MAPPING THE *Plasmodium falciparum* GENOME WITH YEAST ARTIFICIAL CHROMOSOMES

Paul James Meaney

PhD. Thesis

Institute of Cell, Animal and Population Biology
University of Edinburgh.

September, 1995

DECLARATION

The research reported in this thesis is original and my own, except where due acknowledgement is made, and has not been submitted for any other degree.

Paul James Meaney.

DEDICATION

I dedicate this thesis with love to my parents, James and Margaret Meaney, whose constant, unquestioning support carried me through the rough times. One day, folks, I'll explain what all the big words mean. This thesis is also dedicated to Marije. You were there every time I needed you and you are the best thing that ever happened to someone like me.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. David Arnot for giving me the opportunity to continue my studies in Molecular Biology in Edinburgh and for introducing me to the field of genome mapping with artificial chromosomes. In addition to encouraging my development as an independent scientist, Dr Arnot helped shape the course of the project and was always on hand with advice or information. I would also like to thank Dr David Walliker for having tremendous patience with me when I was getting to grips with the completely new field of malariology. Thank you for your wisdom, advice and unwavering generosity.

The number of people who have helped me, supported me and found new reserves of patience within themselves at my constant questioning are numerous. In the 3rd floor lab, I would like to thank Mr Bill Richardson for introducing me to the intricacies of PCR and sequencing. Thanks also for putting up with my eternal questioning, for giving support and help in the final months of the project with YAC mapping, sequencing and PCR and for being a good friend. Thanks also to Dr Ali Sultan, for sparing time to discuss technical and scientific problems during times when his own work was less than successful. Good luck to you, Ali - you deserve it. Grateful thanks also to Ms Awinder K. Sohal, who was constantly supportive during my time as a fledgling Ph.D student, helping me to settle in, introducing me to everyone and showing me the basics of parasite work. I would also like to thank Dr Lisa Ranford Cartwright for her unwavering patience with my endless interrogations about PCR, and for her unparalleled generosity with PCR primers, reagents and technical know- how. Thanks also to Mr. Paul Horrocks, Professor Brian Kilbey and Ms Alison Creasey for use of materials for screening my library and for numerous helpful discussions when my thinking started spinning off at a tangent. In addition I would like to thank Dr J. Rankin Small for being a good friend and for his critical evaluation of this thesis, Mr Dar- Der Ji (best of Luck!!), Mr Bill Gregory for introducing me to the wonders of Digital Image Scanning, Mr Richard Fawcett for endless supplies of parasites and for a valiant attempt at teaching parasite culturing to a total novice, Drs Jane Carlton, Mark Viney, Hamza Babiker, Rachel Taylor and Elanor Riley for various bits of advice and support. In addition, thanks also go to (Sir) Frank Johnston and Mr Graham Brown for some stunning eleventh hour photographic services. Finally I would like to thank Mr Graeme Daniels. Thanks for being there when life hit the skids and for being one of the best.

In the course of the project I was fortunate to spend time in the Molecular Parasitology Lab at the Institute of Molecular Medicine, Oxford. I would like to thank Dr Alister Craig for inviting me and Dr Chris Newbold for putting up with me for three months. It would be fair to say that without your intervention, I would not have accomplished the work presented here. Thanks also to Dr Jeremy Foster for supplying PCR primers as well as numerous useful discussions by phone, letter and e- mail, to Dr Sue Keyes for again holding my hand during the initial stages of my YAC work and making my stay in

Oxford a pleasant one, to Dr Peter Warn for his advice and for being a great extra pair of hands during *the* transformation. Finally I would like to thank Dr Tony Monaco, for his generosity in putting his tremendous YAC laboratory facilities at my disposal, and Dr Nick Fairweather, for his sound troubleshooting advice, expert opinions and for just being a good friend. In addition I would like to thank Dr Michael Lanzer for his advice in YAC transcriptional analysis.

Over the last eight years here in Edinburgh, I have made numerous friends, all of whom have helped me one way or the other. I would like to thank a rather unique bunch of people and some of the best friends I have ever had- Graeme Anderson, Fraser Macrae, Leo James, Stewart Mullinger, Ed Rodesano, Ian Pearmain, Andy Cotgreave, Mike White, David Ness, Paul Riggs, Chris and Jonnie Burden, Raffi Freire and the rest of the Sneeeky's Ultimate Frisbee Team past and present. Thanks a lot guys for being there and for putting up with my temper on the pitch and me off it. In addition to all of the above, special thanks to PC God Andrew Wilson for some superb computer graphics. I would also like to thank Dr Jackie Nairn, Anya Weischermann, Harry Bear, Harry Smythe, Simon Baker, Ed Wilding, Murray Read and Andy James for being great, but especially Colin Maclellan, Holger Beyer and James Shaw, for being some of the best, most supportive friends I have ever had. And again, to Marije. Thanks for listening, bullying me and being there. In Oxford, thanks goes to Jack, Adam, Ralph and Rachel, Steve, George, Gary, Lee, Derek, Richie, Chris, Danny, Gordon and all the Sharks and Monks.

Finally, again, to my family. Your love and support has carried me when this project was running less than smoothly. Your unfailing support gave me the drive and energy to finish these studies. I hope I haven't let you down.

Abstract.

The World Health Organisation estimates that nearly 40% of the worlds population is at risk from malarial infections, leading to nearly 2 million deaths per year. Problems of drug resistance are increasing and consequently necessitate consideration of alternative strategies for disease control. One possible avenue is a greater understanding of the molecular and cellular biology of the parasite, studies which have been hindered by the biological complexities of the organism. Advances in molecular biological techniques, such as PCR, have increased the rate of information production to such an extent that the ability to rapidly nap any new gene sequence can greatly assist information assembly. Yeast Artificial Chromosome cloning vectors, with their capacity to maintain up to 1 Mb of cloned DNA in a stable form, have proved extremely useful in mapping the genomes of higher eukaryotes. These vectors posses features which can circumvent some of the problems associated with classical molecular manipulation in *Plasmodium falciparum.*

The research presented in this thesis is aimed at contributing to genome mapping in *P. falciparum.* The primary objective is the construction of a complete, detailed YAC- based physical map of chromosome 6, with a resolution of 10 Kb. To accomplish this, an 1100 clone YAC library of the *P. falciparum* clone HB3 was constructed. The library contains clones with an average insert size of 100 Kb. Insert DNA is stable when cultured over 100 generations and the library is predicted to have a four to five fold genome redundancy, corresponding to 90% of the genome. Chromosome 6 specific YAC's have been isolated and chromosome mapping initiated. Overlapping YAC's have been identified by using Sequence Tagged Site markers obtained from the 5` and 3` ends of each YAC by Inverse PCR. A total of 700 Kb of *P. falciparum* DNA has been cloned and this has been extensively mapped with seven restriction enzymes. Maps for all available YAC's will be presented. In addition, an attempt has been made to evaluate the degree of stage specific gene expression of cloned DNA within each YAC.

The implication of these findings for genome mapping in *P. falciparum* will be discussed in the thesis.

LIST OF ABBREVIATIONS

(d)dATP	2', 3' dideoxyadenosine- 5' - triphosphate
(d)dCTP	2', 3' dideoxycytidine- 5' - triphosphate
(d)dGTP	2', 3' dideoxyguanosine- 5' - triphosphate
(d)dTTP	2', 3' dideoxythymidine- 5' - triphosphate
Amp	Ampicillin
ARS	Autonomously Replicating Sequence
BSA	Bovine Serum Albumin
cDNA	complementary Deoxyribonucleic Acid
CEN	Centromere
CEPH	Centre d' Étude du Polymorphisme Humaine
dATP	Deoxyadenosine- 5' - triphosphate
dCTP	Deoxycytidine- 5' - triphosphate
dGTP	Deoxyguanosine- 5' - triphosphate
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic Acid
DNAase	Deoxyribonuclease
DTT	Dithiothreitol
dTTP	Deoxythymidine- 5' - triphosphate
EDTA	Ethylenediamine N,N,N',N' Tetra-acetic Acid
FISH	Fluorescence *In Situ* Hybridisation
GBP	Glycophorin Binding Protein
GCG	Genetics Computer Group
GLARP	Glutamine and Alanine Rich Protein
IPCR	Inverse Polymerase Chain Reaction
Kb	Kilobase pair
KAHRP	Knob- Associated Histidine Rich Protein
OD	Optical Density
PBS	Phosphate Buffered Saline
PCI	Phenol : Chloroform : Isoamylalcohol
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gradient Gel Electrophoresis
RNAase	Ribonuclease
RPMI	Rosewell Memorial Park Institute
RT- PCR	Reverse Transcription PCR

STS	Sequence Tagged site
SDS	Sodium Dodecyl Sulphate
TAE	Tris Acetate EDTA
TBE	Tris Borate EDTA
TEL	Telomere
TRAP	Thrombospondin Related Antigen protein
TRP	Tryptophan
URA	Uracil
UV	Ultra Violet
YAC	Yeast Artificial Chromosome
YPD	Yeast extract, Peptone and Dextrose Media

TABLE OF CONTENTS

CHAPTER 1 : INTRODUCTION

CHAPTER 2 : MATERIALS AND METHODS.

1. Materials.

2. Methods.

CHAPTER 1.

INTRODUCTION

1.1 The Global Malaria Situation

In terms of morbidity, mortality and economic effect, malaria is the most significant disease of human populations in the tropics. In man, the disease is caused by four species of the protozoan genus *Plasmodium*: *P. falciparum*, *P. vivax*, *P. malaria* and *P. ovale*. *P. falciparum* accounts for most of the infections in Africa and one third of the infections in the rest of the world. Although few reliable statistics exist on malaria morbidity and mortality in Africa, epidemiological predications indicate that 2 billion people (nearly 40% of the worlds population) in 102 countries are at risk; that at least 300 million people are infected with malaria parasites each year, leading to over 100 million clinical cases and perhaps 2 million deaths (Schuster and Milhous, 1993). In Africa, the majority of cases are in children under the age of 5, with pregnant women being in a particularly high risk group due to the increased risk of severe disease as a result of immunosuppression (Menendez, 1995). The clinical course of even treated infections may last 5 - 15 days, resulting in an incapacity of the patient, leading to a loss of labour productivity. This, coupled with the steadily increasing expense of a complete course of antimalarial treatment (Kondrachine and Trigg, 1995), results in a tremendous cost to developing nations, both socially and economically (WHO 1992).

The clinical symptoms of malaria are caused by the development of the parasite in red blood cells. The main symptom is fever, which shows a clear cyclic pattern in all but *P. falciparum* infections. Falciparum malaria is the most dangerous form of the disease because it can have life threatening complications such as anaemia, respiratory distress syndromes and cerebral malaria. Many infected people are partially protected from these complications by immunity acquired from previous recent infections. Severe complications are found predominantly in relatively non-immune people, notably children, expatriates and migratory workers entering endemic areas, and in the inhabitants of areas with unstable malaria transmission. Infections are also compounded by social geography, including poor housing, crowding, rapid urbanisation and deforestation, war and civil disturbances (Kondrachine and Trigg, 1995).

Two major obstacles to malaria treatment are the spread of drug resistance within a population and spiralling costs. In 1955, the 8th World Health Assembly proposed a program for the world-wide eradication of malaria based upon vector and parasite targets. In 1969, the WHO was forced to revise the program to one of control. Chloroquine resistance first appeared in Thailand and South America in the 1950's (Koella, 1993) although it was not detected in east Africa until the late 1970's (Schapira *et al.*, 1993). Resistance has subsequently spread across the continent and although variable in degree, resistance is generally becoming more pronounced. For example incidence of RIII (high) chloroquine resistance in Malawi has increased from 8.5% to 33% in 9 years. There is still no recorded occurrence of resistance in Central America and the Caribbean. Reports from Thailand demonstrate a 50% inefficacy of mefloquine, but resistance to this drug is not yet an operational problem in Africa due to its limited use. Chloroquine resistance was first reported in *P. vivax* infections in 1989 in Papua New Guinea and

has since been detected in Indonesia and Vanuatu (WHO 1994). This increase in resistance in multiple species of malaria parasites has led to the introduction of first, second and third line treatment strategies, which may in time result in the development and spread of a multidrug resistance phenotype. The application of multiple prophylaxis to control infection also places an added economic burden on the individuals at risk particularly since malaria often occurs in poor and remote areas where people are not often able to afford treatment.

Compounding these problems is the cost of developing new antimalarials. The current estimate is that the cost of developing new drugs has more than doubled in the past ten years. It now takes an average of 231 million US dollars and 12 years to take a chemical from the research laboratory to the field. Only one in every 3000 compounds evaluated by the US Army in primary malaria screens reached a final test system. Approximately half of these were deemed suitable to advance to preclinical toxicology trials and from this final group, only one in five pass Federal Drug Administration regulations (Schuster and Milhaus, 1993). Since malaria is a disease of developing countries, it is difficult for such countries to afford intensive antimalarial research programmes or to pay for expensive new medicines. In addition, current antimalarials are expensive despite WHO recommendations for price regulation. Consequently, alternative approaches to malaria eradication are required. For chemotherapy, possible alternatives include artemisinine, a potent antimalarial derived from the plant *Artemisia anniva* which has demonstrated efficacy in the treatment of cerebral and other severe malarias as well as being effective against other types of resistant parasites (WHO, 1992). In recent years, efforts to develop effective vaccines have resulted in trials of the Spf66 vaccine in Tanzania, which indicated a 30% reduction in the incidence of first episodes of malaria (Alonso *et al.*, 1994), leading to requests for the vaccine from Bolivia, Zaire, Kenya Tanzania, Mozambique and Colombia and the development of a second generation version of the vaccine (Maurice, 1995). However, this promising result was not replicated in a clinical trial in the Gambia (D'Alessandro *et al.*, 1995)

A greater understanding of the molecular biology of the parasite might facilitate novel therapeutics and the design of new drugs. Suitable targets for antimalarial drugs could include metabolic pathways in the parasite that differ from the host, parasite enzymes which display functional and structural differences from their host counterparts or signal transduction pathways within the parasite which could be interrupted. The genes encoding such targets would need to be mapped, cloned, sequenced and expressed. To facilitate such "knowledge intensive" strategies, the ability to position any new genes rapidly onto a detailed map of the parasite genome would greatly streamline the generation, storage and accessibility of such information. Detailed genetic and biochemical "blueprints" are the goal of several new initiatives in the genome mapping of human pathogens. The work presented in this thesis is oriented towards such genome cloning, mapping and genetic charaterisation projects.

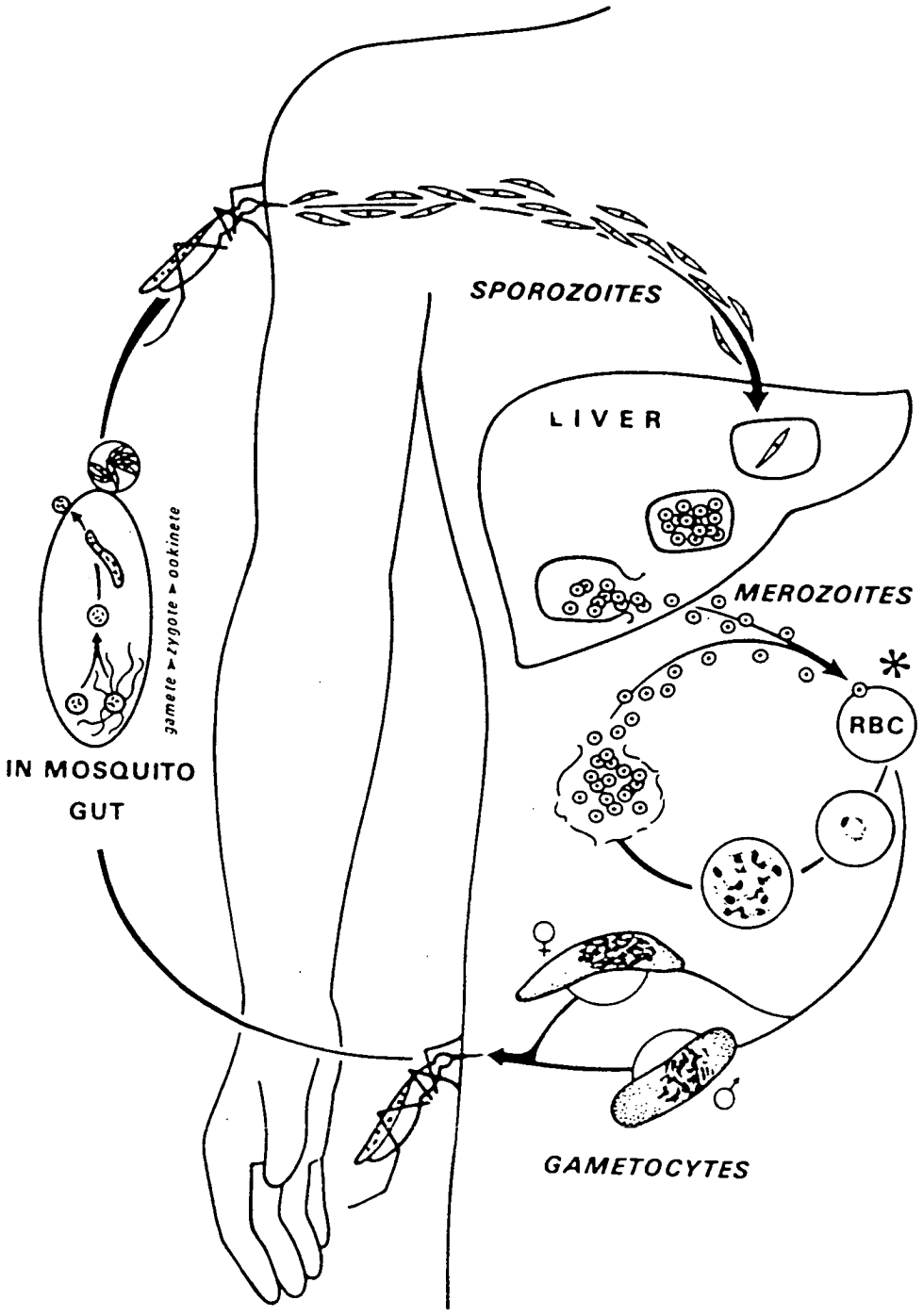
1.2. Biology of Malaria Parasites.

1.2.1 The Life Cycle

To understand some of the problems associated with the investigation of the molecular biology of malaria, the complex life cycle of the parasite must be considered. As shown in figure 1.1., all species of human malaria parasites demonstrate essentially similar characteristics: a sexual phase in certain *Anopheles* mosquitoes and an asexual phase within the vertebrate host.

The sexual fusion of 'male' and 'female' gametes starts after a female *Anopheles* mosquito ingests malaria parasites in a blood meal taken from an infected human host. Digestion of erythrocytes and asexual parasites occurs in the mosquito stomach, whereas gametocytes undergo further development. The haploid male gametocyte undergoes a process of rapid DNA replication in the gut of the mosquito to produce eight flagellated gametic cells. These cells locate and fertilise the female gametes (macrogametes) to form diploid zygotes. Maturation of the zygote results in the first meiotic division event, visualised by electron microscopy shortly after fertilisation (Sinden and Hartley, 1985) which is presumably followed by the second division. This results, over the next 18 - 24 hours, in the formation of an ookinete, which migrates to the epithelial cells of the stomach wall where it undergoes further maturation to form an oocyst. Sporogony, leading to the formation of several thousand haploid sporozoites, occurs within the oocyst which ultimately ruptures to release the sporozoites into the mosquito body cavity. These subsequently migrate into the salivary glands and are injected into the bloodstream of the host when the mosquito feeds.

After passage through the blood, sporozoites enter the parenchymal liver cells (hepatocytes) where they undergo a series of rapid mitoses (pre- erythrocytic shizogony). This results in the formation of around $3x\ 10^4$ merozoites which are released on hepatocyte rupture into the bloodstream. The duration of this phase (between 6 - 16 days), the size of the schizont and the number of merozoites released depends on the species of *Plasmodium*. The released merozoites invade red blood cells and the erythrocytic cycle begins. Maturation of the merozoites involves microscopically distinguishable transformations from ring forms to trophozoite forms and finally mature schizonts. After schizogony, red blood cell lysis occurs which releases between 8 - 32 merozoites into the bloodstream. These merozoites can reinvade new erythrocytes and continue asexual replication, or can differentiate into male and female gametocytes. The above asexual erythrocytic cycle takes 48 hours in *P. falciparum*, *P. ovale* and *P. vivax*. Blood cell rupture is associated with fever, as patients infected with malaria demonstrate schizogonic periodicity i.e. febrile paroxysms associated with a pattern of attacks every third day .


SPOROZOITES
LIVER
MEROZOITES
RBC
IN MOSQUITO
GUT
gamete ▶ zygote ▶ ookinete
GAMETOCYTES

1.2.2 The Genome of Malaria Parasites.

The malaria genome is eukaryotic in nature. Chromosomes are contained within nuclei, possess telomeric structures (Dore *et al.*, 1986; Blackburn, 1990), are associated with histone scaffold proteins (Creedon *et al.*, 1992; Longhurst and Holder, 1995; Bennett *et al.*, 1995) and are ordered into nucleosomes (Cary *et al.*, 1994). *P. falciparum* posesses small nuclear ribonucleoproetins (Francoeur *et al.*, 1985), and some genes possess introns. Messenger RNA contains post-transcriptional modifications such as a 5` cap and 3` polyadenylated tail, characteristic of eukaryotic systems (Hyde *et al.*, 1984; Levitt, 1993).

Due to their small size and the fact that they do not condense during mitosis, the chromosomes of malaria parasites cannot be visualised by conventional light or electron microscopy (Janse, 1993). However, the development of Pulsed Field Gradient gel electrophoresis (see section 6) has allowed the resolution of intact chromosomes on conventional gels (Kemp *et al.*, 1985; Van der Ploeg, *et al.*, 1985). The genome of *P. falciparum* is organised into 14 chromosomes (Langsley *et al.*, 1987; Wellems *et al.*, 1987; Langsley *et al.*, 1988) ranging in size from 630 kb to 3.5 Mb (Foote and Kemp, 1989). This is in agreement with the number of kinetochores observed in serial sections of nuclei (Prensier and Slomianny, 1986). In addition the genome contains 6 kb and 35 kb extrachromosomal elements. The smaller element corresponds to the mitochondrial genome, whereas the larger element contains rRNA genes and is thought to share common ancestry with modern plant plastid DNA (Feagin, 1994).

The exact size of the genome may vary slightly between isolates; however estimates of 1 - 3 x 10^7 bp have been proposed by several groups by the determination of the yield of isolated parasite DNA and by the cloning of size-selected DNA fragments digested with restriction enzymes (Goman *et al.*,1982; Wellems *et al.*, 1987). However, the best estimates are inferred from genome mapping work: a Yeast Artificial Chromosome map for chromosome 2 has revealed an accurate size of 1.03 Mb (Lanzer *et al.*, 1993) and an estimate of 26 Mb for the whole genome has been determined from karyotype analysis (Gu *et al.*, 1990). Thus in comparison with genome sizes of 100 Mb for *Caenorhabditis elegans* and 3000 Mb for human, for both of which YAC based mapping projects are either completed or in progress, the feasibility of a YAC based *P. falciparum* map becomes evident.

The base composition of the malaria parasite genome varies from A+T rich, in the case of *P. knowlesi* and *P. fragile* (McCutcheon *et al.*, 1984) to extremely A+T rich, in the case of *P. falciparum* and *P. lophurae* (Pollack *et al.*, 1982; McCutcheon *et al.*, 1984; Weber *et al.*, 1987). *P. falciparum* demonstrates an average A+T content of 82% (Weber *et al.*, 1987) with noncoding regions showing a higher A+T, on average 86% compared with a 70% value for gene sequences (Weber *et al.*, 1987; Li *et al.*, 1989; Holloway *et al.*, 1990). Repetitive DNA sequences have been reported for several malaria species, with 10% of the *P. falciparum* genome consisting of repetitive DNA (Hough- Evans and Howard, 1982). A 21 bp, tandemly repeated A+T rich repetitive element, *rep20*, has been reported (Aslund *et al.*, 1985) which is present in all parasite chromosomes (Oquendo *et al.*, 1986)

1.2.3 Molecular Analysis of the *P. falciparum* Genome.

Gene cloning in *P. falciparum,* as in other organisms, has been revolutionised by the invention of the Polymerase Chain Reaction (PCR) technique (Saiki *et al.*, 1989). Earlier experiments in gene cloning were almost entirely based upon studies of immunoactive proteins. 'Classical' gene cloning in *P. falciparum* relied upon the screening of expression libraries with either total human sera, identifying such genes as GARP (Triglia *et al.*, 1988b), RESA (Coppel *et al.*, 1984) and ORA (Favaloro *et al.*, 1984) or with specific antibodies raised to a particular malarial antigen, identifying genes including MSA2 (Smythe *et al.*, 1988) and CARP (Wahlgren *et al.*, 1986). Developments in hybridisation technologies permitted the identification of genes with probes designed from published peptide sequences, such as RNA polymerase II (Li *et al.*, 1989) and β Tubulin (Delves *et al.*, 1989) followed by degenerate probes based on consensus sequences with other organisms, such as DNA polymerase α (Ridley *et al.*, 1991) and actin genes (Wesseling *et al.*, 1988). However, in the past four years PCR has been used to identify novel genes by the use of PCR primers designed using published evolutionarily conserved sequences. Consequently, genes can be identified by their evolutionary conservation rather than their immunogenicity. Examples of genes identified by this method include DNA polymerase δ (Fox and Bzik, 1991), a vacuolar ATP- ase (Karcz *et al.*, 1993) and RAN (Sultan *et al.*, 1994), With such a wealth of gene sequence information being generated so rapidly, the need for organisation and genetic mapping becomes of paramount importance.

1.2.4 Mapping of the *P. falciparum* Genome

With the introduction of Pulsed Field Gradient (PFG) Gel electrophoresis (Schwartz and Cantor, 1984) and specifically Contour Clamped Homogenous Electric Fields (CHEF- Chu *et al.*, 1986), it has become possible to visualise and resolve the chromosomes of *P. falciparum* (Gu *et al.*, 1990). Such studies have revealed the highly polymorphic nature of the genome. Chromosomes display considerable size polymorphism both in natural populations of geographically distinct isolates (Kemp *et al.*, 1985; Kemp *et al.*, 1987; Sinnis and Wellems, 1988; Pologe and Ravetch, 1988) as well as in culture (Biggs *et al.*, 1989; Corcoran *et al.*, 1988; Scherf and Mattei, 1992). Such polymorphisms can arise by crossing over events (Sinnis and Wellems, 1988), by chromosome breakage followed by *de novo* telomere addition (Pologe and Ravetch, 1988) or by intrachromosomal recombination within subtelomeric repeat sequences (Corcoran *et al.*,1988). More recently, sequence evidence has demonstrated proof of intragenic recombination within the MSP- 1 gene after a laboratory cross was performed between two isolates (Kerr *et al.*, 1994). In addition, the genome demonstrates a degree of plasticity in response to drug pressure - parasites grown under pyremethamine and chloroquine pressure show chromosome polymorphism as a result of gene amplification (Foote *et al.*, 1989; Wilson *et al.*, 1989; Watanabe and Inselburg, 1994).

As techniques became available for separation of large DNA molecules, linkage and long range physical mapping strategies are being initiated. For physical maps, a combination of a "top down" approach using chromosome- specific markers to probe PFG separated genomic restriction fragments (Watanabe *et al.*, 1994), or a "bottom up" approach, by covering each chromosome in a series of ordered clones (Rubio *et al.*, 1995) can be used to determine high resolution physical maps of all chromosomes. Top- down, low resolution, long- range physical maps now exist for chromosomes 1 and 2 (Corcoran et al., 1988), chromosomes 3, 11 and 12 (Walker- Jonah *et al.*, 1992), chromosome 4 (Sinnis and Wellems, 1988; Watanabe and Inselburg, 1994; Rubio *et al.*, 1995), chromosome 5 (Foote *et al.*, 1989) and chromosome 9 (Shirley *et al.*, 1990). Current genome maps are presented in figure 1.2.

1.2.5 Problems involved in the investigation of *P. falciparum*.

The nature of the life cycle of the parasite and some of the characteristics of the genome have presented problems that have made molecular biological studies of the parasite difficult. Possibly because of the unusually high A+T content of the *P. falciparum* genome, large fragments of DNA are generally unstable when cloned in *E. coli* (Weber, 1988a; Triglia *et al.*, 1992). Until 1978, no *in vitro* culturing system was available. Since then, the development of refined culture systems involving parasitised red blood cells in an RPMI based medium saturated with a gaseous mixture of oxygen, nitrogen and carbon dioxide (Trager and Jensen 1976; Zolg *et al.*, 1982) has meant that quantities of parasite DNA can now be obtained for molecular biological analysis. The complexity of the life cycle coupled with limited selection systems and a rather fragile culture system have meant that classical molecular manipulative techniques involving gene inactivation and complementation have not been possible. A transfection system was developed for *P. gallinaceum* (Goonewardene *et al.*, 1993) leading to the development of a transfection system for *P. falciparum*, meaning that cloned sequences can now be inserted into the parasite and potentially modified (Wu *et al.*, 1995).

Genetic analysis of the parasite has proven difficult. Studies in model organisms such as yeast, *Drosophila* and *Escherichia coli* have been facilitated by such factors as simple life cycles, marker availability and simple selection of mutants. Although two genetic crosses have been performed in *P. falciparum*, by passaging two different, genetically marked parasite clones through mosquitoes and allowing completion of the life cycle through a primate host yielding both recombinant and parental progeny (Walliker *et al.*, 1987; Wellems *et al.*, 1990), the technical difficulty in performing such crosses, the lack of polymorphic markers and the plasticity of the genome *in vitro* has meant that linkage analyses are problematic. However, such studies have demonstrated that the chloroquine resistance locus is unlinked to the multidrug resistance locus, *mdr* (Wellems *et al.*, 1990). A long- range RFLP map of the genome has also been determined (Walker- Jonah *et al.*, 1992). As will be discussed, many of these difficulties can be alleviated by cloning the genome in Yeast Artificial Chromosomes (YAC's). YAC vectors (Burke *et al.*, 1987) can clone large fragments of DNA as stable artificial chromosomes in yeast

1
2
3
4
5
6
7
8
9
10
11
12
13
14

and various PCR or subcloning strategies can be employed to construct multiple, polymorphic, chromosome- specific markers for a more complete linkage map of the genome. In addition detailed physical maps would facilitate novel gene mapping and characterisation, as such information could be quickly mapped and linkage groups established.

1.3. Pulsed Field Gradient Gel electrophoresis.

Electrophoresis using agarose or polyacrylamide gels has been the simplest method of separating DNA fragments. The operating principle is that the negatively charged phosphate backbone of the DNA helix causes migration towards an anode when an electric field is applied. There is a logarithmic relationship between mobility and molecular weight because of the sieving properties of the gel matrix. However, the upper limit to resolution of DNA fragments with conventional electrophoresis is approximately 20 kb (Smith and Cantor, 1986), which renders the system useless for genome- level analysis, where a ten- to a hundred fold increase in the order of magnitude of fragment resolution is required (Barlow and Lehrach, 1987). Several groups therefore developed an alternative approach to electrophoresis, namely Pulsed Field Gradient Gel Electrophoresis, which is capable of separating molecules of up to 2 Mb (Schwartz and Cantor, 1984; Carle and Olson., 1984).

1.3.1 Theory of PFG

PFG electrophoresis relies upon the application of either homogenous or non- homogenous electric fields in alternate directions for defined periods of time ("pulses") to the DNA fragments to be separated. The basic principle of PFG is that each time the field direction changes, DNA molecules spend part of the subsequent pulse time 'relaxing' (i.e. changing shape and orientation in response to the altered field), before attainment of new configurations which permit translation through the gel in the new direction. DNA relaxation times are approximately proportional to DNA size. As relaxation times approach the pulse duration, the net mobility varies sharply and very high electrophoretic resolution can be observed (van Ommen *et al.*, 1986; Muller *et al.*, 1987; Drumm *et al.*, 1988; Fulton *et al.*, 1988). If the reorientation time is slower than the pulse time, then the DNA fragments will no longer respond to the alternating fields and gel migration becomes independent of the pulse applied (Schwartz and Cantor, 1984; Smith and Cantor, 1987). The pulse interval is thus a critical factor in determining the optimal window for resolution. A plot of DNA velocity in a gel against pulse time will yield a curve with a linear domain denoting the maximal window of fragment size separation for that particular switch time. Above and below this window, DNA fragments migrate as compression zones (Birren *et al.*, 1988). Other factors contributing to the separation of fragments are (i) buffer temperature, (ii) agarose concentration, (iii) voltage gradient and (iv) reorientation angle (i.e. the angle between the direction of fragment mobility after field switching). In general, for optimal resolution a low buffer temperature, relatively low

(1.2%) agarose concentration, low voltage gradient and reorientation angles of 90° yield conditions where separation is determined by the switching interval (Birren *et al.*, 1988).

1.3.2 Variants of PFG

Early pulsed field gels depended upon perpendicularly oriented non- uniform electric fields i.e. the field creates a gradient with a net movement in one direction (Schwartz and Cantor, 1984). Intact DNA samples are prepared in agarose to minimise shearing. The theory was first applied in the construction of the Orthogonal Field Alternating Gel (ORFAGE - Carle and Olson, 1984) which demonstrates many of the features of modern pulsed field gels (including buffer recirculation to prevent heating and timed switching of the electrodes). However a major limitation of the system were 'edge effects' - samples loaded in the flanking wells underwent complex trajectories making size comparisons between lanes difficult. The system was refined with the development of the Field Inversion Gradient gel (FIGE- Carle *et al.*, 1985). This differs from ORFAGE in that DNA separation results from periodic inversion of the electric field resulting in a reorientation angle of 180°. The major advantage of FIGE is that DNA samples run as linear tracks, circumventing edge effects. However, a drawback of the system is that it demonstrates a biphasic separation of fragments of different size ranges for a given switching interval e.g. molecules of up to 500 kb and greater than 800 kb were separated but fragments of 600 - 700 kb demonstrated no mobility (Anand, 1986). Such a problem meant that the size of a DNA fragment could not be determined from its mobility. This has been overcome by using computer controlled switching interval ramps within a run.

The most common form of PFG in use today is the Contour Clamped Homogenous Electric Field, or CHEF (Chu *et al.*, 1986). The system relies on having a series of electrodes in a hexagonal array, generating field directions separated by 120°. The net effect is a forward migration, with the leading electrodes having a higher potential than the lagging ones. The design results in DNA fragments separating in a straight track, independent of the location of sample loading on the gel and has been used in the efficient separation of fragments in excess of 5 Megabases (Vollrath and Davis., 1987). Several commercially available CHEF machines exist, the most advanced being the Bio- Rad CHEF Mapper™. With such machines, it is possible not only to vary switch times and field strengths but also field re-orientation angles for increased size resolution.

1.3.3 PFG and Mapping

The development of PFG meant that karyotypes could be assigned to lower eukaryotes, whose chromosomes could now be resolved as individual bands on a gel. Karyotypes have been published for numerous organisms, including *Pseudomonas* (Grothues and Tummler, 1991), *Saccharomyces cerevisiae* (Carle and Olson, 1985), *Candida albicans* (Lasker *et al.*, 1989), *Dictyostelium discoideum*

(Cox *et al.*, 1990), *Neurospora crassa* (Orbach *et al.*, 1988) and *Aspergillus nidulans* (Brody and Carbon, 1989).

The major impact of Pulsed Field Gradient electrophoresis has been the physical mapping of large genomes. Genetic maps can position loci relative to each other by means of linkage probabilities and can provide a resolution of between 5 and 25 centiMorgans (cM), where 1 cM in humans is defined as a 1% incidence of recombination during meiosis and corresponds to approximately 1 Mb. The ability to separate fragments of DNA up to 10 Mb facilitates the first step in the linkage of physical and genetic maps. Physical maps are generated by compiling data resulting from the complete and partial digestion of intact chromosomes with infrequently cutting enzymes followed by subsequent hybridisation to locus-specific probes. By using a variety of enzymes, a ladder of fragments is produced by partial digestion that indicates the linear order of sites along a chromosomal region (Barlow and Lehrach, 1987). An example of such a link is in the mapping of the human Cystic fibrosis locus: linkage analysis demonstrated the presence of two markers, MET and D7S8, which flank the CF gene (Lathrop *et al.*, 1988). A mapping strategy as described using labelled MET and D7S8 probes resulted in the construction of a detailed restriction map of 3Mb encompassing the region containing these markers (Drumm *et al.*, 1988;). Long range chromosomal maps can now be constructed using rare- cutting enzymes and single probes distributed every 1 Mb. Such maps can then be correlated with established Linkage maps (Donis- Keller *et al.*, 1987). Pulsed Field Gels have been used to construct multiple long range maps of defined chromosomal regions, such as a 2.3 Mb map surrounding the murine MHC locus (Muller *et al.*, 1987), a 4 Mb map of the Duchenne Muscular Dystrophy locus (van Ommen *et al.*, 1986) and 12 Mb surrounding the CF locus (Fulton *et al.*, 1988). With YAC's containing contiguous inserts of up to 1.4 Mb, and with restriction maps of multiple enzyme sites internal to the YAC (Nelson *et al.*, 1989), the Human Genome Project goal of constructing a complete physical map of the human genome with a resolution between restriction sites of an average 100 kb is now within reach. Such a strategy can be applied to the physical mapping of any organism's genome cloned in YAC's.

1.4.Yeast Artificial Chromosome Cloning.

1.4.1 Introduction.

Classical gene cloning has relied on the use of small plasmids, bacteriophage vectors and cosmids (Wahl *et al.*, 1987). However the upper limit to cloning in cosmids is approximately 50 kb and this limited capacity resulted in a discrepancy between the centiMorgan resolution of linkage mapping and the resolution of physical mapping. Therefore, a limitation to eukaryotic genome analysis has been the availability of suitable cloning vectors. The genomes of eukaryotes and higher organisms are complex, with some genes spanning hundreds of kilobases (Anand *et al.*, 1991) coupled with regulatory sequences that are too distant to be contained in plasmid or phage vectors (Banerji *et al.*, 1981). Genes

may also be clustered into functional expression units spanning megabase regions of DNA (Kozono *et al.*, 1991; Mendez *et al.*, 1995). These features, together with the fact that certain DNA sequences are particularly difficult to clone in conventional vectors have led to the development of cloning vectors capable of circumventing such problems.

The problem of the discrepancy between physical and genetic maps can be illustrated in the search for the human Cystic Fibrosis (CF) locus. Linkage mapping is based on the fact that genes that are physically close to each other in the genome are likely to be inherited together. Therefore a known locus could effectively tag a disease locus by examination of inheritance of both markers through a pedigree. Extensive mapping defined six markers linked to CF inheritance: MET, COL1A2, TCRB, D7S8, D7S13 and D7S16. Further analysis suggested the order MET - CF - D7S8 (Lathrop *et al.*, 1988). Physical mapping using further markers identified by these original markers demonstrated the MET - D7S8 interval to be 1300 - 1800 kb (Drumm *et al*, 1988; Fulton *et al.*, 1988), a distance too large to be linked by the physical mapping techniques of the time. Further linkage studies were used to identify markers more closely linked to CF. A novel technique of 'chromosome jumping' was developed and over 60 cosmid and jumping clones were isolated (Rommens *et al.*, 1988). The gene was eventually identified by sequencing open reading frames from the cDNA sequences of CF patients (Riordan *et al.*, 1988). Thus a time consuming combination of linkage mapping, physical mapping and molecular manipulative techniques had to be employed to identify this important human disease gene locus.

A major advance in the analysis of complex genomic DNA systems was the description of the Yeast Artificial Chromosome cloning vector - or YAC (Burke *et al.*, 1987). The advantage of YAC's over established cloning vectors was the large insert capacity- the YAC vector could stably propagate genomic DNA of up to 1000 kb. The concept of the YAC is simple. If the basic components that define a chromosome could be inserted into a plasmid vector then that vector should behave as if it were a chromosome when a large segment of foreign DNA is introduced. As will be discussed, the YAC has proved to be a very versatile resource for biologists, and although YAC analysis of human DNA has been associated with extensive problems of cocloning (Green *et al.*, 1991a) and insert instability (Neill *et al.*, 1991), the advantages of this system when applied to the mapping and analysis of other organisms are becoming apparent. With regard to the above example, YAC clones spanning the CF locus were subsequently identified, taking far less time and research effort (Green and Olson, 1990b.).

The work in this thesis details the cloning and mapping of chromosome 6 of the human malaria parasite *Plasmodium falciparum*, a vector- borne parasite common to Africa and south east Asia. As part of the Malaria Genome Mapping Collaboration, the aim of this work is to use YAC's to clone chromosome 6 - specific DNA and arrange them as sets of ordered overlapping clones and to present a detailed physical map of the chromosome. In the rest of this chapter I shall review the current state of YAC applications to the mapping and investigation in the Human Genome and how such strategies might be applied to the mapping of *P. falciparum*.

1.4.2 YAC Components.

To function as an artificial chromosome, YAC's must possess elements that define a true chromosome. Such elements allow (i) stable maintenance of the chromosome within a cell, and (ii) accurate replication and transmission of error- free information to daughter cells. YAC's must also be selectable to discriminate between recombinants and non recombinants. Consequently, YAC's posses centromeres, telomeres, eukaryotic origins of replication and selectable auxotrophy markers. In addition, as a further selective marker, the YAC's posses the *ade2-* colony colour assay mutation. The structure of the YAC vector used in this study, pYAC4 is presented in fig 1.3.

1.4.2.1 Centromeres.

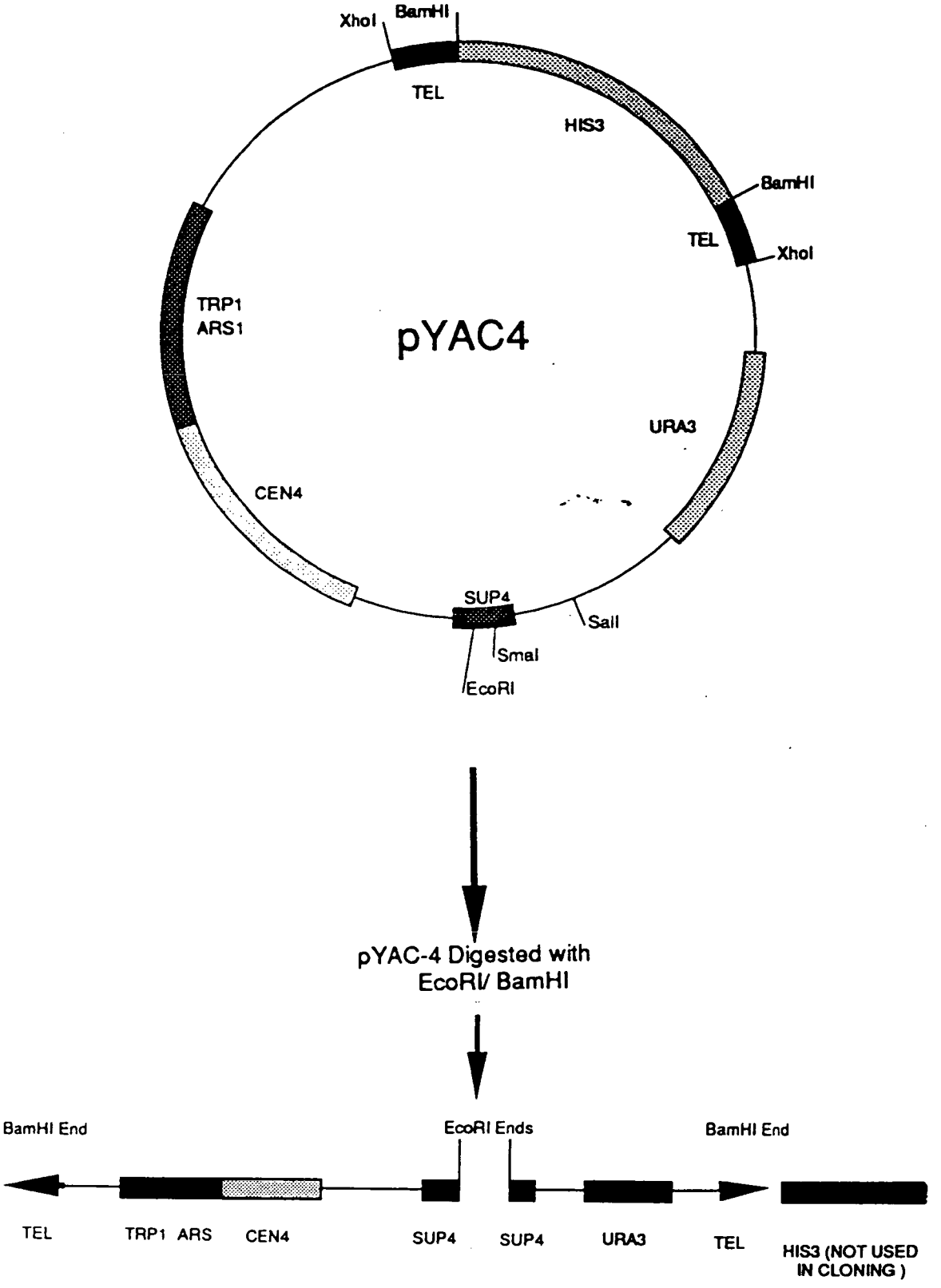
Centromeres of eukaryotic chromosomes regulate and maintain faithful segregation of replicated chromosomes to daughter cells during both meiotic and mitotic cell divisions. The centromere- kinetochore region acts a point of attachment of chromosomes to single microtubules during mitosis (Peterson *et al.*, 1976) and for replicated sister chromatids during both mitotic prophase-metaphase and throughout the first meiotic division (Murray and Szostak, 1985). The centromere is essentially a constricted region of the chromosome to which spindle fibres attach. Centromeres of the budding yeast *Saccharomyces cerevisiae* and fission yeast *Schizosaccharomyces pombe* vary dramatically: the budding yeast centromeres are comparatively small (approximately 125 bp, with little repetitive DNA), whereas those of *S. pombe* vary in size but cover kilobases of DNA and contain several classes of repeated DNA sequences (Clarke, 1990; Murakami *et al.*, 1991).

The *S. cerevisiae* centromere is defined as a 125 bp minimal length sequence (CEN- Cottarel *et al.*, 1989) consisting of a 78- to 86- bp central core region that is >90% A+T (Centromere DNA Element II, CDE II) flanked on one side by an 8 bp partially conserved CDE I sequence (PuTCACPuTG) and on the other by a 26- bp conserved sequence $TGTTT^{T}_{A}$-TGNTTTCCGAAANNNAAAAA (CDE III). CDE III is bilaterally symmetric around the C- residue at position 14. The centromere has been shown by nuclease protection to be contained in a 220- 250 bp fragment protected from cleavage and is flanked by a highly ordered array of nucleosomes (Bloom and Carbon, 1982). Centromeres are functional in either orientation in chromosomes and are not chromosome specific. (Clarke and Carbon, 1983).

All three centromere elements contribute to centromere function. Deletion of CDE1 has only a minor effect on mitotic segregation of the parent chromosome, but a pronounced effect in meiosis I. CDE II is structurally conserved: high A+T content and length, rather than sequence are critical for CDEII function (Panzeri *et al.*, 1985; Cumberledge *et al* 1987). Single base- pair changes within the region of bilateral symmetry as well as mutations at the boundary of CDE III completely inactivate centromere function (Hegemann *et al.*, 1988). In addition to nucleosome proteins, CEN sequences are

Figure 1.3

Diagrammatic representation of the Artificial Chromosome vector pYAC4. The lower diagram corresponds to the linear plasmid produced after digestion with *Bam*HI to expose the telomeres and *Eco*RI to expose the cloning site (reproduced from Techniques for the Analysis of Complex Genomes, Academic Press Ltd., 1992)


XhoI
BamHI
TEL
HIS3
BamHI
TEL
XhoI
TRP1
ARS1
pYAC4
URA3
CEN4
SUP4
SalI
SmaI
EcoRI
pYAC-4 Digested with
EcoRI/ BamHI
BamHI End
EcoRI Ends
BamHI End
TEL
TRP1 ARS
CEN4
SUP4
SUP4
URA3
TEL
HIS3 (NOT USED
IN CLONING)

known to interact with proteins. Centromere Protein 1 (CP1) binds to CDEI and such binding is essential for mitotic function of the centromere (Baker *et al*, 1989). A functionally involved protein binds to CDE III, as mutants lose binding ability, but the role of the protein within kinetochore functions is unknown (Ng and Carbon, 1987).

In the construction of the YAC plasmid, the *S. cerevisiae* chromosome 4 CEN sequence *CEN4* was used (Mann and Davis, 1986).

1.4.2.2 Telomeres.

Telomeres have two roles that assist chromosome function: (i) stabilisation, as chromosomes with broken ends are known to fuse with each other or to be degraded by nucleases, and (ii) duplication, as they permit complete replication of all the genetic information contained within the chromosome, without the loss of the extreme 5' ends of each strand as predicted by the model for linear DNA replication (Blackburn, 1991). Telomeres from different organisms share similar sequence structures. Essentially, telomeric DNA consists of a stretch of simple sequence repeats, the sequence of which varies slightly between eukaryotes. Examples of telomeric repeats include ($G_{1\text{-}3}T$) and ($G_{1\text{-}8}A$) for *S. cerevisiae*, $AGGGTT^{C}_{T}$ for *Plasmodium* species and AGGGTT for vertebrates (Blackburn, 1990). Telomeres posses a G- rich strand that extends 5'- 3' beyond the complementary C- rich strand. This sequence is added by a Telomerase, a novel reverse transcriptase that uses an RNA template to direct synthesis of the G- rich strand. Telomerase action is not sequence specific, but the enzyme is essential for maintenance of chromosome length and cell viability. Telomerase RNA's have been identified from *Euplotes* and *Tetrahymena* (Shippen- Lentz, 1990b; Greider and Blackburn, 1989). Recently in *S. cerevisiae* a gene coding for the RNA component, TLC1, has been identified. This contains the sequence CACCACACCCACACAC, which when transcribed yields the known yeast telomere sequence and when mutated results in different telomeric repeat sequences being added (Singer and Gottschling, 1994). There is some evidence that the telomere may function as a form of biological "clock", since somatic cells gradually lose telomeric DNA at the rate of 15 - 40 bp per year, and cells with shorter telomeres undergo fewer cell divisions. Germline cells undergo no decrease in telomere length (de Lange, 1994).

In the construction of the YAC plasmid, Burke *et al* (1987) used rDNA sequences from *Tetrahymena*, since when cleaved with restriction endonucleases, the ends can act as substrates for the addition of yeast telomeric repeat sequences and functionally heal *in vivo* (Murray *et al.*, 1986).

1.4.2.3 ARS Sequences

Autonomously Replicating Sequences (ARS) are sequences in eukaryotes required for the initiation of bi-directional replication (Brewer and Fangman, 1987; Benbow *et al*, 1992). Unlike prokaryotes, whose binding of the proteins associated with the DNA replication complex (*dna*A, *dna*B

and *dna*C) is highly specific for the *Ori*C origin of replication, the eukaryotic ARS origin seems to initiate replication in compact areas associated with the sequence rather than the sequence itself (Huberman *et al.*, 1988).

ARS sequences were first identified as a *trp*1 associated yeast DNA fragment that had a 1000-fold higher transformation efficiency that other plasmids, a feature that was discovered to be associated with autonomous replication (Stinchcomb *et al.*, 1979). Analysis of the structure of ARS1 revealed the presence of a conserved core consensus sequence (designated the ACS- AT rich Core Consensus Sequence). The ACS consists of an 11 bp 5'- $^{T}_{A}$TTTA $^{T}_{C}{}^{A}_{G}$ TTT$^{T}_{A}$- 3' element that is essential for ARS function (Mahrens and Stillman, 1992). The ACS is flanked by a A+T rich 3' region, approximately 100 bp in size, designated element 'B'. This region is composed of three elements- B1, B2 and B3, with each individual domain important but not essential for origin activity; however the whole B element is essential for origin function (Mahrens and Stillman, 1992.) Footprinting assays revealed a complex of six polypeptides that bind ARS1 with high specificity, named the Origin Recognition Complex (ORC). The ORC requires a functional ACS for binding, recognises the ACS of other ARS elements and shows strong evidence that it is associated with the ARS throughout most of the cell cycle (Bell and Stillman, 1992). B3 has also been shown to bind the transcriptional activator ABF1, consistent with the fact that origins of replication of a range of different eukaryotes appear to be associated with transcription proteins (DePamphilis, 1988). In addition to evidence that ARS activity is essential for the function of the HMR- E silencer this indicates a strong relationship between the control of DNA replication and transcription (Rivier and Rhine, 1992). In addition, the YAC plasmid possesses a copy of the bacterial *Ori* element, for replication in *E. coli* (Burke and Olson, 1991.)

In the construction of the YAC plasmid, Burke *et al* (1987) use a 5.5 kb fragment containing TRP1 and ARS1 from Yrp17 (Tschumper and Carbon, 1980).

1.4.2.4 Selectable Markers.

In order to select for the presence of recombinant linear YAC's, auxotrophic markers are used to inhibit the growth of untransformed host cells on restrictive media. The markers used are TRP1 and URA3, specific for the left and right arms of the plasmid respectively. TRP1 codes for N(5' -phosphoribosyl) anthranilate isomerase, a protein which catalyses the third step in the tryptophan biosynthetic pathway (Dobson, *et al.*, 1983). URA3 codes for OMP decarboxylase, the final enzyme in the pathway for the biosynthesis of pyrimidines (Bach *et al.*, 1979). The host strain used in this study, AB1380, is *ura3 trp1* i.e. lacks functional copies of the two genes, which must be supplied *in trans* by the YAC in order for growth of the cell in conditions of limiting uracil and tryptophan. In addition, the YAC plasmid possesses a copy of the ampicillin resistance gene for maintenance in *E. coli* (Burke and Olson, 1991.)

1.4.2.5 *Ade2* Colour Assay.

To distinguish between recombinant and non- recombinant YAC plasmids, the YAC vector possesses a system that is a variant of the SUP11 suppression of the *ade2* mutation (Heiter *et al.*, 1985). Cloned, ochre- suppressing forms of a $tRNA^{tyr}$ gene cause a variable colour phenotype depending upon its copy number within the cell. Yeast ade2- 101 mutants are defective in 5AIR carboxylase, an enzyme involved in adenine metabolism, and consequently accumulate a red pigment (phosphoribosyl-aminoimidazole). Cells expressing the ochre- suppressing phenotype are white. Such a colour assay was used to investigate centromere function (Hegemann *et al*, 1988). YAC's use a similar system, involving SUP4 and an ade2- 1_{ochre} mutation (Shaw and Olson, 1984). Consequently, the cloning site in pYAC4 has been engineered within the SUP4 locus, so that insertion of foreign DNA will inactivate the suppresser phenotype. Hence recombinant YAC's appear red when grown in adenine- rich environments while non recombinant YAC's appear white. Untransformed yeast cells are also red. When used in combination with auxotrophy selection, this system provides an accurate means of differentiating recombinant YAC's from cloning artefacts.

1.4.2.6 Alternative *S. cerevisiae* Host Strains

The preferred host strain for primary library construction has been the yeast AB1380, with other strains possessing different auxotrophy markers for manipulations. In addition to being *ade2 ura3 trp1*, allowing for complementation by the SUP4, URA3 and TRP1 vector phenotypes, the yeast is also ψ^+, a cytoplasmic determinant that enhances ochre suppression. However, as will be discussed, one of the problems associated with YAC's is the structural instability of certain cloned sequences and the presence of chimeric YAC's i.e. containing two of more non- contiguous stretches of DNA co- cloned in a single YAC. A hypothesis for these occurrences is that the yeast recombinational apparatus mediates recombination events within tandemly repeated homologous sequences within the cloned genome (Green *et al.*, 1991a). Of the RAD excision- repair genes, only RAD1 and RAD52 appear to be directly involved in recombination (Schiestl and Prakash, 1988). The RAD52 gene is necessary for most recombinational and repair processes including gene conversion, spontaneous mitotic and meiotic recombination and in mediating recombination between co- transformed plasmids and host cells (Adzuma *et al.*, 1984). Mutations were constructed which revealed that *rad52* strains increased the stability of YAC's (Neill *et al.*, 1991). Two new recombination- deficient host strains have been constructed: LIV1 (RAD52::LEU2) and LIV2 (Δ RAD52), both of which are available for future library construction and which demonstrate increased ability to stabilise unstable YAC's at the expense of dramatically lower transformation efficiencies (Vilageliu and Tyler- Smith, 1992). Libraries have been constructed in these strains and will be discussed in section 1.6.4.

1.4.2.7 YAC Plasmids

The first plasmid constructed, pYAC2, contained a *Sma*I site for blunt- ended DNA cloning (Burke *et al.*, 1987). Subsequent vectors include pYAC4 which contains an *Eco*RI site within the SUP4 locus and has become the standard cloning vector in use today for general library construction (Anand *et al.*, 1990; Albertsen *et al.*, 1990; Coulson *et al.*, 1991; Larin *et al.*, 1991; Triglia and Kemp, 1991.): pYAC3 has a *Sna*BI site and pYAC5 a *Not*I site. Further derivatives include (I) pYAC- RC (Marchuck and Collins, 1988), a modification of pYAC3 containing a polylinker sequence within the SUP4 intron with the capacity to clone fragments of DNA digested with infrequently cutting enzymes; (ii) pYACneo (Traver *et al.*, 1989) containing the neomycin resistance gene and the ColE1 origin of replication conferring selection in mammalian cells and (iii) pYACneo- Not, a vector for the cloning of telomeres (Cross *et al.*, 1989). This vector possesses a single telomere and requires the addition of a second functional genomic telomere for the chromosome to be maintained *in vivo*. Shero *et al.*, (1991) describe vectors with a second polylinker and a T7 RNA promoter, for riboprobe construction. Finally Smith *et al.*, (1990) describe a modified vector that permits amplification of artificial chromosomes by using a conditional centromere and selecting for expression of the thymidine kinase gene

1.5. YAC's and Mapping.

1.5.1. Introduction.

The physical mapping strategy of the Human Genome project has depended entirely upon reconstructing entire chromosomes as ordered sets of overlapping YAC's. When considering a genome size of 3 x 10^9 base pairs, 60 000 cosmids (insert capacity ~50 kb) or 30 000 P1 clones (insert capacity ~100 kb) would be required for complete genome coverage, compared to 6000 - 2200 YAC clones (for inserts of 500 kb and 1.4 Mb respectively). The viability of such an approach was first tested in genome analyses with model systems such as *Drosophila melanogaster* (Ajoika *et al.*, 1991) and *Caenorhabditis elegans* (Coulson *et al.*, 1988; Coulson *et al.*, 1991). The ability to reconstruct large genome regions faithfully as artificial chromosome clones has meant that in addition to mammalian genome analysis, the genomes of any organisms of interest can be studied. In addition to the libraries discussed, YAC libraries have been constructed for *Arabidopsis thaliana* (Schmidt and Dean, 1993); *Beta vulgaris* (Eyers *et al.*, 1992); *Myxoccocus xanthus* (Kuspa *et al.*, 1989); *Dictyostelium discoideum* (Kuspa *et al.*, 1992) and *Schistosoma mansoni* (Tanaka *et al.*, 1995).

To construct sets of overlapping clones of human DNA, several multiple- coverage YAC libraries exist: the CEPH Library (Albertsen *et al.*, 1990), the CEPH MegaYAC Library (Chunakov *et al.*, 1992), the ICI Library (Anand *et al.*, 1990), the ICRF library (Larin *et al.*, 1991) and the "classic" St. Louis Library (Brownstein *et al.*, 1989). In order to ensure multiple overlapping clones per marker,

more than one copy of the library is usually screened. Initially such libraries were held as super dense filters for hybridisation screening in individual laboratories. Library copies are currently held, in the UK, at the Human Genome Mapping Project centre in Cambridge. Library screening is carried out at the centre using the markers supplied by individual research teams; positive YAC clones are redistributed as stab cultures and agarose blocks for subsequent manipulations. Such a system will be employed for the *P. falciparum* project.

High resolution physical mapping of megabase size genomes involves breaking down the genome into ordered sets of clones. Overlapping clone sets are called contigs. However, in any mapping strategy, there must be some way of ordering clones identified by landmarks into overlaps and to initiate bi-directional chromosomal walking from the start point. The Polymerase Chain Reaction (Saiki *et al.*, 1988) provides a means to accomplish this through the construction of Sequence Tagged Site (STS) markers (Olson *et al.*, 1989). An STS is a single copy sequence identifiable by PCR; if two clones screen positive for an STS then that DNA sequence is common to both YAC's and thus are overlapping. This 'STS content mapping' has been used in the construction of YAC contigs spanning human chromosome 21 (Chumakov *et al.*, 1992) and the euchromatic region of the human Y chromosome (Foote *et al.*, 1992). The major advantage of STS mapping over other mapping strategies (such as linkage mapping (Donis- Keller *et al.*, 1987) or radiation hybrid mapping (O'Connell *et al.*, 1994)) is that the PCR based approach can form a common 'language' for mapping: markers can exist as PCR data in a database, accessible to any researcher, rather than as radiolabelled probes or RFLP's. Database entries would include not only the PCR primer sequence, but also details for the PCR assay. The removal of radionucleotides from mapping projects also contributes to laboratory safety. In addition, source DNA does not have to be common; contigs compiled in different laboratories using YAC libraries constructed with different sources of insert DNA could be screened with the same STS's (Hamvas *et al.*, 1993). PCR based approaches have several advantages over contemporary hybridisation- based methods of YAC identification. PCR is based on a binary score system that makes identification simpler and the reaction lends itself to automation - an important factor when considering the scope of a genome map project (Huxley *et al.*, 1991; Schlessinger *et al.*, 1991). Current genetic maps are therefore employing PCR-based markers to determine linkage analysis so that such maps can integrate with physical maps and *vice versa* (Murray *et al.*, 1994). This approach has already been used in the construction of two *P. falciparum* YAC contigs (Lanzer *et al.*, 1993; Rubio *et al.*, 1995) and forms the strategy adopted for contig assembly in the work described here.

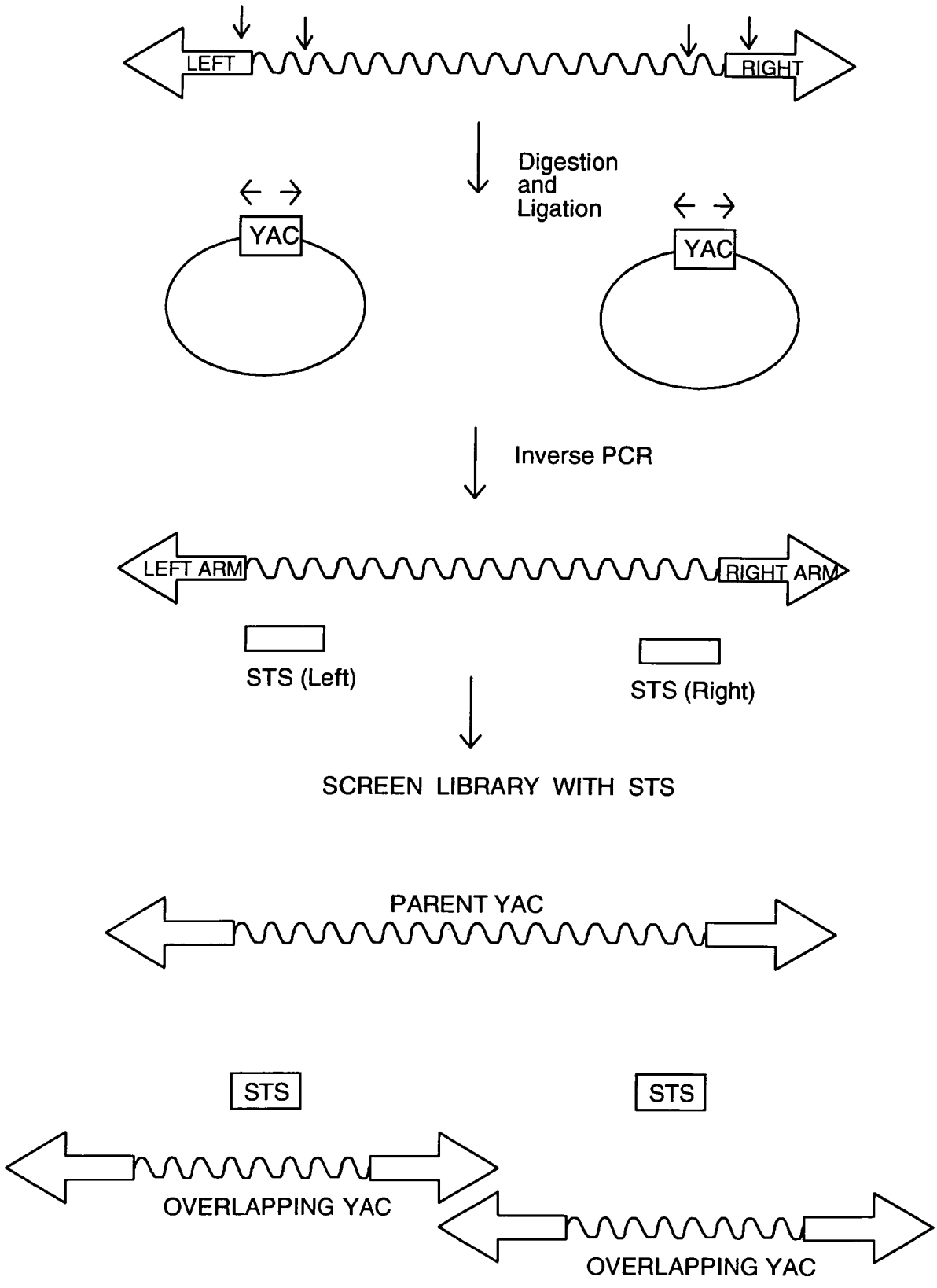
STS content mapping can take two approaches: random or end-clone specific. End- clone mapping involves obtaining a single copy sequence from either one or both ends of the YAC, designing PCR primers based on the sequence data obtained and screening the library with this new STS. Any clones identified by the end- clone STS will be overlapping with the 5' and/ or 3' regions of the YAC, and contig extension continues (Palazzolo *et al.*, 1991; Kere *et al.*, 1992). End rescue techniques have

been used to create many YAC contigs, including the CF locus (Green and Olson, 1990b.) and the human dystrophin gene (Coffey *et al.*, 1992). The principle is illustrated in fig 1.4. The techniques used to rescue the ends of these YAC's to initiate bi-directional walks will be discussed. In essence, for the construction of maps of specific regions, the end rescue strategy is not only a highly accurate method of generating multiple overlapping clones covering a short region of the genome, but also has the advantage of permitting contig coverage of a region where few new YAC's have been recovered by the random STS content approach (Schlessinger *et al.*, 1991).

The random STS approach is to use a defined, limited DNA source, e.g. flow sorted chromosomes (McCormick *et al.*, 1993) or somatic cell hybrids (Lee *et al.*, 1992;), to generate multiple, random, chromosome specific STS's which can then be used to screen a YAC library (Green *et al.*, 1991b; Cole *et al.*, 1991.) This has the advantage in that multiple markers can be generated in areas for which there is little prior STS information. Cole *et al.*, (1992) describe the use of *Alu*- PCR (Nelson *et al.*, 1989) to generate multiple X chromosome- specific STS's from a somatic cell hybrid. *Alu*- PCR is based on the fact that mammalian genomes posses short interspersed repeat DNA sequences (SINES- Singer *et al.*, 1992), of which the major human type is the *Alu* repeat (Jelinek *et al.*, 1982). These repeats consist of two 130 bp tandem repeats and are found ubiquitously in the human DNA.They are believed to be on average approximately 4 kb apart (Britten *et al.*, 1988). PCR primers directed to such repeats should amplify the non- repetitive human DNA within the flanking *Alu* sequences. The amplified product can be labelled directly and used as a probe, or sequence information can be obtained and the product used as an STS (Tagle and Collins, 1992). *Alu*- PCR has been used to generate human chromosome 2 (Liu *et al.*, 1995), 6 (Meese *et al.*, 1992), 7 (Green *et al.*, 1995) and 21 (Chumakov *et al.*, 1992) specific markers .

Another approach was used by Green *et al.*, (1991b), which involved the subcloning of flow sorted chromosome 7 DNA into M13 and bacteriophage λ, and use of the sequence data thus obtained to generate over 100 *de novo* STS markers specific for chromosome 7. A recent protocol has described the construction of M13 libraries prepared from YAC DNA purified from a low melting point PFG (Vaudin *et al.*, 1995). This protocol permits rapid STS construction by omitting the primary cloning step. Recently, SUP- PCR (Hadano *et al.*, 1993) has been described as a method of amplifying YAC DNA using a single unique oligonucleotide as a primer to amplify chromosomal DNA in the absence of any sequence information. The amplified products can be cloned into any *Eco*RI containing vector, thus generating high density markers along the cloned DNA region.

In addition, defined templates can be used to construct chromosome- specific YAC libraries (Lee *et al.*, 1992; McCormick *et al.*, 1993; Green *et al.*, 1995). Such a library will be enriched for a particular chromosome, and so any of the techniques described above could be used for the construction of a complete chromosomal YAC contig and the *de novo* generation of markers. Such libraries have


LEFT
RIGHT
Digestion
and
Ligation
YAC
YAC
Inverse PCR
LEFT ARM
RIGHT ARM
STS (Left)
STS (Right)
SCREEN LIBRARY WITH STS
PARENT YAC
STS
STS
OVERLAPPING YAC
OVERLAPPING YAC

already been constructed for chromosomes 7 (Green *et al.*, 1995), X (Lee *et al.*, 1992), 11 (Qin *et al.*, 1993), 9 (McCormick *et al.*, 1993), 16 and 21 (McCormick *et al.*, 1993).

For general map construction, a combination of both approaches will generally be used to complete a physical map of any region. End rescue clones can be used to extend the contig, whereas repeat sequence PCR or subcloning strategies can be used to initiate contigs, increase marker density within the growing contig or screen for the presence of chimeric inserts within the YACs. Such an approach has been used in the construction of a YAC contig around *P. falciparum* chromosome 2 - overlapping YAC's were identified with chromosome specific markers and gaps were closed by end rescue techniques (Lanzer *et al.*, 1993). The use of defined DNA sources for map construction in *P. falciparum* will be discussed in Chapter 7.

1.5.2 End Clone STS Marker Construction Strategies

One of the key strategies for initiating YAC contig coverage of a defined region is by the isolation of terminal insert DNA at both vector:insert junctions and to use this DNA as an STS probe. Initially, such markers were obtained by digesting YAC DNA, subcloning it into plasmid vectors and subsequently using the appropriate plasmid inserts as hybridisation probes to screen YAC filters for overlapping hybridisation signals (Bronson, *et al.*, 1991; Kozono *et al.*, 1991; Green and Olson, 1991b). However, such methods were labour intensive and time consuming. In response to this, a number of PCR based strategies have been developed to isolate the DNA from YAC ends in a single step and facilitate direct subcloning of the PCR product for sequencing.

1.5.2.1 *Alu*- Vector PCR

The *Alu*- *Alu* PCR strategy has been discussed for the construction of multiple markers internal to the YAC. *Alu*- Vector PCR (Nelson *et al.*, 1989) is a modification of this, in which one set of primers is internal to the YAC vector and the other is specific to an *Alu* Sequence. The end fragment is amplified only in cases where an *Alu* repeat is sufficiently close to the end of a YAC to permit PCR. *Alu*- vector PCR has been used, in conjunction with other techniques, to construct YAC contigs spanning 8 Mb of the human X chromosome (Little *et al.*, 1992), 1.5 Mb across the HLA class II region (Ragoussis *et al.*, 1991) and 2 Mb across the Huntington's disease gene candidate region (Bates *et al.*, 1992; Zuo *et al.*, 1992).

1.5.2.2 Vectorette PCR

An elegant alternative is the vectorette method (Riley *et al.*, 1990). The vectorette is an oligonucleotide cassette which has a non- complementary central core (or "bubble"). YAC DNA is digested and ligated to the cassette, so that in a population of ligated fragments, a proportion will consist of vector:insert:vectorette species. It is this that forms the template for the PCR reaction. A pair of

primers is used, one set specific for the YAC arm, the other specific for the non- complementary bubble of the vectorette. Amplification with the vector primer results in synthesis of a strand containing vector, insert and a sequence of DNA that is now complementary to one strand of the bubble, and it is this that can now be used to prime the PCR reaction with the bubble primer. Consequently, synthesis depends upon a primary round of polymerisation from the vector primer; amplification occurs after this initial step, as polymerisation can then proceed from either the vector or the vectorette. In a similar concept to *Alu*- vector PCR, amplification depends upon the presence of a restriction site close enough to the junction to permit PCR. Vectorette PCR has proven to be a particularly robust technique, having been applied to contig construction across the CF locus (Anand *et al.*, 1991); the D21S13 and D21S16 region on human chromosome 21 (Butler *et al.*, 1992); the Human α5 (IV) collagen gene (COL4A5) (Vetrie *et al.*, 1992) and the human dystrophin gene (Coffey *et al.*, 1992).

1.5.2.3 Inverse PCR (IPCR)

Another method that has been used for the isolation of YAC termini is Inverse- PCR (Ochman *et al.*, 1988; Triglia *et al.*, 1988a). Whereas conventional PCR permits the amplification of DNA segments between regions of known sequence, IPCR facilitates the *in vitro* amplification of DNA flanking a known sequence. Source DNA is digested with restriction enzymes, circularised and amplified with primers whose sequence is synthesised in the opposite orientation to that of conventional PCR. The technique has been modified for YAC's (Silverman *et al.*, 1989) and sets of primers designed which vary depending upon the location of the enzyme cut site both within the YAC vector and that used to digest the insert (Arveiller and Porteous, 1991). IPCR has been used to construct multiple YAC contigs including the fragile X locus (Hirst *et al.*, 1991), one of the major causes of mental retardation in adults; the type 1 Neurofibromatosis gene (Marchuk *et al.*, 1992); the plasminogen activator inhibitor type 2 gene on chromosome 18 (Silverman *et al.*, 1991); the Huntington disease region (Zuo *et al.*, 1992); the region containing the gene for hereditary breast and ovarian cancer, BRCA1 (Albertsen *et al.*, 1994) and in a 1.1 Mb contig across *P. falciparum* chromosome 4 (Rubio *et al.*, 1995). Again, a major limitation of this system is that it relies upon the presence of a restriction site within a region close enough to the junction to permit amplification by PCR.

1.5.2.4 Alternative PCR- based Strategies

In addition to the above methods, there exist a range of novel PCR- based strategies available to researchers for isolating STS's from the termini of YAC's. Junction trapping (Patel *et al.*, 1993) involves the partial digestion of total YAC DNA with *Sau*3A to generate fragments containing the vector- insert junction in a size range that facilitates PCR, cloning of the junction fragments into a vector (pBluescript II KS), and amplification using nested PCR primers (Frohman and Martin, 1989). The first round PCR is carried out using YAC specific primers and the M13 -20 sequencing primer.

Second round PCR uses internal nested primers to reselect desired products. The advantage of this method is that cloning into pBluescript allows direct sequencing of the STS using primers specific to the vector.

1.5.3 YAC's and Human Genome Analysis.

In order to illustrate how YAC's can be used in the mapping of the *P. falciparum* genome, applications to human genome analysis will now be considered. Complete chromosomal physical maps, consisting of overlapping recombinant clones, provide a primary resource for the genetic investigation of any organism. Such maps facilitate correlation of genetic and cytological data, offer immediate access to clones from any region and provide template sources for large scale sequencing. National Institutes of Health goals for the Human Genome project included a 2 - 5 cM genetic map by 1995 and a physical map with a marker resolution of 100 kb by 1998. Initial linkage maps produced by the NIH/ CEPH consortium (1992) and Genethon/ CEPH (Weissenbach *et al.*, 1992) have been subsequently refined (Co- Operative Human Linkage Centre, 1994) to yield a high density genetic map with an average marker density of 0.7 cM. Thus one of the aims of the project has been achieved.

As has been discussed, YAC's positive for a particular marker can be identified if PCR primers exist for that marker (Olson *et al.*, 1989). A first level contig can be constructed by STS content mapping, whereby overlapping YAC's are identified by the presence of a common STS (Foote *et al.*, 1992; Chumakov *et al.*, 1992). Such data demonstrate the presence of gaps in the contig, which can be closed by either (a) screening of an alternative YAC resource with the STS that reveals the gap (Mandel *et al.*, 1992); (b) specific end rescue of the YAC ends 5' and 3' to the gap and walking with the *de novo* generated markers, as used in the contig around *P. falciparum* chromosomes 2 and 4 (Lanzer *et al.*, 1993; Rubio *et al.*, 1993) or (c) using alternative resources such as P1 (Sternberg, 1990), cosmids or Bacterial Artificial Chromosomes (Shizuya *et al.*, 1992) to close the gap (Zuo *et al.*, 1992). Such an approach has resulted in a first generation STS based physical map of the entire genome (Cohen *et al.*, 1993), a contig spanning the entire chromosome 21 (Chumakov *et al.*, 1992) and a 28 Mb contig spanning the Y chromosome euchromatic region (Foote *et al.*, 1992). Such ordering, however, reveals no information about the internal sequence organisation of the YAC clone. A PFG based map approach will reveal STS order within a YAC clone.

A consequence of this STS based approach is the linking of physical and genetic maps, one of the preliminary aims of the project. (Cox *et al.*, 1994). A human linkage map with a marker resolution of 0.7 cM has been constructed and of the 5480 loci present, over 4000 are sequences for which PCR data exist (Cooperative Human Linkage Centre, 1994). These unique STS markers will allow not only correlation between existing physical and genetic maps but may result in the initiation of new physical maps. New markers constructed as discussed can be used to refine and develop new and existing genetic

maps. Similarly the CEPH linkage map for chromosome 16 anchors genetic and physical data with over 90 PCR based markers (Kozman *et al.*, 1995). The discrepancy between the centiMorgan and the kilobase is therefore being resolved.

A great deal of effort has been invested in the automation of many of the procedures mentioned because of the amount of information generated and to facilitate processing of these data. The enormous number of new loci and quantity of genotyopic data generated, in addition to regional physical maps, can only be handled by publicly accessible databases Information currently available in the databases include YAC physical map data, STS loci data and linkage group data. All information is accessible via File Transfer Protocol (FTP) access on the internet or on the relevant homepages on the World Wide Web (WWW). Any new information is added, so that databases are constantly being refined and updated (CHLC: Murray *et al.*, 1994). In addition, YAC manipulations have been simplified by the development of automated workstations, which can be applied to gridding YAC clones into high density filters (Bentley *et al.*, 1992), screening filters by PCR (Sloan *et al.*, 1993) and preparation of YAC DNA from clone collections (MacMurray *et al.*, 1991). The development of automated DNA sequencing (Hunkapiller *et al.*, 1991) has meant that large quantities of sequence data can be generated, thus freeing laboratory time to allow the achievement of more intellectual goals. Automation in the *P. falciparum* Genome Project would greatly increase the rate of YAC clone identification and this is being pursued.

1.6. Problems Associated with YAC Analysis

1.6.1 Introduction.

The analysis of yeast artificial chromosomes is central to the physical mapping of large, complex genomes. In order that (a) such maps can be extended, (b) the information be made available as a resource and (c) genome sequencing and functional analysis can occur, it is necessary that the cloned insert DNA be an accurate representation of the genomic region intended for coverage and that it be present as a single, contiguous stretch of DNA. If the insert contains a segment of non contiguous DNA, then false positive markers will be constructed and time, research facilities and finance will be expended constructing redundant probes. Two of the main sources of concern when initiating any new contig assembly are the presence of unstable YAC's and the presence of chimeric YAC's.

1.6.2 Instability of Cloned DNA Sequences.

Instability can be viewed as any difference between the original DNA sequence and its structure in a YAC clone. Instability can be recognised as multiple bands observed after Southern hybridisation of a PFG filter with a YAC specific probe. Such a banding pattern can arise by two distinct methods: firstly yeast artefacts, such as a yeast cell co-transformed with two or more YAC's (Anand *et al.*, 1991; Zuo *et*

al., 1992; McCormick *et al.*, 1993). Such events can be corrected using standard yeast manipulations, such as meiotic segregation and tetrad analysis to rescue single YAC's (Nemeth *et al.*, 1993).

Alternatively, such a banding pattern may be due to artificial chromosome instability which may be observed as insert deletions, as the YAC may appear to be stable but the insert can no longer be correlated to the genome. Unstable YAC's may be detected as "unclonable" sequences, as the DNA motifs causing the repeats are so unstable that they are deleted at an early stage and appear to be missing from libraries (Bates *et al.*, 1992; Zuo *et al.*, 1992). The use of cosmid contigs to facilitate map closure suggests that regions of the genome difficult to clone in YAC's can be rescued by cloning in other systems (Zuo *et al.*, 1992; Murrell *et al.*, 1995). Since the initial reports of unstable YAC sequences (Albertsen *et al.*, 1990; Wada *et al.*, 1990; Abidi *et al.*, 1990) investigations have revealed several possible mechanisms for instability. One possible cause is recombination between repetitive elements (Neil *et al* 1990) as the introduction of such YAC's into recombination- deficient host strains (Sciestl and Prakash, 1988) appeared to correct instability, although some sequences still appeared unstable (Kouprina *et al.*, 1994).

Another possible source for YAC instability comes from the transformation procedure itself. Stable YAC's demonstrated irreversible structural changes on retransformation (Neil *et al.*, 1990). Similarly analysis of YAC clones of a retransformed random human clone revealed one parental YAC (1050 kb) and four deletion derivatives ranging from ~250 to ~450 kb in size with the derived YAC's being stable (Albertsen *et al.*, 1990). It is more than likely however, only as detailed physical maps are assembled, such discrepancies in YAC clones will become apparent.

1.6.3 Co- Cloning Events

"Chimeric" YAC's are clones containing two or more noncontiguous segments of DNA within the insert and represent the most common artefact found in total genomic YAC libraries. Such YAC's generate spurious map information that complicates the construction of YAC- based physical maps and contigs. Human chimeras can be identified by a comparison of the physical map of the YAC with the genomic region (Anand *et al.*, 1991; Green *et al.*, 1991), by standard hybridisation techniques for libraries derived from heterologous cell lines (Little *et al.* 1992) and by Fluorescent In- Situ Hybridisation, although this can only identify gross YAC chimerism between two or more chromosomes (Selleri *et al.*, 1991; Selleri *et al.*, 1992).

Two possible models have been proposed: firstly, such YAC's result from a coligation event between two non- contiguous fragments of DNA due to production of overlapping ends by enzyme digestion. However, a recent study strongly suggests that chimerism results from an alternate mechanism to coligation (Wada *et al.*, 1994). The alternative hypothesis is that chimeras result from inter- strand homologous recombination. Sequencing a chimera junction revealed the presence of an *Alu* repeat, suggesting that the YAC had been produced as the result of an *in vivo* recombination event

mediated by this repetitive sequence element (Green *et al.* 1991). This is in agreement with the fact that *Alu* repeat sequences appear to induce rearrangements frequently (Britten *et al.*, 1988). In view of the highly recombinogenic nature of free DNA ends in yeast (Orr- Weaver and Szostak, 1983), such a recombination event would be expected to occur efficiently. This is also consistent with the observation that Bacterial Artificial Chromosome (BAC) libraries constructed in *E. coli* have particularly low chimerism frequencies (Shizuya *et al.*, 1992). If coligation were the sole cause of chimerism, comparable rates would have been observed in this bacterial system.

A recent study has pointed to the transformation procedure itself as a possible source for the origin of chimeric YAC's (Larionov *et al.*, 1994). It is proposed that the co- transformation and co-penetration of two YAC's into a single yeast spheroplast form the basis for a recombination event mediated by the yeast, as up to 46% of co- transformed YAC's were chimeric. When the transforming DNA's were heterologous or when a recombination- deficient host strain was used, co- transformation levels were lower than that observed, which suggests links to the low levels of chimerism observed from YAC libraries derived from somatic cell hybrids (McCormick *et al.*, 1993).

Chimera frequencies have been estimated for the currently available YAC libraries used in the mapping of the human genome. The Washington University Library, St. Louis (Brownstein *et al.*, 1989) has been screened and chimerism frequencies of 4 - 60% have been reported (Abidi *et al.*, 1990; Bronson *et al.*, 1991; Marchuk *et al.*, 1992; Selleri *et al.*, 1992). An average chimera frequency of 40% has been obtained (Eubanks *et al.*, 1992; Sliverman *et al.*, 1990; Zuo *et al.*, 1992). The ICRF YAC library (Larin *et al.*, 1991) has been screened and chimerism frequencies of 22% have been reported (Monaco *et al.*, 1992). The CEPH library (Albertsen *et al.*, 1990) has been screened and chimerism frequencies of 50% have been reported (Bates *et al.*, 1992). The CEPH Mega- YAC library, has been shown to contain chimera frequencies of 80% (Anderson, 1993), consistent with the observation that the larger the insert, the greater the probability of chimerism (Schlessinger *et al.*, 1991; Selleri *et al.*, 1992; Wada *et al.*, 1994; Nagaraja *et al.*, 1994). Such observations should serve as a warning to any groups undertaking YAC- based genome mapping projects. However, the Mega- YAC library has been effectively used in the construction of a contig spanning the Wilson's Disease Locus (Petrukhin *et al.*, 1993), so that it is possible that initial concerns have been exaggerated.

1.6.4 Solutions to YAC Instabilities.

YAC instability is apparently dependent upon the yeast recombination apparatus, as *rad1* and *rad52* strains stabilise unstable YAC's (Ling *et al.*, 1993). YAC libraries have been constructed in *rad* host strains for mouse genomic DNA (Chartier *et al.*, 1992) and human genomic DNA (Haldi *et al.*, 1994). Both demonstrate increased clone stability at the expense of reduced transformation efficiencies. If such libraries can ensure accurate map data and in the process minimise time spent investigating and

eliminating chimerism, then an investigation into YAC library construction in recombination- deficient yeast strains would seem to be a priority, not only for HUGO but for any burgeoning Genome Project.

Another option for reducing the percentage of chimerism in any first- level search has been the construction of human YAC libraries derived from specialised cell sources- either flow- sorted human chromosomes or somatic cell hybrids. For libraries constructed from cell hybrids, the majority of the chimerics are interspecies and therefore detectable by hybridisation (Wada *et al.*, 1990; Lee *et al.*, 1992). Such libraries demonstrate chimera frequencies of 10 - 20% indicating that the background of excess heterologous DNA is suppressing homologous recombination (Lee *et al.*, 1992). An alternative has been the use of flow sorted chromosomes as the source of DNA. The nanogram levels of DNA used results in a substantially lower percentage of chimerics when compared to the total genomic libraries (McCormick *et al.*, 1993a.; McCormick *et al.*, 1993b.) However, as the YAC's were constructed from an enriched DNA source, homologous cocloning events would involve fragments from different regions of the chromosome, and may thus remain undetectable by FISH.

In conclusion, for streamlining mapping strategies to maximise resources and minimise time lost to error, an investigation of constructing third generation YAC libraries with (a) *rad52* host strains; (b) flow sorted or cell hybrid DNA as a template to provide minimal amounts of target DNA in a background of excess heterologous DNA; (c) ligation conditions which suppress cocloning; and (d) modified transformation protocols to reduce the possible shearing of DNA would seem to be an invaluable resource in the accurate physical mapping of complex genomes.

1.7. The Malaria Genome Project.

1.7.1 YAC's and *P. falciparum*.

As discussed, the high A+T content of *P. falciparum* results in large DNA regions being unstable in bacteria (Weber et al., 1988a). The YAC cloning system is known to maintain up to 1000 kb of mammalian DNA in a stable form in yeast. Although some sequences do appear to be "unclonable", such areas are thought to be highly repetitive sequences undergoing recombination within the yeast. With the A+T content of *S. cerevisiae* being closer to *P. falciparum* than *E. coli*, initial attempts to clone *P. falciparum* DNA in YAC's showed that the cloned DNA correlated perfectly with the genomic equivalent and that preliminary assessments indicated that clones were stable (Triglia and Kemp, 1991). A four fold library of the parasite strain FCR3 was subsequently constructed and DNA inserts were found to be stable in a continuous culture experiment over 75 cell generations (de Bruin *et al.*, 1991). Consequently, the major limitation to the construction of *P. falciparum* chromosome contigs, that of clone instability, appeared to have been circumvented. With the groundwork for YAC manipulation

having been laid down by Human Genome researchers, chromosome 10 specific YAC's spanning 100 kb were identified and both physically and transcriptionally mapped (Lanzer *et al.*, 1991).

A complete YAC contig spanning 1.03 Mb of chromosome 2 has been reported (Lanzer *et al.*, 1993). YAC clones were identified by PCR using chromosome 2 specific markers, with the remaining gaps being closed by constructing STS markers from the insert termini (method not described) and identifying overlapping clones. As *P. falciparum* chromosomes do not condense, FISH analysis cannot be used to identify chimeras. The only way to identify chimeras is to probe a chromosomal blot. All STS markers constructed mapped to chromosome 2, but this does not reveal the presence of an intra-chromosome chimeric. Comparative restriction mapping with STS markers between YAC and genomic DNA revealed clone integrity; however this cannot reveal the presence of chromosomal rearrangements outwith the area of the STS. A telomere specific YAC library was constructed, based upon previous telomere YAC's (Cross *et al.*, 1989), and successfully demonstrated the cloning of both telomeres. Transcription mapping of the entire chromosome represented by a minimal set of overlapping YAC's, indicated a compartmentalisation of the chromosome in to an active core and silent ends. Telomeric and sub- telomeric YAC's have been used to investigate the structure of the polymorphic subtelomeric regions (de Bruin *et al.*, 1994). The YAC library of the *P. falciparum* clone B8 library has been extended, and clones specific to chromosome 9 have been used to map the break point associated with loss of cytoadherence *in vitro* (Barnes *et al.*, 1994). Finally, both the B8 and FCR3 libraries have been used to define a 1.05 Mb contig spanning chromosome 4, using IPCR to create 12 *de novo* end- clone STS's. The resultant physical map has a resolution of 30 kb using both frequently and infrequently cutting enzymes, with STS markers spaced on average every 50 kb. Again, the question of chimerism remains to be addressed (Rubio *et al.*, 1995)

1.7.2 Aims of the Genome Project

The aim of the *P. falciparum* Genome Project is to generate overlapping YAC contigs of all 14 chromosomes, with concomitant STS generation along each chromosome. Restriction maps at a 20 - 50 kb resolution level are to be constructed. 3D7 was the strain used in the construction of the reference library as it is (i) known to grow well *in vitro*, (ii) it has been used in a genetic cross with clones HB3 and DD2 for which cloned progeny are available, (iii) it is the reference for chromosome numbering on Pulsed field gels and (iv) it has not undergone deletion of any known function (W. T. M. G. C., 1995). The Library has been constructed with *Eco*RI and *Mun*I digested DNA, and ~3600 clones of average insert size 80 kb are available for use. Sequence content of the Library has been estimated by PCR; 50 sequences have tested positive, representing 10 of the 14 chromosomes, however several clones have showed co- transformation events, and based upon the model proposed by Larionov *et al.* (1994), chimeric events might be expected even though the genome apparently lacks highly repetitive dispersed repeats. A screening procedure by which the library is screened at the Resource Centre at the Institute

for Molecular Medicine, Oxford with markers supplied by individual research teams, followed by redistribution of the positive YAC as a stab culture after preliminary assessment has been proposed. It is envisaged that as data accumulates, information will be made available on the Walter and Eliza Hall Institute database, accessible to researchers by FTP access. The project aims for integration of map data by 1996.

1.8. Aims and Objectives of this Thesis.

The aim of this thesis has been to consider the role of genome mapping in elucidating the genetics of *P. falciparum*. The experimental work consisted of the construction of a Yeast Artificial Chromosome library from the *P. falciparum* strain HB3, the assessment of library representation and the evaluation of clone stability at the genomic, chromosomal and nucleotide level. As part of the Malaria Genome Project, this laboratory has also undertaken a collaboration with the Molecular Parasitology Group at the IMM, Oxford to create a detailed physical map of chromosome 6. This project describes the isolation of chromosome 6 specific YAC's from the HB3 library and the generation of end- clone STS markers by IPCR with the subsequent construction of YAC contigs. A detailed physical map of all cloned DNA is presented, with a resolution of 10 kb. In addition, experiments have been undertaken to determine the degree of stage- specific gene transcription within the cloned DNA.

CHAPTER 2.

MATERIALS AND METHODS.

1. Materials

1.1 General Materials

Except where stated, all chemicals were obtained from British Drug House (BDH.) plc., Fisons Scientific Apparatus or Sigma Chemical Co. Ltd., UK and were of Analytical grade. Solvents were obtained from BDH and Sigma and were of either Analytical or Molecular Biology grade. Radioactive nucleotides were supplied by Amersham International plc. The sources of other molecular biological reagents (enzymes etc.,) are quoted in the text.

Centrifugation was carried out either in bench top centrifuges (IEC Centra MP4, Denley BR 104 refrigerated Centrifuges), or microfuges (Hereaus Biofuge 13, IBI IMV- 13). Higher speed centrifugations were performed in a Sorvall RC- 5B refrigerated centrifuge or an ODT- 65 Ultracentrifuge (DuPont Instruments)

1.2 Parasite Sources.

The parasites used in this work were clones 3D7A and HB3. These *P. falciparum* clones are kept in the WHO Registry of Standard Strains of Malaria Parasites in the Centre for Parasite Biology, ICAPB, Edinburgh

1.3. Erythrocytes.

Fresh whole blood, group O Rh+, used specifically for parasite culturing, was obtained from the Edinburgh and South East Scotland Blood Transfusion Service. Whole blood was prepared for use in culturing by pelleting the cells by centrifugation, removing the supernatant, resuspending in a excess volume of incomplete RPMI medium (to remove the citrate anti- coagulant) and spinning for 10 min at 2 000 rpm. After centrifugation the supernatant was removed, including as much of the white blood cell layer as possible to minimise subsequent contamination with human DNA. This procedure was repeated twice after which the red cell pellet was resuspended in a half volume of complete RPMI medium to yield a final haematocrit of 50%. Washed red blood cells were stored at 4° for up to one week.

1.4 Bacterial strains

The following Bacterial strains were used in the course of this study: the Invitrogen INVαF' [F' *end*A1 *rec*A1 *hsd*R17 (r_k^-, k_m^+) *sup*E44 *thi*-1 *gyr*A96 *rel*A1 ϕ80*lac*ZΔM15 Δ(*lac*ZYA- *arg*F)U169 *deo*R λ^-]; and Y1090*hsdR* [*supF hsdR araD*139 Δ*lon* Δ*lac* U169 *rpsL trpC22*::Tn*10*(*tet*^{r}) pMC9] (Young *et al*, 1983).

1.5 Yeast strains.

The Yeast strain used in this study was the AB1380 YAC Host strain (Burke *et al*, 1987) [MATα ψ^+ *ura3 trp1 ade2-1 can1-100 lys2-1 his5*].

1.6 YAC Plasmids

The YAC vectors used in this study were pYAC4 (Burke *et al.*, 1987) and pYAC4- RC (Marchuck and Collins, 1988). The complete sequence of the pYAC4 vector is now available (Kuhn and Ludwig, 1994), and is accessible using GenBank Accession number U01086. The pYAC4 plasmid vector sample used in the actual construction of the HB3 YAC library was a gift from Dr Alister Craig.

1.7 cDNA libraries

Two *P. falciparum* cDNA libraries were screened in the course of this study. These were an asexual cDNA library made from late trophozoites in plasmid pJEF14 (Sultan *et al*, 1994) and a gametocyte cDNA library from clone 3D7 (Alano, *et al* 1991) in λNM 1149.

1.8 Oligonucleotide Primers Used in This Study.

Except where stated, all oligonucleotide primers were synthesised by either the Oswel DNA Service, Department of Chemistry, University of Edinburgh, or Genosys Biotechnologies Inc., Cambridge. Primers supplied were HPLC purified. Primer sequences are given below (+: sense / forward; -: antisense / reverse) and read from 5' to 3'.

1.8.1 Commercial Primers.

λ GT 11 Primers	+ : GGT GGC GAC GAC TCC TGG AGC CCC G
(Young *et al*, 1983.)	- : TTG ACA CCA GAC CAA CTG GTA ATG
M13 sequencing Primers	+ : GTA AAA CGA CGG CCA GT
(New England Biolabs)	- : AAC AGC TAT GAC CAT G
pJFE-14 Primers	+ : TCT AGA GAT CCC TCG ACC T
(Sultan *et al*, 1994)	- : GTA GGT ATG GAA GAT CCC T

1.8.2 YAC Primers.

Inverse PCR Primers	372: GAA TTG ATC CAC AGG ACG GG
(Arveiler and Porteous, 1991)	373: GCC AAG TTG GTT TAA GGC GC
	374: GGA AGA ACG AAG GAA GGA GC
	375: AAA CTC AAC GAG CTG GAC GC
	556: GCC CGA TCT CAA GAT TAC G

YAC Internal Sequencing Primers	left (L Seq) :	CCGAGAACGGGTGCGCATAG
	Right (R seq) :	CCTAGTCCTGTTGCTGCC

YAC Arm Probes	Left (LR1) + :	GTG TGG TCG CCA TGA TCG CG
(Hirst *et al*, 1991)	Left (LP) - :	ATG CGG TAG TTT ATC ACA GTT AA
	Right (RP) + :	GATCATCGTCGCGCTCCAGCGAAAGC
	Right (RR3) - :	CTC GCC ACT TCG GGC TCA

1.9 General Stock Solutions and Media

AHC Medium	1.7g Yeast Nitrogen Base (w/o Amino Acids) 5.0g $(NH_4)_2SO_4$ 10g Acid Hydrolysed Casein 20mg Adenine Hemi- Sulphate make to 1L; pH to 5.8 After autoclaving, add 50 mL sterile 40% (w/ v) glucose
AHC agar	Same recipe as AHC medium add 10 - 20g Bacto- Agar per 1000 mL
Glucose	40g of glucose dissolved per 75 mL, the volume made up to 100 mL then autoclaved
Luria Bertani Medium (LB)	1% Bacto- tryptone 0.5% Bacto- yeast extract 1% NaCL pH to 7.2 with NaOH
LB Agar	LB medium with 1.5 - 2% agar

1.8.3 Gene Sequence Primers.

Table 2.1: gene sequence primers used in this study

Primers listed are read 5' - 3' and are given in the order : upstream (5') and downstream (3'). PCR conditions are quoted in the text

GENE	PRIMER SEQUENCE	REFERENCE
Calmodulin	GGG CTC AGA GTA TTA TTC CAA TGT GCA TGA GGG GGA TGC TTC CAT CTC CAT CTT TAT C	Robson and Jennings (19
GBP 130	GGG CTC AGA GTG TAC ATC TAA AGT TAG G GGG GGA TCC TTA TAT AAA CCT ACA ATT AGC	Lanzer *et al.*, (1992)
MSP 2	GAA GGT AAT TAA AAC ATT GTC GAG TAT AAG GAG AAG TAT G	Ranford- Cartwright *et* (1993)
MSP 1	CAC ATG AAG TTA TCA AGA ACT TGT C GTA CGT CTA ATT CAT TTG CAC G	Ranford- Cartwright *et* (1993)
GLURP	TGA ATT CGA ATA TGT TCA CAC TGA AC Sequence witheld pending publication	Borre *et al.*, (1991)
RAN	GGA GAA TTC GAA AAA AAA TA GGT TTT TCG AAA TTG TAG TT	Sultan *et al.*, (1994)
CS	CCA CAA GTT ACA CTA CAT GG TTA AAG CAA CCA GGG GAT G	Del Portillo *et al.*, (198
SHEBA	ATG AAT ATT CGA AAG TTC TAC ATT TAA GAA TCA TCT CCT TCG TC	Moelans *et al.*, 1991
TRAP	GGA TCC AAA ATA ATG AAT CAT CTT GGG ACCTTTACCACAAGTTACACTACATGGAGACCATTC	Robson *et al.*, (1988)
λ wt 70	CGG TTC TTC AGG AAA CGT G GGG TCC TCA TTT AAT AGG TG	J. Foster, Per communication
MSP 1	AAG AAT TCG TCA AAA AAC CTA GAA GCT TTA G TCG GAT CCA TCA ATT AAA TAT TTG AAA CC	Kimura *et al.*, (1990)

APTase 1	GCT GTA CCA AGA ACA TCT AG CTT CTT CAT CAG GAC TAC TCG	Krishna *et al.*, (1993)
ATPase 2	GGA GAT GTA ATA CCT GTT GTT GG GGT ATT CTG TGG GAC AAT AAT CC	Krishna *et al*, (1994)
ATPase 3	GAA TTC CTA TAT GTG GAA AAA TCC G TAT CTA CAT GAT TTC CAT GC	Krishna *et al.*, (1994)
DNA pol α	ACA AGT GAA CGA GTT GA TAC ATG GTA AAA TCC C	Ridley *et al.*, (1991)
DNA pol δ	CCT GAT TTT CTC ACT GG AGT ACG GTT GCT CCT T	Fox and Bzik (1991)
PCNA	CTC GGA TCC AAG GTT AAT TAA AAA ATG GGA TTT AAT TGT AAC TCC	Kilbey *et al.*, (1993)
Cytochrome b	TCA ACA ATG ACT TTA TTT GG TGC TGT ATC ATA CCC TAA AG	Creasey *et al.*, (1993)
β - Tubulin	CTG ATG AGC ATG GAA TAG ATC CAG CTG TCC CGA CAA GTA CTA CTC	Delves *et al.*, (1989)
Histone 2A	GCC TCA AAG GGA ACT TCA AAT TC ATC TTG ATT GGC AGT ACC AGC TT	Creedon *et al*, 1992
pSC.11.63	ACA TCT ATT AAT AAA ACG ATT ATA TAC TT TGT GAC TGA GGA TAT TGA AAA ATT GAA TGG	J. Foster, *Peronal Communication*
RCC 1	TCT TTT CAT AAT GTT ACA AAT TTT GTA AAT GAC ACC	Ali A. Sultan, PhD Thesis. 1994
Histone 3	GCA AGA ACT AAA CAA ACA GCA AG TGA TCT TTC TCC ACG GAT ACG TC	Longhurst *et al*, 1995

LB top agar	LB medium with 0.7% agar
LB-amp	LB supplemented with 50µg/ mL ampicillin
MilliQ	Double deionised water, autoclaved
RPMI (incomplete)	Add 42 mL Sodium Bicarbonate, 50 mg/ mL gentamycin and 40 mL heat inactivated human serum to 600 mL incomplete RPMI
RPMI (complete)	10.4g RPMI 1640; 5.94g HEPES dissolved in 960 mL dH_20). Filtered through 0.22 mM Nalgene filter, stored at 4°C for four weeks maximum
SOC	2% Bacto- tryptone, 0.5% Bacto- yeast extract 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20 mM glucose
X- gal	Fresh, dry LB- Amp plates, spread with 25 µl X- Gal (40 mg/ mL stock in Dimethyl- formamide)
YPD	10g Yeast Extract 20g Peptone Adjust volume to 1L, pH to 5.8. Autoclave. Add 50 mL of autoclaved 40% (w/ v) glucose

YPD agar — YPD medium with 1.5 - 2% agar.

General Stock Solutions.

PBS — 10X Stock Solution

80g NaCl, 2g KCl, 11.5g Na_2 HPO_4.7 H_2O), 2g KH_2PO_4, dH_20 to 1L, pH to 7.4 Working Solution: (pH 7.4) 137 mM NaCl, 2.7mM KCl, 4.3 mM $Na_2HPO_4.7H_2O$ 1.4 mM KH_2PO_4

SDS — 10X Stock Solution

100g was dissolved in 900 mL H2O, heated to 68°C and the volume adjusted to 1000 mL

Sorbitol — 2M Stock Solution

Add 376.4 g Sorbitol to 800 mL dH20. When dissolved, adjust the volume to 1L and Autoclave.

SSC — 20X Stock Solution

175g NaCl, 88g Na_3 Citrate.$2H_2O$, adjusted to pH 7.0 and made up to 1L with dH_2O. 2X Working Solution: 0.3M NaCl, 0.03M Na_3Citrate.$2H_2O$

TAE — 50X Stock Solution

242g Tris Base, 37.2g Na2 EDTA.2H20, 57.1 mL Glacial Acetic Acid, adjust volume to 1L. Working Solution: Dilute to a 1X solution (40mM Tris Acetate, 2 mM EDTA)

TBE — 10X Stock Solution

108g Tris Base, 55g Boric Acid, 40mL 0.5M EDTA pH 8.0., adjust to 1L with dH_20. Working Solution: dilute to 1X (89 mM Tris base, 89 mM Boric Acid, 2 mM EDTA)

TE — 10X Stock Solution

100 mM Tris- HCl pH 8.0, 10 mMEDTA pH 8.0, Working Solution: dilute 10X 10 mM Tris HCl, 1 mM ETDA

TE^{50} — 1X Stock Solution

10 mM Tris- HCl pH 8.0, 50 mM EDTA pH 8.0

2. Methods.

2.1 Construction of the YAC library.

2.1.1 Preparation of *P. falciparum* DNA.

2.1.1.1 Culturing of Parasites.

Parasite culturing methods used were a modification of those of Trager and Jensen (1976) and Zolg *et al.*, (1982) and are essentially those described by Walliker *et al.* (1987). Parasites stored in liquid Nitrogen were thawed and resuspended in 5 ml of complete RPMI medium containing washed RBC, to initiate culturing at 5% haematocrit in sterile 25ml culture flasks. Flasks were gassed with a mixture of 1% O_2, 3% CO_2 and 96% N_2 at 37°C without disturbing the RBC layer. Media was changed daily (Haynes *et al.*, 1976), again with minimal interruption to the sedimented blood cells. For larger cultures of 15ml volume, parasitised RBC's were added to a final parasitaemia of 1%, with the diluting volume made up of washed blood at 5% haematocrit and complete RPMI medium. Cultures were maintained until a suitable parasitaemia was obtained. Parasitaemia was determined by removing a small amount of blood in a clean, sterile Pasteur pipette, smearing onto a slide, fixing with alcohol and flooding with Giemsa's stain for 30 - 45 minutes. The slide was washed with water and examined microscopically to determine the ratio of parasitised to unparasitised cells. To maintain cultures indefinitely, fresh RBC's were used to dilute the culture every 2 - 3 days, to ensure that the parasites were kept in a healthy state.

2.1.1.2 Harvesting of Parasites.

When culturing multiple flasks of parasites, one culture can lag another in growth efficiency. To harvest parasites at the same parasitaemia, it is advantageous to harvest parasites without extracting the DNA. This is accomplished in essence as described by Arnot *et al* (1993) and Babiker *et al* (1991). Parasitised RBC's were pelleted by centrifugation at 2 000*g* for 10 min at 4°C. Parasites were released from the blood cell envelope by lysis with a five fold excess of 0.1% saponin in 1X PBS. After incubation at room temperature for 10 minutes, the parasites were pelleted by centrifugation at 4 000*g* for 15 min and the supernatant, containing haemoglobin and red cell ghosts, was carefully aspirated off. Parasites could be stored at -70° indefinitely.

2.1.1.3 Preparation of Intact Chromosomes in Agarose Blocks.

Parasite DNA was prepared essentially as described in Babiker *et al.* (1991). Frozen parasites, harvested at 10% parasitaemia, were allowed to thaw and resuspended in 2 volumes of 1X PBS at room temperature. The suspension was fixed by the addition of 2% LMP agarose in 1X PBS at 45°C to a final

agarose concentration of 1.2%. The mixture was divided into 200 µl aliquots and pipetted into the wells of a 1 x 6 x 10 mm eight well mould. The mould was placed at 4°C for 30 min to allow the molten agarose to gel. Once set, the blocks were transferred to a solution of 1% sarkosyl, 0.5*M* EDTA (pH 8.0) for 60 min at 50°C, to allow diffusion of haemoglobin. The blocks were subsequently transferred to a second 1% sarkosyl, 0.5*M* EDTA solution containing 0.25 mg/ml proteinase K (Boehringer) and left to digest overnight at 50°C. The buffer was replaced with fresh sarkosyl/ EDTA/ proteinase K and left overnight at the same temperature. In order to prepare the parasite DNA for subsequent manipulation, e.g. restriction enzyme digestion, the proteinase K still present must be inactivated. This is done essentially as described in Nguyen *et al.* (1990). Blocks are rinsed 3 times in TE to remove the EDTA/sarkosyl and are treated twice for 30 min with 0.04 mg/ml (final concentration) PMSF (freshly dissolved at 40.0 mg/ml in isopropanol) in TE at 50°C. After this last rinse, blocks can be stored in 0.5*M* EDTA at 4°C for long term storage, TE^{50} for short term storage, or directly equilibrated in the buffer required for enzymatic digestion.

2.1.1.4 Preparation of Liquid *P. falciparum* DNA.

The method used was essentially that described by Arnot *et al* (1993). The frozen parasite pellet was allowed to thaw and resuspended in 1 ml of ice cold 1X PBS. The parasite suspension was centrifuged for 10 mins at 12 000 rpm. The supernatant was removed and the pellet resuspended in 600 µl of: 10m*M* Tris- HCl, pH 7.6; 50 m*M*EDTA pH 8.0; 0.1% SDS; 1 mg/ml proteinase K. The resulting viscous lysate was homogenised by 30 passages through a 25 gauge needle and incubated overnight at 50°C. The lysate was extracted 3 times with phenol and then once with chloroform. Extractions were performed by adding an equal volume of phenol to the buffer and vortexing. The resultant colloidal suspension was centrifuged for 15 min at 13 000 rpm, and the upper aqueous layer was carefully removed without drawing up any of the sediment at the layer interface. This supernatant was used for subsequent re- extraction with chloroform to remove traces of phenol. The DNA was precipitated by adding NH_4Ac to 1*M* and an equal volume of isopropanol, mixing thoroughly and leaving at room temperature for 15 minutes. DNA was then pelleted by centrifugation for 5 mins at 10 000 rpm, washed in 70% ethanol, dried in a heating block at 70°C for 10 min and resuspended in 100 µl TE. The DNA in this form could be stored indefinitely at -20°C or -70°C.

2.1.1.5 PFG diagnostics of Intact Chromosomes.

Pulsed field gel operations were performed using a Bio- Rad CHEF DR- II. Refrigeration of the tank was carried out by means of a Grant coolant unit linked to a water bath containing a 1/5 dilution of radiator Anti - freeze. In order to prepare the DNA cell for electrophoresis, the residual buffer in the tank was washed out with 2L of distilled water, and the tank drained. Enough 0.5X TBE buffer to just cover the gel (corresponding to approximately 1.8L, but dependent on the final thickness of the gel) was

poured into the tank and allowed to chill to 5 - 10°C. This is to ensure that the operating temperature of the tank is constant when the gel is loaded and the Power Unit switched on. Chromosome separations were carried out in either 1% Seachem Agarose (IBI) or 1% Rapid PFG agarose (BRL). The latter is used for the separation of larger chromosomes as it reduces the run time and voltage setting. Agarose was added at the final concentration desired to 300 ml of 0.5X TBE buffer and microwaved for 3 minutes The semi- molten agarose was mixed and then re- microwaved for a further three minutes, or until the agarose has melted. By separating the melting process into two separate stages, it reduces the accumulation of superheated agarose, which causes the liquid to boil over. The molten agarose was allowed to cool to 45 - 50°C, and then it was poured into a mold with a compatible 10- well comb. Combs and moulds could be varied to increase the number of samples loaded, with a maximum possible load of 30 samples. Once the agarose had set, 2 mm slices of parasite DNA were cut using a clean, sterile scalpel blade and loaded into the gel wells which were then sealed with LMT agarose. The gel was then loaded into the tank. A buffer circulation flow rate was selected so that the movement of the buffer surface was barely detectable. For the purposes of this study it was not necessary to separate out the larger chromosomes, and to determine the quality of a particular parasite DNA preparation, a 24 hour run using moderately high switch times could be employed. Markers used were intact yeast chromosomes supplied as PFG markers (New England Biolabs). Run conditions used for this purpose were :

Run Time :	24 Hr.
Voltage :	180 Volts
Switch Time :	120 sec. Fixed
Buffer Temp. :	6°C
Re- Orientation Angle :	90°

Upon run completion, the gel is transferred to 500 ml of 0.5X TBE buffer, 50µl of 10 mg/ ml EtBr is added and the gel allowed to shake gently for 10 - 20 minutes. The buffer is removed and the gel is allowed to destain in excess of distilled water for 30 - 60 minutes. Alternatively a longer staining time can be used and the gel left to destain overnight. The gels were then photographed under UV using either a Gel Documentation System (Video copy Processor, Mitsubishi™) or using a Polaroid Camera set-up and 667 or 665 professional film, with an f5.6 / f8 aperture and 1 - 4 sec. exposure.

2.1.2 Preparation of YAC Vector DNA.

2.1.2.1Culturing of Bacteria.

A single bacterial colony containing the YAC plasmid pYAC4, cultured under Ampicillin selection, was inoculated into 50 mls of LB- amp and incubated with shaking at 37°C for 16 - 18 hours. The following day, 5 ml of overnight culture was added to 450 ml of prewarmed Terrific Broth in a 2L

sterile Erlenmeyer flask, and incubated with shaking for 2.5 hrs. 2.2 ml of Chloramphenicol (from a 34 mg/ml stock freshly dissolved in ethanol) was added, to a final concentration of 170 µg/ ml. The culture was incubated overnight with shaking at 37°C .

2.1.2.2 YAC Plasmid DNA Extraction.

The bacterial cells were harvested by centrifugation at 4000 rpm for 15 min in a sorvall GSA rotor. The supernatant was carefully aspirated away, the pellet resuspended in 100 mls of ice cold STE (0.1*M* NaCl, 10 m*M* Tris- HCL pH 8.0, 1m*M* EDTA pH 8.0), and the centrifugation repeated. The washed pellet was resuspended in 10 ml of sterile Lysis Solution I (50 mM glucose, 25 m*M* Tris - HCl, pH 8.0, 10 m*M* EDTA pH 8.0). 1 mL of a freshly prepared solution of lysozyme (10 mg/ml in 10 m*M* Tris- HCl pH 8.0) was added and the suspension incubated for 5 minutes at room temperature. 20 ml of fresh Lysis Solution 2 (0.2N NaOH, 1% SDS) was added and the suspension incubated at room temperature for 5 - 10 minutes. 15 ml of ice cold Lysis Solution 3 (60 ml 5*M* KAc, 28.5 ml dH_20, 11.5 ml Glacial Acetic Acid) was added to the lysis suspension and mixed. The mixture was stored on ice for 10 minutes. The lysate was centrifuged at 4 000 rpm for 15 min at 4°C in a Sorval GSA rotor and the supernatant filtered through four layers of cheesecloth into a 250 ml centrifuge bottle. 0.6 volumes of isopropanol were added, mixed and incubated at room temperature for 10 minutes. The precipitated nucleic acids were recovered by centrifugation at 5 000 rpm for 15 minutes.The supernatant was carefully aspirated away, and the pellet washed with 70% ethanol, allowed to dry and dissolved in 3 ml of TE.

2.1.2.3 Caesium Chloride Gradient Purification of YAC DNA.

To ensure the purity of the plasmid sample, DNA was purified by CsCl gradient equilibrium centrifugation. 1 g of solid CsCl was added per mL of DNA, and the solution warmed to 30°C to aid dissolution. When the salt had dissolved, 0.8 mL of a 10 mg/ mL solution of EtBr was added and mixed. The solution was centrifuged at 8 000 rpm in a sorvall SS34 rotor, and the lower aqueous layer transferred to a clean ultracentrifuge tube. The tube was filled with light paraffin oil and sealed. The DNA was centrifuged in an ODT 65 Ultracentrifuge at 45 000 rpm for 36 hours. Two bands were visible under UV, with the lower band being the desired fraction. The DNA was recovered by first removing the chromosomal DNA to minimise chances of contamination. This was accomplished by immobilising the tube in a clamp/stand, pricking the tube with a needle to break the vacuum seal and inserting a hypodermic needle just below the target band, visualised under UV, and gently extracting. The contaminant DNA was discarded, and the procedure repeated for YAC DNA. Once all samples had been collected, EtBr was removed as described. Cs Cl was removed by dialysis over 48 hours with 2x changes of 3L of 1X TE. Plasmid DNA was analysed both electrophoretically and spectroscopically.

2.1.3 Preparation of Ligation DNA.

The methods used for the preparation of components required for the construction of a YAC library are a modification of those described by Monaco *et al* (1993).

2.1.3.1 Partial Digestion of Intact Chromosomes.

Intact HB3 chromosomes, prepared in agarose as described in section **2.1.1.3**, were washed with five changes of 1X TE on ice for 30 minutes to remove traces of 0.5*M* EDTA. 1ml of partial digest buffer was made up as follows:

714 µl	MilliQ H_20
100 µl	10X *Eco*RI buffer (Boehringer SureCut buffer H)
50 µl	BSA (100X stock, New England Biolabs)
10 µl	1*M* $MgCl_2$
26 µl	Spermidine (0.1M)

190 µl of buffer was added to each tube, assuming that the volume of agarose used would make up the reaction volume to 200 µl. The amount of DNA added per reaction was the equivalent of half a 2 mm slice. Enzyme levels corresponding to 10, 5, 1, 0.5, 0.2 and 0.1 units of *Eco*RI (Pharmacia) were added to the digestion mix. The reaction was left on ice for 1 hr to allow the enzyme to thoroughly diffuse into the block. The reaction was incubated at 37°C for 2 hr, with occasional gentle mixing. The reaction was stopped by the addition of 8 µl of 0.5m EDTA. Digestion products were equilibrated in a bath of 0.5X TBE prior to analysis on a 1% Rapid agarose PFG, with a 20 - 40 sec. ramp at 150V over 22 hours. The PFG mid range Marker II (New England Biolabs) was used which yields a range of marker sizes between 25 - 200 kb

2.1.3.2 Restriction Digestion of YAC Plasmid.

100 µg of pYAC4 DNA at a concentration of 0.5µg/ µl was digested to completion with *Bam*HI (to expose the telomeres) and *Eco*RI (to open the cloning site) in a 350 µl reaction with 250U *Eco*RI (Pharmacia) and 200U *Bam*HI (Pharmacia) with 35 µl of 10X One-Phor-All universal enzyme buffer (Pharmacia). The reaction was incubated for 90 minutes at 37°C. The reaction was heated at 68°C for 10 minutes to inactivate the enzymes.

2.1.3.3 Phosphatase Treatment of Digested YAC Plasmids

The digested YAC DNA is phosphatase treated to prevent the religation of the vector arms, which would reduce transformation efficiency. Calf Intestinal Alkaline Phosphatase (Beohringer), which removes 5` phosphate groups, is added directly to the reaction tube at 0.06U/ µg vector. The reaction is incubated at 37°C for 30 minutes. The DNA is extracted once with phenol/ chloroform/ isoamylalcohol (PCI), and the organic layer re- extracted with TE, to ensure recovery of all the dephosphorylated

plasmid. DNA is precipitated with NaAc to 1*M* and 2 volumes of ethanol, washed, dried and resuspended in 50 µl of TE. To test the efficiency of the dephosphorylation reaction and the ability of the vector ends to ligate, two 20 µl reactions were set up containing 1U T4 Ligase, 2 µl 10X Ligase buffer, 0.5 µg CIP- treated vector DNA. 1U of T4 polynucleotide kinase was added to one tube only, and the results analysed by conventional agarose electrophoresis.

2.1.4 Ligation Reactions.

2.1.4.1 Ligation of Vector DNA to Insert DNA

From the results of the partial digest, the optimum enzyme conditions required to yield fragments of the correct size could be calculated. The next stage is to perform a bulk digest. It was found that two parasite DNA blocks, containing approximately 10 µg of DNA per block, yield enough digested DNA for two transformation reactions. Blocks were washed out 3 x 5 ml of 1X TE, with 2 washes carried out on ice and the last wash at 30°C. Blocks were then sliced into 3mm slivers, and two slivers, containing approximately 2.5 µg of *P. falciparum* DNA per sliver, were placed into sterile eppindorfs. 760 µl of *Eco*RI digest buffer plus enzyme was added, and the reaction incubated for 2 hr at 37°C. EDTA was added to 4 µl per 100 µl buffer and the reaction inactivated by incubating at 37°C for 30 minutes The blocks were washed over 1 hr on ice with three changes of TE[50] buffer. A final wash was carried out for 30 min at 37°C. The blocks were then washed 4 x 30 min with 1X ligation buffer (0.5*M* Tris- HCl pH 7.5; 0,1*M* $MgCl_2$, 0.3*M* NaCl) containing 1X polyamines (diluted from 100x stock: 0.075*M* spermidine-$(HCl)_3$, 0.03*M* spermine- $(HCl)_4$). For the ligation to plasmid DNA the following mix was made:

300 µg digested blocks
360 µl H_2O
40 µl 10X Ligation Buffer (NEB)
15 µl pYAC RI/ BamHI/ Phosphatase treated

The reaction was pre- incubated at 37°C for 45 minutes. To the reaction mix the following were added:

7 µl 0.1*M* ATP
7 µl 0.1*M* DTT
2 µl Ligase (400U/ µl New England Biolabs)

The reaction was incubated for 30 min at 37°C, then overnight at room temperature. The ATP and DTT were added to a final concentration of 0.001*M*.

2.1.4.2 PFG Size fractionation of Ligated Products.

A 1.2% LMP agarose gel was cast and allowed to set. Of the ten wells, the two outer lanes were reserved for markers, the next inner two had a small slice of digested and ligated DNA loaded to serve as an indication of digestion, and the central core of six lanes were completely loaded with all available ligated DNA. Any ligation product lefts over were stored in 0.5*M* EDTA for subsequent re-use. The ligated DNA was separated by a 160 V run with 20 seconds pulse time over 16 hr.

2.1.4.3 Isolation of Size- Selected Ligation Products

On completion of electrophoresis, the central core was excised using a sterile scalpel blade and stored in 0.5X TBE buffer. The flanking regions were stained with ethidium bromide and the results visualised under UV. Under UV, marker lanes were notched with a scalpel at the appropriate DNA sizes, as determined by the markers. The marker lanes were then placed adjacent to the central, unstained core, and the notches were used to facilitate accurate excision of recombinant, size- selected YAC DNA. For the purposes of this study, two bands were excised, corresponding to a size interval of 75 - 125 kb, and 125 - 175 kb. The slices were stored in 50 ml of TE^{50} at 4°C for up to one week.

2.1.5 Transformation.

The transformation reaction is carried out essentially as described in Burgers and Percival, using lyticase (Sigma) to spheroplast cells from the *S. cerevisiae* strain AB1380.

2.1.5.1 Culturing of Host Cells.

From a frozen glycerol stock, AB1380 host cells were streaked to single colonies on a fresh YPD plate, and incubated for 2 - 3 days at 30°C. A single colony was used to inoculate 10 ml of YPD in a sterile 50 ml falcon tube and incubated for 24 hr with angled shaking at 120 rpm. Angled shaking when culturing in these vessels is essential to ensure that the broth is vigorously aerated. If there is little aeration, poor culture growth is seen. 200 µl of cells was diluted into 200 mls sterile YPD and the culture incubated overnight at 30°C. On reaching a specific OD_{600} nm (determined by diluting the culture 1/10 in YPD, referencing against YPD), the culture is aliquotted using sterile technique into 50 ml Falcon tubes. The recommended OD for optimal transformation is between 0.1 and 0.2 (Monaco et al., 1993), however I have found that the OD can be as high as 0. 286 with little loss of transformation efficiency. The YPD plate used for the initial inoculation must be less than a week old. If the plate is older than a week, a loss in transformation efficiency (as determined by controls) is seen. If a transformation is planned, it is advantageous to use freshly growing AB1380, if possible culturing directly from frozen stocks.

2.1.5.2 Transformation of Host Cells with Ligation Products

Prior to transformation, agarose blocks containing recombinant YAC DNA were cut into 1g slices. One slice from each size band was used in the transformation. The blocks were rinsed overnight in 15 mL of Pre- Transformation buffer (0.1*M* Tris pH 7.5, 0.03*M* NaCl, 0.001*M* EDTA, 1X polyamines.) Plugs not used were stored in TE^{50} at 4°C. The buffer was changed twice prior to the transformation. The presence of polyamines throughout the transformation protocol has been reported to eliminate a size bias against the transformation of YAC's greater than 300 kb (Albertsen *et al.*, 1990; Connelly *et al.*, 1991). However, since the average insert size of the transforming DNA is 100kb, the presence of polyamines is to protect DNA in agarose from degradation during the melting step (Larin *et al.*, 1991). The buffer was removed and the plug placed in a sterile eppindorf. The plug was melted at 68°C for 10 - 15 minutes, or until the agarose had completely melted. The eppindorf tube was transferred to a 37°C waterbath for 15 - 25 minutes, to allow the agarose to cool to digestion temperature. At this point, liquid YAC DNA is susceptible to shear damage. 150U of Agarase (Sigma) at 1U/ µl was gently stirred in to the mix. Pipetting up and down was avoided to minimise shear damage. The reaction was incubated for >4 hr at 37°C.

Host cells at the correct OD_{600} were harvested by centrifugation at 3 000 rpm for 5 - 10 minutes on a bench top centrifuge. The medium was decanted, and the cells washed in 20 ml of sterile H_2O, The centrifugation was repeated and the cell pellet resuspended in 20 ml of sterile 1.0*M* sorbitol. Centrifugation was again repeated, and the pellet resuspended in 20 ml SCE (1.0 *M* sorbitol; 0.1*M* sodium citrate, pH treated to 5.8 with citric acid; 0.01*M* EDTA pH 7.5). 46 µl of β- mercaptoethanol was added and 300 µl of cells were removed as a prelyticase control. 400U of Lyticase (Sigma) was added to the tube, mixed gently and the reaction incubated at 30°C. The degree of spheroplast formation was tested at 5 minute intervals by one of two independent methods:

a. Using a bench top spectrophotometer, the OD600 nm of a 1/10 dilution of the spheroplast suspension in distilled water is measured. with the prelyticase aliquot as reference zero. Where the spheroplast value is 1/10 the prelyticase value, spheroplast formation is 90% complete.

b. 10 µl of cells were mixed with 10 µl of 2% SDS and examined microscopically. As the lyticase digests the yeast cell wall, the cell becomes vulnerable to lysis with detergent. Under phase contrast, spheroplasts appear as dark rings ("ghosts"). I have found than this is a much more reliable and accurate method of determining the degree of spheroplasting than spectrophotometry.

Spheroplast formation was taken to 80 - 90% which takes between 10 - 20 minutes. Cells were centrifuged gently at 1 100 rpm on a bench top centrifuge without refrigeration. The SCE was decanted and the cells gently resuspended in 20 ml of 1.0M sorbitol, to maintain spheroplast integrity. Cells were spun again at 1 100 rpm without refrigeration and the pellet resuspended in 20 ml STC (1.0M sorbitol; 0.01M Tris- HCl, pH 7.5; 0.01M $CaCl_2$.). A cell count was taken by making a 1/50 dilution of cells and examining microscopically on a haemocytometer (Advanced Neubeyer). Cells were spun again and then

resuspended in a volume of STC calculated for a final concentration of 4.0 - 6.0 x 10^8 cells/ mL. All subsequent manipulations are carried out in a clean, sterile containment hood. Cells were divided into 150 µl aliquots in sterile 15 ml falcon tubes, and 75 µl of agarase treated DNA was added. DNA and spheroplasts were allowed to interact at room temperature for 10 minutes. 1.5 mL of PEG (20% Polyethylene glycol 6000; 0.01*M* Tris- HCl pH 7.5; 0.01*M* $CaCl_2$; made fresh and filter sterilised) was added and the suspension mixed by gentle inversion. The reaction was left at room temperature for 10 min. The spheroplasts were again centrifuged at 1 100 rpm for 8 minutes at 20°C. The PEG was carefully pipetted away without disturbing the transformed spheroplasts, and the pellet was gently resuspended in 225 µl of SOS (1.0*M* Sorbitol; 25% YPD; 0.0065*M* $CaCl_2$; 10 µg/ ml tryptophan; 1 µg / ml uracil. SOS must be made fresh and filter sterilised.) The mixture was incubated at 30°C for 30 minutes.

While this was being carried out, top regeneration agar was kept molten at 45°C. Regeneration top agar is the same composition as regeneration plates, except at an agar concentration of 7g/ L instead of 15 - 20 g/L. Regeneration media is composed of 1.0M sorbitol, 2% glucose, 1X amino acids, agar as stated. This mix can be autoclaved. Yeast Nitrogen Base w/o amino acids (Difco) is made to a 6.7% 10X stock and tryptophan to a 1 mg/ ml solution Both are sterile filtered, and added to the media after autoclaving, to a final concentration of 0.67% and 20 µg/ ml respectively. 10X amino acid stock solution are made as follows: 300 mg Tyrosine- HCl; 225 mg Adenine- HCl; 225 mg Arginine- HCl; 225 mg Histidine (mono- HCl : mono H_20); 225 mg Methionine; 300 mg Lysine (mono- HCl); 525 mg Phenylalanine; 600 mg Leucine; 1600 mg valine; 300 mg isoleucine.

Aliquots of cells and SOS from each size fraction are pooled into 50 mL Falcon tubes, and 50 mL of (-uracil) regeneration top agar is added, inverted gently to mix and poured onto the surface of pre-warmed 22 x 22 cm (- uracil) rgeneration plates. Once the top agar had set, the plates were incubated upside down for 2 - 3 days at 30°C. As a control of transformation efficiency, 10 ng of YCp50 was included in a separate reaction.

Notes: In addition to the general comments available for carrying out the transformation reaction (Monaco et al., 1993), several points must be stressed in order to ensure that subsequent transformations are carried out smoothly.

1. Agarase is stored at 4°C. When adding agarase to the liquid DNA, ensure that the pipette tip is at room temperature, as adding agarase to the DNA direct from cold will cause the resolidification of agarose. Transforming DNA will remain trapped in agarose beads and will not be incorporated into the cell.

2. Time for optimal spheroplasting will depend on the initial starting OD. If the initial OD is very low, then the time needed to reach optimal spheroplast concentrations will be correspondingly short. Thus SDS/ spectrophotometry time points must be adjusted accordingly.

3. The critical step in the procedure is spheroplasting. It is sometimes advantageous to take reading every two minutes. This can be determined by judging the rate of spheroplasting after 5 or 10 mins. It is useful to have microscope, spectrophotometer and incubator all to hand. Sometimes it is also best to harvest at approximately 75 - 80% spheroplast formation, as the enzyme reaction will still continue while the cells are spinning, and will only cease once the cells have been washed.

4. After spheroplasting, cells are fragile and must be treated delicately. Avoid pipetting up and down, and gently stroke the side of each tube to mix/ resuspend the cells.

2.1.6. YAC Library Preparation.

2.1.6.1 Picking of Primary Transformants.

After transformation, suspected positives must be picked from (- uracil) selection plates onto (- uracil, -tryptophan) double selection plates, which provide maximum stringency for selection of both left and right vector arms. (-U -T) plates are prepared as for (-U) plates, except that tryptophan is not added after autoclaving. Colonies are embedded in top agarose, so lifting is not easily accomplished. The simplest way to do this is in a sterile yeast hood using sterile yellow Gilson tips, and to attempt to scoop out the colony. Once a colony is picked and streaked, the tip is discarded. The suspected recombinant YAC is patch streaked onto -U -T selection plates, and the plates are incubated upside down for 2 - 3 days at 30°C

2.1.6.2 Picking of Secondary Transformants.

Colonies seen after this second round of selection suggest transformants containing a recombinant artificial chromosome. Using a FinnPipette 8- Channel Multipipettor, 150 µl of YPD is pipetted into the wells of an 8 x 12 microtitre dish. Individual YAC's are picked into individual wells with either sterile wooden toothpicks, or sterile yellow Gilson tips. The plates are grown at 30°C for 2 - 3 days.

2.1.6.3 YAC Library Duplication.

In order to create a master and backup copy of the library, the contents of each plate are duplicated. This is done by using a multichannel pipettor to mix 150 µl of YPD containing 40% glycerol into each well, so that the final glycerol concentration is 20%. After mixing, 150 µl of cells in YPD/ glycerol is transferred to a clean, sterile 96- well microtitre dish. Plates are sealed with Plate Sealers (TitreTek), snap frozen on dry ice and stored at -80°C. Libraries can be stored in this form indefinitely.

2.1.6.4 YAC Plate Culturing .

In order to grow the YAC library for subsequent manipulation, plates are allowed to thaw to room temperature. All manipulations were carried out in a sterile containment hood. Replication was carried out using a sterile, 96- prong replicator which is dipped into the thawed library. Each prong is thoroughly coated with YAC cells, removed and placed onto the surface of a prewarmed AHC selection plate, ensuring that the orientation of well A1/ H12 is marked. The baseplate of the replicator is gently pressed to ensure that all pins make contact with the agar surface. The replicator is removed without smearing the transferred cells, and incubated upside down for 2 -3 days at 30°C. After this time a pattern of cells that copy the wells of the microtitre dish should be seen. The opened plates are resealed, snap frozen on dry ice and returned to - 70°C.

2.2 DNA Preparation Methods.

2.2.1 Preparation of Yeast DNA.

2.2.1.1 Culturing of Yeast over Multiple Generations

A single, freshly growing AB1380 yeast colony from a new YPD plate was used to inoculate 10 mL AHC broth in a sterile 50 mL Falcon tube and grown with angled shaking at 120 rpm overnight at 30°C. 10 µl of culture broth was diluted into 990 µl H_2O and vortexed. 2 x 10 µl of cells were analysed microscopically using a haemocytometer to determine an initial cell count (N_i). 10 µl of this initial starting culture of known cell concentration was used to inoculate fresh YPD broth, and the culture incubated again with shaking overnight at 30°C. The following day a cell count was again taken, and the total number of cells was (N_f). The ratio N_f/N_i gives the number of cells theoretically produced by one parental cell dividing to yield two daughter cells. From this value, the number of generations of division undergone by a single cell in that culture could be estimated. Subsequent cell counts were taken every 48 hours.

2.2.1.2 Preparation of Intact YAC DNA in Agarose

10 - 50 mL of sterile YPD was inoculated with a single freshly growing colony of AB1380 and grown overnight with shaking at 30°C. Cells were harvested in a benchtop centrifuge at 10 000 rpm for 10 minutes. The cell pellet was resuspended in 1 ml 50 m*M* EDTA and the suspension transferred to eppindorf tubes. The cells were quickly pelleted by spinning 13 000 rpm for 10 - 20 seconds, and resuspended in 200 µl 50 m*M* EDTA. 100 µl zymolase solution 1 (1 mL SCE : 25 ml 2*M* sorbitol, 1 mL 0.5*M* EDTA, 1.47g sodium citrate - make up to 50 mL with dH_2O; 50 µl β- mercaptoethanol; 10 mg zymolase (ICN)) and 0.5 mL of LMP agarose (1% in 0.125*M* EDTA) was added, the suspension mixed

by pipetting and immediately aliquotted into 200 µl block formers. The moulds were incubated at 4°C for 30 minutes to allow the agarose to set, then removed to 4- 5 mL EDTA solution (9 mL 0.5*M* EDTA pH 8.0, 1 mL 1*M* Tris- HCl pH 8.0, 0.5 mL β- ME) and incubated for > 4 hours at 37°C. The EDTA solution was discarded, taking care not to damage any of the agarose blocks, and 4 - 5 ml proteinase K solution was added (10 mg Proteinase K (Boehringer); 8 mL 0.5M EDTA pH 8.0; 1 ml 10% sodium sarkosyl; 1 ml 1 mg/mL RNAase (Sigma)) and the blocks incubated at 37°C for > 6 hours. YAC blocks in this form are ready for PFG separation, but in order for any subsequent manipulation, the proteinase K must be inactivated. This is done essentially as described as follows.

2.2.1.3 Liquid Preparation of Total YAC DNA

Total genomic and YAC DNA were isolated from a 50 mL overnight culture of transformed AB1380 host cells grown in YPD medium. Selection was not necessary for overnight growth. Cells were harvested by spinning at 3 500 rpm for 10 mins. in a bench centrifuge, washed once with 20 mL sterile H_2O and resuspended in eppindorfs in 0.5 mL breakage buffer (0.9*M* sorbitol; 14 m*M* β- ME; 50 m*M* phosphate buffer pH 7.5) containing zymolase between 2 - 5 mg/ mL and incubated at 37°C for 45 minutes. Spheroplasts were lysed by the addition of 50 µl 10% SDS, 50 µl 0.5*M* EDTA pH 8.0 and 100 µl of a 25 mg/ ml proteinase K solution and incubation at 65°C for 30 mins The sample was extracted once with an equal volume of PCI, and the DNA precipitated by the addition of 10*M* NH_4Ac to 1*M* and an equal volume of ethanol. DNA was usually seen to preciptate immediately, so a -70°C step was unnecessary. DNA was pelleted by spinning for 10 minutes at 13 000 rpm, washed with 70% ethanol, dried on a heating block and resuspended in 200 µl TE. 30 µl of RNAase (10 mg/ mL) was added and the reaction incubated at 65°C for 30 minutes. The DNA was extracted once with PCI and once with chloroform, precipitated with NH_4Ac/ ethanol, washed and resuspended in 200 µl TE. DNA was analysed spectrophotmetrically on a GeneQuant or electrophoretically on a 1% TAE mini-gel.

2.2.1.4 Preparation of Total YAC Plate Pools

The method used for preparing DNA representing the total information content of a single plate was essentially that described in Green and Olson (1990a). YAC's stored in 96 well dishes were grown for 2 - 3 days on AHC selection plates as described in section **2.1.6.4**. Yeast cells from each plate were pooled by 2 successive washes with 10 mL of 40 m*M* EDTA pH 8.0/ 90 m*M* β ME. A glass spreader was used to wash cells together, and was sterilised with alcohol between uses. The total cell poolate was collected in a 50 mL Falcon tube. Cells were harvested in a bench centrifuge for 10 minutes at 2 000 rpm and resuspended in 2 mL of Lysis buffer (1M sorbitol; 100 m*M* sodium citrate pH 7.0; 60 m*M* EDTA pH 8.0; 100 m*M* β ME; zymolase 20-T at 0.5 mg/ mL), and incubated at 37°C for 2 hours. The resulting spheroplasts were harvested as described above and resuspended in 0.7 mL (50 m*M* Tris- HCl pH 7.4; 25 m*M* EDTA pH 8.0; 500 m*M* NaCl; 3 m*M* β ME; 0.1% Nonidet P- 40; 1% SDS) and

incubated at 68°C for 15 minutes. Samples were extracted twice with PCI, precipitated at room temperature with isopropanol and digested overnight with 0.1 mg/ mL RNAase at 37°C. The DNA was re- precipitated with isopropanol, washed with 70% ethanol, dried and resuspended in 500 µl TE. DNA was analysed as described in the following section.

2.2.1.5 UV Spectrophotometric Analysis of YAC DNA

To accurately determine the purity and concentration of DNA in each plate pool preparation, 1 µl of DNA was diluted into 999 µl of sterile H_2O and analysed on a Perkin Elmer Lambda UV/ VIS scanning spectrophotometer across 230, 260 and 280 nm. Graphical data plots yield indications of residual amounts of ethanol, salts and proteins which could have adverse effects on subsequent manipulations, and allow for repurification of any DNA preps.

2.2.1.6 Preparation of YAC Masterpools

In the course of this project, DNA preparations representing total DNA from all 16 YAC plates were generated. In order to minimise time and materials used in the analysis of the DNA, YAC masterpools as described in Green and Olsen (1990a) were created. This was done by using the spectrophotomteric data generated to create master pools representing mixtures of DNA from more than one pool, at a final DNA concentration of 2.5 ng/ µl. The masterpools created are shown in table 2.2.

MASTERPOOL	PLATES
1	1 2 3
2	4 5 6
3	7 8 9
4	10 11 12
5	13 14 15 16

Table 2.2. Construction of the HB3 YAC Library MasterPools used in the course of this study.

2.2.2 Preparation of Bacterial DNA.

2.2.2.1 Mini- Prep Plasmid DNA Preparation

Small scale preparations of plasmid DNA were carried out according to the methods recommended by S. Keyes (personal communication). 1.5 ml of a 5 mL overnight culture grown in LB-

Amp at 37°C was aliquotted into a 1.5 mL eppindorf and the cells harvested at 13 000 rpm for 30 seconds. The supernatant was carefully decanted and the pellet allowed to drain. 300 µl of STET (4g sucrose (Sigma); 2.5 mL triton X- 100, 1.25 mL 2*M* Tris pH 8.0; 5 mL 0.5*M* EDTA pH 8.0; diluted to 50 mL with sterile distilled H_2O) and 25 µl of a 10 mg/ mL solution of Lysosyme (Sigma, freshly made in sterile H_2O and used within 5 min) were added and mixed with the cell pellet by vortexing. The cell suspension was boiled for 45 seconds and the viscous cell lysate centrifuged at 13 000 rpm for 15 minutes. The flocculent pellet was carefully removed and the DNA precipitated with an equal volume of isopropanol. The DNA was washed with 70% ethanol, dried on a heating block and the pellet resuspended in 50 µl TE. 25 µl of a 1/ 1000 dilution of 10 mg/ mL RNAase in sterile water was added to the DNA and the reaction incubated at 68°C for 10 - 20 minutes. 5 µl of plasmid DNA was analysed on a gel. The unused cell culture is viable at 4°C for several days and can be used for a second DNA prep or to re- inoculate fresh LB- Amp to reculture the plasmid. After four days, however, it was noticed that cell integrity seemed to diminish.

2.2.3 Preparation of DNA from a λ library

2.2.3.1 Preparation of Plating Bacteria.

50 mL of sterile LB medium supplemented with 0.2% maltose to induce *lamB*, was inoculated with a single Y1090 colony and cultured overnight with shaking at 37°C. The cells were harvested by centrifugation at 4 000 rpm for 10 min at room temperature. The supernatant was discarded and the cell pellet resuspended in 20 mL of sterile 0.01*M* $MgSO_4$. The cell suspension could be stored in this form at 4°C for up to three weeks.

2.2.3.2 Titring of the Phage Library.

In order to prepare the library for DNA extraction, the phage must be harvested from confluently lysed plates i.e. zones of lysis resulting from the infection of bacteria with a single phage which appear to just touch each other. This will give the optimal phage concentration. To do this, the library must be titred to estimate the pfu concentration (plaque forming unit). Tenfold serial dilutions of the phage stocks in SM were prepared. 0.1 mL of each dilution was mixed with 0.1 mL of plating cells in sterile eppindorfs and mixed. Cells and phage were incubated at 37°C for 20 min to permit phage adsorption. 3 mL of molten LB top agar was added, vortexed and poured onto a prewarmed LB plate. When the surface of the plate was evenly covered with molten agar, the agar was allowed to set and the plates incubated upside down overnight at 37°C. The dilution which yields confluent growth could then be determined.

2.2.3.3 Preparation of DNA from λ libraries.

10^5 pfu of bateriophage were mixed with 0.1 mL of plating bacteria and incubated for 20 min at 37°C. 3 mL of molten LB top agar was mixed with the adsorbed phage/ cell suspension, poured onto the surface of a prewarmed LB plate, allowed to harden and incubated upside down overnight. The following day plates demonstrating confluent lysis were rinsed with 5 mL SM and the top surface scraped off. 0.1 mL of chloroform was added and the suspension incubated with shaking for 15 min at 37°C. The released phage particles were harvested by centrifugation at 4 000 g for 10 min at 4°C. The supernatatant was recovered and one drop of chloroform was added to remove any bacteria still present. The phage suspension could then be stored at 4°C. EDTA from a 0.5*M* stock was added to a final concentration of 20 m*M* and proteinase K was added at 50 µg/ mL. SDS was added to 0.5% final concentration and the reaction incubated for 1 hour at 55°C. The DNA is extracted once with phenol, once with PCI, once with chloroform, precipitated with NaAc and ethanol, washed, dried and resuspended in 0.5 mL of TE.

2.3 Restriction Digestion of DNA

Restriction endonuclease enzymes were obtained from Boehringer Mannheim Chemical Company, New England Biolabs and Pharmacia Biochemical Corp. Restriction endonuclease digestion was performed in either the appropriate 10X buffer supplied by the manufacturer, or in 2X KGB buffer (200 m*M* potassium glutamate; 50 m*M* Tris- acetate pH7.6; 20 m*M* MgAc ; 100 µg/ mL BSA; 1m*M* β-ME ;[McLelland *et al,* 1988]), both to a final reaction concentration of 1X. Reactions were carried out, unless otherwise stated, at 37°C for 1 - 2 hours with 4 - 12 units of enzyme per µg of DNA. Reactions were terminated by the addition of a 1/10 volume of 50 m*M* Na_2 EDTA, 40% w/v sucrose, containing 0.25% bromophenol blue, 0.25% xylene cyanol.

2.4 Agarose Gel electrophoresis.

Agarose gels were used to check the integrity of DNA, to gauge DNA concentrations and to analyse PCR products and restriction digests. 1 - 2% agarose (IBI) was dissolved in the appropriate volume of 1X TAE buffer and microwaved as described in section **2.1.1.5**. After the solution had cooled to around 45°C, Ethidium Bromide (EtBr) was added from a 10 mg/ mL stock to a final concentration of 0.5 µg/ mL and mixed. The gel was poured into a level mould, and allowed to polymerise. Depending on the type of mould and number of wells used, the range of samples that could be analysed on a single gel varied from 6 - 35. The DNA to be analysed was mixed with 3 - 5 µl gel loading buffer (see previous section) and loaded into the submerged gel. Electrophoresis was carried out using a DNA Sub Cell and Power Pac 3000 (Bio Rad) with 1.5 L 1X TAE as buffer. Electrophoresis conditions were determined by

the size of the fragment and the nature of the analysis. Unless stated otherwise in the text, conditions were 80 - 100 volts for 1 - 2 hours. Visualisation was carried out using short wave UV, which causes the ethidium bromide intercalated within the DNA helix to fluoresce. Size estimation of DNA fragments was determined by using a variety of commercially available DNA size standards.

2.5 Southern Blotting Methods

Transfer of DNA fragments from agarose gels onto nylon membranes was performed essentially according to the methods of Southern (1975). We have found that blotting pad capillary transfer generally works best for pulsed field gels, whereas wick capillary transfer will suffice for ordinary agarose gels.

2.5.1 Wick Capillary Transfer

Gels containing fractionated DNA were prepared for transfer by soaking in 0.25 *M* HCl for 15 - 30 minutes, in order to partially depurinate the DNA. Gel were then soaked in 0.4 *M* NaOH for 30 - 60 minutes. A wick was constructed by using an inert support in a 0.4*M* NaOH bath and a strip of 3M chromatography placed over the support and cut to a length that allows both edges to be immersed in the bath. Pieces of 3M paper, cut to the size of the gel and prewetted in a 0.4*M* NaOH were placed on the support. The treated gel was placed on these filters, and a piece of Hybond N^+, prewetted with 0.4*M* NaOH and cut to the correct size, was placed over the gel. Saran wrap (Dow Chemical co.) was placed over the apparatus and cut to expose the gel to prevent evaporation and to focus buffer transfer through the blot. Two sheets of dry 3M paper, a stack of paper towels, a glass plate and a weight were placed on top of the gel. The bath was filled with 1 L of 0.4 *M* NaOH, and the DNA allowed to transfer from 4 hours to overnight. After transfer the blot was neutralised in 2X SSC for 10 minutes, dried and stored in Saran wrap

2.5.2 Blotting Pad Capillary Transfer

Gels are pre-treated for transfer as described above. The gel was placed on a 5 cm stack of blotting pads (Sigma), overlaid with Hybond N^+, Saran wrap, one 3M filter soaked in 0.4*M* NaOH, three dry 3M filters, a stack of paper towels, a glass plate and a weight. The bath is filled with at least 2L of 0.4 *M* NaOH, and the DNA allowed to transfer for a minimum of 24 hours. After transfer, the wells of the gel are marked on the filter with a pencil, for orientation purposes during blotting, neutralised, dried and stored in saran wrap.

2.6 Radioactive Labelling Methods

2.6.1 Preparation of PCR Products for Labelling

PCR fragments separated on an agarose gel were purified for use in a labelling reaction by using the Magic™ PCR prep kit (Promega), according to the manufacturers protocol. The DNA was separated in 1% LMP agarose and a slice containing the DNA fragment was melted in a 1.5 mL eppindorf at 68°C for 10 - 20 minutes. 1 mL of Magic ™ PCR prep resin was added to the melted agarose and vortexed. The mix was pipetted into a syringe barrel, and passed through a mini column. The column was washed with 2 mL of wash solution (80% isopropanol). The syringe was removed and the mini column centrifuged at 13 000 rpm to remove traces of solvent. The minicolumn was transferred to a new eppindorf and 50 µl of TE was added to allow the bound DNA to elute. After 1 minute, the eppindorf was centrifuged for 1 min at 13 000 rpm. 5 µl of purified DNA was analysed on a mini gel.

For smaller PCR products (< 100bp), it was found that the yield was extremely low and that probes generated from this method were poor. To prepare DNA fragments of this size, the direct purification protocol was used. The PCR reaction was transferred to a separate clean eppindorf and was mixed with 100 µl of Direct Purification Buffer ™. 1 mL of Magic ™ prep resin was added and vortexed. PCR product rescue is as described above.

2.6.2 Random prime labelling of DNA

In the course of this study, the random primed labelling reaction as described by Feinberg and Vogelstein (1983) was employed. The principle of random primed labelling is based on the use of random hexanucleotide primers to act as sites of DNA polymerase initiation, and the consequent incorporation of radioactively labelled dNTP's into the nascent DNA strand. Polymerisation reaction are carried out by the Klenow fragment of *E. coli* DNA polymerase I, which retains polymerase and 3` - 5` exonuclease activity, but which lacks the 5` - 3` exonuclease activity, which would degrade the primers and possibly remove the incorporated dNTP's. The template for polymerisation is denatured double stranded DNA, and polymerase initiation occurs along the length of the template. Consequently the size of the template does not affect the reaction, and incorporation is uniform along the length of the template.

The Random Primed DNA Labelling Kit (Boehringer) was used in this study to synthesise all radioactive probes from DNA templates. DNA (25 - 100 ng in TE buffer) was denatured by boiling for 10 min and then snap cooled on ice for 5 minutes, to prevent re- annealing of the DNA strands. The following mix was prepared:

25ng	Denatured DNA
3µl	dATP, dGTP and dTTP (1 µl of each)
2µl	Reaction Mix
5µl	α - [^{32}P] dCTP (1.87 Mbq)
1µl	Klenow enzyme

made up to 20 µl with sterile dH_2O.

The reaction was incubated for 30 min at 37°C, and the reaction terminated by heating to 65°C for 10 minutes Incorporation levels of 1.8 x 10^9 dpm/ µg were seen.

2.6.3 Preparation of cDNA Probes

2.6.3.1 Preparation of poly(A)$^+$ mRNA

In the preparation of messenger RNA, samples are contaminated with other forms of RNA present within the cell, namely ribosomal RNA and tRNA. In order to prepare cDNA probes, mRNA must first be separated from contaminant RNA. The method used here is based on the PolyA Tract® mRNA Isolation System (Promega), which uses biotinylated oligo(dT) primers to hybridise to the poly(A) tract of eukaryotic mRNA's. The hybrids are captured and washed at high stringency using streptavidin coupled to paramagnetic particles (SA-PMP's). A high magnetic field is used to capture the PMP's, and the mRNA is eluted from the solid phase by the simple addition of ribonuclease- free water.

The concentration of RNA was determined spectrophotometrically. 0.1 - 1 mg of RNA was made up to a volume of 500 µl with sterile RNase free water and the template denatured by heating at 65°C for 10 minutes. 3 µl of the Biotinylated- Oligo(dT) probe and 13 µl of 20X SSC was added to the RNA, mixed gently and incubated at room temperature until the reaction had cooled. 0.5X and 0.1X stock solutions of SSC in RNase- free water are prepared. The SA- PMP's are resuspended and then captured in the magnetic field. This is seen as a brown suspension becoming clear as the particles are attracted to the sides of the tube. The supernatant is removed and the particles washed three times with 0.3mL of 0.5X SSC. After each wash, the particles are captured magnetically and the supernatant carefully removed. The washed SA- PMP's are resuspended in 0.1 mL of 0.5X SSC and the entire contents of the annealing reaction are added to the tube. The reaction is incubated at room temperature for 10 mins, to allow the conjugation of the streptavidin and biotin moeities. The SA- PMP's are captured with the magnetic field, and the supernatant is carefully removed. The particles are washed four times with 0.1X SSC (0.3 mL) and after the final wash as much of the aqueous phase is removed as possible. The mRNA is eluted by resuspending the final SA- PMP pellet containing bound poly(A)$^+$ mRNA - biotinylated Oligo(dT) in 0.1 mL of Rnase- free water. The particles are magnetically captured, and the supernatant containing the eluted mRNA transferred to a sterile RNase- free tube. The elution step is repeated, and the two supernatants mixed. The concentration of the eluted RNA is determined spectrophotometrically and the RNA is precipitated with NaAc/ isopropanol overnight, centrifuged, washed, dried and resuspended in 10 µl TE for subsequent manipulation.

2.6.3.2 Reverse- Transcription PCR of poly(A)$^+$ mRNA

The method used is essentially that described by Dr Michael Lanzer (personal communication) and is a modification of that described in Froussard (1992) and in Lanzer *et al* (1993). Fractionated poly(A)$^+$ mRNA in a volume of 10 µl is denatured by heating at 65°C for 10 minutes and kept on ice to prevent re- annealing. To the RNA the following is added:

8µl 5X RT buffer

4µl 0.1*M* DTT

2µl pd$(N)_6$ (10 mg/ ml)

4µl dGTP, dATP, dCTP, DTTP (1 µl each)

3µl Reverse transcriptase 200U/ µl (Superscript, BRL)

12µl H_2O

The reaction is incubated at 42°C for 60 minutes. The reaction is stopped by the addition of 2 µl of 0.5*M* EDTA, 2 µl of 10% SDS, 1.4 µl 10N NaOH and incubating at 65°C for 30 minutes.

20 µl of 1*M* Tris pH7.4, 5 µl of 2N HCl and 28 µl of 1 x TE are added, the reaction split into two aliquots and separated on a Clonetech G30 column (see **2.6.4.2**). The products are collected, mixed, NH_4Ac is added to 1*M* and the cDNA is precipitated with an equal volume of ethanol. The cDNA is resuspended in 12.5 µl of H2O and the following PCR mix is made:

12.5 µl cDNA

2.5 µl 10X PCR buffer (100 m*M* Tris pH 8.4, 500 m*M* KCl, 15 m*M* $MgCl_2$, 0.1% gelatine, 1% Triton X- 100)

2.5 µl dNTP mix (2 m*M* dGTP, 2 m*M* dATP, 2 m*M*dTTP, 1 m*M* dCTP)

2.5 µl Primers pd$(N)_6(G/C)_3$ (1 µg/ µl)

5µl α - [^{32}P] dCTP

1µl *Taq* polymerase (1u / µl)

The reaction was overlaid with a drop of oil. PCR conditions are as follows: 94° 10 sec/ down ramp to 24° over 2 min/ 24° 10 sec/ back ramp to 94° over 2 min 45 sec/ for 30 cycles of amplification. In order to carry out this kind of PCR, a refrigerated PCR thermal cycler is necessary, rather than a fan cooled, because of the low temperatures involved. PCR's were carried out on a Biometra Trio- block, with appropriate radioactive shielding. On completion, 25 µl of TE was added to the PCR reaction, and the products separated on a Clonetech 100 spun column. Probes are then ready for hybridisation.

2.6.4 Removal of Unincorporated Nucleotides.

In general it is possible to use the products of a labelling reaction directly in a hybridisation reaction. However, it is sometimes advantageous to reduce possible background signals by separating the

labelled probe from the excess of unincorporated nucleotides in the reaction mixture. In the course of this study this was performed in two ways:

2.6.4.1 Preparation of Sephadex G50 Columns.

Columns were made by packing a glasswool - plugged disposable pipette with Sephadex G-50 prepared in TE buffer (10 m*M* Tris-HCl, pH 8.0, 1m*M* ETDA pH 8.0) A reservoir of buffer was placed on top of the column to ensure that the column did not dry out and that a constant downward flow was maintained. 5 µl of gel loading dye was mixed in with the labelling reaction and the sample loaded on the column. A recent development from Amersham has been the inclusion of Redivue™ dye which migrates with the labelled probe, and the gel loading dye migrates with the unincorporated nucleotides. As separation progresses down the column. separation can be followed using a Geiger counter. Two distinct radioactive peaks, separated by a trough can detected. The leading peak, identified visually with Redivue dye, corresponds to the probe, and the lagging peak , identified by the blue loading dye, corresponds to unincorporated nucleotides. All of the leading fraction was collected into 1.5 mL eppindorfs and stored at 4°C

2.6.4.2 ChromaSpin Spun Columns

While G50 columns are an accurate method of ensuring removal of unincorporated nucleotides, they are laborious and time consuming. A more rapid method of purifying probes is the use of ChromaSpin Spun Columns (Clonetech Laboratories Inc.). The columns can be used for the purification and size selection of radiolabelled probes, depending on the choice of column used. Separation is based on a matrix of defined pore size. Molecules larger than the pore size are excluded from the resin and migrate rapidly through the column whereas smaller molecules are retained by the column. Columns were used essentially according to manufacturers instructions. Columns were inverted several times to completely resuspend the gel matrix, opened, placed into 1.5 mL collection tubes and centrifuged for 3 minutes at 1 800 rpm to remove excess buffer from the column. The sample volume to be loaded was made up to 50 µl with TE and loaded into the centre of the semi- dry gel bed, taking care not to load any along the inner wall of the column. The column was placed into a second 1.5 mL collection tube and centrifuged for 5 minutes at 1 800 rpm. If separation has occurred correctly, a red dye is visible at the top of the column, and the eluate is clear. The column is discarded safely and the collection tube sealed and stored at 4°C.

2.6.5 Safety Considerations

The radiation used in this study is α - [^{32}P], a β source. β particles are less densely ionising than α particles, but are more penetrating and can be detected by a counter placed 0.5 m away. High energy β particles pose an external hazard, however the energies of the radioactive nucleotides used are

not significantly large and this combined with the inverse square law for radiation means that, in the short term β particles can be handled safely. However, this project has demanded the continuous use of radioactivity, and so the safety concern becomes the cumulative effect of exposure to low energy radiation. To minimise this, basic safety precautions are taken: to prevent ingestion and thus cause an internal hazard, protective clothing is worn over open areas of skin that might come into contact with radiosources; all radioactive manipulations occur behind protective 10 mm plexiglass shields, and all radiosources are stored in shielded containers. Waste matter (tips, columns etc.) are disposed of in a containment hood within a radioactive room.

2.7 Southern Hybridisation.

2.7.1 Prehybridisation

The methods used was that recommended by duPont. All hybridisation reactions were carried out in a hybridisation oven (Hybaid ™). As all probes were detecting homologous sequences, conditions of high stringency were used throughout. To maximise the rate of annealing, hybridisations were carried out in the presence of 10% dextran sulphate, used as an exclusion agent to reduce the operational volume of the reaction and a high ionic strength buffer, NaCl, to carry out hybridisations 20 - 25°C below the T_m.

100 mL of hybridisation buffer as made as follows: 84.6 mL H20, 10g Dextran Sulphate (sodium salt), 10 mL 10% SDS. The buffer was heated at 65°C until it had dissolved, then 5.8g NaCl were added and heated at 65°C for a further 30 minutes. The buffer was aliquotted into 50 mL Falcon tubes which could be stored indefinitely at -20°C. 10 mL of buffer was used to wash southern blots for a minimum of 90 min at 62°C with rotation. For cDNA probes, 20 mg of total yeast RNA (Boehringer) was included in the buffer.

2.7.2 Hybridisation.

To prepare a probe solution depends upon the method used to purify the probe. If a G50 column was used, probe solutions were made as follows: 0.7 mL H_2O, 0.2mL α - [^{32}P] labelled probe, 0.1 mL of 5 mg/ mL sheared and sonicated herring sperm DNA. If the probe was purified from Chromaspin columns, probe solutions were made as follows: 50 µl α - [^{32}P] labelled probe, 0.25 mL denatured carrier DNA, 0.7 mL H_2O. In both cases, the probe solution is boiled for 10 min and kept on ice for 15 minutes. For probes prepared from G50 columns, it is possible to use a 0.5 mL aliquot of probe in a hybridisation reaction, whereas for spun column probes, it is necessary to use the total volume of probe. Probes are added to the hybridisation buffer, and the reaction allowed to incubate with agitation

overnight at 62°C. For cDNA probes, 20 mg of total yeast RNA is included in the probe solution. The radioactive buffer can be stored at 4°C and used in subsequent hybridisation reactions.

2.7.3 Filter washing.

Filters were washed of excess, unbound probe initially in 0.2X SSC at room temperature for 10 minutes. Filters were then washed for 30 min at 62°C with 0.2X SSC / 0.1% SDS and checked for radioactivity in non - DNA regions with a Geiger counter. If counts were detected, the washing step was repeated until the filters were clear. Filters were allowed to dry at room temperature and sealed in Saran wrap.

2.7.4 Filter stripping.

In order to remove reversibly bound probe from filters for subsequent hybridisation with a different probe, filters were washed in excess of 0.4*M* NaOH at 42°C for 30 minutes, followed by neutralising in 0.1X SSC/ 0.1% SDS/ 0.2*M* Tris pH 7.5 for 30 min at 42°C. Filters were exposed overnight to X- Ray film to determine the efficiency of probe removal.

2.7.5 Autoradiography.

Autoradiography was performed using Kodak XAR -5 X- ray film for ^{32}P probes, in a cassette containing intensifying screens. The cassettes were stored at -70°C for 24 - 96 hours depending upon the age and quality of the probe and the quality of DNA transfer to the filters. For ^{35}S gels, cassettes containing Kodak Biomax MR film were stored at room temperature for 24 hours. All films were developed using a X-O'GRAPH X1 automatic X- Ray film processor.

2.8. The Polymerase Chain Reaction (PCR)

2.8.1. General PCR

PCR is an *in vitro* technique by which a target stretch of DNA flanked by regions of known sequence is amplified (Saiki *et al.*, 1988). PCR involves primer sequences annealing to the known flanking sequences and repeated rounds of denaturation, annealing and extension in the presence of all dNTP's and a thermostable DNA Polymerase. The polymerase used was isolated and purified from a thermophilic bacteria, *Thermus aquaticus,* whose enzymes are adapted for optimal activity at elevated temperatures. The target sequence can be amplified as much as 10^9 times.

PCR reactions were performed in sterile 0.5 mL eppindorfs as follows:

0.2µl 100x Primer 5' (10 µ*M*)

0.2µl 100x Primer 3' (10 µ*M*)

0.2µl 100x dNTP's (7.5 m*M* dATP, dTTP, dCTP; 6.5 m*M* dGTP; 1m*M* deazo-dGTP) (Boehringer)

0.2 µl 2% gelatine

2 µl 10X PCR buffer (500 m*M* KCl; 100 m*M* Tris- HCl pH 8.8; 25 mM $MgCl_2$)

0.1 µl Taq polymerase (Cetus corp.)

1 - 4 µl DNA template

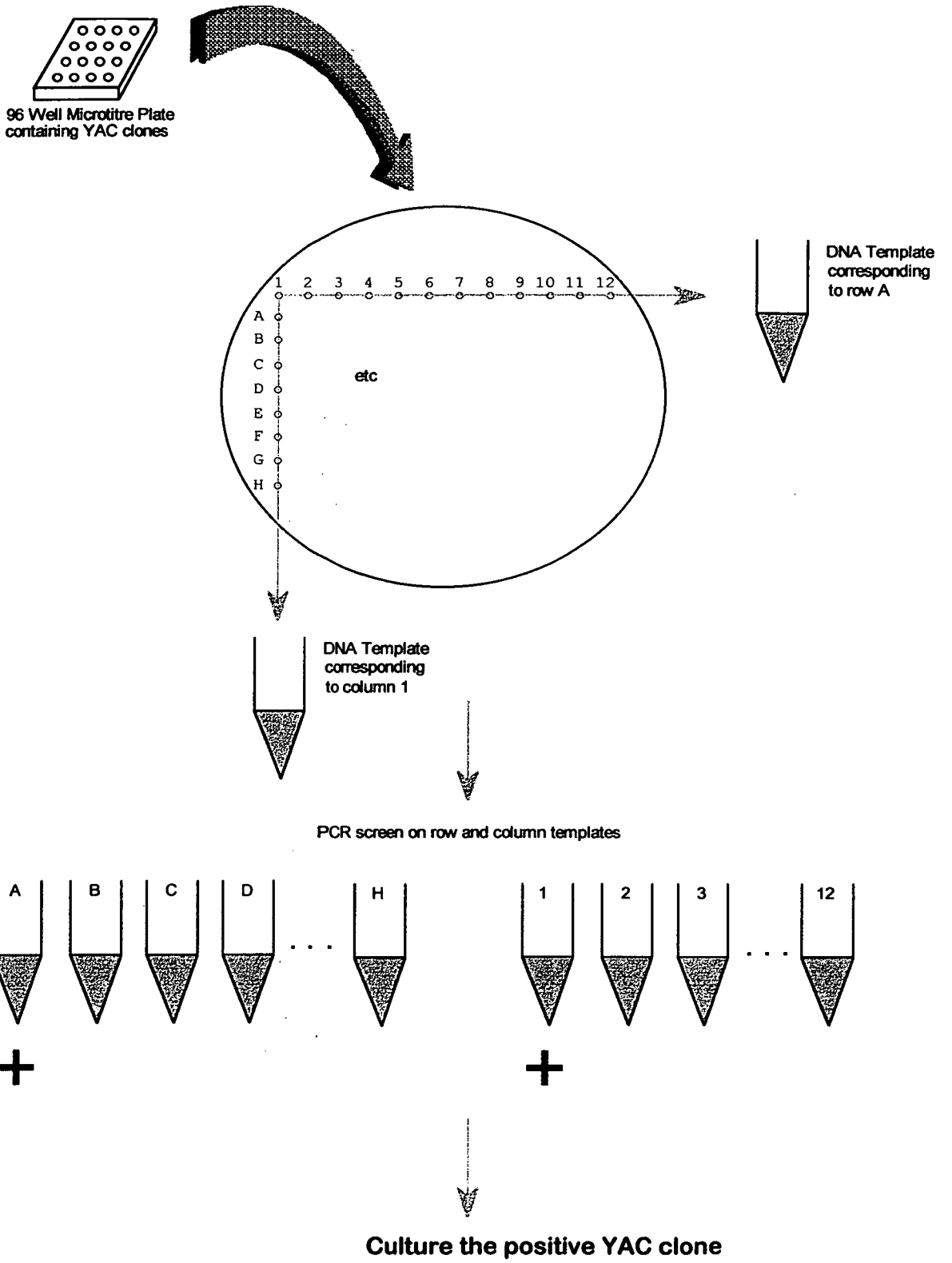
16.1 - 13.1µl H_2O

The reaction was overlaid with a drop of sterile light mineral oil (Sigma), and the reaction carried out on a Hybaid™ Omnigene thermal cycler. The reaction volumes detailed are for 20 µl reactions, and must be upscaled accordingly for 50 µl reactions. To save on materials, the total number of reactions per primer is made as a single premix and then UV irradiated for 5 min to remove any contaminant DNA. The reaction volume is dispensed into eppindorfs and template is added. 7- Deaza-2' dGTP is incorporated into the nucleotide mix, as the base analogue prevents the formation of secondary structures in the newly amplified DNA which could inhibit subsequent rounds of amplification (McConlogue *et al.*, 1988)

2.8.2 YAC Specific PCR

2.8.2.1 Row and Column PCR

In order to identify single YAC's containing target sequences from a bank of 96 clones, (the storage format), row and column PCR was used to identify the exact colony co- ordinates of the positive YAC (Heard *et al.*, 1989). The technique is a simplification of that described by Kwiatkowski *et al.*, (1990) and is illustrated in fig 2.1. YAC plates are cultured as described in section **2.1.6.4.** Using sterile yellow tips a small scraping of a colony was transferred to 100 µl of sterile TE in a 0.5 mL eppindorf, and the tip discarded. This procedure was repeated until samples of all the YAC colonies in the first row were in suspension in the tube. This was repeated for all the rows. and again for all the columns, 20 tubes in all. The tubes were boiled for 10 minutes, and centrifuged 13 000 rpm for 5 minutes. The supernatant was diluted 1/ 10 and 4 µl of this was used in a PCR reaction. It was noted that with this kind of crude DNA preparation, the number of cycles had to be increased to 40 for a signal to be seen and that DNA prepared in this form was stable at - 20°C for several weeks, but after rounds of freeze-thawing, the template became unsuitable for use in a PCR reaction and a new set of templates had to be prepared.


96 Well Microtitre Plate containing YAC clones
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
etc
DNA Template corresponding to row A
DNA Template corresponding to column 1
PCR screen on row and column templates
A
B
C
D
H
1
2
3
12
Culture the positive YAC clone

2.8.2.2 End Rescue By Inverse PCR

Inverse PCR is a useful technique that allows the *in vitro* amplification of DNA flanking a region of known sequence. The technique is based on the digestion of the target DNA with restriction enzymes and recircularisation of the cleavage products, prior to amplification with primers that are in an inverse orientation to conventional PCR (Triglia *et al.*, 1988; Ochman *et al.*, 1988). Primers used were as described in section **1.5.2.**

Intact YAC's in agarose blocks were prepared as described (see section **2.2.1.2**). Blocks were washed out overnight in 1X TE, then 4x 30 min at 50°C. A 4mm slice of agarose was melted at 68°C for 15 - 30 minutes, and 4 - 8 µl of liquid YAC DNA was digested in 20 µl reactions with *Eco*RV (Boehringer), *Rsa*I (Boehringer), *Nhe*I (Boehringer), *Mse*I (Boehringer) and *Taq*I (Boehringer) for 30 min at 37°C, except for *Taq*I which has a reaction temperature of 65°C. Incubating the digest reaction for a longer time has no effect on the efficiency of the ligation and subsequent PCR. Samples are heated at 68°C for 15 min and immediately loaded into the ligation mix:

34µl	dH_2O
5µl	10X Ligation buffer
10 µl	Liquid, digested DNA
0.3µl	T4 DNA Ligase (400, 000U/ µl; NEB)

The reaction was incubated overnight at 14 - 16°C. The following day, ligation products were heated at 68°C for 10 min prior to PCR. The PCR reaction was carried out in a 50µl volume with 1 µl of template and 0.5U of *Taq* polymerase (Cetus). PCR conditions are : (Right Arm) 94° 60s; 52° 60s; 72° 180s for 35 cycles; (Left Arm) 94° 60s; 52° 60s; 72° 120s for 35 cycles.

2.9 Bacterial Recombinant DNA Techniques

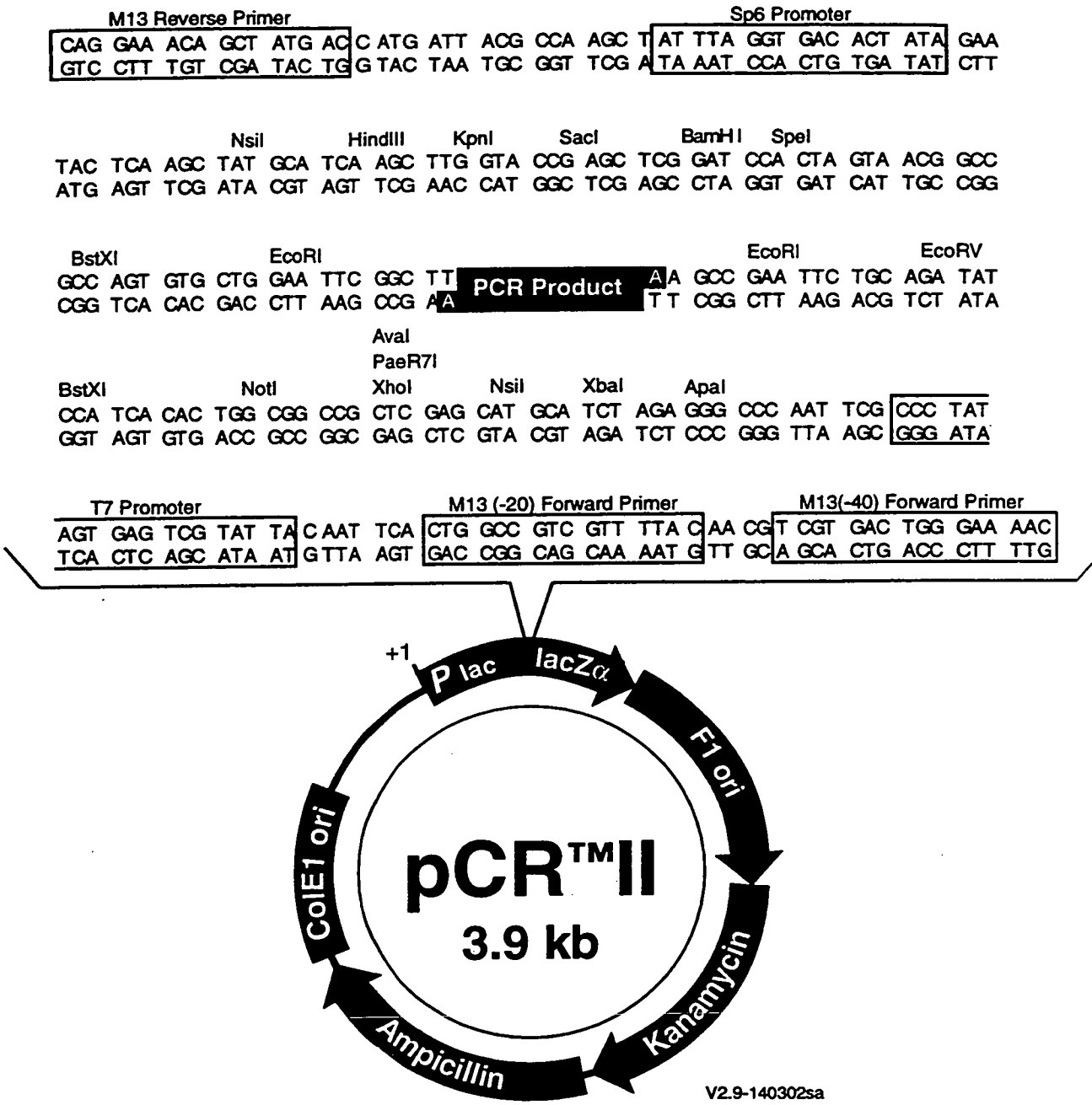
2.9.1 Cloning into the pCR II™ Vector.

PCR products were prepared for sequencing by cloning into the TA Cloning™ System (Invitrogen Corp), a one step method for cloning PCR products (See fig 2.2). The system makes use of the fact that certain DNA polymerases also catalyse non template addition of single nucleotides to the 3` end of all duplex molecules (Clarke, 1988), thus eliminating the need to blunt end or alternately modify the PCR product to facilitate cloning. *Taq* polymerase is known to add a single deoxyadenosine residue to the 3` end of PCR products and the vector has a single T overhang at the cloning site to exploit this (Mead *et al* 1991). This system was used to generate all STS marker stabilates

The ligation and transformation reactions were carried out according to the manufacturers instructions. PCR products were cleaned from the oil overlay and incubated in a ligation reaction with

Figure 2.2

Diagrammatic representation of the pCRII™ plasmid vector map. The sequence above the vector represents the pCRII vector sequence with the PCR product cloning site indicated.


M13 Reverse Primer
CAG GAA ACA GCT ATG AC C ATG ATT ACG CCA AGC T AT TTA GGT GAC ACT ATA GAA
GTC CTT TGT CGA TAC TG G TAC TAA TGC GGT TCG A TA AAT CCA CTG TGA TAT CTT
Sp6 Promoter
NsiI HindIII KpnI SacI BamHI SpeI
TAC TCA AGC TAT GCA TCA AGC TTG GTA CCG AGC TCG GAT CCA CTA GTA ACG GCC
ATG AGT TCG ATA CGT AGT TCG AAC CAT GGC TCG AGC CTA GGT GAT CAT TGC CGG
BstXI EcoRI EcoRI EcoRV
GCC AGT GTG CTG GAA TTC GGC TT PCR Product A A GCC GAA TTC TGC AGA TAT
CGG TCA CAC GAC CTT AAG CCG A A T T CGG CTT AAG ACG TCT ATA
AvaI
PaeR7I
BstXI NotI XhoI NsiI XbaI ApaI
CCA TCA CAC TGG CGG CCG CTC GAG CAT GCA TCT AGA GGG CCC AAT TCG CCC TAT
GGT AGT GTG ACC GCC GGC GAG CTC GTA CGT AGA TCT CCC GGG TTA AGC GGG ATA
T7 Promoter
M13 (-20) Forward Primer
M13(-40) Forward Primer
AGT GAG TCG TAT TA C AAT TCA CTG GCC GTC GTT TTA C AA CG T CGT GAC TGG GAA AAC
TCA CTC AGC ATA AT G TTA AGT GAC CGG CAG CAA AAT G TT GC A GCA CTG ACC CTT TTG
+1
P lac
lacZα
F1 ori
pCR™II
3.9 kb
ColE1 ori
Ampicillin
Kanamycin
V2.9-140302sa

the pCR™ vector as 1:1 to 1:5 molar ratios of vector: insert, depending on the size and degree of amplification of the target DNA sequence. Ligation reactions were set up as follows:

2µl pCR II vector (50 ng)

1 µl PCR product

1µl 10X Ligase Buffer

1µl T4 Ligase (4U/ µl)

6µl H_2O

Ligation reactions were incubated overnight at 14 - 16°C.

2.9.2 Transformation of competent *E. coli*

2 µl of β Mercaptoethanol was gently mixed with thawed competent cells (25 - 50 µl). 2 µl of each ligation mix was added directly to each vial of competent cells, gently mixed and incubated on ice for 30 minutes. The unused ligation mixtures could be stored at -20°C for up to several weeks for repeated transformations. Cells were heat shocked at 42°C for 30 seconds and placed on ice for 2 minutes. 450 µl pre- warmed SOC medium was added and the cells incubated at 37°C for 1 hour with horizontal shaking. Cells were stored on ice for plating. 25 and 100 µl of transformed cells were spread directly onto LB- Amp plates containing 25 µl X -gal per plate, allowed to dry and incubated upside down at 37°C for 24 hours. Plates were then incubated at 4°C for 2 -3 hours to allow the blue selection colour to develop. This last step is useful for eliminating putative false positives.

2.10 Sequencing.

DNA sequencing was performed using the dideoxy chain termination reaction as described by Sanger *et al* (1977) available as the Sequenase™ Version 2.0 Kit (USB Corp.). The method involves the use of nucleotide analogues in a polymerisation reaction using a Histidine[123] mutant of bacteriophage T7 polymerase which has increased processivity (up to ninefold that of native T7) and which lacks the 3` - 5` exonuclease activity (Tabor and Richardson, 1989). The dideoxynucleotide analogues lack the 3`- OH group necessary for chain elongation, and result in the synthesis of a truncated DNA strand. The protocol used includes the modification for sequencing double stranded DNA using 10% DMSO in the reaction to enhance signal intensity, reduce background and prevent the formation of secondary structures within the sequencing template (Winship, 1989)

2.10.1 Preparation of DNA

To prepare double stranded DNA for sequencing, the template is first denatured. To 5 µg of DNA in a 20 µl reaction, 5 µl of 4N NaOH, 20 µl 1 m*M* EDTA and 55 µl H_2O were added, and the reaction incubated at 37°C for 30 minutes. The DNA was ethanol precipitated overnight at -70°C with

1*M* NaAc, centrifuged at 13 000 rpm for 15 minutes, washed, dried and resuspended in half the starting volume (10 µl).

2.10.2 The Sequencing Reaction

The sequencing reaction can be broken down into two steps: (i) labelling, involving the extension of the primer with limiting dNTP's and radioactive label, resulting in the almost total incorporation of radioactivity into all the nascent DNA chains; and (ii) termination, involving the addition of ddNTP's to the reaction to progressively halt DNA synthesis.

The labelling mix was set up as follows:

2 µl	Sequenase Reaction Buffer
1µl	10% DMSO
1µl	Primer (1ng/ µl)
6 µl	Denatured ds DNA

The reaction is heated at 96°C for 5 minutes and snap cooled on ice for 2 minutes. Annealing of primer to template occurs in this stage. Then the following is added to the reaction mix:

1µl	0.1M DTT
1 µl	α - [35S] dCTP (1000ci)
1µl	Labelling Mix (1.5 µ*M* each dNTP) diluted 1/ 5 in H_2O
1µl	Sequenase (diluted 1/ 5 in Enzyme dilution Buffer)

The solution was mixed thoroughly and incubated for 15 minutes at room temperature. 3.5 µl of labelling mix was aliquotted into a pre warmed (37°C) termination mix, corresponding to each appropriate dideoxy NTP, made by mixing 2.5 µl ddNTP with 0.5 µl DMSO. The termination reaction was incubated at 37°C for 5 minutes and the reaction stopped by the addition of 4 µl stop solution (20 m*M* EDTA containing 0.05% Bromophenol blue, 0.05 % xylene cyanol and 95% Formamide). Samples were heated at 85°C for 5 min prior to loading (3 µl per well) into a 6% denaturing polyacrylamide gel.

2.10.3 Poly Acrylamide Gel Electrophoresis

Sequence reaction were separated on a 380 mm x 0.3 mm 6% polyacrylamide gel. Plates used for pouring were thoroughly cleaned with acetone, and the top plate was siliconised with Sigmacote (Sigma) to prevent the gel from sticking. The gel was made as follows:

50g	Urea (Final concentration 8M)
30 mL	dH2O
10 mL	10X TBE

20 mL 30% Acrylamide/ 1.6 % Bis- acrylamide (6% and 0.16% final concentrations)

The mixture was dissolved by mixing under hot water and filtered. 0.8 mL 10% ammonium persulphate and 20 µL TEMED (N, N, N' N' - tetramethylethylenediamine) were added and the mixture poured between the glass plates. After polymerisation, the gel was pre- run for 15 - 30 minutes, the samples loaded and electrophoresis performed at 50W constant power on an IBI Baserunner™. On completion of a run, the gel was dried under vacuum on a gel drier (Model 583 Gel Dryer, BIO - RAD) for 1 - 2 hours at 80°C, and exposed to X- ray film (Kodak Biomax MR) overnight at room temperature. Films were developed in an automated X- ray film processor (X- ograph X1)

2.10.4 Database Searches

Sequences were analysed using the University of Edinburgh Computing facilities. Data analysis was performed using the University of Wisconsin Genetics computer Group (GCG) Sequence Analysis Software Package Version 7.1 (Devereux *et al.*, 1984)

2.11 Mapping of YAC's

Physical mapping of artificial chromosomes is based on the use of probes specific to the left and right arms of the YAC vector (Hirst *et al.* 1991). Total YAC DNA is partially digested with restriction endonucleases, separated on a PFG with appropriate conditions for the selection of fragments of the desired size, transferred onto nylon and subjected to successive rounds of probing and stripping with left arm, right arm and internal STS probes. Data obtained is composited to generate the resultant physical map, whose resolution depends on the number and activity of the endonucleases chosen (Chiu *et al.*, 1994; Gaensler *et al,*. 1991)

The protocol used was a modification of that described by Dr. N. Fairweather (*personal comm.*) and makes use of the fact that T4 polymerase buffer works with almost all restriction enzymes (Monaco *et al*, 1992). The buffer is made as follows (10X stock): 500 m*M* NaCl, 100 m*M* Tris-HCl pH 8.0, 100 m*M* $MgCl_2$). Blocks are washed overnight at 4°C in excess TE, and then treated twice 90 min with two changes of 1X TE at 50°C, to ensure complete removal of the 0.5*M* EDTA storage buffer. The blocks are then washed in 1X TE for 30 min at room temperature. Blocks are cut into 3 mm slices, and half slice of these are used per reaction. In order to obtain a complete range of partial digest fragments, from undigested to complete digest a series of digest reactions containing 15U, 3U, 0.3U and 0.03U per enzyme per YAC were set up as follows in 1.5 mL eppindorfs:

30 µl	YAC gel slice
20 µl	10X T4 polymerase buffer
2 µl	10 mg/ mL BSA (100X, NEB)

20 µl 5 m*M* DTT (diluted from a 1M stock)

120 µl 100X Spermidine (75 m*M*, Sigma)

3µL Restriction endonuclease

The reaction is incubated overnight on ice at 4°C to allow sufficient time for the enzyme to completely saturate the block. The following day, the reaction was incubated at the enzyme digest temperature as per manufacturers instructions for 1 hour. The reaction was terminated by adding 0.5 mL 0.5M EDTA pH 8.0 and the digests could be stored in this form at 4°C for several weeks, or until separation on a PFG. Digests were then separated, visualised, photographed, blotted, probed, stripped and reprobed. This protocol appears to yield better digestion products than other published protocols attempted (Bentley *et al.*, 1993; Hamvas *et al.*, 1994).

CHAPTER 3.

CONSTRUCTION AND CHARACTERISATION OF A YEAST ARTIFICIAL CHROMOSOME LIBRARY OF *P. Falciparum* CLONE HB3.

3.1 Abstract

A Yeast Artificial Chromosome library of the human malaria parasite *P. falciparum* clone HB3 was constructed in the YAC vector pYAC4. HB3 DNA was partially digested with *Eco*RI, ligated to *Eco*RI/ *Bam*HI digested and phosphatase treated vector in agarose, size selected on a pulsed field gel and used to transform spheroplasts of the yeast strain AB1380. Transformation efficiencies for the construction of the library were around 5 x 10^5 transformants / μg. The library contains approximately 300 YAC's of insert size 150 kb and 800 YAC's of average insert 100 kb. Screening of the library with PCR primers showed that of the 29 sequences tested, 26 were represented in the library. Library size estimations indicate a 90% representation for single copy sequences and a four to five fold coverage of the genome. Stability assays indicate that the cloned YAC DNA is an accurate representation of genomic DNA and is stable when cultured over 100 generations.

3.2. Introduction.

The artificial chromosome vector used in this study was pYAC4. In *E. coli*, the vector replicates as a circular 11.3 kb plasmid, which is linearised on digestion with *Bam*HI. When the YAC is digested with *Bam*HI and *Eco*RI, three fragments are produced corresponding to the left and right YAC vector arms and a 1.8 kb "stuffer" fragment. This fragment contains the HIS3 gene which is excluded from the cloning reaction. Digestion with *Bam*HI exposes the telomerase substrate sequences which can be recognised by the *S. cerevisiae* telomerase and converted into functional telomeres *in vivo*. The 6 kb left arm contains ARS1 and CEN4 sequences as well as the selectable marker TRP1. The 3.4 kb right arm contains the URA3 selectable marker (Bach *et al.*, 1979) with the *Eco*RI cloning site located within the SUP4 gene. SUP4 produces a mutant tRNA ^{tyr} which suppresses the ochre mutation in the yeast *ade2* gene. Cells containing an intact copy of SUP4 (i.e. non- recombinant YAC's) are red, whereas YAC's containing an interrupted copy of the gene result in white yeast colonies.

The construction of Yeast Artificial Chromosome libraries can be separated into several distinct phases: (i) partial digestion and size fractionation of the insert DNA, (ii) ligation of vector arms to the DNA and size fractionation of the recombinant YAC's (iii) the transformation of yeast spheroplasts and (iv) the subsequent selection and storage of each YAC clone. The original human YAC libraries were constructed with liquid genomic DNA, size fractionated on sucrose gradients as ligation templates (Burke *et al.*, 1987; Coulson *et al.*, 1988; Garza *et al.*,1989). YAC's constructed by this approach were limited in use due to the relatively small insert sizes (~ 200 kb), possibly due to shear forces acting upon the DNA template. The protocol was modified by the use of high molecular weight DNA embedded in agarose and size fractionated by pulsed field gradient gel electrophoresis (Anand *et al.*, 1989; Anand *et al.*, 1990; Albertsen *et al.*, 1990). This minimised the shear forces on the DNA and yielded higher inserts (McCormick *et al.*, 1990). Further refinements included the addition of polyamines, which acted to eliminate size bias in the transformation of larger YAC's (Connelly *et al.*,

1991) and the inclusion of a second size fractionation step, after the ligation of YAC plasmid to insert DNA (Imai and Olson, 1990; Larin *et al.*, 1991). Such modifications yielded inserts of up to 700 kb (Larin *et al.*, 1991). The final protocol used was a modification of that described by Monaco *et al.* (1993) in that all manipulations - partial digestion, ligation and size- fractionation - occurred in agarose and liquid DNA was only employed prior to the transformation step.

The template DNA used was the parasite clone HB3. The parasite was isolated in Honduras (Vaidya *et al.*, 1993) and is distinct from the WTMGC reference strain 3D7 (Walliker *et al.*, 1987). In addition, both 3D7 and HB3 were the parental strains used for a genetic cross (Walliker *et al.*,1987) for which cloned recombinant progeny exist (Walliker *et al.*, 1987; Sinnis and Wellems, 1988; Wellems *et al.*, 1988,). The large number of putative chromosome- specific polymorphic markers that could be constructed as a direct result of YAC contig analysis and characterisation could form the basis of a genetic approach to gene analysis. In addition, physical map data generated would enhance and complement maps constructed in the reference clone.

3.3. Strategies for the Construction and Screening of the HB3 YAC library

In order to prepare the pYAC4 vector DNA (Burke *et al.*, 1987), plasmid maxipreps were performed and the plasmid DNA separated on caesium chloride gradients. The DNA was spectrophotometrically analysed for purity and was restriction digested to test for spontaneous telomere deltion during prior manipulations. The vector was digested with *Eco*RI and phosphatase treated. HB3 DNA was harvested from parasites and partially digested with *Eco*RI to yield DNA fragments in the desired size range.

The cloning strategy involves the ligation of vector arms to insert DNA in agarose, by allowing the diffusion of restricted YAC plasmid into the agarose slice. Ligations occurred in a molar excess of vector to insert, to (i) maximise ligation of vector to the genomic inserts and (ii) to minimise the incidence of coligation of multiple genomic fragments (Haldi *et al.*, 1994) These recombinant products were size fractionated on a Pulsed field gel, and slices corresponding to 75 - 145kb and 145 - 194 kb were used to transform yeast spheroplasts, prepared using a cell wall lytic enzyme (Lyticase). This enzyme acts to sever glucan chains linking the cell wall to the plasma membrane and to partially degrade the mannoprotein component of the cell wall. In the presence of PEG and Ca^{2+} ions, DNA is taken into the cell (Hinnen *et al.*, 1978; Burgers and Percival, 1987). Recombinant YAC's were screened by selecting first on (-uracil) plates, to allow the reconstitution of the yeast cell wall, then (-uracil/ -tryptophan) plates for selection of inheritance of both YAC arms. Positive YAC's were then grown in a nutrient- rich medium YPD, in 96- well microtitre plates, replicated and stored at -80°C.

To prepare DNA for PCR analysis, each plate was replica stamp cultured on selection plates and total DNA preparations of each plate were combined (Green and Olson, 1990a). Aliquots of DNA from sets of plate pools were combined to provide master pool templates to facilitate PCR analysis (see

fig 3.1). Master pools were screened with multiple PCR primers from divergent sites within the genome so that an estimation of genome representation could be proposed. Such a PCR- based first level library screen has become a standard strategy in the identification of chromosome specific sequences from YAC libraries (Green *et al.*, 1991; Schlessinger *et al.*, 1991; de Bruin *et al.*, 1992).

To test whether genomic sequence information was accurately represented in the cloned DNA, YAC's and genomic DNA identified by a particular marker were digested and probed with the radiolabelled marker. To attempt to evaluate clone stability, a random YAC clone was cultured over 100 generations, and DNA samples from various generation time points were taken. The clone was analysed as intact artificial chromosomes, as YAC's digested with rare cutting enzymes and as loci digested with frequently cutting enzymes.

3.4. Results.

3.4.1 Construction of a Yeast Artificial Chromosome Library of the *P. falciparum* isolate HB3 in the YAC Vector pYAC4

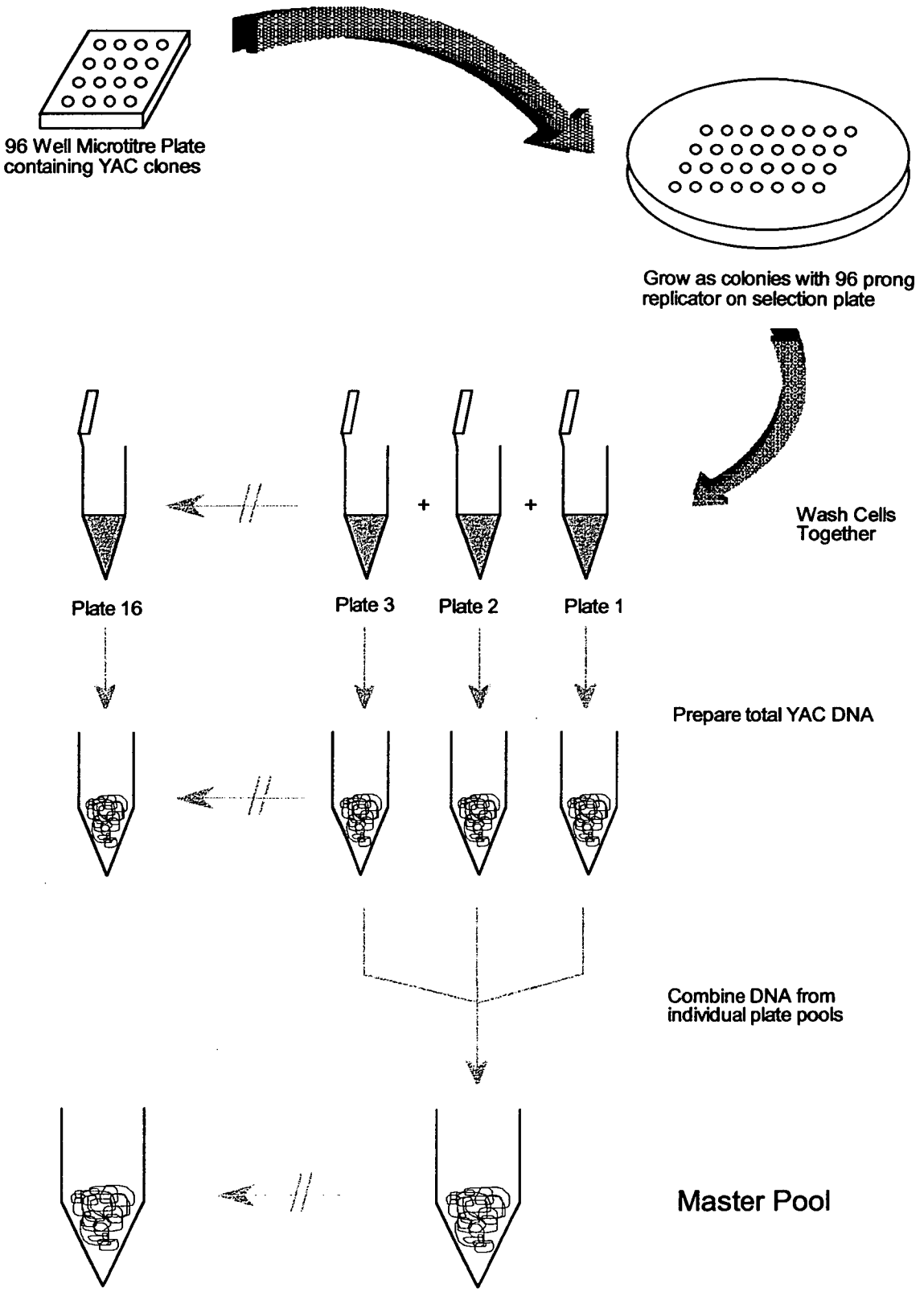
3.4.1.1 Preparation of YAC Vector.

The YAC vector was harvested from *E. coli* and purified on a CsCl gradient. It is known that some isolates of the vector in culture spontaneously delete small amounts of the TEL sequence, and to confirm that the vector used is suitable for cloning, a *Hin*dIII diagnostic digest is performed. Digestion of the YAC vector with *Hin*dIII results in four bands: 3.5 kb, 3.0 kb 1.9 kb and a doublet at 1.5 kb. If the 1.5 kb fragment resolves into two fragments of 1.3 and 1.2 kb then TEL sequences have been lost. The results showed that four bands were visible and that the plasmid was ready for use in a cloning ligation (data not shown).

To prepare the vector for cloning, 100 µg of pYAC4 at 0.43 µg / µl was simultaneously double digested with *Bam*HI and *Eco*RI, phosphatase treated with Calf Intestinal Alkaline Phosphatase (CIP) and purified by phenol/ chloroform extraction and ethanol precipitation. To test the efficiency of the dephosphorylation reaction test ligations were set up in the presence and absence of polynucleotide kinase. Three bands were seen in the kinase$^-$ (K^-) lane, whereas in the kinase$^+$ (K^+) multiple bands were seen above the 6 kb fragments and a smear seen below the 1.5 kb fragment (data not shown). The extracted, digested and CIP treated vector was stored at 4°C ready for use in the ligation reaction

3.4.1.2 Partial Digestion of HB3 DNA.

Parasites were grown in culture as described. On reaching a parasitaemia of 7 - 8%, comprising mostly schizont stages with some ring stages, parasites were harvested and the DNA extracted as described. Samples of HB3 DNA to be used in the preparation of the YAC library were


96 Well Microtitre Plate containing YAC clones
Grow as colonies with 96 prong replicator on selection plate
Wash Cells Together
Plate 16
Plate 3
Plate 2
Plate 1
Prepare total YAC DNA
Combine DNA from individual plate pools
Master Pool

examined on a 24 hour Pulsed Field Gel to determine chromosome integrity. Results indicated little chromosomal degradation.

Preliminary pilot digests were performed as described in Section 2. Due to the comparatively small size of the *P. falciparum* genome (~26 Mb) and the small size of the chromosomes (0.65 - 3.5 Mb, Triglia *et al.*, 1992; fig 3.2), large insert YAC's i.e. over 500 kb would yield little useful map information for the majority of the chromosomes. It was decided to attempt to isolate two YAC sizes : 75 - 150 kb and 150 - 200 kb. Initial partial digests on 3mm x 5mm slices showed that the optimal digest conditions yielding fragments in the desired size range lay around 0.5U *Eco*RI. Secondary partial digests with an enzyme range from 2U to 0.05U refined the digest parameters. As shown in figure 3.3a., the results of the secondary pilot digest showed that 0.2U of *Eco*RI digesting for 2 hours yielded a smear of fragments ranging from 25 - 300 kb

3.4.1.3 Size Fractionation of Recombinant YAC's.

All manipulations of *P. falciparum* DNA were carried out in agarose, with liquid DNA only being used immediately prior to transformation. Large scale partial digests were carried out in sterile plastic Eppindorfs containing 2x 3mm slices of HB3 DNA in a 760 µl reaction with 0.4U of *Eco*RI. Ligation of YAC arms to digested insert DNA occurred in agarose by allowing diffusion of ligase and YAC DNA into the block, and allowing the ligation to occur over a 16 hour period. Samples were loaded into the central 6 lanes of a 1.2% LMP agarose gel and separated out under pulsed field conditions selected for resolution of fragments of up to 300 kb. The results are shown in figure 3.3b. The bands containing recombinant YAC's in the desired size range were excised as described in section 2.1.4.3 and were stored in TE ^{50}. Half of each size band slice was used per transformation. This protocol allows the generation of recombinant DNA ready for transformation in three days.

3.4.1.4 Transformations and Preliminary YAC Clone Characterisations

The transformation procedure was carried out as described in section 2.1.5.2. and is a modification of that described by Burgers and Percival (1987). When the successful yeast transformation was performed, the host AB1380 was grown by culturing a single colony from a freshly growing YPD plate in 10 ml YPD with angled shaking at 30°C for 20 hours. 200 µl of culture was inoculated into 200 ml fresh YPD broth and incubated at 30°C for 27 hours. 100 µl of this culture was used to inoculate 100 ml of fresh YPD and this was incubated overnight at 30°C. The starting cell OD as determined by the OD600 nm of a 1/ 10 dilution of the culture was 0.286, higher than the 0.1- 0.2 range quoted (Monaco et al., 1993).

The gel slice containing recombinant, size- selected YAC's was agarase treated for 3 hours. The lyticase reaction was allowed to proceed for 15 min, taking the reaction to an estimated 70 - 80% spheroplast formation, as determined by SDS lysis. The cell count taken by haemocytometer revealed an

Figure 3.2. The Karyotype of *P. falciparum*

Intact chromosomal preparations of two *P. falciparum* clones, 105/ 12 and 105/ 11, isolated from the Asar region of the Sudan were resolved on Pulsed field gels. Chromosomes are indicated by numbers. Polymorphism is observed between the two isolates in chromosome 4.

A. The smaller *P. falciparum* chromosomes were resolved with Pulsed field gel conditions of 22 hours, 120seconds; 26 hours, 180 seconds; 24 hours 300 seconds in a 0.8% IBI agarose gel at 130V in 0.5X TBE.

B. The larger chromosomes were resolved with Pulsed field conditions of 22 hours, 120 seconds; 26 hours, 180 seconds; 24 hours, 300 seconds; 24 hours, 260 seconds in a 0.8% IBI agarose gel at 130V in 0.5X TBE.

A

B

average of 45 cells per 5 grid squares for a 1/ 50 spheroplast dilution. The total number of cells is calculated as follows:

$$Cell_{tot} = No.\ Cells \times 5 \times 10^4 \times 50 \times Vol_{tot}$$

where: $Cell_{tot}$ = total number of cells in suspension

No. cells = Average number of cells counted in 5 grid squares

5 = Factor to calculate the number of cells in 25 grid squares

10^4 = dilution factor to estimate number of cells per mL

Vol_{tot} = Volume of cell suspension, yielding total cell number

Using the above calculation, the volume of STC required to yield 3 x 10^8 cells/ mL was calculated to be 20mL. Through error, the spheroplast suspension was diluted down to 1.5 x 10^8 cells / mL. To compensate for this, DNA and spheroplasts were allowed to interact for 25 min. After pouring, the transformed spheroplasts were incubated upside down for 3 days, after which time a transformation efficiency of approximately 5 x 10^5 clones per µg was seen, based upon the results of a control transformation with Ycp50 (Hieter *et al.*, 1985). This figure is in good agreement with the efficiency range of 2 - 8 x 10^5 clones per µg quoted (Monaco et al., 1993)

Several random clones were picked, cultured and chromosomal blocks were prepared. The YAC's were separated out on an overnight PFG. The results are indicated in figure 3.4. The smallest size of the yeast chromosome is 220 kb (NEB), and as all YAC's were size selected to be have a maximum insert size of 195 kb, no YAC larger that the smallest yeast chromosome should be visible. The results show individual YAC's of various sizes lower that 220 kb, identifiable as chromosome bands on a Pulsed field gel. From this transformation, 800 Low Insert colonies and 250 High Insert colonies were picked, in addition to 450 YAC's constructed from previous transformation attempts. YAC's were picked into individual wells of a microtitre plate. 16 plates were filled, 11 plates containing low insert and 4.5 containing high insert YAC's. Initial estimates of the Library indicated a size of approximately 1500 clones.

On transferring YAC plates to AHC selection plates for DNA pool construction a number of YAC clones did not grow. The reason for this is unclear, as in selecting for recombinant YAC's, clones had been successively cultured onto (-U) then (-U-T) plates. AHC is a double dropout medium for the selection of both YAC arms. Loss rates varied from approximately 50% to 10% per plate. Plates showing YAC clone loss were replica plated onto AHC and YPD plates. In all cases, each clone grew on YPD, but clone losses were seen on AHC. In addition, all clones on YPD were red, indicating that either the SUP4 gene was inactive, or was not present. Chromosome block preparation from colonies growing on YPD but not AHC as well as true YAC's (growing on both) revealed that the suspect colonies did not

Figure 3.3

A. Pilot digest of *P. falciparum* strain HB3.

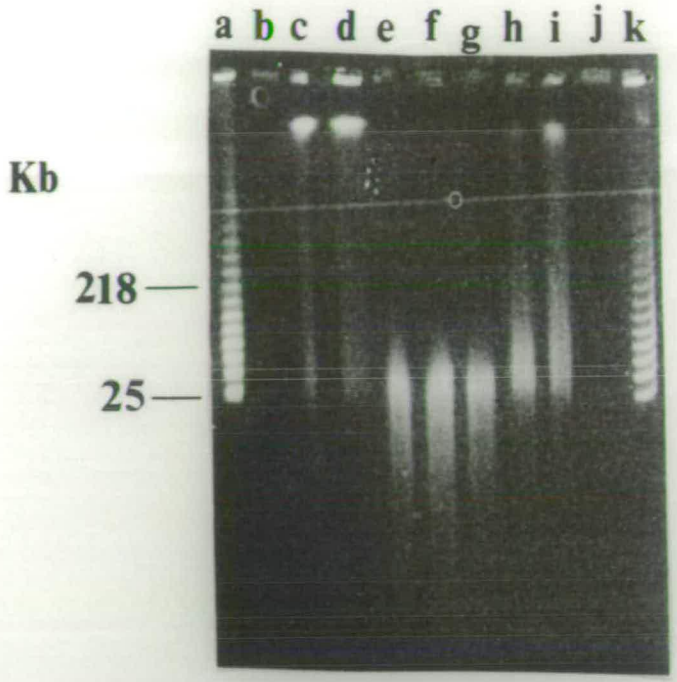
One half of a 2mm sliver of HB3 DNA embedded in agarose was digested with *Eco*RI essentially as described in materials an methods. The DNA was separated on a 1% Rapid Agarose Pulsed Field Gel in 0.5X TBE over 22 hours @ 150v, 20 - 40s ramped.

Lanes a + k : NEB Mid- range PFG marker II
Lane b: blank
Lane c: Negative control (-enzyme)
Lane d: Negative control (not digested)
Lane e: 2U *Eco*RI
Lane f: 1U *Eco*RI
Lane g: 0.5U *Eco*RI
Lane h: 0.2U *Eco*RI
Lane I: 0.1U *Eco*RI

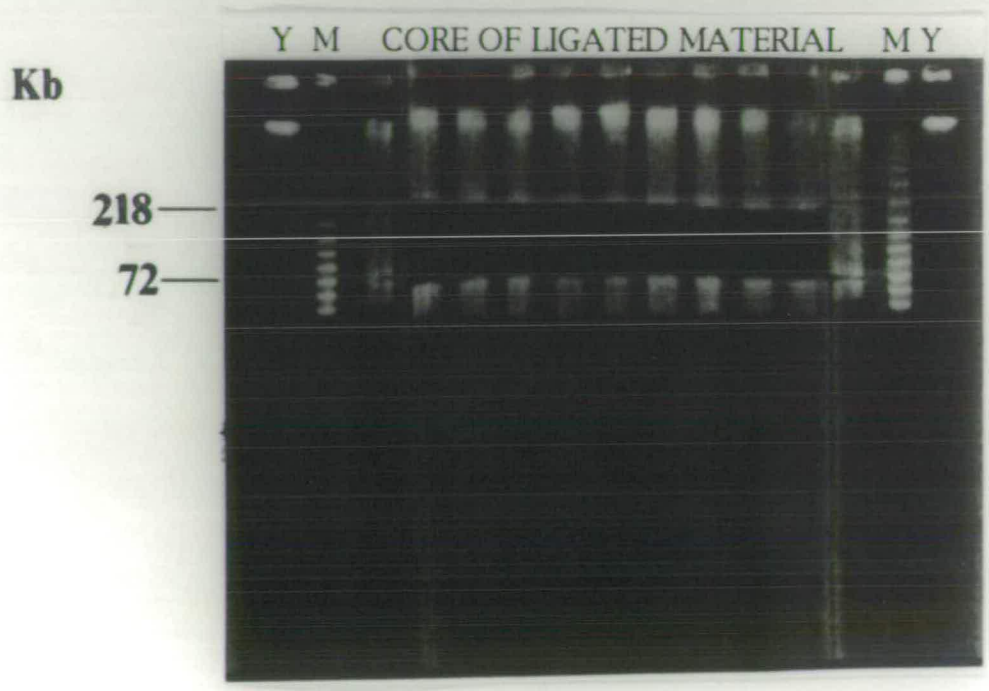
B. Size selection of recombinant YAC's

All digest products were loaded onto the central 8 wells of a 15 well 1.2% LMP agarose Pulsed Field Gel. Gel separation conditions were 16 hours @ 160V with a 20s pulse time. The central core lanes were cut out and the flanking lanes were stained. These were aligned with the unstained core and used to cut out recombinant YAC's of a defined size range. The figure represents the stained central core remnant aligned with the flanking lanes to indicate the size of fragments which have been removed. Y corresponds to Yeast Chromosome PFG marker (NEB) and M corresponds to NEB Mid- range PFG marker II.

A


a b c d e f g h i j k
Kb
218
25

B


Y M CORE OF LIGATED MATERIAL M Y
Kb
218
72

posses artificial chromosomes (data not shown). How certain colonies could survive double selection screening and then fail to grow when a similar selection round is subsequently applied is unknown. Final estimates indicate a library size of approximately 1100 viable clones

3.4.2. Genome Content Analysis of the Library

In order to facilitate clone identification, the PCR- based methods of Green and Olson (1990a) were adopted. Such a technique involves culturing all YAC clones from individual microtitre dishes onto individual AHC plates. Total liquid YAC DNA preps are made and are used as templates in a PCR reaction. To streamline STS isolation, YAC DNA Master Pools are constructed representing DNA samples from several pool sets

3.4.2.1 PCR Screening of YAC MasterPools

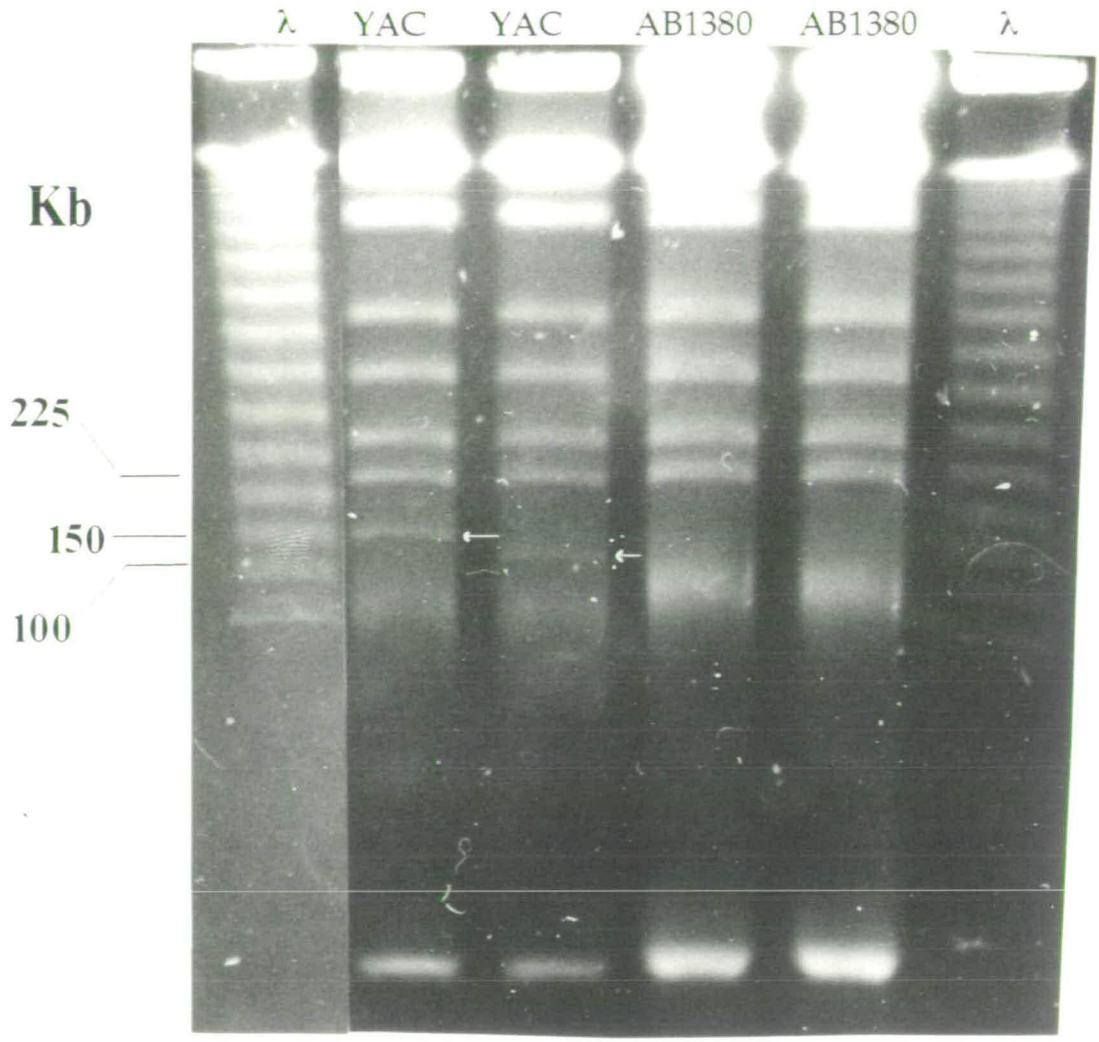
All PCR screening procedures followed essentially the same protocol. Liquid samples representing total YAC DNA from each 96- well microtitre plate were prepared as described in Chapter 2. UV spectrophotometric quantitation of each individual DNA pool revealed OD_{260} nm readings corresponding to between 3350 ng/ µl and 900 ng/ µl. Estimations of sample purity as determined by OD_{260}/ OD_{280} readings varied from 98.8% to 71.1%. As described in section 2.2.1.6, MasterPools at a final DNA concentration of 2.5 ng/ µl were constructed. A schematic screening strategy is illustrated in figure 3.1. To illustrate the YAC identification strategy, the Glycophorin Binding Protein (GBP- Lanzer *et al.*, 1992) will be used as an example. A PCR mix was set up as described in section 2.8.2 and the MasterPools were screened as follows:

10 ng Template DNA	4µl
GBP Primer 5'	0.2µl
GBP Primer 3'	0.2µl
2% Gelatin	0.2µl
dNTP's	0.2µl
10X PCR Buffer	2µl
Taq Polymerase	0.1µl
H_20	13.1µl

The reaction is sealed with a drop of sterile mineral oil. PCR conditions were as follows: (i) Denaturation stage of 94° for 45s; (ii) Annealing stage of 65° for 60s; (iii) Extension stage of 72° for 120s. Amplification was carried out over 31 cycles. Included in the reaction were H_20 and AB1380 negative controls and a HB3 positive control. As shown in figure 3.5a, two clean, single bands are visible in MasterPools 4 and 5, indicating that the GBP gene sequence is contained within the library in

Figure 3.4 YAC's of *P. falciparum*

Several YAC's were randomly picked an prepared as intact chromosomal agarose block preparations. AB1380 corresponds to the Yeast host strain and thus shows the background of yeast chromosomes. YAC corresponds to two randomly picked YAC's which are visible as chromosome-sized bands below the plane of the AB1380 chromosomes, at size markers 150 Kb and 125 Kb. λ corresponds to concatomers of the bacteriophage λ, with each marker representing an interval of approximately 50 Kb. YAC's were separated on a 0.8% IBI agarose Pulsed Field Gel, 0.5X TBE @ 120V. Pulse times were 30 - 70 sec. ramp for 24 hours followed by 80 - 120 sec. ramp for 24 hours


λ
YAC
YAC
AB1380
AB1380
λ
Kb
225
150
100

a minimum of two copies at unidentified locations in plates 10 - 16. The definition of a clean signal is one which produces a single band without smearing in one or more of the test samples and the HB3 control, and causes no amplification of a corresponding sequence of the same size in yeast or water. If a signal in yeast is identified, but is of a different size to the *P. falciparum* STS, then this sequence can be judged acceptable for mapping, as differentiation between yeast and *P. falciparum* sequences is possible. If sequences of the same size are amplified, then this sequence is discarded as an STS.

3.4.2.2 Breakdown of MasterPools to Plate pools

To confirm that the GBP gene sequence identified by the PCR primers is located within one of the plates identified by the MasterPool signal, a second reaction was prepared using MasterPool D and each individual component i.e. plates 10, 11 and 12 as templates. The results in figure 3.5b indicate that again, a positive result is seen in MasterPool D, demonstrating protocol repeatability, but also a single band is seen in plate 10. There are no signals in plates 11 or 12. Consequently, if a YAC containing the GBP gene sequence was desired, subsequent searches would concentrate on plate 10. Plates 11 and 12 would be ignored. This also gives a further indication of the copy number of the STS within the library and gene sequence representation. However, it is unknown how many copies of the STS there are within plate 10. A complete row and column analysis or filter hybridisation approach would have to be used to determine this.

3.4.2.3 Results of YAC Library Analysis by PCR.

The results of screening the YAC library with sequences from random locations within the genome are given in table 3.2. Only PCR signals which conformed to the parameters established in section 3.4.2.1. are included in the table. It should be noted that six chromosome- 6 specific STS markers have been included. These markers have been constructed in the course of this thesis, and data concerning their characterisation will be presented in the following chapter. It should also be noted that marker 63/ 11R has been included. Although the marker appears to identify no YAC pool, the marker was rescued from a YAC identified in plate 11. PCR results show clean signals in all controls, as well as the YAC from which the marker was identified. Consequently, a signal in pool 4 should have been expected. It is assumed therefore, that the PCR conditions for library screening have not been refined for these particular primers, and this is why such data is included. 63/ 11R has no bearing as yet on the analysis of the YAC library.

3.4.2.4 Library Size Estimations

Of the 28 gene sequences used to screen the Library, 26 yielded positive signals conforming to the criteria established in section 3.4.2.1. This leads to an estimated genome representation of 93%. It should be noted that two gene sequences, TRAP and RAN, were not identified, although positive control

Figure 3.5. Masterpool analysis with Glycophorin Binding Protein primers.

A.

Amplification of the GBP gene sequence from within the masterpool templates constructed for rapid library screening and information content analysis. Primers were annealed at 65°C and extended for 120s. 10 µl of PCR product was run on a 1% IBI agarose gel in 1X TAE. The results clearly show amplification signals in pools 4 and 5, which correspond to the control signal.

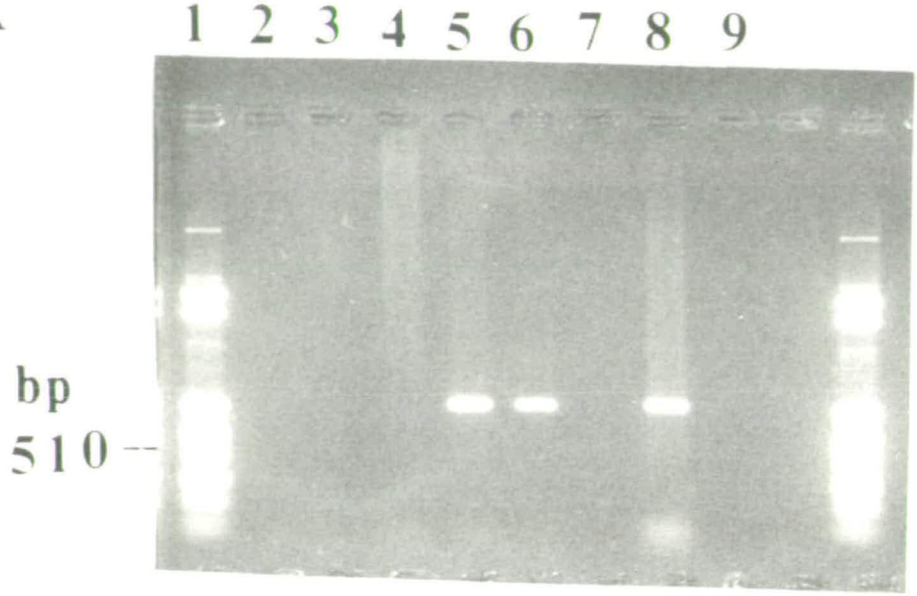
Lane 1: DNA marker (pHLA digested with *Hinf*I)
Lane 2: MasterPool 1 (contains DNA from Plate Pools 1, 2 and 3)
Lane 3: MasterPool 2 (contains DNA from Plate Pools 4, 5 and 6)
Lane 4: MasterPool 3 (contains DNA from Plate Pools 7, 8 and 9)
Lane 5: MasterPool 4 (contains DNA from Plate Pools 10, 11 and 12)
Lane 6: MasterPool 5 (contains DNA from Plate Pools 13, 14, 15 and 16)
Lane 7: negative control - water template
Lane 8: positive control - HB3 template
Lane 9: negative control - AB1380 template

B.

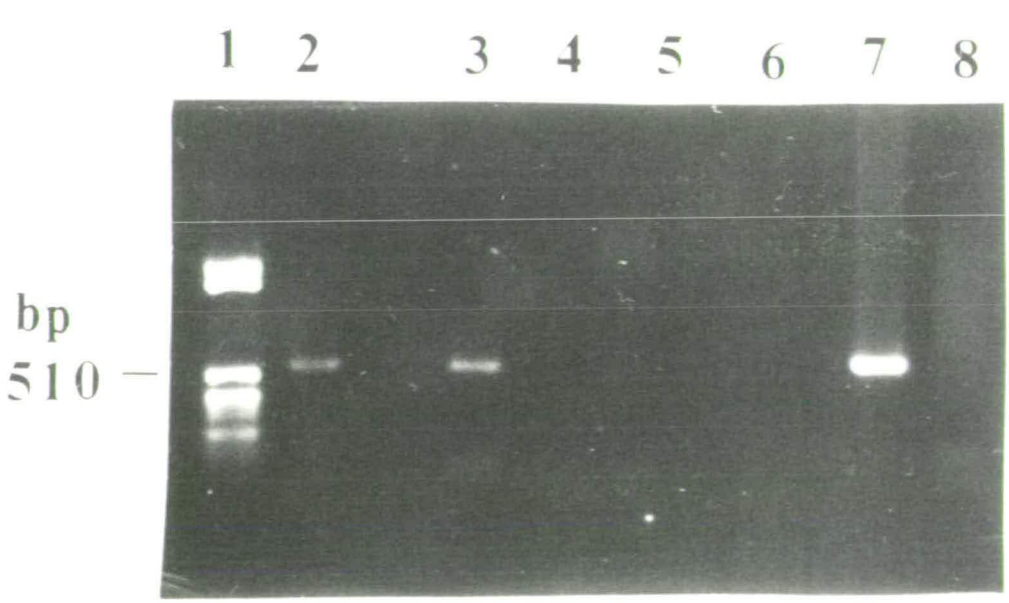
Verification that the masterpool contains the GBP sequence. MasterPool 4 and the Plate pool components were used in a repeat amplification to see wether the screening of the masterpool could in fact identify the plate pool containing the gene sequence. The results show a signal in pool 10. Amplification and electrophoretic conditions are as for figure **A.**

Lane 1: DNA marker
Lane 2: MasterPool 4
Lane 3: Plate Pool 10
Lane 4: Plate Pool 11
Lane 5: Plate Pool 12
Lane 6: negative control - water
Lane 7: positive control - HB3
Lane 8: negative control AB1380

A



B



signals were seen. This could have been due to one of two factors: firstly, by chance, the genome region may not be represented in the library. Alternatively, it is possible that such loci are subtelomeric i.e. lie immediately adjacent to the subtelomeric *rep*20 sequences (Corcoran *et al.*, 1988). Consequently, during digestion such sequences lie on fragments terminating with a single *Eco*RI site, and will thus fail to be included in this standard library. In order to be cloned, a telomere specific library would need to be constructed, using a telomere deleted YAC vector such as pYAC- Neo- Not (Cross et al., 1989), whose function is described in section 1.3.2.7.

In addition, an estimation could be inferred from the approximate size of the clones. the Library consists of 16 x 96 well microtitre plates, with 11 plates containing Low Insert clones of average size ~100 kb and 5 plates containing high insert clones of average insert size ~150 kb. Thus, including loss of YAC's, an estimate of 300 High Insert YAC's and 800 low insert YAC's corresponds to a total amount of cloned DNA of 125 Mb. Estimates of the genome size of *P. falciparum* range from 26 Mb (Gu *et al.*, 1990) to 30 Mb (Wellems *et al.*, 1987). Based on these estimates, the HB3 library described here has a 4 - 5 fold genome coverage, corresponding to a 4 to 5 times redundancy for single copy genes. Such a result is consistent with previous estimates of Library complexity, where libraries with 3 - 4 genome equivalents provide an ~95% probability of finding a given single copy sequence (de Bruin et al., 1992; Riley et al., 1992)

3.4.3. Information Analysis of the Library.

3.4.3.1. Comparison of Cloned YAC DNA to Genomic DNA

Although YAC libraries are known to faithfully copy genomic information, based on a correlation of clone maps with genomic maps (Burke *et al.*, 1987; Green and Olson, 1990; de Bruin *et al.*, 1992), it was decided to confirm that such a correlation could be made with the HB3 library. In order to determine the accuracy of information replication between genomic and cloned DNA, YAC clones identified by the Histone 2A marker were digested with *Eco*RI, *Hin*dIII and *Bam*HI. The results in fig 3.6a clearly show that the same restriction fragment is identified in two different YAC's containing the H2A locus, and that the fragment is the same size as the genomic restriction fragment. Identification of the Histone 2A YAC's will be discussed in chapter 4.

3.4.3.2 Stability of Cloned *P. falciparum* DNA

The major advantage of the YAC system has been the ability to maintain large inserts of cloned DNA in a stable form. YAC clone instability has been well documented (Neill *et al*, 1990; Vilageliu and Tyler- Smith, 1992). Instabilities such as co- transformation events (Zuo *et al.*, 1992),

Table 3.2 Genome representation of the HB3 YAC library

The HB3 YAC library was screened as described in the text. PCR primer sequences are quoted in materials and methods. PCR conditions are available from the relevant references. 10 µl of PCR product was loaded onto an agarose gel and separated for 30 - 120 min, depending upon fragment size. +/ - corresponds to the presence or absence of single bands after visualisation by EtBr staining and UV illumination. PCR primers which could not amplify the HB3 genomic DNA control, amplified the AB1380 negative control or fielded multiple bands were not included. H2/9L, H2/9R, 9R/9L, 63/11R, 87/5R and 5R/6R correspond to chromosome 6 specific STS markers that have been constructed in the course of this thesis. Sequence information and PCR conditions will be presented in chapter 4.

GENE	CHR.	YAC DNA MASTER POOLS					HB3	H20	AB1380
		1	2	3	4	5			
λwt 70	2	+	+	-	+	+	+	-	-
MSP2	2	-	-	+	-	+	+	-	-
CS	3	-	+		-	+	+	-	-
DHFR- TS	4	+	+	-	-	-	+	-	-
POL α	4	+	-	-	+	+	+	-	-
ATPase 1	5	-	-	-	+	-	+	-	-
ATPase2	5	+	+	-	+	+	+	-	-
MDR	5		-	+	+		+	-	-
ASPase	6(?)	-	+	+	+	-	+	-	-
H2/ 9L	6	-	+	+	-	+	+	-	-
H2/ 9R	6	+	+	+	+	+	+	-	-
9R/ 9L	6	+	+	+	+	-	+	-	-
63/ 11R	6	-	-	-	-	-	+	-	-
87/ 5R	6	+	+	-	-	+	+	-	-
5R/ 6R	6	-	+	+	-	-	+	-	-
H2A	6	-	-	+	+	-	+	-	-
H3	6	-	-	+	+	-	+	-	-
pSC11.63	6	-	-	-	+	-	+	-	-
pSC11.87	6	-	+	-	-	-	+	-	-
MSP 1	9	-	+	+	+	+	+	-	-
β- tubulin	10	+	+	-	+	+	+	-	-
GBP	10	+	-	+	+	-	+	-	-
GLURP	10	-	-	-	+	+	+	-	-
POL γ	10	+	+	-	-	+	+	-	-
RAN	11	-	-	-	-	-	+	-	-
ATPase3	12	+	+	-	+	-	+	-	-
PCNA	13	+	+	-	+	+	+	-	-
TRAP	13	-	-	-	-	-	+	-	-
CAL	14	-	-	-	+	+	+	-	-

deletions (Albertsen *et al.*, 1990), apparently unclonable sequences (Bates *et al.*, 1992) and chimerism (Green *et al.*, 1991) have been reported. As a preliminary assessment of the newly constructed YAC library, it was necessary to determine whether randomly picked clones would demonstrate insert stability when cultured over multiple generations. As the cloning system is an artificial chromosome system, it would be expected that information would be accurately replicated and transmitted to daughter cells.

A YAC clone identified by STS pSC11.87 (see chapter 4) was used in the culture experiment. The results of the culturing are given in table 3.3. The generation count was based on the assumption that one cell budding to yield two is generation 1, budding to give 4 is generation 2, budding to give 8 is generation 3 and so on. Samples were taken at 0 and approximately 38, 70 and 110 generations to give an indication of insert stability over 100 generations.

3.4.3.2.1. Genome Level Stability

The results in fig 3.6b. demonstrate that each undigested YAC clone is of the same size, with the slight apparent differences possibly being due to blotting or gel concentration artefacts. This indicates that, in this section of the genome, no major deletions or rearrangements are being carried out. This would have been visible as bands of different or decreasing size on the PFG blot.

3.4.3.2.2 Chromosome Level Stability

To investigate clone stability at the chromosomal level, YAC's were digested with *Sac*I, an enzyme known to cut only once in that clone (see chapter 5). Partial digests of all YAC's were separated on a pulsed field gel and probed with YAC fragment specific to the right arm of the vector. As can be seen in fig 3.7a, in all cases 15U and 3U of enzyme effectively cuts the fragment to produce a single band identified by the right arm probe, whereas 0.3U of enzyme does not cut at all. As the digest fragments are of the same size, this means that no apparent rearrangements, such as duplications or deletions which could be identified as an RFLP, have been introduced as a result of continuous culture

3.4.3.2.3 DNA Level Stability

In order to investigate stability at the gene level, YAC and genomic DNA was digested with frequently cutting enzymes and probed with the STS used initially to identify the YAC, as shown in fig 3.7b. YAC's were digested with *Eco*RI, *Hinf*I and *Nhe*I. The results clearly show that the YAC copy of the sequence after 110 generations in culture is the same size as the genomic copy. The results of the experiments carried out in this section indicate that the cloned YAC DNA in this library is an accurate representation of the genomic sequence and is stably inherited from parent to daughter cells.

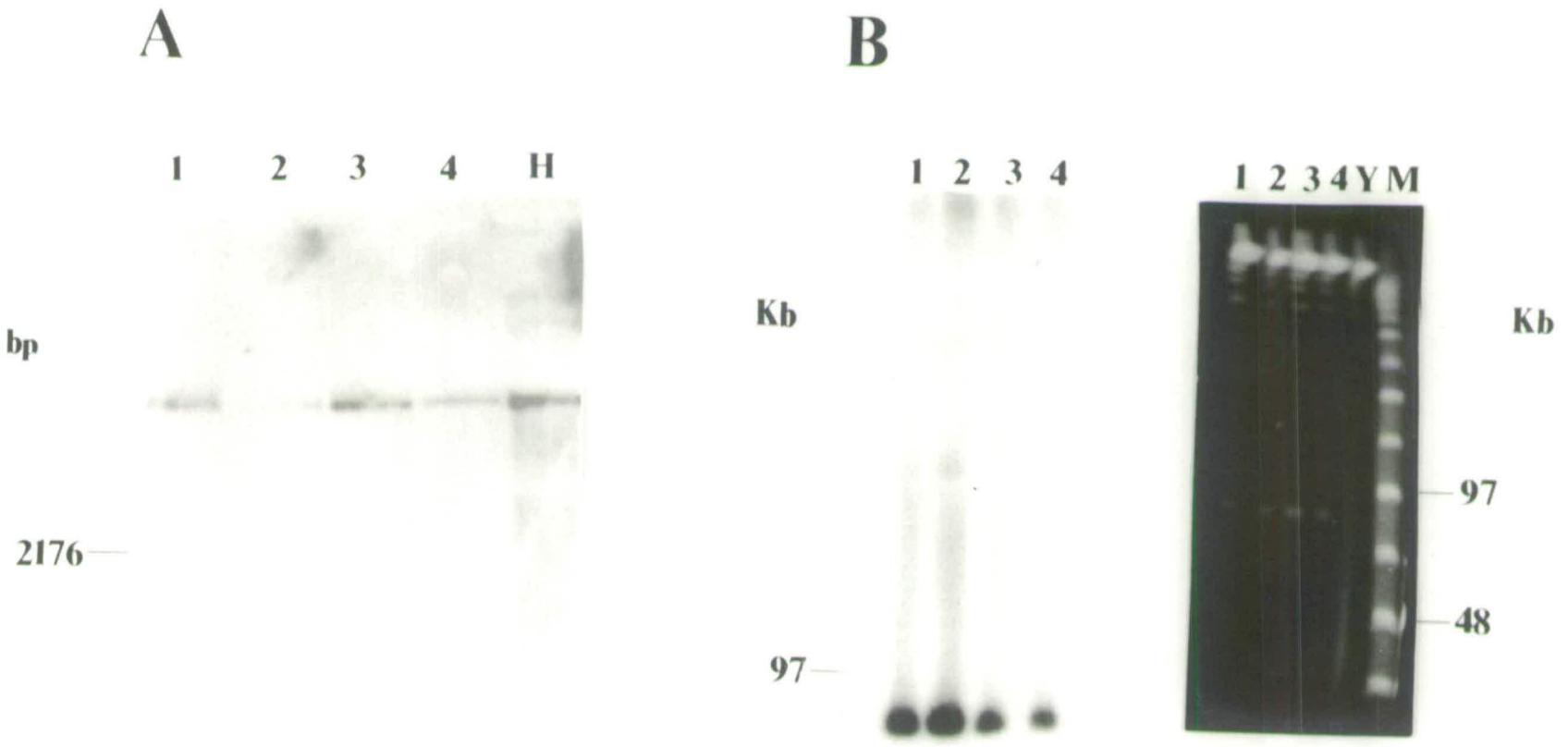
Figure 3.6

A. YAC's containing the H2A gene sequence were digested overnight to completion with 10U *Eco*RI and separated on a 1% IBI agarose gel for 15 hours @ 50V. The gel was transferred to Hybond N^+ ™ membrane and probed with ^{32}-P labelled H2A probe amplified from HB3. ***A*** and ***B*** represent different colony preparations of the same YAC clone.

Lane 1: YAC clone Y10/ H2 - ***A***
Lane 2: YAC clone Y10/ H2 - ***B***
Lane 3: YAC clone Y9/ H2 - ***A***
Lane 4: YAC clone Y9/H2 - ***B***
Lane 5: positive control -HB3

B. YAC samples obtained from various time point during the continuous culture experiment were separated on a 0.9% IBI agarose PFG over 24 hours @ 170V with a 2 - 12 sec ramp. Both the gel and the corresponding blot are presented. The gel was transferred to Hybond N^+™ membrane and probed with a YAC left arm probe.

Lane 1: YAC sample O (corresponding to 0 cell generations)
Lane 2: YAC sample A (corresponding to 40 cell generations)
Lane 3: YAC sample B (corresponding to 70 cell generations)
Lane 4: YAC sample C (corresponding to 110 cell generations)
Lanes Y and M correspond to NEB yeast marker and Mid- range PFG marker II, respectively


A
1
2
3
4
H
bp
2176
B
1
2
3
4
Kb
97
1 2 3 4 Y M
Kb
97
48

TIME	Sample	T_{int}	N_i	N_f (tot)	N_f / N_i	GENS	TOTAL GENS.
1	O	24	6.1×10^5	3×10^7	50	6	6
2		48	3×10^4	7.9×10^8	2.6×10^4	13	19
3		24	7.9×10^5	5×10^7	60	6	25
4	A	24	5×10^4	3.15×10^8	6.3×10^3	13	38
5		24	3.15×10^5	2.35×10^8	750	10	48
6		24	2.35×10^5	5.45×10^8	2.3×10^3	12	60
7	B	24	5.45×10^5	3.6×10^8	660	10	70
8		96	3.6×10^2	4.4×10^8	1.2×10^6	21	91
9		24	4.4×10^5	1.6×10^9	360	9	100
10	C	24	1.6×10^5	2.6×10^8	1.6×10^3	10	110

Table 3.3. Cell count data for the continuous culture experiment.

T_{int} corresponds to the time interval in hours between each cell count. N_i is the number of cells used to inoculate the culture volume, N_f corresponds to the final cell number in the culture; N_f / n_i corresponds to the number of divisions a single cell has undergone within that culture; gens corresponds to the number of generations assuming symmetric budding; total gens corresponds to the cumulative number of cell generations undergone by a single cell from the initial inoculation.

3.5 Discussion.

Yeast artificial chromosomes have the major advantage in genome cloning of the ability to maintain large inserts of DNA in a stable form. This chapter describes the construction of a 4/ 5 fold genomic YAC library of the *P. falciparum* isolate HB3 and the preliminary library assessment strategies that were performed before genomic analysis could occur.

In the construction of the library, the protocol followed (Monaco et al., 1993) recommended two size fractionation steps: a fractionation of partially digested DNA followed by a size selection of recombinant YAC's. The protocol described here represents a refinement of the published protocol in that the first size selection step is omitted, resulting in the ability to prepare recombinant DNA for transformation in three days. By carrying out all manipulations in agarose, shear damage attributed to the low insert size of early libraries is avoided (Brownstein *et al.*, 1989). As stated, the critical step in the transformation is the spheroplast stage. If spheroplasting goes too far or not enough, then transformation efficiencies will be lower and yields of recombinant YAC's will be reduced. Close

Figure 3.7. Investigation of YAC stability over Multiple Generations.

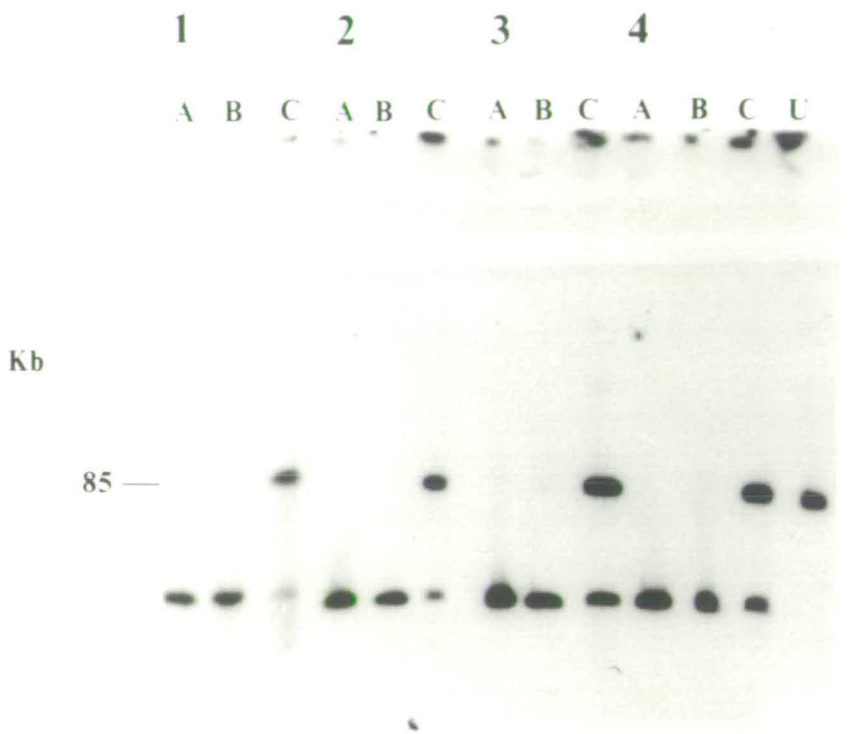
A. YAC clone Y5/ 87 was cultured over 100 generations and YAC's from various time points were digested with *Sac*I as described in Chapter 2 and separated on a 0.9% IBI agarose PFG over 23 hours @ 175V with a 2 - 12 sec ramp. The DNA was transferred to Hybond N^{+}™ membrane and probed with a YAC arm probe using the primers described in Chapter 2. For each YAC clone designated **1, 2, 3** and **4**, ***A*** corresponds to 15U of enzyme; ***B*** corresponds to 3U of enzyme and ***C*** corresponds to 0.3U of enzyme.

Lane Set **1**: YAC sample O (corresponding to 0 cell generations)
Lane Set **2**: YAC sample A (corresponding to 40 cell generations)
Lane Set **3**: YAC sample B (corresponding to 70 cell generations)
Lane Set **4**: YAC sample C (corresponding to 110 cell generations)
U: Undigested control.

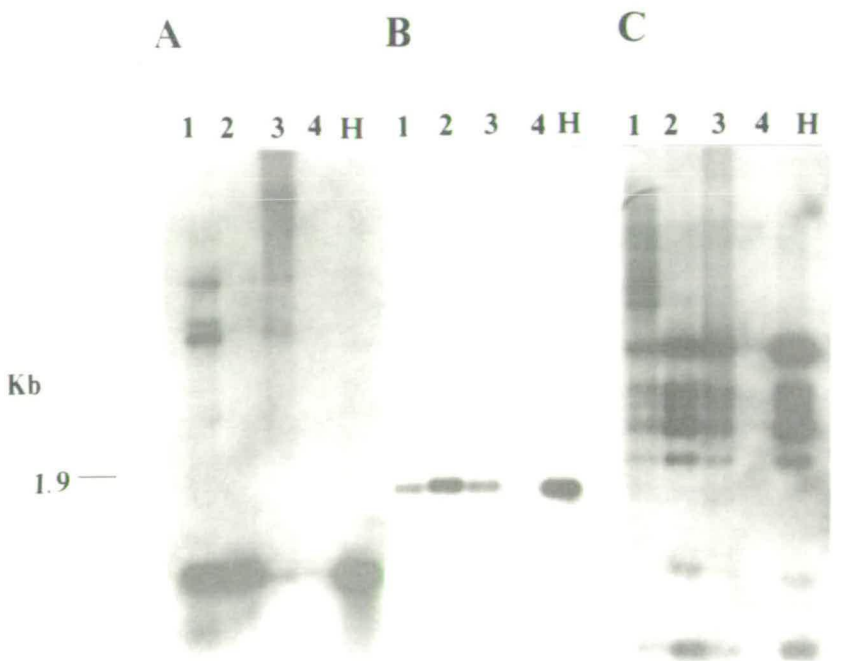
B. YAC's cultured over multiple generations were digested overnight with 20U of *Nde*I, *Hinf*I and *Eco*RI. The digest products were separated out on a 1% IBI agarose in 1X TAE for 15 hours @ 25V. **Set A** corresponds to *Eco*RI digests; **Set B** corresponds to *Hinf*I digests; **Set C** corresponds to *Nhe*I digests. Fragments were blotted onto Hybond N^{+}™ and probed with pSC11.87, amplified using the primers described in Chapter 2 from total HB3 DNA.

Lane 1: YAC sample A
Lane 2: YAC sample B
Lane 3: YAC sample C
Lane 4: YAC sample D
Lane H: positive control HB3

A



B



observance of spheroplast formation is required, with samples being taken as often as two minutes if necessary.

It was also observed that a number of clones failed to grow after selection on AHC plates after a prior round of -U -T selection. Pulsed field gel data and the colour assay on YPD clearly indicates that the clones do not posses YAC's or YAC plasmids. A possible explanation is that such clones contained both unligated YAC arms, which were subsequently lost in liquid culture within the microtitre dishes, leaving in essence untrasformed host cells. This would account for the ability of the cells to grow on rich YPD medium but their inability to grow on selection plates.

Before any genomic analysis can be performed on any new YAC library, it is necessary to be able to identify clones from within the library, to gauge the degree of genome representation and to investigate clone stability. A PCR protocol was developed for use with this HB3 library and the technique proved to be repeatable for different sets of primers. It is now possible to screen master pool templates quickly with primers to any new *P. falciparum* gene sequence which will both allow the isolation of such a sequence on YAC's and refine the estimation of genome representation. To ensure complete genome coverage, a future project would be to construct a telomere library of HB3 in a telomere deficient YAC. This would ensure that any gene sequences proximal to the telomeric sequences would be represented in the library. Similarly, estimates of single copy gene sequence representation should be 4 - 5 library copies per sequence, with a 95% chance of finding a given sequence. In order to improve these odds to 98/ 99%, the complexity should be increased to 10 genome equivalents. Such libraries have a proven increase in the number of positives per marker (Riley *et al.*, 1992). It would be prudent to carry out further transformations to increase the number of recombinant clones. thus increase the genome coverage. It would also be helpful to include a number of " mega- YAC's" i.e. with inserts of >300 kb, which would be particularly useful for contig assembly in the larger *P. falciparum* chromosomes.

CHAPTER 4.

CONSTRUCTION OF A YAC CONTIG OF *P. falciparum* CHROMOSOME 6 USING SEQUENCE TAGGED SITE MARKERS.

4.1 Abstract

Chromosome 6 specific YAC's were identified from within the HB3 YAC library by PCR. Screening involved first identifying the plate containing the YAC, then using a row and column PCR screening method to isolate the individual clone. Five YAC's were identified by chromosome specific markers. Inverse PCR was used to construct STS markers from the ends of the YAC's. Of the ten possible markers that could be isolated, seven were eventually constructed and sequenced and used to identify overlapping YAC's. Five new overlapping YAC's were identified, corresponding to a total of 750 kb of chromosome 6 DNA cloned in YAC's. Three new STS markers were rescued from the ends of these YAC's and used to screen the library. The location of these YAC's are presented.

4.2. Introduction.

The analysis of Yeast Artificial Chromosome clones is central to the study of the complex genomes of organisms such as man and *P. falciparum*. The large insert size that can be cloned in YAC's makes them the vector of choice for the construction of detailed, large scale physical maps of genomic DNA. The analysis of YAC's can be divided into three sections: (i) identification of YAC clones; (ii) assembly of overlapping clones into contigs and (iii) physical mapping of cloned DNA within a YAC. This chapter will deal with the identification and contig asembly of chromosome 6 specific YAC's.

YAC clone identification from within a library can be performed by either PCR or probe hybridisation. Hybridisation- based analysis involves culturing the YAC clones on either ordinary (Brownstein *et al.*, 1989) or super dense (Bentley *et al.*, 1992) filters, followed by colony lysis, immobilisation of DNA onto the membrane and hybridisation with radiolabelled probes. YAC identification by PCR is becoming the method of choice for most genome projects, as such an approach is non- radioactive, more rapid when searching for a single clone and facilitates a Sequence Tagged Site (STS) approach for contig assembly (Green and Olson, 1990a.).

Contig assembly is carried out by bidirectional walking from the ends of each YAC identified in a primary library screen. Novel PCR- based techniques exist to construct end clone sequences, such as Inverse PC, Vectorette PCR, *Alu*- Vector PCR and Junction Trapping. The objective of each technique is to generate sequence data from the extreme 5' and 3' ends of each YAC clone insert. PCR primers designed from this data can be used to rescue this STS sequence and then used to rescreen the library in a second level screen. Any new YAC's identified by this STS must therefore be overlapping with the immediate ends of the parent YAC, and this initiates bidirectional chromosome walking.

This study described the cloning and characterisation of YAC's specific to chromosome 6 of the malaria parasite *P. falciparum*. Previous mapping studies have shown that the only gene sequences known to map to the chromosome are Ag844, Histone 2A (Creedon et al., 1992), plasmids pSC11.63 and pSC11.87, and λW70. Due to the limited number of markers available and the lack of antigen gene sequences identified, no physical mapping studies had been performed on this chromosome (Triglia *et*

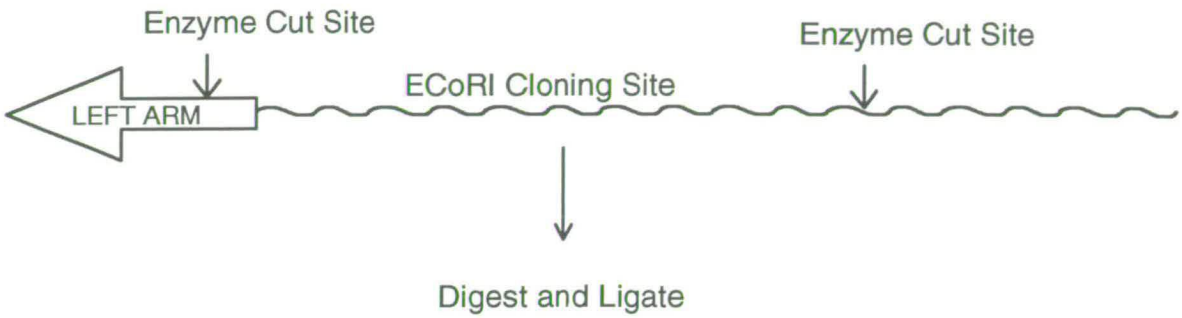
al., 1992). The chromosome is being mapped as part of the Wellcome Trust Malaria Genome Mapping Consortium, which aims to provide complete detailed physical maps of all 14 *P. falciparum* chromosomes (WTMGMC, 1995). However, the comparative lack of mapped markers for which PCR data exist necessitates a strategy whereby new YAC's are identified either from STS markers constructed *de novo* from YAC end rescues or from random, chromosome specific markers generated from isolated chromosome 6 DNA. For the puroposes of rapid information assembly, an STS- end rescue approach to contig assembly was employed.

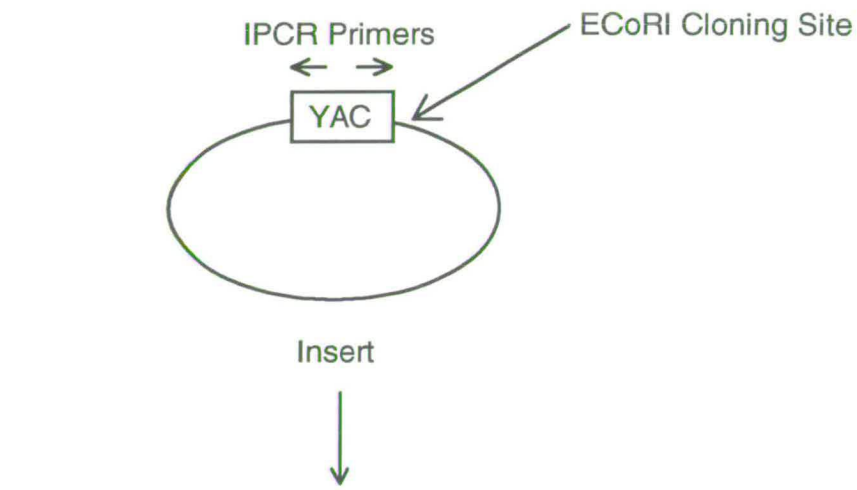
4.3. Strategies used for YAC Clone Identification and Contig Assembly.

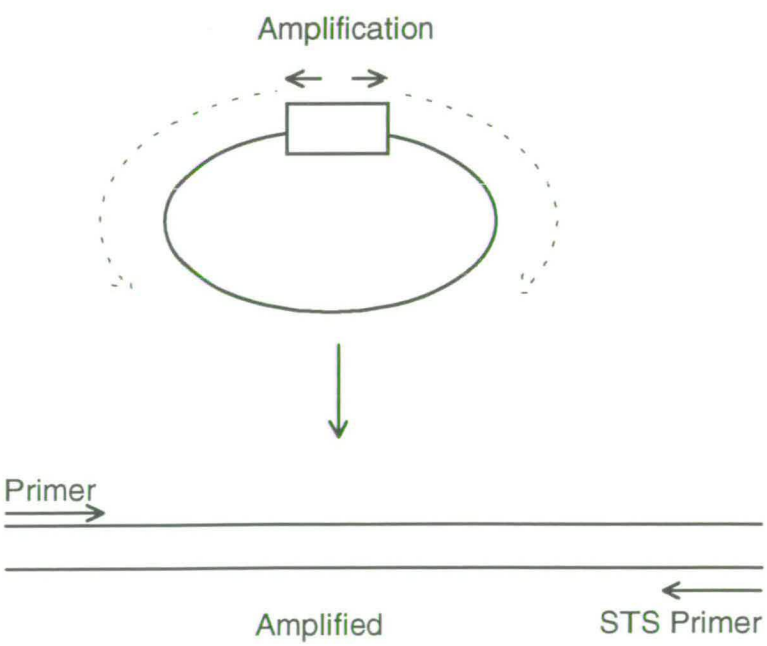
The plate location of chromosome specific YAC's were identified by screening the plate pools described (see Chapter 2) with PCR primers derived from the chromosome 6 STS's. Once a plate location had been assigned for a particular marker, the location of the YAC clone within that plate was determined using row and column PCR with a crude lysate DNA as a template. The schematic for the identification of clones by row and column PCR is given in chapter 2, figure 2.1. Row and column PCR results in a co- ordinate location for each YAC clone, given as a well location in rows A -H and columns 1 - 12. In order to confirm that the identified well (a) contains the YAC and (b) is not contaminated with other YAC's by cross contamination during library duplication, the YAC clone was cultured to obtain single colonies. DNA from several randomly picked colonies were used as templates in a PCR reaction. Positive clones were again cultured to single colonies, and DNA from these were again checked by PCR. Colonies screening positive were again deemed genetically pure and were used for mapping and contig assembly.

In order to rescue the ends of each YAC, the technique of Inverse PCR (IPCR), using primers designed by Arveiller and Porteous (1991) was employed (see fig 4.1). The success of such a technique depends on the presence of an enzyme cut site within the insert close enough to the vector to be amplified by PCR. Generally combinations of enzymes or techniques (such as vectorette PCR) are used to rescue as many YAC ends as possible. The strategy adopted here was to try to increase the number of enzymes available for cloning rather than develop Vectorette PCR technology. The recent publication of the entire sequence of pYAC4 (Kuhn and Ludvig, 1994) and the availability of the sequence in the EMBL database facilitated the identification of appropriate enzymes for use in IPCR.

IPCR rather than direct sequencing from the vector through into the insert was chosen for several reasons: direct sequencing would only generate 200 - 300 bp of sequence data, which depending upon the A+T content of the insert may be useless for primer design. IPCR products could be cloned and sequenced and 100 - 200 bp of sequence data from both the 5' and 3' ends of the PCR product could be obtained. This gives more scope for the construction of relatively G+C rich primers for the second level screen and allows for the construction of larger STS's. The G+C content of the IPCR fragment can also give an idea of whether or not the fragment is a coding or non- coding sequence. In addition, cloning of


Enzyme Cut Site
Enzyme Cut Site
LEFT ARM
EcoRI Cloning Site
Digest and Ligate


IPCR Primers
EcoRI Cloning Site
YAC
Insert


Amplification
Primer
Amplified
STS Primer

the IPCR product into a plasmid vector allows the fragment to be used as a radiolabelled probe, either by directly labelling from the vector or amplifying the insert followed by labelling. Level two screenings are carried out essentially the same way as primary screens: plate identification followed by clone identification and purification. IPCR has successfully been applied in the construction of YAC contigs in humans (Hirst *et al.*, 1991; Marchuk *et al.*, 1992) and *P. falciparum* (Rubio *et al.*, 1995).

4.4 Results.

4.4.1. Isolation of Single Chromosome 6 Specific YAC's From Within the Library.

4.4.1.1. STS Marker Histone 2A

Chromosome 6 is an unmapped chromosome. Very few genes have been positioned and there are only 6 anonymous phage or plasmid clones assigned to this chromosome (Triglia *et al.*, 1992). One of the genes that has been identified is Histone 2A (H2A, Creedon *et al.*, 1992). Histones are amongst the most conserved proteins known,with PfH2A demonstrating 54- 75% homology to the histones of fungi, plants, vertebrates and invertebrates. The gene is present in single copy, similar to the histone 2A genes from other lower eukaryotes (Creedon *et al.*, 1992)

To identify YAC's containing the H2A sequence, PCR primers based on the published sequence were designed. The 5' primer extends from position 34 to position 57; the 3' primer extends from position 392 to position 370. This predicts a PCR product size of 358 bp. Sequence of the primers used is given in table 2.1, PCR conditions are described in the text.

4.4.1.1.1. Location within the Library

All 16 plate pools were screened with the H2A primers. The results are shown in fig 4.2a. Positive signals are seen in pools 9 + 10 as well as the HB3 control. Faint residual signals are seen in all other pools as well as the AB1380 control. This may be ascribed to slight PCR amplification of a low homologous primer binding sequence within yeast genomic DNA due to the high degree of conservation of the H2A sequence. Homology studies based on comparisons between the *P. falciparum* DNA sequence and a yeast DNA sequence demonstrated 68.7% homology over the central 300 bp core region for both 400 bp sequences. The flanking 5` and 3` 50 bp regions were dissimilar, as the GCG software employed did not identify enough homology between these sequences to perform a comparison. The H2A primers used were designed from 5` 34 - 55bp and 3` 369 - 382 bp. The 3' primers should therefore be non homologous to yeast, however, the 5' primers could concievable bind slightly to genomic DNA. PCR conditions for the reaction were (i) denature at 94° for 60s; (ii) anneal at 55° for 60s; (ii) 72° at 60s for 30 cycles. Increasing the annealing temperature reduced the number of background bands, but did not

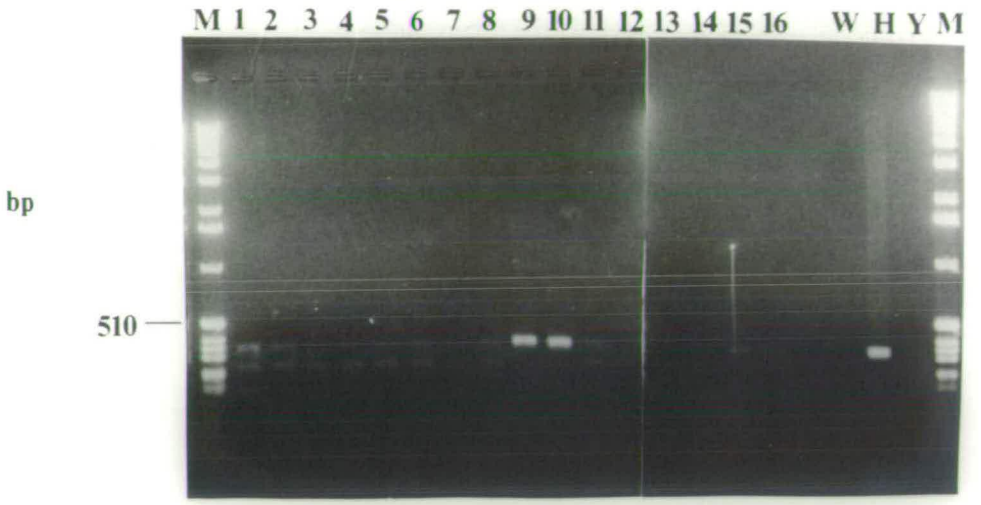
Figure 4.2. Screening the HB3 YAC plate pools with H2A Primers

A. All 16 HB3 YAC plate pools were screened with the H2A primers, under amplification conditions of 55°C annealing and 30 cycles. 10µl of PCR product was loaded onto a 1% agarose gel for separation and subsequent visualisation after EtBr staining and UV illumination. A signal of ~350 Bp is revealed in plates 9 and 10.

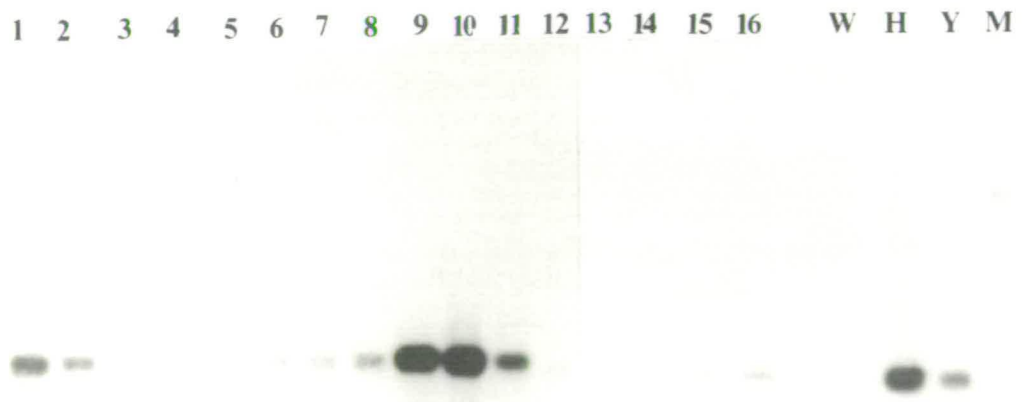
Lane M: DNA Marker BRL I Kb Ladder
Lanes 1 - 16: HB3 YAC library plate pools 1 - 16.
Lane W: negative control - water
Lane H: positive control - HB3
Lane Y: negative control - AB1380

B. Corresponding Autoradiograph of the above gel. DNA fragments were transferred to Hybond N^{+TM} and probed with the H2A product amplified from total HB3 DNA using the primers described in Chapter 2. The filters were exposed to Kodak X- omat AR film overnight and developed.

A


M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 W H Y M
bp
510

B


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 W H Y M

completely eradicate the faint signal in the AB1380 control. Based upon the repeatability of this signal in plates 9 and 10, these were taken as being the location of the H2A gene within the library. The blotting and probing of the filter with the H2A gene sequence indicates a possible location in plates 1 and 2 (fig 4.2b), however these results were discarded in favour of the strong signals from plates 9 and 10.

4.4.1.1.2 Location within the Plate

The principle behind row and column PCR is illustrated in chapter 2, figure 2.1. The results in figure 4.3 demostrate the results of row and column analysis on plate 9 with the H2A primers. The conditions used were 94° 60s; 52° 60s; 72° 60s for 40 cycles. The results clearly show a positive signal in row A and in column 1. For row and column PCR, the number of amplification cycles had to be increased to 40, possibly due to inhibitory factors within the template DNA preparation. For plate 10, row and column analysis revealed positive signals in row A and column 11 (data not shown). These results indicate that only two copies of the gene are represented in the library. This is slightly lower than the predicted 4 fold frequency for single copy sequences based on a 4x library coverage and can be ascribed to either (a) statistical variance or (b) more clones than estimated had been lost.

4.4.1.1.3 Purification of H2A YAC's

To confirm that the wells containing the H2A YAC's had not been cross contaminated with other YAC's during library duplication, both YAC clones were cultured to yield single colonies and liquid DNA PCR templates were made from four random colonies. The results in figure 4.4 show that all four of the YAC's cultured from plate 9 contained the H2A sequence. Only two of the four YAC's cultured from plate 10 contained the H2A gene. One of the YAC's that screened positive, 10- 2 was again cultured to single colonies and four random templates for PCR were prepared. The results show that all four colonies contain the H2A gene, and one of these was selected for use in YAC mapping. The plate 9 YAC used for subsequent manipulation was designated Y9/H2. The plate 10 YAC used for subsequent manipulation was designated Y10/H2.

4.4.1.1.4 Blotting Studies on H2A YAC's

To confirm the presence of the H2A locus on the two identified YAC's, intact chromosome preparations in agarose were made of Y9/H2, Y10/H2 as well as two randomly chosen YAC's and the AB1380 yeast host strain. The results in figure 4.5 show a clear signal in both Y9/H2 and Y10/H2 and no signal in any of the other lanes. This confirms the PCR data in that both YAC's contain the Histone 2A gene. The estimated size of the YAC's are approximately 160 kb for Y9/H2 and 150 kb for Y10/H2. Thus accounting for the 11 kb of vector DNA, this corresponds to inserts of approximately 150 and 140 kb.

Figure 4.3. Row and column analysis with H2A primers

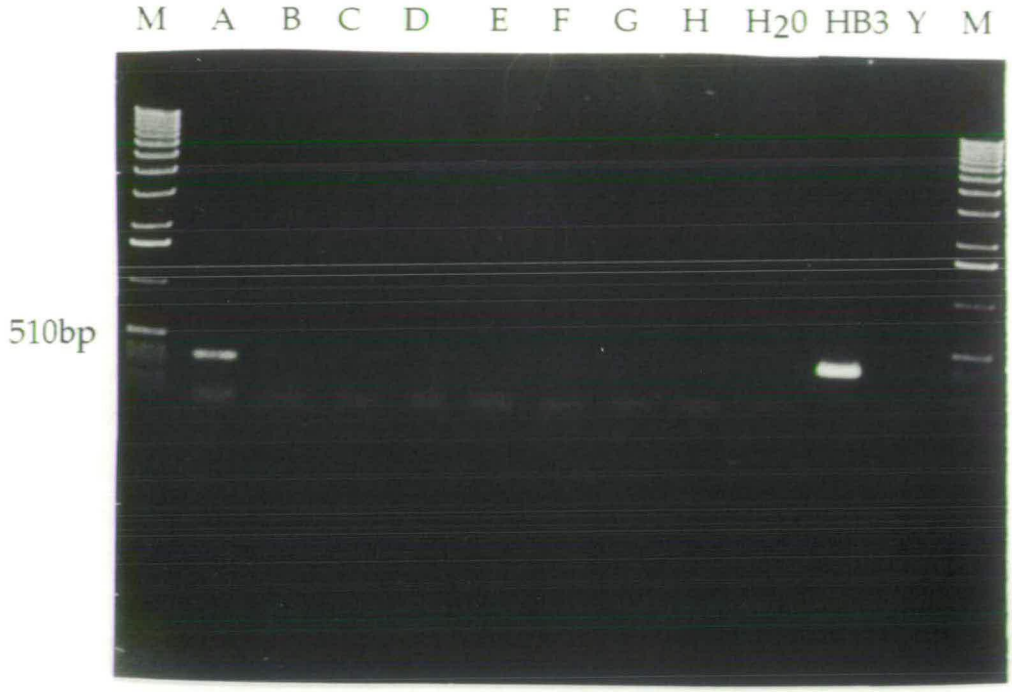
A. Results of amplification with H2A primers on DNA templates corresponding to the total DNA per row of the 96 well microtitre plate designated number 10. Amplification was carried out under annealing temperatures of 52°C for 40 cycles. 10 µl was loaded onto a 1% IBI agarose gel in 1X TAE and after separation, the products visualised after EtBr staining and UV illumination.

Lane M: DNA size standards - BRL 1 Kb Ladder
Lanes A - H: DNA templates corresponding to rows A - H
Lane H_20: negative control - water
Lane HB3: positive control - HB3
Lane Y: negative control - AB1380

B. Results of amplification with H2A primers on DNA templates corresponding to the total DNA per column of the 96 well microtitre plate designated number 10. Amplification conditions were as for section A. DNA fragments were visualised as described above.

Lane M: DNA size standards - BRL 1 Kb Ladder
Lanes 1 - 12: DNA templates corresponding to columns 1 - 12.

A



B

4.4.1.2. STS Marker pSC11.63

One of the anonymous markers assigned to chromosome 6 is the plasmid pSC11.63 (Triglia *et al.*, 1992). Sequence data was obtained for the plasmid and primers were designed (J. Foster, personal communication.) The primer sequence is given in chapter 2, Table 1. Library screening and YAC clone identification and purtification strategies are essentially as for Histone 2A.

4.4.1.2.1. Location within the Library

The HB3 YAC pools were screened with the pSC11.63 primers. An amplification signal of ~220 bp was seen in master pool 4 using PCR amplification signals of 94° 60s; 52° 60s; 72° 60s for 30 cycles of amplification and 10 ng of template and 0.5U Cetus Amplitaq™ polymerase. Screening of plate pools 10, 11 and 12 using the same PCR conditions and 4 µl of template revealed amplification of pools 11 and 12, with clean signals in the water HB3 and AB1380 controls. The YAC is therefore present in a minimum of two copies within the library.

4.4.1.2.2 Location within the Plate

Plates 11 and 12 were cultured and row and column DNA templates were prepared as described. PCR conditions were 94°, 60s; 52°, 60s; 72°, 60s for 40 cycles. The result of plate 11 showed clear signals in row E cloumn 8. The results of plate 12 showed clear signals in row C column 3. To confirm that location, the PCR was repeated on those four templates. A signal corresponding to the 220bp band of Boehringer DNA Marker VI was seen in all four lanes. The STS sequence is only present in two copies within the library, which is lower than the predicted frequency for this library.

4.4.1.2.3 Purification of the 11.63 YAC's

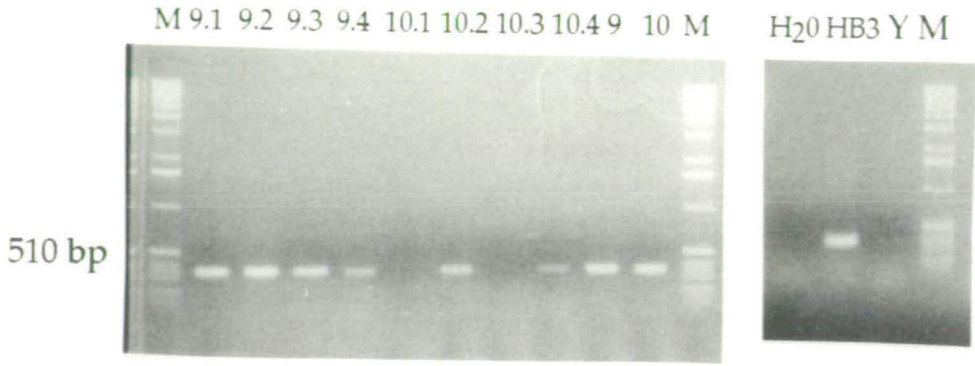
Both suspected YAC's were cultured to single colonies on AHC selection plates. Liquid DNA templates for PCR were prepared from three single colonies of each YAC. Each colony should have arisen from a single source cell and thus should just contain a single, genetically pure YAC. A PCR reaction on YAC samples 11.1 - 11.3 and 12.1 - 12.3 showed positive signals in all lanes, with no signal in the AB1380 or H_2O control lanes. PCR conditions were 94° 60s; 52°, 60s; 72°, 60s for 35 cycles with 1µl of a 1/ 40 template dilution and 0.5U Cetus Amplitaq™ polymerase. YAC's 12.3 and 11.1 were cultured on AHC plates to single colonies and these were used as YAC sources for further manipulations. YAC 11 was designated Y11/63 and YAC 12 was designated Y12/63. Blotting studies on both YAC's separated on Pulsed field gels for 18 hours at 170 Volts with a 10 - 20 seconds ramp indicate sizes of approximately 95 kb for Y11/ 63 and 80 kb for Y12/ 63.

Figure 4.4. Colony purification of YAC's

A. YAC's Y9/ H2 and Y10/ H2 were cultured to single colonies and used as templates in a PCR reaction. Amplification conditions were 58°C annealing temperature with 30 cycles of amplification. 9.1, 9.2, 9.3 and 9.4 correspond to DNA prepared from four single colonies of YAC Y9/ H2. 10.1, 10.2, 10.3 and 10.4 correspond to DNA prepared from four single colonies of YAC Y10/ H2. 9 and 10 correspond to the original YAC template prepared from the parent colony. H_2O, HB3 and Y correspond to water negative control, HB3 positive control and AB1380 negative control respectively. The DNA size standard used was the BRL 1 Kb Ladder.

B. Corresponding Autoradiograph of the above gel. DNA was transferred onto HybondN^{+}™ and probed with a H2A probe amplified form total HB3 DNA with the primers described in chapter 2, washed , exposed and developed as described in materials and methods.

A


M 9.1 9.2 9.3 9.4 10.1 10.2 10.3 10.4 9 10 M
H20 HB3 Y M
510 bp

B

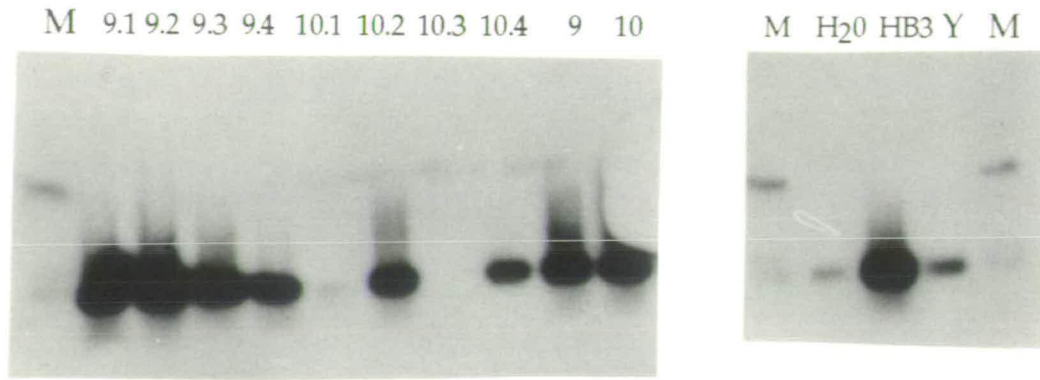
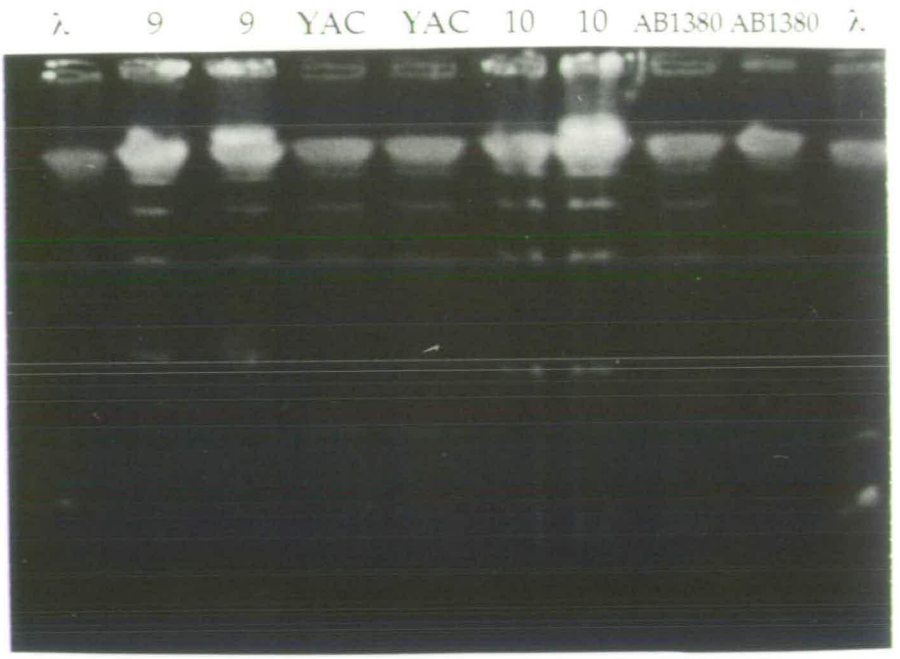

M 9.1 9.2 9.3 9.4 10.1 10.2 10.3 10.4 9 10
M H20 HB3 Y M

Figure 4.5. Southern blot of YAC PFG

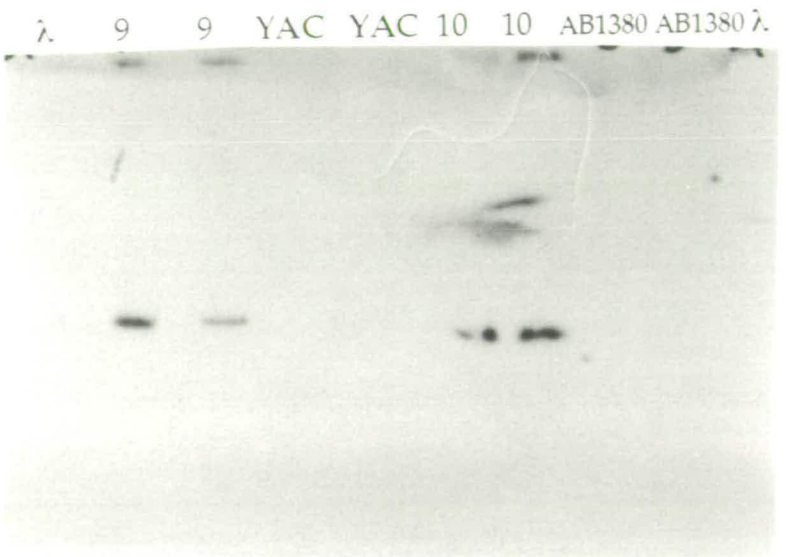
A. YAC's Y9/ H2 (Lanes 9) and Y10/ H2 (Lanes 10) were resolved from the background of the AB1380 chromosomes on a 1% IBI agarose PFG in 0.5X TBE for 20 hours @ 170V with a 10 - 20 sec ramp. Included were AB1380 chromosomal preparations so that the HB3 YAC's could be visualised against the yeast background and two randomly chosen YAC's as a control for blotting studies. The DNA size standard used was concatomers of bacteriophage λ. The gel was stained with EtBr, destained with water and photographed under UV illumination

B. Corresponding Autoradiograph of the above gel. The DNA fragments were blotted onto Hybond N^{+}™ membranes and probed with a H2A probe amplified from total HB3 DNA, using the primers given in chapter 2.

A


λ 9 9 YAC YAC 10 10 AB1380 AB1380 λ
150 Kb

B


λ 9 9 YAC YAC 10 10 AB1380 AB1380 λ

4.4.1.3 STS Marker pSC11.87

The second anonymous marker known to map to chromosome 6 is pSC11.87 (Triglia *et al.*, 1992). The marker was sequenced and PCR primers constructed (J. Foster, Personal communication). The sequence of the primers used is given in Table 2.1.

4.4.1.3.1 Location within the Library

The primers amplifying pSC11.87 were used to screen the HB3 Masterpools. PCR screening of 10 ng of template with amplification conditions 94° 60s; 52° 60s; 72° 60s for 30 cycles gave a clean signal of approximately 300 bp in Masterpool 2. Subsequent PCR reactions localised this to plate pool 5. Again, the number of positive clones identified is lower than the expected frequency.

Row and column DNA templates were constructed from plate 5, and used in a PCR reaction with the 11.87 primers under amplification conditions 94° 60s; 52° 60s; 72° 60s for 40 cycles. The results showed clean signals in row A and column 5. No signal was seen in the AB1380 control, and a clear positive signal was seen in the HB3 control.

4.4.1.3.2 Purification of the 11.87 YAC's

The suspected YAC clone was cultured on AHC plates at 30° C for three days and liquid DNA templates for PCR were prepared from four single colonies. The PCR showed that three of the four templates yielded positive signals, one (YAC 5.1) did not. One of the positive YAC's (designated YAC Y5/ 87) was cultured on AHC plates and used as a source of YAC DNA for future manipulations. Pulsed field gel analysis of the YAC indicated a clone size of approximately 85 kb.

4.4.2. Level One Sequence Tagged Site Marker Construction.

4.4.2.1. Inverse PCR results

The sequence of the YAC plasmid around the IPCR primer sites is indicated in figure 4.6. The concept of inverse PCR is based on two primers orientated in opposite directions. As indicated in figure 4.6a and 4.6b, primer pairs 372 and 373 for left arm rescues and 374 and 556 for right arm rescues are used. YAC DNA is digested with the enzymes *Taq*I, *Eco*RV and *Rsa*I for left arm rescues and *Nde*I, *Taq*I, *Rsa*I and *Eco*RV for the right arm. *Eco*RV, *Rsa*I and *Taq*I are published rescue enzymes (Areveiller and Porteous, 1992), but *Nde*I was first used in the course of this study. The location of primer 372 is 365 - 384 and primer 373 is 512 - 521. Because the primers are in the opposite orientation, the intervening 128 bp of YAC plasmid DNA is not amplified. Sequencing through from primer 373 will reveal the *Eco*RI cloning site of the YAC and will thus serve as a marker to identify the start of the *P. falciparum* insert. Sequencing through using primer 372 will reveal the enzyme cut site

Fig 4.6a. sequence of pYAC4 left arm showing the *Eco*RI cloning site and the Inverse PCR primer locations and sequence as well as the IPCR enzyme restriction sites and the location of the internal sequencing primer.

TTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACG 69
CAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 138

*Rsa*I *Eco*RV
TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCAT 207

CGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGG 276
<¬Internal Sequencing Primer

*Taq*I
AGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTA 345

CGCGATCATGGCGACCACACCCCGTCCTGTGGATCAATTCCCTTTAGTATAAATTTCACTCTGAACCATC 414
<¬PRIMER 372

< NON - AMPLIFIED >
TTGGAAGGACCGGTAATTATTTCAAATCTCTTTTTCAATTGTATATGTGTTATGTTATGTAGTATACTC 483

TTTCTTCAACAATTAAATACTCTCGGTAGCCAAGTTGGTTTAAGGCGCAAGACTTTAATTTATCACTAC 552
PRIMER 373 ¬>

GGAATTC
*Eco*RI Cloning Site

Fig 4.6b. Sequence of the pYAC4 right arm showing the *Eco*RI cloning site as well as the location and sequence of the Inverse PCR primer sequences, the internal sequencing primer and the location of the restriction enzyme sites used for IPCR.

GGAATTCC **GTAATCTTGAGATCGGGCGTTC** GATCGCCCCGGGAGATTTTTTTGTTTTTTATGTCTTCC**A** **621**

***Eco*RI Cloning Site** **<¬ PRIMER 556**

TTCACTTCCCAGACTTGCAAGTTGAAATATTTCTTTCAAGGGAATTGATCCTCTACGCCGGACGCATCG **690**

..... // .....

NON- AMPLIFIED

CCTTTCAATTCAATTCATCATTTTTTTTTTATTCTTTTTTTTGATTTCGGTTTCTTTGAAATTTTTTTG **1794**

ATTCGGTAATCTCCGAACAGAA**GGAAGAACGAAGGAAGGAGCA**CAGACTTAGATTGGTATATATACG**CA** **1863**

PRIMER 374 ¬>

***Nde*I**

TATGTAGTGTTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGCA **1932**

***Taq*I**

GGAAACGAAGATAAAT**CATGTC**GAAAGCTACATATAAGGAACGTGCTGCTACTCAT**CCTAGTCCTGTT** **2001**

1989 Internal

*Rsa*I

GCTGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTC**GTAC**CA **2070**

Sequencing Primer¬>

***Eco*RV**

CCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGTTTACTAAAAACACATGTG**GATATC** **2139**

used to isolate the IPCR fragment. Similarly, although the right arm primers are separated by 1236 bp of YAC DNA, because of the inverse orientation, this is not amplified. An example of some of the STS markers produced in the course of this study is given in figure 4.7.

4.4.2.2 Sequencing of Left Arm Rescues

The five chromosome 6 specific YAC's so far identified were subject to left arm rescue with *Taq*I, *Rsa*I and *Eco*RV. Inverse PCR conditions established for left arm rescues were (i) denaturation at 94° for 60s; (ii) primer annealing at 52° for 60s; (iii) extension at 72° for 180s; amplified for 35 cycles. The 3 min extension time was used as the size of fragment to be amplified was unkown. Results of IPCR with primers 372 and 373 showed that for YAC Y9/ H2A a 400 bp *Rsa*I band and a 350 bp *Eco*RV band were produced. For YAC Y10/ H2A a 350 bp band was produced with *Rsa*I. YAC 12/ 63 showed an ~520 bp *Eco*RV band, an ~ 500 bp *Rsa*I band and an ~ 580 bp *Taq*I band. No left arm rescue was possible for any enzyme combination for YAC 11/ 63. For YAC Y5/ 87 an ~650 bp band was observed with *Taq*I. All of these products were cloned into the pCRII ™ vector for sequencing.

Initial sequencing was carried out with the M13 sequencing primers (NEB). An example of typical sequence data obtained for a left arm rescue is given in figures 4.8 and 4.11. Depending upon the orientation of the STS insert within the vector, either the forward or reverse primer would sequence through into the *Eco*RI cloning site. If the STS was constructed using either *Eco*RV or *Rsa*I, then the internal sequencing primer designed in this study was used and will be discussed in section 4.4. The full sequence of all primers designed for use in this first level screen is given in section 4.5.

4.4.2.3 Sequencing of Right Arm Rescues

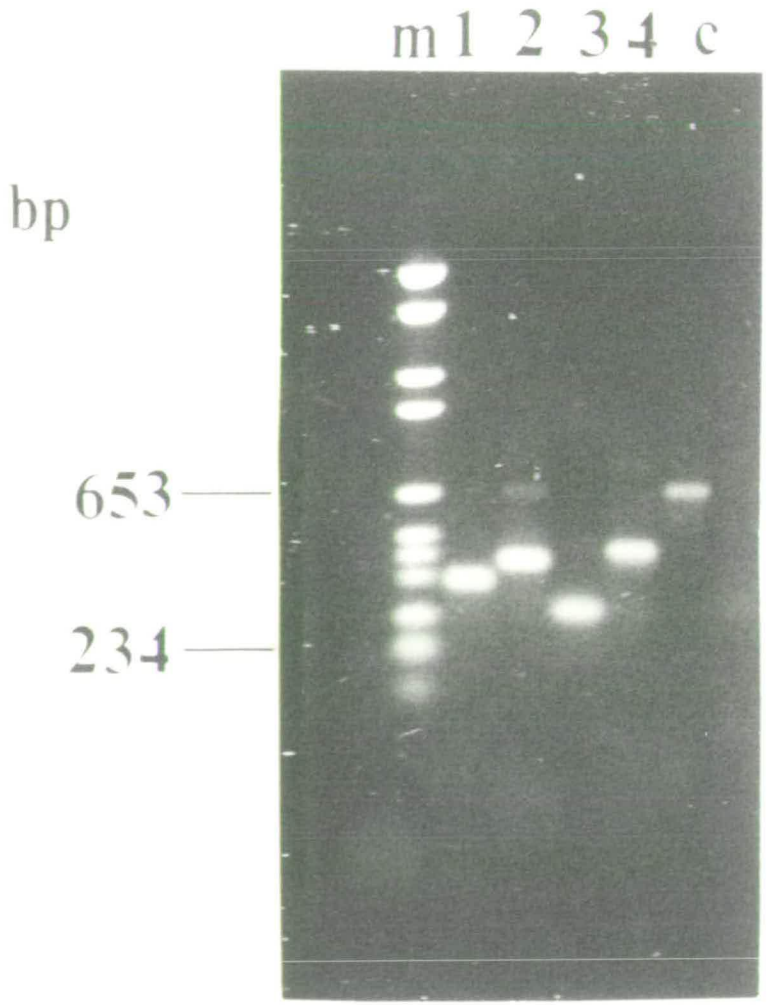
Inverse PCR reactions were also performed on all five chromosome 6 specific YAC's so far identified to isolate right arm end clone STS's. End clone rescues were performed with *Taq*I, *Eco*RV, *Rsa*I, *Nde*I and *Mae*II under standard amplification conditions of (i) denaturation at 94° for 60s; (ii) primer annealing at 52° for 120 s; (iii) extension at 72° for 180s; amplification performed for 35 cycles. A 400 bp *Rsa*I and a 450 bp *Mae*II fragment was identified for YAC Y9/ H2A. No rescue was possible for YAC Y10/ H2A. Two ~400 bp *Rsa*I and *Nde*I fragments were obtained for Y12/ 63 and an ~ 800 bp fragment was obtained with the *Nde*I digest of Y11/ 63. Two 450 bp *Rsa*I and *Nde*I rescues were performed on YAC Y5/ 87.

All suspected STS markers were cloned into the Invitrogen pCRII ™ vector for sequencing. Initial sequencing was carried out using standard M13 forward and reverse primers. An example of sequence information obtained for an *Nde*I rescue is given in figure 4.9. As in section 4.3, depending upon the orientation of the STS inserted within the pCRII™ vector depends upon whether the sequence reads through into either the *Eco*RI cloning site of the YAC or through into the arm itself and the

Figure 4.7. *An example of STS markers isolated in the course of this study.*

STS markers rescued form the ends of YAC's by Inverse PCR were cloned into the Invitrogen pCRII™ vector. 10µl aliquots of positive left arm IPCR signals were resolved on a 1.5% IBI agarose gel in 1X TAE for 12 hours @ 20V and photographed after EtBr staining under UV illumination.

Lane m: DNA size standards Boehringer DNA marker VI
Lane 1: Left arm *Eco*RV IPCR rescue of YAC Y9/ H2
Lane 2: Left arm *Rsa*I IPCR rescue of YAC Y9/ H2
Lane 3: Left arm *Rsa*I IPCR rescue of YAC Y10/ H2
Lane 4: Left arm *Rsa*I IPCR rescue of YAC Y12/ 63
Lane c: negative water control.


m 1 2 3 4 c
bp
653
234

enzyme cut site used to rescue the YAC. Full sequences of all primers used in a level one screen are given in section 4.5.

4.4.2.4 Sequencing using Internal Primers

As shown by the sequence of pYAC4 in figures 4.6a and 4.6b, the *Taq*I enzyme used to rescue the left arm is only 20 bp away from the end on primer 373. In addition, the *Nde*I, *Taq*I and *Mae*II enzymes used for right arm end clone rescue are 27, 117 and 137 bp away from primer 374. All of these enzymes yield sequence information for STS construction from a single 6 hour sequencing run, assuming an average of 250 - 300 base pairs of readable sequence. However, most of the STS markers rescued were with the enzymes *Rsa*I and *Eco*RV. These enzyme cut sites are very A+T rich (GATC and GATATC for *Rsa*I and *Eco*RV respectively) and would be expected to occur frequently in the *P falciparum* genome. However, a major disadvantage is the distance away from the IPCR primer sites in both arms: in the left arm *Eco*RV is 174 bp and *Rsa*I is 200 bp away from primer 372; in the right arm *Rsa*I is 230 bp and *Eco*RV is 300 bp away from primer 374. This means that up to 300 bp of YAC sequence has to be read before the *P. falciparum* insert is reached. To circumvent this, YAC specific left and right arm internal sequencing primers were designed, and their sequence is given in Chapter 2. The left arm primer extends from position 254 to 276 thus reducing the distance to the *Rsa*I site to 88 bp and *Eco*RV to 66 bp. The right arm primer extends from position 1989 to 2007, thus reducing the distance to the *Rsa*I site to 60 bp and *Eco*RV to 132bp. An example of sequence data obtained with M13 and internal primers is given in figures 4.10, 4.11 and 4.12

4.4.2.5 Design of STS Primers

The design of STS primers for use in a level one library screen are given in table 4.1. Due to the extremely high A+T of the *P. falciparum* genome, PCR primers were designed with as high a G+C content as possible. Primers of lower G+C content were compensated for by an increase in length. During data compilation it was noted that in sequencing 87/ 5L, the *Taq*I site used to rescue the YAC was undetectable, as there was an upstream 10 bp deletion of the YAC arm sequence CGCATCAGCT. Sequencing from the other side of the STS revealed a normal M13 - YAC - *Eco*RI - Insert sequencing pattern. However, it was also discovered that a second STS, marker 63/ 12L, had exactly the same deletion, and exactly the same insert sequence. Sequencing from Primer 373 of each *Taq*I rescued STS revealed identical inserts also. To test which sequence was correct, an *Rsa*I STS of 63/ 12L was sequenced. This showed that the primer 373 insert sequence was correct, but that the primer 372 sequence was radically different, and that the correct YAC arm sequence with *Rsa*I site was observed. Thus the sequence data for 63/ 12L was composited from *Taq*I and *Rsa*I rescues.

Fig 4.8a. Sequence of Left arm *Taq*I rescue of YAC Y6A/ 9L reading into the pCRII™vector with the M13 forward primer.

Fig 4.8b. Sequence of Left arm *Taq*I rescue of YAC Y6A/ 9L reading into the pCRII™ vector with the M13 reverse primer

Fig 4.8a.

M13 *Eco*RI YAC ARM SEQUENCE YAC *Eco*RI
GGATATCTGCA**GAATTC**GGCTT GCCAATTGGTTAAGGCGC AAGACTTTAATTTATCACTACG**GAATTC**
M13 SEQUENCE PRIMER 373 YAC ARM SEQUENCE

AAAAAAAAAAAAAAA**GAATAAGAGATGACAATAA**TATATATTATAGTGTGTTTATATTATAGTTCATAACATTTC
INSERT STS PRIMER 1

ATTATTTTACTTTTT

Fig 4.8b.

M13*Eco*RI YAC ARM SEQUENCE *Taq*I
CCGCCAGTGTGTGCTG**GAATTC**GGCTT GAATTGATCCACAGGACGGGGT GTGGTCGCCATGATCGCGTAGTCGA
M13 SEQUENCE PRIMER 372 YAC ARM SEQUENCE

GGAATGAAACAT**GGTTATCTATCTATATAAGG**AGATCATATATATATTCTACTTTTTAATTATATAAATTAATTTC
INSERT STS PRIMER 2

CAAAGAAACTATATTACTTAAAATCTTGAT

Fig 4.9a. Sequence of the right arm NdeI rescue of YAC Y11/ 63 reading into the pCRII™ vector with the M13 reverse primers

Fig 4.9b. sequence of the right arm *Nde*I rescue of YAC Y11/63 reading into the pCRII™ vector with the M13 forward primers.

Fig 4.9a

M13 *Eco*RI YAC *Eco*RI INSERT
GCCAGTGTGCT**GGAATTC**GGCTT GCCGATCTCAAGATTAC**GGAATTC** ATTAATAATATTTTTCGTAAATT
M13 SEQUENCE PRIMER 556 YAC SEQUENCE

CAGTTGCGAGATTCGATGAT TTTCATTTTATCATTATCATTAA
STS PRIMER 1

Fig 4.9b

M13 *Eco*RI
CACTATAGGGCGAATTGGGCCCTCTACATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCA**GAATTC**GGC
M13 SEQUENCE

*Nde*I
TT GGAAGAACGAAGGAAGGAGC ACAGACTTAGATTGGTATATATA **CGCATATG** AACAATGACATATATCATTAT
PRIMER 374 YAC ARM SEQUENCE INSERT

AATAATAATAGAAACGATGATATACTTTTAATTTT**GGACAGTCAATTGATACCAGC**ATTCTATTTATACATGATT
STS PRIMER 2

ATTCATATCAAAGTCTCTCGTATGATCTCGTGAATATTAATACAATATATG

Fig 4.10a. Sequence of Right Arm rescue of YAC 9R with *Rsa*I and sequencing using the right arm internal primers. The other STS primer would have been sequenced with conventional M13 primers, as the sequence would have shown M13: YAC: *Eco*RI cloning site: Insert as shown in fig 4.9

Fig 4.10b. Sequence of Left Arm rescue of YAC Y9/ H2 with *Rsa*I and sequencing using the left arm internal sequencing primers. The other STS primer would have been sequenced with conventional M13 primers, as the sequence would again have showed M13: YAC; *Eco*RI cloning site: Insert as shown in fig 4.8

Fig 4.10a

*Rsa*I

TTGGATGTTC**GTAC** TTTAATTTTTTTTTT **AAACATTCCTAATATTTGACAC** ATATATATAGATATGTATGTATA

YAC SEQUENCE — INSERT — STS PRIMER

Fig 4.10b

*Rsa*I

CCCGCAAGAGGGCCGGCA**GTAC** ATATTATTTGT **CGAGGTATATTTGTTGATATCC** TTATTTTTTTCGTTCATTATT

YAC ARM SEQUENCE — INSERT — STS PRIMER

TATAAATTATTTAACATAATAATTTATTTAAGAAAT

4.4.3. Isolation of Overlapping YAC's using STS's.

4.4.3.1. Location within the Library

As a first level check, all STS markers were amplified on controls to test the efficacy of the

YAC	ARM	STS	SEQUENCE	LEN.	% G+C	STS size
Y9/ H2	LEFT	H2/ 9L	CAAAAATTAAGGTTAGAGGGG CGAGGTATATTTGTTGATATCC	21 22	38 36	~100 bp
Y9/ H2	RIGHT	H2/ 9R	ATATGTTGAAAAATAAGTTTAGC AAACATTCCTAATATTTGACAC	23 22	22 27	~100 bp
Y5/ 87	RIGHT	87/ 5R	CACAACCAATAGGTGTATG GATCATAATAATAACACTGAAGAAG	19 25	40 28	~250 bp
Y5/ 87	LEFT	87/ 5L	CACCGATTTGATTATTTGGATC CAACAATGAGTTTATATTAAAGG	22 23	36 26	N/ A
Y11/ 63	RIGHT	63/11R	CAGTTGCGAGATTCGATGAT GGACAGTCAATTGATACCAGC	20 21	45 47	~700 bp
Y12/ 63	LEFT	63/12L	CAACAAATGAGTTTATATTAAAGG CAAAGAAAATAAAGATCCTTCG	24 22	25 32	~250 bp
Y12/ 63	RIGHT	63/12R	AACATCTCAAAATATTACAACAC TTATGAGTATTGGCTTGAA	23 19	26 32	84 bp

Table 4.1. *Sequence and Designation of YAC's and the STS markers used in a Level One screen of the HB3 YAC Library. LEN corresponds to length of primer and is given in base pairs. The orientation of each primer sequence is given as identified by the M13 forward and reverse primer*

primers and PCR conditions. As shown in figure 4.13a, primers 63/ 11R and 87/ 5R were screened on HB3, 3D7, AB1380, H20 and Parent YAC controls. Three positive results and two negative are the criterion for a successful STS marker. 63/ 11R amplified an ~700 bp signal under PCR conditions 94° 60s; 52° 60s; 72° 60s for 35 cycles. 87/ 5R amplified an ~250 bp signal under 94° 60s; 55° 60s; 72° 60s for 35 amplification cycles. H2/ 9R amplified an ~100 bp signal under 94° 60s; 50° 60s; 72° 60s for 35 amplification cycles; H2/ 9L amplified an ~ 100 bp signal under 94° 60s; 52° 60s; 72° 60s. No amplification signal was seen with either 63/ 12L or 63/ 12R.

Library screens were performed essentially as described in section 3. A MasterPool screen was omitted at this stage in favour of a direct plate pool screen. In addition to the controls included as shown in figure 4.13b., as a secondary control each STS should amplify the original pool from which it was identified. Pool screening with 87/ 5R amplified positive signals in pools 2, 3, 5, 6 and 16. Trace signals were seen in pools 9, 10, 11, 12 and 13. Water and yeast controls were clear, so the results could be explained by gel loading error or actual PCR signals. STS H2/ 9R amplified signals in pools 2, 5, 9 and 14. H2/ 9L amplified pools 3, 6, 9, and 13. STS 63/ 11R failed to amplify the pools despite numerous successful control amplifications under various PCR conditions. The successful amplification of different

Figure 4.11. Sequencing Gel for YAC rescues.

Both autoradiographs were scanned into an Apple Macintosh using an Epson GT 8500 Scanner and Adobe Photoshop imaging software. Image manipulation was performed using Persuasion 2.1 software. Scanned images were printed on an Apple Laserwriter Pro.

Figure 4.11 A.

YAC Y9/ H2 left arm digested with *EcoR*I and sequenced with the M13 forward primers (NEB). The primer reveals sequence leading into the *EcoR*I cloning site of the YAC. *P. falciparum* insert sequence is visible as sequence divergence beyond this point.

Figure 4.11B.

YAC Y9/ H2 left arm digested with *Rsa*I and sequenced with M13 reverse primers (NEB). The primer reveals YAC sequence and the location of the internal sequencing primer used in this study is indicated.

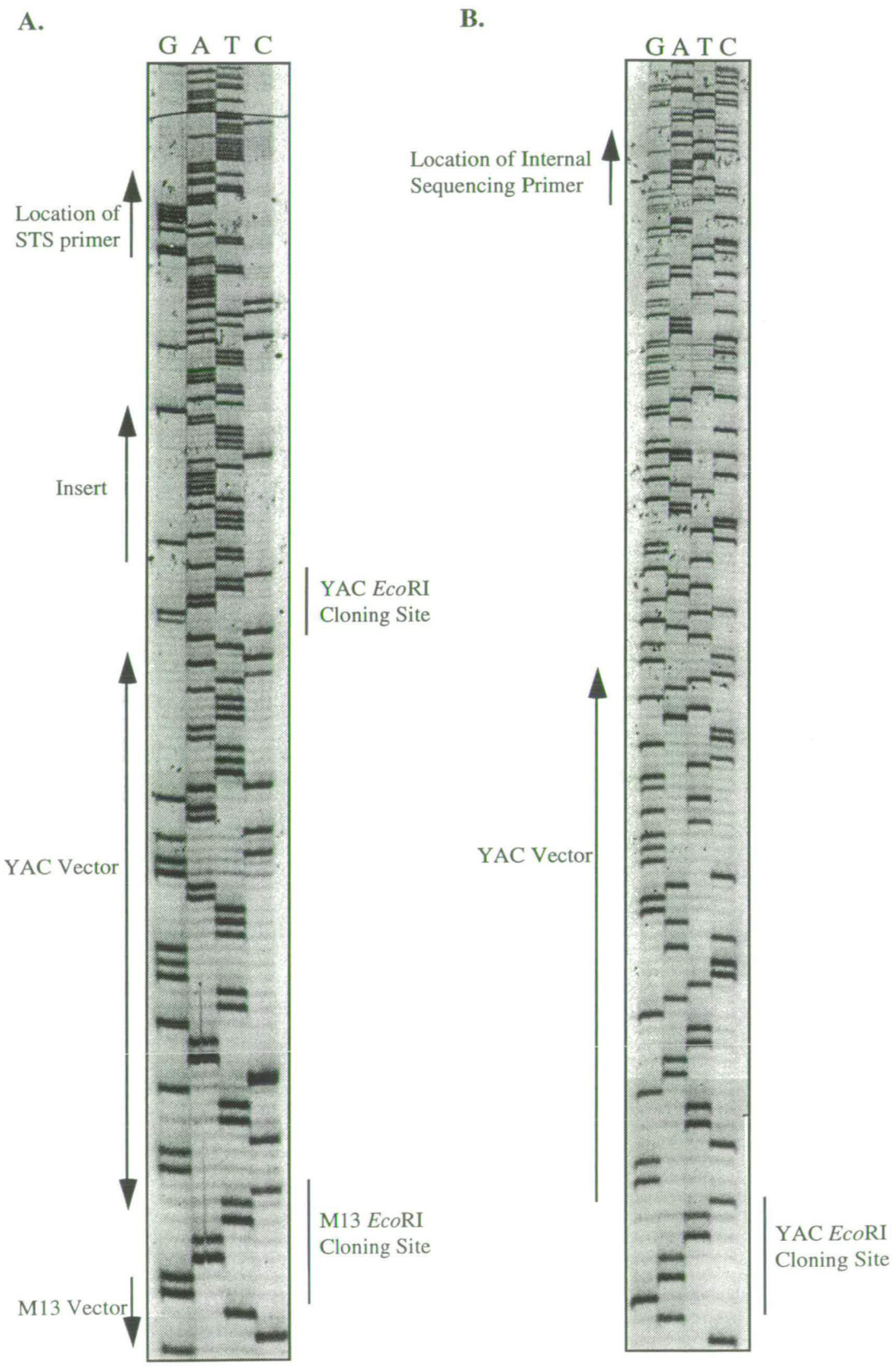
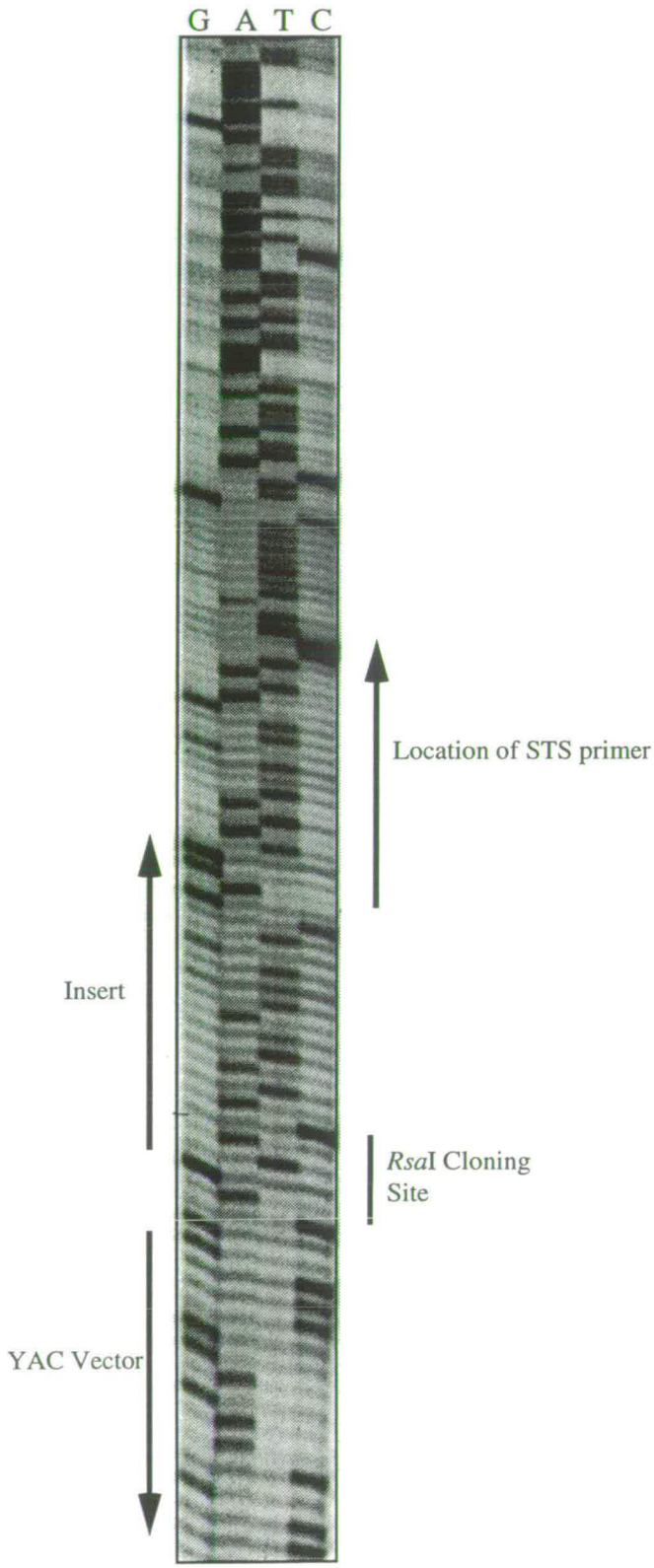

A.
G A T C
Location of
STS primer
Insert
YAC *Eco*RI
Cloning Site
YAC Vector
M13 *Eco*RI
Cloning Site
M13 Vector
B.
G A T C
Location of Internal
Sequencing Primer
YAC Vector
YAC *Eco*RI
Cloning Site

Figure 4.12. Sequencing Gel for Internal Sequencing primers

Scanning of the gel and digital manipulation of the image was performed as described for figure 4.12. The figure shows YAC Y9/ H2 digested with *Rsa*I and sequenced with the Left Arm internal sequencing primers. The *Rsa*I cut site and malarial insert sequence are indicated.


G A T C
Location of STS primer
Insert
RsaI Cloning Site
YAC Vector

pools per STS primer indicates that the STS sequences are apparently unique, as highly repetitive DNA would have been expected to show amplification in all 16 pools. In addition, the number of YAC pools identified per STS is in accordance with the frequency expected for a 4/ 5 times genome coverage.

4.4.3.2. Identification and Purification of Level 1 YAC's

For STS 87/ 5R, plates 2 and 6 were cultured and row and column DNA templates were prepared as described. PCR conditions under 40 cycles of amplification revealed positive signals in rows C and H, columns 4 and 12 of plate 6; rows A and B, columns 1, 3 and 7 of plate 2. This indicates four possible locations in plate 6 (C4, C12, H4 and H12), and six possible locations in plate 2 (A1, A3, A7, B1, B3 and B7). Culturing of individual clones for PCR templates revealed positive signals in 6- 4/H and 2- A1. Both of these were cultured to single colonies and three clones randomly picked. Each clone was positive, so colonies 6.1 and 2.1A were taken and cultured for YAC stocks. YAC 6.1 is designated Y6/ 5R and YAC 2.1A is designated Y2/ 5R. Pulsed field gel analysis indicate the YAC sizes to be approximately 60 kb for Y6/ 5R and 100 kb for Y2/ 5R

For STS H2/ 9R, plate 2 was cultured and row and column DNA preps were made. PCR under 40 cycles of amplification revealed a clear location in column 10, row B. The clone was cultured to single colonies, and PCR amplification of templates from three random colonies were all positive under 30 cycles of amplification. The YAC identified was designated Y2/ 9R and was shown to be approximately 60 kb.

For STS H2/ 9L, plates 6 and 13 were cultured and row and column templates prepared. PCR carried out with 40 amplification cycles revealed signals in plate 6, row A columns 1, 2, 4, 5 and 7 corresponding to five possible locations. Signals were also observed in plate 13, row F columns 1, 6, 7 and 8. PCR performed on individual templates confirmed signals in 6- B/ 2 and 13 - F/ 8. Both colonies were cultured to obtain single colonies and all the plate 6 clones were positive. The plate 13 clones were only amplified on addition of undiluted template.YAC 6- B/ 2 was designated Y6/ 9L and YAC 13 -F/ 8 was designated YAC Y13/ 9L Pulsed field gel analysis indicates an insert size of approximately 40 kb for YAC Y13/ 9L and 70 kb for YAC Y6/ 9L.

4.4.3.3 Blotting Studies on YAC's

To confirm that the STS markers identified mapped to chromosome 6, a Pulsed field gel of intact HB3 chromosomes resolving chromosome 6 was blotted and probed with all available markers. The results show that H2A, 87/ 5R, 11.63, H2/ 9L and 63/ 11R all map to the same region of the genome (data not shown). This indicates that the STS markers rescued do not come from inter-chromosomal chimeric YAC's. Presumably because of the low STS marker size, it was not possible to purify H2/ 9R and pSC11.87 for labelling.

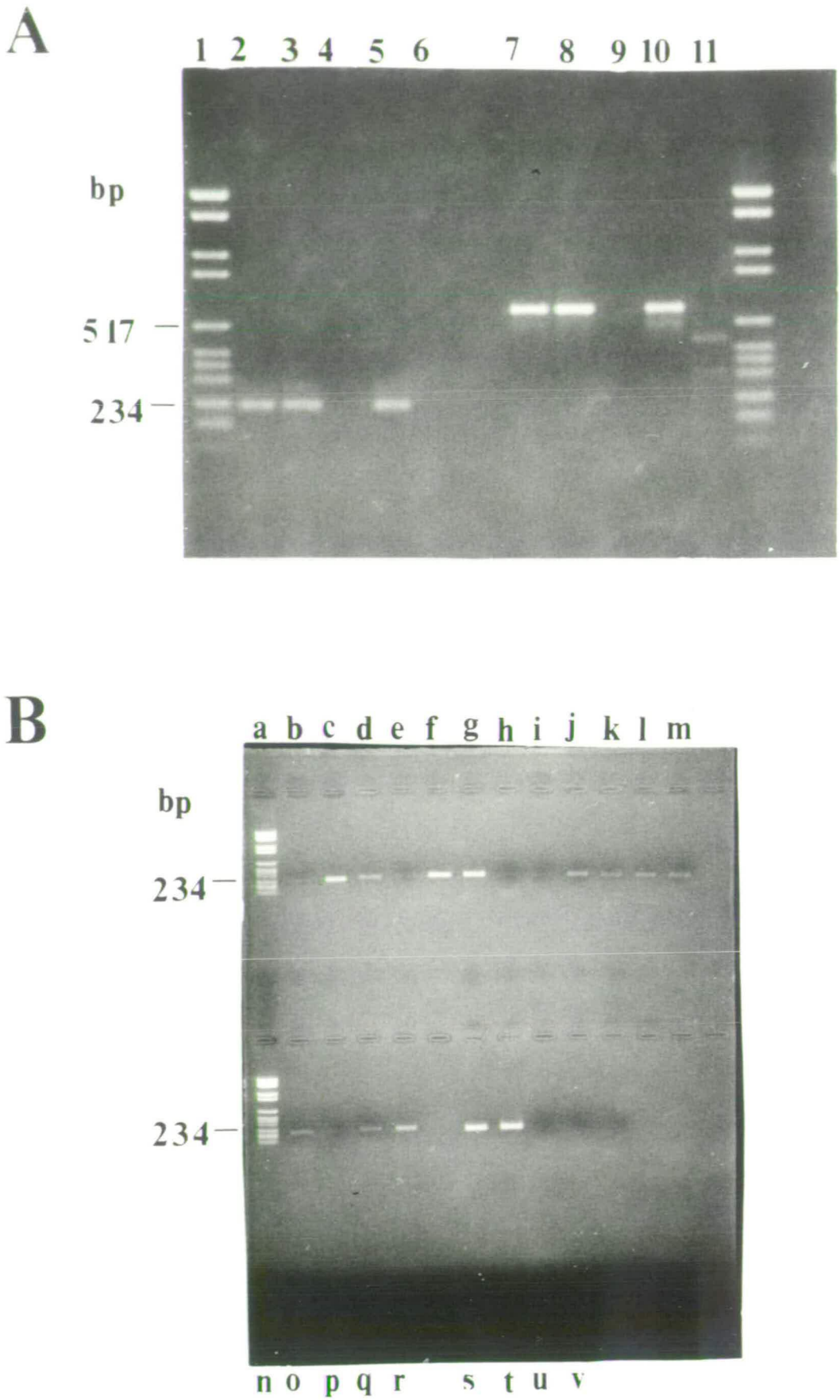
Figure 4.13 Analysis of STS markers constructed on this study.

A. STS primers 63/ 11R and 87/ 5R, constructed in this study were used to screen all positive and negative controls available to optimise amplification conditions. 10µl of PCR product were resolved on a 2% IBI agarose gel in 1X TAE over 3 hours @ 50V, stained with EtBr and photographed under UV illumination.

Lane 1: DNA size standard - Boehringer DNA Marker VI
Lane 2: 87/ 5R amplified on total 3D7 DNA (positive control)
Lane 3: 87/ 5R amplified on total HB3 DNA (positive control)
Lane 4: 87/ 5R amplified with a water template (negative control)
Lane 5 : 87/ 5R amplified on a YAC Y5/ 87 template (positive control)
Lane 6: 87/ 5R amplified on an AB1380 template (negative control)
Lane 7: 63/ 11R amplified on total 3D7 DNA (positive control)
Lane 8: 63/ 11R amplified on total HB3 DNA (positive control)
Lane 9: 63/ 11R amplified with a water template (negative control)
Lane 10:63/ 11R amplified with a YAC Y11/ 63 template (positive control)
Lane 11: 63/ 11R amplified with an AB1380 template (negative control)

B. Amplification with the HB3 YAC library MasterPools with the 87/ 5R primers. Amplification occurred under 55°C annealing with 35 cycles. 10µl of PCR product was resolved on a 1% IBI agarose gel in 1X TAE and photographed under UV illumination.

Lane a: DNA size standard- Boehringer DNA marker VI
Lanes b - r : HB3 YAC library plate pools 1 - 16
Lane s: total 3D7 DNA (positive control)
Lane t: total HB3 DNA (positive control)
Lane u: total AB1380 DNA (negative control)
Lane v: water template (negative control).


A
1 2 3 4 5 6 7 8 9 10 11
bp
517
234
B
a b c d e f g h i j k l m
bp
234
234
n o p q r s t u v

4.4.4. Level Two Sequence Tagged Site Marker Construction.

The YAC's identified in the Level One screen were subjected to standard inverse PCR to rescue Level Two STS markers from the first generation YAC's. Contig extension can therefore take place by identifying new YAC's within the library with these primers. To ensure that the map is extending outwards, the inclusion of the original parent YAC as a negative control confirms contig extension.

4.4.4.1. Inverse PCR Results

Digestion of YAC's Y2/ 9R, Y2/ 5R and Y6/ 5R with *Rsa*I followed by amplification with left arm primers revealed a 650 bp band for Y2/ 9R and an ~300 bp band for Y2/ 5R. No amplification was possible for Y6/ 5R. Digestion of YAC Y6/ 9L revealed a clean 650 bp band with *Taq*I. No left arm rescue was possible for YAC Y13/ 9L. Digestion of the Level one YAC's with all available IPCR enzymes followed by Right Arm amplification rescue revealed a 700 bp *Eco*RV band for Y6/ 5R. An 800 bp *Eco*RV band was observed for YAC Y13/ 9L, no amplifications were seen with the remainder of the level One YAC's.

All PCR products were cloned into the pCRII™ vector for sequencing with standard M13 primers (NEB) as well as the internal sequencing primers (See Section 4.4).

4.4.4.2. Design of STS Primers.

The criterion for the Level One STS primer design was applied here also. From the sequence data obtained, primers with as high a G+C content as possible were selected. Sequencing of the 300 bp *Rsa*I band from Y2/ 5R revealed an insert too small and with too high an A + T content for useful STS primer design. Sequencing reactions with M13 forward, reverse and YAC internal was performed per marker and the results are presented in table 4.2. It was discovered during sequencing that DNA prepared from the InvF' clones containing the Y6/ 5R *Eco*RV band did not posess any insert. All attempts to re- PCR the STS from the original IPCR ligation mix failed and these experiments were not pursued.

YAC	ARM	STS	SEQUENCE	LEN	% G + C	STS Size
Y6/ 9L	LEFT	9L/ 6L	AAGAATAAGAGATGACAATAA GGTTATCTATCTATATAAGG	21 20	24 30	N/ A
Y6/ 5R	RIGHT	5R/ 6R	GGTGCTATATTTGGATGATATC AAGGATTGGGATTTGATTCC	22 20	36 40	~230 bp
Y2/ 9R	LEFT	9R/ 9L	CTTGTTACAAATAGCGAAGG AAATAACTATGAGCTTTTGC	20 20	40 30	~ 150 bp

Table 4.2. Sequence and designation of the Level two STS markers constructed in this study.

4.4.5. Identification of Overlapping YAC's using Level Two STS's .

STS markers constructed in the level one screen were used in a level two screen of the HB3 YAC library. The plate locations of level two YAC's were identified, but no clone identification or purification studies have been performed as yet.

4.4.5.1 Location within the Library

STS marker 9R/ 9L, identified from the right arm of a YAC containing STS H2/ 9R was used to screen all plate pools. Amplification conditions of 94° 60s; 50° 60s; 72° 60s for 35 cycles showed an ~150 bp signal in pools 2, 5, 6, 9 and 10. The signal from pool 2 is expected as the YAC was isolated from there. PCR on 3D7, HB3 and Y2/ 9R templates were positive. This indicates a minimum of four copies of the STS within the library, which is in good agreement with the frequency predicted. Interestingly, PCR tests on YAC's Y9/ H2 and Y10/ H2 were negative, indicating that the orientation of the insert within Y2/ 9R results in the contig extending outwards. If a left arm rescue could be performed, this would increase marker density within the Y9/ H2 - Y10/ H2 - Y2/ 9R contig.

STS marker 5R/ 6R was used to screen all plate pools. Amplification conditions of 94° 60s; 50° 60s; 72° 60s for 35 cycles showed an ~230 bp band in plates 4, 6, 9 and 14. 3D7, HB3 and Y6/ 5R were all positive. The signal in pool 6 is expected as this YAC was identified from here. This indicates a minimum of four copies of the STS within the library, and is in good agreement with the predicted copy number frequency. Again, the parent YAC of the contig, Y5/ 87, was included as a control and the negative result indicates that the insert of Y6/ 5R is not overlapping back onto itself. Thus contig extension will proceed outwards from this contig.

In screening the library with the 9L/ 6L primers, no signal was seen in either pools or controls for various combinations of annealing temperature, amplification cycle number and Mg^{2+} concentration. PCR conditions remain to be optimised for this set of primers.

4.5. Discussion

To assemble a physical map of any chromosome as an ordered series of overlapping artificial chromosomes, chromosome specific YAC's must first be identified from within the library. This was performed for chromosome 6 using a PCR- based approach. Once YAC's had been identified, markers were constructed from the ends of each YAC by Inverse PCR and used to screen for overlapping YAC's.

Three markers were known to exist for chromosome 6 and were used to identify preliminary chromosome 6 specific YAC's. Up to two YAC's were identified per marker, a figure considerably lower than the 4 copy frequency predicted from library size estimations. Such sequences may be underrepresented in genomic libraries or alternatively the estimate of the library size may be somewhat overestimated. However, STS markers isolated by end clone rescue were present in much higher frequencies than the 4/ 5 copy number per marker indicated. End insert rescues were performed and

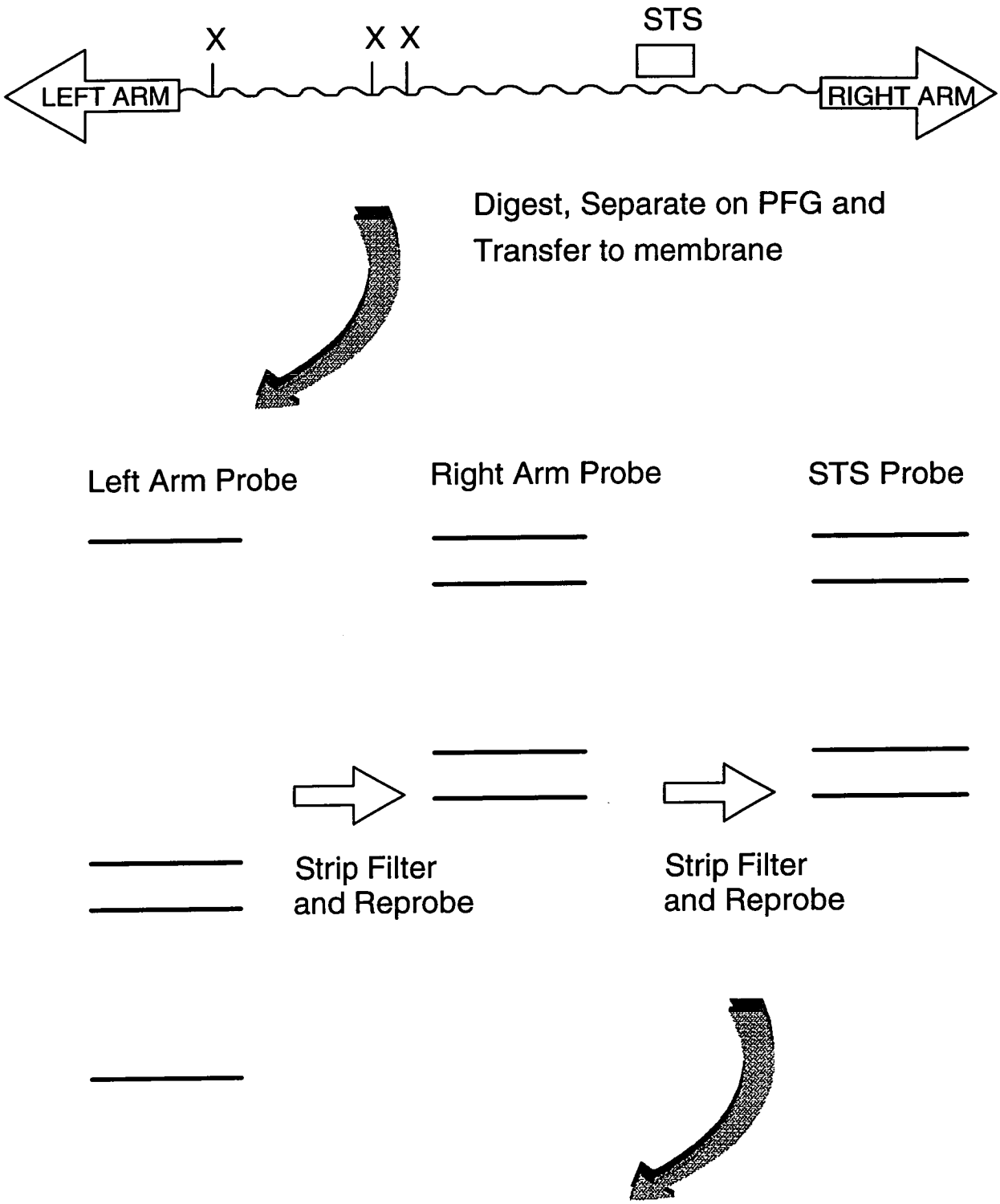
used to identify overlapping YAC's. Three contigs were initiated: a contig seeded by the H2A marker, yielding two H2A YAC's, a YAC overlapping the right arm (Y2/ 9R) and two YAC's overlapping the left arm (Y6/ 9L and Y13/ 9L). The second contig consists of a seed pSC11.87 YAC and two YAC's overlapping the right arm (Y2/ 5R and Y6/ 5R). The final contig is two YAC's identified by the marker pSC11.63. This corresponds to a total of 750 kb, approximately half of the chromosome. However, it remains to be determined exactly how much chromosomal coverage the contigs provide.

End insert rescue was performed using inverse PCR. During the course of the project, 10 YAC's were identified corresponding to 20 end inserts to be rescued. Useful IPCR sequence data was obtained from 10 YAC ends by this method alone, corresponding to ten new chromosome 6 specific markers constructed for these small contigs alone. This success rate is in agreement with previously published studies which have had to use a combination of techniques to rescue both ends of each YAC isolated (Marchuk *et al.*, 1992; Bates *et al.*, 1992; Albertsen *et al.*, 1994). In addition, the STS marker sequences described are in good agreement with regards to length, G+C content and STS size for the other published *P. falciparum* STS markers (Lanzer *et al.*, 1993; Rubio *et al.*, 1995). Both describe STS markers ranging in size from 59 bp to 1200 bp. The sequence data presented here, in addition to the conditions for amplification, means that any future research project involving chromosome 6 has ready access to a number of chromosome specific markers.

Future work would be to refine the IPCR protocol by increasing the range of enzymes available for rescue or to use a second PCR based method of end clone rescue (such as vectorette PCR) in parallel to ensure the maximum possible chance of obtaining end clone information Such techniques could be used to attempt end clone rescue for the YAC ends not identified. In addition, future work on contig construction would involve isolating the level 2 YAC's identified, rescuing ends as described and going on to perform level three STS construction. Finally, due to the absence of STS markers for a level 1 search, an investigation into methods of constructing multiple chromosome 6 specific markers, by either subcloning YAC inserts or subcloning total chromosome 6 DNA, would orient the mapping strategy towards an STS- content based approach and should greatly streamline information assembly.

CHAPTER 5.

CONSTRUCTION OF A PHYSICAL MAP OF *P. falciparum* CHROMOSOME 6.


STS
X
X X
LEFT ARM
RIGHT ARM
Digest, Separate on PFG and
Transfer to membrane
Left Arm Probe
Right Arm Probe
STS Probe
Strip Filter
and Reprobe
Strip Filter
and Reprobe
Data Assembly and Map Construction

5.1.Abstract

All YAC's isolated in the course of this study were restriction enzymes mapped by partial digestion with seven enzymes: *Nhe*I, *Pst*I, *Pvu*II, *Xho*I and *Sac*I which have a hexanucleotide ($^{G}/_{C}$)($^{G}/_{C}$)($^{A}/_{T}$) target sequence, and *Nar*I and *Eag*I, which have a hexanucleotide ($^{G}/_{C}$)($^{G}/_{C}$)($^{G}/_{C}$) target sequence. Each partial digest was resolved on an overnight pulsed field gel and transferred to a membrane. Each filter was successively hybridised with probes specific to the left arm, right arm and internal STS marker. Physical maps of all cloned YAC's are presented. Left and right map information proved consistent for the assembly of final composite maps, however problems were encountered in assembling putative overlapping YAC's into contigs. One of the YAC pairs identified, Y9/H2 and Y10/H2, are possibly chimeric, based upon the presence of a heterologous *Xho*I site. Implications of chimerism within the Genome Project are discussed.

5.2. Introduction.

Complete physical maps or 'contig' maps consisting of overlapping recombinant DNA clones spanning an entire genome are a primary guide for investigating the organisation of the genetic material of an organism and the information encoded. Such physical maps facilitate correlation to genetic linkage maps, offer immediate access to clones from any region and assist in genomic sequencing and expression studies of the insert DNA. As discussed, extensive restriction mapping of YAC inserts by Southern analysis with rare- cutting enzymes and PFG electrophoresis with multiple, locus- specific probes can be used to construct detailed restriction maps by relating the position of the gene landmarks to the locations of restriction sites within the insert. Such procedures also check correlation of the YAC information with the genomic copy and allows estimation of YAC clone integrity. In addition, physical maps can be obtained by assessing the STS content of each YAC, where YAC's containing common markers are found to overlap and thus contig assembly occurs. As overlapping YAC's identified by end-rescue STS screening automatically contain markers that can be positioned on a restriction map of the insert, physical maps constructed this way can correlate an STS based physical map strategy with a restriction site- based approach. Such approaches have been used to create long range and short range physical maps in contigs spanning the human Dystrophin (Coffey *et al.*, 1992), Duchenne Muscular Dystrophy (Monaco *et al.*, 1992) and Multiple Endocrine Neoplasia (Brooks- Wilson *et al.*, 1993) loci, as well as in a YAC contig spanning *P. falciparum* chromosome 4 (Rubio *et al.*, 1995)

One of the advantages of the YAC is the ability to construct physical maps in the absence of any STS information. Such a strategy is based on the use of probes specific to the left and right arms of the vector, with the result that the restriction fragments so produced are based upon proximity to the vector arms. A combination of both patterns produced must be used to determine a composite restriction map of the insert (illustrated in figure 5.1). Any STS markers subsequently identified can be positioned

on the map. Such a strategy has been used in the construction of maps spanning the human Acidic Fibroblast growth factor 1 (Chiu *et al.*, 1994), Huntingdon's disease (Bates *et al.*, 1992) and Type 1 Neurofibromatosis (Marchuk *et al.*, 1992) genes. Due to the relatively few markers mapped to *P. falciparum* chromosome 6 (Triglia *et al.*, 1992), this strategy is used for maps constructed in the course of this study.

5.3. The Mapping Strategy.

In order to construct physical maps of YAC insets, there are three parameters which have to be optimised in order to yield maximal information. Such conditions will depend upon the size of the insert and on the genome composition of the organism under study. These are: (i) choice of enzymes used for map construction; (ii) partial digestion and size- fractionation of DNA and (iii) serial filter hybridisation with vector specific probes. Due to the technical limitations of obtaining optimal partial digests, effective resolution of all bands on gels and complete transfer of DNA fragments onto the membrane, all restriction sites may not be detected. Comparison of the restriction sites in overlapping YAC's, mapping of saturated markers to the genome and employing different PFG resolution conditions per YAC will ensure completion of the map.

5.3.1 Choice of Enzymes

As with any map strategy, the choice of enzymes will affect the resolution of the resulting map. As YAC cloning involves breaking down the chromosome into an ordered series of clones, enzymes such as hexa- and octanucleotide $^{G}/_{C}$ rich sequence specific enzymes which have been used in long range map construction of higher eukaryotic genomes become uninformative at this level (Fulton *et al.*, 1988; Albertsen *et al.*, 1994). For example *Sma*I, which has a recognition sequence of CCC$^{\downarrow}$GGG, would be expected to cut infrequently in the A+T rich genome of *P. falciparum*. *Sma*I has been shown to cut chromosome 4 once (Watanabe *et al.*, 1994), chromosome 11 three times (Walker- Jonah *et al.*, 1992), chromosome 5 three times (Foote *et al.*, 1989) and twice in chromosome 2 (Corcoran *et al.*, 1988). With reference to chromosome 5, size polymorphism of 1.6 - 2.4 Mb is observed between isolates, however the 5` *Sma*I sites demonstrate conserved locations at 375 kb and 700 kb from the 5` telomere (Foote *et al.*, 1989). Assuming a chromosome size of 2.0 Mb, it would take a minimum ordered set of 20 YAC clones with inserts of 100 Kilobases to cover the chromosome, and thus *Sma*I would be expected to cut only three of these, the fourth and seventh YAC from the 5` telomere, assuming perfect overlaps. YAC map information available from such enzymes would thus be limited. In order to gain maximum site information per insert, a range of hexanucleotide restriction enzymes with 4 $^{G}/_{C}$ sites and 2 $^{A}/_{T}$ sites were chosen. By increasing the number of A+T recognition sites, the restriction frequency within a defined area should increase. This has been shown for chromosome 4, as *Pvu*II cuts 18 times and *Nhe*I cuts 19

times (Watanabe *et al.*, 1994). The final selection of enzymes were: *Pst*I (CTGCA$^{\downarrow}$ G), *Nhe*I (G$^{\downarrow}$ CTAGC), *Sac*I (GAGCT$^{\downarrow}$ C), *Xho*I (C$^{\downarrow}$ TCGAG), *Pvu*II (CAG$^{\downarrow}$ CTG), *Eag*I (C$^{\downarrow}$ GGCCG) and *Nar*I (GG$^{\downarrow}$ CGCC). *Nar*I and *Eag*I are included as 'landmarking' enzymes - enzymes cutting infrequently which can link short range maps to any long range maps compiled.

5.3.2 Partial Digestion of YAC DNA

The production of a optimal partial digest is critical to the construction of a map. Partial digests have the advantage of yielding a ladder of fragments, corresponding to all DNA band in the interval between completely digested and undigested. Probes on adjacent restriction fragments can be linked by common hybridisation to one of the partial fragments. In order to yield such a ladder, the DNA must be cleaved efficiently and the enzyme levels titred to optimise digest conditions. As discussed in chapter 2, preincubation of agarose blocks for 18 hours on ice at 4°C with enzyme and buffer ensures that each block is thoroughly saturated with inactive enzyme and that most restriction sequences will have an enzyme bound. When the blocks are incubated at a permissive temperature, digestion at all sites begins simultaneously. This technique proved more efficient than digesting DNA with high levels (e.g. 40U) of enzyme or by varying the digestion time with fixed units of enzyme. Optimal partial digest titres were observed with 15U, 3U, 0.3U and 0.03U of enzyme. In general, 15U corresponded to a complete digest, 0.03U corresponded to an undigested control with 3U and 0.3U yielding partial digests.

Once partial digest conditions had been established, the fragments must be resolved on a pulsed field gel. Switch time conditions were selected to ensure minimal resolution of the yeast chromosomes and maximal resolution of the fragments produced by digestion. For YAC's of 80 - 150 kb, run conditions resulting in yeast chromosomes resolving as a compression zone, with the remainder of the gel area left for the resolution and mapping of YAC fragments were a 2 - 12 second switch time ramped over 24 hours, as shown in figure 5.2a. For smaller YAC's pulse times were dropped.

5.3.3 Mapping Probes.

After resolution, partial digest fragments are transferred to nitrocellulose for serial hybridisation. Filter hybridisation and probe stripping are carried out essentially as described in materials and methods. Probes used are a 330 bp left arm probe and a 365 bp right arm probe, amplified from the pYAC- RC plasmid using primers described in chapter 2. The efficiency of filter stripping must be confirmed prior to subsequent rehybridisation by exposing the filter to X- ray film for 48 hours. If trace signals can be seen, the filter must be restripped and reexposed, as spurious signals affect map assembly data. An example of signals produced by left and right arms are shown in figure 5.2b. In order to calculate the exact size of the fragments produced, a standard log curve of DNA mobility (in cm) is calculated, based on separation of molecular weight markers of known size. The mobility of restriction fragments can be measured, and the fragment size in kilobases can be determined. Maps are assembled

Figure 5.2. Physical mapping of *P. falciparum* YAC's

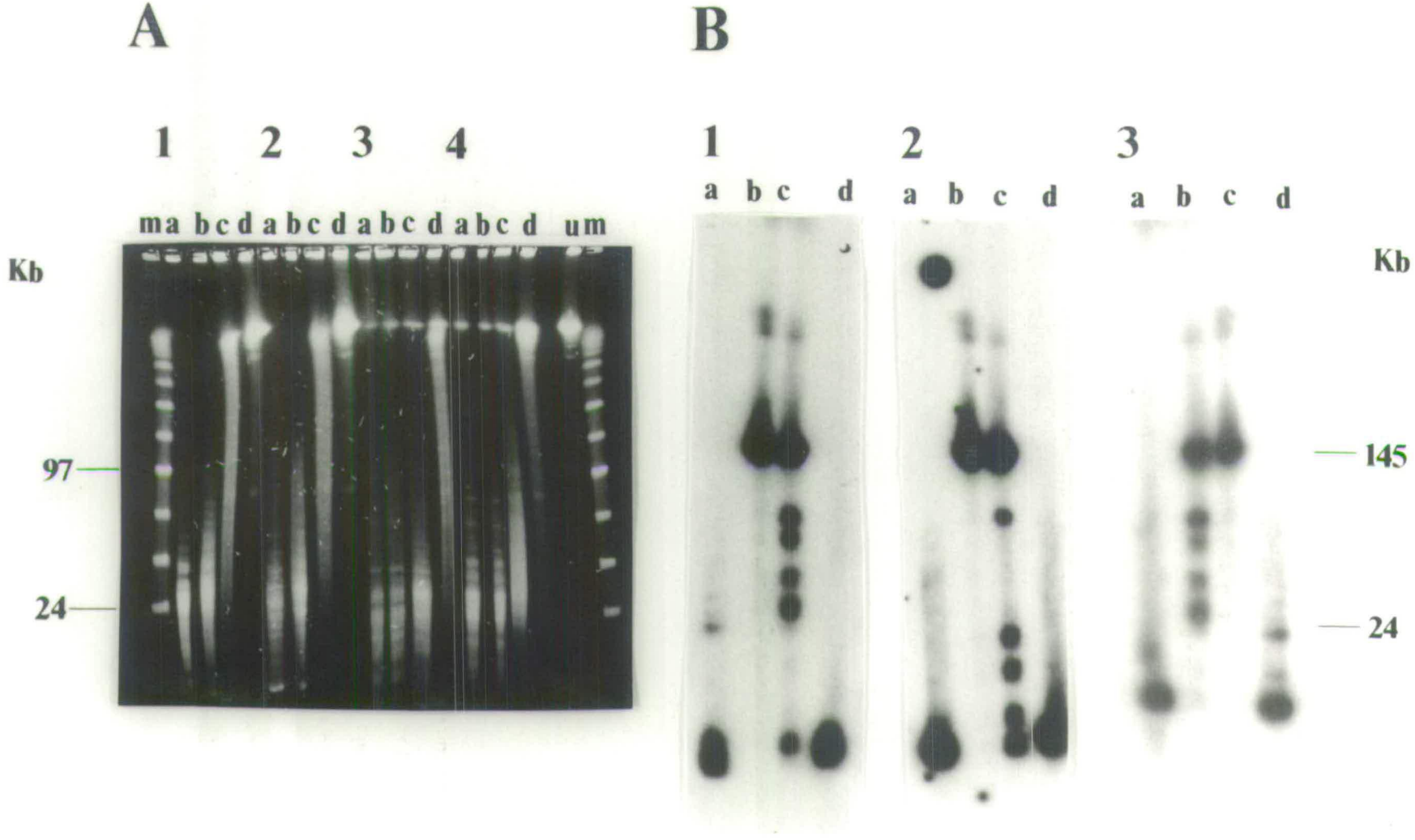
Physical mapping in Artificial Chromosomes is based on the partial digest of YAC's with infrequently cutting enzymes and the subsequent transfer to membranes and serial hybridisation with left arm, right arm and internal gene sequence probes.

A. YAC Y5/ 87 was partially digested essentially as described in Chapter 2. Restriction fragments were resolved on a 0.9% IBI agarose gel in 0.5X TBE for 24 hours @ 175V with a 2 - 12 sec ramp. Gel lanes designated **a** correspond to digests with 15U of enzyme; **b** correspond to 3U enzyme; **c** correspond to 0.3U enzyme and **d** to 0.03U enzyme.

Lane m:	DNA size standards NEB Mid range PFG marker II
Lane Set **1**:	YAC Y5/ 87 partially digested with NheI
Lane Set **2**:	YAC Y5/ 87 partially digested with SacI
Lane Set **3**:	YAC Y5/ 87 partially digested with PstI
Lane Set **4**:	YAC Y5/ 87 partially digested with XhoI
Lane u:	Undigested YAC Y5/ 87 control

B. YAC Y10/ H2 was digested with NheI and resolved on a 0.9% IBI agarose gel in 0.5X TBE for 24 hours @ 175V with a 2 - 12 sec. ramped pulse time. Fragments were transferred onto Hybond N^{+TM} and successively probed with ^{32}P- labelled YAC arm specific probes amplified from pYAC- RC template with primers described in Chapter 2. Between each hybridisation, filters were stripped as described and exposed to X- ray film for 24 - 72 hours to confirm the efficiency of probe removal. Lanes marked **a** correspond to digest with 15U enzyme; **b** to digestion with 0.3U enzyme; **c** to digestion with 0.03U enzyme; **d** to digestion with 3U of enzyme.

Lane Set 1:	YAC Y10/ H2 probed with Left Arm specific Probe
Lane Set 2:	YAC Y10/ H2 probed with Right arm specific Probe
Lane Set 3:	YAC Y10/ H2 probed with internal STS marker H2A


A
1
2
3
4
ma bcd abcd abcd abcd um
Kb
97
24
B
1
2
3
a b c d a b c d a b c d
Kb
145
24

by merging data for left and right arm probes. Although such data should correlate for sites located internal to the fragment, the information is particularly useful for mapping sites proximal to the insert termini. Enzyme sites located close to the 5` end may not be visible as small right arm probe fragments, as such fragments may have run off the gel, but will be visible as very large left arm probe fragments. YAC physical mapping is therefore a composite of both strategies.

5.4. Results.

5.4.1 General Mapping Considerations

In constructing the maps, several general points should be mentioned. Firstly, maps presented are of sites which correlate for both sets of hybridisation filters. During data assembly, it was observed that sites appeared in one set of filters which could not be correlated with the other. Such data could have arisen due to poor labelling of one set of probes, by removal of probes from filters due to stringent washing procedures by loss of fragments from the gel due to PFG resolution conditions. Second, log curves were constructed by calculating the reduction factor of gel photographs by a comparison of 1cm on a gel photograph with 1cm on a ruler. The ratio obtained was used to convert distance migrated by DNA size standards on the photograph to the actual distance migrated on the original gel. Consequently measurement error means that values obtained for DNA migration are not absolutely accurate. This, coupled with errors in compiling curves with lines of best fit, means that fragment sizes have an approximate error of $\pm$ 5 kb. For map construction the 5' end is defined as 'left' and the 3' end is defined as 'right'.

5.4.2 Physical Maps of *P. falciparum* Chromosome 6

A schematic illustration of the YAC's identified in the course of this study on the basis of STS content is presented in figure 5.3. Physical mapping of each YAC will be presented in detail in addition to ordering of YAC's into overlaps

5.4.2.1 Physical Map of YAC's Y9/H2 and Y10/H2.

The size of Y9/H2 is estimated at 160 kb. *Nar*I digestion reveals a 50 kb fragment that hybridises with the left arm probe and a 110 kb fragments hybridising with the right arm probe. Therefore *Nar*I maps to a site 50kb from the 5' end. *Xho*I digestion reveals a 55kb fragment with a left arm probe and 100 kb fragment with the right arm. Therefore *Xho*I maps to a site 55 kb from the 5` end. *Sac*I digestion reveals a single site 70kb from the left arm and 80 kb from the right arm and thus maps

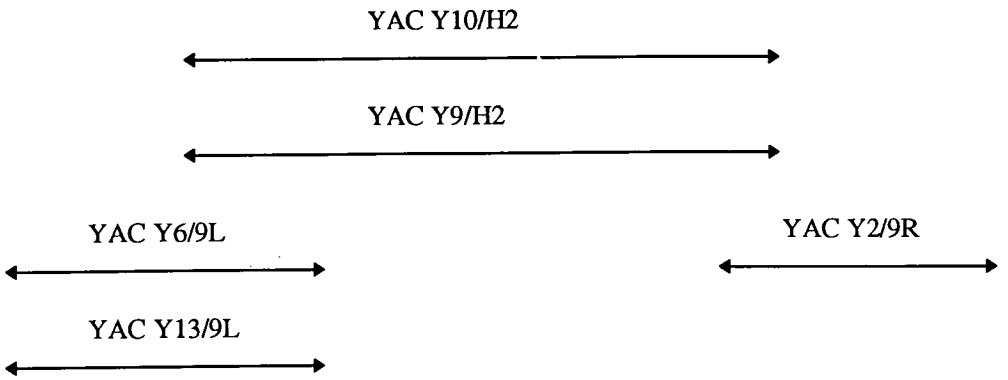
Figure 5.3 Schematic of YAC contigs constructed in this study.

This illustrates how the YAC's have been assembled based on their STS content. The figure is a diagrammatic representation and is not drawn to any scale. The YAC's have been identified by the following STS markers:

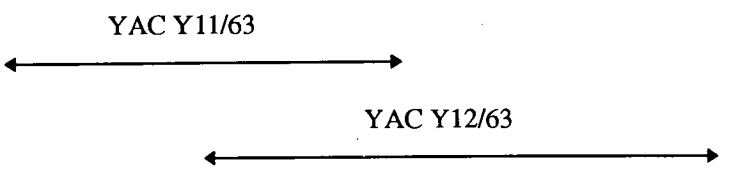
YAC		STS
H2A	-	Y9/H2, Y10/H2
H2/9L	-	Y6/9L, Y13/9L
H2/9R	-	Y2/9R
pSC11.63	-	Y11/63, Y12/63
pSC11.87	-	Y5/87
87/5R	-	Y2/5R

A, B and **C** refer to the sets of contigs that have been constructed based on primary STS screens and overlap assemblies with end- rescued STS markers.

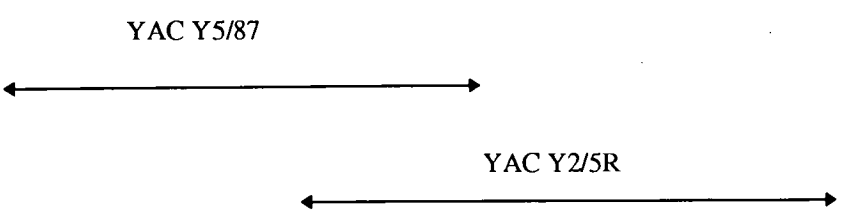
A.


YAC Y10/H2
YAC Y9/H2
YAC Y6/9L
YAC Y2/9R
YAC Y13/9L

B.


YAC Y11/63
YAC Y12/63

C.


YAC Y5/87
YAC Y2/5R

to a single site 70 kb from the 5` end. There are two *Nhe*I sites identified by the left arm probe at 80 and 90 kb. Two corresponding sites are identified by the right arm probe at 70 and 60 kb. Thus although both fragments indicate a YAC size of 150 kb, the 10 kb interval is constant and so the discrepancy can be attributed to errors in fragment size measurement. Therefore *Nhe*I maps to two sites located 80 and 90 kilobases from the 5` end. *Pvu*II digestion reveals sites 130 and 80 kb from the left arm and 30 and 70 kb from the right, therefore sites are assigned 80 and 130 kilobases from the 5` end. *Pst*I sites have been identified at 110kb, 95 kb, 50kb and 30 kb from the right end. Left arm probes reveals two sites 40 and 50 kb from the left end. The right arm fragments are strong signals, whereas the left arm fragments are comparatively weak. *Pst*I maps to two confirmed sites, 40 and 50 kilobases from the 5' end, the other sites have been omitted from the map pending further analysis.

H2A maps to a 50 kilobase *Nar*I fragments and is thus located in the 5` 50 kb. In addition the STS maps to 80 and 130 kb *Pvu*II fragments and 90 and 80 KB *Nhe*I fragments, indicating that the marker is in the 5' region of the YAC. The lack of markers in this region means that the marker can only be assigned to a 50 kilobase *Nar*I fragment. Map data for this YAC is presented in figure 5.4.

The size of Y10/H2 is estimated at 150 kilobases. Digestion with *Nar*I reveals a single site 50 kb from the left arm and 100 kb from the right arm and is therefore assigned a map location 50 kilobases from the 5` end. Digestion with *Nhe*I reveals multiple sites within the insert: left arm probes reveal sites at 35 kb, 80 kb, 90 kb, 110 kb and 120 kb. Right arm studies confirm sites at 35 kb, 45 kb, 65 kb and 120 kb, all of which, within $\pm$ 5 kb correspond with the left arm restriction patterns. *Nhe*I sites are therefore located approximately 30, 80, 90, 110 and 120 kilobases from the 5` end. A single *Sac*I site is observed 100 kb from the left arm and 55 kb from the right arm. Therefore, within the limits of experimental error, *Sac*I cuts once within the YAC 50 kilobases from the 5` end. *Pvu*II yields multiple fragments: sites are observed 40 kb, 50 kb, 80 kb and 100 kb from the left arm. Corresponding patterns are observed 55 kb, 70 kb, 90 kb and 110 kb from the right arm. With the exception of the site located at 50 kb left and 90 kb right, all sites correspond to the 150 kb size of the uncut YAC. The sites demonstrated show an error of 10 kb which may be accounted for by the errors in graph construction. *Pvu*II cuts at 40, 80 and 100 kilobases from the 5` end. There are no *Pst*I or *Xho*I sites within the insert.

H2A maps to a 100 kilobase *Nar*I fragment, which places the marker in the 3` 100 kb of the insert. In addition, the marker identifies the 120, 110, 90, 80 and 40 kilobase *Nhe*I fragments, but not the 35 kilobase fragment. This places H2A on a 30 kilobase *Nar*I/ *Nhe*I fragment. In addition H2A identifies 40, 55, 80 and 110 kb *Pvu*II fragments representing the 110 kilobase fragment from the right arm, the fragment obtained between the 100 kb and 50 kb sites from the 5` end, the 80 kb fragments from the 5' end and a 40 kb fragment produced by digestion at the 40 and 80 kb sites from the 5` end. Thus H2A maps to a 30 kilobase *Nar*I - *Nhe*I/*Pvu*II fragment. Map data for this YAC is presented in figure 5.5.

Figure 5.4 Physical Map of Y9/H2.

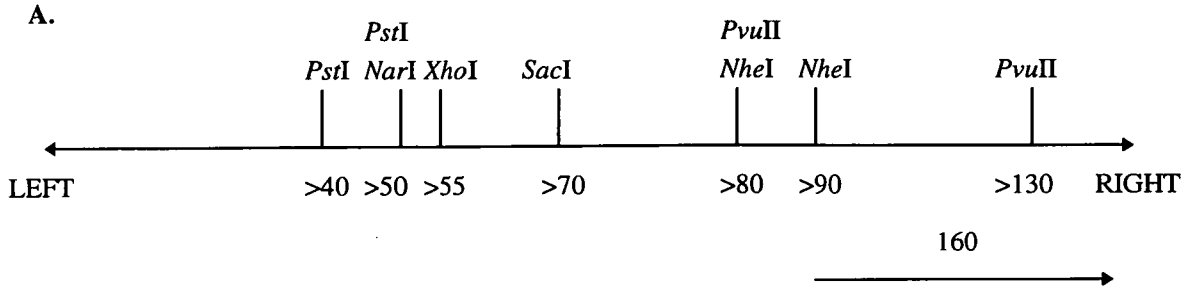
The size of the YAC is indicated in 5.4A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the H2A marker is indicated in figure 5.4C. The figures are as follows:

5.4A. Left Arm probe

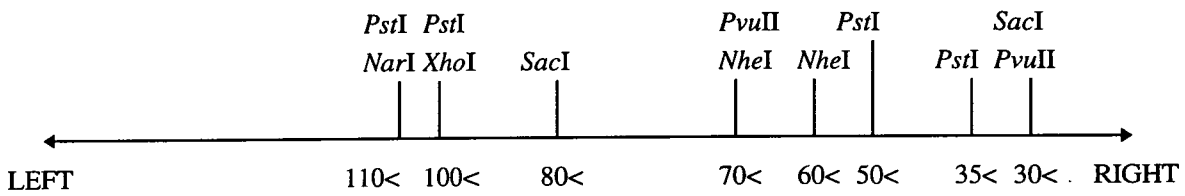
5.4B. Right Arm Probe.

5.4C. Composite Map based on both sets of probe data.

A.


PstI
PstI NarI XhoI
SacI
PvuII
NheI NheI
PvuII
LEFT
>40 >50 >55
>70
>80 >90
>130
RIGHT
160

B.


PstI PstI
NarI XhoI
SacI
PvuII
NheI NheI
PstI
SacI
PstI PvuII
LEFT
110< 100<
80<
70<
60< 50<
35< 30<
RIGHT

C.

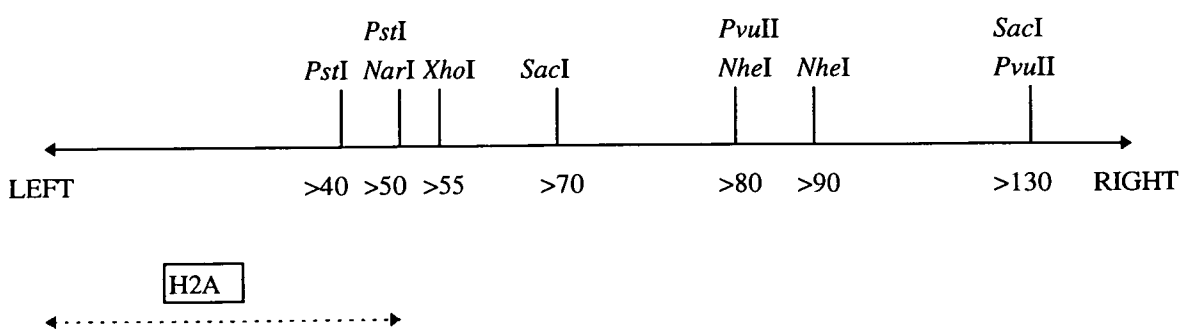

PstI
PstI NarI XhoI
SacI
PvuII
NheI NheI
SacI
PvuII
LEFT
>40 >50 >55
>70
>80 >90
>130
RIGHT
H2A

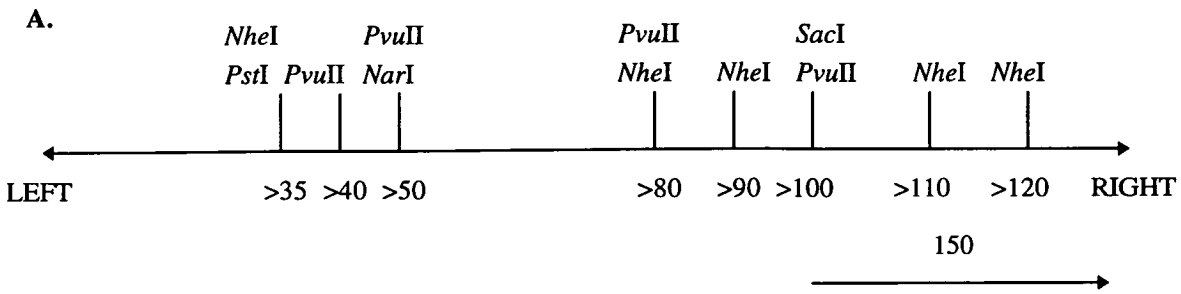
Figure 5.5 Physical Map of Y10/H2.

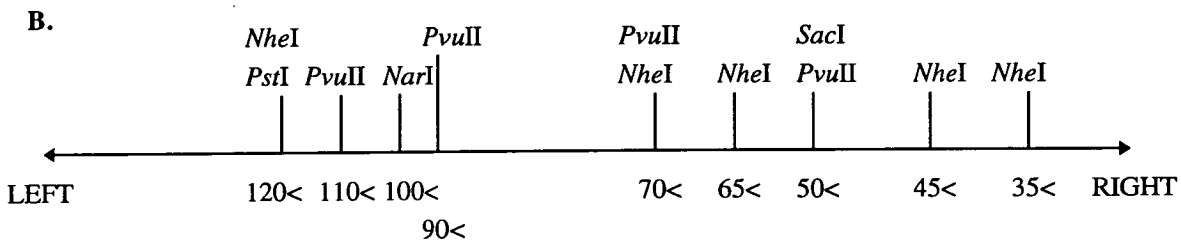
The size of the YAC is indicated in 5.5A. Figures underneath the restricion sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the H2A marker is indicated in figure 5.5C. The figures are as follows:

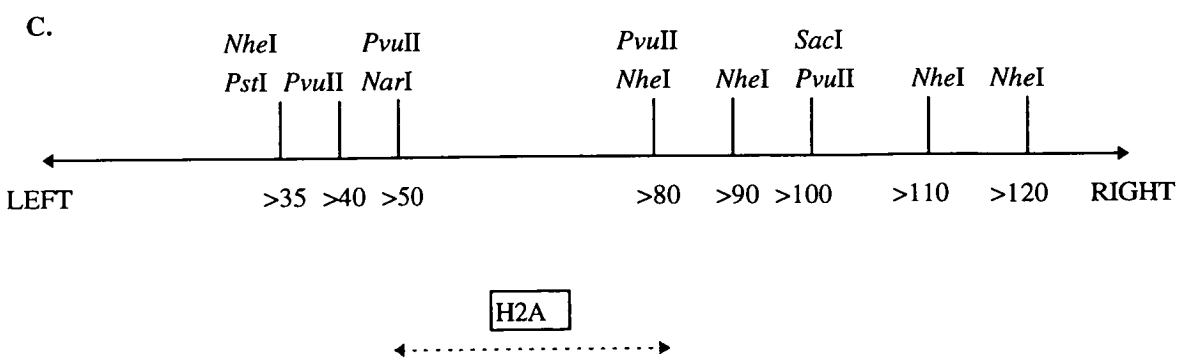
5.5A. Left Arm probe

5.5B. Right Arm Probe.

5.5C. Composite Map based on both sets of probe data


A.
NheI
PstI
PvuII
PvuII
NarI
PvuII
NheI
NheI
SacI
PvuII
NheI
NheI
LEFT
>35
>40
>50
>80
>90
>100
>110
>120
RIGHT
150


B.
NheI
PstI
PvuII
NarI
PvuII
PvuII
NheI
NheI
SacI
PvuII
NheI
NheI
LEFT
120<
110<
100<
90<
70<
65<
50<
45<
35<
RIGHT


C.
NheI
PstI
PvuII
PvuII
NarI
PvuII
NheI
NheI
SacI
PvuII
NheI
NheI
LEFT
>35
>40
>50
>80
>90
>100
>110
>120
RIGHT
H2A

Overlap assembly of the two YAC's is problematic. The presence of a *Nar*I site in both YAC's would seem to be a good landmark site. However an *Xho*I site is present 10 kilobases from the *Nar*I site in Y9/H2 whereas *Xho*I does not cut at all within YAC Y10/H2. In addition, there are multiple *Pst*I sites close to the *Nar*I site in Y9/H2 whereas there are none within close proximity in YAC Y10/H2. By comparison of the maps presented in figure 5.4, the lack of overlaps, in addition to the *Xho*I and *Pst*I data would suggest that either one or both of the YAC's identified by H2A are chimeric. Either analysis of a third YAC identified by H2A, repetition of existing map probe to confirm site location or mapping of H2A to genomic DNA with enzymes used to derive the maps would be necessary to identify the erroneous YAC.

5.4.2.2 Physical Map of YAC's Y11/63 and Y12/ 63

Y11/63 has an estimated size of 85 kilobases. Digestion with *Nhe*I reveals multiple sites 10, 20, 30, 65 and 80 kb from the left arm. The 80 kb fragment is undetectable on the right arm filter, which reveals sites at 30, 55, 65 and 80 kilobases, the latter site undetectable on the left arm filter. *Nhe*I therefore maps to sites 5, 20, 30, 60 and 80 kilobases from the 5` end. *Xho*I cuts once, 40 kb from the left arm and 45 kb from the right, leaving a single site 40 kilobases from the 5` end. *Sac*I cuts once, 75 kb from the left and 10 kb from the right arms, leaving a single site 75 kilobases from the 5` end. *Pst*I has cut sites at 30 kb and 80 kb from the left arm, and 55 kb from the right, yielding a pattern of sites 30 and 80 kilobases from the 5` end of the YAC. STS marker pSC11.63 identifies a 74 kb *Sac*I fragment and a 44 kb *Xho*I fragment, which localises the marker to the 3` end on the YAC. The STS also identifies a 55 kb *Pst*I fragment, which approximates to the fragment between the 5` 80 kb and the 5` 30 kb sites. In addition, the marker identifies a 55 kb and 30 kb *Nhe*I fragments. The 55 kb fragment is between 5` 80 kb and 5` 30 kb, but the 30 kb fragment could be either the 5` 60 kb site containing the right arm, or an internal fragment from 5` 60 to 5` 30 kb. The best mapping location possible is that pSC11.63 maps to a 30 kilobase *Xho*I - *Sac*I fragment and it remains unknown whether the marker is 5` or 3` of the 5` 60 kb *Nhe*I site. Maps for YAC Y11/63 are presented in figure 5.6.

YAC Y12/63 has an estimated size of 80 kilobases. *Xho*I cuts once leaving a single site 40 kilobases from the left and right arms. *Nhe*I shows multiple sites within the insert, yielding a pattern at 75, 50, 35 and 20 kb from the right arm and 30, 45, 60 and 70 kb from the left arm. Both 75 and 70 kb sites are visible only on one filter. The final order of *Nhe*I sites are at 5, 30, 45, 60 and 70 kilobases from the 5` end. *Sac*I cuts twice, at 55 kb from the right (corresponding to 25 kb from the left) and 60 kb from the left (corresponding to 20 kb from the right). *Pst*I cuts three times - 70 kb and 65 kb from the right arm and 70 kb from the left, yielding a pattern of cut sites 10, 15 and 70 kilobases from the 5` end. pSC11.63 identifies a 25 kilobase *Sac*I fragment, positioning the marker on the immediate 5` end. In addition, the marker identifies a 30 kilobase *Nhe*I fragment and a 40 kilobase *Xho*I fragment confirming the location of the marker to the terminal 5` end. The marker also identifies 55 kilobase *Pst*I fragment,

Figure 5.6 Physical Map of Y11/63.

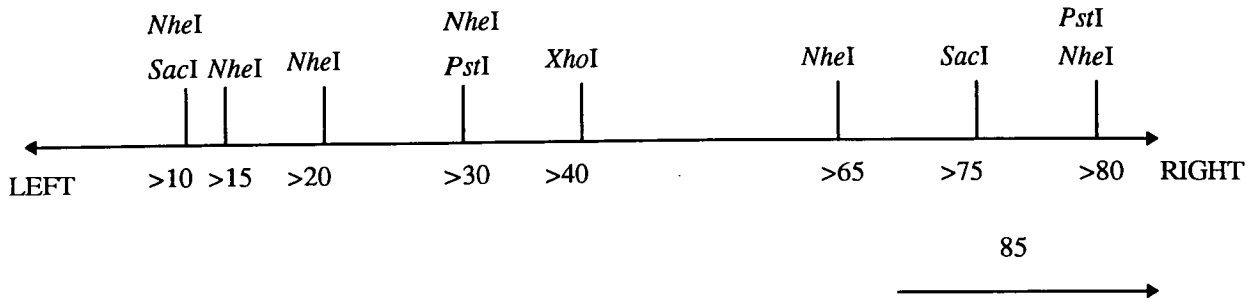
The size of the YAC is indicated in 5.6A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the pSC11.63 marker is indicated in figure 5.6C. The figures are as follows:

5.6A. Left Arm probe

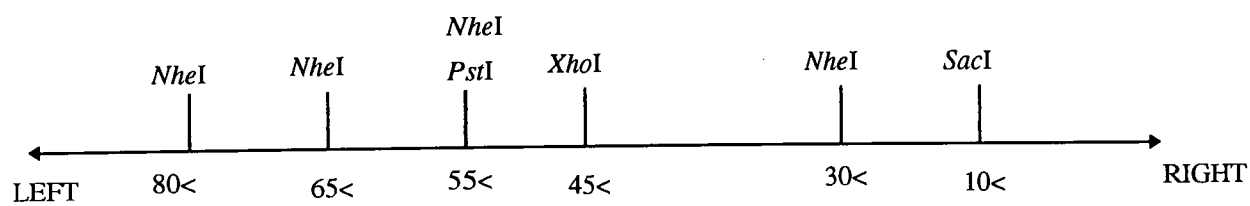
5.6B. Right Arm Probe.

5.6C. Composite Map based on both sets of probe data

A.


NheI
SacI NheI NheI
NheI
PstI
XhoI
NheI
SacI
PstI
NheI
LEFT
>10 >15 >20
>30
>40
>65
>75
>80
RIGHT
85

B.


NheI
NheI
NheI
PstI
XhoI
NheI
SacI
LEFT
80<
65<
55<
45<
30<
10<
RIGHT

C.

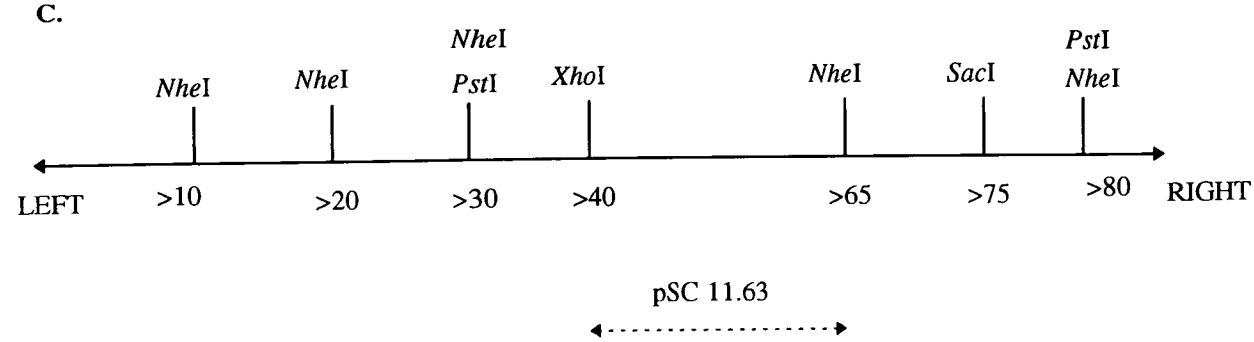

NheI
NheI
NheI
PstI
XhoI
NheI
SacI
PstI
NheI
LEFT
>10
>20
>30
>40
>65
>75
>80
RIGHT
pSC 11.63

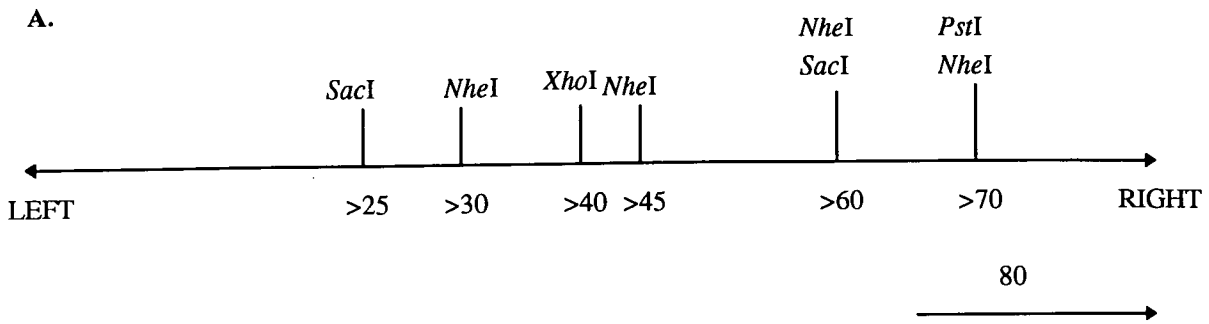
Figure 5.7 Physical Map of Y12/63.

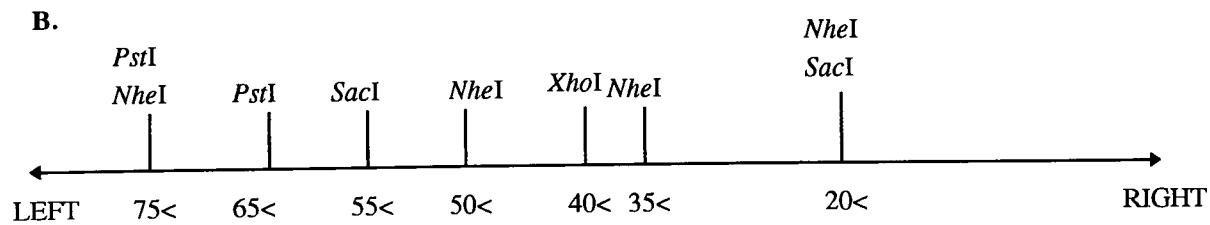
The size of the YAC is indicated in 5.7A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the pSC11.63 marker is indicated in figure 5.7C. The figures are as follows:

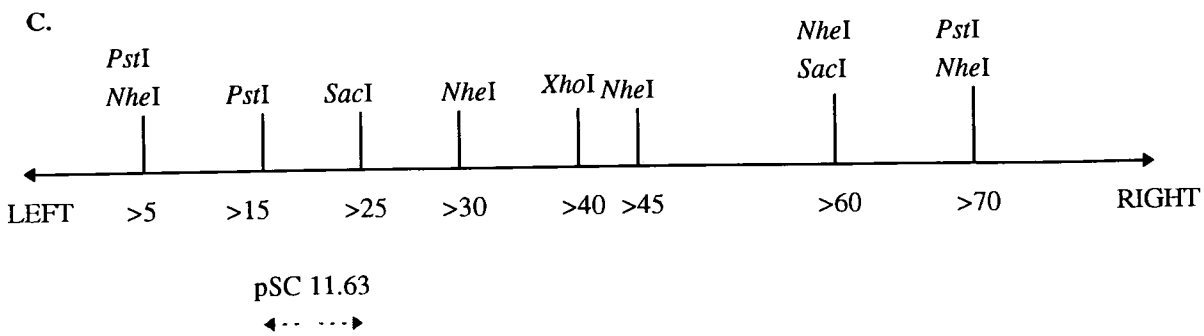
5.7A. Left Arm probe

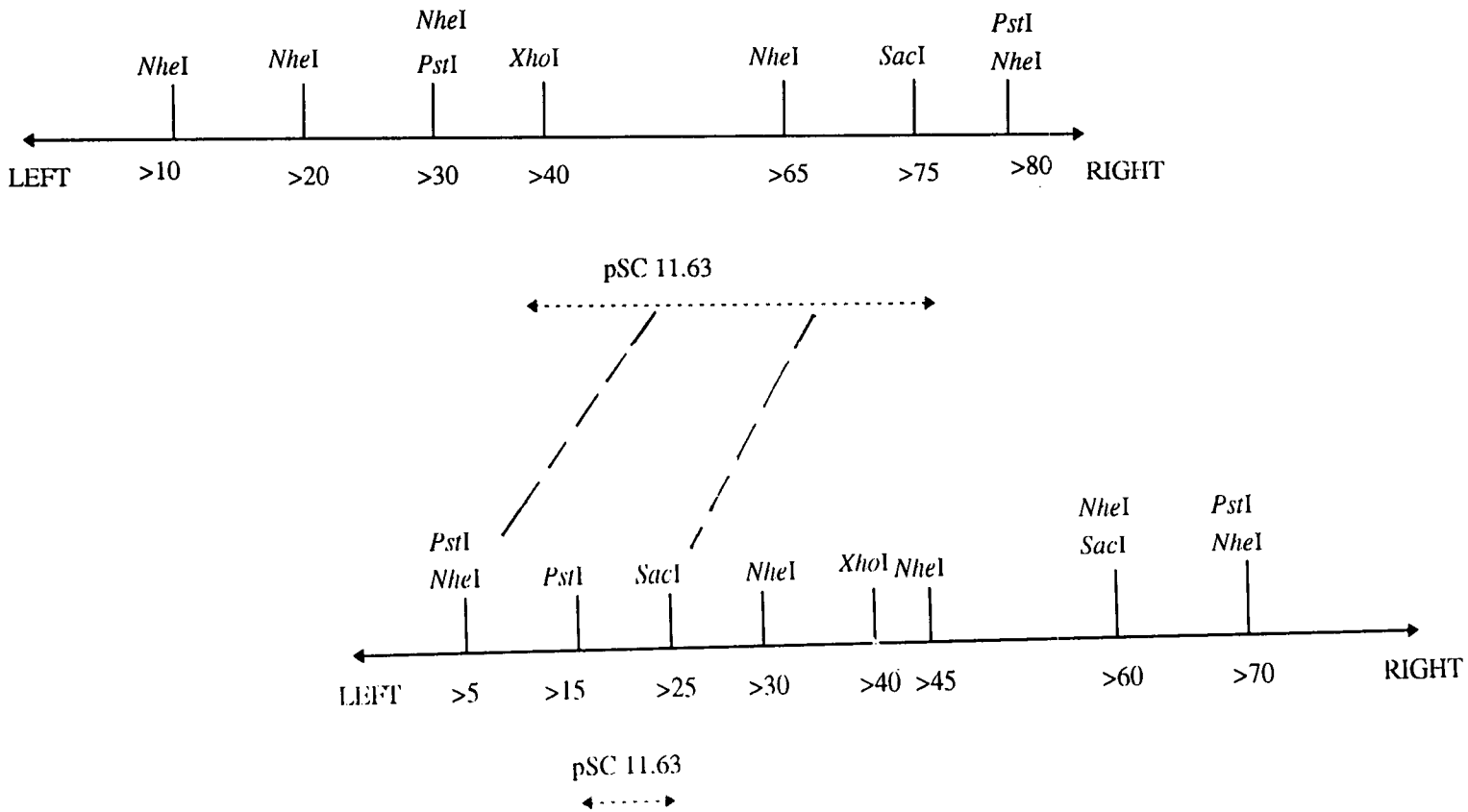
5.7B. Right Arm Probe.

5.7C. Composite Map based on both sets of probe data


A.
SacI
NheI
XhoI
NheI
NheI
SacI
PstI
NheI
LEFT
>25
>30
>40
>45
>60
>70
RIGHT
80


B.
PstI
NheI
PstI
SacI
NheI
XhoI
NheI
NheI
SacI
LEFT
75<
65<
55<
50<
40<
35<
20<
RIGHT


C.
PstI
NheI
PstI
SacI
NheI
XhoI
NheI
NheI
SacI
PstI
NheI
LEFT
>5
>15
>25
>30
>40
>45
>60
>70
RIGHT
pSC 11.63


NheI
NheI
NheI
PstI
XhoI
NheI
SacI
PstI
NheI
LEFT
>10
>20
>30
>40
>65
>75
>80
RIGHT
pSC 11.63
PstI
NheI
PstI
SacI
NheI
XhoI
NheI
NheI
SacI
PstI
NheI
LEFT
>5
>15
>25
>30
>40
>45
>60
>70
RIGHT
pSC 11.63

corresponding to the fragment produced between 5` - 70 kb and 5` - 15 kb. Thus the marker maps to a 10 kilobase *Sac*I - *Pst*I fragment. Maps for Y12/63 are presented in figure 5.7.

The optimal overlap that can be estimated for the two YACs is a direct overlap between the extreme 5` and 3` ends. YAC Y11/63 has a 3` order *Xho*I - *Nhe*I/pSC11.63 - *Sac*I - *Pst*I/ *Nhe*I. YAC Y12/63 has a 5` order *Nhe*I - *Pst*I- pSC11.63 - *Sac*I. The *Nhe*I and *Sac*I sites overlap, but the *Pst*I site is anomalous. No explanation can be offered for such results, other than the *Pst*I site is erroneous. this indicates an overlap of 30 kilobases between the 3` end of YAC Y11/63 and the 5` end of YAC Y12/63. This indicates a contig of 135 kilobases for both YAC's. Composite maps for YAC's Y11/63 and Y12/63 are presented in figure 5.8.

5.4.2.3 Physical Map of YAC Y5/87.

The YAC has an estimated size of 80 kilobases. *Pst*I is seen to produce four clear bands with the left arm probe, but none is seen with the right. The clarity of the left arm signals would seem to indicate that further investigation is needed with right arm probes to clarify this. *Pst*I signals are seen 10, 15, 25 and 30 kilobases from the left arm. *Nhe*I cuts four times, at 25, 30, 35 and 40 kilobases from the left arm. Two bands are visible with the right arm probe at 40 and 45 kb, corresponding to the 40 and 35 kb left arm. However, in view of the inefficiency of the right arm data for this YAC. the left signals will be taken as accurate. *Sac*I cuts once, producing a 40 kilobase fragment identifiable by both left and right arms. STS marker pSC11.87 identifies a 40 kb *Sac*I fragment, which could be either 5` or 3`. In addition a 60 kilobase *Pst*I fragment is identified, which could correspond to the 3` end of the insert. 35, 40, 55 and 60 kb *Nhe*I inserts are identified, which would seem to indicate that the marker is on the 3` 40 kilobase *Nhe*I/*Sac*I fragment, as larger *Nhe*I sites are being identifies than seen with the left arm probe. The final order of sites in this YAC is given in figure 5.9.

5.4.2.4 Physical Map of YAC's Y6/9L and Y13/9L

Both YAC's were identified with the STS H2/9L, the marker obtained form the left arm rescue of YAC Y9/ H2. YAC Y6/9L has an estimated size of 85 kilobases. *Sac*I has multiple sites within the insert 35, 45, 60 and 75 kb from the left arm, and 5, 20, 35 and 45 kb from the right. All these sites align, yielding a final pattern of *Sac*I 5` - 35, 45, 60 and 75 kilobases. *Nhe*I has multiple cut sites within the insert, at 20, 45, 50 and 60 kb from the left arm corresponding to 15, 30, 35, 65 and 80 kb from the right. This latter site is not visible with a left arm probe. This yields a final digest pattern for *Nhe*I of 5` 5, 20, 45, 50, and 65kb. *Pvu*II cuts twice within the insert, 50 kb and 70 kb from the left arm, corresponding to 11 kb and 30 kb from the right. This yields a pattern of 5` 50 and 70 kilobases for *Pvu*II. No data was obtained for *Xho*I, *Nar*I or *Eag*I. STS H2/9L maps to a 50 kilobase *Pvu*II fragment, which maps the marker to the 5` fragments. In addition the marker identifies a 15 kilobase *Nhe*I fragment, produced by the cut sites at 5` 45 kb and 5` 60 kb. In addition, the marker identifies 20, 40

Figure 5.9 Physical Map of Y5/87.

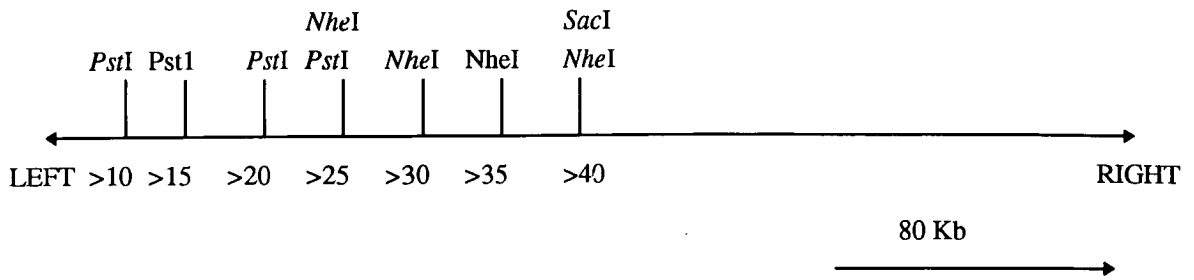
The size of the YAC is indicated in 5.9A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the pSC11.87 marker is indicated in figure 5.9C. The figures are as follows:

5.9A. Left Arm probe

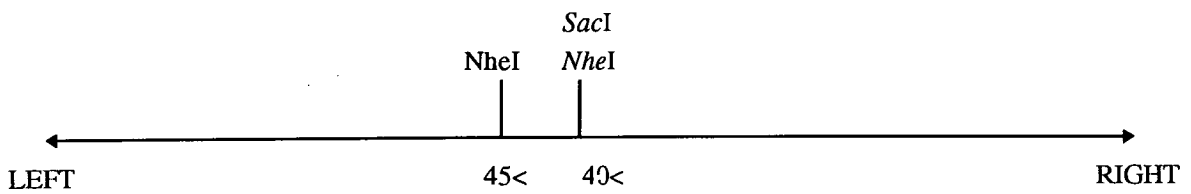
5.9B. Right Arm Probe.

5.9C. Composite Map based on both sets of probe data

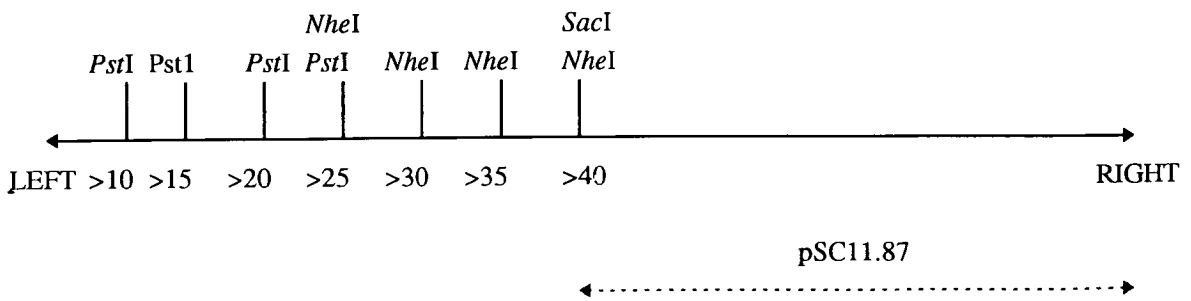
A.


NheI
SacI
PstI Pst1 PstI PstI NheI NheI NheI
LEFT >10 >15 >20 >25 >30 >35 >40
RIGHT
80 Kb

B.


SacI
NheI NheI
LEFT
45<
40<
RIGHT

C.


NheI
SacI
PstI Pst1 PstI PstI NheI NheI NheI
LEFT >10 >15 >20 >25 >30 >35 >40
RIGHT
pSC11.87

and 50 kb *Sac*I fragments indicating the marker to be present on a 5 kilobase *Pvu*II - *Sac*I fragment. Physical maps for this YAC are presented in figure 5.10.

YAC Y13/9L is estimated to be 90 kilobases. *Sac*I has multiple sites within the insert 20, 40, 55 and 70 kb from the left arm and 35, 50 and 70 kb from the right. This results in a final digest pattern of 5` 20, 40, 55 and 70 kilobases. *Eag*I and *Nar*I cut once, both at 70 kilobases from the left arm. Due to gel separation, the corresponding 20 kb fragment is not detected. *Nhe*I cuts once, defining a 60 kb left arm and 30 kb right arm fragment. *Pst*I cuts once, defining a 75 kb right arm fragment that has no corresponding band in the left arm filter. *Xho*I cuts twice yielding a 50 kb and 85 kb left arm fragments. Neither band is visible in the right arm filter, which due to the signal intensity in the left arm filter could be ascribed to hybridisation difficulties. Due to technical limitations, no marker assignment could be performed for this YAC. Physical maps for this YAC are presented in figure 5.11.

Overlap assembly of the YAC's with the parent YAC Y9/H2 cannot be performed, as there is not enough map data in the 5' 50 kilobases of Y9/H2 for an accurate overlap to be made possible. Isolation of further YAC's spanning this region or subcloning into cosmids to provide a higher resolution physical map of this region would be required to order the contig.

5.4.2.5 Physical Map of YAC Y2/9R

The estimated size of the YAC is 80 kb. *Sac*I cuts once, defining a 25 kb fragment from the left arm and a corresponding 55 kb fragment from the right arm. *Nhe*I cuts once, defining a 50 kb left arm fragment and a 35 kb right arm fragment. *Pst*I cuts three times, defining fragments 60, 70 and 75 kb from the left arm, and corresponding 10, 15 and 25 kb fragments from the right arm, which align given experimental error. The final order of sites 5` to 3` is, *Sac*I (25 kb), *Nhe*I (50 kb), *Pst*I (60 kb), *Pst*I (70 kb) and *Pst*I (75kb). Despite numerous attempts, no internal STS data was available. Future work would involve hybridisation of these filters with the H2/9R STS marker. Map data for this YAC is presented in figure 5.12. Overlap assembly cannot be performed due to the lack of physical map information in the 5` 25 kb of YAC Y2/9R. Isolation of YAC's spanning this region or subcloning strategies would be required to complete this contig.

5.4.2.6 Physical Map of YAC Y2/5R

The size of the YAC is estimated at 100 kilobases. Only two enzymes appear to cut within this insert: *Nhe*I cuts four times defining fragments 15, 20, 40 and 50 kb from the left arm and corresponding 50, 55, 80 and 90 kb from the right arm. *Pst*I cuts twice, 70 and 80 kb from the left arm, corresponding to 25 kb and 20 kb from the right. this leaves a final order of sites 5` *Nhe*I (15 kb), *Nhe*I (20 kb), *Nhe*I (45 kb), *Nhe*I (50 kb), *Pst*I (70 kb) and *Pst*I (80 kb). The STS marker used to identify the YAC, 87/5R, maps to a 70 kilobase *Pst*I fragment, locating the marker to the 5` *Pst*I fragment. In addition, the STS defines 50 and 30 kilobase *Nhe*I fragments, which would place the marker between

Figure 5.10 Physical Map of Y6/9L.

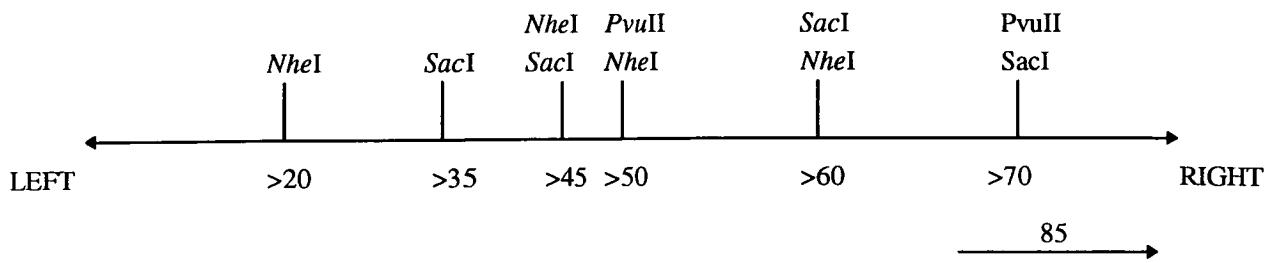
The size of the YAC is indicated in 5.10A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the H2/9L marker is indicated in figure 5.10C. The figures are as follows:

5.10A. Left Arm probe

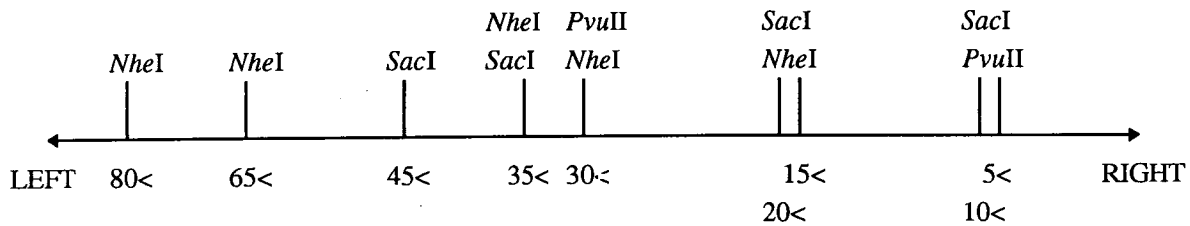
5.10B. Right Arm Probe.

5.10C. Composite Map based on both sets of probe data

A.


NheI
SacI
NheI SacI
PvuII NheI
SacI NheI
PvuII SacI
LEFT
>20
>35
>45
>50
>60
>70
RIGHT
85

B.


NheI
NheI
SacI
NheI SacI
PvuII NheI
SacI NheI
SacI PvuII
LEFT
80<
65<
45<
35<
30<
15<
20<
5<
10<
RIGHT

C.

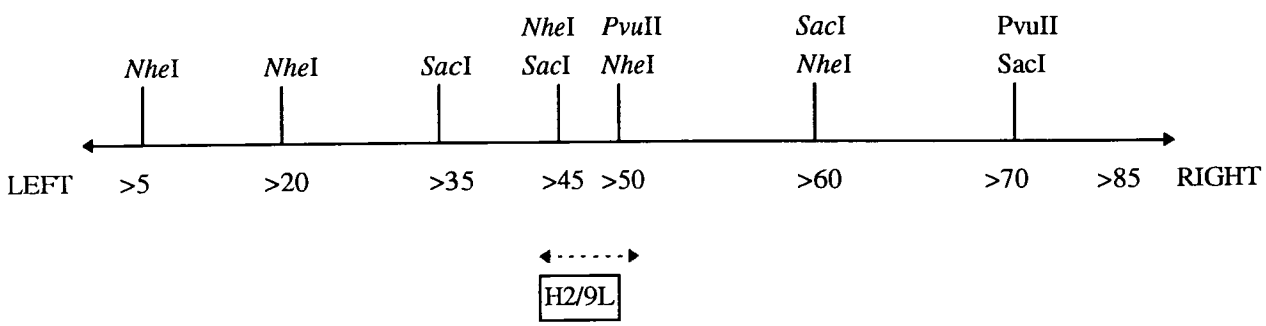

NheI
NheI
SacI
NheI SacI
PvuII NheI
SacI NheI
PvuII SacI
LEFT
>5
>20
>35
>45
>50
>60
>70
>85
RIGHT
H2/9L

Figure 5.11 Physical Map of Y13/9L.

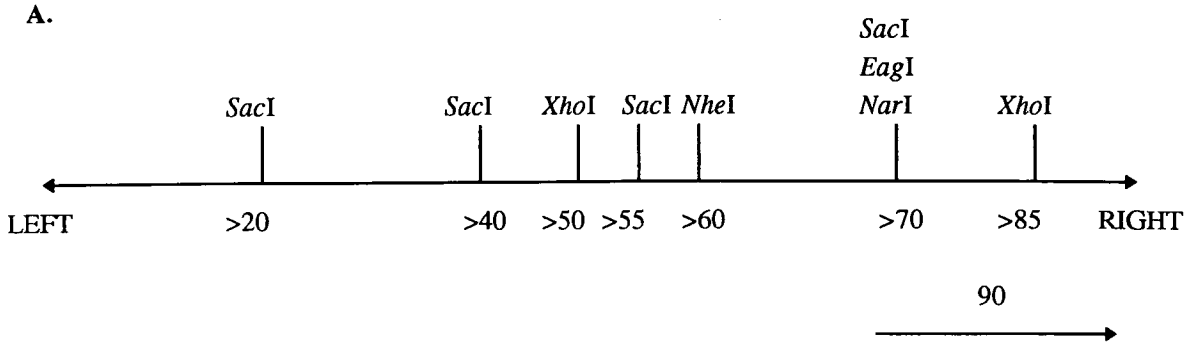
The size of the YAC is indicated in 5.11A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. No location for the H2/9L marker could be determined. The figures are as follows:

5.11A. Left Arm probe

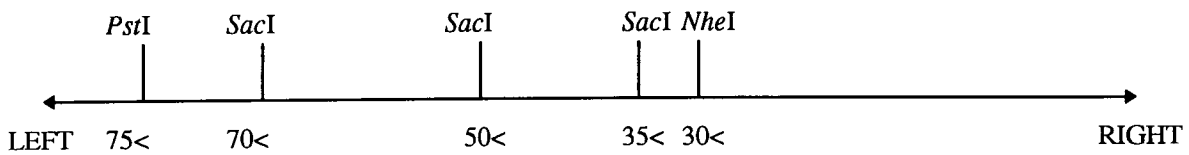
5.11B. Right Arm Probe.

5.11C. Composite Map based on both sets of probe data

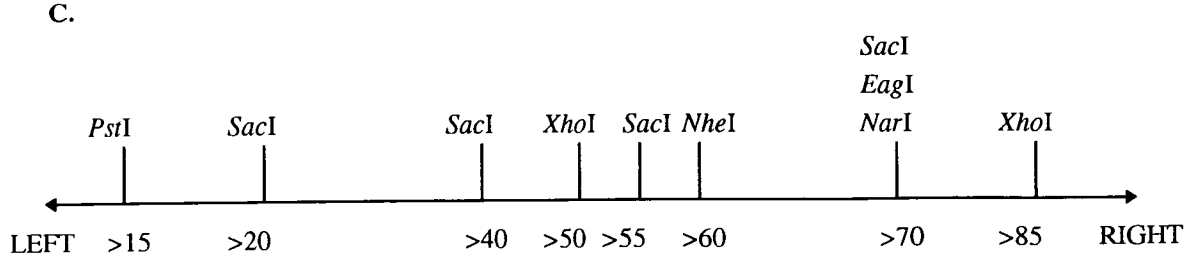
A.


SacI
EagI
NarI
SacI
SacI
XhoI
SacI
NheI
XhoI
LEFT
>20
>40
>50
>55
>60
>70
>85
RIGHT
90

B.


PstI
SacI
SacI
SacI
NheI
LEFT
75<
70<
50<
35<
30<
RIGHT

C.


SacI
EagI
NarI
PstI
SacI
SacI
XhoI
SacI
NheI
XhoI
LEFT
>15
>20
>40
>50
>55
>60
>70
>85
RIGHT

*Nhe*I sites at 5` 20 kb and 5` 50 kb. Thus the marker maps to a 25 kilobase *Nhe*I fragment. Physical map data is presented in figure 5.13. Overlap assembly between the YAC and the parent YAC Y5/87 cannot be performed due to the lack of physical map data in the 3' 40 kilobases of the YAC

5.5. Discussion.

One of the advantages of the Artificial Chromosome cloning system is the ability to generate physical map data in the absence of any marker information. The presence of genomic DNA markers has in the past facilitated the construction of long range physical maps. Probes specific to the left and right vector arm allow composite maps of any YAC insert to be constructed. If any marker is subsequently identified then that marker can be rapidly mapped to a restriction fragment.

The work presented details the construction of high resolution physical maps of all cloned *P. falciparum* HB3 YAC's. Several features became apparent. Due to the small insert size of YAC's, conventional enzymes used in map construction e.g. *Sma*I, *Not*I and *Apa*I proved uninformative for artificial chromosome mapping. In order to generate useful information from the YAC's, enzymes which recognise hexanucleotide sequences containing 4 $^{G}/_{C}$ and 2 $^{A}/_{T}$ cut frequently enough to generate maps with a 10 kilobase resolution, and clustering or dispersal of sites may give indications of the presence of gene sequences. Despite technical difficulties associated with physical mapping in YAC's, map construction per *P. falciparum* YAC involving left and right sequence data was, in general, reasonably internally consistent, with few anomalous sites being observed. Such erroneous results could be ascribed to inefficient partial digestion, poor resolution due to inefficient pulsed field gel resolution or poor transfer of DNA fragments to membranes. In addition, inadequate labelling of probes and inefficient removal of probes prior to successive hybridisation resulting in apparent presence of restriction sites can also give rise to YAC mapping errors, and may have contributed to spurious results, especially after a third successive hybridisation of a single filter.

More problematic was the assembly of YAC's into overlapping contigs. Two YAC's identified by the H2A locus showed evidence of chimerism, although the marker rescued from the left arm mapped to chromosome 6 (data not shown). The presence of an *Xho*I site close to the gene locus in one YAC which is not detected in the other indicates that the YAC contains non- contiguous DNA. An investigation of chimerism within the HB3 YAC's will be necessary before any maps can be presented with any degree of confidence. If co- cloning events become more common, then second generation library construction strategies might have to be considered, e.g. constructing YAC libraries from defined DNA sources (i.e. isolated chromosomes) or in recombination deficient host strains. Only one set of YAC's have been assembled into overlaps, those identified by the marker pSC11.63. More YAC's would need to be identified per marker, or alternatively an STS content approach to map construction using

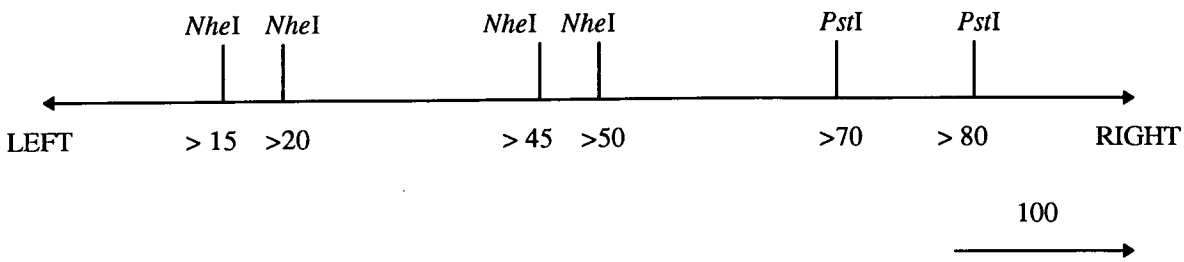
Figure 5.12 Physical Map of Y2/9R.

The size of the YAC is indicated in 5.12A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. Due to technical difficulties, no internal STS map position was determined. The figures are as follows:

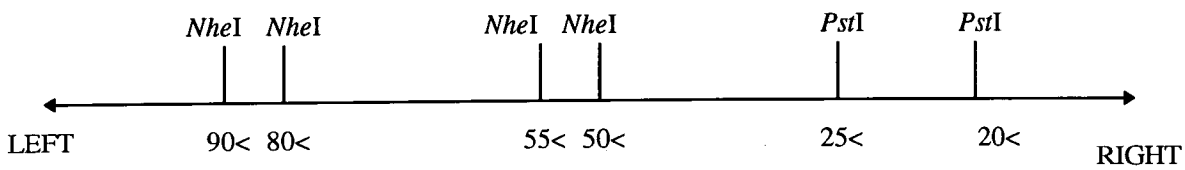
5.12A. Left Arm probe

5.12B. Right Arm Probe.

A.


NheI NheI NheI NheI PstI PstI
LEFT
> 15 >20 > 45 >50 >70 > 80
RIGHT
100

B.


NheI NheI NheI NheI PstI PstI
LEFT
90< 80< 55< 50< 25< 20<
RIGHT

C.

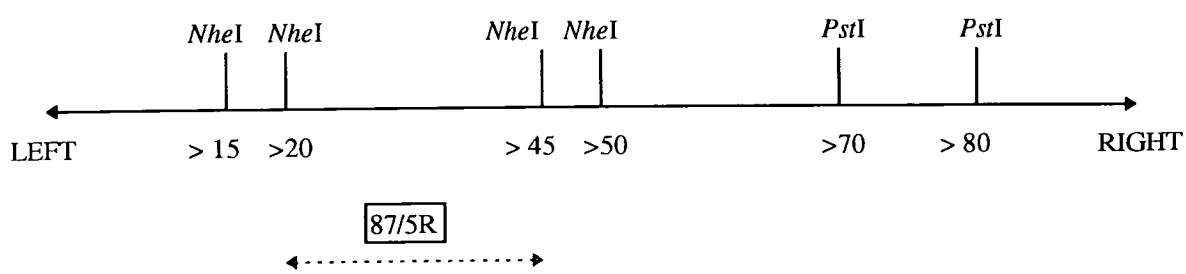

NheI NheI NheI NheI PstI PstI
LEFT
> 15 >20 > 45 >50 >70 > 80
RIGHT
87/5R

Figure 5.13 Physical Map of Y2/5R.

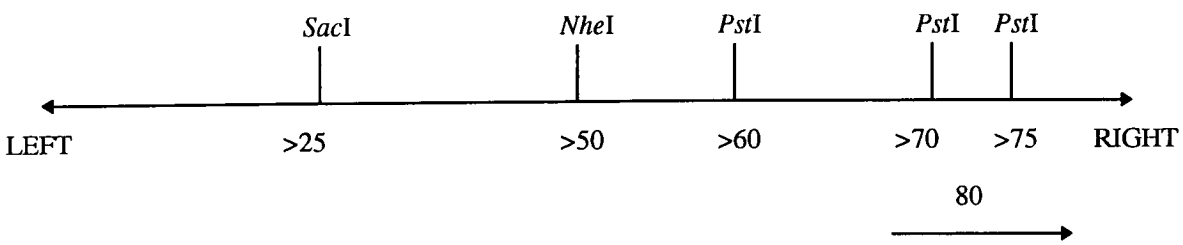
The size of the YAC is indicated in 5.13A. Figures underneath the restriction sites indicate the size in Kb of that site from the respective YAC arm. The suspected location of the 87/5R marker is indicated in figure 5.13C. The figures are as follows:

5.13A. Left Arm probe

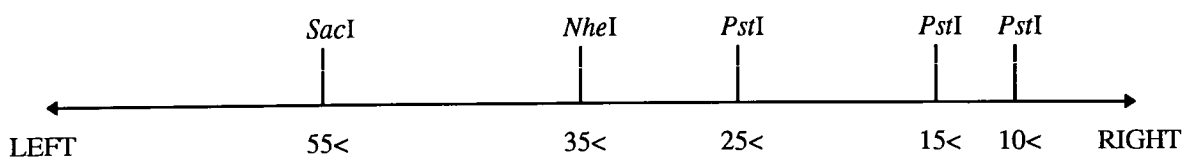
5.13B. Right Arm Probe.

5.13C. Composite Map based on both sets of probe data

A.


SacI
NheI
PstI
PstI
PstI
LEFT
>25
>50
>60
>70
>75
RIGHT
80

B.


SacI
NheI
PstI
PstI
PstI
LEFT
55<
35<
25<
15<
10<
RIGHT

multiple chromosome- specific markers would be required to rapidly identify chimerism or deletions within YAC's.

In conclusion, YAC mapping has been used to efficiently construct highly detailed maps for *P. falciparum* chromosome 6, which will allow the rapid positioning of any newly identified markers onto small restriction fragments and thus will enable researchers to not only be provided with the YAC containing the gene, but also a detailed map location of the gene on this chromosome.

CHAPTER 6.

AN INVESTIGATION OF GENE ORGANISATION AND EXPRESSION WITHIN *P. falciparum* CHROMOSOME 6.

6.1 Histone Gene Organisation on Chromosome 6.

6.1.1. Abstract.

PCR primers designed from the published Histone 3 sequence were used to screen the HB3 YAC MasterPools. Results indicated the gene sequence is present in multicopy within the library and amplifies pools 9 and 10, known to contain Histone 2A. Further PCR analysis indicated the primers amplified both YAC's containing the Histone 2A locus and did not amplify non- H2A containing YAC's. Hybridisation analysis with labelled H3 probes confirmed that H3 did not cross- hybridise to H2A and was subsequently used to probe the H2A- containing YAC digest filters. The results showed that H3 maps to the same 30 kb *Nhe*I/ *Nar*I fragment in YAC Y10/H2 as H2A, and to the same 50 kb *Xho*I/ *Nar*I fragment in YAC Y9/H2 as H2A. This would seem to indicate a functional clustering of the histone genes on chromosome 6 in *P. falciparum*

6.1.2. Introduction.

The arrangement of gene sequences into clustered units is one of the features which distinguish the arrangement of genetic information within prokaryotes from that of eukaryotes. Clustering can be a response in the organisation of developmentally expressed genes, as in the α- like globin genes (Lauer *et al.*, 1980). Alternatively, clustering can be a mechanism to create large amounts of messenger RNA for proteins required in abundance in the cell. Examples of such genes include tRNA genes (Hovemann *et al.*, 1980; Fostel *et al.*, 1984), rRNA genes (Worton *et al.*, 1988), and histones (Hentschel and Birnsteil, 1981). In addition, genes which demonstrate common transcriptional expression demonstrate clustering, as in *Drosophila melanogaster* heat shock proteins (Corces *et al.*, 1980). Investigations into gene organisation in *P. falciparum* have demonstrated evidence of gene clustering, including a cluster of nine tRNA genes between copies of rRNA genes in each inverted repeat of the 35 kb extrachromosomal element (Wilson *et al.*, 1991; Gardner *et al.*, 1994a). In addition, rRNA genes are clustered as sets of 18S, 5.8S and 25S genes, with each set repeated 4 - 8 times at various chromosomal locations throughout the genome (Weber *et al.*, 1988a), and small rRNA fragments have been found to be present on the extrachromosmal 6 kb element (Feagin, 1994), with six fragments homologous to the small subunit rRNA and seven fragments homologous to the large subunit rRNA

Histones are ubiquitous in eukaryotes. Their products, a family of five small highly conserved proteins (H1, H2A, H2B, H3 and H4) serve to organise DNA into its fundamental unit, the nucleosome, and subsequently into higher order chromatin structures. The structures of histones, particularly H3 and H4, are generally highly conserved between organisms, even between animal and plants, and reflect the fundamental role performed by the proteins within the cell (Maxson *et al.*, 1983). H2A, H2B, H3 and H4

form an octamer core around which DNA coils to form the nucleosome, with H1 involved in chromatin formation (Creedon *et al.*, 1992). Copy number of the histone gene cluster varies dramatically between organisms, from a repeat frequency of 400 copies for *Psammechinus miliaris* (Hentschel and Birnsteil, 1981) and 100 copies for *Drosophila melanogaster* (Maxson *et al.*, 1983) to 10 copies for chickens (Harvey *et al.*, 1981). The high copy number of the histone gene clusters in organisms such as sea urchins amphibians and *Drosophila* may be due to the need for rapid cell division and accompanying DNA synthesis during early development. In mammals and most other higher eukaryotes, cell division during early embryogensis is not so rapid and fewer copies of the histone genes are needed. During several stages of the *P. falciparum* life cycle, cell division is also very rapid i.e. during male gametogenesis, exflagellation (3 nuclear doublings in 10 mins), erythrocytic schizogony (5 doublings in ~ 5 hours) and exo- erythrocytic schizogony. Sequences corresponding to histones 2A (Creedon *et al.*, 1992), 2B (Bennett *et al.*, 1995) and H3 (Longhurst and Holder, 1995; Bennett *et al.*, 1995) have been reported for *P. falciparum*. The genes demonstrate features common to all other histones identified - H2A and H3 lack introns and posses a 5` polyadenylation signal. No data was presented for H2B. Both H2A and H3 are present in single copy and have been mapped to chromosome 6 whereas H2B has been mapped to chromosome 11. H3 has been reported to be present in multicopy (Bennett *et al.*, 1995), which in view of the well documented stoichiometric nature of the histone gene family is a curious observation (Maxson *et al.*, 1983). YAC's known to contain the H2A sequence were analysed on the basis of their STS content for sequences corresponding to H3.

6.1.3. Mapping of the Histone 3 Gene to HB3 YAC's.

PCR primers designed from the published sequence were used to screen all 16 plate pools of the HB3 YAC library. Using PCR conditions of annealing at 55°C for 60 seconds over 35 cycles of amplification, positive 300 bp signals were observed in pools 9, 10 and 11 (data not shown). This corresponds to the results observed for the H2A primers, which detected signal in pools 9 and 10. Standard controls of HB3 and 3D7 were positive, AB1380 and water were negative. The observance that H3 amplified the same plates as H2A led to a second round of amplification which included YAC DNA as template. DNA from YAC's containing the H2A gene sequence, Y9/H2 and Y10/H2, were included, as well as YAC's overlapping the 5` and 3` arms of YAC Y9/H2 which should not contain the H2A sequence and should therefore act as negative controls. The results in figure 6.2.1A demonstrate that the H3 primers amplify a sequence in YAC's Y9/H2 and Y10/H2, but do not amplify Y13/9L or Y2/9R. This indicates that H3 is on the same region of the chromosome as H2A.

6.1.4. Mapping the H3 gene Within the YAC's.

Before mapping experiments could be performed it was necessary to confirm that the H3 gene did not cross hybridise with the H2A gene. H3 primers were used to amplify the gene from a genomic

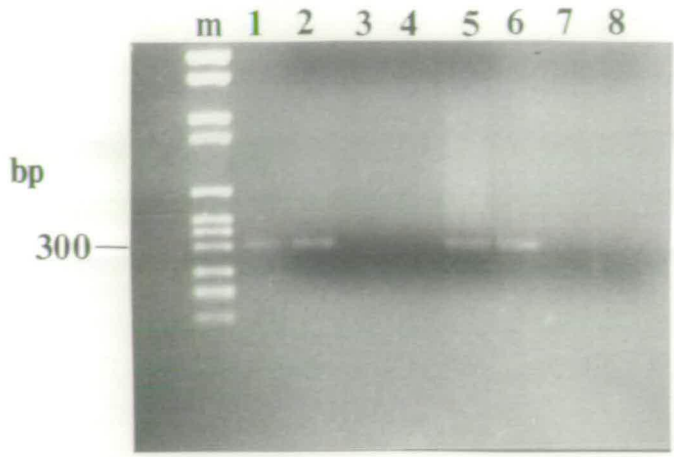
Figure 6.1.1. Analysis of Histone Gene Organisation in *P. falciparum*.

A. PCR reaction with H3 primers with various YAC DNA as templates. Amplification occurred under 550C annealing conditions for 35 cycles. 10 µl of PCR product was resolved on a 1% IBI agarose gel in 1X TAE and photographed under UV.

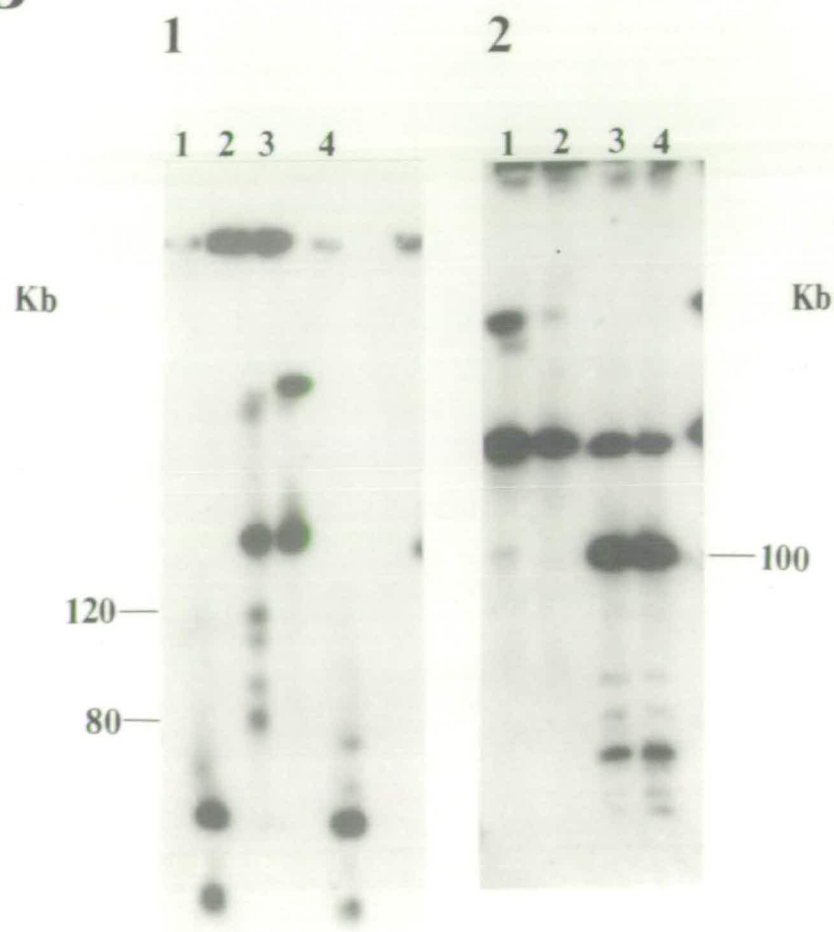
Lane m: DNA Size standard Marker VI (Boehringer)
Lane 1: amplification with H2A- containing YAC Y9/H2 as template
Lane 2: amplification with H2A- containing YAC Y10/H2 as template
Lane 3: amplification with non- H2A containing YAC Y2/9R as template
Lane 4: amplification with non- H2A containing YAC Y13/9L as template
Lane 5: amplification with HB3 genomic DNA
Lane 6: amplification with 3D7 genomic DNA
Lane 7: amplification with AB1380 genomic DNA
Lane 8: amplification with water control.

B. Physical mapping of H3 onto the H2A YAC Y10/H2 digest filters. (1): Y10/H2 digested with *Nhe*I and probed with H3. Lanes 1 - 4 correspond to 15, 0.3, 0.03 and 3 U of enzyme. Pulsed field gel conditions for resolution were 0.9% IBI Agarose in 0.5X TBE with a 2 - 12 second ramp for 24 Hours at 175V.(2) YAC Y10/H2 digested with *Nar*I and probed with H3. The 100 Kb fragment is clearly visible, as well as anomalous smaller fragments. Lanes **1 - 4** correspond to digestion with 0.03, 0.3, 3 and 15U of enzyme. Fragments were resolved on a 1% IBI agarose pulsed field gel with a 2- 12 second ramp for 24 hours at 175V

A



B



HB3 DNA template. The amplified product was resolved on a 1% gel, blotted and probed with a ^{32}P labelled H2A probe. The results demonstrated that H2A does not cross hybridise with H3 (data not shown) and therefore probing of the YAC digest filters could proceed without generating erroneous data due to cross hybridisation. This result is confirmed by comparison of the H3 and H2A sequences. Bestfit analyses of both DNA sequences contained in the EMBL database demonstrated an 80% similarity between the H2A 3` sequence from position 409 to 473 and H3 3` sequence from 553 to 604. No sequence homology was observed in the 5` region of both genes, therefore 5` primers would not be complementary. In addition the H2A and H3 reverse primers were designed outside this region of homology, from positions 383 - 404 (H2A) and positions 529 - 553 (H3). Consequently probes amplified by either primer pair would be expected to be non- complementary.

Filters containing digests of YAC's Y9/H2 and Y10/H2 were probed as described in chapter 2. The results of the experiments showed that in YAC Y9/H2, H3 maps to 50 kb *Xho*I and *Nar*I fragments, similar to H2A, as well as 80 and 130 kb *Pvu*II fragments. This positions the gene on the extreme 5' 50 *Xho*I/ *Nar*I fragment of the YAC. However, certain anomalies were observed - the gene sequence also identifies a 45 kb *Nar*I fragment in addition to 120 and 30 kb *Eag*I fragments. From prior studies with left, right and H2A probes, *Eag*I is known not to cut at all within the insert. This result cannot be explained. Thus both H2A and H3 map to the same 50 kb *Xho*I/ *Nar*I fragment in YAC Y9/H2. As shown in figure 6.1.2B, H3 maps to the 120, 110, 90 and 80 kb *Nhe*I fragments in YAC Y10/H2, as well as 75, 60 and 40 kb fragments, which correspond to internal *Nhe*I fragments. The 35 kb fragment is not identified. Thus both H2A and H3 map to the same 45 kb *Nhe*I fragment. In addition, H3 identifies a 100 kb *Nar*I fragment, similar to H2A. However, anomalies again exist: H3 identifies two *Eag*I fragments, although by previous left and right arm probe experiments, *Eag*I does not cut within the insert. In addition as shown in figure 6.1.1.B, the probe identifies multiple 20, 30 and 40 kb *Nar*I fragments, even though this enzyme has been shown to cut only once within the insert. Despite these anomalies, Histones 3 and 2A can be mapped to the same 30 kb *Nhe*I/ *Nar*I restriction fragment. Map locations are shown in figure 6.1.2.

6.1.5 Discussion.

Previous studies have shown that of the histone genes identified in *P. falciparum*, 2A and 3, are present in single copy and map to chromosome 6 and 2B maps to chromosome 11. YAC mapping studies have positioned both genes on the same minimal 30 kb *Nhe*I/ *Nar*I fragment and 50 kb *Nar*I/ *Xho*I in YAC's identified on the basis of their H2A STS content. In all higher eukaryotes investigated, histone genes are arranged into clusters, although the genes can be on different strands and thus are transcribed in different orientations. Each cluster has a variable number of repeats depending upon the organism. Examples of cluster arrangements include a 450 fold repeat of H1(>) H4(>) H2B(>) H3(>) H2A(>) in *Psammechinus miliaris*, a 20 - 50 fold repeat of H3(>) H4(>) H2A(<) H2B(>) in *Xenopus*

laevis, a 60 fold repeat of H4(>) H2A(<) H2B(>) H1(>) and H3(>) in *Xenopus borealis* and a 10- fold repeat of H3(<) H2A(>) H4(<) H1 H2A(<) H2B(>) in chickens. Lower eukaryotes such as *Saccharomyces cerevisiae* indicate that the four histones are arranged into linked sets of H2A/H2B and H3/H4, although both pairs are not contiguous within 5 kb (Hereford *et al.*, 1979). Such evidence of gene conservation would seem to cast doubt on recent reports that H2B is present as a single copy gene on a different chromosome to H2A (Bennett *et al.*, 1995). Data presented in this thesis would seem to indicate clustering of these genes in *P. falciparum*, however further work would involve isolation of the other histone genes and hybridisation of the labelled gene products to the YAC digest filter and further fine structure mapping to determine the degree of clustering.

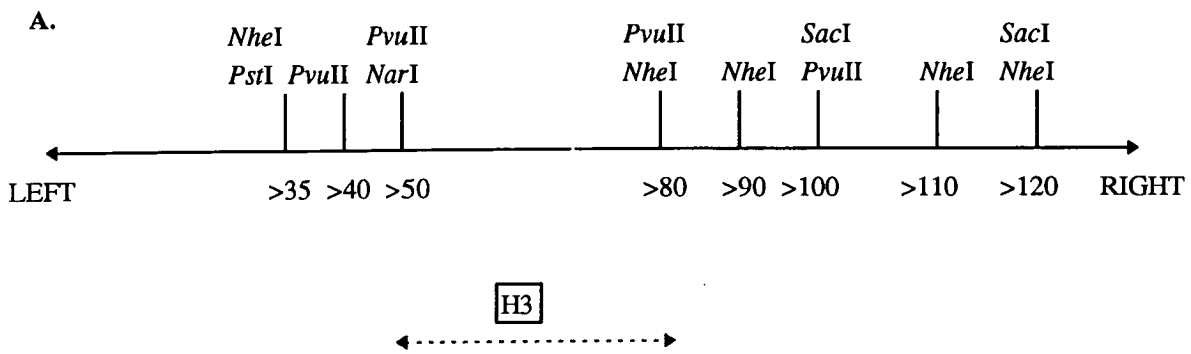
Transcriptional studies have indicated that H2A mRNA is detectable in total RNA preparations of gametocyte, zygote and erythrocyte stages in *P. falciparum* (Creedon *et al.*, 1992). In yeast and other eukaryotes, histone mRNA synthesis is tightly coupled with DNA synthesis and replication, and is consequently largely restricted to S phase (Marashi *et al.*, 1982). The link between histone production and DNA synthesis suggests that the mechanisms of histone gene regulation are linked with that of DNA replication, which implies a cell cycle control mechanism. DNA synthesis in *P. falciparum* occurs during erythrocytic and exo- erythrocytic schizogony, microgamete formation, post fertilisation and during sporozoite formation. During sporogony and schiozogony, large amounts of DNA synthesis occurs, resulting in the formation of several thousand haploid sporozoites per oocyst and around 3 x 10^4 merozoites within hepatocytes. Such a large amount of DNA synthesis would require a corresponding production of large quantities of histone proteins, if the first- level chromatin structures demonstrated for asexual blood stages also exist in sexual stages (Cary *et al.*, 1994). It is known that during embryogenesis in echinoderms, eggs enter a period of extremely rapid cell division in the course of which hundreds of daughter cells are produced by repeated division of egg cytoplasm. Large quantities of histones are used in chromatin formation and in order to provide this, eukaryotes demonstrate histone synthesis that is uncoupled from the cell cycle (Maxson *et al.*, 1983). A combination of mobilisation of stored histone mRNA transcripts and increased transcription from embryo histone repeated clusters accounts for such an increase in histone production (Goustin *et al.*, 1981). The *P. falciparum* histone genes are apparently single copy, therefore to provide large amounts of mRNA a strong promoter/ enhancer system must be in operation. If such enhancer binding sites could be interrupted, or the cell cycle altered in such a way as to block histone synthesis, then this may be a novel target for potential chemotherapeutic agents although the possible high evolutionary conservation between host and parasite targets may prove problematic.

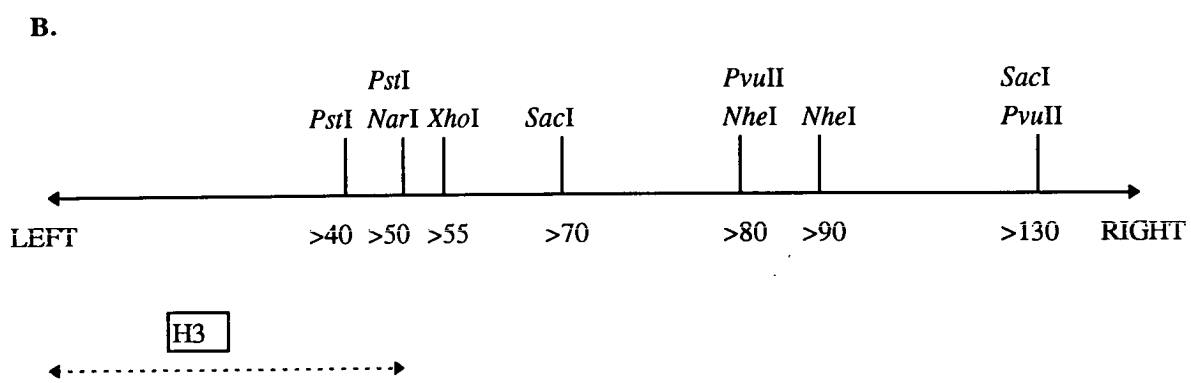
Figure 6.1.2. Location of the H3 gene on YAC's Y9/H2 and Y10/H2

Mapping of the genes is based on the physical map data described in the text.

A. Mapping the H3 gene to YAC Y10/H2

B. Mapping the H3 gene to YAC Y9/H2


A.
NheI
PstI
PvuII
PvuII
NarI
PvuII
NheI
NheI
SacI
PvuII
NheI
SacI
NheI
LEFT
>35
>40
>50
>80
>90
>100
>110
>120
RIGHT
H3


B.
PstI
PstI
NarI
XhoI
SacI
PvuII
NheI
NheI
SacI
PvuII
LEFT
>40
>50
>55
>70
>80
>90
>130
RIGHT
H3

6.2 Transcriptional Analysis of Chromosome 6 YAC's.

6.2.1 Abstract.

An investigation into stage specific gene transcription within the cloned HB3 YAC's was performed using radiolabelled cDNA probes derived from total asexual, trophozoite, schizont and gametocyte stage RNA. Probes derived from trophozoite cDNA identified two *Hind*III fragments, approximately 10 and 12 kb in size in three of the cloned YAC's. Two of the YAC's are overlapping, and consequently the gene sequence is probably contained in the common area of overlap.

6.2.2. Introduction.

The life cycle of the malaria parasite involves a number of specific stages in both sexual and asexual cycles of development. The sexual cycle involves maturation of male and female gametes, zygote fusion with subsequent maturation and differentiation to yield sporozoites. The asexual cycle involves exo- erythrocytic and erythrocytic schizogony, involving maturation of sporozoites through ring, troph and schizont stages culminating in the release of mature merozoites and the differentiation of merozoites into male and female gametocytes. In both stages, the parasite appears as a series of microscopically and morphologically distinct forms, with each form requiring specific genes expresed under some form of stage specific regulation. By examining filters of RNA prepared from sexual and asexual stages, specific gene transcription patterns can be observed. Examples of such sexually expressed genes include pfg 27/25, which is expressed only in sexual stages (Alano *et al.*, 1991), genes such as PCNA (Kilbey *et al.*, 1993) and Ran/TC4 (Sultan *et al.*, 1994) which are expressed during the asexual stage, and the serine-rich antigen SERP- H (Knapp *et al.*, 1991), which is only expressed in blood stages. Some evidence for co- ordinate regulation of gene transcription comes from a study of the serine repeat antigen, SERA (Fox and Bzik, 1994), which demonstrates co- ordinate stage- specific regulation of expression with MSA 2, a gene known to be closely linked to SERA (Lanzer *et al.*, 1993).

Several possible mechanisms for control of differential gene expression could exist in *P. falciparum*. Translational regulation may occur as ribosomes are known to be synthesised at different points in the life cycle (Gunderson *et al.*, 1987). Post- transcriptional regulation occurs in several genes, for example ribosomal RNA expression. One rRNA unit, designated A, is transcribed predominantly in the asexual stages whereas the other form, designated C, is mainly restricted to sporozoites. Switching of expression of one form to another involves the control of RNA processing, with A- form RNA being broken down in gametocytes resulting in an accumulation of C- form RNA (Waters *et al.*, 1989). GBP130, KAHRP and CS are apparently transcriptionally regulated, as gene transcription is sensitive to α- amanitin, indicating the involvement of a drug sensitive RNA Polymerase, possibly Polymerase II (Lanzer *et al.*, 1992). Both actin I and II, and α tubulin demonstrate similar expression patterns. Actin I encodes a 2.5 kb transcript and is expressed both sexually and asexually whereas Actin II encodes a 1.9

kb transcript and expression is restricted to the sexual stages (Wesseling *et al.*, 1989). Similarly, α tubulin encodes two forms: αI is transcribed at high levels in both stages, whereas α II is transcribed predominantly in the sexual blood stages. β- tubulin encodes three forms, one of which is expressed as a 2.7 kb transcript throughout the blood stages, whereas two larger transcripts are observed in the sexual stage. Variation in transcript size has been shown not to be due to differential intron processing and may be a result of variable transcriptional initiation or differential 3` modification (Delves *et al.*, 1990).

The identification of transcriptional signals in Plasmodium is limited to structural analysis, mainly because the lack of an *in vivo* transfection system or *in vitro* transcription system preclude a classical mutational approach in the definition of flanking 5` or 3` sequences responsible for the developmental regulation of gene expression. The recent report of a transfection system in *P. falciparum* (Wu *et al.*, 1995) may allow questions involving the mechanism of gene regulation to be addressed.

6.2.3. Strategies for Transcription Analysis.

To investigate stage specific gene expression within the cloned HB3 YAC's, two strategies were adopted. The first strategy was essentially that described by Lanzer *et al.*, (1992, 1993). Most eukaryotic mRNA's posses a relatively long polyA stretch (up to 200 residues) at the 3` end. The addition of this poly (A) tail is catalysed by the enzyme poly(A) polymerase and the reaction occurs in the nucleus after transcription. Not all mRNA's, however, posses a poly(A)$^+$ tail. This poly(A)$^-$ fraction is a mixture of mRNA's that have lost the poly(A) tract during extraction, and a percentage of mRNA molecules that do not posses the poly(A). A predominant component of this fraction is the histone mRNA's, but there is considerable overlap between the poly(A)$^+$ and poly(A)$^-$ species. It is thus possible for a gene to be represented in both forms.

The poly(A) region of mRNA can bind by base pairing to immobilised oligo(dT) on a solid support and this reaction can be used to isolate poly(A)$^+$ mRNA. The most advanced mRNA purification methods involve Biotinylated oligo(dT) primers which are used to bind to poly(A)$^+$ mRNA. The hybrids are captured and washed at high stringency using streptavidin coupled to paramagnetic particles together with a magnetic separation stand. mRNA is eluted from the streptavidin coated particles, reverse transcribed to cDNA and amplified using universal degenerate primers in the presence of α ^{32}P-dCTP. The labelled cDNA is separated on a spun column and is used to probe filters containing YAC DNA digested with *Dra*I or *Hind*III.

Another method used was the preparation of total cDNA from libraries derived from trophozoite enriched parasites (a gift from Dr Alister Craig; Sultan *et al.*, 1994) and gametocytes (Alano *et al.*, 1989), and amplification of inserts with vector specific primers in the presence of α- ^{32}P dCTP. The labelled cDNA's were separated from unincorporated nucleotides and used to probe YAC digest filters.

6.2.4. Results.

6.2.4.1 Expression within Total and Schizont RNA Extractions.

Total RNA was prepared from both fractions (a gift from Dr D. Arnot), reverse transcribed and amplified as described. An indication of incorporation was obtained from Radiation scintillation counters. After removal of unincorporated nucleotides, the count was extremely low. Filters containing total YAC DNA digested with *Dra*I and *Hind*III were probed with total and schizont cDNA probes. After gentle washing, no probe signal was seen on any filter. This cannot be ascribed to transcriptionally silent domains within the YAC, as no signal was seen even in the 3D7 control lanes. The negative result can be ascribed to the low levels of starting RNA resulting in a poor hybridisation probe. The PolyATtract™ protocol requires starting RNA levels of between 0.1 - 1 mg RNA. Starting amounts were 130 µg total asexual RNA and 20 µg schizont RNA. Even with an increased number of amplification cycles, there probably was not enough starting material for isolation of poly(A)$^+$ mRNA, as the quantities used were approaching minimum threshold levels as recommended by the manufacturers. Future experiments would be to repeat the procedure with 1 mg of starting RNA and hybridising filters with the new probes.

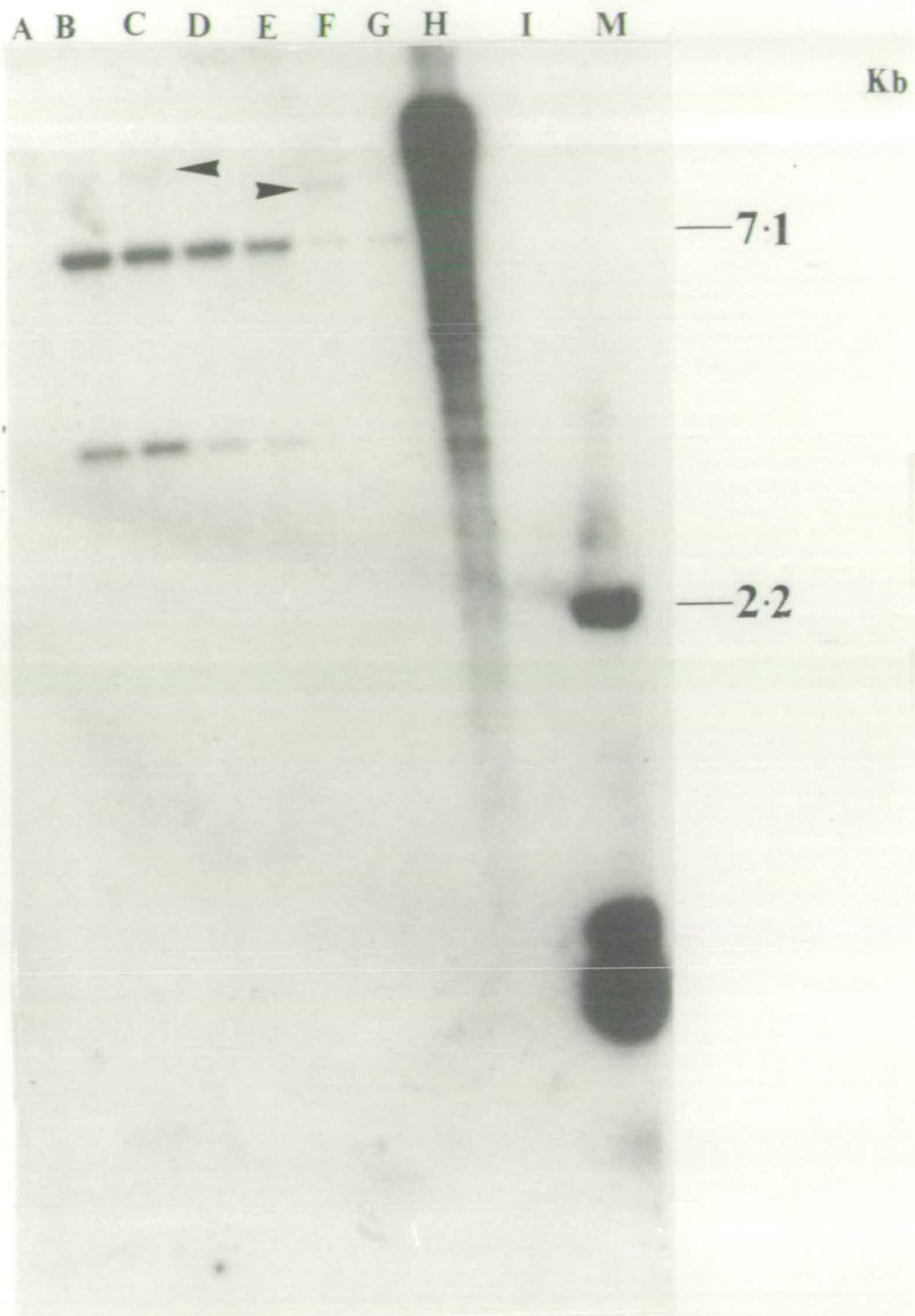
6.2.4.2 Gene Expression within the Trophozoite Stage.

A cDNA library constructed from predominantly trophozoite RNA in the plasmid vector pJFE14 was directly amplified with primers designed to both vector arms. 10ng of the cDNA library template was used directly in a PCR reaction. Amplification conditions were 94°C denaturing for 5 min; 40°C annealing for 1 min; 72°c extension for 2 min for two cycles; followed by a repeat with a 1 min denature for 28 cycles. The labelled probe was separated from unincorporated nucleotides and used to probe YAC digest filters .The results are shown in figure 6.2. Two sets of yeast genes are visible at approximately 7 kb and 5 kb, as the same signal is amplified in all YAC's and in the AB1380 yeast control. Two sets of signals are observed. A signal of approximately 12 kb is observed in YAC's Y11/63 and Y12/63, which are YAC's known to overlap by the 5` and 3` 30kb containing the pSC11.63 STS. No signal is seen in lanes containing the H2A YAC's Y9/H2 and Y10/H2. A second smaller signal of approximately 10 kb is seen in the lane containing the YAC Y5/87. No signal was seen with the same probe when used when filters containing the remaining YAC's were hybridised. These would seem to be genuine signals as (1) they are not present in any of the negative controls used (ii) they are not uniformly present (iii) the same signal is seem in two YAC's known to be overlapping and (iv) the signals are of different sizes.

Figure 6.2 Transcriptional Expression within HB3 Artificial Chromosomes.

YAC DNA was digested to completion overnight with 10U of HindIII. 10µl of DNA was loaded onto a 1% IBI agarose gel in 1X TAE and the fragments resolved over 25 hours @ 15V. The DNA was transferred onto Hybond N^+ and the fragments probed with amplified, labelled trophozoite DNA as described. 10mg of total yeast RNA was included in the hybridisation buffer. The lanes are as follows:

Lane A: Total Human DNA
Lane B: YAC Y12/63
Lane C: YAC Y11/63
Lane D: YAC Y10/H2
Lane E: YAC Y9/H2
Lane F: YAC Y5/87
Lane G: AB1380 Genomic DNA
Lane H: Total 3D7 DNA
Lane I: Total Simian DNA
Lane M: Molecular Weight Marker VI (Boehringer)


A B C D E F G H I M
Kb
7·1
2·2

6.2.4.3. Gametocyte Gene Expression.

The cDNA library constructed from gametocytes in λNM 1149 was first titred as described. The results are shown in table 6.1. The phage titre is determined as follows:

$$\text{Phage concentration (in pfu/ µl)} = \frac{\text{No. of plaques x 10 (dilution Factor)}}{\text{10 (for 1 µl)}}$$

The mean titre for the gametocyte cDNA library was determined to be 2.1×10^3 pfu / µl.

Plate	Dilution	Number of Plaques
1	Neat	Indeterminate
2	10-1	~ 2200
3	10-2	370
4	10-2	15
5	10-4	1

Table 6.1. Phage titre for the gametocyte cDNA library in λ NM 1149.

DNA from the Library was prepared as described. Commercial primers for the λ vector were used to amplify inserts in the presence of radiolabelled nucleotides. Alter nucleotide removal, a low signal was observed with a scintillation counter. Hybridisation of filters revealed no signal even in control lanes. The problem can be ascribed to inefficient probing due to poor template preparation. Future experiments would be to repeat the procedure with improved template for labelling.

6.2.5. Discussion.

To investigate gene transcription within the cloned YAC's, a variety of techniques were performed. Due to technical limitations involving purity and quantity of starting material, no results were obtained for investigations into stage specific expression within total sexual schizont and gametocyte stages. Positive signals were observed with trophozoite cDNA probes. The signals seen were of the same size in both Y11/63 and Y12/63 YAC's, indicating that the complete gene sequence is contained within both common inserts, which is known from mapping studies to be a terminal 30 kb *Nhe*I/ *Sac*I fragment. The cDNA probes identified corresponding genomic *Hind*III digest fragments of 12 and 10 kb. However, no estimation of the size of the open reading frame can be inferred - in cloning PCNA, the 825 bp ORF was contained on a 4.4 kb *Hind*III fragment (Kilbey *et al.*, 1993), so fragment size does not necessarily correlate with ORF size. However, the one of the largest gene sequences

identified is the RNA Polymerase III gene, which is present as a single 8.5 kb mRNA fragment containing 4 introns (Li *et al.*, 1991). Therefore it is highly probable that the fragment contains the 5` regulatory elements and possibly distant enhancer sequences as well. Thus the advantages of YAC's in investigation of gene expression are becoming apparent. YAC's were used to transciptionally map *P. falciparum* chromosome 2 (Lanzer *et al.*, 1993). Transcriptional analysis of overlapping YAC's demonstrated an actively expressed central domain flanked by transcriptionally silent, polymorphic ends. If such a model is true for all chromosomes, then these YAC's could map to the central core of the chromosome. Future work would involve (a) repeating YAC analysis with probes derived form other stages of the parasite life cycle and (b) completing chromosomal coverage in YAC's and analysing a minimal set of overlapping YAC's with cDNA probes to gauge transcriptionally active domains within the chromosome.

CHAPTER 7.

DISCUSSION

7.1. Objectives of the *P. falciparum* Genome Project.

The work presented in this thesis has been to enhance and complement work being carried out by the Malaria Genome Mapping Consortium. This involved the construction of a yeast artificial chromosome library of the *P. falciparum* clone HB3 and an attempt to provide a complete, detailed physical map of chromosome 6. Due to technical complexities and a lack of starting chromosome specific markers for contig coverage, a complete map was not produced. However, foundations have been laid which should enable map completion by 1996.

The ultimate aim of the Project is to provide complete, detailed YAC- based physical maps of all fourteen *P. falciparum* chromosomes with a resolution of 15 - 30 kb. Once such maps are completed and YAC clones characterised, it should be possible to identify a minimal set of clones that completely cover a chromosome from telomere to telomere. DNA pools representing entire chromosomes could be constructed from these YAC's. Fourteen such pools would be representative of the entire *P. falciparum* genome. Any new gene sequence identified could be accurately mapped to a chromosome by a simple PCR based screening procedure, eliminating dubious chromosomal assignments due to poor PFG resolution. Once a newly identified gene is positioned within the genome, a second round of PCR screening on liquid YAC DNA representing a minimal linear set of YAC's covering a chromosome could be used to identify the YAC which contains the gene. Once this is obtained, the gene could be labelled and used to probe a set of reference filters containing that YAC digested with Map Reference enzymes. The data obtained could be used to position the gene on the physical map. Filters could then be stripped, ready for subsequent reprobing. As the map is extended, linkage groups could become apparent, as well as clustering of genes. The data obtained would provide an important resource for the benefite of future work on parasite molecular genetics.

YAC filters and chromosomal YAC pools could be screened with PCR primers supplied by individual laboratories. Once positive YAC's have been identified and the above work performed, data could be distributed to (a) the original research group and (b) a Malaria Genome Database. Information would be disseminated in the form of (i) accurate chromosomal assignments, (ii) map locations in kb, (iii) proximity to known STS markers, (iv) as a complete physical map of the region (v) PCR primer sequence information and amplification conditions and (vi) as YAC DNA preps in liquid form and as intact chromosomal agarose blocks. Such information would be supplied to the original researchers and, with their permission, to any other group of interest.

The work carried out here describes the isolation and mapping of chromosome 6 specific YAC's by a primary screen with chromosome 6 specific markers, construction of end- clone STS markers and their subsequent use in level one and level two library screens. However, before the application of the information generated can be discussed, limitations must first be mentioned.

7.2. Limitations of the *P. falciparum* Genome Project.

In the course of this study, YAC contigs were extended by constructing STS markers from the ends of each YAC and using such markers in a second round library screen, in essence, an End- Clone Specific approach (see chapter 1). However, such an approach is laborious, time consuming, not completely successful and results in contig extension, one YAC at a time. A more efficient strategy for the *P. falciparum* genome project would be the application of an STS content approach. As discussed, such an approach was adopted for the map of chromosome 2, in which all available markers were used in a level one screen and any resulting gaps, identified by an absence of PCR signals, were covered by an end rescue approach. However such a strategy cannot work for the less well characterised chromosomes, for example 6 and 12. An STS content approach involves screening the library with multiple specific STS markers, and identifying overlapping YAC's on the basis of their common information content. For such an approach to be informative (i) the library needs to be large enough to produce multiple YAC's per marker and of a high quality so that there are few anomalies per YAC, and (ii) there must be multiple specific markers for which PCR data exists.

There are several methods by which multiple markers can be constructed. Firstly, intact chromosomes can be resolved on a Pulsed field Gel, excised and subcloned either into plasmids or M13 for sequencing. However, a problem is the contamination of the desired chromosomal band with DNA from other chromosomes or with digestion products due to poor sample preparation. Alternatively, the YAC itself could be subcloned, which would increase marker saturation of a defined area and facilitate isolation of multiple YAC's for that area. However, the insert DNA must be known to be completely contiguous with the genomic copy for this strategy to be efficient. In addition, chromosome specific YAC libraries could be constructed, which would lend itself to an end clone specific map strategy, as the library is enriched for the chromosome of interest and thus fewer clones would have to be screened.

An alternative and biologically more informative approach would be to utilise the Expressed Sequence Tag resource being developed at the Department for Infectious Diseases, Florida for library screening (Chakrabati *et al.*, 1994). An EST can serve the same purpose as a random genomic STS and is simply a segment of a cDNA clone corresponding to an mRNA. This has the advantage that the marker identifies an expressed gene (Adams *et al.*, 1991). EST markers have been constructed for human (Adams *et al.*, 1991) and *C. elegans* genomes (McCombie *et al,* 1992), and have been used to map new genes to YAC's. To date, 673 EST markers have been created for *P. falciparum* as a Mung Bean Nuclease library, corresponding to 215 kb of sequence data and identifying 51 novel genes (Reddy *et al.*, 1993). In addition 550 random cDNA clones have been sequenced generating 389 EST markers (Chakrabati *et al.*, 1994). Analysis of the Mung Bean Nuclease library has identified genes from the erythrocytic, exo- erythrocytic and sexual stages of parasite development and has led to the cloning and characterisation of a *P. falciparum* aspartic proteinase (Dame *et al.*,1994). The library constructed by

this approach will facilitate access to genes from all stages of development, rather than tagging gene expression from specific stages represented by cDNA libraries. This "whole genome" strategy would be an essential requirement for constructing a database for gene expression with which to investigate the biology of the parasite. If these markers can be assigned to a chromosome, then this would in essence be a resource of over 400 expressed gene markers which would facilitate an STS content approach to mapping of the *P. falciparum* genome, in addition to a concomitant investigation of gene clustering and genome organisation. The EST resource would also drastically reduce the number of end- clone rescue steps needed to close contigs. Such a valuable resource should not be ignored by the Genome Consortium.

In addition, the present library constructed for genome analysis is limited in that telomeric and subtelomeric sequences are underrepresented. In order to ensure complete chromosomal coverage and gene representation, a telomere specific library would need to be constructed. Such libraries can be constructed in YAC vectors with a single arm containing CEN, ARS, TEL, URA and TRP sequences, with the second functional telomere provided by the cloned DNA sequence.

Another problem for the genome project is the lack of information management resources. Physical data will be generated in the form of restriction maps, RFLP's, STS content maps and EST expression maps. As EST's and STS's become submitted to the Malaria Genome Database, there must be some way to allow researchers to visually co- ordinate all data sets and so compare all map data simultaneously. The Human Genome Organisation has developed a graphical user interface program, Chromoscope, which is a networked client- server application linking GenBank, Medline and other databases which allows visualisation and retrieval of map data derived from different techniques. Chromosome maps can be aligned and common markers oriented (Boguski and Schuler, 1995). Such a resource would be useful for ordering *Plasmodium* gene clusters onto specific restriction fragments. In addition, information access necessitates the installation of a World Wide Web homepage, where each chromosome could be assigned a different site, with maps being constantly refined and updated. As more and more gene based markers accrue, homology information and links to scientific literature will greatly amplify the value of these markers.

In the course of this study, two artificial chromosomes containing the histone 2A locus were identified. Detailed physical mapping using probes specific to the left and right vector arms have shown that while the internal map data is reasonably consistent and positions have been assigned for the STS, the YAC's cannot be arranged into overlaps. This is due to the presence in one YAC of an *Xho*I site approximately 10 kb from the suspected location of the gene. This evidence suggests that one or both YAC inserts contain noncontiguous DNA resulting from either a co-cloning or deletion event. Correlation with the genomic copy is problematic because the size of restriction fragments produced in a genomic digest will differ depending upon the proximity of sites in the genome outside the region cloned in the YAC. The simplest way to confirm which YAC is chimeric is to isolate further YAC's and map

each insert. In addition, I observed YAC instability, visualised as bands of decreasing size after a YAC Pulsed field gel was Southern blotted and probed with YAC specific sequences (data not shown). In each case, the deletion event resulted in a second band slightly smaller than the parent YAC. As discussed, such anomalies can be the result of either (i) coligation of DNA during the digestion/ ligation step of the transformation protocol or (ii) recombination between repetitive elements mediated by the yeast host cell recombination apparatus (Green *et al.*, 1991). Studies on the *P. falciparum* genome have indicated few dispersed repetitive elements (Weber *et al.*, 1988a; Weber *et al.*, 1988b) and preliminary assessments of YAC libraries and published maps show little evidence of chimerism. However a recent study has pointed to the transformation procedure itself as a possible source for chimera formation, by inducing recombination between two co- penetrating YAC's sharing a region of homology (Larionov *et al.*, 1994). In addition to the well documented chimera frequencies of the human YAC libraries, chimera frequencies of approximately 5% for *C. elegans* (Coulson *et al.*, 1991), 20% for *Arabidopsis* (Schmidt and Dean, 1993) and 5% for Schistosomes (Tanaka *et al.*, 1995) YAC libraries indicate that even when the degree of dispersed repetitive elements is comparatively low (4 - 6% of the total genome in Schistosomes- Simpson *et al.*, 1982), YAC anomalies are still observed. This means that each YAC isolated will have to be extensively mapped for deletions or chimeras by either (i) extensive physical mapping of the insert or (ii) employing a detailed STS- content based approach to mapping. Published physical maps for chromosomes 2 (Lanzer *et al.*, 1993) and 4 (Rubio *et al.*, 1994) should be re-examined in view of the anomalies observed in the HB3 library, which was constructed by a more advanced method than that used for the the FCR3 library used in map construction (De Bruin *et al.*, 1992).

If such anomalies are common, it may be necessary to investigate second generation library construction strategies. One option would be to construct YAC libraries with lower chimera frequencies. As discussed, such libraries could be constructed from (i) *rad1/ rad52* host strains or (ii) from defined DNA sources such as isolated chromosomes. In view of the laborious nature of YAC library construction, screening, clone manipulation, and time already spent on the project, such an approach would be impractical. An alternative would be to utilise the recently developed Bacterial Artificial Chromosome (BAC) vector. Libraries constructed in such vectors could enhance existing YAC libraries if unclonable gaps become apparent. BAC's are a modification of plasmids containing the bacterial F factor (O'Connor *et al.*, 1989). F plasmids are maintained in low copy number in the bacterial cell, which reduces opportunities for recombination and in addition can maintain fragments up to 1 Mb. The F factor operon contains genes which regulate plasmid replication (*oriS* and *repE*) and control copy number (*parA* and *parB*). The BAC vector contains these genes as well as antibiotic resistance markers for selection. The advantages of the BAC system are: (i) very high transformation efficiencies using conventional methods (e.g. electroporation); (ii) stable maintenance of eukaryotic DNA of up to 300 kb; (iii) very low frequency of co-cloning compared to the YAC system and (iv) no incidences of

cotransformation, as *parA* and *parB* are involved in the exclusion of extraneous F factors (Shizuya *et al.*, 1992). Such a system would include its own control of cocloning, based onthe transformation model (Larionov *et al.*, 1994). Another main advantage is the BAC vector lends itself to standard plasmid and bacterial manipulations (Woo *et al.*, 1994). BAC' s have already been used to construct specific libraries for human chromosome 2 (Wang *et al.*, 1994) and chromosome 22 specific BAC's (Kim *et al.*, 1994).

7.3. Applications of the *P. falciparum* Genome Project.

At a first glance, the major application of the Genome Project would be in the mapping and characterisation of novel genes. New gene sequences could be quickly mapped, streamlining the production of information. In addition, the maps and marker data produced could facilitate the search for new genes. Fine structure genetic mapping of chromosomes should often reveal not only the presence of novel genes, but possibly the target sequence itself. Such positional cloning can not only result in the identification of the genetic and/ or biochemical trait under study, but also results in an increased marker saturation for that area. However, such limited applications would mean that the project is under-exploited. The extensive sequence and map information generated could allow some of the more complex questions about the genome of *P. falciparum* to be answered.

7.3.1. Genome Organisation.

Little is known about the complete organisation of genetic information within the *P. falciparum* genome. As discussed, stages at which individual genes have been expressed have been investigated in detail. However little is known about such expression patterns within a chromosomal context. From the YAC map of chromosome 2 (Lanzer et al., 1993), it was shown that the chromosome is arranged into a transcriptionally active central domain and silent subtelomeric domains. The Knob Associated Histidine Rich protein (KAHRP - Triglia *et al.*, 1987) and a membrane protein PfEMP3 were known to demonstrate co- ordinated transcription and translation (Pasloske *et al.*, 1994). Transcriptional mapping of chromosome 2 YAC's showed a telomere proximal gene cluster of KAHRP flanked by PfEMP3 and GLARP. Analysis of YAC's with cDNA probes derived from poly(A)$^+$ mRNA demonstrated that the genes were co- ordinately expressed in the blood stage and were transcribed monocistronically (Lanzer *et al.*, 1994). This implies that the intergenic regions of these genes must contain the minimal necessary elements for transcription initiation and termination. However, there may be common regulators (e.g. silencers or enhancers) which could control developmental regulation of the entire gene cluster. Such investigations could now be performed with the recently developed transfection system. Investigations into co- ordinate regulation of gene transcription and clustering of expressed genes could be performed for the entire genome in YAC contigs when they are completed for all chromosomes.

7.3.2. Gene Families.

Little is also known about the arrangement of genes into gene families. The 5.8s rRNA gene is divided into at least five classes on different chromosomes (Shippen- Lentz *et al.*, 1990b) and five members of the heat shock gene family of proteins have been discovered, all located on different chromosomes (Sharma, 1992). In addition, the tubulin gene family has had several members cloned, with individual genes showing both constitutive and developmental regulation of gene expression (Delves *et al.*, 1990). No investigation has been performed as to the mechanisms of co- ordinate regulation of expression. Recently a multigene family comprising over 140 members of an expressed blood stage antigen has been identified (Carcy *et al.*, 1994). The gene is a homologue of the *Babesia* multigene family encoding the 60 KDa apical RAP1 antigen, deduced by the presence of a consensus amino acid motif. Such a family is similar to the large gene repertoire used to generate antigenic diversity in African trypanosomes (Borst, 1991). Completed YAC contigs of all different chromosomes would allow such gene family members to be mapped, subcloned and sequenced and the transfection system could be used to investigate how such genes are co- ordinately regulated by identifying potential regulatory elements. New gene families could also be quickly identified and mapped. Such widely spread genes may however act as sites for chimera formation, depending upon the degree of homology of individual family members.

7.3.3. Gene Transcription

Little information is available for sequences which regulate gene transcription. Given the extreme A+T base composition of the *P. falciparum* genome in the intergenic regions of the genome (90%), sequences resembling classical TATA transcription initiation motifs (consensus TATA $^{A}/_{T}$ A) may be very abundant upstream of the coding region of both constitutive and developmentally expressed genes (Delves *et al.*, 1990) It is difficult to see how the *Plasmodium* transcriptional apparatus could differentiate between classical TATA sequences representing initiation domains and those arising due to base composition. It may be that transcription initiation differs in *P. falciparum* compared to other eukaryotes. A gene encoding a TATA binding protein required for RNA Polymerase II initiation has been cloned (McAndrew *et al.*, 1993), which reveals a highly divergent primary amino acid sequence compared with TBP proteins from other organisms. However, key domains important for protein function are retained. Such differences may be due to the extreme genome composition of the parasite. The cloning of genes from YAC's as intact copies containing flanking regulatory elements, coupled with the application of the transfection system, would facilitate investigations into gene regulation. If a transcriptional map is obtained, then the upstream sequences of genes expressed at different developmental stages could be compared, and stage specific initiation sequences identified. In addition, subcloning and transfection of gene clusters on YAC's could facilitate investigation of elements involved in global gene regulation.

All of the above questions could be addressed by the construction of detailed EST- based physical and transcriptional maps of each chromosomal YAC contig. Once all EST's representing the total genetic complement of the organism are positioned on a final YAC map, sequential hybridisation of northern filters will determine when a particular gene is expressed during the life cycle. Such information may not only reveal the presence of gene clusters, but also the arrangement of such clusters into co- ordinately regulated sets of mono- or polycistronic expression units. A possible method for performing such large scale Northern analyses would be to use chemiluminescent probes. Individual EST's could be tagged with different probes which fluoresce under a particular UV wavelength. RNA filters could then be probed with multiple probes, and each could be tested for hybridisation by altering the UV wavelength. As will be discussed, the information may also help to elucidate the possible evolution of the parasite.

7.3.4. Recombination.

Data generated can also be used to investigate chromosome rearrangement and recombination. HB3 was shown to lack a subtelomeric region of chromosome 13 normally containing the histidine rich protein PfHRPIII gene. This domain in HB3 was found to be deleted by a duplicative interchromosomal transposition event involving 100 - 200 kb of subtelomeric DNA from chromosome 11, resulting in a duplication of RESA- 2 and Pf332, The inheritance of this duplication was followed by analysis of the progeny of a cross performed between HB3 and Dd2 (Wellems *et al.*, 1990). Three of the fourteen progeny analysed showed a transposition event from chromosome 11 to 13, indicating a relatively high rate of recombination between subtelomeric regions (Hinterberg *et al.*, 1994), Because YAC libraries now exist for HB3 and 3D7, such an event can now be followed in the progeny of this cross (Walliker *et al.*, 1987) to see how the recombination frequency compares. In addition if polymorphic markers are obtained, then the breakpoint can be cloned and sequenced by comparing detailed STS/ EST- based YAC maps of chromosome 13 in HB3 and 3D7, As YAC libraries exist for all parental strains used to carry out both genetic crosses which have been performed to date, the number of putative polymorphic markers that can be constructed can be used to provide detailed information concerning the conservation of linkage groups within the recombinant progeny as well as frequencies of recombination within a chromosome. Such information may indicate which gene linkages are highly conserved and may thus be important for the parasite, and may lead to the cloning and characterisation of recombinational "hot-spots". Detailed recombination maps could also be used to determine wether recombinant progeny have arisen due to independent assortment or by strand exchange at meiosis.

7.3.5. Evolution of *P. falciparum*.

Studies on the molecular evolution of the parasite have determined two basic rationales. Firstly, analysis of small subunit rRNA (SSU rRNA) gene sequences indicate that *P. falciparum* is a relatively

recent addition to the human population, although opinion is divided as to whether the parasite is derived from avian or primate stock (Waters *et al.*, 1991; Escalante *et al.*, 1994). Secondly evidence is accumulating for the inheritance of the extrachromosomal elements of *P. falciparum* from a plastid ancestor. Investigations of the 35 kb extrachromosomal element revealed the presence of an *rpoB* gene related to plastid not eubacterial genes (Gardner *et al.*, 1994b), and a 470 amino acid ORF was identified which demonstrated 50% identity with the genome of the red algae *Antithamnion* (Williamson *et al.*, 1994). In addition, analysis of the putative 6 kb mitochondrial genome of *P. falciparum* has shown that the rRNA genes are fragmented and interspersed with other protein genes, characteristic of the green algae *Chlamydomonas reinhardtii* (Feagin, 1992). Analysis of the nuclear genome demonstrated that the Hsp90 gene was more closely related to plants than to vertebrates or other parasite Hsp90's (Bonnefoy *et al.*, 1994). Studies on the *P. falciparum* glycolytic pathway revealed an enolase gene containing a pentapeptide insertion sequence EWGWS located in a domain conserved with plants, which have a similar motif EWGWC. Such homology may only be observable because of the extreme antiquity of the glycolytic pathway (Read *et al.*, 1994).Such nuclear homologies may have resulted from transposition events in the evoutionary history of the parasite. If a complete detailed EST map of the *P. falciparum* genome is constructed, then new genes may be identified whose conservation and consensus motifs refine and expound the theories of Plasmodium evolution. In addition, if complete transcriptional maps become available for each chromosome, the arrangement of stage specific gene transcripts per chromosome can give an indication as to wether in the past a chromosome was expressing purely sexual or asexual genes. In order for the malaria parasite to have evolved a biphasic life cycle, one possible explanation is that the organism results from the fusion of two ancestors, one vector borne, the other host borne. If one chromosome appears to be predominantly contain sexual or asexual genes, even though recombination events during evolution may have shuffled genes between chromosome, then such genes and gene organisation may provide evidence for such a fusion event having occurred.

7.4 Conclusion.

The *P. falciparum* Genome Project is a means to not only streamline the mapping and investigation of novel genes, but to also investigate some of the fundamental questions of the malaria genome. However, in order to address these problems, the fundamental limitations of the project need to be considered. All YAC's must be screened for chimeras and instabilities, library construction protocols refined and more markers made available for screening. Once these problems are overcome, the technology will be available for a more detailed molecular analysis of this important human pathogen.

REFERENCES

Abidi, F.E., Wada, M., Little, R.D. and Schlessinger, D (1990) Yeast artificial chromosomes containing human Xq24-Xq28 DNA: library construction and representation of probe sequences.
Genomics **7** : 363 - 376

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) Complementary DNA sequencing: Expressed sequence tags and the human genome project.
Science **252** : 1651 - 1656

Adzuma, K., Ogawa, T. and Ogawa, H. (1984) Primary structure of the RAD52 gene in *Saccharomyces cerevisiae.*
Mol. Cell. Biol. **4** : 2735 - 2744

Ajioka, J.W., Smoler, D.A., Jones, R.W., Carulli, J.P., Velleck, 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 chromsosomes.
Chromosoma **100** : 495 - 509

Alano, P., Premawanasa, S., Bruce, M.C. and Carter, R. (1991). A stage specific gene expressed at the onset of gametocytogenesis in *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **49** : 81 - 88

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., LePaslier, D. and Cohen, D. (1990) Construction and characterisation of a yeast artificial chromosome library containing seven haploid human genome equivalents
Proc. Nat. Acad. Sci. **87** : 4256 - 4260

Albertsen, H.M., Smith, S.A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C.S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Mei- Sheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B.A.J. and White, R. (1994) A physical map and candidate genes in the *BRCA1* region on chromosome 17q12- 21.
Nature Genet. **7** : 472 - 479

Alonso, P.L., Smith, T., Armstrong Schellenberg, J.R.M., Masanja, H., Mwankusye, S., Urassa, H., Bastos de Azevedo, I., Chongela, J., Kobero, S., Menedez, C., Hurt, N., Thomas, M.C., Lyimo, E., Weiss, N.A., Hayes, R., Kituya, A.Y., Lopez, M.C., Kilama, W.L., Teuscher, T. and Tanner, M. (1994) Randomised trial of efficacy of SPf66 vaccine against *Plasmodium falciparum* malaria children in southern Tanzania.
Lancet **344** : 1175 - 1181

Anand, R. (1986) Pulsed field gel electrophoresis: a technique for fractionating large DNA molecules.
TIG **11** : 278 - 283

Anand, R., Riley, J.H., Butler, R., Smith, J.C. and Markham, A.F. (1990) A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage.
Nuc. Acids. Res. **18** : 1951 - 1956

Anand, R., Pgilvie, D.J., Butler, R., Riley, J.H., Finniear, R.S., Powell, S.J., Smith, J.C. and Markham, A.F. (1991) A yeast artificial chromosome contig encompassing the Cystic fibrosis locus.
Genomics **9** : 124 - 130

Anderson, C. (1993) Genome shortcut leads to problems.
Science **259** : 1684 - 1687

Arnot, D.E., Roper, C. and Bayoumi, R.A.L. (1993). Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates.
Mol. Biochem. Parasitol. **61** : 12 - 24

Arveiler, B. and Porteous, D. J. (1991). Amplification of end fragments of YAC recombinants by inverse- polymerase chain reaction.
Technique **3** : 24 - 28

Aslund, L., Franzen, L., Weshin, G., Persson, T., Wigzell, H. and Pettersson, U (1985) Highly reiterated non- coding sequence in the genome of *Plasmodium falciparum* is composed of 21- base pair tandem repeats.
J. Mol. Biol. **185**: 509 - 516

Babiker, H.A., Creasey, A.M., Bayoumi, R.A.L., Walliker, D. and Arnot, D.E.A. (1991) . Genetic diversity of *Plasmodium falciparum* in a village in eastern Sudan. 2. drug resistance, molecular karyptypes and the *mdr1* genotype of recent isolates.
Trans. Royal. Soc.Trop. Med. Hyg. **85** : 578 - 583

Bach. M.L., Lacroute, F. and Botstein, D. (1979) Evidence for transcriptional regulation of orotidine - 5' - phosphate decarboxylase in yeast by hybridisation of mRNA to the yeast sequence gene cloned in *Escherichia coli.*
Proc. Nat. Acad. Sci. **76** : 386 - 390

Baker, R.E., Hayes, M.F. and O'Brien (1989) Purification of the yeast centromere binding protein CP1 and a mutational analysis of its binding site.
J. Biol. Chem. **264** : 10843 - 10850

Banerji, J., Rusconi, S. and Schaffner, W. (1981) Expression of a β- globin gene is enhanced by remote SV40 DNA sequences.
Cell **27** : 299 - 308

Barlow, D.P. and Lehrach, H. (1987) Genetics by gel electrophoresis: the impact of pulsed field gel electrophoresis on mammalian genetics.
TIG **3** : 167 - 171

Barnes, D.A., Thompson, J., Triglia, T., Day, K. and Kemp, D.J. (1994) Mapping the genetic locus implicated in cytoadherence of *Plasmodium falciparum* to melanoma cells.
Mol. Biochem. Para. **66** : 21 - 29

Bates, G.P., Valdes, J., Hummerich, H., Baxendale, S., Le Paslier, D.L., Monaco, A.P., Tagle, D., MacDonald, M.E., Altherr, M., Ross, M., Brownstein, B.H., Bentley, D., Wasmuth, J.J., Gusella, J.F., Cohen, D., Collins, F. and Lehrach, H. (1992) Characterisation of a yeast artificial chromosome contig spanning the Huntingdons disease gene candidate region.
Nature Genet. **1** : 180 - 187

Bell, S.P. and Stillman, B. (1992) ATP- dependent recognition of eukaryotic DNA replication by a multiprotein complex.
Nature **357** : 128 - 134

Benbow, R.M., Zhao, J. and Larson, D.D. (1992) On the nature of origins of DNA replication in eukaryotes.
BioEssays **14** : 661 - 669

Bennett, B.J., Thompson, J. and Coppel, R.L. (1995) Identification of *Plasmodium falciparum* histone 2B and histone 3 genes.
Mol. Biochem. Parasitol. **70** : 231 - 233

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

Bentley, K.L., Bradshaw, S.M. and Ruddle, F.H. (1993) Physical linkage of the murine hox- b cluster and nerve growth factor receptor on yeast artificial chromosomes.
Genomics **18** : 43 - 53

Biggs, B.A., Kemp, D.J. and Brown, G.V. (1989) Subtelomeric chromosome deletions in field isolates of *Plasmodium falciparum* and their relationship to loss of cytoadherence *in vitro.*
Proc. Natl. Acad. Sci. **86** : 2428 - 2432

Birren, B.W., Lai, E., Clark, S.M., Hood, L. and Simon, M. L. (1988) Optimised conditions for pulsed field gel electrophoretic separations of DNA.
Nuc. Acids. Res. **16** : 7563 - 7582

Blackburn, E.H. (1990) Telomeres and their synthesis.
Science **249** : 489 - 490

Blackburn, E.H. (1991) Structure and function of telomeres.
Nature **350** : 569 - 573

Bloom, K. and Carbon, J. (1982) Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes.
Cell **29** : 305 - 317

Boguski, M.S. and Schuler, G.D. (1995) ESTablishing a human transcript map.
Nature Genet. **10** : 369 - 372
Borre, M.B., Dziegiel, M., Hogh, B., Petersen, E., Rieneck, K., Riley, E., Meis, J.F., Aikawa, M., Nakamura, K., Masakazu, H., Wind, A., Jacobsen, P.H., Cowland, J., Jepsen, S., Axelsen, N.H. and Vuust, J. (1991) Primary structure and localisation of a conserved immunogenic *Plasmodium falciparum* glutamate rich protein (GLURP) expressed in both preerythrocytic stages of the vertebrate life cycle.
Mol. Biochem. Parasitol. **49** : 119 - 132
Bonnefoy, S., Attal, G., Langsley, G., Tekaia, F. and Mercereau - Puijalon, O. (1994) Molecular chacterisation of the heat shock protein 90 gene of the human malaria parasite *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **67** : 157 - 170
Borst, P. (1991) Molecular genetics of antigenic variation.
Imm. Today. **10** : A29 - 33
Brewer, B.J. and Fangman, W.L. (1987) The localisation of replication origins on ARS plasmids in *S. cerevisiae.*
Cell **51** : 463 - 471
Britten, R.J., Baron, W.F., Stout, D.B. and Davidson, E.H. (1988) Sources and evolution of human *Alu* repeated sequences.
Proc. Nat. Acad. Sci. **85** : 4770 - 4774
Brody, H. and Carbon, J. (1989) Electrophoretic karyotype of *Aspergillus nidulans.*
Proc. Nat. Acad. Sci. **86** : 6260 - 6263
Bronson, S.K., Pei, J., Taillon- Miller, P., Chorney, M.J., Geraghty, D.E. and Chaplin, D.D. (1991) Isolation and charaterisation of yeast artificial chromosome clones linking HLA- B and HLA- C loci.
Proc. Nat. Acad. Sci. **88** : 1676 - 1680
Brooks- Wilson, A.R., Lichter, J.B., Ward, D.C., Kidd, K.K. and Goodfellow, P.J. (1993) Genomic and yeast artificial chromosome long range physical maps linking six loci in 10q11.2 and spanning the multiple endocrine neoplasia type 2A (*MEN2A*) region.
Genomics **17** : 611 - 617
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
Burgers, P. M. J. and Percival, K. J. (1987). Transformation of yeast spheroplasts without cell fusion.
Anal. Biochem. **163** : 391 - 397
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. and Olson, M.V. (1991) Preparation of clone libraries in yeast artificial chromosome vectors.
in Methods in Enzymology **194** : 251 - 270
Butler, R., Ogilvie, D.J., Elvin, P., Riley, J.H., Finniear, R.S., Slynn, G., Morten, J.E.N., Markham, A.F. and Anand, R. (1992) Walking, cloning and mapping with yeast artificial chromosomes: a contig encompassing D21S13 and D21S16.
Genomics **12** : 42 - 51
Carcy, B., Bonnefoy, S., Guillotte, M., Le Scanf, C., Grellier, P., Schrevel, J., Fandeur, T. and Mercereau- Puijalon, O. (1994). A large multigene family expressed during the erythrocytic schizogony of *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **68** ; 221 - 233
Carle, G.F. and Olson, M.V. (1984) Separation of chromosomal DNA molecules from yeast by orthoganal field alternation gel electrophoresis.
Nuc. Acids. Res. **14** : 5647 - 5664
Carle, G.F. and Olson, M.V. (1985) An electrophoretic karyotype for yeast.
Proc. Nat. Acad. Sci. **82** : 3756 - 3760

Carle, G.F., Frank, M. and Olson, M.V. (1986) Electrophoretic separations of large DNA molecules by periodic inversion of the electric field.
Science **232** : 65 - 68
Cary, C., Lamont, D., Dalton, J.P. and Doerig, C. (1994) *Plasmodium falciparum* chromatin: nucleosomal organisation and histone- like proteins.
Parasitol. Res. **80** : 255 - 258
Chartier, F.L., Keer, J.T., Sutcliffe, M.J., Henriques, D.A., Mileham, P. and Brown, D.M. (1992) Construction of a mouse yeast artificial chromosome library in a recombination- deficient strain of yeast.
Nature Genetics **1** : 132 - 136
Chakrabarti, D., Reddy, G.R., Dame, J.B., Almira, E.C., Laipis, P.J., Ferl, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M. (1994) Analysis of expressed sequence tags from *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **66** : 97 - 104
Chiu, I.M., Gilmore, E.C., Liu, Y. and Payson, R.A. (1994) Construction of a yeast artificial chromosome contig encompassing the human acidic fibroblast growth factor (FGF1) gene: toward the cloning of the ANLL/ MDS tumor- suppresor gene.
Genomics **20** : 552 - 560
Chu, G., Vollrath, D. and Davis, R.W. (1986) Separation of large DNA molecules by contour- clamped homogenous electric fields.
Science **234** : 1582 - 1585
Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularue, P., Grinas, L., Bougeleret, L., Bellanne- Chanelot, 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., Reithman, 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. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q.
Nature **359** : 380 - 387
Clark, J.M. (1988). Novel non- templated nucleotide addition reactions catalysed by prokaryotic and eukaryotic DNA polymerases.
Nuc. Acid. Res. **16** : 9676 - 9686
Clarke, L. and Carbon, J. (1983) Genomic subsitutions of centromeres in *Saccharomyces cerevisiae.*
Nature **365** : 23 - 28
Clarke, L. and Carbon, J. (1985). The structure and function of yeast centromeres.
Ann. Rev. Genet. **19** : 29 - 56
Coffey, A.J., Roberts, R.G., Green, E.D., Cole, C.G., Butler, R., Anand, R., Giannelli, F and Bentley, D.R. (1992) Construction of a 2.6 Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS based approach.
Genomics **12** : 474 - 484.
Cohen, D., Chumakov, I. and Weissenbach, J. (1993) A first- generation physical map of the human genome.
Nature **366** : 698 - 701
Cole, C.G., Dunham, I., Coffey, A.J., Ross, M.T., Meier- Ewert, S., Bobrow, M. and Bentley, D.R. (1992) A random STS strategy for construction of YAC contigs spanning defined chromosomal regions
Genomics **14** : 256 - 262
Connelly, C., McCormick, M.K., Shero, J. and Hieter, P. (1991) Polyamines eliminate an extreme size bias against transformation of large yeast artificial chromosome DNA.
Genomics **10** : 10 - 16
Cooperative Human Linkage Centre (1994) A comprehensive human linkage map with centimorgan density.
Science **265** : 2049 - 2054
Coppel, R.L., Cowman, A.F., Anders, R.F., Bianco, A.E., Sain, R.B., Lingelbach, K.R., Kemp, D.J. and Brown, G.V. (1984) Immune sera recognise on erythrocytes a *Plasmodium falciparum* antigen composed of repeat amino acid sequences.
Nature **310**: 789 - 792

Corces, V., Holmgren, R., Freund, R., Morimoto, R. and Meselson M. (1980) Four heat shock proteins of *Drosophila melanogaster* coded within a 12- kilobase region in chromosome subdivision 67B.
Proc. Natl. Acad. Sci. **77** : 5390 - 5393
Corcoran, L.M., Thompson, J.K., Walliker, D. and Kemp, D.J. (1988) Homolgous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum.*
Cell **53**: 807 - 813
Cottarel, G., Shero, J.H., Hieter, P. and Hegemann, J.H. (1989) A 125- base pair *CEN6* DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae.*
Mol. Cell. Biol. **9** : 3342 - 3349
Coulson, A., Waterston, R., Kiff, J., Sulston, J. and Kohara, Y. (1988) Genome linking with yeast artificial chromosomes.
Nature **335** : 184 - 186
Coulson, A., Kozone, Y., Lutterbach, B., Shownkeen, R., Sulston, J. and Waterston, R. (1991) YAC's and the *C. elegans* genome.
Bioessays **13**: 413 - 417
Cox, E.C., Vocke, C.D., Walter, S., Gregg, K.Y. and Bain, E.S. (1990) Electrophoretic karyotype for *Dictyostelium discoideum.*
Proc. Nat. Acad. Sci. **87** : 8247 - 8251
Cox, D.R., Green, E.D., Lander, E.S., Cohen , D. and Myers, R.M. (1994) Assessing mapping progress in the human genome project.
Science **265** : 2031 - 2032
Creasey, A.M., Ranford - Cartwright, L.C., Moore R.J., Williamson, D.H., Willson, R.J.M., Walliker, D. and Carter, R. (1993) Uniparental inheritance of the *mitochondrial* gene cytochrome b in *Plasmodium falciparum.*
Curr. Genet. 360 - 364
Creedon, K.A., Kaslow, D.C., Rathod, P.K. and Wellems, T.E. (1992). Identification of a *Plasmodium falciparum* histone 2A gene.
Mol. Biochem. Parasitol. **54** : 113 - 116
Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I. and Cooke, H.J. (1989) Cloning of human telomeres by complementation in yeast.
Nature **338** : 771 - 774
Cumbereledge, S. and Carbon, J. (1987) Mutational analysis of meiotic and mitotic centromere function in *Saccharomyces cerevisiae.*
Genetics **117** : 203 - 212
Dame, J.B., Reddy, G.R., Yowell, C.A., Dunn, B.M., Kay, J. and Berry, C. (1994) Sequence, expression amd modeled structure of an aspartic proteinase from the human malaria parasite *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **64** : 177 - 190
D'Alessandro, U., Leach, A., Drakeley, J., Bennett, S., Olaleye, B.O., Fegan, G. W., Jawara, M., Langerock, P., George, M.O., Targett, G.A.T. and Greenwood, B.M. (1995) Efficacy trial of malaria vaccine SPf66 in Gambian infants.
Lancet **346**: 462 - 467
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 **14** : 332 - 339
De Bruin, D., Lanzer, M. and Ravetch, J.V. (1994) The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements.
Proc. Natl. Acad. Sci. **91** : 619 - 623
de Lange, T. (1994) Activation of telomerase in a human tumour.
Proc. Nat. Acad. Sci. **91** : 2882 - 2885

Delves, C.J., Ridley, R.G., Goman, M., Holloway, S.P., Hyde, J.E. and Scaife, J.G. (1989) Cloning of a beta tubulin gene from *Plasmodium falciparum.*
Mol. Microbiol **3** : 1511 - 1519
Delves, C.J., Alano, P., Ridley, R., Goman, M., Holloway, S.P., Hyde, J.E. and Scaife, J.G. (1990) expression of α and β tubulin genes during the asexual and sexual blood stages of *Plasmodium falciparum*
Mol. Biochem. Parasitol. **43** : 271 - 278.
Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX.
Nuc. Acid. res. **12** : 387 - 395
Dobson, M.J., Tuite, M.F., Mellor, J., Roberts, N.A., King, R.M., Burke, D.C., Kingsman, A.J. and Kingsman, S.M. (1983) Expression in *Saccharomyces cerevisiae* of human interferon- alpha directed by the TRP1 5' region.
Nuc. Acids. Res. **11**: 2287 - 2302
Donis- Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., Botstein, D., Akots, G., Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller- Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, R.G., Crooks, S.M., Lincoln, S.E., Daly, M.J. and Abrahamson, J. (1987) A genetic linkage map of the human genome.
Cell **51** : 319 - 337
Dore, E., Pace, T., Ponzi, M., Scotti, R. and Frontali, C. (1986) Homologous telomeric sequences are present in different species of the genus *Plasmodium.*
Mol. Biochem. Parasitol. **21** : 121 - 227
Drumm, M.L., Smith, C.L., Dean, M., Cole, J.L., Iannuzzi, M.C. and Collins, F.S. (1988) Physical mapping of the Cystic Fibrosis region by pulsed field gel electrophoresis.
Genomics **2** : 346 - 354
Escalante, A.A. and Ayala, F.J. (1994) Phylogeny of the malaria genus *Plasmodium*, derived from rRNA gene sequences.
Proc. Natl. Acad. Sci. **91** : 11373 - 11377
Eyers, M., Edwards, K. and Schuch, W. (1992) Construction and characterisation of a yeast artificial chromosome library containing two haploid *Beta vulgaris L.* genome equivalents.
Gene **121** : 195 - 201
Eubanks, J.H., Altherr, M., Wagner- McPherson, C., McPherson, J.D., Wasmuth, J.J. and Evans, G.A. (1992) Localisation of the D5 dopamine receptor gene to human chromosome 4p15.1 - p15.3, centromeric to the Huntingdon's Disease locus.
Genomics **12** : 510 - 516
Favaloro, J.M., Marshall, V.M., Crewther, P.E., Copple, R.L., Kemp, D.J. and Anders, R.F. (1989) cDNA sequence predicting an octapeptide repeat antigen of *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **32**: 297 - 300
Feagin, J.E.(1992) The 6 Kb element of *Plasmodium falciparum* encodes mitochondrial cytochome genes.
Mol. Biochem. Parasitol. **52** : 145 - 148
Feagin, J.E. (1994) The extrachromosomal DNA's of apicomplexan parasites
Ann. Rev. Micro. **48** : 81 - 104
Feinberg, A. P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity.
Anal. Biochem. **132** : 6 - 13
Foote, S. J. and Kemp, D. J. (1989) Chromosomes of malaria parasites.
TIG **5** : 337 - 342
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

Foote, S., Vollrath, D., Hilton, A. and Page, D.C. (1992) The human Y chromosome: overlapping DNA clones spanning the euchromatic region.
Science **258** : 60 - 64
Fostel, J., Narayanswami, S., Hamkalo, B., Clarkson, S.G. and Pardue, M.L. (1984) chromosomal location of a major tRNA gene cluster of *Xenopus laevis.*
Chromosoma **90** : 254 - 260
Fox, B.F. and Bzik, D.J. (1991) The primary structure of *Plasmodium falciparum* DNA polymerase δ is similar to drug sensitive δ- like viral DNA polymerases.
Mol. Biochem. Parasitol. **49** : 289 - 296
Fox, B.A. and Bzik, D. J. (1994) Analysis of stage- specific transcripts of the *Plasmodium falciparum* serine repeat antigen (SERA) gene and transcription from the SERA locus.
Mol. Biochem. Parasitol. **68** : 133 - 144
Francoeur, A.M., Gritzmacher, C.A., Peebles, C.L.,Reese, R.T. and Tan, E.M. (1985) Synthesis of small nuclear ribonucleoprotein particles by the malarial parasite *Plasmodium falciparum.*
Proc. Nat. Acad. Sci. **82** : 3635 - 3639
Frohman, M.A. and Martin, G.A. (1989) Rapid amplification of cDNA ends using nested primers.
Technique **1** : 165 - 170
Froussard, P. (1992) A random- PCR method (rPCR) to construct whole cDNA library from low amounts of RNA.
Nuc. Acid. Res. **20** : 2900
Fulton, T.R., Bowcock, A.M., Smith, D.R., Daneshvar, L., Green, P., Cavalli- Sforza, L.L. and Donis- Keller, H. (1988) A 12 megabase restriction map of the cystic fibrosis locus.
Nuc. Acids. Res. **17** : 271 - 284
Gaensler, K.M.L., Burmeister, M., Brownstein, B.H., Taillon- Miller, P. and Myers, R.M. (1991). Physical mapping of yeast artificial chromosomes from the human β- globin gene region.
Genomics **10** : 976 - 984
Gardner, M.J., Preiser, P., Rangachari, K., Moore, D., Feagin, J.E., Williamson, D.H. and Wilson, R.J.M. (1994a) Nine duplicated tRNA genes on the plastid- like DNA of the malaria parasite *Plasmodium falciparum.*
Gene. **144** : 307 - 308
Gardner, M.J., Goldman, N., Barnett, P., Mappre, P.W., Rangachari, K., Strath, M., Whyte, A., Williamson, D.H. and Wilson, R.J.M. (1994b) Phylogenetic analysis of the *rpoB* gene from the plastid- like DNA of *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **66** : 221 - 231
Garza, D., Ajioka, J.W., Burke, D.T. and Hartl, D.L. (1989) Mapping the *Drosophila* genome with yeast artifical chromosomes.
Science **246** : 641 - 646
Goman, M., Langsley, G., Hyde, J.E., Yankovsky, N.K., Zolg, J.W. and Scaife, J.G. (1982) The establishment of genomic DNA libraries for the human malaria parasite *Plasmodium falciparum* and identification of individual clones by hybridisation.
Mol. Biochem. Parasitol. **5** : 391 - 400
Goonewardene, R., Daily, J., Kaslow, D., Sullivan, T.J., Duffy, P., Carter, R., Mendis, K. and Wirth, D. (1993) Transfection of the malaria parasite and expression of firefly luciferase.
Proc. Natl. Acad. Sci. **90** : 5234 - 5236
Goustin, A.S. (1981) Two temporal phases for the control of histone gene activity in cleaving sea urchin embryos *S. purpatus.*
Dev. Biol. **82**: 163 - 175
Green, E.D. and Olson, M.V. (1990a.) Systematic screening of yeast artificial- chromosome libraries by use of the polymerase chain reaction.
Proc. Nat. Acad. Sci. **87** : 1213 - 1217.
Green, E.D. and Olson, M.V. (1990b.) Chromosomal region of the Cystic Fibrosis gene in artificial chromosomes: a model for human genome mapping.
Science **250** : 94 - 98

Green, E.D., Reithman, H.C., Dutchik, J.E. and Olson, M.V. (1991a.) Detection and characterisation of chimeric yeast artificial chromosomes.
Genomics **11** : 658 - 669
Green, E.D., Mohr, R.M., Idol, J.R., Jones, M., Buckingham, J.M., Deaven, L.L., Moyzis, R.K. and Olson, M.V. (1991b) Systematic generation of sequence- tagged sites for physical mapping of human chromosomes: application to the mapping of human chromosome 7 using yeast artificial chromosomes.
Genomics **11** : 548 - 564
Green, E.D., Braden, V.V., Fulton, R.S., Lim, R., Ueltzen, M.S., Peluso, D.C., Mohr- Tidwel, R. M., Idol, J.R., Smith, L.M., Chumakov, I., Le Paslier, D., Cohen, D., Featherstone, T. and Green, P. (1995) A human chromosome 7 yeast artificial chromosome (YAC) resource: construction, characterisation and screening.
Genomics **25** : 170 - 183
Greider, C.W. and Blackburn, E.H. (1989) A telomeric sequence in the RNA of *Tetrahymera* telomerase required for telomere repeat synthesis.
Nature **337** : 337 - 33
Grothues, D. and Tummler, B. (1991) New approaches in genome analysis by pulsed field gel electrophoresis: application to the analysis of *Pseudomonas* species.
Mol. Microbiol. **5** : 2763 - 2776
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.
Ex. Parasit. **71** : 189 - 198
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
Hadano, S., Ishida, Y., Bates, G.P., Nagayama, T., Kanazawa, I., Lehrach, H. and Ikeda, J- E. (1993). Generation of high- density markers from yeast artificial chromosome DNA by single unique primer- polymerase chain resction.
GATA **10** : 105 - 108
Haldi. M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D. and Lander, E.S. (1994) Large human YAC's constructed in a *rad52* strain show a reduced rate of chimerism.
Genomics **24** : 478 - 484
Hamvas, R.M.J., Larin, Z., Brockdorff, N., Rastan, S., Lehrach, H., Chartier, F.L. and Brown, S.D.M. (1993) YAC clone contigs surrounding the *zfx* and *pola* loci on the mouse X chromosome.
Genomics **17** : 52 - 58
Hamvas, R.M.J., Francis, F., Cox, R.D., Nizetic, D., Goldsorthy, M.E., Brown, S.D.M. and Lehrach, H.R. (1994). Rapid restriction analysis of YAC clones.
Nuc. Acid. Res. **22** : 1318 - 1319
Harvey, R.P., Krieg, P.A., Robins, A.J., Coles, L.S. and Wells, J.R.E. (1981) Non- tandem arrangement and divergent transcription of chicken histone genes.
Nature **294** : 49 - 53
Haynes, D.J., Diggs, C.L., Hines, F.A. and Desjardines, R.E. (1976) Culture of of human malaria parasites *Plasmodium falciparum.*
Nature **263** : 767 - 769
Heard, E., Davies, B., Feo, S. and Fried, M. (1989) An improved method for the screening of YAC libraries.
Nuc. Acids. Res. **17** : 5861
Hegemann, J.H., Shero, J.H., Cottarel, G., Philippsen, P. and Hieter, P. (1988) Mutational analysis of centromere DNA from chromosome VI of *Saccharomyces cerevisiae.*
Mol. Cell. Biol. **8** : 2523 - 2535
Hereford, L.M., Fahrner, K., Woolford, J. Jr., Rosbash, M. and Kaback, D.B. (1979) Isolation of yeast histone genes 2A and 2B.
Cell **18** : 1261 - 1271
Hentschel, C.C. and Birnsteil, M.X. (1981) The organisation and expression of histone gene families.

Cell. **1981** : 301 - 313

Hieter, P., Mann, C., Snyder, M. and Davis, R.W. (1985) Mitotic stability of yeast chromosomes: a colony colour assay that measures nondisjunction and chromosome loss.
Cell **40** : 381 - 392

Hinnen, A., Hicks, J, B. and Fink, G, R. (1978) Transformation of yeast.
Proc. Natl. Acad. Sci **75**: 1929 - 1933

Hinterberg, K., Mattei, D., Wellems, T.E. and Scherf, A. (1994) Interchromosomal exchange of a large subtelomeric segment in a *Plasmodium falciparum* cross.
EMBO **13** : 4174 - 4180

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.
Nuc. Acid. Res. **19** : 3283 - 3288

Holloway, S.P., Gerousis, M., Delves, C.J., Sims, P.F.G., Scaife, J.G. and Hyde, J.E. (1990) The tubulin genes of the human malaria parasite *Plasmodium falciparum,* their chromosomal location and sequence analysis of the α- tubulin II gene.
Mol. Biochem. Parasitol. **43** : 257 - 270

Hough- Evans, B.R. and Howard, J. (1982) Genome size and DNA complexity of *Plasmodium falciparum*
Biochemica et Biophysica Acta. **698:** 56 - 61

Hovemann, B., Sharp, S., Yamada, H. and Soll, D. (1980) Analysis of a *Drosophila* tRNA gene cluster.
Cell **19** ; 889 - 895

Huberman, J.A., Zhu, J., Davis, L.R. and Newlan, C.S. (1988) Close association of a DNA replication origin and an ARS element on chromosome III of the yeast *Saccharomyces cerevisiae.*
Nuc. Acids. Res. **16** : 6373 - 6384

Hunkapiller, T., Kaiser, R.J., Koop, B.F. and Hood, L. (1991) Large- scale and automated DNA sequence determination.
Science **254** : 59 - 67

Huxley, C., Hagino, Y., Schlessinger, D. and Olson, M.V. (1991a.) The humanHPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion.
Genomics **9** : 742 - 750

Hyde, J.E., Goman, M., Hall, R., Osland, A., Hope, I.A., Langsley, G., Zolg, J.W. and Scaife, J.G. (1984) Characterisation and translation studies of messenger RNA from the human malaria parasite *Plasmodium falciparum* and construction of a cDNA library.
Mol. Biochem. Parasitol. **10** : 269 - 285

Imai, T. and Olson, M.V. (1990) Second- generation approach to the construction of yeast artificial chromosome libraries
Genomics : 297 - 303

Janse, C.J. (1993) Chromosome size polymorphism and DNA rearrangements in *Plasmodium.*
Parasitol. Today **9** : 19 - 22

Jelinek, W.R. and Schmid, C.W. (1982) Repetitive sequences in eukaryotic DNA and their expression.
Ann. Rev. Biochem. **51** : 813 - 814

Karcz, S.R., Herrmann, V.R. and Cowman, A.F. (1993) cloning and characterisation of a vacuolar ATPase A subunit homlogue from *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **58**: 333 - 334

Kemp, D.J., corcoran, L.M., Coppel, R.L., Stahl, H.D., Bianco, A.E., Brown, G.V. and Anders, R. F. (1985) Size variation in chromosomes from independent cultured isolates of *Plasmodium falciparum.*
Nature **315** : 347 - 350

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.
Proc. Natl. acad. Sci. **84** : 7672 - 7676

Kere, J., Nagaraja, R., Mumm, S., Ciccodicola, A., D'Urso, M. and Schlessinger, D. (1992) Mapping human chromosomes by walking with sequence tagged sites from end fragments of yeast artificial chromosome inserts.
Genomics **14** : 241 - 248
Kerr, P.J., Ranford- Cartwright, L.C.and Walliker, D. (1994) Proof of intragenic recombination in *Plasmodium falciparum.*
Mol. Biochem. Parastiol. **66**: 241 - 248
Kilbey, B.J., Fraser, I., McAleese, S., Goman, M., Ridley, R.G. (1993) Molecular characterisation and stage specific expression of proliferating cell nuclear antigen (PCNA) from the malarial parasite *Plasmodium falciparum.*
Nuc. Acids. Res. **21** : 239 - 243.
Kim, U- J., Shizuya, H., Birren, B., Slepak, T., de Jong, P. and Simon, M.I. (1994) Selection of chromosome 22- specific clones from human genomic Bac library using a chromosome- specific cosmid library pool
Genomics **22** : 336 - 339.
Kimura, E., Mattei, D., Mana di Santi, S. and Scherf, A. (1990) Genetic diversity in the major merozoite surface antigen of *Plasmodium falciparum*: high prescence of a third polymorphic form detected in strains derived form malaria parasites.
Gene **91** : 57 - 62.
Knapp, B., Nau, U., Hundt, E. and Kupper, H.A. (1991) A new blood stage antigen of *Plasmodium falciparum* highly homologous to the serine- stretch protein SERP.
Mol. Biochem. Parasitol. **44** : 1 - 14
Koella, J.C. (1993) Epidemiological evidence for an association between chloroquine resistance of *Plasmodium falciparum* and its immunological properties.
Parasitol. Today. **9** : 105 - 108
Kondrachine, A.V. and Trigg, P. (1995) Malaria: hope for the future
World Health **48** : 26 - 27
Kouprina, N., Eldarov, M., Moyzis, R., Resnick, M. and Larionov, V. (1994) A model system to asses the integrity of mammalian YAC's during transformation and propagation in yeast.
Genomics **21** : 7 - 17
Kozman, H.M., Keith, T.P., Donis- Keller, H., White, R.L., Weissenbach, J., Dean, M., Verngaud, G., Kidd, K., Gusella, J., Royle, N.J., sutherland, G.R. and Mulley, J.C. (1995) The CEPH consortium linkage map of human chromosome 16.
Genomics **25** : 44 - 58
Kozono, H., Bronson, S.K., Taillon- Miller, P., Moorti, M.K., Jamry, I. and Chaplin, D.D. (1991) Molecular linkage of the HLA- DR, HLA- DQ and HLA- DO genes in yeast artificial chromosomes.
Genomics **11** : 557 - 586.
Krishna, S., Cowan, G. M., Meade, J.C., Wells, R.A., Stringer, J.R. and Robson, K.J. (1993) A family of cation ATPase- like molecules from *Plasmodium falciparum.*
J. Cell. Biol. **120** : 385 - 398
Krishna, S., Cowan, G.M., Robson K. and Meade, J.C. (1994) *Plasmodium falciparum*: further characterisation of putative cation ATPases.
Ex. Para. **78** : 113 - 117
Kuhn, R.M. and Ludwig, R.A. (1994) Complete sequence of the yeast artificial chromosome cloning vector pYAC4.
Gene **141** : 127 - 127
Kuspa, A., Vollrath, D., Cheng, Y. and Kaiser, D. (1989) Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes.
Proc. Natl. Acad. Sci. **86**: 8917 - 8921
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
Kwiatkowski, T.J., Zoghbi, H.Y., Ledbetter, S.A., Ellison, K. A. and Chinault, C.A. (1990) Rapid identification of yeast artificial chromosome clones by matrix pooling and crude lysate PCR.

Nuc. Acid. Res **18** : 7191 - 7194
Langsley, G., Sibilli, L., Mattei, D., Falanga, P and Mercereau- Puijalon, O. (1987) Karyotype comparison between *P. chabaudi* and *P. falciparum*: analysis of a *P. chabaudi* cDNA containing sequences highly repetitive in *P. falciparum.*
Nuc. Acids. Res. **15** : 2203 - 2211
Langsley, G. and Ponnundurai, T. (1988) *Plasmodium falciparum*: analysis of karyotpye polymorphism using chromosome specific probes
Ex. Parasitol. **65**: 168 - 173
Lanzer, M., de Bruin, D. and Ravetch, J.V. (1992) Transcription mapping of a 100 Kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates.
EMBO J. **11** :1949 - 1955
Lanzer, M., de Bruin, D. and Ravetch, J.R. (1993) Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome.
Nature **361** : 654 - 657
Lanzer, M., DeBruin, D., Wertheimer, S.P. and Ravetch, J. (1994) Transcriptional and nucleosomal characterisation of a subtelomeric gene cluster flanking a site of chromosomal rearrangements in *Plasmodium falciparum.*
Nuc. acids. Res. **22** : 4176 - 4182
Larin, Z., Monaco, A.P. and Lehrach, H. (1991) Yeast Artificial Chromosome libraries containing large inserts from mouse and human DNA.
Proc. Nat. Acad. Sci. **88** 5109 - 5113
Larionov, V., Kouprina, N., Nikolaishvili, N. and Resnick, M.A. (1994) Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YAC's)
Nuc. Acids. Res. **22** : 4154 - 4162
Lasker, B.A., Carle, G.F., Kobayashi, G.S. and Medoff, G. (1989) Comparison of the separation of *Candida albicans* chromosome- sized DNA by pulsed field gel electrophoresis techniques.
Nuc. Acids. Res. **17** : 3783 - 3793
Lathrop, G.M., Farrall, M., O'Connell, P., Wainwright, B., Leppert, M., Nakamura, Y., Lench, N., Kruyer, H., Dean, M., Park, M., Vande Woude, G., Lalouel, J- M., Williamson, R. and White, R. (1988) Refined linkage map of chromosome 7 in the region of the cystic fibrosis gene.
Am. J. Hum. Genet. *42* : 38 - 44
Lauer, J., Shen, C- K, J. and Maniatis, T. (1980) The chromosomal arrangement of human α- like globin genes: sequence homology and α- globin gene deletions.
Cell **20** : 119 - 130
Lee, J.T., Murgia, A., Sosnoski, D.M., Olivos, I.M. and Nussbaum, R.L. (1992) Construction and characterisation of a yeast artificial chromosome library for Xpter- Xq27.3: A systematic determination of cocloning rate and X- chromosome representation.
Genomics **12** : 526 - 533
Levitt, A (1993) RNA processing in malaria parasites.
Parasitol. Today **9** : 465 - 468
Li, W- B., Bzik, D.J., Gu, H., Tanaka, M., Fox, B.F. and Inselburg, J. (1989) An enlarged subunit of *Plasmodium falciparum* RNA polymerase II defines conserved and variable polymerase domains.
Nuc. Acids. Res. **17** : 9621 - 9636
Li, W., Bzik, D.J., Tanaka, M., Gu, H., Fox, B.F. and Inselburg, J. (1991) Characterisation of the gene encoding the largest subunit of Plasmodium falciparum RNA Polymerase III.
Mol. Biochem. Parasitol. **46** : 229 - 240
Ling, L.l., Ma, N. S- F., Smith, D.R., Miller, D.D. and Moir, D.T. (1993) Reduced occurrence of chimeric YAC's in recombination deficient hosts.
Nuc. Acids. Res. **21** : 6045 - 6046
Little, R.D., Pilia, G., Johnson, S., D'Urso, M. and Schlessinger, D. (1992) Yeast artificial chromosomes spanning 8 megabases and 10 - 15 centimorgans of human cytogenetic band Xq26.
Proc. Nat. Acad. Sci. **89** : 177 - 181

Liu, J., Stanton, V.P., Fujiwara, T.M., Wang, J- X., Rezonzew, R., Crumley, M.J., Morgan, K., Gros, P., Housman, D. and Schurr, E. (1995) Large- scale cloning of human chromosome 2- specific yeast artificial chromosomes (YACs) using an interspersed repetitive sequences (IRS)- PCR approach.
Genomics **26** : 178 - 191
Longhurst, H.J. and Holder, A.A. (1995) The sequence of *Plasmodium falciparum* histone H3
Mol. Biochem. Parasitol. **69** : 111 - 113
MacMurray, A.J., Weaver, A., Shin, H- S. and Lander, E.S. (1991) An automated method for DNA preparation from thousands of YAC clones.
Nuc. Acids. Res. **19** : 385 - 390
Mandel, J- M., Monaco, A.P., Nelson, D.L., Schlessinger, D. and Willard, H. (1992) Genome analysis and the human X chromosome.
Science **258** : 103 - 109
Mann, C. and Davis, R.W. (1986) Structure and sequence of the centomeric DNA of chromosome 4 in *Saccharomyces cerevisiae.*
Mol. Cell. Biol. **6** : 241 - 245
Marahrens, Y. and Stillman, B. (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements.
Science **255** : 817 - 823
Marashi, F., Baumbach, L., Rickles, R., Sierra, F., Stein, J.L. and Stein, G.A (1982) Histone proteins in HeLa S3 cells are synthesised in a cell- cycle stage specific manner.
Science **215** : 683 - 685
Marchuk, D. and Collins, F.S. (1988) pYAC- RC, a yeast artificial chromosome vector for clonoing DNA cut with infrequently cutting restriction endonucleases.
Nuc. Acids.Res. **16** : 7743
Marchuk, D., Tavakkol, R., Wallace, M.R., Brownstein, B.H., Taillon- Miller, P., Fong, C- T, Leguis, E., Andersen, L.B., Glover, T.W. and Collins, F. (1992)
A yeast artificial chromosome contig encompassing the type 1 Neurofibromatosis gene.
Genomics **13** : 672 - 680
Maurice, J.(1995) Malaria vaccine raises a dilemma
Science **267** : 320 - 323
Maxson, R., Cohn, R. and Kedes, L. (1983) Expression and organisation of histone genes.
Ann. Rev. Genet. **17** : 239 - 277
McAndrew, M.B., Read, M., Sims, P.F.G. and Hyde, J.E. (1993). Characterisation of the gene encoding an unusually divergent TATA- binding protein (TBP) from the extremely A plus T rich human malaria parasite *Plasmodium falciparum.*
Gene **124** : 165 - 171
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. (1992) *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues.
Nature genetics **1** : 124 - 131
McConologue, L., Brow, M. A. D. and Innis, M.A. (1988) Structure- independent DNA amplification by PCR using 7- deaza- 2'- deoxyguanosine.
Nuc. Acid. Res. **16** : 9869
McCormick, M.K., Antonarakis, S. and Hieter, P. (1990) YAC cloning of DNA embedded in an agarose matrix.
GATA **7**: 114 - 118
McCormick, M.K., Campbell, E., Deaven, L. and Moyzis, R. (1993a) Low- frequency chimeric yeast artificial chromosome libraries from flow- sorted human chromosomes 16 and 21.
Proc. Nat. Acad .Sci . **90** : 1063 - 1067
McCormick. M.K., Buckler, A., Bruno, W., Campbell, E., Shera, K., Torney, D., Deaven, L., Moyzis, R. (1993b) Construction and characterisation of a YAC library with a low frequency of chimeric clones from flow- sorted human chromsosome 9.
Genomics **18** : 553 - 558

McCutcheon, T.F., Dame, J.B., Miller, L.H. and Barnwell, J. (1984) Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA
Science **225** : 808 - 811
McLelland, M., Hanish, J., Nelson, M. and Patel, Y. (1988) KGB: a single buffer for all restriction endonucleases.
Nuc. Acids. Res. **16** : 364 - 366
Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. and Smith, L.M. (1991). A universal method for the direct cloning of PCR amplified nucleic acid.
Bio/Technology **9** : 657 - 663
Meese, E.U., Meltzer, P.S., Ferguson, P.W. and Trent, J.M. (1992) *Alu*- PCR: characterisation of a chromosome 6- specific hybrid mapping panel and cloning of chromosome- specific markers.
Genomics **12** : 549 - 554
Mendez, M.J., Abderrahim, H.M., Noguchi, M., David, N.E., Hardy, M.C., Green, L.L., Tsuda, H., Yoast, S., Maynard- Currie, C.E., Garza, D., Gemmill, R., Jakobovits, A. and Klapholz, S. (1995) Analysis of the structural integrity of YAC's comprising human immunoglobulin genes in yeast and embryonic stem cells.
Genomics **26** : 294 - 307
Menendez, C. (1995) Malaria during pregnancy: a priority area of malaria research and control
Parasitol. Today **11** : 178
Moelans, I.I.M.D., Meis, J.F.G.M, Kocken, C., Konings, R.N.H. and Schoenmakers, J.G.G. (1991) A novel protein antigen of the malaria parasite *Plasmodium falciparum* located on the surface of gametes and sporozoites.
Mol. Biochem. Parasitol. **45** : 193 - 204
Monaco, A.P., Walker, A.P., Millwood, I., Larin, Z. and Lehrach, H. (1992) A yeast artificial chromosome contig containing the complete Duchenne Muscular dystrophy gene.
Genomics **12** : 465 - 473
Monaco, A.P., Larin, Z. and Lehrach, H. (1993) Construction of Yeast Artificial Chromosome libraries by Pulsed- Field Gel electrophoresis.
Methods in Molecular Biology **12** 225 - 234
Muller, U., Stephan, D., Phillipsen, P. and Steinmetz, M. (1987) Orientation and molecular map position of the complement genes in the mouse MHC.
EMBO J. **6** : 369 - 373
Murakami, S., Matsumoto, T., Niwa, O. amd Yanagida, M. (1991) Structure of the fission yeast centromere cen3: direct analysis of the reiterated inverted region .
Chromosoma **101** : 214 - 221
Murray, A.W. and Szostak, J.W. (1985). Chromosome segregation in mitosis and meiosis.
Ann. Rev. Cell biol. **1** : 289 - 315
Murray, A.W., Schultes, N.P. and Szostak, J.W. (1986) Chromosome length controls mitotic chromosome segregation in yeast.
Cell **45** : 529 - 536
Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M., Weissenbach, J., Gyapay, G., Dib, C., Morrissette, J., Lathrop, G.M., Vignal, A., White, R., Matsunami, N., Gerken, S., Melis, R., Albertsen, H., Plaetke, R., Odelberg, S., Ward, D., Dausset, J., Cohen, D. and Cann, H. (1994) A comprehensive human linkage map with centimorgan density.
Science **265** : 2049 - 2054
Murrell, J., Trofatter, J., Rutter, M., Cutone, S., Stotler, C., Rutter, J., Long, K., Turner, A., Deaven, L., Buckler, A. and McCormick, M.K. (1995) A 500- kilobase region containing the Tuberous Sclerosis locus (TSC1) in a 1.7- Megabase YAC and cosmid contig.
Genomics **25** ; 59 - 65
Nagaraja, R., Kere, J., MacMillan, S., Masisi, M.W.J., Johnson, D., Molini, B.J., Halley, G.R., Wein, K., Trsgnich, M., Eble, B., Railey, B., Brownstein, B.H. and Schlessinger, D. (1994) Characterisation of four human YAC libraries for clone size, chimerism and X chromosome sequence representation.

Nuc. Acids. Res. **22** : 3406 - 3411
Neill, D.L., Villasante, A., Fisher, R.B., Vetrie, D., Cox, B. and Tyler- Smith, C. (1990) Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors.
Nuc. Acids. Res. **18** : 1421 - 1428
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.
Proc. Nat. Acad. Sci. **86** : 6686 - 6690
Nemeth, A.H., Hunter, N., Coleman, M.P., Borts, R.H., Louis, E.J. and Davies, K.E. (1993) Rescue of a single yeast artificial chromosome from a cotransformation event utilising segregation at meiosis
GATA **10** : 123 - 127
Ng, R. and Carbon, J. (1987) Mutational and *in vitro* protein binding studies on centromere DNA from *Saccharomyces cerevisae*
Mol. Cell. Biol. **7** : 4522 - 4534
Nguyen, C., Djabali, M., Roux, D. and Jordan, B.R. (1990) Very high molecular weight DNA for pulsed field gel studies can be obtained routinely from conventional frozen blood aliqouts.
Nuc. Acid. res. **19** : 407
NIH/ CEPH Collaborative Mapping Group (1992) a comprehensive genetic linkage map of the human genome.
Science **258** : 67- 77
O'Connell, P., Albertsen, H.A., Matsunami, N., Taylor, T., Hundley, J.E., Johnson- Pais, T.L., Reus, B., Lawrence, E., Ballard, L., White, R. and Leach, R.J. (1994) A radiation hybrid map of the *BRCA1* region.
Am. J. Hum. Genet. **54** : 526 - 534
O'Connor, M., Peifer, M. and Bender, W. (1989) Construction of large DNA segments in *Escherichia coli.*
Science **244** : 1307 - 1312
Ochman, H., Gerber, A. S. and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction.
Genetics **120** : 621 - 623
Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989) A common language for the physical mapping of the human genome.
Science **245** : 1434 - 1435
Oquendo, P., Goman, M., MacKay, M., Langsley, G., Walliker, D and Scaife, J.G. (1986) Characterisation of a repetitive DNA sequence from the malaria parasite *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **18**: 89 - 101
Orbach, M.J., Vollrath, D., Davis, R.W. and Yanofsky, C. (1988) An electrophoretic karyotype of *Neurospora crassa.*
Mol. Cell. Biol. **8** : 1469 - 1473
Orr- Weaver, T.L. and Szostak, J.W. (1983) Yeast recombination: the association between double strand gap repair and crossing over.
Proc. Nat. Acad. Sci. **80** : 4417 - 4421
Palazzolo, M.J., Sawyer, S.A., Martin, C.M., Smoller, D.A. and Hartl, D.L. (1991) Optimised strategies for sequence- tagged site selection in genome mapping.
Proc. Nat. Acad. Sci. **88** : 8034 - 8038.
Panzeri, L., Landonio, L., Stotz, A. and Philippsen, P. (1985) Role of conserved sequence elements in yeast centromere DNA.
EMBO J. **4** : 1867 - 1874
Pasloske, B.L., Baruch, D.I., Ma, C., Tarschi, T.F., Gormley, J.A. and Howard, R.J. (1994) PfEMP3 and HRPI: co- expressed genes localised to chromosome 2 of *Plasmodium falciparum.*
Gene **144** : 131 - 136
Patel, K., Sheer, D. and Hampton, G.M. (1993) "Junction trapping" a simple PCR- based method for the isolation of YAC- insert termini.
GATA **10** : 42 - 48

Petruhkin, K., Fischer, S.G., Pirastu, M., Tanzi, R.E., Chernov, I., Devoto, M., Brzustowicz, L.M., Cayanis, E., Vitale, E., Russo, J.J., Matseoane, D., Boukhgalter, B., Wasco, W., Figus, A.L., Loudianos, J., Cao, A., Sternleib, I., Evgrafov, O., Parano, E., Pavone, L., Warburton, D., Ott, J., Penchaszadeh, G.K., Scheinberg, I.H. and Gilliam, T.C. (1993) Mapping, cloning and genetic characteriastion of the region containing the Wilson disease gene.
Nature Genet. **5** : 338 - 343

Prensier, G. and Slomianny, C. (1986) The Karyotype of *Plasmodium falciparum* determined by ultrastructural serial sectioning and 3D reconstruction.
J. Parasitol. **72** : 731 - 736

Pollack, Y., Katzen, A.L., Spira, D. and Golenser, J. (1982) The genome of *Plasmodium falciparum* I: DNA base composition.
Nuc. Acids. Res. **10** : 539 - 546

Pologe, L.G. and Ravetch, J. (1988) Large deletions result from breakage and healing of *P. falciparum* chromosomes.
Cell **16** : 869 - 874

Qin, S., Zhang, J., Isaacs, C.M., Nagafuchi, S., Sait, S.N.J., Abel, K.J., Higgins, M.J., Nowak, N.J. and Shows, T.B. (1993) A chromosome 11 YAC library
Genomics **16** : 580 - 585

Ragoussis, J., Monaco, A., Mockridge, I., Kendall, E., Campbell, R.D. and Trowsdale, J. (1991) Cloning of the HLA class II region in yeast artificial chromosomes.
Proc. Nat. Acad. Sci. **88** : 3753 - 3757

Ranford- Cartwright, L.C., Balfe, P., Carter, R. and Walliker, D. (1993) Frequency of cross-fertilisation in the human malaria parasite *Plasmodium falciparum.*
Parasitology **107** : 11 - 18

Read, M., Hicks, K.E., Sims, P.F.G. and Hyde, J.E. (1994) Molecular characterisation of the enolase gene from the human malaria parasite *Plasmodium falciparum*: evidence for ancestry within a photosynthetic lineage.
Eur. J. Biochem. **220** : 513 - 520

Reddy, G.R., Chakrabarti, D., Schuster, S.M., Ferl, R.J., Amira, E.C. and Dame, J.B. (1993) Gene sequence tags from *Plasmodium falciparum* genomic DNA fragments prepared by the "genease" activity of mung bean nuclease
Proc. Nat. Acad. Sci. **90** : 9867 - 9871

Ridley, R.G., White, J.H., McAleese, S.M., Goman, M., Alano, P., de Vries, E. and Kilbey, B.J. (1991) DNA polymerase δ : gene sequences from *Plasmodium falciparum* indicate that this enzyme is more highly conserved than DNA polymerase α.
Nuc. Acids. Res. **19** : 6731 - 6736

Riley, J., Butler, R., Ogilvie, D., Jenner, D., 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.
Nuc. Acids. res. **18** : 2887 - 2890

Riley, J.H., Ogilvie, D. and Anand, R. (1992) Construction, charaterisation and screening of YAC libraries.
in Techniques for the Analysis of Complex Genomes Academic Press

Riordan, J.R., Rommens, J.M., Kerem, B- S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J- L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S. and Tsui, L- C.(1989) Identification fo the cystic fibrosis gene: cloning and characterisation of the complementary DNA
Science **245** : 1066 - 1073

Rivier, D.H. and Rine, J. (1992) An origin of DNA replication and a transposon silencer require a common element.
Science **256** : 659 - 663

Robson, K.J.H., Hall, J.R.S., Jennings, M.W., Harris, T.J.R., Marsh, K., Newbold, C.I., Tate, V.E. and Weatherall, D.J. (1988) A highly conserved amino acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite.
Nature **335** : 79 - 83

Robson, K.J.H. and Jennings, M.W. (1991) The structure of the calmodulin gene of *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **46** : 19 - 34
Rommens, J.M., Iannuzzi, M.C., Kerem, B- S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L- C. and Collins (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping.
Science **245** : 1059 - 1066
Rubio, J.P., Triglia, T., Kemp, D.J., De Bruin, D., Ravetch, J. and Cowman A.F. (1995) A YAC contig map of *Plasmodium falciparum* chromosome 4: characterisation of a DNA amplification between two recently separated isolates.
Genomics **26** : 192 - 198
Saiki, R.K., Gelfland, D.H., Stoffel, S., Scharf, Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich H.A. (1988) Primer directed amplification of DNA with a thermostable DNA polymerase.
Science **239** : 487 - 491
Sanger, F., Nicklen, S. and Coulson, R. (1977) DNA sequencing with chain terminating inhibtors.
Proc. Nat. Acad. Sci. **74** : 5463 - 5467.
Schapira, A., Beales, P.F. and Halloran, M.E. (1993) Malaria: living with drug resistance
Parasitol. Today **9** : 168 - 174
Scherf, A. and Mattei, D. (1992) Cloning and charaterisation of chromosome breakpoints of *P. falciparum*: breakage and new telomere formation occurs frequently and randomly in subtelomeric genes.
Nuc. Acids. Res. **20**: 1491 - 1496
Schiestl, R.H. and Prakash, S. (1988) RAD1, an excision repair gene of *Saccharomyces cerevisiae*, is also involved in recombination.
Mol. Cell. Biol. **8** : 3619 - 3626
Schlessinger, D., Little, R.A., Freue, D., Abidi, F., Zucchi, I., Porta, G., Pilia, G., Nagaraja, R., Johnson, S.K., Yoon, J-Y., Srivastava, A., Kere, J., Palmieri, G., Ciccodicola, A., Montanaro, V., Romano, G., Casamassimi, A. and D'Urso, M. (1991) Yeast artificial chromosome- based genome mapping: some lessons from Xq24 - q28
Genomics **11** : 783 - 793
Schuster, B.G. and Milhous, W.K. (1993) Reduced resources applied to anti- malarial drug development.
Parastiol. Today **9** : 167 - 168
Schmidt, R. and Dean, C. (1993) Towards construction of an overlapping YAC library of the *Arabidopsis thaliana* genome.
BioEssays **15**: 63 - 69
Schwartz, D.C. and Cantor, C.R. (1984) Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis.
Cell **37** : 67 - 75
Selleri, L., Hermanson, G.G., Eubanks, J.H. and Evans, G.A. (1991) Chromosomal *in situ* hybridisation using yeast artificial chromosomes.
GATA **8** : 59 - 66
Selleri, L., Eubanks, J.H., Giovannini, M., Hermanson, G.G., Romo, A., Djabali, M., Maurer, S., McElligott, D.L., Smith, M.V. and Evans, G.A. (1992) Detection and characterisation of "chimeric" yeast artificial chromosomes by fluorescent *in situ* suppression hybridisation.
Genomics **14** : 536 - 541
Sharma, Y.D. (1992) Structure and possible function of heat shock proteins in *falciparum* malaria.
Comp. Biochem. Phys. **102** : 437 - 444
Shaw, J. and Olson, M.V. (1984). Effects of altered 5' flanking sequences on the *in vivo* expression of a *Saccharomyces cerevisiae* $tRNA^{tyr}$ gene.
Mol. Cell. Biol. **4** : 657 - 665
Shero, J.H., McCormick, M.K., Antonarakis, S.E. and Hieter, P. (1991) Yeast artificial chromosome vectors for efficient clone manipulation and mapping.
Genomics **10** : 505 - 508

Shippen- Lentz, D. and Blackburn, E.H. (1990a.) Functional evidence for an RNA template in telomerase
Science **247** :546 - 552
Shippen- Lentz, D., Afroze, T. and Vezza, A.C. (1990b.) Heterogeneity and expresion of the *Plasmodium falciparum* 5.8s ribosomal RNA genes.
Mol. Biochem. Parasitol. **38** : 113 - 120
Shirley, M.W., Biggs, B.A., Forsyth, K.P., Heidi, J.M., 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.*
Mol. Biochem. Parasitol. **40** : 137 - 146
Shizuya, H., Birren, B., Kim, U- J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300 kilobase- pair fragments of human DNA in *Escherichia coli* using an F- factor- based vector.
Proc. Nat. Acad. Sci. **89** : 8794 - 8797
Silverman, G.A., Ye, R.D., Pollock, 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
Proc. Nat. Acad. Sci. **86** : 7485 - 7489.
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.
Proc.Nat. Acad. Sci. **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
Simpson, A.J.G., Sher, A. and McCutcheon, T.F. (1982) The genome of *Schistosoma mansoni*: isolation of DNA, it's size, bases and repetitive sequences.
Mol. Biochem. Parasitol. **6** : 125 - 137
Sinden, R. E. and Hartley, R.H. (1985) Identification of the meiotic division of malarial parasites.
J. Protozool. **32** : 742 - 744
Singer, M.F. (1982) SINE's and LINE's: highly interspersed repeated short and long interspersed sequences in mammalian genomes.
Cell **28** : 433 - 434
Singer, M.S. and Gottschling, D.E. (1994) *TLC1*: Template RNA component of *Saccharomyces cerevisiae* telomerase.
Science **266** : 404 - 409
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
Sloan, D.D., Blanchard, M.M., Burough, F.W. and Nowotny, V. (1993) Screening yeast artificial chromosome libraries with robot- aided automation.
GATA **10** : 128 - 143
Smith, C.L. and Cantor, C.R. (1986) Pulsed field gel electrophoresis of large DNA molecules.
Nature **319** :701 - 703
Smith, D.R., Smyth, A.P. and Moir, D.T. (1990) Amplification of large artificial chromosomes.
Proc. Nat. Acad. Sci. **87** : 8242 - 8246.
Smythe, J.A., Coppel, R.L., Brown, G.V., Ramasamy, R., Kemp, D.J. and Anders, R.F. (1988) Identification of two integral membrane proteins of *Plasmodium falciparum.*
Proc. Natl. Acad. Sci. **85** : 5195 - 5199
Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis.
J. Mol. Biol. **98** : 503 - 517
Sternberg, N.L. (1990) Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 Kilobase pairs.

Proc. Nat. Acad. Sci. **87** : 103 - 107

Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979) Isolation and characterisation of a yeast chromosomal replicator.
Nature **282** : 39 - 43

Sultan, A.A., Richardson, W.A., Alano, P., Arnot, D.A and Doerig, C. (1994) Cloning and charaterisation of a *Plasmodium falciparum* homologue of the *Ran/TC4* signal transducing GTPase involved in cell cycle control.
Mol. Biochem. Parasitol. **65** : 331 - 338

Tabor, S. and Richardson, C.C. (1989). Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by *in vitro* mutagenesis.
J. Biol. Chem. **264** : 6447 - 6458

Tagle, D.A. and Collins, F.S. (1992) An optimised *Alu*- PCR primerpair for human- specific amlpification of YAC's and somatic cell hybrids.
Hum. Mol. Genet. **1** : 121 - 122

Tanaka, M., Hirai, H., LoVerde, P.T., Nagafuchi, S., Franco, G.R., Simpson, A.J.G. and Pena, S. D. (1995) Yeast artificial chromosome (YAC)- based genome mapping of *Schistosoma mansoni*
Mol. Biochem. Parasitol. **69** : 41 - 51

Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture.
Science **193** : 673 - 675

Traver, C.N., Klapholz, S., Hyman, R.W. and Davis, R.W. (1989) Rapid screening of a human genomic library in yeast artificial chromosomes for singly copy sequences.
Proc. Nat. Acad. Sci. **86** : 5898 - 5902.

Triglia, T., Stahl, H.D., Crewther, P.E., Scanlon, D., Brown, G.V., Anders, R.F. and Kemp, D.J. (1987) Complete sequence of the gene for the knob- associated histidine rich protein from *Plasmodium falciparum.*

Triglia, T., Peterson, M.G. and Kemp, D.J. (1988a) A procedure for the *in vitro* amplification of DNA segments that lie outside the boundaries of known sequence.
Nuc. Acid Res. **16** : 8186

Triglia, T., Stahl, H- D., Crewther, P.E., Silva, A., Anders, R.F. and Kemp, D.J. (1988b) Structure of a *Plasmodium falciparum* gene that encodes a glutamic acid rich protein (GARP)
Mol. Biochem. Parasitol. **31:** 199 - 202

Triglia, T., and Kemp, D.J. (1991) Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes.
Mol. Biochem. Parasitol. **44** : 207 - 212

Triglia, T., Wellems, T.E. and Kemp, D.J. (1992) Towards a high- resolution map of the *Plasmodium falciparum* genome.
Parasitol Today **8** : 225 - 229

Vaidya, A.B., Morrisey, J., Plowe, C.V., Kaslow, D.C. and Wellems, T.E. (1993) Unidirectional dominance of cyutoplasmic inheritance in two genetic crosses of *Plasmodium falciparum.*
Mol. Cell. Biol. **13** : 7349 - 7357

van der Ploeg, L.H.T., Smits, M., Ponnundurai, T., Vermeulen, A., Meuwissen, J.H.E.T. and Langsley, G. (1985) Chromosome- sized DNA molecules of *Plasmodium falciparum.*
Science **229** : 658 - 661

van Ommen, G.J.B., Verkerk, J.M.H., Hofker, M.H., Monaco, A.P., Kunkel, L.M., Ray, P., Worton, R., Wieringa, B., Bakker, E. and Pearson, P.L. (1986) A physical map of 4 million bp around the Duchenne Muscular Dystrophy Gene on the Human X- chromosome.
Cell **47** : 499 - 504

Vaudin, M., Roopra, A., Hillier, L., Brinkman, R., Sulston, J., Wilson, R.K. and Waterston, R.H. (1995) The construction and analysis of M13 libraries prepared from YAC DNA.
Nuc. Acids. Res. **23** : 670 - 674

Vetrie, D., Flinter, F., Bobrow, M. and Harris, A. (1992) Construction of a yeast artificial chromosome contig encompassing the human α5(IV) collagen gene (COL4A5)
Genomics **14** : 634 - 642

Vilageliu, L. and Tyler- Smith, C. (1992) Structural instability of YAC clones and the use of recombination deficient yeast host strains.
in *Techniques for the Analysis of Complex Genomes.* Academic Press, 1992
Vollrath, D. and Davis, R.W. (1987) Resolution of DNA molecules greater than 5 megabases by contour- clamped homogenous electric fields.
Nuc. Acids. Res. **16** : 7865 - 7876
Wada, M., Little, R.D., Abidi, F., Porta, G., Labella, T., Cooper, T., Valle., G.D., D'Urso, M. and Schlessinger, D. (1990) Human Xq24 - Xq28: approaches to mapping with yeast artificial chromosomes.
Am. J. Hum. Genet. **46** : 95 - 106
Wada, M., Abe, K., Okumura, O., Taguchi, H., Kohno, K., Imamoto, F., Sclessinger, D. and Kuwano, M. (1994) Chimeric YAC's were generated at unreduced rates in conditions that suppress coligation.
Nuc. Acids. Res. **22** : 1651 - 1654
Wahl, G.W., Lewis, K.A., Ruiz, J.C., Rothenberg, B., Zhao, J. and Evans, G.A. (1987) Cosmid vectors for rapid genomic walking, restriction mapping and gene transfer.
Proc. Nat. Acad, Sci. **84** : 2160 - 2164
Wahlgren, M., Aslund, L., Franzen, L., Sundvall, M., Wahlin, B., Berzins, K., McNicol, L.A., Bjorkman, A., Wigzell, H., Perlmann, P. and Pettersson, U. (1986) a *Plasmodium falciparum* antigen containing clusters of asparagine residues.
Proc. Natl Acad. Sci. **83**: 2677 - 2681
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.
Mol. Biochem. Para. **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.
Wang, M., Chen, X- N., Shouse, S., Manson, J., Wu, Q., Li, R., Wrestler, J., Noya, D., Sun, Z-G., Korenberg, J. and Lai, E. (1994) Construction and characterisation of a human chromosome 2-specific BAC library.
Genomics **24** : 527 - 534.
Watanabe, J. and Inselburg, J. (1994) Establishing a physical map of chromosome No. 4 of *Plasmodium falciparum.*
Mol. Biochem. Para. **65** : 189 - 200
Waters, A.P., Syin, C. and McCutcheon, T.F. (1989) Developmental regulation of stage specific ribosome populations in *Plasmodium.*
Nature **342** : 438 - 440
Waters, A.P., Higgins, D.G. and McCutcheon, T.F. (1991) *Plasmodium falciparum* appears to have arisen as a result of a lateral transfer between human and avian hosts.
Proc. Natl. Acad. Sci. **88** : 3140 - 3144
Weber, J.L. (1987) Analysis of sequences for the extremely A+T rich genome of *Plasmodium falciparum.*
Gene **52** : 103 - 109
Weber, J.L. (1988a) Molecular biology of malaria parasites.
Ex. Para. **66** : 143 - 170
Weber, J.L. (1988b). Interspersed repetitive DNA from *Plasmodium falciparum.*
Mol. Biochem. Parasitol. **29** : 117 - 124
Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. and Lathrop, M. (1992) A second-generation linkage map of the human genome.
Nature **359** : 794 - 801

Wellems, T.E., Walliker, D., Smith, C., de Rossario, V.E., Maloy, W.L., Howard, R.J., Carter, R. and McCutchan, T.F. (1987) A histidine- rich protein gene marks a linkage group favoured strongly in a genetic cross of *Plasmodium falciparum.*
Cell **49**: 633 - 642
Wellems, T.E., Panton, L.J., Gluzman, I.Y., de 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
Wesseling, J.G., deRee, J.M., Ponnudrai, T., Smits, M.A. and Schoenmakers, J.G.G. (1988) Nucleotide sequence and deduced amino acid sequence of a *Plasmodium falciparum* actin gene.
Mol. Biochem. Parastiol. **27**: 313 - 320
Wesseling, J.G., Snijders, P.J.F., van Someren, P., Jansen, J., Smits, M.A. and Schoenmakers, J.G. (1989) Stage specific expression and genomic organisation of the actin genes of the malaria parasite *Plasmodium falciaprum.*
Mol. Biochem. Parasitol. **35** : 167 - 176
WHO (1992) Tropical Disease Research Eleventh Programme Report, World Health Organisation, Geneva, Switzerland.
WHO (1994) Antimalarial drug policies: data requirements, treatment of uncomplicated malaria and management of malaria in pregnancy, Geneva, Switzerland.
Williamson, D.H., Gardner, M.J., Preiser, P., Moore, D.J., Rangachari, K. and Wilson, R.J.M. (1994) The evolutionary orogin of the 35 Kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophtye ancestry.
Mol. Gen. Genet. **243** : 249 - 252
Wilson, C.M., Serrano, A.E., Wasley, A., Bogenschutz, P., Shankar, A.H. and Wirth, D.F. (1989) Amplification of a gene related to mammalian mdr genes in drug resistant *Plasmodium falciparum.*
Science **248** : 1184 - 1186
Wilson, R.J.M., Gardner, M.J., Williamson, D.H. and Feagin, J.E. (1991) Have malaria parasites three genomes?
Parasitol. Today. **7** : 134 - 136
Winship, P.R. (1989) an improved method for directly sequencing PCR amplified material using dimethyl sulphoxide.
Nuc. Acid. Res. **17** : 1266
WTMGC - Wellcome Trust Malaria Genome Collaboration (1995) The *Plasmodium falciparum* genome project: a resource for researchers.
Parasitol. Today **11** : 1 - 4
Woo, S- S., Jiang, J., Gill, B.S., Paterson, A.H. and Wing, R.A. (1994) Construction and characterisation of a bacterial artificial chromosome library of *Sorghum bicolor*
Nuc. Acids. Res. **22** : 4922 - 4931
Worton, R.G., Sutherland, J., Sylvester, J.E., Williard, H.F., Bodrug, S., Dube, I., Duff, C., Kean, V., Ray, P.N. and Schmickel, R.D. (1988) Human ribosomal RNA genes: orientation of the tandem array and conservation of the 5` end.
science : **239** : 64 - 68
Wu, Y., Sifri, D., Lei, H- H., Su, X- Z. and Wellems, T.E. (1995) Transfection of *Plasmodium falciparum* within human red blood cells
Proc. Natl. Acad. Sci. **92**: 973 - 977
Young, R.A. and Davis, R.W. (1983) Yeast RNA plymerase II genes : isolation with antibody probes.
Science **222** : 778 - 782.
Zolg, J.W., Macleod, A.J., Dickson, I.H. and Scaife, J.G. (1982) *Plasmodium falciparum*: modifications of the in vitro culture conditions improving parasitic yields.
J. Parasitol. **68** : 1072 - 1080
Zuo, J., Robbins, C., Taillon- Miller, P., Cox, D.R. and Myers, R.M. (1992) Cloning of the Huntingdon disease region in yeast artificial chromosomes.
Hum. Mol. Genet. **1** : 149 - 159

" This is The End.
My only friend....
The End."

Jim Morrison.