in kilobases (kb). Below each panel is a diagrammatic representation of the genotype of the three stocks.

Turner, 1990; Sternberg and Tait, 1990). In order to distinguish between the first and other hypotheses, the DNA content of progeny clones have been measured and a summary of the data is given in Table 4. The majority (21/24) of the progeny clones have DNA contents equal to or between the parental values and three progeny clones have raised DNA contents. Thus the major class of progeny clones could not be derived by fusion of the

Table 4. DNA contents of hybrid progeny clones

Clone	DNA content (mean)	Reference
Parental		
247	1.01	Wells *et al.*, 1987;
386	1.00	Le Page *et al.*, 1988
Hybrid		
(i) 723 vi-L	1.51	
(ii) 9 metacyclic clones	1.04 (1.09-0.94)	
Parental		
247	1.08	Kooy, 1991
927	0.9	
Hybrid		
(i) 2 bloodstream clones	0.96 (1.01-0.91)	
(ii) 5 metacyclic clones	0.96 (1.01-0.85)	
Parental		
058	2n	Gibson *et al.*, 1991
196	2n	
Hybrid		
(i) 5 bloodstream clones	2n	
(ii) 2 bloodstream clones (W3 and W5)	3n	

The DNA content figures refer to the DNA content (measured by microfluorimetry) and flow cytometry relative to an internal DNA standard. The mean for each group of clones has been taken. The values are given relative to those of the parental stocks which are interpreted as having a DNA contact of twice the haploid value (2n) (Gibson *et al.*, 1991).

diploid parental stocks but a small but significant class show raised DNA contents. One possible explanation of these observations is that meiosis is imperfect leading to the non-disjunction of some paired chromosomes resulting in progeny clones that are triploid for these chromosomes (Tait and Turner, 1990; Sternberg and Tait, 1990). Evidence for triploidy has been provided by molecular karyotype analysis of the progeny clones W3 and W5 (Gibson *et al.*, 1991); the molecular karyotype of the hybrid progeny clone 723vi-L has not been determined.

INHERITANCE OF KARYOTYPE

The chromosome of *T. brucei* can be separated by pulsed-field gel electrophoresis (PFGE); three `classes' of chromosome have been identified, the mini-chromosomes (50–100 kb), the intermediate size chromosomes (150–800 kb) and the large chromosomes (>900 kb). Using a range of separation conditions and a series of gene probes 80% of the chromosomes have been separated and identified and some 19 bands of DNA resolved (Van der Ploeg *et al.*, 1989). Further studies identified 14 of these bands as 7 pairs of homologous

chromosomes, each pair differing by as much as 20% in size (Gottesdiener *et al.*, 1990). The ability to define and separate pairs of homologous chromosomes, as well as the intermediate and mini chromosomes, allows the determination of the inheritance of these chromosomes in the progeny clones from crosses.

The single study on the mini-chromosomes suggests that not all these chromosomes are inherited by the hybrid progeny and that novel size chromosome bands are generated (Wells *et al.*, 1987). Analysis of the intermediate chromosomes shows that they are not inherited in a strictly mendelian manner (Gibson, 1989; Le Page *et al.*, 1988) with some hybrids inheriting all parental chromosomes, others lacking any contribution from one parent and others inheriting a variable number of intermediate chromosomes from either parent. These results suggest that the intermediate chromosomes are not segregating as diploid homologues and therefore could be considered as haploid chromosomes inherited in a random fashion.

The inheritance of the larger chromosomes has been studied in more detail using a combination of PFGE and the identification of specific chromosomes by probing blots with single or closely linked multiple gene probes (Gibson *et al.*, 1991; Gibson, 1989; Gibson and Garside, 1991). Analysis of the hybrid progeny clones from the cross between stocks J10 and 196 showed that they had recombinant non-parental karyotypes on the basis of ethidium bromide staining of pulsed-field gels (Gibson, 1989) and this was also shown in the progeny clones of another cross (stock 058 ⟩ stock 196) (Gibson and Garside, 1991). Similar results have been obtained in our own laboratory and a sample of the data is shown in Figure 2 for the progeny of a cross between STIB 247 and STIB 386. All of the six hybrid progeny clones show non-parental karyotypes (F9/45 clones 1, 11, 10, 9, 7, 2) when pulsed-field gels are stained with ethidium bromide. While most of the chromosome bands can be accounted for as having been derived from one or other parent, two progeny clones (9 and 10) show a novel non-parental chromosome band at a size of approximately 1.2 Mb.

Analysis of blots of such pulsed-field gels with gene probes has shown that homologous chromosomes differ in size by several hundred kilobases. The inheritance of such homologues in the progeny of crosses shows that one homologue is inherited from each parent. This has been shown for the cross between stocks 058 and 196 for the chromosomes carrying the β-tubulin and PGK loci (Gibson, 1989) and similar results from our own laboratory are shown in Figure 2 for PGK. This probe hybridizes to two similar sized chromosomes in stock STIB 386 and to two resolved chromosomes in stock STIB 247; the six progeny clones show inheritance of alternative single homologues from each parent producing all four possible recombinants (ac, ad, bc or bd, Figure 2). This result convincingly shows chromosome segregation and recombination, providing further direct evidence for the occurrence of meiosis. A further two crosses have been analyzed using markers for pairs of homologous chromosomes and have given similar results (A. Tait, N. Buchanan, G. Hide and C. Turner, unpublished data).

While the above results establish a general picture of a normal diploid mendelian genetic system, a number of less conventional findings have been made by analysis of the karyotype of the hybrid progeny clones from crosses, although none of these observations contradict the evidence for meiosis. It is clear from the data of Gibson and colleagues (Gibson *et al.*, 1991) that some hybrid progeny clones can be trisomic for a number of chromosomes as discussed earlier. Novel non-parental size chromosomes are

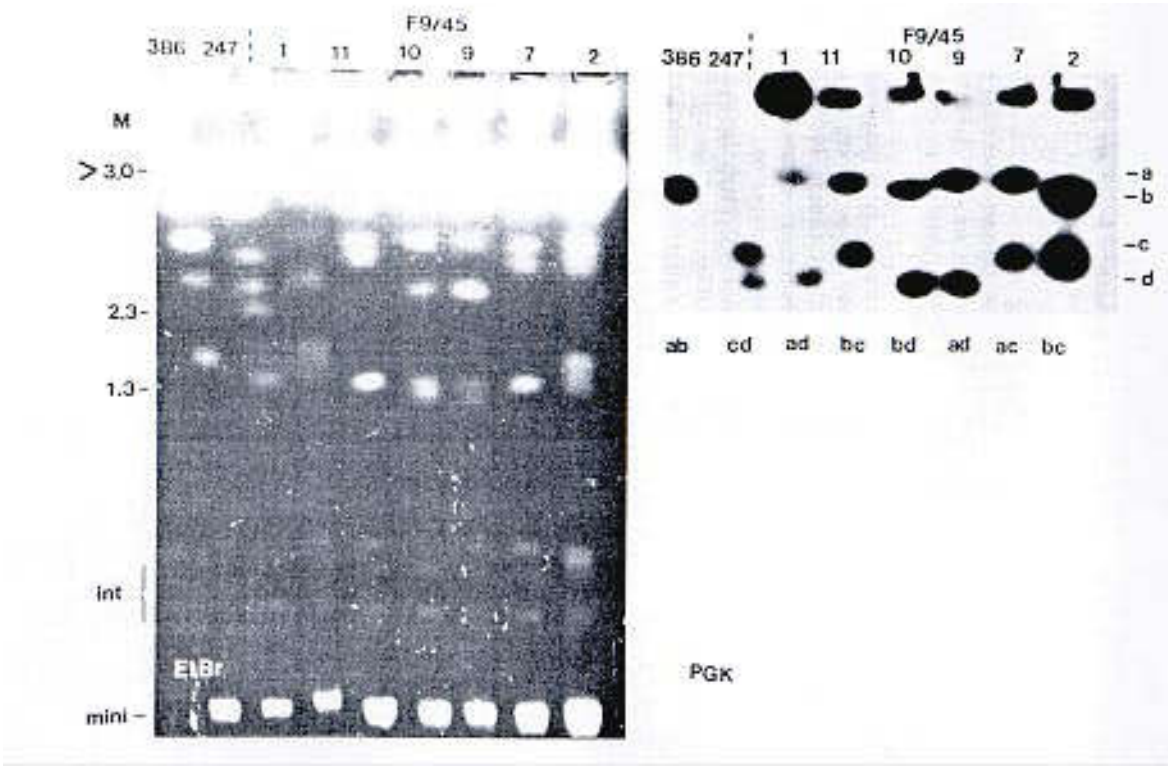

Figure 2. The karyotype of the parental and F_1 progeny clones from a cross between STIB 247 and STIB 386. The left hand panel shows the ethidium bromide-stained pattern of the chromosomes after separation by PFGE while the right hand panel shows a Southern blot of the same gel probed with the PGK gene probe. The sizes of the DNA bands are indicated in megabases (Mb) determined by running a standard *Saccharomyces cerevisiae* chromosome preparation. The F_1 progeny clones (F9/45 mCl 2, 4, 7, 9, 10, 11) are described in Sternberg *et al.*, 1989. M-mini-chromosomes; int—intermediate chromosomes.

observed in the progeny clones when two pairs of homologous chromosomes are examined (detected respectively by the phospholipase C gene probe or by the PGK and β-tubulin gene probes) and it has been suggested that such size changes are not specifically associated with genetic exchange (Gibson, 1989; Gibson *et al.*, 1991; Gibson and Garside, 1991). We have undertaken a similar study using gene probes which recognize five pairs of homologous chromosomes ranging in size from 1.5 Mb to >3.5 Mb and have analyzed the progeny clones from two crosses in detail (STIB 386 × STIB 247 and STIB 247 × TREU 927/4). We have not detected any size changes in the chromosomes detected by PGK (Figure 2) or those detected by a protein kinase gene KIN-2 (J.C. Mottram, unpublished data). However size changes have been observed in the pair of homologues detected by the cysteine proteinase gene probe (Mottram *et al.*, 1989) (CP) and in the two pairs of homologues detected by the RNA polymerase II gene probe (Evers *et al.*, 1989). An example of these data is illustrated in Figure 3 using the CP gene probe and Southern blots of pulsed-field gels of the progeny from the cross between STIB 386 and STIB 247. The parental stocks show hybridization signals to single ethidium bromide staining bands which differ in size by ~500 kb; two of the progeny clones (Cl 2 and 9) show inheritance of one homologue from each parent while the remaining progeny clones (4, 7, 10, 11) show inheritance of the STIB 247 parental homologue but no chromosomal band of equivalent size to the STIB 386 parental homologue, instead a smaller chromosome of non-parental size is observed. This analysis presumes that the two homologues detected by CP are of the same size in each parental stock. If one homologue is inherited from each parent in the progeny clones this would suggest that the STIB 386 parental homologue has undergone a reduction in size of a similar degree in each independently isolated progeny clone. Given that such size changes have been observed in a total of five pairs of homologous chromosomes, it is clear that these are frequent events. The question remains as to the mechanism by which these size changes arise: do they arise at mitosis during vegetative growth within the parental stock, do they arise at meiosis or do they arise post meiotically when homologous chromosomes of diverse origin are brought together by mating? At the present time these questions cannot be definitively answered, although Gibson *et al.* (1991) have shown a chromosomal size alteration in a progeny clone of parental genotype suggesting that size changes can occur mitotically. We have analyzed the karyotype of STIB 386 and 247 after prolonged bloodstream and procyclic passages as well as after tsetse fly transmission and have failed to observe chromosome size changes (A. Tait and C. Turner, unpublished data) suggesting that the size changes result from mating. If changes in the size of chromosomes occur at appreciable frequencies during mitosis it would be predicted that the karyotype of cloned stocks would change during passage with the result that several chromosome bands would hybridize to any particular gene probe. Such results have not been reported, favouring the hypothesis that these size alterations are generated as a consequence of mating and may provide one of the mechanisms of generating the high level of karyotype diversity observed between stocks.

A further observation concerns the linkage relationships between the loci coding for PGK, glucose phosphate isomerase (GPI) and β-tubulin; blots of pulsed-field gels suggest that in one stock these loci are located on a single linkage group while in other stocks the GPI locus is located on a separate linkage group (Gibson and Garside, 1991). These results

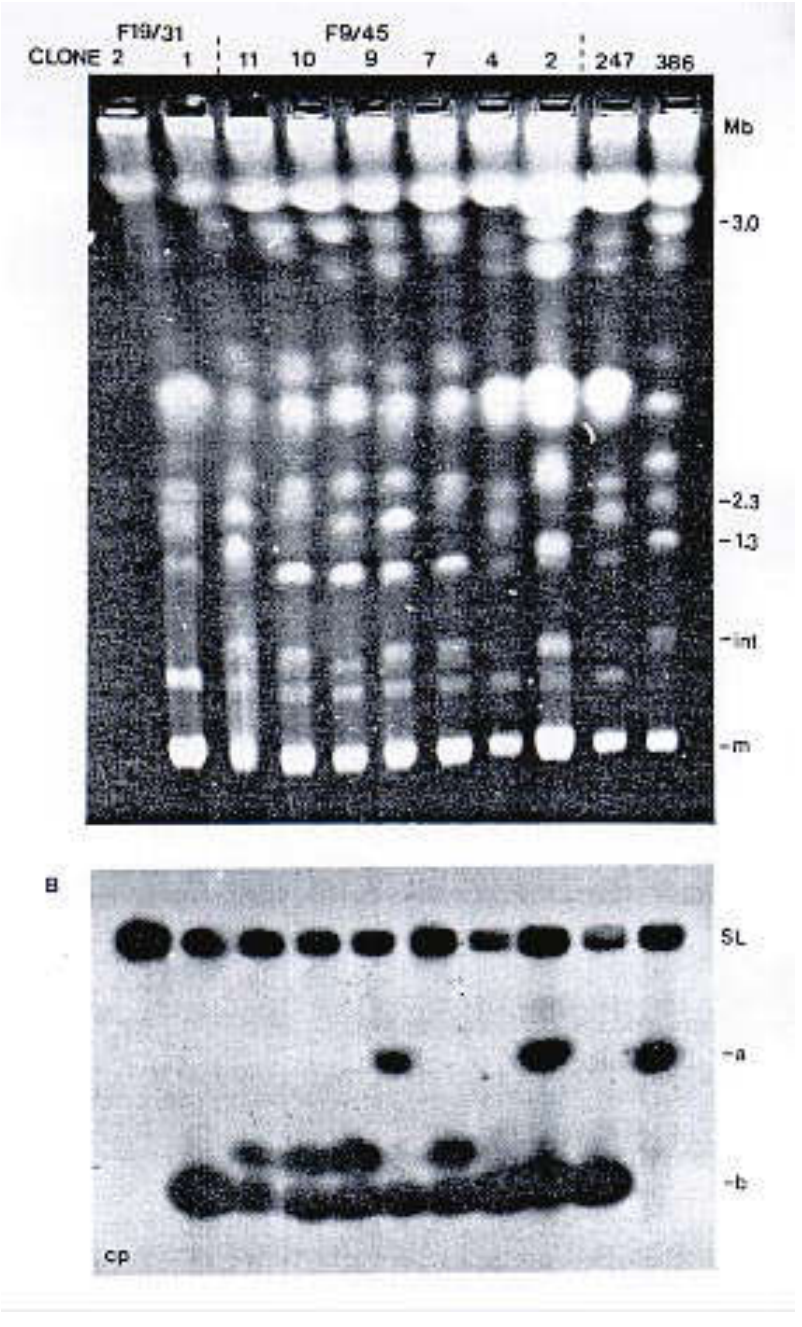

Figure 3. The karyotype of parental and F_1 progeny clones from a cross between STIB 247 and STIB 386 illustrating chromosomal size changes. The top panel (A) shows an ethidium bromide-stained pulsed-field gel separation of the chromosomes separated under longer pulse times than Figure 2. The gel was blotted and probed with the CP gene probe and the blot is shown in the lower panel (B). The parental and F_1 progeny clones are as described in Figure 2. SL=application slot.

could be explained if in one stock non-homologous chromosomes are of the same size but in the second stock differ in size. We have shown this to be the case in STIB 247 (unpublished data).

INHERITANCE OF THE KINETOPLAST

Using restriction fragment length polymorphisms (RFLPs) in the maxi circle DNA, the hybrid progeny clones from two crosses have been analyzed (Sternberg *et al.*, 1987, 1989; Gibson, 1989); the results show that the inheritance is uniparental—a single maxi-circle genotype is inherited by each progeny clone and no recombination between maxi-circle genomes has been demonstrated. The inheritance of the mini-circles has been examined in one cross (Gibson and Garside, 1990) and, in contrast to the maxi-circle, each progeny clone inherits mini-circles from both parents. These results have been interpreted as showing that when the two parental stocks (or their haploid products) fuse, the kinetoplast also fuses and the maxi-circles segregate in subsequent mitotic divisions while the mini-circles do not and remain a mixture of both parental types (Gibson and Garside, 1990). This interpretation is strengthened by the k-DNA analysis of two hybrid clones which are identical in nuclear genotype, have a mixed mini-circle genotype but differ in their maxi-circle genotype. This suggests that these two clones are the daughter cells from the same mating in which the kinetoplasts from both parents must have fused with subsequent segregation of the maxi-circle (Gibson and Garside, 1990).

MATING TYPES AND SELF-FERTILIZATION

As further crosses are undertaken, it is pertinent to ask whether all stocks can mate with each other or whether there are mating barriers determined by some system of mating types. In order to examine these questions we have undertaken three types of experiment.

We have screened metacyclic clones derived from five single stock tsetse transmissions for loci which are heterozygous in the parental stock. If self-fertilization took place one would predict that some of the progeny clones would be homozygous at the loci screened. All progeny clones derived from single stock transmissions remain heterozygous, thus suggesting that self-fertilization does not take place. Stocks STIB 247, STIB 386 and TREU 927/4 have been mixed in all three pairwise combinations and transmitted through tsetse flies (Sternberg *et al.*, 1989; Turner *et al.*, 1990); F_1 hybrid progeny are produced by all combinations suggesting that there are no barriers for mating between stocks. If this were so, one would predict that self-fertilization would take place in single stock transmissions. A possible explanation for these apparently contradictory results could be provided if a mating type system existed and at least one of the three stocks was heterozygous for alleles at a mating type locus. Such a hypothesis would predict that at least one of the three parental stocks would undergo self-fertilization.

To examine this possibility we have analyzed the progeny clones from a mixed tsetse transmission of stocks STIB 247 and STIB 386. A total of 27 metacyclic clones were analyzed for iso-enzyme markers which were homozygous but different for isocitrate

dehydrogenase (ICD) and alkaline phosphatase (AP) in the parental stocks; 22 of the progeny clones were shown to be identical to the STIB 247 parent. When ten of these clones were analyzed for an iso-enzyme marker and a RFLP (detected by probe pBE9) for which STIB 247 was heterozygous, four of the clones were homozygous for one or other marker. The molecular karyotype of each of the ten clones was also examined and four clones shown to have an altered karyotype. The alteration of karyotype occurred in three pairs of homologous chromosomes detected by specific gene probes (A. Tait, N. Buchanan, G. Hide and C.M.R. Turner, unpublished data); an example of the data for one pair of homologues is illustrated in Figure 4. Two of the clones (54 and 58), by comparison to the STIB 247 parent have lost a chromosome band (B2) but show an increase in ethidium bromide staining of a second band (B1). In the parental stocks (STIB 247), B1 and B2 hybridize to a phospholipase C gene probe (PLC) (Carrington *et al.*, 1989) suggesting that they represent a pair of homologous chromosomes (Figure 4). In the progeny clones 54 and 58 only band B1 hybridizes to the PLC probe and shows an increased intensity of hybridization which, coupled with the increased staining by ethidium bromide suggests that two copies of this homologue are present.

Overall, the results show that five of the ten `parental' progeny clones are recombinant and are homozygous for certain heterozygous markers and for certain chromosomes. Such genotypes/phenotypes are those predicted if these clones were the products of self-fertilization. The results of this analysis are summarized in Table 5 and are consistent with 247 having undergone meiosis followed by fusion of the meiotic products to yield recombinants. Whether such self-fertilization requires the presence of another stock (STIB 386) or can occur when STIB 247 is tsetse-transmitted as a single stock remains to be determined. The fact that 50% of the clones analyzed are recombinant suggest that mating may take place at a higher frequency than previously considered and the observation that a high proportion of these events involve self-fertilization suggests that the interpretation of the genetic structure of natural populations of *T. brucei* may be complex.

CONCLUSIONS AND PROSPECTS

The data obtained to date suggest that *T. brucei* has a very flexible non-obligatory genetic system involving cross and self-fertilization together with mechanisms producing high levels of homologous chromosome recombination which lead to changes in size of the chromosomes. A number of questions remain to be resolved. At what stage does mating take place? Does meiosis precede fusion or does it occur after fusion of the diploid parental cells? At what stage and by what mechanism do the high levels of homologous chromosomal recombination take place? Is there a system of mating types? Current research is aimed at answering these questions.

While understanding the mechanisms and the system of gene exchange is an important goal, a second objective is to develop a simple system of genetic analysis that could be used to determine the genetic basis of a range of phenotypes of relevance to disease such as drug resistance, virulence and fly transmissibility. The existing system is highly labour intensive and the development of an *in vitro* system using transfected selective markers would potentially circumvent many of the current problems.

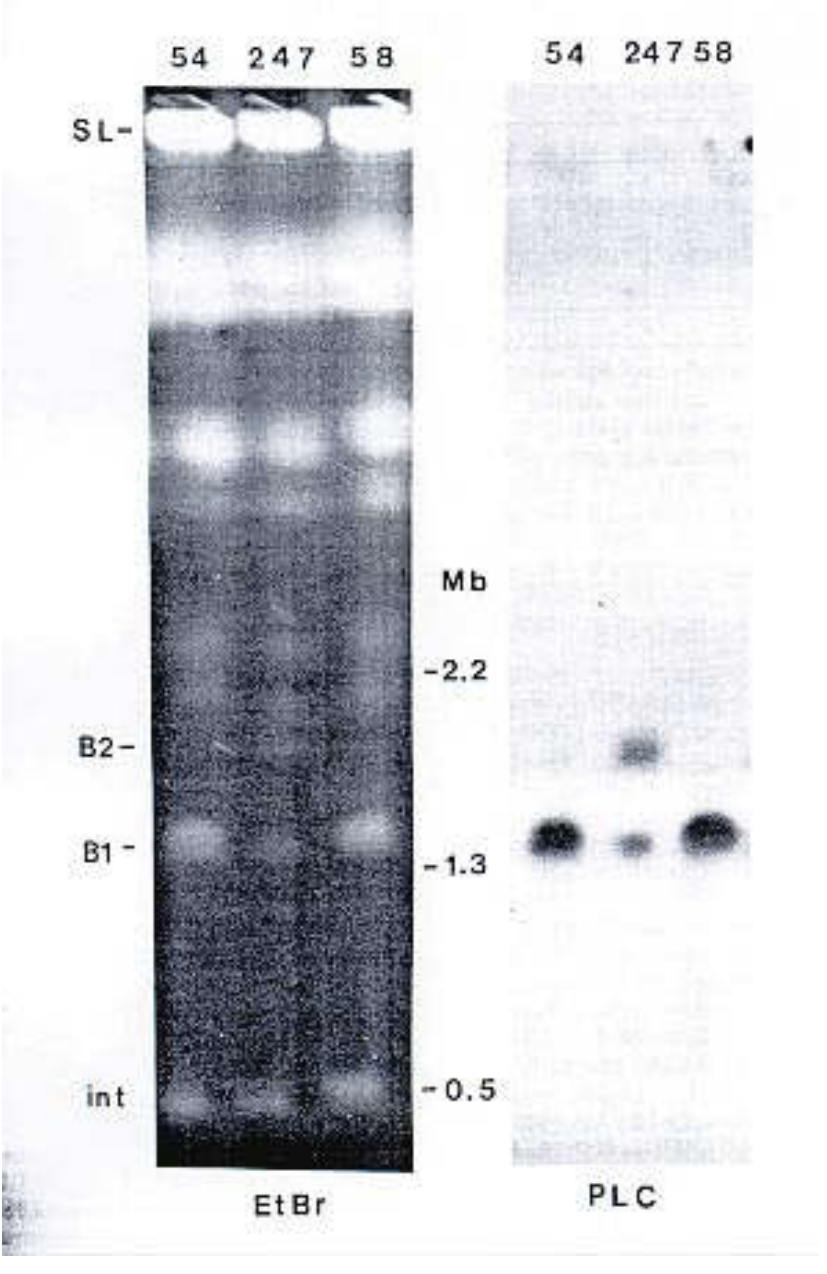

Figure 4. Karyotype of parental type clones derived from tsetse fly transmission. The left hand panel shows an ethidium stained pulsed-field gel of DNA from the parental stock, STIB 247, and two metacyclic clones (54 and 58) of parental genotype. The bands are designated as follows: int—intermediate chromosomes; B1 and B2—smallest large' chromosomes. The positions of the standard markers are indicated in megabases (Mb). The right hand panel shows a Southern blot of the pulsed-field gel probed with the PLC gene probe (A. Tait, N. Buchanan, G. Hide and C.M.R. Turner, unpublished data).

Table 5. Phenotype, genotype and karyotype of five recombinant progeny clones derived from the mixed infection of a tsetse fly with two different stocks of *T. brucei.*

Clone	Tyr^3	pBE9	1	2	3	4	7	8
54	2–4	+–	++	–	+	+	+	+
53	2–4	– –	+	+	++	–	+	+
57	2	+–	+	+	+	+	++	–
58	4	+–	++	–	+	+	+	+
35	4	nd	nd	nd	nd	nd	nd	nd
247	2–4	+–	+	+	+	+	+	+

All five progeny clones were identical to STIB247 when screened for homozygous markers. The enzyme phenotype and the genotype (Tyr^3) determined by the probe pBE9 is indicated using the nomenclature illustrated in Figure 1. The pairs of homologous chromosomes which differ in size are indicated as 1, 2; 3, 4; and 7, 8 respectively. The presence of a single copy of each is indicated as + while two copies of a single homologue are indicated as ++ and absence of a homologue by –.

ACKNOWLEDGEMENTS

We are grateful to the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (TDR) and the Wellcome Trust for financial support of work conducted in our laboratories. M. Turner is a Royal Society University Fellow.

REFERENCES

CARRINGTON, M., BULOW, R., REINKE, H. and OVERATH, P. 1989. Sequence and expression of the glycosyl-phosphatidylinositol-specific phospholipase C of *T. brucei. Molecular and Biochemical Parasitology* 33: 289–296.

CIBULSKIS, R.E. 1988. Origins and organisations of genetic diversity in natural populations of *T. brucei. Parasitology* 96: 29–57.

EVERS, R., HAMMER, A., KOCK, J., JESS, W., BORST, P., MEMET, S. and CORNELISSEN, A.W.C.A. 1989. *Trypanosoma brucei* contains two RNA polymerase II sub-unit genes with an altered C-terminal domain. *Cell* 56: 585–597.

GIBSON, W. and GARSIDE, L. 1990. Kinetoplast mini-circles are inherited from both parents in genetic hybrids of *T. brucei. Molecular and Biochemical Parasitology* 42: 45–54.

GIBSON, W.C. 1989. Analysis of a genetic cross between *T. b. rhodesiense* and *T. b. brucei. Parasitology* 99: 391–402.

GIBSON, W.C. and GARSIDE, L. 1991. Genetic exchange in *T. brucei:* variable chromosomal location of housekeeping genes in different trypanosome stocks. *Molecular and Biochemical Parasitology* 45: 77–90.

GIBSON, W.C., GARSIDE, L. and BAILEY, M. 1991. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *T. b. rhodesiense* and *T. b. brucei. Molecular and Biochemical Parasitology* 52: 189–200.

GIBSON, W.C., MARSHALL, T.F. de C. and GODFREY, D.G. 1980. Numerical analysis of enzyme polymorphisms: a new approach to the epidemiology and taxonomy of the subgenus *Trypanozoon. Advances in Parasitology* 18: 175–246.

GOTTESDIENER, K., GARCIA-ANOVEROS, J., LEE, M.G. and Van der PLOEG, L.H.T. 1990. Chromosome organisation of the protozoan *T. brucei*. *Molecular and Cellular Biology* 10: 6079–6083.

JENNI, L., MARTI, S., SCHWEIZER, J., BETSCHART, B., Le PAGE, R.W.F., WELLS, J.M., TAIT, A., PAINDAVOINE, P., PAYS, E. and STEINERT, M. 1986. Hybrid formation between African trypanosomes during cyclical transmission. *Nature* 322: 173–175.

KOOY, F. 1991. Cellular differentiation and genetic aspects of the trypanosome life cycle. PhD Thesis, University of Utrecht.

Le PAGE, R.W.F., WELLS, J.M., PROSPERO, T.D. and STERNBERG, J. 1988. Genetic analysis of hybrid *T. brucei*. In: Turner, M.J. and Arnot, D., eds. *Molecular Genetics of Parasitic Protozoa*. Cold Spring Harbour: Cold Spring Harbour Publications, pp. 65–71.

MOTTRAM, J.C., NORTH, M.J., BARRY, J.D. and COOMBS, G.H. 1989. A cysteine protease c-DNA from *T. brucei* predicts an enzyme with an unusual C-terminal extension. *Federation of European Biochemical Societies Letters* 258: 211–215.

PAINDAVOINE, P., ZAMPETTI-BOSSELER, F., PAYS, E., SCHWEIZER, J., GUYAUX, M., JENNI, L., and STEINERT, M. 1986. Trypanosome hybrids generated in tsetse flies by nuclear fusion. *EMBO Journal* 5: 3631–3636.

SCHWEIZER, J., POSPICHAL, H., HIDE, G., BUCHANAN, N., TAIT, A. and JENNI, L. 1993. Analysis of a genetic cross between two East African *T. brucei* clones. *Parasitology*, submitted.

SCHWEIZER, J., POSPICHAL, H. and JENNI, L. 1991. Hybrid formation between African trypanosomes *in vitro*. *Acta Tropica* 49: 237–240.

STERNBERG, J. and TAIT, A. 1990. Genetic exchange in African trypanosomes. *Trends in Genetics* 6: 317–322.

STERNBERG, J., TURNER, C.M.R., WELLS, J.M., RANFORD-CARTWRIGHT, L.C., Le PAGE, R.W.F. and TAIT, A. 1989. Gene exchange in African trypanosomes; frequency and allelic segregation. *Molecular and Biochemical Parasitology* 34: 269–280.

STERNBERG, J.M., TAIT, A., HALEY, S., SCHWEIZER, J. and JENNI, L. 1987. The inheritance of chromosomal and kinetoplast variants in *T. brucei*. In: *Molecular Strategies of Parasitic Invasion*. New York: Alan R. Liss Inc., pp. 253–258.

TAIT, A. 1980. Evidence for diploidy and mating in trypanosomes. *Nature* 287: 536–538.

TAIT, A. 1983. Sexual processes in the kinetoplastida. *Parasitology* 86: 29–57.

TAIT, A. and TURNER, C.M.R. 1990. Genetic exchange in *T. brucei*. *Parasitology Today* 6: 70–75.

TIBAYRENC, M., KJELLBERG, F. and AYALA, F.J. 1990. A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences of the USA* 87: 2414–2418.

TURNER, C.M.R., STERNBERG, J., BUCHANAN, N., SMITH, E., HIDE, G. and TAIT, A. 1990. Evidence that the mechanism of gene exchange in *T. brucei* involves meiosis and syngamy. *Parasitology* 101: 377–386.

Van der PLOEG, L.H.T., SMITH, C.L., POLVERE, R.I. and GOOTESDIENER, K.M. 1989. Improved resolution of chromosome sized DNA from *T. brucei* stock 427-60. *Nucleic Acid Research* 17: 3217–3227.

WELLS, J.M., PROSPERO, T.D., JENNI, L. and Le PAGE, R.W.F. 1987. DNA contents and molecular karyotypes of hybrid *T. brucei*. *Molecular and Biochemical Parasitology* 24: 103–116.

The identification of genes of importance in *Trypanosoma brucei brucei* through the use of transfection techniques

N.B. Murphy, A. Muthiani, P. Pandit, S. Kemei, S.K. Moloo, K. Hirumi, H. Hirumi and A.S. Peregrine

Continued developments in the techniques of molecular biology and their application to the study of trypanosomes now offer several approaches to identify genes which control traits of importance in the epidemiology of trypanosomiasis and in the control of this important group of diseases. The disease complex is responsible for significant decreases in productivity and losses in livestock in approximately one-third of the African continent, thus retarding agricultural and economic development. The recent progress reported on transfection methods for *Trypanosoma brucei brucei* adds to the list of techniques that will aid in the identification of genes controlling important traits. This paper outlines the importance of adopting and exploiting such techniques in studies on trypanosomes and describes how current developments can rapidly lead to a more comprehensive understanding of these parasites in order to control the diseases they cause.

INTRODUCTION

The advent of molecular biology techniques and their application to a diverse variety of organisms has rapidly advanced our understanding of processes involved in genetic control and differentiation in eukaryotic organisms (see Williams, 1991; Wolk, 1991; Demetri, 1992; Graf, 1992; Oren, 1992; for reviews). The accelerating pace of technological development in the biological sciences is encouraging a trend towards the analysis of organisms at the most detailed level possible, and large programs to map and sequence the genomes of an increasing number of organisms are becoming more common and acceptable (Aldhous, 1992; Anderson and Aldhous, 1991, 1992; Bodmer, 1992). Although a large body of data is expected from such programs, which will eventually lead to a better understanding of the diversity and complexity of biological organisms, the time scales and efforts involved, particularly when applied to organisms which contribute to the severe levels of malnutrition and lack of sustainable growth in the developing countries of the world, require careful consideration on the merits of investing in large genome mapping projects for protozoan parasites. One such collection of organisms, which are responsible for a group of diseases (trypanosomiases) in man and his domesticated livestock, are the protozoan parasites of the genus *Trypanosoma* (Brady, 1991). These parasites are grouped into five well-differentiated subgenera *(Nannomonas, Duttonella, Trypanozoon, Pycnomonas* and *Schizotrypanum)* and ten less well-defined species (Losos, 1986). Three of these subgenera *(Nannomonas, Duttonella* and *Trypanozoon)* are widespread in Africa where they cause considerable losses in productivity in domestic animals (WHO, 1979; WHO 1986). Two of the subspecies in

the *Trypanozoon* subgenus *(T. b. rhodesiense* and *T. b. gambiense)* are, in addition, responsible for life-threatening diseases in man (WHO, 1986).

Currently, control of trypanosomiasis in domestic livestock involves eradication of the insect vector, the tsetse fly *(Glossina* sp.), the exploitation of trypanotolerant livestock and the administration of trypanocidal compounds (Morrison *et al.,* 1981). Chemotherapeutic and chemoprophylactic agents are widely used across the African continent. However, the efficacy of many of the compounds is unfortunately being impeded by the emergence of drug-resistant trypanosomes (Leach and Roberts, 1981). Whilst all the compounds used in domestic livestock have been on the market for over 30 years, very little is known about their modes of action. Furthermore, even less is known about the molecular basis of drug resistance in African trypanosomes.

In addition to drug resistance, different trypanosome isolates, even of the same genotype, can express markedly different phenotypes. The phenotypes considered to be of primary importance in the control of trypanosomiasis in the field are, in addition to drug resistance, virulence, pathogenicity, host specificity and tsetse transmissibility. An understanding of the molecular basis of these different phenotypes may allow better use of the current trypanocides, may lead to a more informed approach to the control measures which should be adopted under specific circumstances, and may aid in the development of newer therapeutic agents.

Although members of the *Trypanozoon* subgenus can undergo sexual recombination (Tait, 1980, 1983; Jenni *et al.,* 1986; Gibson, 1989; Sternberg *et al.,* 1989; Sternberg and Tait, 1990; Tait and Turner, 1990), direct genetic analysis of trypanosomes is complicated by the fact that they are diploid, and no haploid stage has yet been identified. The mapping of the genome of trypanosomes is further complicated by its fluidity, since the gene switching mechanisms resulting in antigenic variation of the major surface protein, the mechanism by which these parasites survive in the mammalian host, can result in alterations, sometimes very pronounced, in genomic DNA (Van der Ploeg *et al.,* 1992). The result of these rearrangements is that each trypanosome isolate has a distinctive and unique karyotype, leading to problems in the choice of isolate for mapping purposes, in addition to problems in making direct comparisons between isolates. Trypanosomes are also unusual in that they contain a large number of chromosomes ranging in size between 50 kb and over 5 Mb (Van der Ploeg *et al.,* 1989). Many of the small and intermediate-sized chromosomes are haploid, while the largest size classes are diploid (Gottesdiener *et al.,* 1990). Further to this, homologous chromosomes tend to differ in size, whereas heterologous chromosomes can be of similar size (Gibson and Garside, 1991). Finally, many housekeeping genes located on the megabase size class of chromosomes have a copy number of greater than 2n and tend to be arranged in tandem arrays, with the number of tandem repeats differing between homologous chromosomes (Gibson *et al.,* 1991).

These features of the trypanosome genome make it difficult to envisage a rapid solution to the disease problems these organisms cause through genomic mapping of a representative trypanosome parasite. It is therefore important to develop and utilize alternative strategies for the identification of genes controlling important traits. This paper outlines the potential in using the recently developed transfection technology for trypanosomes for the identification of genes controlling traits of importance, with two examples of how this might be productively exploited.

CURRENT DEVELOPMENTS IN STABLE TRANSFECTION METHODS FOR *TRYPANOSOMA BRUCEI BRUCEI*

The recent development of transfection systems for *Trypanosoma brucei brucei* now offer opportunities for the genetic manipulation of these parasites (Eid and Sollner-Webb, 1991; Lee and Van der Ploeg, 1990, 1991; ten Asbroek *et al.*, 1990). The majority of reports on the stable transformation of *T. b. brucei* have utilized *in vitro*-generated procyclic forms, a stage in the life-cycle which occurs in the tsetse fly, the insect vector, and which is non-infective for mammalian hosts (Tobie, 1958). In the absence of autonomously replicating vectors, stable transfectants have been selected *in vitro* on the basis of resistance to aminoglycoside antibiotics, principally G418 (Geneticin®), resulting from integration by homologous recombination of *T. b. brucei* genes containing the bacterial *neo* gene, encoding neomycin phosphotransferase II (NPTII) (Lee and Van der Ploeg, 1990, 1991; ten Asbroek *et al.*, 1990; Eid and Sollner-Webb, 1991). However, many of the more desirable and informative studies involving the life-cycle stage occurring in the mammalian host require bloodstream-form transfectants. A report by Carruthers and Cross (1992) describes a system for high efficiency clonal growth and selection of stable bloodstream- and procyclic-form transfectants on agarose plates, which is a significant advancement in the limited range of tools available for the genetic analysis of trypanosomes. The requirement for pre-adaptation of bloodstream-form parasites in this system, however, can be a disadvantage for some desired studies, and many such adapted parasites lack the ability to undergo cyclical development through tsetse flies and sexual recombination. In addition, *in vitro* cultivation of bloodstream forms is generally more difficult than that of procyclic forms and the numbers of bloodstream-form parasites that can be propagated in such systems are considerably lower than numbers of procyclic forms (Brun and Jenni, 1985; Brun and Schonenberger, 1979; Hirumi and Hirumi, 1989). It would therefore be useful to be able to select for bloodstream-form transfectants *in vivo*, particularly since most populations of *T. b. brucei* grow well in rodents. Thus, in order to extend the use of *T. b. brucei* transfection systems for both functional studies on bloodstream forms and studies on sexual recombination, in which recombinants contain selectable markers from both parental populations, we defined conditions for the *in vivo* selection of bloodstream-form transfectants which are resistant to G418 (Murphy *et al.*, 1993).

A SYSTEM FOR THE *IN VIVO* SELECTION OF BLOODSTREAM-FORM TRANSFECTANTS

We have utilized the pleomorphic parasite *T. b. brucei* GUTat 3.1, which has the capacity to undergo cyclical development (Onyango *et al.*, 1966). Bloodstream-form GUTat 3.1 trypanosomes were transformed *in vitro* to procyclic forms as described (Brun and Schonenberger, 1979; Fish *et al.*, 1989) and cultures maintained for two months prior to their use for transfection experiments. Procyclic trypanosomes were then grown to between mid and late log-phase (5–8 × 10^6/ml), harvested and resuspended at 5 × 10^7/ml in bicine-buffered saline containing 2% (w/v) glucose (Balber *et al.*, 1979). The β-tubulin gene-targeting plasmid, pUCTbneo3 (ten Asbroek *et al.*, 1990), digested with the restriction

enzymes *Sac*I and *Sal*I to release the *T. b. brucei* β-tubulin fragment, which contains the bacterial *neo* gene, was added at a final concentration of 8 μg/ml and the mixture incubated for 10 minutes at room temperature. The trypanosome/DNA mixture (0.8 ml) was then transferred to a 0.4 cm BioRad electroporation cuvette and subjected to two 300 V pulses at 960 μF capacitance, using a BioRad gene pulser with a capacitance extender. The mixture was maintained at room temperature for a further 10 minutes before being transferred to procyclic culture medium (Fish *et al.*, 1989) at a concentration of 2×10^6 living trypanosomes/ml. Following an overnight incubation at 27 °C, transfected trypanosomes were selected by transferring the electroporated parasites to fresh procyclic culture medium containing G418 at a concentration of 25 μg/ml, with a cell density of 2×10^6 viable trypanosomes/ml. After incubation for four days in this concentration of G418, no dividing trypanosomes could be detected by microscopic examination and by day 7 the majority of trypanosomes were dead. Cultures were then diluted by 100 × and 1000 × in 24-well tissue culture plates containing 1 ml of medium (plus 25 μg/ml of G418) in each well. The cultures were then maintained at 27 °C in a humidified atmosphere of 5% CO_2 in air and monitored for the next seven days. Trypanosomes from culture wells found to contain actively dividing trypanosomes were then transferred to larger-volume cultures, expanded separately and subsequently analyzed. Each expanded population was designated GUTat 3.1/BBR1 to 22. On the basis of the number of wells containing G418-resistant trypanosomes, the generation frequency of stable transfectants was estimated to be approximately one to two in 10^5 trypanosomes.

To establish that the G418-resistant trypanosomes arose through integration of the *neo* gene into the *T. b. brucei* tubulin gene cluster, parasites from the expanded cultures were harvested and embedded in 0.7% low melting point agarose and prepared for pulsed-field gel electrophoresis, as described (Van der Ploeg *et al.*, 1984). The chromosome-sized DNA molecules within the agarose blocks were fractionated on a 1.5% SeaKem GTG agarose (FMC BioProducts) gel in 1 × TBE (90 mM Tris-borate, pH 8.3, 2 mM EDTA) at 120 V, using a Pulsaphor gel electrophoresis unit (LKB) with a hexagonal electrode array. The DNA was then transferred to Nytran filters (Schleicher and Schuell) by Southern blotting in 20 × SSC (Sambrook *et al.*, 1989). Thereafter, the blots were hybridized with a ^{32}P-labelled fragment of the *neo* gene to establish, firstly, whether the *neo* gene had integrated into the *T. b. brucei* genome and, secondly, into which chromosome it had integrated. Hybridization was observed to either one of two different-sized DNA molecules, of 1.5 or 2.2 Mb, for each G418-resistant clone tested, whereas no hybridization to chromosomes of the parental parasite was observed (Figure 1B). Differences observed in hybridization signals between the transfectants were due to differences in sample loading and the efficiency of transfer of the larger size-classes of chromosomes to the filter. Figure 1C shows that the β-tubulin gene co-localizes to either of the two chromosome-sized DNA molecules of 1.5 or 2.2 Mb to which the *neo* probe hybridized to for the transfectants. However, in contrast to the *neo* probe, the β-tubulin probe hybridized to both the parental and transfected parasites. Of 10 transfectants tested, all showed localization of the *neo* gene to either of the two chromosome-sized DNA molecules without bias for either one (data not shown). In further studies, we tested for co-localization of the *neo* gene with the tubulin gene cluster by restriction enzyme analysis. Disruption of the tubulin gene cluster

through integration of the neo gene should be evident by NcoI digestion as the coding region of the neo gene contains this restriction enzyme site. Restriction fragments were separated on a 1.5% agarose gel in a similar manner to the chromosome-sized DNA molecules. The gel was then blotted and the resultant Southern blot hybridized with both the neo and b-tubulin probes as described above. Figure 1D shows the autoradiograph following hybridization with the b-tubulin probe. The additional hybridizing fragments show that the b-tubulin cluster has been disrupted in the transfectants, thereby establishing co-localization of the neo gene with the tubulin gene cluster.

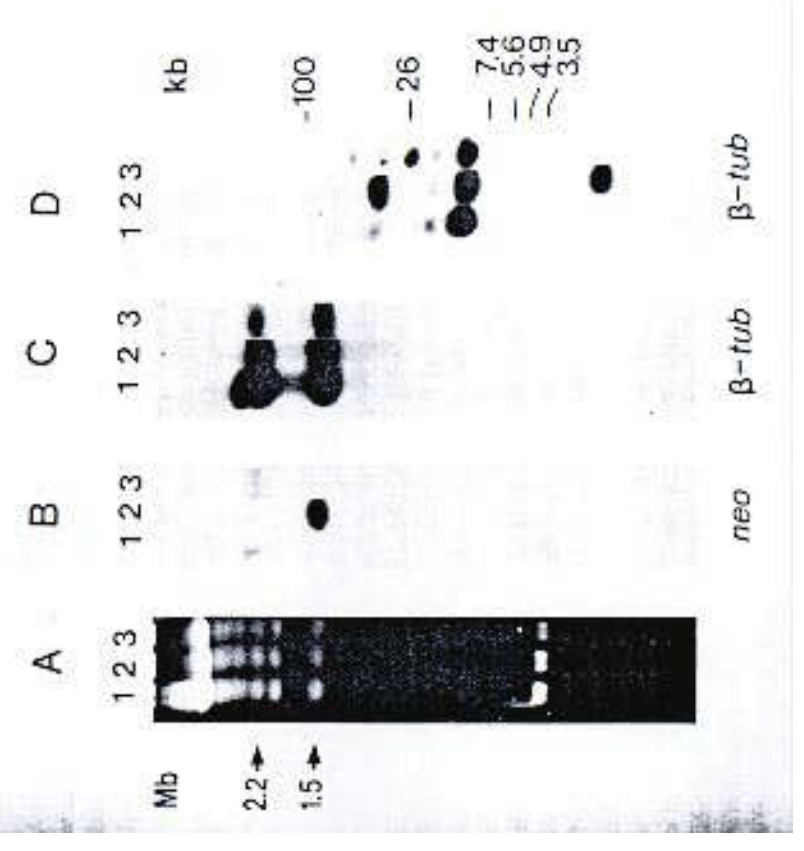

Figure 1. Integration of the β-tubulin gene-targeting plasmid, pUCTbneo3, into the genome of transfected *T. b. brucei* GUTat 3.1 parasites. Panel A shows an ethidium-bromide stained gel of separated chromosome-sized DNA molecules of the parental GUTat 3.1 parasite (lane 1) and two independent transfectants GUTat 3.1/BBR3 and GUTat 3.1/BBR7 (lanes 2 and 3, respectively). DNA from the gel was transferred to a nylon membrane and hybridized with the *neo* gene probe (*neo*, panel B) or β-tubulin gene probe (β-*tub*, panel C) labelled with ^{32}P-dCTP. Restriction enzyme digestion with *Nco*I on agarose-embedded DNA from the parental and transfected parasites demonstrates that the *neo* gene has integrated into the β-tubulin gene cluster. In panel D, *Nco*I-digested DNA of the parental (lane 1) and two transfectants (lanes 2 and 3) was separated by pulsed-field gel electrophoresis, transferred to a nylon membrane and hybridized with the β-tubulin gene probe which shows disruption of this cluster in the transfectants.

As mentioned previously, *in vitro*-derived procyclic trypanosomes are classically non-infective for mammalian hosts. However, we have found that in many procyclic cultures of *T. b. brucei*, a few mammalian infective forms are generated (A. Muthiani and N.B. Murphy, unpublished results). In order to ascertain the levels of resistance to G418 expressed by the transfected trypanosomes *in vivo*, it was necessary to generate mammalian infective forms. Approximately 5×10^7 *in vitro*-derived procyclic forms of one of the transfectants, GUTat 3.1/BBR3, were inoculated into each of five mice. Ten days later, all five mice became parasitaemic. Blood from the infected mice was collected for the generation of bloodstream-form GUTat 3.1/BBR3 stabilates (Dar *et al.*, 1972). The remaining trypanosomes were expanded further by needle passage to naive sublethally irradiated (650 rad) mice. When the level of parasitaemia in these animals had attained at least 100/field at × 250 magnification, the mice were killed and blood collected into sodium citrate (3% [w/v] final concentration). Presence of the *neo* gene was confirmed by Southern blot analysis, following purification of the parasites on a DE-52 column (Lanham and Godfrey, 1970) and isolation of the trypanosome DNA (data not shown).

To characterize the sensitivity of the trypanosome populations to G418 *in vivo*, bloodstream forms of *T. b. brucei* GUTat 3.1 and *T. b. brucei* GUTat 3.1/BBR3 were expanded separately in sublethally irradiated mice. Prior to the first peak of parasitaemia, trypanosomes were collected and aliquots containing 10^6 trypanosomes were inoculated intraperitoneally into mice. Twenty-four hours following infection, the mice were divided into groups and treated with G418 at doses of 10, 20, 30, 40, 50 or 80 mg/kg body weight (bw) by inoculating intraperitoneally 0.2 ml of the drug in sterile water. At 24 and 48 hours following the first treatment, G418 was administered to animals in each group at the same dose as before, resulting in three treatments per mouse. Repeated drug treatments were necessary to ensure complete elimination of non-transfected GUTat 3.1 parasites from the mice. Mice were then monitored daily, for 33 days, for the presence of parasites by microscopic examination of wet-blood films. Animals found to be parasitaemic were recorded and then removed from the experiment. The results of the drug treatments for the parental population, GUTat 3.1, and transfected trypanosome, GUTat 3.1/BBR3, are summarized in Table 1. For GUTat 3.1/BBR3, there was no significant delay in the onset of parasitaemia in any of the mice treated with G418, as compared to the control group (data not given); all treated mice became parasitaemic. At the highest drug dose, 80 mg/kg bw, some of the mice died from drug toxicity. In contrast to GUTat 3.1/BBR3, with the parental clone, GUTat 3.1, there was a delay in first detection of trypanosomes of three to four days in mice that became parasitaemic. At the lowest drug dose used (10 mg/kg bw) all mice infected with the parental clone became parasitaemic. In contrast, at doses of 20 and 30 mg/kg bw, only 8/10 and 2/9 mice, respectively, became parasitaemic. At the higher doses of 40 and 50 mg/kg bw, none of the mice was parasitaemic. Therefore, on the basis of these results, doses of G418 of between 40 and 80 mg/kg bw for three consecutive days are sufficient to eliminate all non-transfected *T. b. brucei* parasites from infected mice. However, at the highest dose of 80 mg/kg bw, 50% of the mice died from drug toxicity. Since treatment with 40 mg/kg bw effectively abolished all the sensitive parental trypanosomes, without any significant deleterious effect on the mice, we have maintained this as our selection treatment dose.

Table 1. *In vivo* sensitivity of *Trypanosoma brucei brucei* GUTat 3.1 and GUTat 3.1/BBR3 to G418.

	GUTat 3.1	GUTat 3.1/BBR3
10 mg kg^{-1} bw	10/10*	12/12
20 mg kg^{-1} bw	8/10	25/25
30 mg kg^{-1} bw	2/9	ND
40 mg kg^{-1} bw	0/9	23/23
50 mg kg^{-1} bw	0/10	ND
80 mg kg^{-1} bw	ND	7/7

bw Body weight
ND Not done
* Number of parasitaemic mice following treatment/number of mice infected

The clear differences in the levels of resistance *in vivo* between the transfectant, GUTat 3.1/BBR3, and the parental population, GUTat 3.1, demonstrated that it should be possible to directly select for transfectants in mice. This information on the levels of resistance of bloodstream-form *T. b. brucei* transfectants to G418 can now be utilized in the development of efficient stable transfection techniques for bloodstream parasites. Such a selection system would be beneficial in circumstances in which it is not possible to utilize an *in vitro* selection system, and would also overcome the requirement for the parasites to be pre-adapted to *in vitro* culture systems prior to transfection. The ability to select for transfectants of *T. b. brucei in vivo* also offers new possibilities for studies on genetic recombination in these parasites. Since the current transfection techniques facilitate targeting of foreign genes to specific loci, genetic crosses between, for example, trypanocide-resistant field isolates and laboratory-generated transfected G418-resistant parasites should also be possible.

To test the possibility of using bloodstream-form transfectants for the selection of parasites which have undergone genetic exchange, a genetic cross between GUTat 3.1/BBR3 and a multidrug-resistant field isolate, CP 2469 (Kaminsky *et al.*, 1990) was carried out. The relative sensitivities of both populations to G418 and the trypanocide diminazene were determined in mice. A dose of 20 mg/kg bw of diminazene was sufficient to eliminate all infections with GUTat 3.1/BBR3 but did not eliminate any infections with CP 2469, whereas three treatments at 40 mg/kg bw of G418 was sufficient to eliminate all infections with CP 2469, but not GUTat 3.1/BBR3. Therefore, each population has a selectable single phenotype which is lacking in the other, and hybrids could therefore be selected on the basis of having both characteristics. Because of the differences in both drugs in structure and mode of action, synergistic effects due to treatments with both drugs would not be expected to occur.

Two groups of sublethally irradiated rats were infected with *T. b. brucei* clones GUTat 3.1/BBR3, or IL 3565, a clone of CP 2469. At peak parasitaemia, the rats were killed and heparinized blood from the two groups was collected separately. Fifty ml of the infected blood from the two groups was mixed, and 400 teneral male *Glossina morsitans centralis* (tsetse flies) were fed *in vitro* on the blood. On day 30 following feeding, the surviving

380 tsetse were induced to probe onto warmed slides at 37°C to identify those with mature infections. Seven tsetse (1.84%) showed metacyclics in their salivary probes. The seven tsetse with mature infections were fed singly on sublethally irradiated Swiss white mice twice weekly until day 85 when all the seven infected tsetse had died. Mice were treated with diminazene at 20 mg/kg bw on day 1 and G418 at 40 mg/kg bw on days 5, 6 and 7 post infection. Breakthrough infections were observed from four separate feedings: day 39, from flies 2, 4 and 5, day 46 from fly 5, day 53 from fly 3 and day 64 from flies 1, 2 and 5. This represents at least one breakthrough infection from five of the seven flies fed on mice. Non-treated control infections were carried out on three separate occasions: day 46 when all seven flies were fed producing seven infections, day 67 when four flies were fed producing four infections, and day 81 when two flies were fed producing two infections. Stabilates were prepared from all mice that became parasitaemic. The different populations are currently being analyzed to determine whether they contain true recombinants.

In these studies we were somewhat surprised at the potency of G418 against *T. b. brucei in vivo;* even at high levels of parasitaemia, nearly all the parasites disappeared from the bloodstream within 24 hours following the first drug treatment at 40 mg/kg bw (N.B. Murphy and A.S. Peregrine, unpublished observations). Furthermore, by comparison to other organisms (Davies and Jimenez, 1980), bloodstream-form *T. b. brucei* parasites are extremely sensitive to G418 and show a significantly greater level of sensitivity (approximately 100-fold) to G418 *in vitro* compared to procyclic forms (N.B. Murphy, unpublished observations). The reasons for the differing sensitivities of the two life-cycle stages of *T. b. brucei* to G418 may be related to differences in growth rates or differences in their rates of endocytosis (Webster and Fish, 1989). Other salivarian trypanosomes such as *T. congolense* are as sensitive to G418 *in vivo* as *T. b. brucei* (N.B. Murphy and A.S. Peregrine, unpublished observations). The reasons for the relatively high levels of sensitivity of trypanosomes to G418 in comparison to other eukaryotes are unclear, but since G418 is an aminoglycoside, it may be related to the unusual structure of their ribosomes (White *et al.,* 1986). This possibility should be investigated further since it may present some potential for the generation and testing of potent inhibitors of trypanosome ribosome function, with minimal inhibition of host function, thus offering new prospects for the control of trypanosomes and the treatment of the disease in the future.

IDENTIFICATION OF *TRYPANOSOMA BRUCEI BRUCEI* SEQUENCES INVOLVED IN THE ESTABLISHMENT/MAINTENANCE OF INFECTION

The changes in the proliferative capabilities of trypanosomes are influenced, at least in part, by host factors. It would be desirable to determine how such factors might interact with parasite molecules and influence the establishment and maintenance of infections. To fully exploit transfection technology for the identification of genes controlling such phenotypes, the ability to generate mutants is required. Of particular relevance to such studies is the generation of *T. b. brucei* GUTat 3.1 mutants through extended growth *in vitro* and manipulation of the incubation temperature and hemen concentration in the culture media. These mutants can be propagated in *in vitro* culture systems, but fail to

establish infections in mammalian hosts with inocula of up to 10^7 parasites. They therefore represent ideal material for testing systems for the generation of revertants, since there is a direct selection system for such revertants through growth in a mammalian host. The generation of mutants with easily identifiable phenotypes is an important tool in genetic analysis. However, the ability to generate revertants and map the genomic location at which an alteration has occurred which correlates with the change in phenotype is considerably more powerful.

Karyotype analysis has revealed a consistent alteration involving two chromosomes of approximately 1.8 Mb between the parental and attenuated populations (see Figure 2A). The significance of these alterations in karyotype is difficult to determine, and a concerted effort to analyze this might lead to information on somatic alterations in the genome which are unrelated to the attenuated phenotype. We therefore tested whether it would be possible to generate virulent revertants of the attenuated mutant through the use of transfection technology. Although transfection vectors capable of autonomous replication are not yet available for trypanosomes, stable integrative transformation is thought to occur exclusively through homologous recombination (Eid and Sollner-Webb, 1991). We reasoned that the generation of transfected revertants through the use of total DNA from a virulent parasite would result from homologous recombination events which correct the mutation(s). Total DNA from a well characterized virulent parasite, ILTat 1.1 (Miller and Turner, 1981) was prepared and sheared to an average size of 30 kb. This DNA was added to 5×10^7 live attenuated trypanosomes, suspended in bicine-buffered saline, at a concentration of 20 μg/ml, and subjected to a single pulse of 550 volts with a capacitance of 650 μF. The electroporator was a home-made apparatus with two platinum wires as electrodes separated by a gap of 0.8 cm. As a control, 20 μg of salmon-sperm DNA was electroporated into 5×10^7 attenuated parasites. Following a 10 min incubation at room temperature, two groups of five sublethally irradiated mice (650 rad) were infected with the electroporated trypanosomes. On day 10 following infection, one of the mice from the population transfected with the ILTat 1.1 DNA developed a high parasitaemia. Two additional mice in this group subsequently developed high parasitaemias, whereas the control group remained aparasitaemic. On reinfection of non-irradiated mice with 10^5 parasites from the three breakthrough populations, all mice developed high parasitaemias within five days. The karyotypes of the parental virulent GUTat 3.1 parasite, the avirulent mutant and the three independent revertants were compared (Figure 2A). The karyotypes of the revertants were identical to the attenuated population and differed to the virulent parental, thus establishing that the revertants are derived from the attenuated population and are not a result of contamination by the virulent parental parasite. We had therefore generated three independent revertant populations through transfection of total DNA from a virulent population.

Several DNA probes have been used in an attempt to identify the region of the genome which has undergone an alteration in these parasites, but without success (data not shown). Recently, a system of arbitrary primer (AP) PCR, otherwise known as randomly amplified polymorphic DNA (RAPD), has been developed and applied to the characterization of the genomes of various organisms, including trypanosomes (Welsh and McClelland, 1990; Williams *et al.,* 1990; Waitumbi and Murphy, 1993; Dirie *et al.*, 1993). Each DNA band of the resultant fingerprints represents priming of a different locus, and hence marker, in

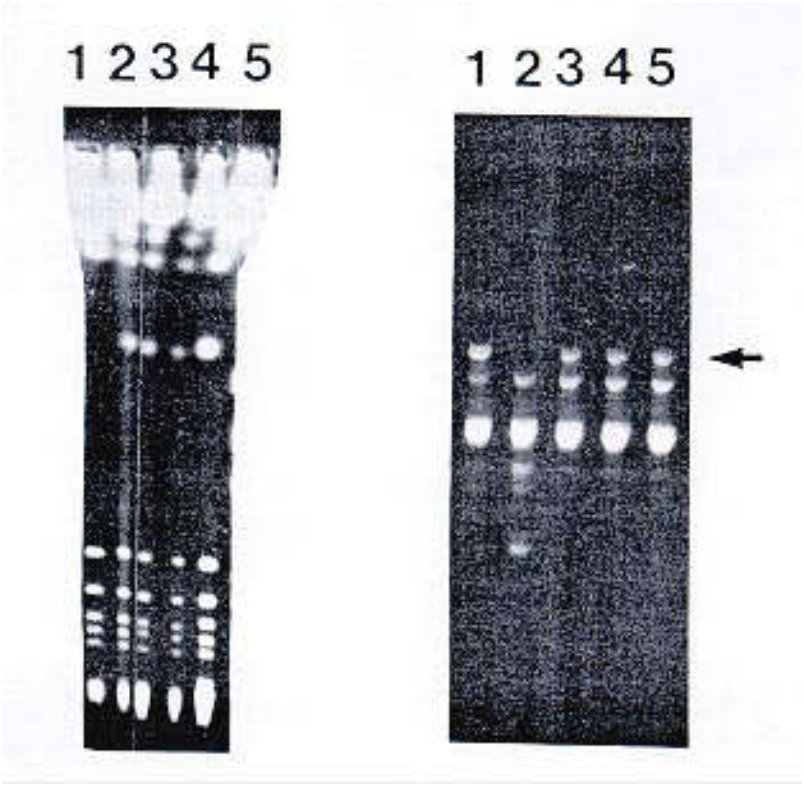

Figure 2. DNA analysis showing differences between the parental, virulent *T. b. brucei* GUTat 3.1 isolate (lane 1, panels A and B), the avirulent mutant (lane 2, panels A and B) and the three independent transfected revertants (lanes 3–5, panels A and B). Panel A shows a pulsed-field gel of separated chromosome-sized DNA molecules from each isolate and panel B shows the products of AP-PCR reactions on total DNA from each isolate with oligonucleotide ILO 873. The arrow points to the PCR product which is generated from total DNA of the virulent parental and transfectants, but not from the avirulent mutant.

the genome, and the application of many primers can result in the generation of a well saturated and generally evenly spread set of markers throughout the genome. We therefore applied the AP-PCR technique, using over 100 primers, on DNA from the virulent parental, the avirulent mutant and the three revertant transfectants in an attempt to identify the genomic location where the alteration occurred. One of these primers, ILO 873, amplified a fragment in the virulent parental and the three transfectants but not in the avirulent mutant (Figure 2B). This PCR product has been cloned and sequenced at both ends. One end of the fragment contains sequences of the TRS1/ingi transposable element (Kimmel *et al.*, 1987; Murphy *et al.*, 1987), but the sequence identity of the other end has not yet been established.

CONCLUSIONS

We have outlined two examples of how transfection technology can be exploited to rapidly lead to the identification of important genes in trypanosomes. Such studies should help in increasing our understanding of trypanosomes and ultimately aid in controlling the diseases that this important group of parasites causes. The advantages of this type of approach over the direct mapping approach is illustrated by the rapidity with which regions of the genome controlling important phenotypic traits have been and can be identified. It is hoped that such approaches will not be neglected, but rather will be more seriously considered in future studies on these parasites.

REFERENCES

ALDHOUS, P. 1992. Britain plans large-scale sequencing center. *Science* 256: 958.

ANDERSON, C. and ALDHOUS, P. 1991. Genome project. Secrecy and the bottom line. *Nature* 354: 96.

ANDERSON, C. and ALDHOUS, P. 1992. Human genome project. Still room for HUGO? *Nature* 355: 4–5.

ten ASBROEK, A.L.M.A., OUELLETTE, M. and BORST, P. 1990. Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. *Nature* 348: 174–175.

BALBER, A.E., BANGS, J.D., JONES, S.M. and PROIA, R.L. 1979. Inactivation or elimination of potentially trypanolytic complement-activating immune complexes by pathogenic trypanosomes. *Infection and Immunity* 24: 617–627.

BODMER, W. 1992. Genome research in Europe. *Science* 256: 480–481.

BRADY, J. 1991. Seeing flies from space. *Nature* 351: 695.

BRUN, R. and JENNI, L. 1985. Cultivation of African and South American trypanosomes of medical and veterinary importance. *British Medical Bulletin* 41: 122–129.

BRUN, R. and SCHONENBERGER, M. 1979. Cultivation and *in vitro* cloning of procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Tropica* 36: 289–292.

CARRUTHERS, V.B. and CROSS, G.A.M. 1992. High-efficiency clonal growth of bloodstream- and insect-form *Trypanosoma brucei* on agarose plates. *Proceedings of the National Academy of Sciences of the USA* 89: 8818–8821.

DAR, F.K., LEMECHA, H. and MULATU, W. 1972. Cryopreservation of pathogenic African trypanosomes *in situ:* metacyclic and bloodstream forms. *Journal of Protozoology* 19: 494–497.

DAVIES, J. and JIMENEZ, A. 1980. A new selective agent for eukaryotic cloning vectors. *Journal of Tropical Medicine and Hygiene* 29: 1089–1092.

DEMETRI, G.D. 1992. Hematopoietic growth factors: current knowledge, future prospects. *Current Problems in Cancer* 16: 177–259.

DIRIE, M.F., MURPHY, N.B. and GARDINER, P.R. 1993. DNA fingerprinting of *Trypanosoma vivax* isolates rapidly identifies intraspecific relationships. *Journal of Eukaryotic Microbiology* 40: 132–134.

EID, J. and SOLLNER-WEBB, B. 1991. Stable integrative transformation of *Trypanosoma brucei* occurs exclusively by homologous recombination. *Proceedings of the National Academy of Sciences of the USA* 88: 2118–2121.

FISH, W.R., MURIUKI, C., MUTHIANI, A.M., GRAB, D.J. and LONSDALE-ECCLES, J.D. 1989. Disulfide bond involvement in the maintenance of the cryptic nature of the cross-reacting determinant of metacyclic forms of *Trypanosoma congolense. Biochemistry* 28: 5415–5421.

GIBSON, W.C. 1989. Analysis of a genetic cross between *T. b. rhodesiense* and *T. b. brucei*. *Parasitology* 99: 391–402.

GIBSON, W.C. and GARSIDE, L. 1991. Genetic exchange in *T. brucei:* variable chromosomal location of housekeeping genes in different trypanosome stocks. *Molecular and Biochemical Parasitology* 45: 77–90.

GIBSON, W.C., GARSIDE, L. and BAILEY, M. 1991. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *T. b. rhodesiense* and *T. b. brucei. Molecular and Biochemical Parasitology* 52: 189–200.

GOTTESDIENER, K., GARCIA-ANOVEROS, J., LEE, M.G., and Van der PLOEG, L.H.T. 1990. Chromosome organisation of the protozoan *T. brucei. Molecular and Cellular Biology* 10: 6079–6083.

GRAF, T. 1992. Myb: a transcriptional activator linking proliferation and differentiation in hematopoietic cells. *Current Opinions in Genetics and Development* 2: 249–255.

HIRUMI, H. and HIRUMI, K. 1989. Continuous cultivation of *Trypanosoma brucei* bloodstream forms in a medium containing a low concentration of serum protein without feeder cell layers. *Journal of Parasitology* 75: 985–989.

JENNI, L., MARTI, S., SCHWEIZER, J., BETSCHART, B., Le PAGE, R.W.F., WELLS, J.M., TAIT, A., PAINDAVOINE, P., PAYS, E. and STEINERT, M. 1986. Hybrid formation between African trypanosomes during cyclical transmission. *Nature* 322: 173–175.

KAMINSKY, R., GUMM, I.D., ZWEYGARTH, E. and CHUMA, F. 1990. A drug incubation infectivity test (DIIT) for assessing resistance in trypanosomes. *Veterinary Parasitology* 34: 335–343.

KIMMEL, B.E., ole MOIYOI, O.K. and YOUNG, J.R. 1987. `Ingi', a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. *Molecular and Cellular Biology* 7: 1465–1475.

LANHAM, S.M. and GODFREY, D.G. 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. *Experimental Parasitology* 28: 521–535.

LEACH, T.M. and ROBERTS, C.J. 1981. Present status of chemotherapy and chemoprophylaxis of animal trypanosomiasis in the eastern hemisphere. *Pharmacology and Therapeutics* 13: 91–147.

LEE, G.-S.M. and Van der PLOEG, L.H.T. 1990. Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei. Science* 250: 1583–1587.

LEE, G.-S.M. and Van der PLOEG, L.H.T. 1991. The hygromycin B-resistance-encoding gene as a selectable marker for stable transformation of *Trypanosoma brucei. Gene* 105: 255–257.

LOSOS, G.J. 1986. Trypanosomiases. *Infectious Tropical Diseases of Domestic Animals.* London: Longman, pp. 182–318.

MILLER, E.N. and TURNER, M.J. 1981. Analysis of antigenic types appearing in first relapse populations of *Trypanosoma brucei. Parasitology* 82: 63–80.

MORRISON, W.I., MURRAY, M. and McINTYRE, W.I.M. 1981. Bovine trypanosomiasis. In: Ristic, M. and McIntyre, I., eds. *Diseases of Cattle in the Tropics.* The Hague, Netherlands: Martinos Publishers, pp. 469–497.

MURPHY, N.B., MUTHIANI, A.M. and PEREGRINE, A.S. 1993. Use of an *in vivo* system to determine the G418 resistance phenotype of bloodstream-form *Trypanosoma brucei brucei* transfectants. *Antimicrobial Agents and Chemotherapy* 37: 1167–1170.

MURPHY, N.B., PAYS, A., TEBABI, P., COQUELET, H., GUYAUX, M., STEINERT, M. and PAYS, E. 1987. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. *Journal of Molecular Biology* 195: 855–871.

ONYANGO, R.J., Van HOEVE, K. and de RAADT, P. 1966. The epidemiology of *Trypanosoma rhodesiense* sleeping sickness in Alego Location, Central Nyanza, Kenya. 1. Evidence that cattle may act as reservoir hosts of trypanosomes infective to man. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 60: 175–182.

OREN, M. 1992. p53: the ultimate suppressor gene? *FASEB Journal* 6: 3169–3176.

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T. 1989. *Molecular Cloning,* Second Edition. Cold Spring Harbor: Cold Spring Harbor Laboratory.

STERNBERG, J. and TAIT, A. 1990. Genetic exchange in African trypanosomes. *Trends in Genetics* 6: 317–322.

STERNBERG, J., TURNER, C.M.R., WELLS, J.M., RANFORD-CARTWRIGHT, L.C., Le PAGE, R.W.F. and TAIT, A. 1989. Gene exchange in African trypanosomes: frequency and allelic segregation. *Molecular and Biochemical Parasitology* 34: 269–280.

TAIT, A. 1980. Evidence for diploidy and mating in trypanosomes. *Nature* 287: 536–538.

TAIT, A. 1983. Sexual processes in the kinetoplastida. *Parasitology* 86: 29–57.

TAIT, A. and TURNER, C.M.R. 1990. Genetic exchange in *T. brucei*. *Parasitology Today* 6: 70–75.

TOBIE, E.J. 1958. The cultivation of *Trypanosoma congolense in vitro*. *Journal of Parasitology* 44: 241–242.

Van der PLOEG, L.H.T., GOTTESDIENER, K.M. and LEE, M.G.S. 1992. Antigenic variation in trypanosomes. *Trends in Genetics* 8: 452–457.

Van der PLOEG, L.H.T., SCHWARTZ, D., CANTOR, C.R. and BORST, P. 1984. Antigenic variation in *Trypanosoma brucei* analysed by electrophoretic separation of chromosome-sized DNA molecules. *Cell* 37: 77–84.

Van der PLOEG, L.H.T., SMITH, C.L., POLVERE, R.I. and GOTTESDIENER, K.M. 1989. Improved resolution of chromosome sized DNA from *T. brucei* stock 427-60. *Nucleic Acids Research* 17: 3217–3227.

WAITUMBI, J.N. and MURPHY, N.B. 1993. Inter- and intra-species differentiation of trypanosomes by genomic fingerprinting with arbitrary primers. *Molecular and Biochemical Parasitology* 58: 181–186.

WEBSTER, P. and FISH, W.R. 1989. Endocytosis by African trypanosomes. II. Occurrence in different life-cycle stages and intracellular sorting. *European Journal of Cell Biology* 49: 303–310.

WELSH, J. and McCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 7213–7218.

WHITE, T.C., RUDENKO, G. and BORST, P. 1986. Three small RNAs within the 10-kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. *Nucleic Acids Research* 14: 9471–9489.

WHO. 1979. *The African Trypanosomiasis. Report of a Joint WHO Expert Committee and FAO Expert Consultation. Technical Report Series 635*. Geneva: World Health Organization.

WHO. 1986. *Epidemiology and Control of African Trypanosomiasis. Report of a WHO Expert Committee. Technical Report Series 739*. Geneva: World Health Organization.

WILLIAMS, J.G. 1991. Regulation of cellular differentiation during *Dictyostelium* morphogenesis. *Current Opinions in Genetics and Development* 1: 358–362.

WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSKI, J.A. and TINGEY, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531–6535.

WOLK, C.P. 1991. Genetic analysis of cyanobacterial development. *Current Opinions in Genetics and Development* 1: 336–341.

Chromosome polymorphism in *Trypanosoma brucei brucei* and the selection of chromosome-specific markers

S.E. Melville, M.P. Barrett, J.P. Sweetman, J.W. Ajioka and R.W.F. Le Page

The first direct evidence for mating and gene exchange in African trypanosomes was presented in 1986 and it is now possible to perform crosses by mixed passage through laboratory-bred tsetse flies. Hybrid progeny are rescued and clonal populations established. When the karyotypes of a number of hybrid progeny were compared, they were found to differ both from each other and from the parents in the number and size of their chromosomes. This kind of variation is also observed when the karyotypes of field isolates from tsetse-transmitted populations are compared.

However, the investigation of genome structure and chromosome inheritance in hybrids and also in field isolates is hampered by a paucity of suitable chromosome markers. Without sufficient linked markers, it is not possible to begin to investigate the content of the extra DNA in `homologues' of different sizes, the maintenance or loss of linkage groups or the locations of chromosome breakage and recombination events. To this end we have concentrated on a single pair of chromosomes and have isolated gene and VNTR markers. However, because of the lack of cytogenetic techniques applicable to the study of trypanosome chromosomes, it is necessary to reconstruct the chromosome with overlapping DNA clones in order to prove the linkage of markers and to order them along the chromosome. We have succeeded in selecting pulsed-field gel band-specific subgenomic libraries as a means of simplifying this task. The cloning of the chromosome has the added advantage that the ends of these clones may serve as precisely positioned physical markers.

Preliminary data on the structure and inheritance of the chromosomes is necessary for the development of a suitable approach to the analysis of such a plastic genome.

GENOME STRUCTURE: SPECIES-SPECIFIC PROBLEMS FOR PHYSICAL MAPPING

The nuclear genome of the African trypanosome is a highly plastic structure, as observed by pulsed-field gel electrophoresis (PFGE). This clearly has implications for the physical mapping of the genome. The chromosomes may be divided into three size classes: approximately 100 mini-chromosomes of about 50 kb, 1–8 intermediate chromosomes of between 200 and 700 kb and 16–18 megabase chromosomes ranging from 1 to 5.7 Mb (Van der Ploeg *et al.*, 1984a; Wells *et al.*, 1987; Gottesdiener *et al.*, 1990; reviewed in Tait and Turner, 1990). However, the number and sizes of the intermediate and megabase chromosomes differ. The karyotype varies considerably not only between species and subspecies, but also between independently isolated field stocks (Van der Ploeg *et al.*, 1984b; Gibson and Borst, 1986; Gibson, 1989; Gibson and Garside, 1991). The most striking polymorphism is observed in the megabase chromosomes, which exhibit multiple differences in size and number. It has been proposed that sexual recombination may be involved in the generation of the observed polymorphism.

It was first demonstrated in 1986 that African trypanosomes can undergo a process of mating and gene exchange (Jenni *et al.,* 1986). Several separate crosses have now been performed (Paindevoine *et al.,* 1986; Sternberg *et al.,* 1988; Sternberg *et al.,* 1989; Gibson, 1989; Turner *et al.,* 1990; reviewed in Tait and Turner, 1990; Gibson and Garside, 1991) and these have involved *Trypanosoma brucei brucei* from East and West Africa, *T. b. gambiense* and *T. b. rhodesiense.* Genetic exchange is therefore possible between subspecies and between strains originating from widely separated geographic areas.

While the feasibility of performing laboratory crosses increases the scope for genetic analysis of these organisms, the process remains a time-consuming and skilled process for which only a few laboratories are equipped. It is not feasible to retrieve and clone large numbers of F_1 hybrids; nor is it possible to ensure that those retrieved are indeed F_1 and not F_2 (Sternberg *et al.,* 1989). Therefore, genetic analysis of African trypanosomes remains a difficult task.

A cross between two cloned stocks of trypanosomes (STIB247L and TREU927) was carried out by Prof. A. Tait and Dr. C.M.R. Turner of Glasgow University as described by Turner and colleagues (1990). These field isolates are both subspecies *T. b. brucei* and both originate from East Africa, but their karyotypes differ considerably. Even greater karyotypic complexity was observed when the megabase chromosomes of cloned hybrid trypanosomes derived from this laboratory cross were studied using PFGE. The range of chromosome size and number is illustrated in Figure 1. Such karyotypic variability clearly poses quite specific problems for the creation of physical maps of the nuclear genome of these organisms.

The megabase chromosomes were shown to be diploid and to contain housekeeping genes and both basic copy and expression-linked variable surface glycoprotein (VSG) genes (Gibson *et al.,* 1985; Van der Ploeg *et al.,* 1984c). Six chromosome pairs were identified in the laboratory stock 427/60 using a range of markers (Gottesdiener *et al.,* 1990). These pairs of chromosomes were shown to contain long stretches of homology and are referred to tentatively as homologues although they may differ in size by 10–20%. We have also found that apparently homologous chromosomes differ in size in both the parents and the progeny. This observation must also be taken into account in the construction of a physical map and the location of markers.

The lack of cytogenetic procedures complicates the study of chromosome inheritance and polymorphism in *T. b. brucei;* although long-range restriction fragment length polymorphism (RFLP) mapping procedures can provide an initial guide to the differences between homologous chromosomes, it is clear that a systematic analysis of chromosomal breakpoints and behaviour can best be approached by developing sets of suitably dispersed but linked markers for individual chromosomes.

Although PFGE provides a method of separating the chromosomes of trypanosomes, not all bands in an agarose gel represent a single chromosome. Some bands are more intense than others and may contain comigrating DNA molecules of different content. This severely limits the usefulness of Southern hybridization as a method for showing the linkage of markers. The most likely homogeneous band is that of least staining intensity with a homologue known to migrate to another position in the gel. However, definitive proof of the linkage of markers has to come from the `reconstruction' of the chromosome from clones of overlapping DNA fragments. For this reason, the

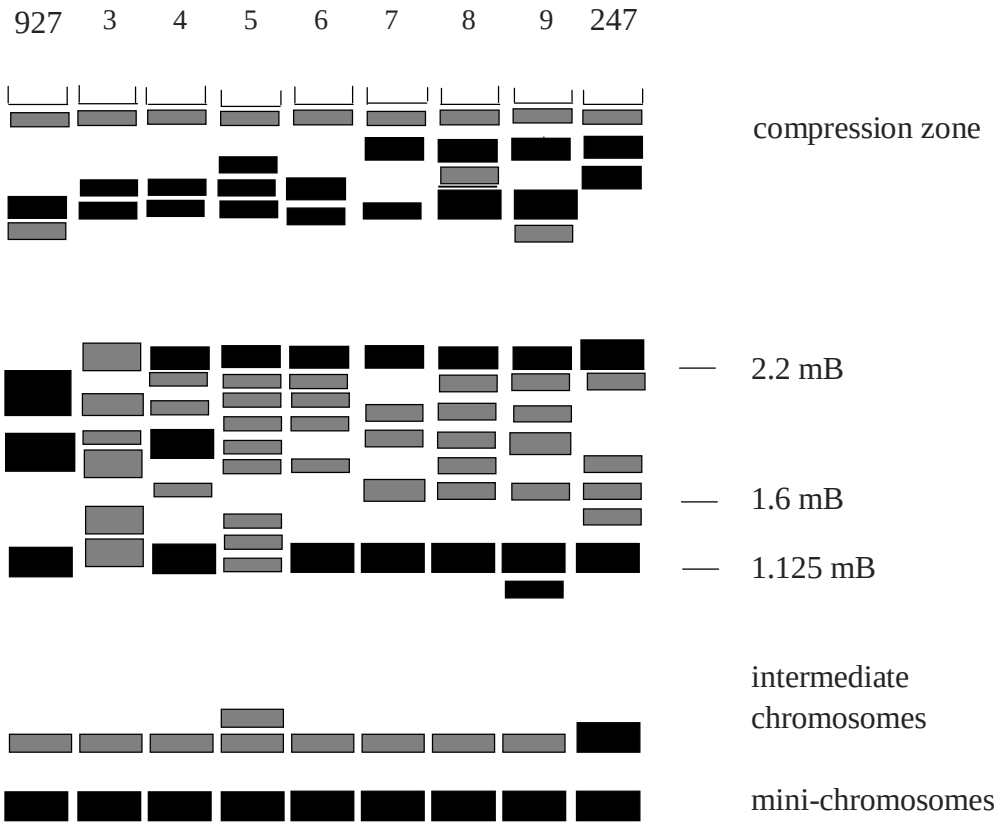

Figure 1. The megabase chromosomes of *T. b. brucei*. The chromosomes of the parent trypanosome clones (TREU927 and STIB247L) and of seven hybrid clones derived from a laboratory cross (F532/72 mcl 3-9) (see Turner *et al.*, 1990) are shown in diagrammatic form. These chromosomes were visualized by PFGE (CHEFDRII, Biorad) using pulse times ranging from 700 to 1400 sec at 2.5 V cm^{-1}. The sizes given are derived from *Saccharomyces cerevisiae* (YNN295) chromosomes. Considerable chromosome polymorphism is evident.

construction of a physical map using cloned DNA is more suitable than, for example, restriction site mapping.

AN APPROACH TO THE PHYSICAL MAPPING OF THE TRYPANOSOME GENOME: CHROMOSOME-SPECIFIC LIBRARIES

For the purpose of physical mapping with cloned DNA it is clearly of advantage to use vectors which will accommodate large pieces of foreign DNA such as cosmids (Ish-Horowicz and Burke, 1981), bacteriophage P1 (P1s) (Sternberg, 1990) or yeast artificial chromosomes (YACs) (Burke *et al.*, 1987). Data on the content and behaviour of chromosomes, such as the varying sizes of apparent homologues, will be obtained

considerably more rapidly if single chromosomes can be cloned and mapped individually. The clear method of choice for the preparation of a chromosome-specific subgenomic library of overlapping clones involves excising a chromosome band from an agarose gel, partially digesting the DNA and cloning these fragments into the chosen vector. In practice, this procedure works quite well with vectors which replicate small fragments of foreign DNA, but the difficulty of this procedure increases with the size of the fragments to be cloned: it is difficult to obtain large fragments in sufficient concentration and the necessity to size-select, to avoid chimaeric clones, reduces the concentration further. This method proved to be very difficult using cosmid vectors, and is probably impossible with bacteriophage P1 or YACs. Thus we have employed the alternative strategy of selecting PFGE band-specific subsets of cosmid clones containing *T. b. brucei* DNA as a means of simplifying chromosome mapping (S.E. Melville, J.P. Sweetman and R.W.F. Le Page, in preparation). This method makes chromosome walking considerably less cumbersome: such small libraries may be kept as ordered arrays, simplifying the preparation of filters and the retrieval of clones, and facilitating mapping in laboratories without access to high-tech robotics for the production of high density filters. The study of a single chromosome pair has helped us to prepare for the species-specific problems to be encountered in a large-scale physical mapping project (see J. Blackwell, this volume).

A cosmid library of *T. b. brucei* DNA was plated, the colonies lifted and replica filters created. The radioactive chromosomal probe was prepared by PFGE-fractionation and hybridized to the cosmid DNA immobilized on the filter, identifying clones containing DNA derived from that pulsed-field gel band and from its homologue. These clones were then stored individually in a much smaller, more manageable, subgenomic library.

This very simple idea depends on two factors: efficiently labelling the correct amount of chromosomal DNA, and the amount of dispersed repeated DNA in the genome. If the chromosomes of the trypanosome have sequence in common, then the chromosomal probe will select cloned DNA from all the different chromosomes which contain these sequences. In anticipation of this problem, the cosmid colony lifts were prehybridized with sheared genomic trypanosome DNA. The highly repetitive portion of the genomic DNA in the prehybridization mixture will hybridize more rapidly than the single copy DNA (Sealey, 1985) since it is present, by definition, in higher copy number. This prehybridization procedure allows highly repeated sequence in the unlabelled genomic DNA to hybridize to homologous DNA in the cosmid clones, thus reducing the level of signal obtained from hybridization of these same sequences in the radioactive chromosomal probe to the cosmid clones.

The estimated haploid nuclear genome of *T. b. brucei* is 40 Mb and the kinetoplast approximately 3.3 Mb (Borst *et al.*, 1982). A 5 × haploid genome equivalent of cosmids containing average inserts of 35 kb is therefore 5950 cosmid clones. The chromosome from which the probe was made is approximately 1.5 Mb: if the cosmid inserts average 35 kb, then a 5 × genome equivalent to cover this one chromosome is 214 clones. Table 1 shows the average number of clones obtained in various probing experiments using a range of genomic DNA concentrations in the prehybridization solution.

The primary observation arising from these experiments is that even when no genomic trypanosome DNA was included in the prehybridization solution, the number of clones

Table 1. Selecting cosmid clones derived from a single *T. b. brucei* chromosome and its homologue. The average number of clones retrieved after prehybridization with genomic trypanosome DNA and hybridization with a radioactive chromosomal probe are given. The number expected is estimated according to the size of the chromosome and the average insert DNA length (see text).

Concentration of genomic DNA	Number of clones retrieved				Number
μg/ml)	Strong	Intermediate	Weak	Total	expected
0.0	33	71	126	230	214
0.5	33	67	138	238	214
1.5	34	59	116	209	214
5.0	21	22	145	188	214

selected by the chromosomal probe was approximately that expected. The number of cosmid clones to which the chromosomal probe hybridized did not change significantly when the filters were prehybridized with 0.5 or 1.5 μg ml^{-1} of genomic trypanosome DNA. This indicates that the pulsed-field gel band chosen for the preparation of probe does indeed contain a single species of chromosome. This result also suggests that it may not be necessary to attempt to mask repeated sequences in the genome before hybridization. However, there are various possible ways to account for this result.

1. There is very little dispersed repeated DNA in the genome of *T. b. brucei*.
2. There are dispersed repeats but fortuitously the chromosome from which the probe was prepared contains little repeated DNA in common with other chromosomes.
3. The dispersed repeated DNA is underrepresented in the cosmid library, either because it is not readily cloned, or because it is associated with the subtelomeric regions which are known to lack restriction sites.
4. The mobility of the chromosome in the pulsed-field gel is affected by structural features, or other characteristics, and the actual size of the chromosome differs substantially from the estimate of 1.5 Mb.

Secondly, it was observed that the signals obtained on the autoradiographs varied greatly in their strength. Very strong signals obtained from complex probes such as this commonly signify repetitive sequences. These sequences may include chromosome-specific mini-satellite repeats and the tandemly reiterated genes which are common in the trypanosome genome (Clayton, 1988). Weaker signals probably represent cosmids containing only single copy sequence. Signal strength will also be affected by the length of the insert in the cosmid.

The number of strongly hybridizing colonies did not differ significantly in the first three experiments. However, when the concentration of unlabelled genomic DNA was increased to 5 μg ml^{-1}, the numbers hybridizing clearly decreased to below that expected. Using replica filters, it was shown that the decrease was due to the loss of hybridization to cosmid clones for which only a weak signal was obtained in other experiments using less genomic DNA in the prehybridization solution. Cosmid clones which had previously shown strong hybridization gave signals of reduced intensity. This suggests that the concentration of

genomic DNA in the prehybridization solution was too high to selectively mask dispersed repeats and was beginning to mask *bona fide* chromosome-specific clones. A concentration of 5 μg ml^{-1}, for example, is very high compared to that used to protect repeated sequences in the human genome (Sealey, 1985).

A convenient number of clones to work with for manual screening is at most 384, for these can be stocked in four microtitre plates and may all be plated onto 22 × 22 cm plates using a 96-well replicator and lifted on a single filter. For a chromosome of 1.5 Mb, this represents an 8.5 × genomic library. During a chromosome walk involving thus far 12 cosmid clones, only once was an overlapping clone not retrieved from the chromosome-specific library. The random library was then probed again and a suitable clone obtained. The selected library has been probed with six markers known to be specific to the chromosome used as a probe and in each case three or more cosmid clones containing the marker sequence were retrieved. All these results strongly suggest that a useful, essentially chromosome-specific library was selected.

The chromosomal probe was then hybridized to a Southern blot of a PFGE separation of the chromosomes of the same trypanosome stock (Figure 2). The probe hybridized to the mini-, the intermediate and most of the megabase PFG bands. This indicates that there is sequence in common between these chromosomes. They are all known to contain the telomeric repeats. However, the probe has shown only a low level of hybridization to two of the megabase chromosomes. This suggests that these chromosomes have considerably less sequence in common, possibly only the telomeric repeats. This provides some evidence of dispersed repetitive DNA in the trypanosome genome. However, these sequences may not occur on all chromosomes and may be located for the most part in the subtelomeric regions.

Known sequences which may be suitably masked by this method include the 70 bp repeats upstream of the variable surface glycoprotein genes (Liu *et al.,* 1983) and the randomly inserted transposable elements RIME and INGI (Hasan *et al.,* 1984; Kimmel *et al.,* 1987). As more information is obtained regarding the types of repeated sequences which may interfere with this procedure, it will certainly be more efficient to use cloned or amplified DNA of the specific repeated sequences in the prehybridization solution instead of genomic DNA.

This method of clone selection would probably prove inefficient for libraries constructed in plasmid or bacteriophage λ libraries as the small insert DNA will bind insufficient probe. However, libraries constructed in YAC and P1 vectors should prove ideal.

Chromosome walking using radioactive probes created from the ends of the cosmid inserts is very simple using a small library in an ordered array. It is not yet known if there are regions of the chromosome which are not represented, other than the subtelomeric and telomeric regions. However, the library does contain clones from both homologues and it is prudent to include a Southern blot of a PFGE separation of the chromosomes in all walking steps: if there is a clean breakpoint at which the sequence of the `homologues' diverges, this will become clear. In addition, there are genes which exist at more than one locus and which hybridize to two chromosome pairs, such as the PARP gene (Mowatt and Clayton, 1987). Clones deriving from all four chromosomes are likely to be included in the library but Southern hybridization to PFGE separations should show if the walk has gone astray.

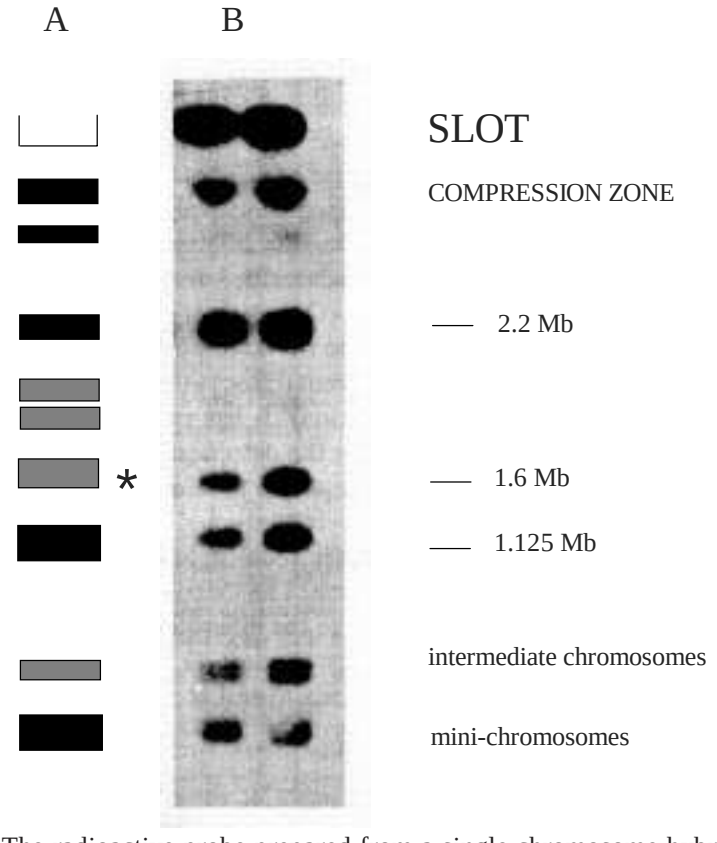

Figure 2. The radioactive probe prepared from a single chromosome hybridizes to most but not all other chromosomes within the same strain. Panel A shows the megabase chromosomes of *T. b. brucei* clone F532/72 mcl 7 in diagrammatic form (see also Figure 1) (*chromosome from which the probe was prepared). Panel B shows the result of Southern hybridization of the chromosomal probe to a PFGE separation of the chromosomes of F532/72 mcl 7.

THE SELECTION OF MARKERS FOR THE ANALYSIS OF CHROMOSOME STRUCTURE AND INHERITANCE

A systematic analysis of chromosomal breakpoints and meiotic behaviour can best be provided by developing sets of suitably dispersed, linked physical markers for individual chromosomes. The cloning of the chromosome in cosmids, P1s or YACs provides simple markers in the form of end-probes. These can be converted into polymerase chain reaction (PCR) based sequence-tagged sites (STSs) by sequencing and oligonucleotide synthesis. Further STSs internal to the cosmid, P1 or YAC clones may be generated by subcloning into plasmids, selecting PCR-sized inserts, checking for repetitive sequences and sequencing across the cloning junctions (see Blackwell *et al.*, this volume).

However, for the analysis of meiotic recombination and chromosome inheritance amongst the progeny derived from laboratory crosses, polymorphic markers are required.

These may also be required when mapping a chromosome using a chromosome-specific library: if some of the extra DNA content in the larger of a pair of homologues is in the form of DNA insertions throughout the chromosome, then polymorphic markers will allow us to look at this by distinguishing between clones from different chromosomes. RFLP markers have been used extensively in the analyses of crosses but newer methods make it possible to detect any single base substitution (see Blackwell *et al.*, this volume). This greatly increases the efficiency of polymorphic marker selection.

Other types of polymorphic markers have been described, for example variable numbers of tandem repeats (VNTRs) and microsatellites (CA repeats) (Jeffries, 1986; Weber and May, 1989). We have located a VNTR sequence on the chromosome under study: it is chromosome-specific and polymorphic in the parental stocks (M.P. Barrett, S.E. Melville, J.P. Sweetman and R.W.F. Le Page, in preparation). It has proven useful not only as a marker for chromosome inheritance, but also for the detection of trisomy in some of the hybrid progeny of the cross. Trisomy has been reported previously (Le Page *et al.*, 1988; Gibson *et al.*, 1992) and may be a common occurrence amongst the progeny derived from laboratory crosses. It remains to be seen whether such organisms exist in the wild, or whether they require human intervention to survive. The location of further chromosome-specific VNTR markers would allow us to analyze the inheritance of each of the parental chromosomes in the hybrid progeny of laboratory crosses.

ACKNOWLEDGEMENTS

This work was supported by the Overseas Development Administration, UK. We are indebted to Prof. A. Tait of the Wellcome Unit of Molecular Parasitology and Dr. C.M.R. Turner of the Department of Zoology, University of Glasgow, for the cloned stocks of field isolates STIB 247-L and TREU 927 and the cloned hybrid stocks derived from a laboratory cross. We also thank Prof. Tait for his advice on karyotype analysis by pulsed-field gel electrophoresis.

REFERENCES

BORST, P., Van der PLOEG, L.H.T., Van HOEK, J.F., TAS, J. and JAMES, J. 1982. On the DNA content and ploidy of trypanosomes. *Molecular and Biochemical Parasitology* 6: 13–23.

BURKE, D.T., CARLE, G.F. and OLSON, M.V. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.

CLAYTON, C.E. 1988. The molecular biology of the kinetoplastidae. In: Rigby, P.W.J., ed. *Genetic Engineering* 7. London: Academic Press, pp. 1–56.

GIBSON, W. and GARSIDE, L.H. 1991. Genetic exchange in *Trypanosoma brucei brucei:* variable chromosomal location of housekeeping genes in different trypanosome stocks. *Molecular and Biochemical Parasitology* 45: 77–90.

GIBSON, W., GARSIDE, L. and BAILEY, M. 1992. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei. Molecular and Biochemical Parasitology* 52: 189–200.

GIBSON, W., OSINGA, K.A., MICHELS, P.A.M. and BORST, P. 1985. Trypanosomes of the subgenus *Trypanozoon* are diploid for housekeeping genes. *Molecular and Biochemical Parasitology* 16: 231–242.

GIBSON, W.C. 1989. Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei. Parasitology* 99: 391–402.

GIBSON, W.C. and BORST, P. 1986. Size-fractionation of the small chromosomes of *Trypanozoon* and *Nannomonas* trypanosomes by pulsed-field gradient gel electrophoresis. *Molecular and Biochemical Parasitology* 18: 127–140.

GOTTESDIENER, K., GARCIA-ANOVEROS, J., LEE, M.G.-S. and Van der PLOEG, L.H.T. 1990. Chromosome organisation of the protozoan *Trypanosoma brucei. Molecular Cell Biology* 10 (11): 6079–6083.

HASAN, G., TURNER, M.J. and CORDINGLEY, J.S. 1984. Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei. Cell* 37: 333–341.

ISH-HOROWICZ, D. and BURKE, J.F. 1981. Rapid and efficient cosmid vector cloning. *Nucleic Acids Research* 9: 2989.

JEFFRIES, A.J., WILSON, V., THEIN, S.L., WEATHERALL, D.J. and PONDER, B.A.J. 1986. DNA `fingerprints' and segregation analysis of multiple markers in human pedigrees. *American Journal of Human Genetics* 39: 11–24.

JENNI, L. MARTI, S., SCHWEIZER, J., BETSCHART, B., Le PAGE, R.W.F., WELLS, J.M., TAIT, A., PAINDAVOINE, P., PAYS, E. and STEINERT, M. 1986. Hybrid formation between African trypanosomes during cyclical transmission. *Nature* 322: 173–175.

KIMMEL, B.E., ole-MOIYOI, O.K. and YOUNG, J.R. 1987. Ingi, a 5.2 kb dispersed sequence element from *Trypanosoma brucei* that carries half a smaller mobile element at either end and has homology with mammalian LINEs. *Molecular Cell Biology* 7: 1465–1475.

Le PAGE, R.W.F., WELLS, J.M., PROSPERO, T.D. and STERNBERG, J., 1988. Genetic analysis of hybrid *T. brucei.* In: Turner, M.J. and Arnot, D., eds. *Current Communications in Molecular Biology: Molecular Genetics of Parasitic Protozoa*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, pp. 65–71.

LIU, A.Y.C., Van der PLOEG, L.H.T., RIJSEWIJK, F.A.M. and BORST, P. 1983. The transposition unit of VSG gene 118 of *Trypanosoma brucei:* presence of repeated elements at its border and absence of promoter-associated sequences. *Journal of Molecular Biology* 167: 57–75.

MOWATT, M.R. and CLAYTON, C.E. 1987. Developmental regulation of a novel repetitive protein of *Trypanosoma brucei. Molecular Cell Biology* 8: 2838–2844.

PAINDAVOINE, P., ZAMPETTI-BOSSELER, F., PAYS, E., SCHWEIZER, J., GUYAUX, M., JENNI, L. and STEINERT, M. 1986. Trypanosome hybrids generated in tsetse flies by nuclear fusion. *EMBO Journal* 5: 3631–3636.

SEALEY, L. 1985. Removal of repeated sequences from hybridization probes. *Nucleic Acids Research* 13 (6): 1905–1922.

STERNBERG, J., TAIT, A., HALEY, S., WELLS, J.M., Le PAGE, R.W.F., SCHWEIZER, J. and JENNI, L. 1988. Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. *Molecular and Biochemical Parasitology* 27: 191–200.

STERNBERG, J., TURNER, C.M.R., WELLS, J.M., RANFORD-CARTWRIGHT, L.C., Le PAGE, R.W.F. and TAIT, A., 1989. Gene exchange in African trypanosomes: frequency and allelic segregation. *Molecular and Biochemical Parasitology* 34: 269–280.

STERNBERG, N. 1990. Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobase pairs. *Proceedings of the National Academy of Sciences of the USA* 89: 103–107.

TAIT, A. and TURNER, C.M.R. 1990. Genetic exchange in *Trypanosoma brucei. Parasitology Today* 6 (3): 70–75.

TURNER, C.M.R., STERNBERG, J., BUCHANAN, N., SMITH, E., HIDE, G. and TAIT, A. 1990. Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitology* 101: 377–386.

Van der PLOEG, L.H.T., CORNELISSEN, A.W.C.A., BARRY, J.D. and BORST, P. 1984a. Chromosomes of Kinetoplastida. *The EMBO Journal* 3 (13): 3109–3115.

Van der PLOEG, L.H.T., CORNELISSEN, A.W.C.A., MICHELS, P.A.M. and BORST, P. 1984b. Chromosome rearrangements in *Trypanosoma brucei. Cell* 39: 213–221.

Van der PLOEG, L.H.T., SCHWARTZ, D.C., CANTOR, C.R.C. and BORST, P. 1984c. Antigenic variation in *Trypanosoma brucei* analysed by electrophoretic separation of chromosome-sized DNA molecules. *Cell* 37: 77–84.

WEBER, J.L. and MAY, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44: 388–396.

WELLS, J.M., PROSPERO, T.D., JENNI, L. and Le PAGE, R.W.F. 1987. DNA contents and molecular karyotypes of hybrid *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* 249: 103–106.

Summary of discussion

Chairperson: *Prof. D. Walliker*
Rapporteurs: *Drs. P. Toye and O. ole-MoiYoi*

Five presentations were made in this session. The first presentation was made by Dr. P. Gardiner (ILRAD), who gave an overview of the enormous differences among the trypanosomes and described interesting and wide-ranging biochemical processes present in these organisms. Dr. P. Majiwa (ILRAD) focused on the complexity of classifying the *Trypanosoma congolense* group of parasites and described the methodologies, such as simple morphometry, isoenzyme patterns, RFLP analysis and pattern of hybridization of repetitive DNA sequences, for characterization of these parasites. Prof. A. Tait (Glasgow University) first presented evidence for genetic exchange in *T. brucei* spp. and then showed the inheritance patterns of chromosomes by using chromosome-specific markers. He also discussed mating behaviour and selfing in *Trypanosoma brucei* spp., i.e. mating between members of a single stock. Dr. S. Melville (Cambridge University) described production of subgenomic libraries specific for pulsed-field gel bands, which simplifies chromosome mapping, to identify and characterize chromosomes of trypanosomes. A novel approach to rapid identification of genes controlling important characteristics in the trypanosome genome was described by Dr. N. Murphy (ILRAD).

There was a short discussion on each of the papers presented and then a more general discussion followed. This centred round the occurrence of polymorphisms, genetic exchange and the significance of these on important traits, such as drug resistance in trypanosomes. It was pointed out that the wide polymorphisms observed in trypanosomes was not peculiar to this parasite, as numerous other organisms show similar polymorphisms. Examples cited were *Theileria, Plasmodia* and *Candida.* One concern raised with regard to defining polymorphisms in field populations of trypanosomes was lack of correlation between molecular fingerprint and function. It was recognized that characterization of field populations will provide better knowledge of the variability of the parasites that exist in nature and improve our understanding of the epidemiology. However, because there is a sense of urgency and resources are limited, there is a need to set priorities and clearly define goals for such studies. One suggestion was that a more systematic approach involving the generation of chromosome specific markers and construction of a physical map would be more productive, although this would be a long-term study. In contrast, in the short term, a laboratory-based analysis using genetic crosses to examine specific traits, such as drug resistance, coupled with DNA transfection technology could yield useful information rapidly. The advantages and disadvantages of both the broad, systematic approach and a specific approach using laboratory-based analysis of crosses and transfection technology were discussed.

With regard to the specific approach, one question raised was that although DNA can be transfected into parasites, the knowledge of the genome structure is still necessary to

understand the integration of the foreign DNA into the genome. This is particularly true for multigenic traits, such as is likely for resistance to the trypanocide diminazene aceturate. Developing transfection vectors for studies on functions of important genes was considered time consuming; in the long run, it may be quicker to use genetic analysis and genome maps to achieve the same result. One possibility was to use a combination of approaches where genetic maps can be resolved around key biological traits such as trypanotolerance of N'Dama cattle and drug resistance. However, the crucial biological traits that need to be studied must be determined. It was recognized that the task of constructing a complete physical map of the trypanosome genome was going to be much harder than that for *T. parva* and *Plasmodium falciparum* because of the larger size of the trypanosome genome. Nonetheless, a `gold standard' genome map of *T. brucei* would be a valuable resource and may also be useful for *T. congolense.* In terms of resources required to map eukaryotic genomes, it has been estimated that a *P. falciparum* contig map would take five laboratories three years to complete. It was pointed out that an overlapping contig map of *Saccharomyces pombe* took one post-doctoral fellow only two years to complete. After a lengthy discussion, it was agreed that both approaches have a place and in some ways are synergistic.

TECHNIQUES FOR PHYSICAL AND FUNCTIONAL ANALYSIS

Markers for mapping trypanotolerance genes

S.J. Kemp

Several West African *Bos taurus* cattle breeds show a remarkable ability to survive and be productive in the face of a trypanosome challenge which would quickly kill susceptible cattle. This phenomenon, termed trypanotolerance, offers a sustainable approach to improving cattle productivity in the tsetse-infested areas of Africa. The best characterized of the trypanotolerant breeds is the N'Dama. We are adopting a `genome analysis' approach in an attempt to identify the genetic basis for trypanotolerance in the N'Dama.

A cross between N'Dama and trypanosusceptible Boran *(Bos indicus)* cattle has been established using embryo-transfer to generate large full-sibling families. The F_1 generation is almost complete and the first F_2 calves were born in late 1992.

These will be challenged with *Trypanosoma congolense* and, in collaboration with several other laboratories, they will be genotyped with a large array of genetic markers. Correlations will then be sought between marker inheritance and trypanotolerance status. These markers fall into three main categories.

- Randomly amplified DNA polymorphisms (RAPDs).
- Randomly identified polymorphic dinucleotide repeat sequences (microsatellites).
- Polymorphisms in specific genes.

RANDOMLY AMPLIFIED DNA POLYMORPHISM MARKERS

Randomly amplified polymorphic DNA is detected by polymerase chain reaction (PCR) amplification of genomic DNA using short arbitrary primers (Welsh and McClelland, 1990; Williams *et al.*, 1990). This generates a complex `fingerprint' which is often highly polymorphic. Markers of this type have the advantage that they require no prior sequence information and so may be readily applied to any organism. The usefulness of RAPD markers for detecting polymorphism in N'Dama and Boran has been demonstrated. An important refinement to RAPD analysis, bulked segregant analysis (Michelmore *et al.*, 1991) is presently being tested in crosses of inbred mice which differ in survival time following *T. congolense* challenge. Bulked segregant analysis may also have an application in the analysis of crosses between outbred, but distinct, cattle breeds such as N'Dama and Boran.

MICROSATELLITE MARKERS

Microsatellites are regions of short sequence repeats which appear to be scattered through mammalian genomes and which are highly polymorphic (Weber and May, 1989; Beckman and Weber, 1992). PCR primers are designed to specifically amplify single microsatellites

and polymorphism is revealed by high resolution electrophoresis. It is expected that such markers, generated here and elsewhere, will provide the bulk of markers on the bovine genome map.

SPECIFIC GENES

The third category of markers being sought and applied at the International Laboratory for Research on Animal Diseases (ILRAD) utilizes polymorphisms in specific `candidate' genes. These genes are selected from the databases on the basis of their physical location and their potential involvement in trypanotolerance. An example is the gamma-crystallin gene which is located in an important linkage group on bovine chromosome 8. PCR primers are designed to amplify untranslated regions of these genes from genomic DNA of both N'Dama and Boran cattle and the amplified fragments are sequenced. Sequence polymorphisms are detected and PCR-based assays developed to rapidly screen animals for this polymorphism.

It is hoped that this approach will, in the short term, provide markers of genes controlling trypanotolerance and, in the longer term, allow the identification and study of these genes. This process will be greatly facilitated by the known conservation of synteny between the human, murine and bovine genomes.

REFERENCES

BECKMAN, J.S. and WEBER, J.L. 1992. Survey of human and rat microsatellites. *Genomics* 12: 627-631.

MICHELMORE, R.W., PARAN, I. and KESSELI, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis—a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the USA* 88: 9828–9832.

WEBER, J.L. and MAY, P.E. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44: 388–396.

WELSH, J. and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 7213–7218.

WILLIAMS, J.G.K., KUBELIK, A.R., LIVAK, K.J., RAFALSKI, J.A. and TINGEY, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531–6535.

Derivation of large insert clones from human chromosome 22 using the bacteriophage P1 vector system

N.S. Shepherd, J.N. Coulby and S.L. Ackerman

The creation of recombinant molecules to form reference libraries is basic to understanding the biology of an organism at the molecular level. Since fewer genomic libraries are likely to be generated for protozoan parasites than for `model' organisms such as the human or mouse, it is very important to define the advantages and disadvantages of the various approaches. In this report, we illustrate the use of the relatively new P1 vector cloning system to generate recombinant clones of a given region of a genome. The example provided is the generation of large insert clones of human chromosome 22 from a hamster/human hybrid cell line. The advantages of the P1 cloning system and alternative approaches to obtain P1 recombinant clones of the region are discussed.

INTRODUCTION

Human chromosome 22 is one of the smallest chromosomes, representing approximately 2% of the estimated 3×10^6 kb of the human haploid genome (Kaplan *et al.,* 1987). This small size, together with the many pathologic conditions associated with alterations of the chromosome, make it an ideal model system for studies of chromosome organization, studies of constitutional and acquired chromosomal rearrangements, as well as studies to identify and characterize an estimated 600 (or more) human genes on the chromosome. These studies would be facilitated by the construction of a set of genomic clones and an ordering of the clones into overlapping contiguous subsets (contigs) corresponding to their native chromosomal location. Clearly, the larger the size of the insert DNA present in each recombinant clone, the less effort involved in developing the overlapping physical map. Although the yeast artificial chromosome (YAC) cloning system offers the advantage of extremely large inserts from several hundred kilobases to greater than a megabase (Burke *et al.,* 1987; Albertsen *et al.,* 1990; Anand *et al.,* 1989; Abidi *et al.,* 1990; Larin *et al.,* 1991; Bellanné-Chantelot *et al.,* 1992), the disadvantages of that system have become apparent. Genomic reference libraries created in YAC vectors often have a relatively high percentage of chimeric clones (i.e., two noncontiguous genomic fragments present on the same clone) (Abidi *et al.,* 1990; Bellanné-Chantelot *et al.,* 1992). Furthermore, the inability to separate the YAC DNA from that of the host makes techniques such as restriction mapping, subcloning of fragments and reintroduction into mammalian cells somewhat cumbersome. Nevertheless, YAC reference libraries continue to play a major role in genome mapping studies and the polymerase chain reaction (PCR) has increased their utility as a quick source of region specific probes. Several reference libraries specific for human chromosome 22 have already been generated using DNA obtained from either

flow-sorted chromosome 22 or a somatic cell hybrid containing the chromosome (McDermid *et al.*, 1989; de Jong *et al.*, 1989; Kawasaki *et al.*, 1992). However, these existing chromosome-specific libraries are less than optimal for several reasons. First, because they were generated in plasmid, λ and cosmid type vectors with insert sizes ranging from several thousand base pairs to approximately 45 kb, these libraries are less useful for building a physical map. In addition, this size limitation may result in individual clones containing only a portion of a gene since human genes are often 40 kb or larger. Furthermore, the instability of cosmid clones propagated as multi-copy plasmids in the host bacterial cell has been documented (Kim *et al.*, 1992). We chose to explore the use of the recently described bacteriophage P1 vector system as a means to obtain clones specific to human chromosome 22. This system, first described by Sternberg (1990), offers several advantages (Sternberg, 1992). Firstly, it allows the cloning of 75–100 kb inserts, a size significantly greater than that of cosmids. Secondly, the cloning efficiency (10^4 to 10^5 clones/μg of vector arms) is sufficiently high such that libraries of relatively large genomes are possible (Sternberg *et al.*, 1990; Pierce *et al.*, 1992a; N.S. Shepherd, B.D. Pfrogner, J.N. Coulby, S.L. Ackerman, G. Vaidyanathan, R.H. Sauer, T.C. Balkenhol and N. Sternberg, personal communication). Thirdly, P1 recombinant clones can be maintained as single copy plasmid replicons in the cell, thus certain genomic sequences may be more stable when cloned in the P1 vector as compared to a multi-copy plasmid or cosmid vector. Finally, unlike YACs, the recombinant DNA can be easily recovered and separated from that of the host. In addition, the new vector derivative, pAd10-SacBII, provides a positive selection for recombinant clones, easy excision of the cloned insert, and the generation of riboprobes from the ends of the insert (Pierce *et al.*, 1992b).

We initiated projects to explore three different methods of creating the desired chromosome-specific library in the P1 bacteriophage vector system. In this report we describe the preparation of P1 recombinant clones from a somatic cell hybrid which contains chromosome 22 as the sole component of human DNA, and the subsequent screening of clones to identify those containing human DNA. In the discussion we compare this approach to that of cloning from DNA purified from flow-sorted chromosome 22 or isolation of chromosome-specific clones from a total human P1 recombinant library.

MATERIALS AND METHODS

Preparation of High Molecular Weight DNA

Approximately 6×10^7 cells of the KG1 hybrid (GM10888 of The Human Genetic Mutant Cell Repository of Coriell Institute in Camden, New Jersey, USA—NIH publication number 92-2011, 1992) were trypsinized, harvested by gentle centrifugation, and washed several times in prewarmed phosphate buffered saline (PBS without magnesium and calcium, Flow Labs, McLean, Virginia 22102, USA). The final cell pellet was resuspended in 500 μl warm PBS, 0.25 ml of 10 mg/ml proteinase K (Boehringer Mannheim GmbH); 1.5 ml 250 mM EDTA pH8 and 0.25 ml 10% SDS. Following incubation for two hours at 60 °C, 0.8 ml 4M NaCl was added and the solution was stored overnight at 4 °C. The solid white solution was then dialized for several days at

4 °C against sterile TE buffer (10 mM Tris pH8, 1 mM EDTA) to completely eliminate the proteinase K and SDS. Restriction enzyme digestion, size fractionation on a sucrose gradient, and the concentration and dialysis of fractions before cloning was essentially as described previously (Sternberg, 1990; Sternberg *et al.*, 1990).

Preparation of Vector Arms

DNA of the pAD10-SacBII vector (Pierce *et al.*, 1992b) was prepared as described in Pierce and Sternberg (1992) from the bacterial strain NS3622, which contains the vector in the host NS3607 (Pierce *et al.*, 1992b). Dephosphorylated vector arms with *Bam*HI ends at the cloning site were made as described previously (Pierce *et al.*, 1992a) with only slight modifications. In brief, 5 μg vector DNA was digested to completion at the unique *Sca*I site (Pharmacia enzyme), treated with bacterial alkaline phosphatase (Gibco BRL, 25 units/μg vector) for an hour at 65 °C. The DNA was extracted with phenol/chloroform, ethanol precipitated and resuspended in TE buffer. The vector cloning site was generated by digesting the purified DNA with *Bam*HI (New England Biolabs) followed immediately by dephosphorylation with calf intestinal phosphatase (Boehringer Mannheim, 0.002 units/μg vector) for 30 minutes at 37 °C. After phenol/chloroform extraction and ethanol precipitation, the vector arms were resuspended in TE buffer at a concentration of 0.5 μg/μl and stored at 4 °C.

In vitro Ligation and P1 Packaging Reactions

DNA ligations using a three-fold excess of dephosphorylated vector arms to human DNA insert were routinely performed with 0.5 μg vector in a final volume of 16 μl using T4 DNA ligase (New England Biolabs). Prior to addition of enzyme and buffer, the two DNAs were preincubated at 37 °C for 30 minutes and then cooled on ice. Following overnight incubation at 16 °C, the reaction mixture was heated at 70 ⌈C for 10 minutes, stored at 4 °C, and used within a week.

In vitro packaging of the ligation products was carried out with bacterial packaging extracts obtained from Du Pont/NEN (pacase extract lot #LA136 and head/tail extract lot #PB093) as suggested by supplier and Sternberg (1990), with 2–4 μl of the ligation products per packaging reaction. The packaged phage were stored in the absence of chloroform at 4 °C and routinely used within one week. Between 3 and 15 μl of each phage lysate was used for infection of 100 μl of the *Escherichia coli* plating host NS3529 *[recA-, mcrA-, Δ(hsdR, hsdM, mcrB, mrr)* (λ*imm*λ*LP1)* (λ*imm434-P1: Cre+)]* prepared as described previously (Pierce and Sternberg, 1992). After phage adsorption at 37 °C for 15 minutes without shaking in a 1.5 ml tube heat block (Eppendorf #5436), 0.8 ml L broth (Sambrook *et al.*, 1989) was added and incubation continued at 37 °C for 35 minutes with moderate shaking. Cells were centrifuged for 30 seconds and all but approximately 50 μl of the supernatant was removed. Cells were resuspended and the entire contents of each tube was plated on a single LB plate supplemented with 25 μg/ml kanamycin and 5% sucrose, and incubated for a minimum of 18 hours at 37 °C.

Colony Hybridization

Colonies were transferred to nylon-backed, nitrocellulose NitroPure membranes (Micron Separations Inc., Westboro, Massachusetts, USA). The original plate was reincubated for four to six hours to regenerate the transferred colonies. The filters were processed by standard colony lysis and hybridization conditions (Sambrook *et al.*, 1989). Prehybridization was three hours at 68 °C in 6 × SSC, 1× Denhardt's solution with 100 μg/ml of non-radioactive, sheared, salmon sperm DNA. Hybridization was overnight under similar conditions but in addition contained 10% dextran sulfate. The ^{32}P-labelled hybridization probes were prepared by random primer labelling (Feinberg and Vogelstein, 1983; Boehringer Mannheim kit) DNA prepared from human foreskin fibroblasts (Viromed Lab., Minnetonka, Minnesota, USA) or a CHO cell line. Approximately 1–3 × 10^6 cpm of probe was used per ml of hybridization buffer. The filters were washed for two to three hours at 68 °C with a solution of 0.1% SDS and SSC 2 × (initial)–0.5 × (final), were dried at room temperature, and exposed to Kodak Xomat-AR film.

RESULTS

Preparation of P1 Recombinant Clones from the KG1 Hybrid

The high molecular weight DNA required for the P1 cloning process was isolated from cells of the hamster/human somatic cell hybrid KG1. This cell line, also referenced as GM10888 (The Human Genetic Mutant Cell Repository of Coriell Institute, Camden, New Jersey, USA—NIH publication number 92-2011, 1992), has been documented to have human chromosome 22 as the only human DNA (Lichter *et al.*, 1990). To generate the appropriately sized, random fragments for cloning, the DNA was partially restricted with *Sau*3AI, size fractionated once through a 5–30% linear sucrose gradient (Sternberg *et al.*, 1990), and analyzed by field inversion gel electrophoresis (Figure 1). Fractions containing a large proportion of DNA in the 50–150 kb size range were dialysed, concentrated with butanol, spot dialysed and stored at 4 °C. The genomic DNA fractions were ligated to the pAd10sacBII vector arms, packaged into the P1 phage head *in vitro*, and used to infect the *E. coli* plating host (Sternberg, 1990). By a site-specific recombination system, the injected DNA is circularized to form a single copy plasmid replicon within the host (Sternberg, 1990, 1992). Cells containing recombinant molecules were selected for by overnight growth at 37 °C on LB plates supplemented with kanamycin and 5% sucrose (Pierce *et al.*, 1992b).

Identification of Human Clones

Since the hamster and human genome are similar in size, the human chromosome 22 component of the KG1 hybrid is only about 2%. To identify those P1 recombinant clones containing human DNA, colonies were transferred to nitrocellulose filters, processed for colony hybridization, and hybridized to a radioactive probe prepared from total human

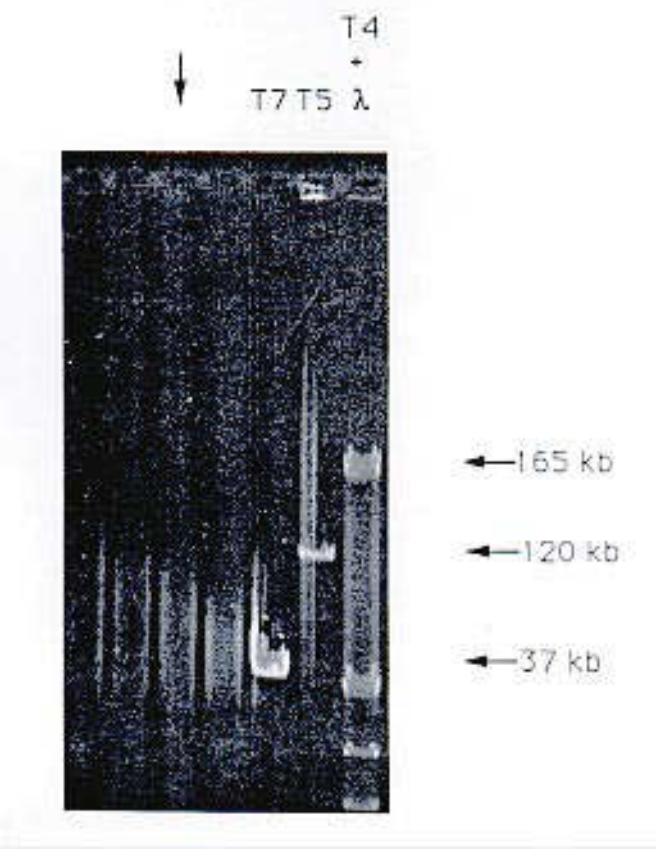

Figure 1. Pulsed-field gel electrophoresis for size analysis of genomic DNA insert fractions. The first four lanes contain consecutive fractions taken from a 5–30% sucrose gradient separation of *Sau*I3AI partially digested genomic DNA. The vertical arrow indicates the fraction from which clones KG1-01 to KG1-25 were obtained. Electrophoresis was with 1% SeaKem GTG-grade agarose (FMC) and 0.5 × TBE buffer (Sambrook *et al.*, 1989) at room temperature using a Hoefer PC750 pulse controller (100 volts, 0.3 sec forward and 0.1 sec reverse, ramp .15 for 17 hours). DNA size standards are 37 and 120 kb of bacteriophage T7 and T5 genomes respectively (Sigma); and the 165 kb of bacteriophage T4 genome (Sigma) in the same lane as a *Hind*III digest of λ DNA (only 23, 9.4 and 6.6 kb fragments are shown; BRL).

DNA. Only 0.4% of the total clones prepared with the KG1 genomic DNA as insert hybridized strongly to the probe (Figure 2). This low value was not due to a high proportion of clones without insert DNA. Twenty-five strongly hybridizing positives were picked, labelled DMPC-KG1-01 through -25 and streaked to purify to single colony. The purified clones were retested by a secondary hybridization for the presence of human DNA (Figure 3a). Although the intensity of hybridization varied, all seemed to contain human DNA with the exception of KG1-13. As expected, four random clones used for the control hybridization did not hybridize strongly to the human probe and most likely contain hamster DNA as insert.

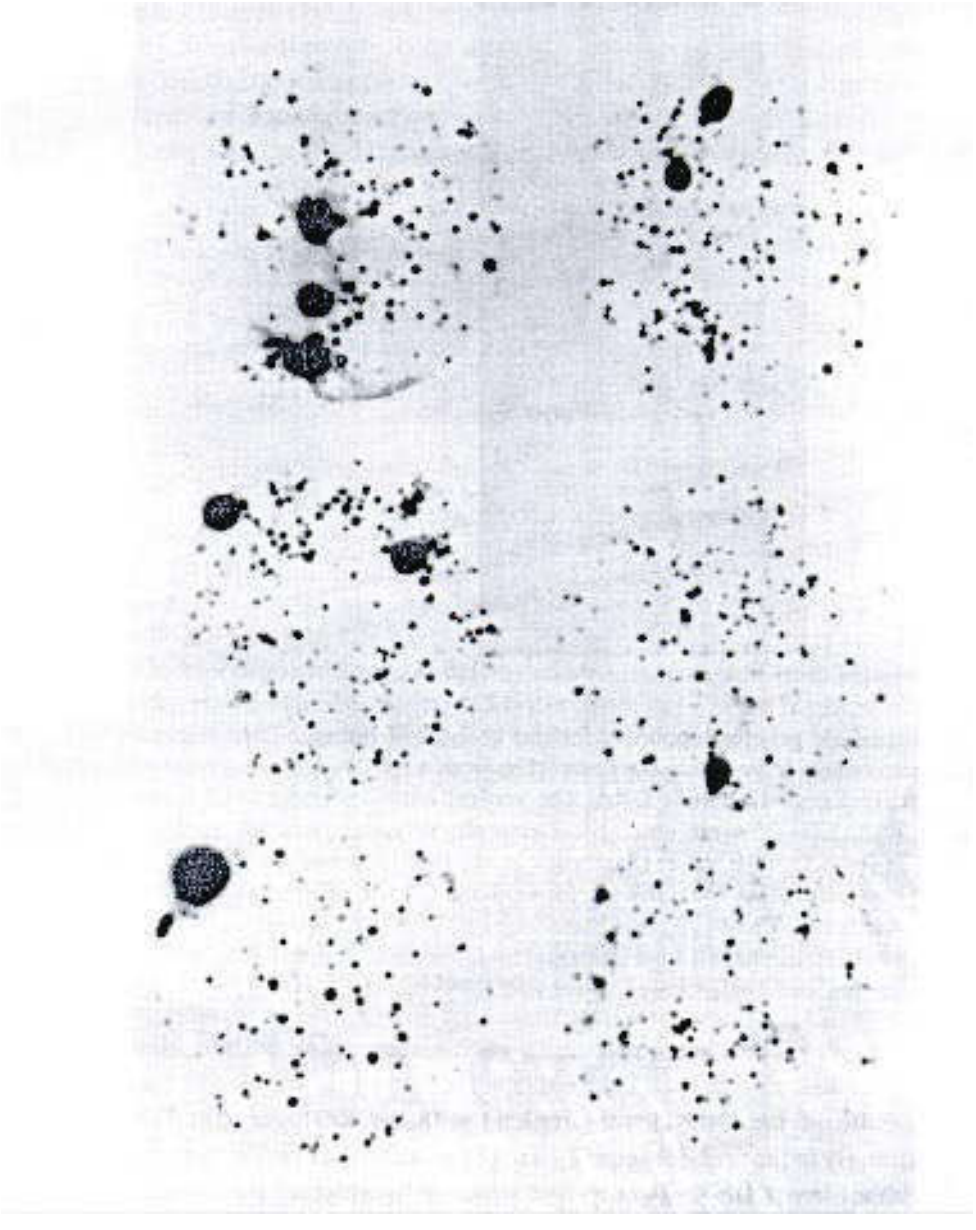

Figure 2. Initial identification of human recombinant clones using colony hybridization. P1 recombinant clones prepared from the KG1 hybrid DNA were transferred to a nitrocellulose filter and probed with 32 P-labelled total human DNA. Positives were selected and rescreened as indicated in the text.

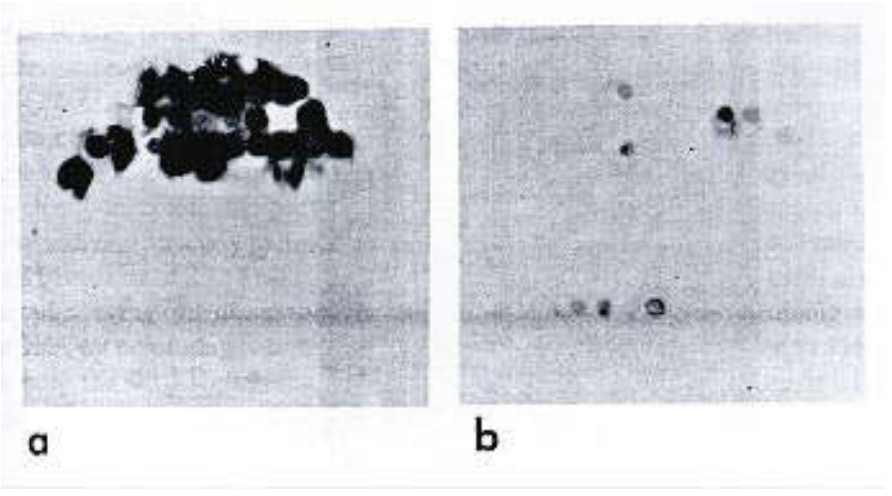

Figure 3. Colony hybridization of recombinant clones KG1-01 through KG1-25 with a ^{32}P-labelled hybridization probe from (a) total human DNA or (b) total hamster DNA.

Testing for Chimeric Clones

The use of many YAC genomic libraries is hampered by a relatively high level of chimerism, that is, two non-contiguous genomic fragments present on the same cloned insert (Abidi *et al.,* 1990; Bellanné-Chantelot *et al.,* 1992). However, there are no reports in the literature with respect to the degree of such scrambling when cloning with the P1 vector system. In principle, the human clones obtained from the hamster/human hybrid should be easy to test, for any small insert fragments that are ligated with human DNA during the *in vitro* ligation step are likely to be of hamster origin and thus detected by using total hamster DNA as a probe. Figure 3b shows the results obtained when KG1-01 through KG1-25 were hybridized to a radioactive probe prepared from total hamster DNA. Several of the 24 clones that hybridized to human DNA in Figure 3a hybridized to hamster DNA to various extents. For example, KG1-12 hybridized very strongly while KG1-01, 17 and 23 showed weak hybridization. These clones may be chimeric, containing both human and hamster sequences. However, the four randomly chosen clones present at the bottom of the filter as well as the KG1-13 clone (most likely clones with hamster DNA) hybridized poorly to the hamster probe. We therefore believe that an alternative explanation, that of weak cross-hybridization due to similar repetitive sequences shared between the two species, is possible. Finally, it should be apparent that the frequency of clone scrambling is likely to be dependent upon how prevalent small fragments are in the insert fraction of DNA and will be highly variable from the preparation of one library to that of the next. All precautions should be taken to eliminate small sized fragments of insert DNA before the ligation step.

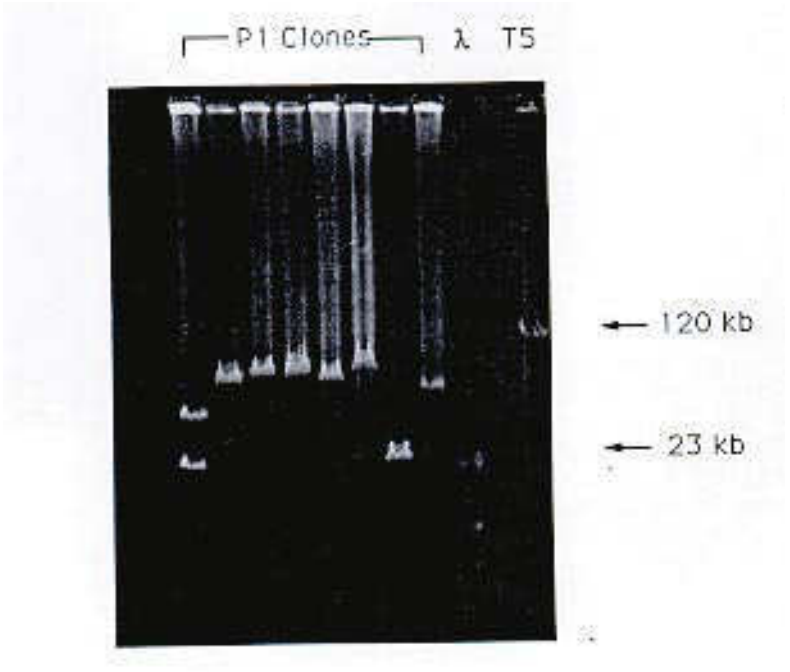

Figure 4. DNA analysis of eight P1 recombinant clones prepared from the KG1 hybrid. Plasmid DNA was prepared, digested with *Not*I restriction enzyme which cuts once in the P1 vector sequence, and subjected to field inversion gel electrophoresis as in Figure 1. Size markers were λ *Hind*III digest (only 23, 9.4 and 6.6 kb fragments are shown; BRL) and DNA of the 120 kb T5 bacteriophage genome (Sigma).

P1 Clone Analysis

Plasmid DNA of P1 recombinant clones from the KG1 hybrid was prepared, digested with *Not*I, and analyzed by field inversion gel electrophoresis (Figure 4) using methods described previously (Pierce *et al.*, 1992a; N.S. Shepherd, B.D. Pfrogner, J.N. Coulby, S.L. Ackerman, G. Vaidyanathan, R.H. Sauer, T.C. Balkenhol and N. Sternberg, personal communication). The estimated insert size of KG1-01 to -25 was determined by summing the restriction fragment sizes and subtracting the 18 kb of vector sequence present in every clone. The insert sizes varied from approximately 14 to 90 kb, with 12 clones greater than 70 kb (KG1-01, 02, 03, 04, 06, 10, 11, 12, 16, 17, 19 and 24). It should be emphasized that optimal size fractionation of the input DNA is required, for although the pAd10-SacBII vector was designed to eliminate cloning of small inserts (Pierce *et al.*, 1992b), such fragments may be cloned if present in the fraction.

DISCUSSION

We have demonstrated how P1 recombinant clones may be obtained for a specific region of the human genome by capitalizing upon the use of a hybrid cell line containing only

human chromosome 22. This general approach is certainly not new, but the utility of using the P1 vector system is quite apparent when one considers that cosmid type vectors would require more than twice the number of clones to be identified for comparable coverage of the chromosome. If the size of human chromosome 22 is taken to be 2% of the human genome (2% of 3×10^6 kb = 6×10^4 kb), then 750 P1 recombinant clones with an average insert size of 80 kb would give approximately a one-fold coverage of the chromosome. Although costly and time-consuming, generation of several fold coverage of the chromosome is clearly an obtainable goal by this approach. However, a major concern arose when clones hybridizing strongly to human DNA appeared at a frequency of 0.4% rather than the expected frequency of 2%. This suggested that a library generated by this method may result in a biased coverage of the chromosome, since many clones not hybridizing strongly to human DNA yet belonging to human chromosome 22 may be missed in the picking process. Furthermore, the weak signal obtained using total hamster DNA as probe made it impossible to eliminate unwanted clones using this probe. These factors together with the consideration of alternative approaches (see below) led us to the ultimate decision to stop short of obtaining complete chromosome coverage by this approach.

A second method to generate chromosome 22-specific P1 clones was to use DNA isolated from flow-sorted chromosome 22. Although the preparation of flow-sorted chromosomes is expensive and not universally available, the approach seemed promising since unlike cloning from the KG1 hybrid, the majority of the clones generated by this technique should contain only chromosome 22 DNA. Although preparation of high molecular weight DNA from flow-sorted chromosome 22 material was possible (S.L. Ackerman, unpublished results), the following steps of partial digestion and subsequent size fractionation with such small amounts of DNA made it difficult to obtain a large number of clones approaching the optimal insert size of the P1 vector cloning system. Furthermore, there was a potential problem of contamination of the chromosome 22 fraction by fragmentation of larger chromosomes during the sorting process. Even if the chromosome separation was performed with the KG1 hamster/human hybrid cell line such that the unwanted chromosome fragments would be of hamster origin, it would be difficult to identify the hamster clones by hybridization (Figure 3b).

The third method for obtaining the chromosome-specific clones was to generate a library of P1 recombinant clones from total human DNA and then to screen the library for clones containing chromosome 22 sequences. We have recently finished the construction of such a reference library providing an estimated threefold coverage of the genome. The library, designated DMPC-HFF #1, was prepared from primary, human foreskin fibroblast cells and consists of approximately 130–140,000 recombinant clones stored in 1,500 microtitre dishes with one clone per well (N.S. Shepherd, B.D. Pfrogner, J.N. Coulby, S.L. Ackerman, G. Vaidyanathan, R.H. Sauer, T.C. Balkenhol and N. Sternberg, personal communication). The average insert size of the clones is estimated to be 80 kb. The library may be screened using PCR or colony hybridization using probes generated for specific loci previously mapped to the chromosome (for example, see Delattre *et al.*, 1991). These P1 clones may then be used to obtain overlapping, adjacent clones from the library after generating end-labelled hybridization probes or new PCR primers. Alternatively, the library may be screened by hybridization with more complex probes generated by amplifying the chromosome 22 containing regions of DNA present

on other DNAs using primers specific for the human *Alu* repetitive element (Nelson *et al.,* 1991). A prime source of DNAs for generation of such complex probes would be large restriction fragments (for example from flow-sorted chromosomes as in Kawasaki *et al.,* 1992), somatic cell hybrids containing defined portions of human chromosome 22 as their only human component, or YAC clones already mapped to the chromosome. In this manner, the P1 clones identified would already be classified as to sub-chromosomal location thus facilitating the assembly of clones into an ordered array (contig building). The approach of creating chromosome-specific sub-libraries from the total human library has several advantages. As mentioned above, it has the advantage of obtaining groups of clones that can be immediately assigned to a region of the chromosome. A second advantage is that unlike the limited and costly source of DNA obtained from flow-sorted chromosomes, a source of total human DNA is easily obtained and allows an optimization of the cloning technique to obtain the best possible clones for the library. Finally, unlike clones prepared from DNA of the KG1 hybrid, all of the clones are of potential interest since they all originate from human DNA. The use of relatively unique DNAs as probes makes it less likely that the resulting library is biased in coverage, for a wide variety of probes are already available. In summary, we believe that the best approach is that of creating chromosome-specific sub-libraries from a total human genomic library. The DMPC-HFF #1 library is currently being distributed to several major laboratories throughout the world to make this possible.

ACKNOWLEDGEMENTS

We thank Dr. J. Biegel, Cytogenetics Laboratory and Chromosome 22 Center, Children's Hospital of Philadelphia, Pennsylvania, for the gift of the KG1 cells. We also thank Drs. J. Rosenbloom and B. Shenker of the University of Pennsylvania for the flow-sorted chromosome 22 material. We especially appreciate the advice of Dr. N. Sternberg and the support of human genome research within The Du Pont Merck Pharmaceutical Company by Dr. M. Pearson. S.L. Ackerman received funds from the Chromosome 22 Center grant, NIH HG00425.

REFERENCES

ABIDI, F.E., WADA, M., LITTLE, R.D. and SCHLESSINGER, D. 1990. Yeast construction and representation of probe sequences. *Genomics* 7: 363–376.

ALBERTSEN, H.M., ABDERRAHIM, H., CANN, H.M., DAUSSET, J., Le PASLIER, D. and COHEN, D. 1990. Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proceedings of the National Academy of Sciences of the USA* 87: 4256–4260.

ANAND, R., VILLASENTE, A. and TYLER-SMITH, C. 1989. Construction of yeast artificial libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucleic Acids Research* 17: 4325–4333.

BELLANNÉ-CHANTELOT, C., LACROIX, B., OUGEN, P., BILLAULT, A., BEAUFILS, S., BERTRAND, S., GEORGES, I., GILBERT, F., GROS, I., LUCOTTE, G., SUSINI, L., CODANI, J.-J., GESNOUIN, P., POOK, S., VAYSSEIX, G., LU-KUO, J., RIED, T., WARD, D., CHUMAKOV, I.,

LE PASLIER, D., BARILLOT, E. and COHEN, D. 1992. Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* 70: 1059–1068.

BURKE, D.T., CARLE, G.F. and OLSON, M.V. 1987. Cloning of large exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.

DELATTRE, O., AZAMBUJA, C.J., AURIAS, A., ZUCMAN, J., PETER, M., ZHANG, F., HORSCAYLA, M.C., ROULEAU, G. and THOMAS, G. 1991. Mapping of human chromosome 22 with a panel of somatic cell hybrids. *Genomics* 9: 721–727.

FEINBERG, A.P. and VOGELSTEIN, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* 132: 6–13.

de JONG, P.J., YOKOBATA, K., CHEN, C., LOHMAN, F., PEDERSON, L., McNINCH, J. and van DILLA, M. 1989. Human chromosome-specific partial digest libraries in λ and cosmid vectors. *Cytogenetics and Cell Genetics* 51: 985.

KAPLAN, J.C., AURIAS, A., JULIER, C., PRIEUR, M. and SZAJNERT, M.J. 1987. Human chromosome 22. *Journal of Medicine Genetics* 24: 65–78.

KAWASAKI, K., MINOSHIMA, S., KUDOH, J. and SHIMIZU, N. 1992. Shotgun polymerase chain reaction: construction of clone libraries specific to a *Not*I fragment of flow-sorted human chromosome 22. *Genomics* 13: 109–114.

KIM, U.-J., SHIZUYA, H., de JONG, P.J., BIRREN, B. and SIMON, M.I. 1992. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Research* 20: 1083–1085.

LARIN, Z., MONACO, A.P. and LEHRACH, H. 1991. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proceedings of the National Academy of Sciences of the USA* 88: 4123–4127.

LICHTER, P., LEDBETTER, S.A., LEDBETTER, D.H. and WARD, D.C. 1990. Fluorescence *in situ* hybridization with *Alu* and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. *Proceedings of the National Academy of Sciences of the USA* 87: 6634–6638.

McDERMID, H.E., BUDARF, M.L. and EMANUEL, B.S. 1989. Toward a long-range map of human chromosomal band 22q11. *Genomics* 5: 1–8.

NELSON, D.L., BALLABIO, A., VICTORIA, M.F., PIERETTI, M., BIEW, R.D., GIBBS, R.A., MALEY, J.A., CHINAULT, A.C. WEBSTER, T.D. and CASKEY, C.T. 1991. *Alu* primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human) chromosome: identification of clones associated with a disease locus. *Proceedings of the National Academy of Sciences of the USA* 88: 6157–6161.

PIERCE, J.C. and STERNBERG, N.L. 1992. Using bacteriophage P1 system to clone high molecular weight genomic DNA. *Methods in Enzymology* 216: 549–574.

PIERCE, J.C., SAUER, B. and STERNBERG, N. 1992a. A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy. *Proceedings of the National Academy of Sciences of the USA* 89: 2056–2060.

PIERCE, J.C., STERNBERG, N. and SAUER, B. 1992b. A mouse genomic library in the bacteriophage P1 cloning system: organization and characterization. *Mammalian Genome* 3 (10): 550–558.

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Laboratory Press.

STERNBERG, N. 1990. Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobase pairs. *Proceedings of the National Academy of Sciences of the USA* 87: 103–107.

STERNBERG, N., RUETHER, J. and de RIEL, K. 1990. Generation of a 50,000 member human DNA library with an average insert size of 75–100 kb in a bacteriophage P1 cloning vector. *The New Biologist* 2: 151–162.

STERNBERG, N.L. 1992. Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends in Genetics* 8: 11–16.

Yeast artificial chromosomes: cloning and analysis of complex genomes

G.A. Silverman, K.M. Lia and S.S. Schneider

Yeast artificial chromosome (YAC) cloning systems permit the propagation of 0.1 to 2 megabases (Mb) of foreign DNA in yeast. They are considered to be an ideal tool for genomic analysis by filling the gap between higher-resolution bacterial cloning and lower-resolution genetic mapping. YAC cloning has assumed a prominent role in the efforts to assemble contiguous sets of clones (contigs) spanning the chromosomes of lower and higher eukaryotes. Recent examples include the cloning of the euchromatic region of the human Y chromosome (~40 Mb) and the entire long-arm of human chromosome 21 (~50 Mb). YACs are also assuming a role in the strategies being developed to solve other aspects of genome analysis. Several techniques using YACs are now available for 1) cloning and characterizing functional domains such as telomeres, centromeres and amplicons, 2) isolating new genes and polymorphic DNA fragments, 3) re-constructing and manipulating (mutating) large DNA molecules, and 4) transferring genomic fragments to appropriate hosts for expression and complementation studies.

INTRODUCTION

The advent of yeast artificial chromosome (YAC) cloning systems (Burke *et al.,* 1987) provides a means to span the gap between higher-resolution molecular cloning and lower-resolution genetic mapping and cytogenetics. The cloning capacity of YACs exceeds, by 50- to 100-fold, that of conventional prokaryotic cloning systems (e.g. cosmid, λ phage and plasmid vectors). Indeed, YAC libraries are facilitating the analysis of complex genomes such as *Arabidopsis thaliana* (Guzman and Ecker, 1988), *Caenorhabditis elegans* (Coulson *et al.,* 1988), *Daucus carota* (Guzman and Ecker, 1988), *D. discoideum* (Kuspa *et al.,* 1992), *Dictyostelium melanogaster* (Garza *et al.,* 1989), *Homo sapiens* (Brownstein *et al.,* 1989), *Lycopersicon esculentum* (Martin *et al.,* 1992), *Mus musculus* (Burke *et al.,* 1991), *Myxococcus xanthus* (Kuspa *et al.,* 1989), *Plasmodium falciparum* (Triglia and Kemp, 1991), *Schizosaccharomyces pombe* (Maier *et al.,* 1992), and *Zea mays* (Gupta and Hoo, 1991). Although YACs are becoming one of the preferred large capacity cloning systems for mapping entire genomes, they are emerging also as versatile tools useful in other aspects of genome analysis (Hieter *et al.,* 1990). The purpose of this review is to illustrate how YACs provide insight into the organization, content and function of complex genomes.

HOUSEKEEPING: YAC VECTORS, LIBRARY CONSTRUCTION, SCREENING AND CHARACTERIZATION

Vectors

The initial series of pYAC vectors constructed by Burke *et al.* (1987) vary at the restriction enzyme recognition sequence located at the cloning site. In summary, these vectors are

constructed on a pBR322 backbone and contain a cloning site within the *SUP4* gene (this provides colour selection in the presence of the *ade2-1* mutation), a left vector arm *(TEL, CEN4, ARS1, TRP1,* Amp^r, *ori),* a right vector arm *(TEL, URA3)* and a stuffer fragment *(HIS3)* (Figure 1).

Several modifications of the pYAC vector include the insertion of rare-cutting restriction enzyme recognition sequences into the cloning sites, T3 and T7 promoters flanking the cloning site and an *ori* sequence and neomycin resistance (Neo^r) gene into the left vector arm.

Besides the pYAC series, at least two other unique YAC vectors are available. The pJS97 and pJS98 vectors (pJS97 = *CEN* arm) contain multiple cloning sites flanked by T7 promoters. To permit plasmid-rescue of the terminal portions of the genomic insert

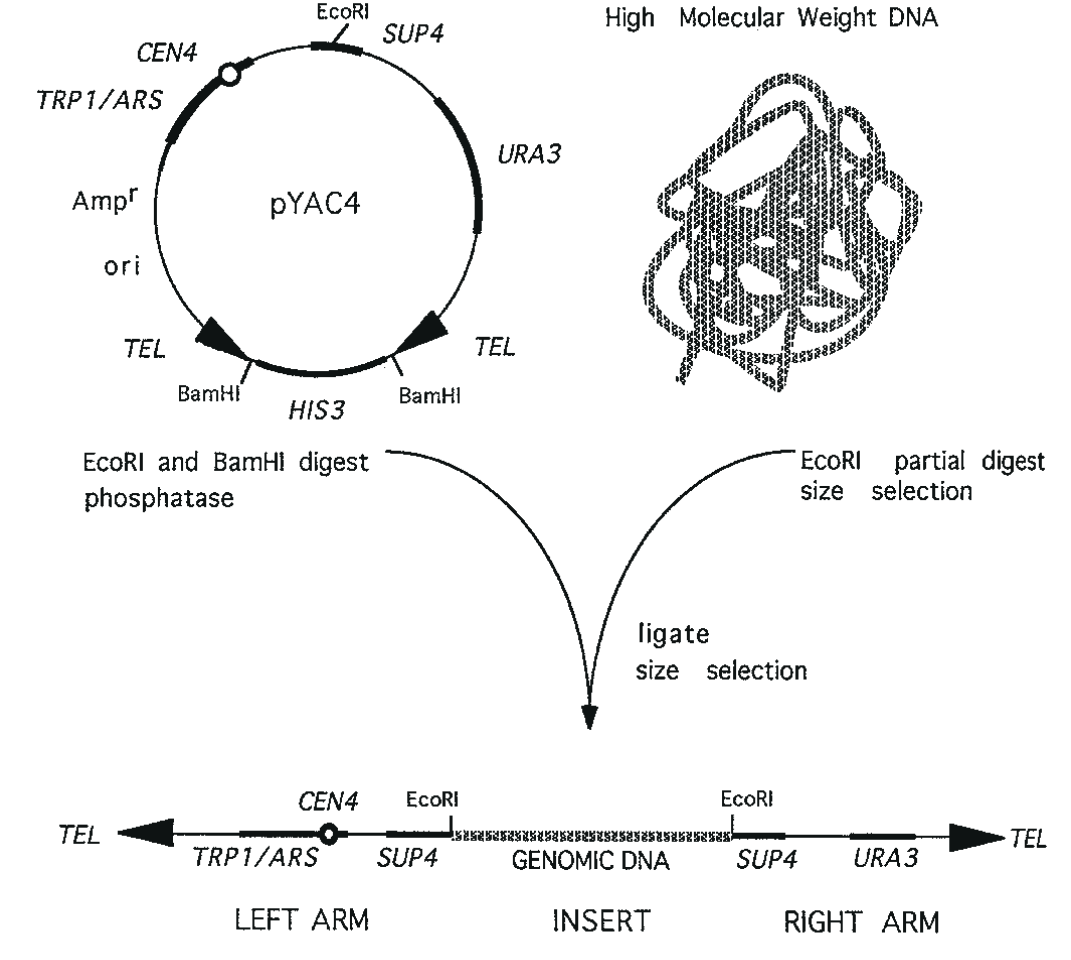

Figure 1. YAC vector and construction of a yeast artificial chromosome containing a genomic insert.

(McCormick *et al.*, 1990), each vector arm contains the elements that facilitate plasmid rescue in *Escherichia coli*—a second polylinker (for plasmid circularization), *ori* sequences and the Amp^r gene. The pCGS966 vector permits an increase in copy number by 10- to 20-fold (Smith *et al.*, 1990). The unique components of this vector include a heterologous thymidine kinase (TK) gene and a *GAL1* promoter. The *GAL1* promoter is adjacent to the *CEN4* sequences. Transcription through the *CEN4* sequence inactivates the centromere and allows for an increase in copy number. Maximal amplification is achieved by incubating yeast in media containing thymidine, galactose, sulfanilamide and methotrexate. The latter two agents inhibit *de novo* and salvage (but not TK) pathways of dTMP synthesis and provide the selection pressure necessary to increase copy number.

Library Construction

In the case of pYAC4 (Figure 1), the vector is prepared for ligation by double digestion with *Eco*RI (cloning site) and *Bam*HI (stuffer fragment) (Burke and Olson, 1991). To prevent re-ligation, the fragments are treated with alkaline phosphatase. The pYAC4 vector, in large excess, is ligated to size-selected genomic DNA that has been partially digested with *Eco*RI. *Saccharomyces cerevisiae* strain AB1380 *(Mat a, ψ+, ura3-52, trp1, ade2-1, can1-100, lys2-1, his5)* is transformed by the spheroplast method. Transformation efficiencies are in the order of 10^2–10^3/μg DNA.

Newly constructed libraries benefit most from performing all manipulations (e.g. digestion, ligation) of genomic DNA in low-melting agarose (McCormick *et al.*, 1990). Moreover, size selection for high-molecular-weight DNA by pulsed-field gel electrophoresis (PFGE), either before and/or after ligation, has resulted in a steady increase in mean insert size from ~200 to over 1000 kb (Chumakov *et al.*, 1992b).

Screening

Individual clones are identified by colony hybridization to gridded arrays (conforming to multimers of a 96-well microtitre plate) of YACs growing on nylon filters (Brownstein *et al.*, 1989). However, as the complexity of libraries increases, more efficient screening methods are desired. Green and Olson (1990b) demonstrated that YACs can be isolated using the polymerase chain reaction (PCR). PCR assays, using pools of DNA derived from different sets of 1920 clones, allow for identification of a minimal number of plates that contain the YAC of interest. Filters representing these plates are screened by colony hybridization. Alternatively, more complex pooling schemes permit the entire screening process to be performed using the PCR (Bentley *et al.*, 1992).

Bentley *et al.* (1992) describe a semi-automated method to create high density filters. Using a commercially available robotics work-station, 1536 clones are gridded to a single 80 × 120 mm filter. Several copies of a ~20,000 clone library are gridded in approximately three hours. In turn, yeast filters can be used for colony hybridization or as a source of DNA for PCR pooling schemes.

Characterization

The initial characterization of YACs determines whether yeast faithfully replicates foreign genomic DNA. Human plasminogen-activator inhibitor type 2 (PAI2) and factor IX probes detect the appropriate size restriction fragments in digests of YAC DNA (Brownstein *et al.,* 1989). Numerous reports, using other probes, confirm these observations (Little *et al.,* 1989). In part, this substantiates the fidelity of YAC cloning.

The physical mapping of individual clones is achieved by multiple approaches (Nelson, 1990). Internal restriction sites are mapped by indirect end-labelling techniques (Burke *et al.,* 1987). This is achieved by partial digestion and electrophoresis of YAC DNA, followed by hybridization with a vector-arm probe. Alternatively, the location of rare-cutting restriction sites are identified by hybridizing digested YAC DNA, separated by PFGE, to internal and vector-arm probes (Silverman *et al.,* 1989). Finally, YACs can be `fingerprinted' by either hybridizing or amplifying their DNA with repetitive sequences such as human *Alu* (Bellanné-Chantelot *et al.,* 1992; Nelson *et al.,* 1991) and L1 (Bellanné-Chantelot *et al.,* 1992).

GENOME ANALYSIS: ORGANIZATION

For the purposes of this review, genome organization is defined as the positional relationships between the individual components such as genes, introns, telomeres and centromeres which comprise a chromosome. Understanding these positional relationships will lead to the generation of complete physical maps of entire genomes.

Multiple strategies are used to generate physical maps of chromosomes. Mapping by `bottom-up' approaches rely on the assembly of an overlapping set of contiguous DNA clones. In part, the advantage of YACs over conventional cloning vectors lies in the smaller number of repetitive cloning steps necessary to build contigs spanning individual chromosomes (see below). Chromosomal mapping by `top-down' approaches rely on the ability to visualize the hybridization of cloned DNA sequences to their respective chromosomes. The hybridization of YAC DNA to human metaphase (Montanaro *et al.,* 1991) or *Drosophila* polytene (Garza *et al.,* 1989) preparations is useful in ordering probes within defined chromosomal regions. The mapping of human interphase chromosomes by fluorescence *in situ* hybridization (FISH) can be used to order probes separated by as little as 50 kb of DNA. However, the more commonly used metaphase preparations only provide accurate positional information when probes are separated by >1000 kb of DNA. Thus, an integrative mapping strategy that employs FISH or genetic methods to anchor candidate clones along the chromosomes and contig assembly techniques to close the gaps between these points may prove to be an efficient approach (Ewens *et al.,* 1991).

The ability to simultaneously clone and map large segments of the genome represents a major advantage of YAC cloning. Furthermore, YACs can be used to clone chromosomal domains such as amplicons (Schneider *et al.,* 1992), telomeres (Riethman *et al.,* 1989) and AT-rich regions (Triglia and Kemp, 1991), which are difficult to isolate in *E. coli*. However, relative to prokaryotic cloning systems, significant disadvantages still

exist. For instance: 1) library construction, maintenance and screening are more difficult; 2) rare clones can undergo rearrangement or deletion (Neil *et al.,* 1990); 3) copy number is low (1 copy/haploid genome); 4) purification schemes yield scant amounts of YAC DNA; 5) subcloning is required for more detailed analysis; and 6) co-transformants (yeast containing more than 1 YAC) and chimeras (a single YAC containing non-contiguous DNA segments) hinder contig assembly (Green and Olson, 1990a; Silverman *et al.,* 1991). In some libraries, 30–50% of the clones may be chimeras. Although the etiology is unclear, the co-ligation of DNA fragments and the co-transformation of YACs (with subsequent homologous recombination) are implicated in this process. Evidence for the latter event is derived from analysis of a chimeric YAC spanning a portion of the human CFTR gene (Green *et al.,* 1991). DNA sequencing of the chimeric junction reveals an *Alu* fragment dissimilar to that normally present in that portion of the CFTR gene. In this case, *in vivo* recombination between *Alu* fragments on two separate YACs appears to result in the chimeric clone.

Chromosomal Walking

The generation of contig maps in probe-poor regions can be achieved by directed chromosomal walks between flanking landmarks (Figure 2A). For example, two human genes, the BCL2 protooncogene and PAI2, are linked to a 1200 kb *Mlu*I fragment (Silverman *et al.,* 1991). No other known probes map to this fragment. First, YACs containing these genes are isolated from a human genomic library. Next, the genomic inserts adjacent to the cloning sites are isolated by an inverse PCR technique. In turn, these probes are used to re-screen the library for the next set of overlapping clones. After several rounds of end-probe isolation and library re-screening, the walks converge. Overall, 16 clones spanning ~2 Mb are assembled using 11 end-fragments. Similar walks are used to clone genomic regions spanning the Huntington's disease (Zuo *et al.,* 1992), type 1 neurofibromatosis (NF-1) (Marchuk *et al.,* 1992), HLA (Bronson *et al.,* 1991), familial adenomatous polyposis (Hampton *et al.,* 1992), myotonic dystrophy (Buxton *et al.,* 1992), Fragile X (Hirst *et al.,* 1991) and Wilm's tumour (Bonetta *et al.,* 1990) loci.

The success of chromosomal walking depends on the ability to isolate new probes from the ends of YAC inserts. Many techniques are available. PCR-based approaches include inverse (Silverman *et al.,* 1989), vectorette (Riley *et al.,* 1990) and vector-*Alu* PCR (Nelson *et al.,* 1989). These procedures generate small fragments (<2000 bp) which are used as probes or sequencing templates. The PCR methods are fast and efficient. However, success rates vary from 60–90% for any one procedure. Non-PCR-based approaches include the rescuing, in *E. coli,* of a plasmid containing the vector arm and adjacent insert (Burke *et al.,* 1987) and the screening of YAC sub-libraries with vector arm probes (Bronson *et al.,* 1991). Plasmid-rescue is a simple technique that yields fragments up to 20 kb in size. Initially, the inability to plasmid-rescue sequences from the *URA3* side of the pYAC vector limited the usefulness of this technique. However, homologous recombination can be used to retro-fit the *URA3* arm with the elements necessary for plasmid-rescue in *E. coli* (Hermanson *et al.,* 1991). Newer vectors, such

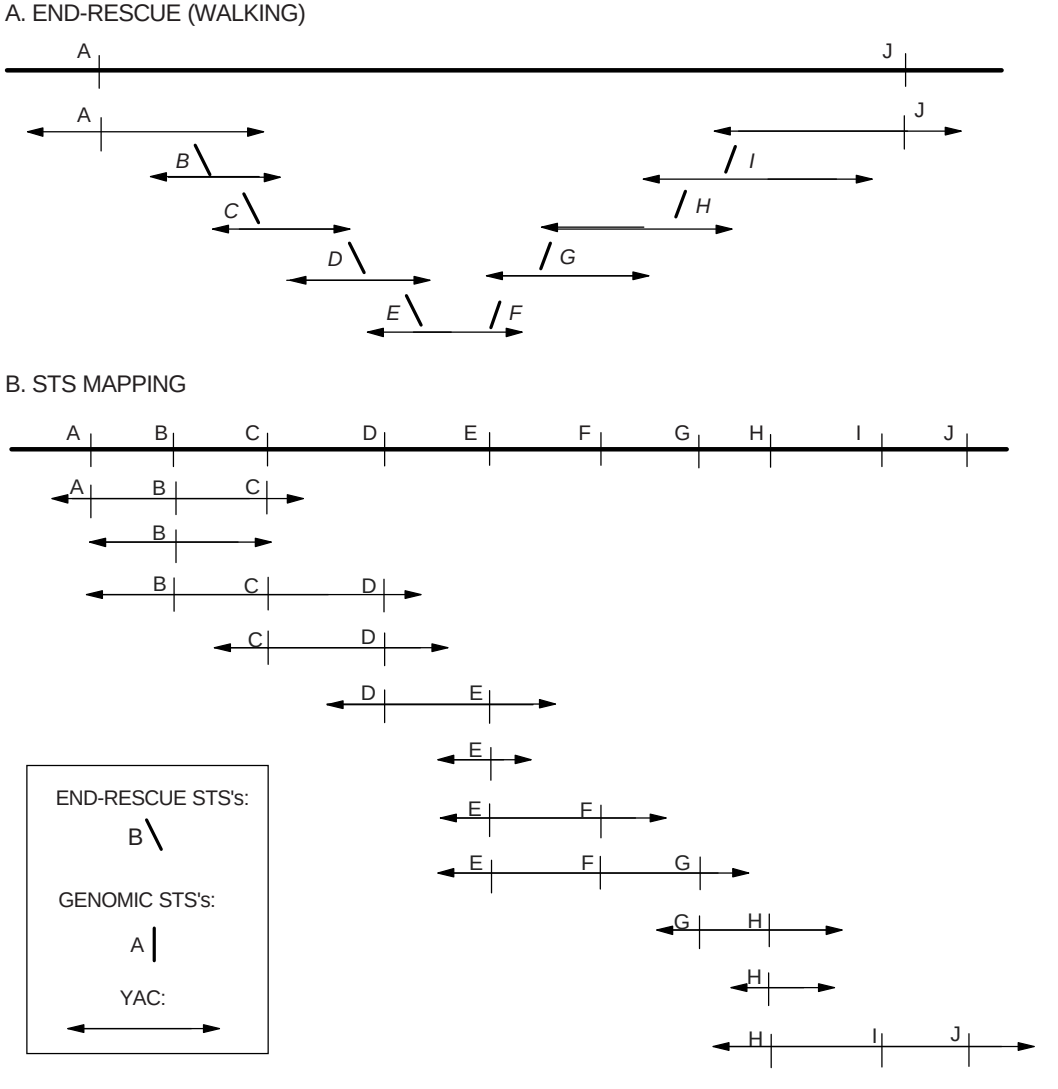

Figure 2. YAC contig construction. A: Chromosomal walking using end-rescue probes or STSs. Genomic markers (A,J) flanking an area of interest are used to isolate their cognate YACs. Probes isolated from the ends of these YACs *(B-I)* are used to rescreen the library. Repetitive rounds of walking will converge to yield a complete map. B: STS mapping yields a set of YAC clones in which order and overlap are determined by probe (A-J) content.

as pJS97/98 (Hieter *et al.*, 1990) are designed to allow plasmid-rescue from both sides. The screening of plasmid or λ phage YAC sub-libraries with vector arm probes is a reliable technique. However this method may be too labour-intensive and time-consuming for analysis of large numbers of clones.

Direct genomic DNA sequencing, using labelled vector-arm primers and a linear amplification scheme, may be the most efficient means to obtain end-fragment information (Coulson *et al.,* 1991). At this time, however, the lack of widespread use makes it difficult to assess the relative utility of this technique.

Mapping by an STS-based Approach

The paradigm for sequence tagged site (STS) based mapping was brought forth by Green and Olson (1990a). In part, the success of this approach is dependent upon a collection of closely spaced (~100–500 kb apart) STSs (Figure 2B). Each STS is a small, unique DNA sequence that is detectable by a PCR assay. YAC clones are aligned by comparing the STS content of each clone (Figure 2B). Using 16 STSs and 30 YACs, an ~1500 kb contig spanning the CFTR gene is assembled (Green and Olson, 1990a). To date, the most dramatic use of this technique is the construction of ~40 Mb size YAC contigs spanning both the euchromatic region of the human Y chromosome (Foote *et al.,* 1992) and the long arm of human chromosome 21 (Chumakov *et al.,* 1992a). For the Y chromosome, ~200 STSs are used to order ~200 YACs; whereas for chromosome 21, 191 discreet STSs are used to order ~800 YACs.

Mapping by DNA Fingerprinting

Different types of λ phage and cosmid fingerprinting methods are used to create partial or complete contig maps of the *C. elegans*, *S. cerevisiae*, and *E. Coli* genomes. The construction of YAC contigs by DNA fingerprinting poses unique problems due to the large insert size, the high percentage of chimerism in some libraries, genetic polymorphism, and the non-uniform distribution of repetitive elements. Considering these factors, Bellanné-Chantelot *et al.* (1992) have fingerprinted 22,000 YACs from a large insert (mean size = 810 kb) human genomic YAC library by hybridizing YAC DNA restriction fragments to either a LINE-1 (L-1) or an *Alu* repeat probe. Overlap between clones is assessed by a statistical model. On the basis of these calculations, the library contains 1175 to 1241 contigs of 4.3 to 5.7 clones, respectively. The grouping of contigs based on a statistical model is validated by FISH. Probes derived from ten random contigs map to specific regions on metaphase chromosomes. Using this library, the authors estimate that ~50% of the human genome is covered in total, with ~30% covered in contigs spanning at least 3 Mb.

GENOME ANALYSIS: CONTENT

The identification of new genes is a major goal of all genome projects. YACs are proving to be a useful reagent in these efforts by providing large segments of cloned substrate for further analysis. For example, clusters of unmethylated CpG motifs (CpG islands) frequently occur in the promoter region of housekeeping and some tissue-specific genes (Bird, 1986). Also, unmethylated CpG motifs are within the sequence recognized by many of the rare-cutting restriction enzymes. Since *S. cerevisiae* do not methylate this motif,

digestion of YAC DNA with rare-cutting restriction endonucleases reveals the presence of CpG clusters. Although not all clusters of rare-cutting restriction sites represent true CpG islands within their genome of origin, they provide clues as to the possible location of new genes. Unique DNA sequences isolated from these islands can be used to garner additional evidence as to the presence of a new gene (Elvin *et al.,* 1992). For instance, a DNA probe can be examined for conserved sequences by hybridizing to restriction fragments from different species (zoo blots) (Monaco *et al.,* 1986). Alternatively, selected fragments from the YAC can be `assayed' for gene function using open-reading frame vectors (Weinstock, 1987) or vectors which `trap' promoters, exons or pA signals (Hochgeschwender and Brennan, 1991).

Gel-purified YAC DNA can be labelled and used to screen cDNA libraries for the presence of new transcripts. This method relies on the pre-hybridization of the DNA probe with vector arm and genomic DNA (C_0t = 125–250) (Elvin *et al.,* 1990). This minimizes the signals due to hybridization of vector arms and repetitive sequence, respectively. Using this technique, human aldose reductase (Elvin *et al.,* 1990) and NF-1 (Wallace *et al.,* 1990) cDNAs were obtained. However, the sensitivity of this technique is variable. Indeed, the YAC probe identifies only 10% of aldose reductase cDNAs contained within the library (Elvin *et al.,* 1990).

Recombination-based techniques can be exploited to help identify new genes. One technique takes advantage of the high frequency of homologous recombination in *S. cerevisiae* (Reeves *et al.,* 1990). A cDNA library is constructed using a bacterial plasmid that contains a yeast selectable marker and a sequence which facilitates telomere formation. Yeasts are transformed with plasmids linearized at a site between the cDNA and telomere sequence. A homologous recombination event will insert the entire plasmid into the YAC, providing the YAC contains sequences homologous to the cDNA. This will yield a new yeast prototrophy and a new YAC fragmented (terminated) near the plasmid insertion site (Figure 3C).

New genes can be isolated from YACs by a direct, cDNA selection technique (Lovett *et al.,* 1991; Parimoo *et al.,* 1991). Purified YAC DNA is fixed to a solid support such as a nylon membrane. The DNA is hybridized with a cDNA mixture amplified from a cDNA library. This hybridization is performed after blocking either the DNA filter or the cDNA probe with a complex series of DNAs (e.g. repetitive sequence or total genomic DNA, yeast DNA, rDNA, poly(dI)-poly(dC), and/or plasmid DNA). After removing the unbound cDNAs, the bound sequences are eluted, re-amplified, and subcloned into λ phage vectors. Analysis of the `selected' library reveals enrichment of rare sequences ($1/10^5$) by ~1000-fold (Lovett *et al.,* 1991; Parimoo *et al.,* 1991). In a modification of this method (Morgan *et al.,* 1992), cDNAs are amplified and blocked with repetitive DNA. Next, the cDNAs are hybridized to biotinylated, purified YAC DNA. Finally, the bound species are purified by binding of the duplex to streptavidin coated beads. The bound cDNAs are eluted, re-amplified and subjected to several additional rounds of this purification scheme. An enrichment of 100,000-fold is achieved (Morgan *et al.,* 1992).

Molecular cloning by genetic complementation is a powerful method to isolate new genes in prokaryotes and lower eukaryotes. The ability to transfer large YACs, intact, into mammalian cells (Huxley and Gnirke, 1991) enhances the possibility of using this technique to discover new genes in higher eukaryotes (see below).

Studies examining the genetic elements and factors which control, for example, 1) RNA transcription, processing and transport; 2) nuclear organization; 3) DNA replication and repair; 4) homologous and non-homologous recombination; 5) functional chromosome domains such as telomeres, centromeres, autonomous replicating sequences and

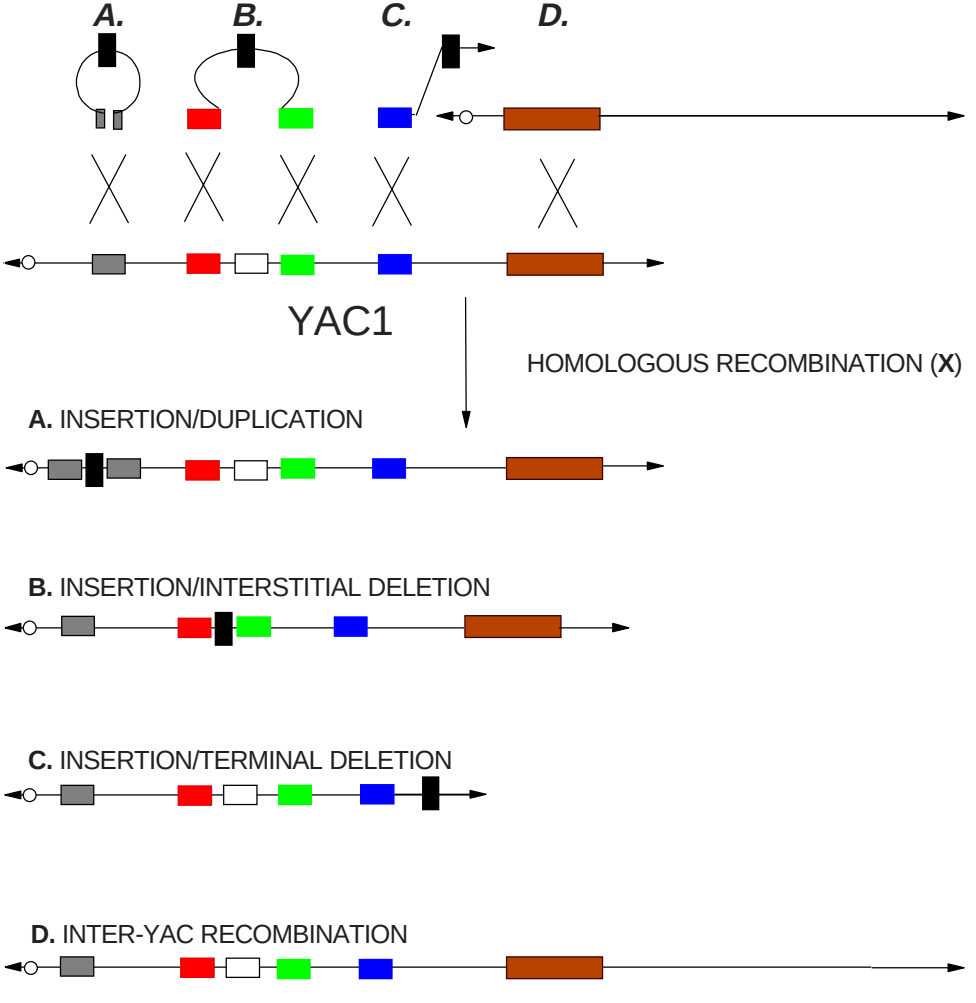

Figure 3. Modification of YACs by homologous recombination. *Upper panel.* YAC1 is modified by a homologous recombination event between a yeast integrating plasmid *(A-C)* or another YAC *(D)*. This yields: *(A)* insertion of a selectable marker or other gene (black rectangle) and duplication of the targeting sequence, *(B)* insertion of a selectable marker and interstitial deletion of a segment (white rectangle), *(C)* terminal deletion (fragmentation) with addition of a selectable marker and a new telomere and *(D)* a recombinant YAC.

amplicons; 6) higher-order chromatin structure; 7) transposable elements; and 8) spindle formation and other mitotic events, would benefit from the ability to manipulate and transfer sub-chromosomal DNA fragments into the appropriate hosts. The large cloning capacity of YACs makes them the ideal vectors for these purposes. Furthermore, the high recombination frequency in yeast (~3 kb/cM) provides a convenient and reliable method to create insertions, deletions and point mutations, *in vivo* (Hieter *et al.,* 1990; Reeves *et al.,* 1990; Rothstein, 1991).

YAC Modification by Homologous Recombination

Manipulations in yeast are achieved by transformation with yeast integration plasmids (YIPs) (Figure 3). In the absence of an autonomously replicating sequence, YIPs transform yeast almost exclusively via homologous recombination (Rothstein, 1991). YIPs must contain a targeting sequence (a segment of DNA homologous to a portion of a YAC) and selectable markers. By linearizing the YIP prior to transformation, the exposed ends of the targeting sequence help direct recombination to the homologous site within the YAC by at least 1000-fold (Hieter *et al.,* 1990). YIPs can be constructed to insert point mutations, disruptions, selectable markers, regulatory sequences, and/or other functional domains (Figure 3A). YIPs can be configured also to generate defined interstitial (Figure 3B) and terminal (Figure 3C) deletions.

Homologous recombination can be used to re-construct extensive genomic regions containing large genes (Figure 3D) such as dystrophin (Den Dunnen *et al.,* 1992), CFTR (Green and Olson, 1990a), NF-1 (Marchuk *et al.,* 1992) and BCL2 (Silverman *et al.,* 1990). In the latter case, two YACs spanning portions of the ~230 kb BCL2 gene are introduced into the same cell by mating the yeast. The resultant diploids are induced to undergo meiosis and sporulation. The products of meiosis are analyzed after tetrad dissection. Meiotic recombination within the 60 kb region of overlap yields a single 800 kb clone containing the entire BCL2 gene.

Homologous recombination between YACs can also occur during mitotic cross-over (Ragoussis *et al.,* 1992). Overlapping YACs are introduced into the same cell by protoplast fusion. This creates an isosexual diploid cell that can recombine its YACs after ultraviolet irradiation and mitosis.

Transfer of YACs from *Saccharomyces cerevisiae* to a Mammalian Host

The introduction of YACs into mammalian cells is achieved by protoplast fusion (Figure 4). Anonymous YACs, 360 kb (Pavan *et al.,* 1990) and 450 kb (Pachnis *et al.,* 1990) in size, are retrofitted with the aminoglycoside phosphotransferase (Neo^r) gene (this gene confers resistance to the neomycin analog, G418, and serves as a positive selection marker for many transfected cells). Yeast spheroplasts are prepared and fused to mammalian cells in the presence of polyethylene glycol (Figure 4). The fusion products are cultured in conventional tissue culture media containing toxic levels of G418. After three to four weeks, G418-resistant clones are expanded and analyzed. Among 25%

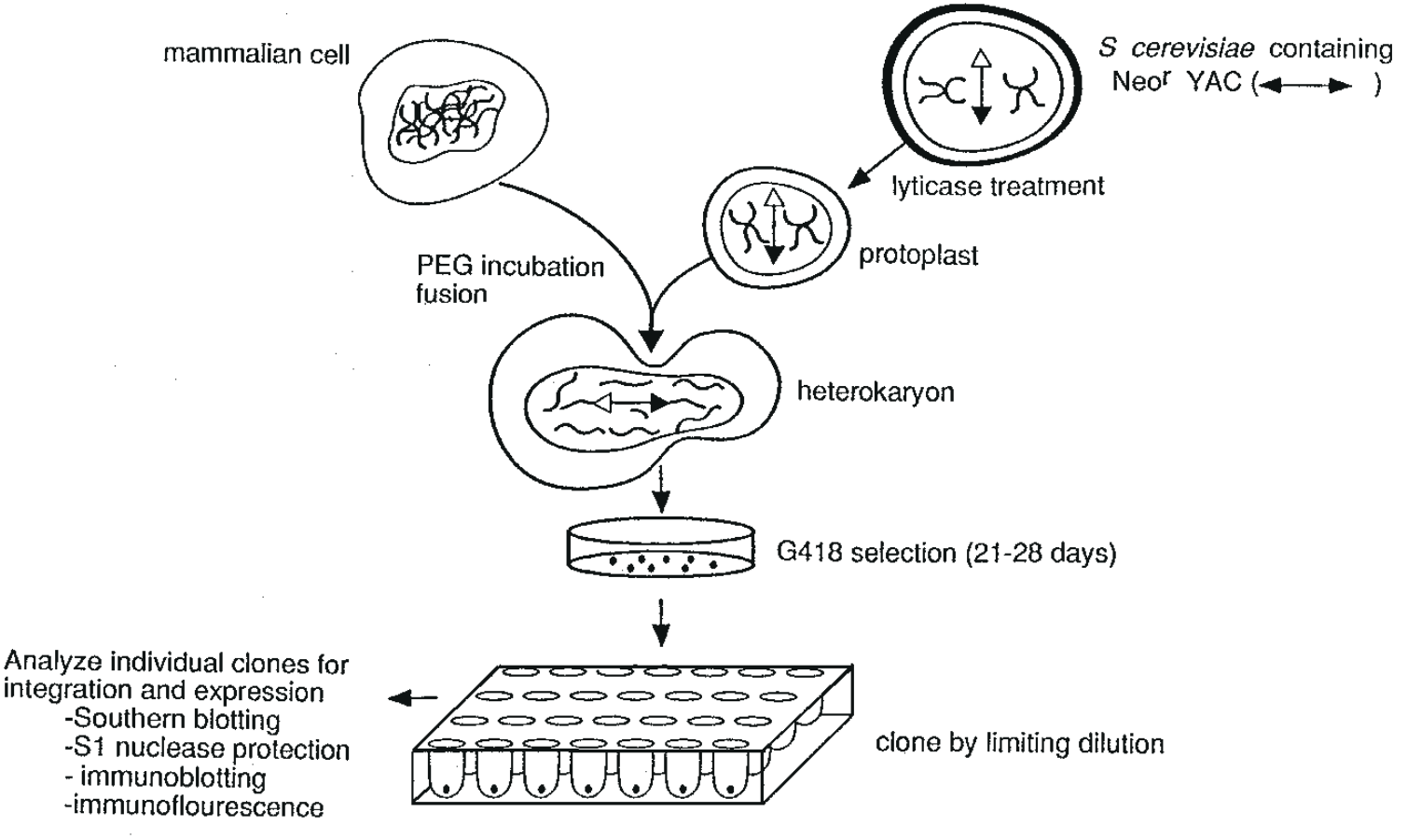

Figure 4. Transfecting retro-fitted YACs into mammalian cells by protoplast fusion. A yeast protoplast containing a YAC (modified by introducing a Neo^r gene) is fused to a mammalian cell in the presence of polyethylene glycol. Fusion products are selected in the presence of G418. Resistant clones arise after three or four weeks. After cloning by limiting dilution (if necessary), cells can be analyzed for integration and gene expression.

(Pavan *et al.*, 1990) to 50% (Pachnis *et al.*, 1990) of the G418-resistant clones contain an intact YAC. In these studies, copy number is not assessed. However, studies using both a 360 kb and an 800 kb YAC show that most clones contain only one or two copies. In other transfections using protoplast fusion methods, genes encoded on the YAC, such as HPRT (Huxley *et al.*, 1991) and GART (Gnirke *et al.*, 1991), can complement defects in mutant rodent cell lines. These findings support attempts to define new genes within YACs by complementing defects in well-defined mutants. Many of the G418-resistant clones obtained by protoplast fusion contain variable amounts of yeast DNA (Gnirke *et al.*, 1991; Huxley *et al.*, 1991). In addition, transformation efficiencies are target cell-dependent with some lines yielding 1–50 G418-resistant clones per 10^6 cells (Huxley and Gnirke, 1991).

Other methods are used to introduce YACs into mammalian cells. Calcium phosphate/precipitation of a 40 kb YAC yields transient (D'Urso *et al.*, 1990) and stable (Elicieri *et al.*, 1991) clones expressing the G6PD gene. Electroporation of high-molecular-weight, total yeast DNA, which contains a 310 kb YAC, leads to the transient expression of the leukocyte common antigen (LCA, CD45) in approximately 5% of the cells (Fernandez-Luna *et al.*, 1991). Although the gene spans ~120 kb, the size of the fragment that enters the cells is not determined. Mouse fibroblasts, defective for both α(1)I collagen alleles, are transfected, via lipofection, with a gel-purified 150 kb YAC containing a mouse collagen gene *(Col1a1)* (Strauss and Jaenisch, 1992). However, only 3/18 G418-resistant clones express the collagen gene, and only 2/18 contain the entire YAC. Thus, the low percentage of collagen expressing clones may be due to incomplete incorporation of the gene into the host genome or a positional effect. Interestingly, only ~50% of G418-resistant clones also containing a stable BCL2 integrant express detectable levels of BCL2 mRNA and protein. This suggests that YACs as large as 800 kb are still subject to the local regulatory factors and chromatin structure present at the site of integration. Finally, gel purified 35 kb (Schedl *et al.*, 1992) and 100 kb (Gnirke and Huxley, 1991) YAC DNA fragments can be microinjected into mouse pronuclei or mutant rodent cell lines, respectively.

The transformation efficiency of microinjected YACs is approximately five-fold lower than that observed with plasmid vectors (Gnirke and Huxley, 1991). The utility of microinjection using larger YACs awaits purification schemes that yield pieces of DNA larger than 100 kb (Gnirke and Huxley, 1991). These results have obvious implications for studying large genomic fragments in transgenic models (Schedl *et al.*, 1992). In summary, YACs can be used to transfer large genomic segments into mammalian hosts. At the present time, protoplast fusion appears to be the best method to transfer the largest YACs intact. However, as newer schemes yield higher concentrations of purified intact YAC DNA, other transfection strategies may prove to be more efficient.

REFERENCES

BELLANNÉ-CHANTELOT, C., LACROIX, B., OUGEN, P., BILLAULT, A., BEAUFILS, S., BERTRAND, S., GEORGES, I., GILBERT, F., GROS, I., LUCOTTE, G., SUSINI, L., CODANI, J.-J.,

GESNOUIN, P., POOK, S., VAYSSEIX, G., LU-KUO, J., RIED, T., WARD, D., CHUMAKOV, I., Le PASLIER, D., BARILLOT, E. and COHEN, D. 1992. Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* 70: 1059–1068.

BENTLEY, D.R., TODD, C., COLLINS, J., HOLLAND, J., DUNHAM, I., HASSOCK, S., BANKIER, A. and GIANNELLI, F. 1992. The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. *Genomics* 12: 534–541.

BIRD, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature (London)* 321: 209–213.

BONETTA, L., KUEHN, S.E., HUANG, A., LAW, D.J., KALIKIN, L.M., KOI, M., REEVE, A.E., BROWNSTEIN, B.H., YEGER, H., WILLIAMS, B.R.G. and FEINBERG, A.P. 1990. Wilms tumor locus on 11p13 defined by multiple CpG island-associated transcripts. *Science* 250: 994–997.

BRONSON, S.K., PEI, J., TAILLON-MILLER, P., CHORNEY, M.J., GERAGHTY, D.E. and CHAPLIN, D.D. 1991. Isolation and characterization of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. *Proceedings of the National Academy of Sciences of the USA* 88: 1676–1680.

BROWNSTEIN, B.H., SILVERMAN, G.A., LITTLE, R.D., BURKE, D.T., KORSMEYER, S.J., SCHLESSINGER, D. and OLSON, M.V. 1989. Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* 244: 1348–1351.

BURKE, D.T. and OLSON, M.V. 1991. Preparation of clone libraries in yeast artificial chromosome vectors. In Guthrie, C. and Fink, G.R., eds. *Methods in Enzymology (Vol. 194): Guide to Yeast Genetics and Molecular Biology.* San Diego: Academic Press, Inc., pp. 251–270.

BURKE, D.T., CARLE, G.F. and OLSON, M.V. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.

BURKE, D.T., ROSSI, J.M., LEUNG, J., KOOS, D.S. and TILGHMAN, S.M. 1991. A mouse genomic library of yeast artificial chromosome clones. *Mammalian Genome* 1: 65.

BUXTON, J.P., SHELBOURNE, P., DAVIES, J., JONES, C., PERRYMAN, M.B., ASHIZAWA, T., BUTLER, R., BROOK, D., SHAW, D., de JONG, P., MARKHAM, A., WILLIAMSON, R. and JOHNSON, K. 1992. Characterization of a YAC and cosmid contig containing markers tightly linked to the myotonic dystrophy locus on chromosome 19. *Genomics* 13: 526–531.

CHUMAKOV, I., RIGAULT, P., GUILLOU, S., OUGEN, P., BILLAULT, A., GUASCONI, G., GERVY, P., LEGALL, I., SOULARUE, P., GRINAS, L., BOUGUELERET, L., BELLANNÉ-CHANTELOT, C., LACROIX, B., BARILLOT, E., GESNOUIN, P., POOK, S., VAYSSEIX, G., FRELAT, G., SCHMITZ, A., SAMBUCY, J.-L., BOSCH, A., ESTIVILL, X., WEISSENBACH, J., VIGNAL, A., RIETHMAN, H., COX, D., PATTERSON, D., GARDINER, K., HATTORI, M., SAKAKI, Y., ICHIKAWA, H., OHKI, M., Le PASLIER, D., HEILIG, R., ANTONARAKIS, S. and COHEN, D. 1992a. Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* 359: 380–386.

CHUMAKOV, I.M., Le GALL, I., BILLAULT, A., OUGEN, P., SOULARUE, P., GUILLOU, S., RIGAULT, P., BUI, H., de TAND, M.-F., BARILLOT, E., ABDERRAHIM, H., CHERIF, D., BERGER, R., Le PASLIER, D. and COHEN, D. 1992b. Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genetics* 1: 222–225.

COULSON, A., KOZONO, Y., SHOWNKEEN, R. and WATERSTON, R. 1991. The isolation of insert-terminal YAC fragments by genomic sequencing. *Technique* 3: 17–23.

COULSON, A., WATERSTON, R., KIFF, J., SULSTON, J. and KOHARA, Y. 1988. Genome linking with yeast artificial chromosomes. *Nature* 335: 184–186.

D'URSO, M., ZUCCHI, I., CICCODICOLA, A., PALMIERI, G., ABIDI, F.E. and SCHLESSINGER, D. 1990. Human glucose-6-phosphate dehydrogenase gene carried on a yeast artificial chromosome encodes active enzyme in monkey cells. *Genomics* 7: 531–534.

Den DUNNEN, J.T., GROOTSCHOLTEN, P.M., DAUWERSE, J.G., WALKER, A.P., MONACO, A.P., BUTLER, R., ANAND, R., COFFEY, A.J., BENTLEY, D.R., STEENSMA, H.Y. and Van OMMEN, G.J.B. 1992. Reconstruction of the 2.4 Mb human DMD-gene by homologous YAC recombination. *Human Molecular Genetics* 1: 19–28.

ELICIERI, B., LABELLA, T., HAGINO, Y., SRIVASTAVA, A., SCHLESSINGER, D., PILIA, G., PALMIERI, G. and D'URSO, M. 1991. Stable integration and expression in mouse cells of yeast artificial chromosomes harboring human genes. *Proceedings of the National Academy of Sciences of the USA* 88: 2179–2183.

ELVIN, P., BUTLER, R. and HEDGE, P.J. 1992. Transcribed sequences within YACs: HTF island cloning and cDNA library screening. In: Anand, R., ed. *Techniques for the Analysis of Complex Genomes*. London: Academic Press Limited, pp. 155–171.

ELVIN, P., SLYNN, G., BLACK, D., GRAHAM, A., BUTLER, R., RILEY, J., ANAND, R. and MARKHAM, A.F. 1990. Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Research* 18: 3913–3917.

EWENS, W.J., BELL, C.J., DONNELLY, P.J., DUNN, P., MATALLANA, E. and ECKER, J.R. 1991. Genome mapping with anchored clones: theoretical aspects. *Genomics* 11: 799–805.

FERNANDEZ-LUNA, J.L., MATTHEWS, R.J., BROWNSTEIN, B.H., SCHREIBER, R.D. and THOMAS, M.L. 1991. Characterization and expression of the human leukocyte-common antigen (CD45) gene contained in yeast artificial chromosomes. *Genomics* 10: 756–764.

FOOTE, S., VOLLRATH, D., HILTON, A. and PAGE, D.C. 1992. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* 258: 62–66.

GARZA, D., AJIOKA, J.W., BURKE, D.T. and HARTL, D.L. 1989. Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* 246: 641–646.

GNIRKE, A. and HUXLEY, C. 1991. Transfer of the human HPRT and GART genes from yeast to mammalian cells by microinjection of YAC DNA. *Somatic Cell and Molecular Genetics* 17: 573–580.

GNIRKE, A., BARNES, T.S., PATTERSON, D., SCHILD, D., FEATHERSTONE, T. and OLSON, M.V. 1991. Cloning and *in vivo* expression of the human GART gene using yeast artificial chromosomes. *EMBO Journal* 10: 1629–1634.

GREEN, E.D. and OLSON, M.V. 1990a. Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* 250: 94–98.

GREEN, E.D. and OLSON, M.V. 1990b. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the USA* 87: 1213–1217.

GREEN, E.D., RIETHMAN, H.C., DUTCHIK, J.E. and OLSON, M.V. 1991. Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* 11: 658–669.

GUPTA, M. and HOO, B. 1991. Entire maize chloroplast genome is stably maintained in a yeast artificial chromosome. *Plant Molecular Biology* 17: 361–369.

GUZMAN, P. and ECKER, J.R. 1988. Development of large DNA methods for plants: molecular cloning of large segments of *Arabidopsis* and carrot DNA into yeast. *Nucleic Acids Research* 16: 11091–11105.

HAMPTON, G.M., WARD, J.R.T.J., COTTRELL, S., HOWE, K., THOMAS, H.J.W., BALLHAUSEN, W.G., JONES, T., SHEER, D., SOLOMON, E., FRISCHAUF, A.-M. and BODMER, W.F. 1992. Yeast artificial chromosomes for the molecular analysis of the familial polyposis *APC* gene region. *Proceedings of the National Academy of Sciences of the USA* 89: 8249–8253.

HERMANSON, G.G., HOEKSTRA, M.F., McELLIGOTT, D.L. and EVANS, G.A. 1991. Rescue of end fragments of yeast artificial chromosomes by homologous recombination in yeast. *Nucleic Acids Research* 19: 4943–4948.

HIETER, P., CONNELLY, C., SHERO, J., McCORMICK, M.K., ANTONARAKIS, S., PAVAN, W. and REEVES, R. 1990. Yeast artificial chromosomes: promises kept and pending. In: Davies, K.E. and Tilghman, S.M., eds. *Genome Analysis Volume 1: Genetic and Physical Mapping*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 83–120.

HIRST, M.C., RACK, K., NAKAHORI, Y., ROCHE, A., BELL, M.V., FLYNN, G., CHRISTADOULOU, Z., MacKINNON, R.N., FRANCIS, M., LITTLER, A.J., ANAND, R., POUSTKA, A.-M., LEHRACH, H., SCHLESSINGER, D., D'URSO, M., BUCKLE, V.J. and DAVIES, K.E. 1991. A YAC contig across the fragile X site defines the region of fragility. *Nucleic Acids Research* 19: 3283–3288.

HOCHGESCHWENDER, U. and BRENNAN, M.B. 1991. Identifying genes within the genome: new ways for finding the needle in a haystack. *BioEssays* 13: 139–144.

HUXLEY, C. and GNIRKE, A. 1991. Transfer of yeast artificial chromosomes from yeast to mammalian cells. *BioEssays* 13: 545–550.

HUXLEY, C., HAGINO, Y., SCHLESSINGER, D. and OLSON, M.V. 1991. The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion. *Genomics* 9: 742–750.

KUSPA, A., MAGHAKIAN, D., BERGESCH, P. and LOOMIS, W.F. 1992. Physical mapping of genes to specific chromosomes in *Dictyostelium discoideum*. *Genomics* 13: 49–61.

KUSPA, A., VOLLRATH, D., CHENG, Y. and KAISER, D. 1989. Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. *Proceedings of the National Academy of Sciences of the USA* 86: 8917–8921.

LITTLE, R.D., PORTA, G., CARLE, G.F., SCHLESSINGER, D. and D'URSO, M. 1989. Yeast artificial chromosomes with 200- to 800-kilobase inserts of human DNA containing HLA, Vk, 5S, and Xq24-Xq28 sequences. *Proceedings of the National Academy of Sciences of the USA* 86: 1598–1602.

LOVETT, M., KERE, J. and HINTON, L.M. 1991. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proceedings of the National Academy of Sciences of the USA* 88: 9628–9632.

MAIER, E., HOHEISEL, J.D., McCARTHY, L., MOTT, R., GRIGORIEV, A.V., MONACO, A.P., LARIN, Z. and LEHRACH, H. 1992. Complete coverage of the *Schizosaccharomyces pombe* genome in yeast artificial chromosomes. *Nature Genetics* 1: 273–277.

MARCHUK, D.A., TAVAKKOL, R., WALLACE, M.R., BROWNSTEIN, B.H., TAILLON-MILLER, P., FONG, C.-T., LEGIUS, E., ANDERSEN, L.B., GLOVER, T.W. and COLLINS, F.S. 1992. A yeast artificial chromosome contig encompassing the type I neurofibromatosis gene. *Genomics* 13: 672–680.

MARTIN, G.B., GANAL, M.W. and TANKSLEY, S.D. 1992. Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. *Molecular Genes and Genetics* 233: 25–32.

McCORMICK, M.K., SHERO, J.H., CONNELLY, C.J., ANTONARAKIS, S.E. and HIETER, P. 1990. Methods for cloning large DNA segments as artificial chromosomes in *S. cerevisiae*. *Technique* 2: 65–71.

MONACO, A.P., NEVE, R.L., COLLETTI-FEENER, C., BERTELSON, C.J., KURNIT, D.M. and KUNKEL, L.M. 1986. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* 323: 646–650.

MONTANARO, V., CASAMASSIMI, A., D'URSO, M., YOON, J.-Y., FREIJE, W., SCHLESSINGER, D., MUENKE, M., NUSSBAUM, R.L., SACCONE, S., MAUGERI, S., SANTORO, A.M., MOTTA, S. and DELLA VALLE, G. 1991. *In situ* hybridization to cytogenetic bands of yeast artificial chromosomes covering 50% of human Xq24-Xq28 DNA. *American Journal of Human Genetics* 48: 183–194.

MORGAN, J.G., DOLGANOV, G.M., ROBBINS, S.E., HINTON, L.M. and LOVETT, M. 1992. The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Research* 20: 5173–5179.

NEIL, D.L., VILLASANTE, A. and FISHER, R.B. 1990. Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Research* 18: 1421–1428.

NELSON, D.A., BALLABIO, A., VICTORIA, M.F., PIERETTI, M., BIES, R.D., GIBBS, R.A., MALEY, J.A., CHINAULT, A.C., WEBSTER, T.D. and CASKEY, C.T. 1991. *Alu*-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. *Proceedings of the National Academy of Sciences of the USA* 88: 6157–6161.

NELSON, D.L. 1990. Current methods for YAC clone characterization. *Genetic Analysis, Techniques and Applications* 7: 100–106.

NELSON, D.L., LEDBETTER, S.A., CORBO, L., VICTORIA, M.F., RAMIREZ-SOLIS, R., WEBSTER, T.D., LEDBETTER, D.H. and CASKEY, C.T. 1989. *Alu* polymerase chain reaction: a method

for rapid isolation of human-specific sequences from complex DNA sources. *Proceedings of the National Academy of Sciences of the USA* 86: 6686–6690.

PACHNIS, V., PEVNY, L., ROTHSTEIN, R. and COSTANTINI, F. 1990. Transfer of a yeast artificial chromosome carrying human DNA from *Saccharomyces cerevisiae* into mammalian cells. *Proceedings of the National Academy of Sciences of the USA* 87: 5109–5113.

PARIMOO, S., PATANJALI, S.R., SHUKLA, H., CHAPLIN, D.D. and WEISSMAN, S.M. 1991. cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proceedings of the National Academy of Sciences of the USA* 88: 9623–9627.

PAVAN, W.J., HIETER, P. and REEVES, R.H. 1990. Modification and transfer into an embryonal carcinoma cell line of a 360-kilobase human-derived yeast artificial chromosome. *Molecular and Cellular Biology* 10: 4163–4169.

RAGOUSSIS, J., TROWSDALE, J. and MARKIE, D. 1992. Mitotic recombination of yeast artificial chromosomes. *Nucleic Acids Research* 20: 3135–3138.

REEVES, R.H., PAVAN, W.J. and HIETER, P. 1990. Modification and manipulation of mammalian DNA cloned as YACs. *Genetic Analysis, Techniques and Applications* 7: 107–113.

RIETHMAN, H.C., MOYZIS, R.K., MEYNE, J., BURKE, D.T. and OLSON, M.V. 1989. Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial chromosome vector. *Proceedings of the National Academy of Sciences of the USA* 86: 6240–6244.

RILEY, J., BUTLER, R., OGILVIE, D., FINNIEAR, R., JENNER, D., POWELL, S., ANAND, R., SMITH, J.C. and MARKHAM, A.F. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Research* 18: 2887–2890.

ROTHSTEIN, R. 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods in Enzymology* 194: 281–301.

SCHEDL, A., BEERMANN, F., THIES, E., MONTOLIU, L., KELSEY, G. and SCHUTZ, G. 1992. Transgenic mice generated by pronuclear injection of a yeast artificial chromosome. *Nucleic Acids Research* 20: 3073–3077.

SCHNEIDER, S.S., HIEMSTRA, J.L., ZEHNBAUER, B.A., TAILLON-MILLER, P., Le PASLIER, D., VOGELSTEIN, B. and BRODEUR, G.M. 1992. Isolation and structural analysis of a 1.2 megabase *N-myc* amplicon from a human neuroblastoma. *Molecular and Cellular Biology* 12: 5563–5570.

SILVERMAN, G.A., GREEN, E.D., YOUNG, R.L., JOCKEL, J.I., DOMER, P.H. and KORSMEYER, S.J. 1990. Meiotic recombination between yeast artificial chromosomes yields a single clone containing the entire BCL2 protooncogene. *Proceedings of the National Academy of Sciences of the USA* 87: 9913–9917.

SILVERMAN, G.A., JOCKEL, J.I., DOMER, P.H., MOHR, R.M., TAILLON-MILLER, P. and KORSMEYER, S.J. 1991. Yeast artificial chromosome cloning of a two-megabase-size contig within chromosomal band 18q21 establishes physical linkage between BCL2 and plasminogen activator inhibitor type-2. *Genomics* 9: 219–228.

SILVERMAN, G.A., YE, R.D., POLLACK, K.M., SADLER, J.E. and KORSMEYER, S.J. 1989. Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proceedings of the National Academy of Sciences of the USA* 86: 7485–7489.

SMITH, D.R., SMYTH, A.P. and MOIR, D.T. 1990. Amplification of large artificial chromosomes. *Proceedings of the National Academy of Sciences of the USA* 87: 8242–8246.

STRAUSS, W.M. and JAENISCH, R. 1992. Molecular complementation of a collagen mutation in mammalian cells using yeast artificial chromosomes. *EMBO Journal* 11: 417–422.

TRIGLIA, T. and KEMP, D.J. 1991. Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Molecular Biochemistry of Parasitology* 44: 207–212.

WALLACE, M.R., MARCHUK, D.A., ANDERSEN, L.B., LETCHER, R., ODEH, H.M., SAULINO, A.M., FOUNTAIN, J.W., BRERETON, A., NICHOLSON, J., MITCHELL, A.L., BROWNSTEIN, B.H. and COLLINS, F.S. 1990. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* 249: 181–186.

WEINSTOCK, G.M. 1987. Use of open reading frame expression vectors. *Methods in Enzymology* 154: 156–163.

ZUO, J., ROBBINS, C., TAILLON-MILLER, P., COX, D.R. and MYERS, R.M. 1992. Cloning of the Huntington disease region in yeast artificial chromosomes. *Human Molecular Genetics* 1: 149–159.

Yeast artificial chromosome cloning and contig construction in protozoan parasites

J.W. Ajioka, J. Espinoza and J.T. Swindle

Setting up a laboratory for genome analysis must necessarily be considered as an integrated system. Starting from arrayed whole genomic libraries and partially normalized cDNA libraries, the construction of ordered chromosome-specific clone sets (essentially contig construction), to data analysis and storage in a modified ACeDB computer program should be organized as an integrated system. The genomes of the protozoan parasites *Toxoplasma, Leishmania* and *Trypanosoma* have unique characteristics which make certain cloning and contig construction strategies more or less efficient than others. For this purpose, a combined expression-tagged sites (ETS)/double-end clone-limited strategy for contig construction is the most efficient as it will also include biological information. Finally, the yeast artificial chromosome (YAC) cloning system faithfully and stably replicates *Trypanosoma cruzi* DNA as demonstrated by the analysis of the calmodulin-ubiquitin locus.

INTRODUCTION

Virtually all aspects of genome analysis, from gene expression to physical mapping of chromosomes, will benefit by employing an integrated systematic approach. All current genome projects, including the prototypical *Caenorhabditis elegans* project, are set up as a system where genetic and physical information are processed and compiled into a central computerized ACeDB (AC. elegans Data Base) style database (Sulston *et al.*, 1992). However, the `flow through' of information in terms of efficiency and accuracy is dependent upon the specific methods by which data is generated. Each genome will present different advantages and problems. The genomes of protozoan parasites have the advantage of being small and containing less dispersed repetitive sequences than higher eukaryotes such as humans or even the fruit fly *Drosophila* (Ajioka *et al.*, 1991). On the other hand, some protozoan parasites, such as the trypanosomatids, carry redundant genes as tandem arrays. These properties are highly relevant to two aspects of genome analysis, the generation of whole genomic libraries and construction of chromosomal contigs (continuous stretches of DNA). Finally, genetic information such as ETS (Venter *et al.*, 1992) and restriction fragment length polymorphisms (RFLPs) can also serve as tools for the construction of the contigs, thus automatically forming a genome map (Ajioka *et al.*, 1991).

The haploid genome sizes of *Toxoplasma, Trypanosoma* and *Leishmania* are approximately all about 3–5 $\times 10^7$ bp (see Blackwell *et al.*, this volume). This suggests that certain cloning vehicles such as cosmids, P1 (see Shepherd *et al.* and Melville *et al.*, this volume) and YACs (see G. Silverman *et al.*, this volume) will be appropriate as they all carry relatively large fragments of DNA. These vehicles make the storage, manipulation and alignment of clones a fairly simple exercise as the number of clones involved is very low. For trypanosomatids, because of tandemly arrayed genes, the YAC system may be the

`workhorse' for genomic cloning as it appears to faithfully and stably replicate these kinds of DNA structures. In contrast, data from genes that have been cloned suggest that, at a molecular level, *Toxoplasma* is a relatively conventional eukaryote. Coding sequences are separated by *cis* introns and there is little base composition or codon bias. However, given the current limitations of the genetic mapping to 1 Mb (5 cM) (Sibley *et al.*, 1992), a genomic analysis using physical mapping techniques is crucial for any molecular study which relies on genetic mapping data for cloning. So, for different reasons, YAC or P1 vectors are necessary for genomic map construction. Arrayed whole genomic libraries can be subdivided into chromosome-specific ordered sets of clones and contigs constructed using a mixed ETS/clone-limited strategy.

The utility and efficiency of different methods of contig construction depend upon the genome and what information can be usefully generated in concert. For protozoan parasites, the fact that the chromosomes can be electrophoretically separated allows the generation of chromosome-specific subsets of clones (see Melville *et al.*, this volume). Therefore, although fingerprinting methods may be faster (see Smith *et al.*, this volume), the small number of clones allows less efficient but more informative methods to be employed. The sequence-tagged sites (STS) (Olson *et al.*, 1989) content mapping, and the variant ETS mapping, have been employed to align clones and add genetic information to the contig map. These methods use known DNA sequences to design oligonucleotide primers for a polymerase chain reaction (PCR) assay, where all clones which contain the sequence will give a specific product. Thus, the STS content of each clone provides the necessary information for aligning them relative to one another. The STS method can either be done with random sequence or with a sequence known to be on the ends of specific clones. Moreover, this process can be `clone limited' confined to a specific subset of clones, starting with one (or two via both ends of a clone) STS assay and systematically tagging unidentified clones until all the clones in the set are marked (Palazzolo *et al.*, 1991). This `clone limited' strategy is four to five times more efficient—requiring fewer STS assays at contig construction than a random approach. However, this is only true for contig construction at nearly 100% coverage of a genome or a chromosome. For up to about 80% coverage both strategies are about the same. Therefore, in order to add genetic information, ETSs in the form of chromosome-specific cDNAs can be used initially. After about 80% coverage, a clone limited strategy should be employed.

Some of these methods have been employed to generate YAC libraries and YAC-telomere libraries in *T. cruzi* and *Leishmania* respectively. A brief outline of the methods and results are presented below.

MATERIALS AND METHODS

Preparation of DNA

High molecular weight *T. cruzi* DNA was prepared by washing approximately 5×10^8 cells twice in 1 × PBS and resuspending the cells in 4 ml of SCE (1.0 M sorbitol, 0.1 M sodium citrate, 60 mM EDTA, pH = 7.0). In a 125 ml flask, 7 ml of lysis buffer (3% sodium lauryl sarcosinate, 0.5 M Tris-HCL, 0.2 M EDTA, pH = 9.0) was slowly added

while gently swirling. After the addition of the lysis buffer, the solution was incubated at 65 °C for 15 minutes and rapidly cooled to room temperature, and loaded onto a 15–20% sucrose pad (0.8 M NaCl, 20 mM Tris-HCL, 10 mM EDTA, pH = 8.0). The gradient was spun for three hours at 26,000 g. The DNA was recovered from the bottom of the gradient by pipetting 5 ml with the large end of a 5 ml glass pipet. The DNA solution was dialyzed against TE overnight and concentrated to 1 ml final volume by dialysis in 20% PEG.

YAC Library Construction, Storage and Screening

The library was constructed using the vector pYAC4 in the *Saccharomyces cerevisiae* strain AB1380. The protocal used was essentially as described in Burke *et al.* (1987) with the following changes: 1) the DNA size fractionation both before and after ligation was performed on a 1% low melting point agarose gel, 0.5⎫TBE run at 200 V, 30 second pulse time for 18 hours; and 2) after ligation with the pYAC4, the resulting size-fractionated gel slice was melted at 68 °C, cooled to 40 °C, treated with beta-Agarose (New England Biolabs) to the manufacturer's specifications and transformed into the yeast strain AB1380. Two thousand recombinant clones were picked and patched onto *-ura, -trp* selective plates and individually stored in 96 well plates and microtubes. The ordered arrays of the 2000 clones were printed onto nylon filters, which were subsequently screened with the ubiquitin gene probe.

YAC-Telomere Library Construction

The YAC-telomere library fo *Leishmania peruviana* was constructed essentially as in Chang *et al.* (1989) with the following changes: 1) the vector used was pJS97 (BRL); 2) the *L. peruviana* DNA was digested to completion with *Bgl*II and ligated into the *Bgl*II cloning site of pJS97; 3) the ligation mix was transformed into the *S. cerevisiae* strain YPH 252 and selected for *ura+*; and 4) one hundred clones were screened by hybridization with the trypanosomatid-specific telomere repeat $(CCCTAA)_n$.

RESULTS

YAC Cloning and Analysis of the *Trypanosoma cruzi cub* Locus

Recent data suggests that the *T. cruzi* genome consists of 15–20 chromosomes, ranging from 200 kb to greater than 2 Mb in size. Thus a haploid genome content would be about 30 Mb. In a recent collaboration with J. Swindle at the University of Tennessee Medical School (J.T. Swindle, J.W. Ajioka, J.Y. Ajioka and D. Gillespie, unpublished data), 2000 recombinant YAC clones were picked and stored as an array. Given an estimate of the genome size of *T. cruzi* to be about 3×10^7 bp, and the average insert size of the library to be 1.5×10^5 bp, 200 clones would represent one genome equivalent. Assuming random (Poisson) cloning, 2000 clones represent a 10-fold coverage, providing better than a 99%

chance of recovering any given part of the genome (Sokal and Rohlf, 1981). The library was screened with a ubiquitin probe, which should identify clones containing either of the two ubiquitin loci, defined as the 2.8 and 2.65 loci from diagnostic restriction fragments. Eight ubiquitin positive clones were recovered from the screen, five of which were analyzed for size and continuity. The clone sizes ranged from 50 to 200 kb. From the analysis of clones containing the tandemly arrayed 2.8 ubiquitin gene locus, it appears as though the YAC cloning procedure faithfully replicates the genomic DNA sequence. Moreover, the representation of the 2.8 locus (eight clones) is well within the random (Poisson) cloning expectation. In contrast, the 2.65 locus is completely absent from the library, where the probability of this occurring at random is less than one percent. This is likely due to the possibility that the 2.65 locus is in or adjacent to a region which is deficient in restriction sites, in this case *Eco*R1. If, for example, one end of the cluster is very near telomeric or centromeric highly repetitive sequences lacking in such restriction sites, the cluster cannot be cloned by this method. These data demonstrate both the power and limitations of the YAC cloning system as applied to trypanosomes. As the ubiquitin loci are the largest tandemly arrayed genes reported to date in any trypanosome, it is likely that most gene clusters will be stable in YACs.

YAC TELOMERE CLONING IN *LEISHMANIA PERUVIANA*

Since some sequences, particularly those in regions lacking restriction sites such as telomeric regions, will not conventionally clone, alternative methods must be employed for full chromosomal coverage. YAC-telomere cloning (Chang *et al.*, 1989) for such purposes has been successful for cloning chromosome ends in *Leishmania peruviana.* Five *bona fide* clones as identified by hybridization with a *Leishmania* specific telomere probe out of 100 screened were recovered. Partial restriction maps show a conserved order of several different enzyme sites, suggesting that there are related sub-telomeric sequences between chromosomes.

CONCLUSION

The results of the YAC cloning for *T. cruzi* and the preliminary results for the YAC-telomere cloning in *L. peruviana* show that material from these cloning strategies accurately and stably reflect the source genomic DNA organization. Since it is difficult if not impossible to analyze tandemly arrayed genes or loci near the ends of chromosomes by conventional means, YAC technology is an essential tool for the cloning and analysis of such regions of protozoan parasite genomes.

REFERENCES

AJIOKA, J.W., SMOLLER, D.A., JONES, R.W., CARULLI, J.P., VELLEK, A.E.C., GARZA, D., LINK, A.J., DUNCAN, I.W. and HARTL, D.L. 1991. *Drosophila* genome project: one-hit coverage in yeast artificial chromosomes. *Chromosoma* 100: 495–509.

BURKE, D.T., CARLE, G.F. and OLSON, M.V. 1987. Cloning large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236: 806–812.

CHANG, J.F., SMITH, C.L. and CANTOR, C.R. 1989. Isolation and characterization of a human telomere. *Nucleic Acids Research* 17: 6109–6127.

OLSON, M.V., HOOD, L., CANTOR, C. and BOTSTEIN, D. 1989. A common language for physical mapping of the human genome. *Science* 245: 1434–1440.

PALAZZOLO, M.J., SAWYER, S.A., MARTIN, C.H., SMOLLER, D.A. and HARTL, D.L. 1991. Optimized strategies for sequence-tagged-site selection in genome mapping. *Proceedings of the National Academy of Sciences of the USA* 88: 8034–8038.

SIBLEY, L.D., LEBLANC, A.J., PFEFFERKORN, E.R. and BOOTHROYD, J.C. 1992. Generation of a restriction length polymorphism linkage map for *Toxoplasma gondii*. *Genetics*, 132: 1003–1015.

SOKAL, R.R. and ROHLF, F.J. 1981. In: *Biometry.* San Francisco, California: Freeman and Co., pp. 82–94.

SULSTON, J., DU, Z., THOMAS, K., WILSON, R., HILLER, L., STADEN, R., HALLORAN, N., GREEN, P., THIERRY-MIEG, J., QIU, L., DEAR, S., COULSON, A., CRAXTON, M., DURBIN, R., BERKS, M., METZSTEIN, M., HAWKINS, T., AINSCOUGH, R. and WATERSTON, R. 1992. The *C. elegans* genome sequencing project: a beginning. *Nature* 356: 37–41.

VENTER, J.C., ADAMS, M.D., MARTIN-GALLARDO, A., McCOMBIE, W.R. and FIELDS, C. 1992. Genome sequence analysis: scientific objectives and practical strategies. *TIBTECH* 10: 8–11.

Summary of discussion

Chairperson: Dr. J. Ajioka
Rapporteurs: Dr. N. Murphy and Mr. R. Skilton

This session comprised three presentations on recent developments in characterizing and analyzing complex genomes. The first presentation by Dr. S. Kemp (ILRAD) outlined ILRAD's genome analysis program which aims to identify the genetic basis for trypanotolerance in N'Dama cattle. There are two parts to this project; the generation of an F_2 population of cattle segregating the trypanotolerance trait and the generation of polymorphic genetic markers, in collaboration with other laboratories, to generate a <20 cM map of the bovine genome. A recent development in the production of polymorphic genetic markers is the generation of randomly amplified DNA polymorphisms (RAPDs) through use of short oligonucleotide primers of arbitrary sequences in the polymerase chain reaction technique. The exploitation of this technique together with bulk segregant analysis may lead to a more rapid identification of genetic markers for trypanotolerance.

Dr. N. Shepherd (Du Pont Merck Pharmaceutical Company) summarized work with the bacteriophage P1 vector cloning system and the use of a new vector derivative which provides a positive selection for recombinant clones and allows easy excision of cloned inserts. A notable development has been the generation of an arrayed P1 library of primary human foreskin fibroblast cells with a three times coverage of the genome and one recombinant clone in each well of a set of microtitre dishes. The entire library is also represented on several high-density colony hybridization filters which are available for use by the research community and can be used to regroup clones into chromosome specific sub-libraries.

In the third and final presentation in this session, Dr. G. Silverman (Harvard Medical School) described the exploitation of yeast artificial chromosome (YAC) cloning systems which permit the propagation of 0.1 to 2 Mb of foreign DNA in yeast. YAC cloning is now assuming a prominent role in efforts to assemble contiguous sets of clones spanning the chromosomes of lower and higher eukaryotes. In addition, YACs have now evolved into highly versatile tools used in both the physical and functional analyses of complex genomes.

The resulting discussion focused on how the various techniques might be exploited for genome analysis of protozoan parasites. The reproducibility of RAPD fingerprints was questioned and, although variation does occur among different runs with the same oligonucleotides and target DNA, intra-batch variation is generally little. The application of this technique to bovine genome analysis and to the analysis of protozoan parasite genomes of interest to ILRAD was recognized as a powerful method that should continue to be exploited.

The advantages and disadvantages of different cloning systems were then discussed. It was recognized that YAC vectors, despite their ability to accommodate large DNA

fragments (0.2 to 5 Mb), may not be the vector of choice for the generation of libraries of some protozoan parasite genomes since, for example, many trypanosome chromosomes are smaller than the minimum-sized DNA accommodated by these vectors. However, telomeric sequences from protozoan parasites have proven refractory to cloning in bacterial systems and YAC vectors have proven extremely useful for the cloning and characterization of such sequences. Some discussion on the relative stability of cloned fragments in various vector systems ensued, and although there are many claims that YAC vectors have the best record for stability, there are conflicting reports regarding this issue. However, the relative stability of cloned DNA fragments can be verified by comparative analysis between restriction maps of genomic and cloned copies. In general, each library has inherent problems and in any genome project a number of different libraries are required. Cosmid libraries, the uses of which were not presented, would be useful for trypanosome genome analysis, since these parasites contain many mini chromosomes of between 50 and 100 kb. Bacteriophage P1 libraries would be useful for the generation of an arrayed library of both trypanosome and *Theileria* genomes and YAC vectors could be used to clone telomeric sequences. A further advantage in using the P1 vector system is the ability to use bacterial transposons, such as Tn5, with selectable antibiotic resistance markers to generate random insertions in genes and, by the transfection methods now available for trypanosomes, introduce the resultant recombinant DNA molecules back into the parasite genome for functional studies. Directed experiments with shuttle vectors which can be propagated in both *Escherichia coli* and trypanosomes would be advantageous, but since such vectors are not yet available, the standard vectors available should be utilized. It was noted that future developments would require the generation and exploitation of vectors for which expression of key genes could be regulated. To this end, work on the isolation and characterization of developmentally regulated genes for protozoan parasites is required, since there is very little understanding of how genes are regulated in these organisms. The exploitation of arbitrary primers on cDNA from different life-cycle forms of trypanosomes is currently being developed and tested at ILRAD for the rapid identification of developmentally regulated genes. Initial results show it to be an extremely powerful system for such studies.

ROUND TABLE DISCUSSION

Summary of round table discussion

Chairpersons: Prof. J. Blackwell and Dr. A. Teale
Rapporteurs: Drs. V. Nene and R. Bishop

The final session of the workshop began with short presentations by Drs. A. Teale and T. Dolan on major research projects of ILRAD's Trypanosomiasis and Tick-Borne Diseases programs, respectively. The subsequent discussions were focused on ILRAD's needs in the field of genome analysis within the context of its short- and long-term objectives.

Dr. Teale briefly outlined the projects within the Trypanosomiasis Program and highlighted the projects under which research in genome analysis was performed. He emphasized that given the complex and unique biological features of the trypanosome, vaccine development was a long-term objective and that the current research focused mainly on generating sophisticated but field-applicable tools for better understanding of the epidemiology of the disease, defining immune responses to the parasite, identifying factors involved in the pathology of the disease, characterizing the molecular mechanisms involved in drug resistance and exploring the potential of trypanotolerant cattle.

Dr. Dolan pointed out that under the Tick-Borne Diseases Program the main focus of research was on *Theileria parva*, the causative agent of East Coast fever. He briefly described the projects within the three main research program areas: epidemiology/biology, antigens and vaccine development. He emphasized that in view of significant progress made towards the identification of protective antigens of *T. parva,* the main thrust of the program in the short- to medium-term was on the development of a subunit vaccine and generation of improved diagnostic methods for elucidating the epidemiology of the disease. He urged that any commitment of research in the area of genome analysis should bear this priority in mind.

The discussion that followed these presentations was structured into five main subject areas (see below) and was aimed at giving ILRAD a set of recommendations for genome analysis.

1. Statements of objectives (existing and potential) of protozoan genome research.
2. Statements of planned activities of different groups in the area of protozoan genome research.
3. Consideration of advantages and disadvantages of different markers.
4. Consideration of resource development such as libraries and databases.
5. Consideration of different types and values of maps.

The global objectives of protozoan genome analysis research differ from ILRAD's objectives. The global rationale for protozoan research is to generate mapped markers and identify genes associated with important biological traits. There is also a fundamental interest in generating information on protozoan genome composition and of their

evolutionary relationship with other higher and lower eukaryotic organisms. Besides this, it was thought that research into the `unknown' might yield an unexpected `fall-out' as has occurred previously in similar projects.

ILRAD's interests in the analysis of the *Theileria* and *Trypanosoma* genomes were more specific. Polymorphic markers are required to identify and characterize these parasites. These markers also form the basis for studying the epidemiology of the diseases the parasites cause in the field. The identification of genes controlling important biological traits is also important, but not a major priority.

The advantages of different markers such as sequence-tagged sites, antigen genes, microsatellites, linking clones and randomly amplified polymorphic DNAs were discussed. It was clear that all these markers are extremely useful in genome analysis. Each provides different information and together they complement each other. For example, a set of polymorphic markers in *T. parva* was extremely useful in characterizing parasite populations and enabled comparative analysis of different parasite stocks and strains. The use of markers mapped on the genome also allows the measurement of genetic distances and defines the extent of genetic recombination. A mapped marker with a linkage association in one organism can be used to examine homologous regions in organisms of other species. It was recognized that all types of DNA markers improve the resolution of physical and genetic maps and are powerful tools for the identification of genes.

With regard to resource development, the potential application of different types of libraries and maps was discussed. There was general agreement that with all protozoan parasites a combination of libraries was necessary and a bottom-up approach to obtaining contig maps of protozoan parasites provides an important resource for basic research on their biology. The use of a common database format was an important consideration in setting up a resource for genome analysis.

The group also emphasized that DNA transfection technology, which can be applied to trypanosomes, should be exploited. Transfection and mapping are not mutually exclusive and the former approach may in some cases provide a more direct and rapid way of identifying genes of interest.

RECOMMENDATIONS

Recommendations

After the final session, a round-table discussion was held on the relevance of different levels of genetic analysis in *Trypanosoma* and *Theileria* species in relation to ILRAD's short- and long-term goals. The following recommendations were made.

GENERAL

1. Genome analysis should be conducted using an integrated multi-disciplinary approach involving generation of polymorphic markers, building of physical and genetic maps, production of mutants and revertants, construction of appropriate libraries, sequencing and establishment of a common computer database.
2. Since no one group of scientists can participate in all the components of genome analysis research, an informal network should be established to co-ordinate, collaborate and exchange information on closely related parasitic protozoa.
3. The choice and line of approach towards aspects of genome analysis should be dictated by the underlying objectives of each project and the nature of the genome to be studied. For example, the study of mutants and transfection technology to generate revertants is more relevant to genetic analysis of *Trypanosoma* than to that of *Theileria.*

SPECIFIC

1. ILRAD should continue to conduct research on genome analysis of *Trypanosoma* and *Theileria* species in view of the expertise, facilities and resources available and the significant progress already made in mapping and characterizing the genomes of these two important genera of protozoa. The factors identified as unique to ILRAD are the ability to maintain parasites *in vitro* and *in vivo,* to transmit them through their natural and intermediate hosts under experimental conditions and the accessibility to the field situation where parasites are naturally endemic.
2. The main objectives of the genome analysis research at ILRAD should be to identify in *Trypanosoma* species and *Theileria parva* the genes related to important biological traits such as virulence, infectivity, transmissibility, drug resistance and antigen diversity. All of these have relevance to improved control methods for these organisms.
3. ILRAD should continue to generate large numbers of genomic markers for both parasites. In addition to the currently available markers, which are based mainly on repetitive sequences and polymorphic gene sequences, the powerful technique of AP-PCR should be exploited to generate polymorphic markers. These will have applications in characterizing laboratory and field isolates, studying population genetics, identifying recombinants, studying genes of important biological traits and improving the resolution of physical maps.

4. ILRAD should concentrate on generating appropriate reference libraries for both *Trypanosoma congolense* and *Theileria parva* parasites as a resource and as a basis for long-term studies on the biology of the two parasites:
   i. for *T. parva,* which has a relatively small-sized genome (10^7 bp), it was recommended that a P1 arrayed library be constructed and the recombinant P1 clones be `ordered' to the level of assigning them to the *Sfi*I restriction fragments.
   ii. due to the much larger genome size (about eight times greater than that of *T. parva)* of trypanosomes, it was agreed that an ordered YAC library and an arrayed P1 library should be constructed.
   iii. For both parasites a telomere YAC library should also be constructed for the complete representation of both genomes.
   iv. the reference parasites to be used for this work should be a *Theileria parva* Muguga clone and the *Trypanosoma congolense* clone IL3000.
5. A common parasite genome database should be installed at ILRAD using the software program ACeDB, which was originally developed for *Caenorhabditis elegans* and is used globally for various genome programs. This database will form the repository for all the information on the molecular biology of these two parasites. The use of a common software program will enable scientists to analyze and compare related genomes of different species.
6. ILRAD should continue its efforts with the DNA transfection technique in trypanosome research in order to identify genes of interest, especially those controlling drug resistance. The focus of research should be the analysis of mutants and transfection-induced revertants by AP-PCR technology.
7. In *Theileria parva,* one of the most important biological traits is the presence of different immunological strains. Generation of several recombinants from two immunologically different parasite clones and *in vitro* screening of these recombinants with an array of polymorphic markers may lead to identification of the gene locus or the gene(s) itself involved in provoking cytotoxic T lymphocyte responses.

This workshop, held specifically to discuss genomic analysis of protozoan parasites, was the first of its kind. It was successful in bringing together ILRAD and international scientists involved in various aspects of genome analysis and provided a forum for intensive but informal discussion. The recommendations will be useful in planning and formulating future research priorities in genome analysis within the context of ILRAD's mandate.

APPENDIX:
LIST OF PARTICIPANTS

List of participants

Agaba, M.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Ajioka, J.
University of Cambridge Clinical School
Department of Medicine, Level 5
Addenbrooke's Hospital
Hills Road
Cambridge CB2 2QQ
UK

Bishop, R.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Blackwell, J.
University of Cambridge Clinical School
Department of Medicine, Level 5
Addenbrooke's Hospital
Hills Road
Cambridge CB2 2QQ
UK

Dolan, T.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Doyle, J.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Gardiner, P.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Gray, R.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Kariuki, D.
Kenya Agricultural Research Institute
P.O. Box 32
Kikuyu
Kenya

Kemp, S.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Khan, B.
Kenya Medical Research Institute
P.O. Box 54840
Nairobi
Kenya

Lessard, P.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Majiwa, P.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

McKeever, D.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Melville, S.
Microbiology & Parasitology Division
Department of Pathology
Cambridge University
Cambridge CB2 1QP
UK

Mining, S.
Kenya Agricultural Research Institute
P.O. Box 32
Kikuyu
Kenya

ole-MoiYoi, O.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Moloo, D.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Morzaria, S.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Murphy, N.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Musoke, A.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Ndolo, T.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Nene, V.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Perry, B.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Shepherd, N.
Du Pont Merck Pharmaceutical Company
Experimental Station
Wilmington
Deleware 19880-0328
USA

Silverman, G.
Joint Programme in Neonatology,
Harvard Medical School
300 Longwood Avenue, Enders 970
Boston
Massachusetts 02115
USA

Skilton, R.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Smith, C.
529 Stanley Hall
Department Molecular and Cell Biology,
University of California
Berkeley
California 94720
USA

Tait, A.
Wellcome Unit of Molecular Parasitology
Glasgow University
Bearsdon Road
Glasgow G61 1QH
UK

Teale, A.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Toye, P.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Walliker, D.
Division of Biological Sciences
Institute of Cell Animal and Population Biology
University of Edinburgh
Edinburgh
Scotland
UK

Wellems, T.
Genetics and Pharmacology Section
Laboratory of Malaria Research
National Institute of Allergy and Infectious Diseases
Bethesda
Maryland 20892
USA

Williams, D.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya

Young, A.
International Laboratory for Research on Animal Diseases
P.O. Box 30709
Nairobi
Kenya