Genome analysis and DNA marker-based characterisation of pathogenic trypanosomes

DOCTORAL THESIS

Utrecht University

The Netherlands

Edwin Chukwura Agbo

2003

The printing of this thesis was financially supported by:

ID-Lelystad B.V.
Utrecht University
Novartis
Bayer B.V.
Pfizer Animal Health B.V.
V.M.D. Laboratories Belgium
Biomedical Primate Research Centre, Rijswijk
Alfasan Nederland B.V.

To my family

Genome analysis and DNA marker-based characterisation of pathogenic trypanosomes

Genoom analyse en DNA merker-gebaseerde karakterisatie van pathogene trypanosoma

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. W. H. Gispen,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op
27 maart 2003 des middags te 14.30 uur

door

Edwin Chukwura Agbo

geboren 21 december 1966 te Mbu-Amon, Nigeria

Promoters: Prof. dr. E. Claassen
Department of Immunology
Erasmus University Rotterdam

Prof. dr. J. P. M. van Putten
Department of Infectious Diseases and Immunology
Utrecht University

Prof. dr. P. Büscher
Department of Parasitology
Institute of Tropical Medicine, Antwerp

Co-promoter: Dr. M. F. W. te Pas
Division of Animal Sciences
Institute for Animal Science and Health, Lelystad

The research in this thesis was performed at the Institute for Animal Science and Health (ID-Lelystad), Lelystad, The Netherlands, and at the Department of Infectious Diseases and Immunology of Faculty of Veterinary Medicine, Utrecht University, The Netherlands.

ISBN 90-393-3306-8

Cover design by Fred van Welie
Layout by Diny Heuckeroth
Printed by Universal Press, Veenendaal

CONTENTS

PREFACE

In this thesis is described a molecular approach to diagnosis and control of pathogenic trypanosomes. Recent developments in genome-wide molecular techniques are providing useful advances for fine-scale genotyping in various taxa. An attempt has been made to evaluate existing molecular tools and to address the need for multi-locus fine-scale genotyping approach for molecular marker-assisted analysis. This tool has been used to uncover cryptic genetic variation of trypanosome strains, to infer phylogenetic relationships and to characterise specific trypanosome phenotypic traits, including the role of differential DNA methylation in modulating human serum response properties of trypanosomes

Objectives and outline of the thesis

Three severe epidemics of African trypanosomosis over the last century jolted scientists and clinicians into (i) active screening (using epidemiological and microbiological data) to identify cases, (ii) surveillance to monitor disease prevalence and, (iii) the application of effective treatments against clinical cases of trypanosomosis. With the success of these measures, the disease had practically disappeared between 1960 and 1965. However, it reappeared in endemic form interspersed by periods of epidemic episodes in several foci following a relaxation of the control measures over the last thirty years. Unfortunately, appropriate effective therapies are now severely limited due mainly to the increasing problem of parasite resistance to available drugs, in addition to high frequency of severe adverse effects of the drugs, and the long duration and high cost of treatments. Furthermore, there is the need for a robust fine-scale diagnostic system for epidemiological follow-up. It is important to identify human-infective isolates in the field to determine the relevance of animal reservoirs and the prevalence of human-infective trypanosomes in tsetse flies.
Therefore, it is evident that controlling trypanosomosis requires innovative diagnostic assay development strategies, as well as new approaches and

targets for developing effective drugs and a possible vaccine. Whole-genome methods are changing the scope of genetic and genomic analysis of *T. brucei* towards the discovery of these lead molecules. A careful analysis of the genetic make-up based on signature profiling of individuals or populations may provide such powerful tools.

This thesis describes the results of genome analyses of pathogenic trypanosomes, with special emphasis on *T. brucei*. The aims were to develop and evaluate molecular marker systems for generating sensitive tools for molecular typing of strains, taxonomy and analysis of the relationships between genetic variations and their consequent functional effects, to enhance our understanding of important traits. The data presented herein have added to knowledge on such parasite parameters as allelic genetic differences, population structure, and molecular basis of human serum resistance/sensitivity of pathogenic trypanosomes.

CHAPTER 1

Trypanosome genomics, the challenge of identifying vaccine targets and the prospects for control of trypanosomosis: a Review

Eddy C. Agbo [1], Phelix A.O. Majiwa [2], Philippe Büscher [3],
Eric Claassen [4] and Marinus F.W. te Pas [1]

[1] Division of Animal Sciences, Institute for Animal Science and Health
(ID-Lelystad), Edelhertweg 15, P.O. Box 65,
8200 AB Lelystad, The Netherlands
[2] International Livestock Research Institute,
P.O. Box 30709, Nairobi, Kenya
[3] Institute of Tropical Medicine, Department of Parasitology,
Nationalestraat 155, B-2000 Antwerp, Belgium
[4] Department of Immunology, Erasmus University Rotterdam,
Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Submitted for publication

1.1. Summary

There have been tremendous advances in our knowledge of trypanosome pathobiology yet many aspects remain unclear. Currently, the genome of *Trypanosome brucei* is being sequenced and, with other genome-wide analysis methods including microarrays, may provide novel insights for better understanding of the parasite, facilitating the development of effective controls. An important new challenge investigators face is how to utilise these tools to study a parasite that possesses many genetic peculiarities. This review summarises our current understanding of African trypanosome genetics and links the impact of genome analysis to the prospects for developing possible vaccine targets and new tools for rationale drug design and specific diagnostics.

1.2. Introduction

The tsetse fly-transmitted trypanosomes are unicellular protozoan parasites that cause acute to chronic debilitating disease called sleeping sickness in humans and nagana in cattle. Molecular epidemiology and genetic analysis have confirmed the classification of *Trypanosoma brucei* into human-infective subspecies - *T. b. gambiense* and *T. b. rhodesiense*, and *T. b. brucei* that is non-infective to humans because the subspecies is lysed by normal human serum. There are about 500,000 cases of human disease in sub-saharan Africa (WHO, 2001), while the animal disease contributes to severely limit livestock production in a third of the continent.

The fight against African trypanosomosis has relied heavily on chemotherapy and vector control strategies, and although therapy and prophylaxis have been most widely applied, their effectiveness remain unsatisfactory (reviewed by Keiser et al., 2001; Anene et al., 2001). Initial anti-trypanosomal drugs were dominated by salts of three compounds: diamine, phenanthridine and phenanthridine-aromatic amide; all were beset by their inability to cross the blood-brain barrier in sufficient quantity to

prevent relapses of late-stage cases of sleeping sickness in humans. This situation was markedly improved following the introduction of Melarsoprol but severe adverse effects frequently occur (Kuzoe, 1993; WHO, 1998). No less remarkable, however, is the versatility shown by the parasites in their ability to develop refractoriness to the treatment of the disease with these drugs. In the past 35 years, cymelarsan, a trivalent water-soluble analogue of the arsenical melarsoprol, and D, L-α-difluoromethyl-ornithine (DFMO) (Eflornithine®) are the only new trypanocides commercially made available for veterinary and human use. Unfortunately, they are also toxic to humans, and trypanosomes are showing increasing resistance to these drugs. Moreover, DFMO, which is the newest trypanocide, appears to be ineffective against *T. b. rhodesiense* and thus can be used in humans only for treatment of the disease caused by *T. b. gambiense* (Bacchi et al., 1980). Even then, the drug is less efficient in children than in adults (Milford et al., 1993). Short-term measures such as chemical modification of existing compounds (Werbovetz et al., 1996) or combination therapy (Jennings, 1993) could have significant impact on disease control in the immediate future.

The need for additional targets and new, less-toxic therapeutics is widely acknowledged (WHO, 2001). There is consensus that vaccination will provide the most effective means for the control of trypanosomosis, but no vaccine is available. In addition, sensitive molecular tools are also required for fine-scale identification and more detailed characterisation of all *T. brucei* subspecies, and for epidemiological investigations of African trypanosomosis. Such specific diagnostic tool will also provide necessary adjunct to the effective use of any new drug or vaccine products. Recent developments towards developing specific markers through fine-scale trypanosome genome analysis (Agbo et al., 2001; Agbo et al., 2002; Agbo et al., In press) indicate that specific molecular diagnosis of *T. brucei* subspecies is feasible. There are high expectations that genomics will offer new leads to rationale vaccine design and accelerate the discovery of novel drug targets, as well as the development of practical and robust diagnostic

assays. This review describes the latest genomics-based technologies and discusses their potential contribution to the control of trypanosomosis caused by *T. brucei*.

1.3. Genomics and target discovery for vaccine and drug development

1.3.1 Trypanosome genetics

Traditionally, the search for novel genes required for survival or virulence of any organism was based on several genetic methods involving genetic crosses, gene knockouts and induced mutations of the genome, followed by screening for the relevant phenotype. In trypanosomes, these approaches have been hampered by the inability to generate sufficient number of progeny clones from genetic crosses and the lack of a fine genetic map. These have affected the ability to determine precisely statistical significance of segregating phenotypes in relation to the number of loci or alleles determining the trait of interest (Gibson and Stevens, 1999; MacLeod et al., 1999; Tait et al., 2002).

1.3.2. Trypanosome genome sequencing

There is now a significant investment in sequencing the genome of *T. brucei* (http://www.tigr.org/tdb/e2k1/tba1/;http://www.sanger.ac.uk/ Projects/ T_brucei). According to these databases, at the current rate of progress, it is estimated that the full genomic sequence of this parasite will be available during the early part of 2004 at the latest. The challenge then will be to convert this sequence information to the functions of the full gene complement of this organism and to assess their potential value as targets for integrated control tools. Already, the extensive genome sequence information available in public databases is such that any known *T. brucei* gene can be identified (at >90% chance) by searching these databases, using a variety of criteria built into the search tools.

Two complementary approaches are being followed in sequencing the trypanosome genome: genomic DNA sequencing, and EST (expressed sequence tag, or expressed gene) sequencing. Trypanosome genomic sequence data reveal that genes are tightly packed on chromosomes (El-Sayed et al., 2000) and appear to be unidirectional over long regions, which could permit better design of future experimental research. Initial data emerging from the *T. brucei* 'shot-gun' genome-sequencing project indicate that the overall sequence organisation of chromosomes is distinct from that of other higher eukaryotic organisms. There are, however, a number of similarities with those of the other lower eukaryotes, particularly the protozoan parasites (Lanzer et al., 1993; Ravel et al., 1996). Several genes encoding housekeeping proteins, enzymes and structural proteins are arranged in tandem repeats, making the sequencing difficult because of ambiguity in distinguishing big DNA clones containing repetitive DNA clusters that may have originated from different chromosomes. In addition, size differences have been observed between chromosome homologues within a single strain of *T. brucei*, which are primarily due to differences in the number of repeat units contained within the repetitive elements on the different chromosome homologues, particularly those found at the telomeric ends. Given that only a few specific genes are known in *T. brucei*, it can be speculated that many targets will remain undiscovered among the genes of unknown function, underscoring the importance of starting functional characterisation of the genome of this parasite, using the available sequence data as the starting point.

One of the most intriguing observations from a majority of genome sequencing projects completed so far is that a significant proportion (at least 40%) of genes identified have unknown functions. In other words, these genes have been identified solely by sequence determination combined with the fact that they either are homologous to other genes identified in other organisms, but whose functions remain unknown or have open reading frames which potentially encode functional proteins. Such genes will need to be characterised by genetic or biochemical analysis.

Expressed sequence tag (EST) analysis provides a rapid means of gene discovery and comparative data on expression levels. An example of the power of EST sequencing is provided by the efforts to sequence approximately 4500 randomly selected cDNAs from *T. brucei*, a process that successfully identified more than 1000 unique gene transcripts (El-Sayed et al., 1995; El-Sayed and Donelson 1997; Djikeng et al., 1998). Prior to this effort, there were only a handful of *T. brucei* genes in public databases. Thus, EST sequencing rapidly identified a large number of candidate genes for further investigations. About 10-15% of these ESTs are similar to a known gene of at least one other organism, and about 50% are novel (El-Sayed et al., 2000; see: www.ncbi.nlm.gov/dbEST/index.html). In addition, EST analysis offers the possibility of identifying the complete open reading frame (ORF) by re-constructing an overlapping consensus sequence from multiple ESTs. Fully sequenced ESTs can also be used to identify alternatively spliced messages and gene families consisting of related genes. In addition, an extensive comparative EST analysis of *T. brucei* and several other protozoan parasites might reveal interesting evolutionary or functional trends among the taxa.

While EST sequencing provides a rapid means of gene discovery and a range of markers for gene-based sequence tagged site (STS) mapping, it is rather biased in that the ESTs originate from cDNAs synthesised from mRNA present in a synchronous population of trypanosomes of a particular life-cycle stage. A majority of the ESTs was generated from the bloodstream forms of the parasite, a very small number from procyclic culture forms and none from the infective metacyclic forms (El-Sayed et al., 1995; El-Sayed and Donelson 1997; Djikeng et al., 1998). This urges for an investigation of the EST content of either a mixture of trypanosomes of all life-cycle stages, or ideally, but more labour intensive, all life-cycle stages individually. Considering the size of *T. brucei* genome, it can be estimated that a substantial number of gene transcripts are not represented among the ESTs currently available in the public domain databases. These represent approximately 1000 unique gene transcripts, from the 6000

genes indirectly estimated to be present in *T. brucei* (Djikeng et al., 1998). However, even after all the possible ESTs of *T. brucei* would have been described, there will still be a significant proportion that encodes proteins of either unknown nature or function. Among these are likely to be the genes that are specific to this parasite because they are essential to its parasitic lifestyle. These genes or their products might be the best candidates for drug and/or vaccine design precisely because they are likely to be highly directed to phenotypic or disease processes unique to the parasite, but would have to be highly expressible and/or accessible to the immune system. Furthermore, since these genes probably have no human counterpart these vaccines and drugs may be highly specific for the parasites expressing less or no adverse effects to humans.

1.3.3. Trypanosoma genome variability

Variations at multiple genetic loci have been analysed for novel allelic differences among *T. brucei* subspecies using restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP) and minisatellite marker analysis system (Agbo et al., 2002; MacLeod et al., 2000; Hide et al., 1994). The AFLP technique has also been applied to identify a number of polymorphic markers among parental lines of *T. brucei* and for the construction of a genetic map (Tait et al., 2002; see: http://www.gla.ac.uk/ibls/II/cmrt). The genetic analysis of significant trait(s) requires that linkage analysis be performed using molecular markers (in a dense genetic map) that co-segregate with the traits in the progeny of a cross. Such markers might be relevant for the identification of targets for the development of drugs and/or vaccines.

To date, there has been a paucity of published data on the analysis of the inheritance of specific trypanosome traits, except a preliminary analysis of resistance to lysis by human serum (Gibson and Mizen, 1997). It is now known that human serum resistance of *T. b. gambiense* and some *T. b. rhodesiense* strains involve a complex of genes, possibly from multiple loci and alleles. This could confound classical genetic approaches to the

analyses and urges for global genome-wide investigations to identify the genes involved. Also, the functional complexity of trypanosomes far exceeds that which may be indicated by the genome data of the single clone currently being sequenced, mainly due to their immense genome plasticity and alternative splicing that result in extensive polycistronic pre-mRNAs. Therefore, forging a link among the genomic sequences of the three trypanosomatids - *T. brucei*, *T. cruzi* and *Leishmania spp* - would broaden the list of target candidates that can then be rapidly validated by other means. The gene categories generated by this approach enable such a pre-selection of target candidates on a whole-genome scale; in other words, targets can be defined according to the characteristics of specific phenotypic trait(s) for possible broad-spectrum applications. Comparative analysis of genomes of such closely-related organisms may generate testable hypothesis of genes that might be responsible for differences in tropism and pathologies between parasite strains or species. Genomic regions that are different among *T. brucei* variants, for example, can be rapidly identified and evaluated for relevance for the development of drugs and/or vaccines. Such comparative studies are already gaining appeal as templates in homology modelling and molecular dynamics study, to elucidate the functional importance and therapeutic potential of target enzymes in trypanosomes (Hide et al., 1990; Stevens and Tibayrenc, 1995; De Greef et al., 1989).

1.4. Genomics and diagnostic assay development

Success of vaccination may depend upon highly reliable diagnosis of parasites, which may be difficult, especially in case of closely-related subspecies. However, genomics will also unravel the genome differences underlying subspecies differences that can be used for proper diagnosis. The development of subspecies-specific markers for molecular diagnosis will be discussed first.

1.4.1. Genetic markers and molecular epidemiology

Single nucleotide polymorphisms (SNPs) are densely spaced polymorphism throughout the genome (see http://snp.cshl.org), with efficient analysis platforms, which promise to allow disease association studies even without carefully defined family or pedigree studies. The SNPs are expected to be the mainstay of efforts not only in mapping of disease associations but also in attempting to identify the genetic determinants of the different reactions to different drugs or compounds. Many methods used in the recent past employed a subset of SNPs, namely those detected by the appearance or disappearance of restriction sites.

Restriction endonuclease-based genome analysis methods have been used to search for markers that may be used to simultaneously differentiate *T. brucei* subspecies (Agbo et al., 2001; Agbo et al., 2002; Hide et al., 1994; Hide et al., 1990; Stevens and Tibayrenc, 1995; Agbo et al., In press). Single nucleotide polymorphisms (SNPs) that are identified as candidate markers can be cloned and converted into PCR-based sequence-tagged sites, and the diagnostic value of each one can subsequently be evaluated in larger population studies, to validate an association of gene with specific parasite traits. The recent development of a finer-scale multi-endonuclease-based genotyping approach, which simultaneously accesses multiple independent allelic sites within the genome, offers a tool for expanding the number of identifiable single nucleotide and restriction fragment polymorphisms (Agbo et al., In press). These tools will also increase the chance of detection of subtle variations in restriction patterns, and help to establish the parameters of any future large-scale efforts to accumulate SNPs in *T. brucei*. However, bioinformatics will provide a key link to the underlying genes and the interpretation of any associations these may have to relevant phenotypes.

One approach to the development of specific assays for *T. brucei* subspecies is the identification of genetic markers linked to traits specific to the respective subspecies. These could correspond to genes that are

specific to a subspecies, and could thus be used for identification on the basis of the trait encoded by the gene. An approach that exploited differential screening of cDNA libraries using radioactive cDNA from human serum-sensitive or resistant clones of *T. b. rhodesiense* led to the identification of the serum-resistant-associated (SRA) gene transcript (De Greef et al., 1989). The gene appears to be present only in most *T. brucei rhodesiense* strains and undetectable in the serum-sensitive *T. brucei brucei* (Rifkin et al., 1994). A recently developed PCR assay based on the SRA gene appears to be specific for the identification of human-infective *T. b. rhodesiense* and could serve as a useful tool for studies on molecular epidemiology of the disease caused by this subspecies (Welburn et al., 2001; Gibson et al., 2002). The value of the SRA gene for the study of human serum resistance will be enhanced by the identification of specific ligands in normal human serum with which the protein product of this gene interacts to cause the effect, but no signal peptide for the gene has been identified. Based on such ligands, *in vitro* assays can be designed whereby compound libraries are screened for small molecules that competitively displace the putative ligand. This offers a basis for developing *T. b. rhodesiense*-specific drug and/or vaccine targets.

1.5. Genomics and identification of drug and vaccine targets

1.5.1. Defining gene function

With the genome sequencing initiatives providing putative targets in profusion, some notion of function beyond what can be discerned from homology is crucial to the decision to continue with a putative target. Knowing when and where a gene is expressed can be an important input to this process. It is apparent from the genomes whose sequences have been completed that it is usually possible to assign functions to, at best, 60% of the genes identified in a particular organism. In addition to categorising genes solely by sequence-based comparisons, selection of putative targets can be more precisely addressed through direct studies of gene expression, at either the mRNA or protein level. Trypanosomes rely heavily

on gene regulation at the post-transcriptional or post-translational levels. Thus, the level of complexity resulting from co- or post-translational modification events can only be dissected and understood through qualitative and quantitative studies of gene expression at the level of the functional proteins themselves. This urges for proteomics investigations of the parasite genome, a subject that is outside the scope of this review.

Recent studies have begun to generalise the notion of an intimate correlation between RNA-triggered silencing and gene activity (Ngo et al., 1998; Wirtz et al., 1999; LaCount et al., 2000; Wang et al., 2000; Drozdz et al., 2002). Due to its hostile environment with dangerous opportunities for unwanted gene expression, trypanosomes have a vehement response to dsRNA (Ngo et al., 1998; Drozdz et al., 2002). This may pose an immense challenge to putative drug delivery machineries. Recently, Ullu et al. (2002) discussed how this association has begun to answer a number of questions relating to mechanisms of gene silencing in trypanosomes. Procedures based on silencing of gene expression could be a valuable tool for functional genomics studies in trypanosomes, providing valuable information about gene function even in cases where only a partial loss-of-function is generated. This might allow the identification of candidate targets for drug development and vaccine design. On the other hand, the assessment of differential gene expression based on DNA methylation has so far not been explored in *T. brucei*. Trypanosome genome data so far indicate that the genome is composed of relatively small number of genes and a large fraction of repetitive DNA. From detailed knowledge of methylation processes in other taxa (reviewed by Martienssen and Richards, 1995; Martienssen, 1998), it is tempting to speculate that most of these repetitive sequences are more heavily methylated relative to functional genes. It is therefore, foreseeable that *T. brucei* genes unique to a subspecies or trait can be enriched-for by the exclusion of repeat sequences in genomic shotgun libraries.

1.5.2. Microarrays (transcriptome profiling)

The strength of microarray technologies lies in their ability to follow genome-wide gene expression patterns to identify several hundreds of genes that may be involved in specific phenotypic traits in an organism (Schena et al., 1995). One important factor in determining the value of microarray is the selection of DNA hybridisation targets. In a pilot study, inserts from a *T. brucei* genomic library were arrayed randomly to generate 'shotgun' microarrays (El-Sayed and Donelson, 1997). To measure variation in expression of genes in the bloodstream and procylic (insect vector) stages of the parasite, the arrays were probed with differentially labelled cDNAs prepared from total RNA isolated from parasites at defined developmental stages. While many clones displayed constitutive levels of expression in both stages of the parasite, a substantial number of test clones showed differential gene expression.

A major limitation of transcriptome analysis in trypanosomes is, in addition to those associated with this approach in general, the level of gene regulation at the mRNA level. If what is known about resistance genes in other taxa is anything to go by, then transcriptome profiling used to isolate more highly expressed genes on the basis of their gene product or patterns of expression would not be useful for isolating trypanosome resistance genes, e.g. human serum or drug resistance genes. Almost all the resistance genes isolated so far from eukaryotes are expressed at low levels and are not greatly induced upon infection. Furthermore, in trypanosomes, extensive co- and post-transcriptional modification events occur, e.g. due to alternative splicing, which result in a diversity of protein products from a single open reading frame. In addition, telomeric sequences tend to be under-represented in standard genomic libraries, a situation that would, for example, limit the value of DNA microarrays for the study of antigenic variation in these parasites. Therefore, a concerted effort to clone telomeric sequences via other means has been proposed (Rudenko, 2000).

Following the identification of several potentially useful targets for rationale drug or vaccine design, their downstream development into 'leads' could be greatly facilitated if efficient technologies for mode-of-action studies were available. For example, exploring drug-induced alterations in gene expression by microarray may provide insight into the induction of genes known to encode pathways affected by a potential drug target. The results should, in principle, provide possible correlation between changes in gene expression and the mode of action of the drug. However, the uses of microarray analysis need not be limited to gene expression. For example, oligonucleotide arrays can be used for the rapid screening of SNPs and thus for exploring subspecies variation.

1.6. Vaccine targets

The development of a broad-spectrum vaccine against trypanosomes meets formidable obstacles primarily because of the biology of this parasite and immune responses exhibited by the hosts in the course of infection. Immune responses to the VSG (the most abundant surface antigen of the parasite) are not protective because of the vast repertoire of immunologically distinct variants. Furthermore, outcome of the disease in an individual host is dependent on its genetic background and that of parasite strains responsible for the infection. Although the possibility of anti-disease vaccine for trypanosomosis has been suggested (Authie, 1994), it can be inferred that developing a protective broad-ranging anti-trypanosome vaccine using conventional vaccine systems currently seems unfeasible. The unique aspects of the biochemistry of trypanosomes make rationale vaccine design an attractive approach, but targets must be selected carefully.

Trypanosome genomics has had a major impact on identification of potentially useful targets, including some enzymes, transporters and metabolites that may be unique to the parasites (reviewed by Keiser et al., 2001; Barrett et al., 1999), and their potential for disease prevention or

therapeutics is being pursued. For example, reverse genetics tool based on RNAi has been used to study the possible therapeutic value of the conserved glycolipid (glycosylphosphatidylinositol, GPI) anchor, which attaches the VSGs to the cell membrane (Bohme and Cross, 2001). It turns out that the high conservation of the VSG GPI signal sequence is not necessary for efficient expression and GPI attachment, suggesting that the GPI anchor alone might not be a high-value vaccine target. Similarly, many surface molecules, including some that might be expected to be major structural components of the coat, can be mutated or completely eliminated without altering the morphology or growth rate of trypanosomes in culture, reviewed by Landfear and Ignatushchenko, 2001). Although this may suggest that the GPI molecules might not be the first choice for vaccine development, they may offer unique genetic options for identifying genes involved in the GPI pathway. Other key invariant surface glycoproteins, transporters and receptor complexes have been identified (Seyfang and Duszenko, 1993; Borst et al., 1998; Pays and Nolan, 1998), but only a few have been tested. For example, the VSG molecules themselves contain structurally conserved C-termini (Carrington et al., 1991; Carrington and Boothroyd, 1996), but these are so few and are generally so small as to be immunologically insignificant. Also, the flagellar pocket, a membrane invagination around the base of the flagellum harbours a number of invariant receptor molecules that could be targets for vaccine design (da Cunha et al., 1989; Engman et al., 1989; Paindovoine et al., 1992; Nozaki et al., 1996). Although some of these molecules are distributed over the entire surface of the cell, they are inaccessible to the host immune response. On the other hand, analogues to myristate, the only fatty acid component of the glycolipid anchor, appeared to have trypanocidal activity *in vitro* (Doering et al., 1994), but bioavailability of active compounds in the cell will need to be enhanced for sustained activity *in vivo*.

A pragmatic model would be to ‘traffic’ toxic vaccine molecules into these invariant target ligands. For example, RNA aptamers have been exclusively expressed in the flagellar pocket (Homann and Goringer, 2001), thus

making the approach attractive for rationale vaccine design. Another possibility could be an agent interacting with the 35 nucleotide sequences at the extreme 5' termini of all mRNAs in trypanosomes but not in mammalian host mRNA. From these studies, knowledge has also been gained on how inhibitory a compound needs to be in order to be an effective trypanocide. The ability to generate such genetically manipulated parasites can also provide insight into whether specific proteins are targets of existing drugs, or effective ligands for a possible vaccine.

1.7. Conclusion

There is an enormity of problems to consider and large hurdles to cross in the development tools for a trypanosome vaccine or new therapeutics. These include the enormity of different possible antigenic specificity's to cover the requirement for durable specific immune response and the immense cost and long-term requirement for development of most vaccines. It is our opinion that the next phase of genomic technologies will accelerate the pace at which lead targets can be identified and validated.

References

Agbo, E. C., B. Duim, P. A. O. Majiwa, P. Büscher, E. Claassen, and M. F. W. te Pas. Multiplex-endonuclease genotyping approach (MEGA): a tool for the fine-scale detection of unlinked polymorphic DNA markers. Chromosoma, In press.

Agbo, E. C., P. A. O. Majiwa, E. Claassen, and M. H. Roos. 2001. Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterisation. Exp. Parasitol. **99:**123-131.

Agbo, E. C., P. A. O. Majiwa, H. J. H. M. Claassen, and M. F. W. te Pas. 2002. Molecular variation of *Trypanosoma brucei* subspecies as revealed by AFLP fingerprinting. Parasitology **124:**349-358.

Anene, B. M., D. N. Onah, and Y. Nawa. 2001. Drug resistance in pathogenic African trypanosomes: what hope for the future? Vet. Parasitol. **96:**83-100.

Authie, E. 1994. Trypanosomiasis and trypanotolerance in cattle: a role for congopain? Parasitol. Today **10:**360-364.

Bacchi C. J., H. C. Nathan, S. H. Hunter, P. P. McCann, and A. Sjoerma. 1980. Polyamine metabolism: a potential therapeutic target in trypanosomes. Science **210:**332-224.

Barrett, M. P., J. C. Mottram, and G. H. Coombs. 1999. Recent advances in identifying and validating drug targets in trypanosomes and leishmanias. Trends Microbiol. **7:**82-88.

Bohme, U., and G. A. M. Cross. 2001. Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. J. Cell Sci. **115:**805-816.

Borst, P., W. Bitter, P. A. Blundell, I. Chaves, M. Cross, H. Gerrits, F. van Leeuwen, R. McCulloch, M. Taylor, and G. Rudenko. 1998. Control of VSG gene expression sites in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **91:**67-76.

Carrington, M., and J. Boothroyd. 1996. Implications of conserved structural motifs in disparate trypanosome surface proteins. Mol. Biochem. Parasitol. **81:**119-126.

Carrington, M., N. Miller, M. Blum, I. Roditi, D. Wiley, and M Turner. 1991. Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. J. Mol. Biol. **221:**823-835.

Da Cunha e Silva, N. L., A. Hasson-Voloch, and W. de Souza. 1998. Isolation and characterisation of a highly purified flagellar membrane fraction from trypanosomatids. Mol. Biochem. Parasitol. **37:**129-136.

De Greef, C., H. Imbrechts, G. Matthyssens, N. Van Meirvenne, and R. Hamers. 1989. A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol. **36:**169-176.

Djikeng, A., C. Agufa, J. E. Donelson, and P. A. Majiwa. 1998. Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. Gene **221:**93-106.

Doering, T. L., T. Lu, K. A. Werbovetz, G. W. Gokel, G. W. Hart, J. I. Gordon, and P. T. Englund. 1994. Toxicity of myristic acid analogs toward African trypanosomes. Proc. Natl. Acad. Sci. USA **91:**9735-9739.

Drozdz, M., I. Quijada, and C. E. Clayton. 2002. RNA interference in trypanosomes transfected with sense and antisense plasmids. Mol. Biochem. Parasitol. **121:**149-152.

El-Sayed, N. M., and J. E. Donelson. 1997. A survey of the *Trypanosoma brucei rhodesiense* genome using shotgun sequencing. Mol. Biochem. Parasitol. **84:**167-178.

El-Sayed, N. M., C. M. Alarcon, J. C. Beck, V. C. Sheffield, and J. E. Donelson. 1995. cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. Mol. Biochem. Parasitol. **73:**75-90.

El-Sayed, N. M., P. Hegde, J. Quackenbush, S. E. Melville, and J. E. Donelson. 2000. The African trypanosome genome. Int. J. Parasitol. **30:**329-345.

Engman, D.M., K. H. Krause, J. H. Blumin, K. S. Kim, L. V. Kirchhoff, and J. E. Donelson. 1989. A novel flagellar Ca2+-binding protein in trypanosomes. J. Biol. Chem. **264:**18627-18631.

Gibson, W. C., and J. Stevens. 1999. Genetic exchange in the *Trypanosomatidae*. Adv. Parasitol. **43:**1-45.

Gibson, W. C., and W. H. Mizen. 1997. Heritability of the trait for human infectivity in genetic crosses of *T. brucei* spp. Trans. R. Soc. Trop. Med. Hyg. **91:**236-237.

Gibson, W., T. Blackhouse, and A. Griffiths. 2002. The human resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infec. Genet. Evol. **1:**207-214.

Hide, G., P. Cattand, D. LeRay, J. D. Barry, and A. Tait. 1990. The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. Mol. Biochem. Parasitol. **39:**213-225.

Hide, G., S. C. Welburn, A. Tait, and I. Maudlin. 1994. Epidemiological relationships of *Trypanosoma brucei* stocks from South-east Uganda: evidence for different populations structures in human infective and non-human infective isolates. Parasitology **109:**95-111.

Homnann, M., and H. U. Goringer. 2001. Uptake and intracellular transport of RNA aptamers in African trypanosomes suggest therapeutic "piggy-back" approach. Bioorg. Med. Chem. **9:**2571-2580.

Jennings FW. 1993. Combination chemotherapy of CNS trypanosomiasis. Acta Tropica **54:**205-213.

Keiser, J., A. Stich, and C. Burri. 2001. New drugs for the treatment of human African trypanosomiasis**:** research and development. Trends Parasitol. **17:**42-49.

Kuzoe, F. A. S. 1993. Current situation of African trypanosomiasis. Acta Tropica. **54:**153-162.

LaCount, D. J., S. Bruse, H. K. Hill, and J. E. Donelson. 2000. Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. Mol. Biochem. Parasitol. **111:**67-76.

Landfear, S. M., and I. Ignatushchenko. 2001. The flagellum and flagellar pocket of trypanosomatids. Mol. Biochem. Parasitol. **115:**1-17.

Lanzer, M., D. de Bruin, and J. V. Ravetch. 1993. Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome. Nature **371:**654-657.

MacLeod, A., A. Tweede, S.C. Welburn, I. Maudlin, C. M. R. Turner, and A. Tait. 2000. Minisatellite marker analysis of *Trypanosoma brucei*: a reconciliation of clonal, panmictic and epidemic population genetic structures. Proc. Natl. Acad. Sci. USA **97:**13442-13447.

MacLeod, A., C. M. R. Turner, and A. Tait. 1999. A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. Mol. Biochem. Parasitol. **102:**237-248.

Martienssen, R. 1998. Transposons, DNA methylation and gene control. Trends Genet. **14:**263-264.

Martienssen, R. A., and E. J. Richards. 1995. DNA methylation in eukaryotes. Curr. Opin. Genet. Dev. **5:**234-242.

Milord, F., L. Loko, L. Ethier, B. Mpia, and J. Pepin. 1993. Eflornithine concentrations in serum and cerebrospinal fluid of 63 patients treated for *Trypanosoma brucei gambiense* sleeping sickness. Trans. R. Soc. Trop. Med. Hyg. **87:**473-477.

Ngo, H., C. Tschudi, K. Gull, and E. Ullu. 1998. Double-stranded RNA induced mRNA degradation in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA **95:**14687-14692.

Nozaki, T., P. A. Haynes, and G. A. Cross. 1996. Characterisation of the *Trypanosoma brucei* homologue of a *Trypanosoma cruzi* flagellum-adhesion glycoprotein. Mol. Biochem. Parasitol. **82:**245-255.

Paindavoine, P., S. Rolin, S. Van Assel, M. Geuskens, J. C. Jauniaux, C. Dinsart, G. Huet, and E. Pays. 1992. A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. Mol. Cell Biol. **12:**1218-1225.

Pays, E., and D. P. Nolan. 1998. Expression and function of surface proteins in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **91:**3-36.

Ravel, C., P. Wincker, C. Blaineau, C. Britto, P. Bastien, and M. Pages. 1996. Medium-range restriction maps of five chromosomes of *Leishmania infantum* and localisation of size-variable regions. Genomics **35:**509-516.

Rifkin, M. R., C. De Greef, A. Jiwa, F. R. Landsberger, and S. Z. Shapiro. 1994. Human serum-sensitive *Trypanosoma brucei rhodesiense*: a comparison with serologically identical human serum-resistant clones. Mol. Biochem. Parasitol. **66:**211-220.

Rudenko, G. 2000. The polymorphic telomeres of the African trypanosome *Trypanosoma brucei*. Biochem. Soc. Trans. **28:**536-540.

Schena, M., D. Shalon, R. W. Davis, and P. O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:**467-470.

Seyfang, A., and M. Duszenko. 1993. Functional reconstitution of the *Trypanosoma brucei* plasma-membrane D-glucose transporter. Eur. J. Biochem. **214:**593-597.

Stevens, J. R., and M. Tibayrenc. 1995. Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterised by RAPD analysis and isoenzymes. Parasitology **110:**181-186.

Tait, A, D. Masiga, J. Ouma, A. MacLeod, J. Asse, S. Melville, G. Lindegard, A. McIntosh, and M. Turner. 2002. Genetic analysis of phenotype in *Trypanosoma brucei***:** a classical approach to potentially complex traits. Philos. Trans. R. Soc. Lond. B. **357:**89-99.

Ullu, E., A. Djikeng, H. Shi, and C. Tschudi. 2002. RNA interference**:** advances and questions. Philos. Trans. R. Soc. Lond. B. **357:**65-70.

Wang, Z., J. C. Morris, M. E. Drew, P. T. Englund. 2000. Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. J. Biol. Chem. **275:**40174-40179.

Welburn, S. C., K. Picozzi, E. M. Fevre, P. G. Coleman, M. Odiit, M. Carrington, and I. Maudlin. 2001. Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. Lancet **358:**2017-2019.

Werbovetz, K. A., C. J. Bacchi, P. T. Englund. 1996. Trypanocidal analogs of myristate and myristoyllysophosphatidyl-choline. Mol. Biochem. Parasitol. **81:**115-118.

Wirtz, E., S. Leal, C. Ochatt, and G. A. M. Cross. 1999. A tightly regulated inducible expression system for dominant negative approaches in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **99:**89-101.

World Health Organisation. 1998. Control and surveillance of African trypanosomiasis, *In* Report of WHO expert committee. Technical Report Series **881**, I-VI:1114.

World Health Organisation. 2001. *In* African trypanosomiasis, Fact Sheet number **259** (WHO publications, Geneva).

CHAPTER 2

Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterisation

Eddy Chukwura Agbo [a], Phelix A.O. Majiwa [b], Eric J.H.M. Claassen [a] and Marleen H. Roos [a]

[a] Institute for Animal Science and Health (ID-Lelystad), Edelhertweg 15, 8200 AB Lelystad, The Netherlands
[b] International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya

Published in *Experimental Parasitology* 2001; **99:**123-131.

2.1. Abstract

We have evaluated whether sequence polymorphisms in the rRNA intergenic spacer region can be used to study the relatedness of two subspecies within the *Trypanosoma brucei*. Thirteen *T. brucei* isolates made up of 6 *T. b. brucei* and 7 *T. b. gambiense*, were analysed using a restriction fragment length polymorphism (RFLP). By PCR-based restriction mapping of the ITS1-5.8S-ITS2 ribosomal repeat unit, we found a fingerprint pattern that separately identifies each of the two sub-species analysed, with unique restriction fragments observed in all but one of the *T. b. gambiense* 'human' isolates. Interestingly, the restriction profile for a virulent, group 2 *T. b. gambiense* 'human' isolate, revealed an unusual RFLP pattern different from the profile of other 'human' isolates. Sequencing data from 4 representatives of each of the two subspecies indicated that the intergenic spacer region had a conserved ITS-1, and a variable 5.8S with unique transversions, insertions or deletions. The ITS-2 regions contained a single repeated element at similar positions in all isolates examined, but not in two of the 'human' isolates. A unique 4 bp $\{C_3A\}$ sequence was found within the 5.8S region of 'human' *T. b gambiense* isolates. Phylogenetic analysis of the data suggests that their common ancestor was an animal pathogen, and that 'human' pathogenicity might have evolved secondarily. Our data show that cryptic species within *T. brucei* group can be distinguished by differences in the PCR-RFLP profile of the rDNA repeat.

2.2. Index Descriptors and Abbreviations

Trypanosoma brucei subgenus; differentiation sequence; RFLP, restriction fragment length polymorphism; rRNA, ribosomal RNA; tRNA, transfer RNA; 5.8S ribosomal repeat; ITS1, first internal transcribed spacer; ITS2, second internal transcribed spacer; PCR, polymerase chain reaction; EDTA, ethylene di-amine tetra-acetic acid.

2.3. Introduction

Trypanosoma brucei is a unicellular, obligate parasite that is transmitted by tsetse flies. The disease trypanosomosis (Kassai et al., 1988) is generally an acute to chronic and debilitating one, designated 'nagana' in cattle and 'sleeping sickness' in humans. Infectivity for humans distinguishes *T. b. gambiense* and *T. b. rhodesiense* from *T. b. brucei*, which does not infect humans. It is generally thought that differences between *T. b. gambiense* and *T. b. rhodesiense* stem from adaptation of *T. brucei* to two different ecosystems in West Africa, where the main host is man, and East Africa, where man is only an incidental host (Gibson, 1986). *Trypanosoma b. gambiense* and *T. b. rhodesiense* have animal reservoirs, which constitute sources of infection for the tsetse fly vector (Uilenberg, 1998; Njagu et al., 1999).

Genetic exchange in *T. brucei* has been demonstrated in the laboratory (Jenni et al., 1986; Gibson, 1989), and crosses between human-infective and non-human infective sub-species have shown that the trait for human infectivity is heritable (Gibson, 1989; Gibson and Mizen, 1997). The ability to distinguish human infective from non-human infective isolates of *T. brucei* has important implications for studies of the transmission dynamics and anthropo-zoonotic potential of this parasite. Current parasitological classification of *T. brucei* into sub-species is problematic due to lack of stable characteristics that distinguish human infective from non-human infective trypanosomes. Both subspecies show uniformity in microscopic appearance, although there are variations in clinical outcome of the diseases they cause. Resistance to lysis by factors in human serum is one of the main distinguishing phenotypes, which indicate that the differences in disease syndrome result from complex interactions between the host and the parasite.

Over the past two decades, most investigators have focused on either the detection of trypanosome infections or definition of species rather than

subspecies. These studies have been based upon the comparison of isoenzymes (Gibson et al., 1980; Betshcart et al., 1983; Tait et al., 1984; Richner et al., 1989), molecular analysis of the genomic or mitochondrial DNA by RFLP and PCR-RAPD or on the use of microsatellite and minisatellite DNA probes (Kukla et al., 1987; Waitumbi and Murphy, 1993; Majiwa et al., 1994; Masake et al., 1997; Donelson and Artama, 1998; Morlais et al., 1998; Biteau et al., 2000). Several studies have been conducted to identify molecular differences between isolates of *T. brucei* subspecies that are human infective and those that are non-human infective, using Southern blot hybridisation analysis with repetitive sequences (Gibson et al., 1985a; Paindovoine et al., 1986; Hide et al., 1990, 1998), chromosome analysis (Brun and Jenni, 1987), isoenzyme analysis (Paindovoine, 1989; Mathieu-Daude and Tibayrenc, 1994) or frequency comparisons (Gibson et al., 1985b). Most of the previous studies have focused on intra-specific characterisation of *T. b. brucei* and *T. b. rhodesiense* stocks, with few reports on *T. b. gambiense* characterisation. Hide et al. (1990) reported genotypic differences between the West African *T. b. gambiense* and the East African *T. b. rhodesiense* isolates. The data suggested that the geographic location alone is insufficient to account for the differences observed.

For developing practical diagnostic and epidemiological markers for characterising *T. brucei* into subspecies in the West Coast of Africa where *T. b. gambiense* is prevalent, it is desirable to have a direct rapid method. A PCR-based method, which could be applied directly to the typing of clinical or field samples to identify a particular ‘human’ or animal isolate affords such a simple and rapid technique. The intergenic spacer separating the 18S and 28S rRNA genes is involved in the processing of pre-RNA (Young and Steitz, 1978) and may contain coding sequences for tRNA (Morgan et al., 1977; Young et al., 1979; Bacot and Reeves, 1991). The region is otherwise considered to be largely non-functional and under minimal selective pressure. Therefore, it is likely to contain sequence variations from which taxonomically useful information might be obtained. We have

evaluated whether PCR-based restriction fragment length polymorphism (PCR-RFLP) and sequence analyses of the internal transcribed spacers (ITS-1, ITS-2) and the intervening 5.8S region, could be used to identify and discriminate among *T. b. gambiense* and *T. b. brucei* isolates. Here we report on putative subspecies-specific markers that could be useful in the identification and characterisation of these subspecies of *T. brucei*.

2.4. Materials and Methods

2.4.1. Trypanosome isolates

Sources of the parasite isolates used in the study are listed in Table 1. Isolates were derived directly from their host of origin and had been identified previously as belonging to either of the two subspecies (Rickman and Robson, 1970). The *T. b. gambiense* isolates were those known to belong to either group 1 or 2 (Gibson, 1986). Three of the *T. b. brucei* isolates (STIB 215, B8/18 clone B, KP2 clone 7) have repeatedly been shown to be sensitive to lysis by human serum. After preliminary expansion and cloning of all isolates in laboratory rodents, genomic DNA was isolated from the trypanosomes essentially as described by Van der Ploeg et al. (1982).

2.4.2. PCR primer design and DNA amplification

T. brucei 18S, 5.8S and 28S rDNA sequence data (AJ009141, AJ009142, D89527, M12676, U22315, U22316, U22317, U22318, U22319, U22320 and U75507) were retrieved from GenBank, and aligned using MegaAlign (DNASTAR Inc., 1993). Based on consensus sequences in the 5' and 3' ends of the 18S and 28S primers, designated ED1-F (CAC CGC CCG TCG TTG TTT C) and ED1-R (TTC GCT CGC CGC TGA CTG A), were designed. A search for homology to known DNA sequences of other organisms was made using the BLAST program (Altschul et al., 1990). The ITS1, ITS2 and intervening 5.8S and flanking sequences at the 5' and 3'

ends of the 18S and 28S sequences, respectively, were amplified using the 18S-F and 28S-R primer pair.

The PCR contained 100ng each of ED1-F and ED1-R, 2.5mM dNTP's, 1xPCR buffer, 2mM $MgCl_2$, 1.25U AmpliTaq Gold (Perkin Elmer, Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands) and 10ng genomic DNA in 50 μl. The reaction profile consisted of 10 min at 95°C to activate *Taq* polymerase, followed by 35 cycles of 95°C for 1 min, 56 °C for 30 s, 72°C for 2 min, and a 10 min final extension at 72°C. Five μl of PCR product was electrophoresed in a 1.5% agarose gel (Ophaero Pronarose D-1, Burgos, Spain) with 1xTris-borate-EDTA (TBE) buffer containing 1μg of Ethidium bromide.

2.4.3. Sequencing of purified PCR products

The PCR products were separately purified using Boehringer High Pure PCR products purification kit (Roche Molecular Biochemicals, Mannheim, Germany). These were sequenced with both PCR primers, using an ABI Prism™ dye terminator cycle sequencing kit (Applied Biosystems, Foster City, Calif., USA). To facilitate the sequencing protocol, we utilised the internal primers, ITS1-F-INT (ATG GAA TGC GTA TCT CTC TAT) and 28S-R-INT (GGC GCC AAA ATA CAT GCG T). The relative positions of the primers in the *T. brucei* rRNA gene are indicated in Fig. 1. Sequences obtained from 4 representative isolates for each of the two subspecies (Table 1), were edited using Lasergene (DNASTAR Inc., 1993) and then aligned using Clustal V (Higgins et al., 1991). The alignment was improved using SeqMan™II (DNASTAR Inc., 1993). Phylogenetic analysis was performed using MegAlign™(DNASTAR Inc., 1993).

2.4.4. PCR-RFLP

A 10μl sample of one set of PCR products was digested with the restriction enzymes *DdeI*, *HhaI*, *AluI* and *MseI* for 3h at 37°C. The digests were analysed by electrophoresis in 3% MetaPhor gel with 1xTBE buffer

containing 1 μg Ethidium bromide. The 100 bp ladder (MBI Fermentas, Vilnius, Lithuania) was used as DNA size marker.

Table 1. Trypanosome isolates used in these studies

Isolate	Origin	Year	Original host
Trypanosoma brucei brucei			
1. WaTat	Uganda	1982	Deer
2. H3 *	Luangwa valley, Zambia	1974	Lion
3. STIB215 *	Serengeti, Tanzania	1971	Lion
4. B8/18 *	Nsukka, Nigeria	1962	Pig
5. KP2, clone 7 *	Côte d'Ivoire	1982	Fly (*G. palpalis*)
6. 1902 (sindo)	Kenya	1971	Waterbuck
T. brucei gambiense			
7. A006	Fontem, Cameroon	1988	Human
8. Boula	Bouenze, Congo	1989	Human
9. 1898	D.R. Congo	1974	Human
10. Dal972 *	Daloa, Côte d'Ivoire	1978	Human
11. Suzena *	Yambio, Sudan	1982	Human
12. NW2 *	Uganda	1992	Human
***T. brucei gambiense*, virulent group 2**			
13. TH2 (78E) *	Koudougou, Côte d'Ivoire	1978	Human

* Isolates sequenced

2.5. Results

2.5.1. RFLP of T. brucei rDNA ITS1-5.8S-ITS2 region

The ribosomal ITS1-5.8S-ITS2 region, 1282 - 1295 bp, was amplified using ED1-F and ED1-R. Agarose gel electrophoresis of the PCR products showed that a single amplimer, with insignificant length variation, was obtained for each isolate. The PCR products obtained using genomic DNA from *T. brucei* isolates, the 6 of animal origin and 7 of human origin, were subjected to endonuclease digestion in order to search for polymorphic restriction enzyme sites that could allow the discrimination of the isolates. The *DdeI*, *HhaI*, *AluI* and *MseI* restriction fragments migrated as 8 - 10 electrophoretic bands. Approximately 380 bp *DdeI* and 350 bp *HhaI* fragments (Fig. 2A), 350 bp *AluI* and 230 bp *MseI* fragments (Fig. 2B) were present in all the 'human' isolates tested, except in the Group 2 virulent *T. b. gambiense,* and were absent in all animal isolates.

2.5.2. Analysis of the ITS1, 5.8S and ITS2 region

DNA sequencing of the internal transcribed spacer (ITS1 and ITS2) and the 5.8S rRNA gene was performed directly on the PCR products. The sequences were assembled to form the complete sequence of the region and the respective sequences ranged from 1282 - 1295 bp. The results confirmed preliminary analysis of PCR amplimer patterns obtained by agarose gel electrophoresis, that all 'human' and animal isolates studied possessed similar domain lengths. Much of the sequences of the *T. b. brucei* and *T. b. gambiense* were alignable, except for a unique {C_3A} insertion, in the 5.8S region of *the T. b. gambiense* isolates, at sequence position 562 - 565 (Fig. 3), and several insertions, deletions and transversions at different positions of the 5.8S and ITS2 (see full sequence data). The end points of the 5.8S rRNA gene were determined by comparative alignment. The sequence is approximately 170 bp in length and identical in all 6 *T. b. brucei* and 7 *T. b. gambiense* isolates, except for a 4 bp deletion/insertion sequence at the central portion of the region. The

ITS2 of the *T. b. brucei* and *T. b. gambiense* isolates revealed considerable sequence variability in an area between the 3' end of the 5.8S rRNA gene and the central region.

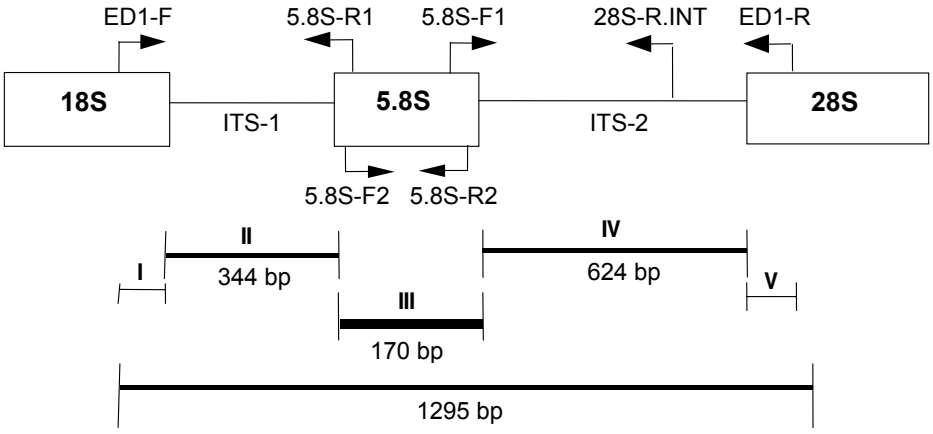


Fig. 1. Representative scheme of amplification of the ITS1, 5.8S and ITS2 repeat region (of isolate TH2). Arrows indicate the relative location and direction of the primers used for the amplification and sequencing of the respective ITS regions. The bars represent the amplification products. The sequenced region, subdivided into 5 regions, is indicated as I - V.

Five different regions with respect to rRNA gene identity can be distinguished (Fig. 1). Region I contains 126 bp of the 3' end of the 18S rDNA (46.2% G + C) which was strictly conserved between the subspecies. Region II corresponds to 344 bp of ITS1 (37% G + C) while Region III (170 bp in length, 60.7% G + C) corresponds to the 5.8S rRNA coding sequence. A unique 4 bp {C_3A} sequence was found within the 5.8S region of human-infective *T. b gambiense* isolates at sequence positions 562-566. ITS2 covers region IV (605 bp, 41% G + C) and consists of one central

repeat element in the four animal and one of the 'human' isolates (ranging from 89 - 99 bp). The 'human' isolate, Suzena, was found to contain two repeat elements in the region (61 and 89 bp), while no repeats were found in Dal 972 and TH2. Region V covers a 72 bp sequence at the 5' end of 28S rRNA gene. Phylogenetic analysis classified the isolates into either of two subspecies, with branches leading to the *T. b. brucei* subspecies relatively longer than those leading to the *T. b. gambiense* subspecies (Fig. 5A).

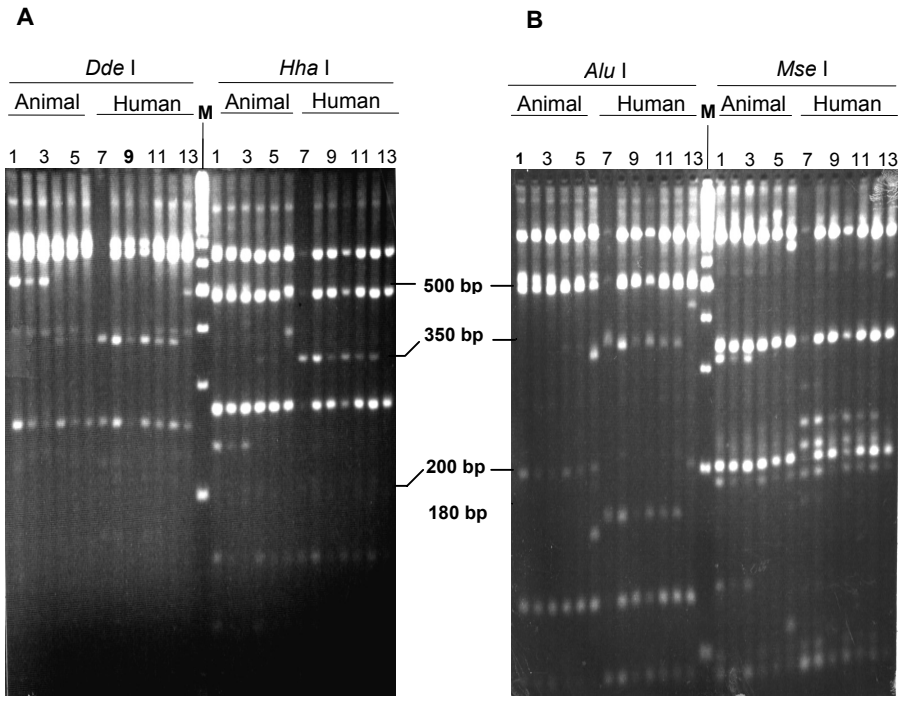


Fig. 2. PCR-RFLP analysis of 6 animal and 7 human isolates reveals a clearly differentiating pattern between the two groups. (A) PCR products amplified with primers ED1-F and ED1-R, and digested with *Dde*I and *Hha*I. (B) PCR products digested with *Alu*I and *Mse*I. 'M' is 100 bp ladder (New England Biolabs). Details on DNA samples (1 - 13) are as listed in Table 1. Arrows indicate polymorphic bands.

	Isolate			584
T.b. g -	NW2	TCATTTCATTG	**CCCA**	ATCTTTGAACGCAAACGGC
T.b. g -	Dal972	TCATTTCATTG	**CCCA**	ATCTTTGAACGCAAACGGC
T.b. b -	B8/18	TCATTTCATTG	- - - -	ATCTTTGAACGCAAACGGC
T.b. g -	TH2	TCATTTCATTG	**CCCA**	ATCTTTGAACGCAAACGGC
T.b. g -	Suzena	TCATTTCATTG	**CCCA**	ATCTTTGAACGCAAACGGC
T.b. b -	KP2	TCATTTCATTG	- - - -	ATCTTTGAACGCAAACGGC
T.b. b -	STIB215	TCATTTCATTG	- - - -	ATCTTTGAACGCAAACGGC

Fig. 3. Alignment region showing unique {C_3A} sequence in the 5.8S region of animal isolates. T. b. b - *Trypanosoma brucei brucei*; T. b. g - *T. b. gambiense*. The numbers in bold on top of the aligned sequences indicate the relative position of the regions.

2.6. Discussion

A relatively simple and rapid typing method based on PCR-RFLP patterns has been used to assess the relatedness, and to differentiate, between *T. b. brucei* and *T. b. gambiense*. The data presented demonstrate that the ITS spacers and the intervening 5.8S of *T. brucei* contain sufficient nucleotide sequence polymorphisms and intraspecific conservation which could be used for discriminating between *T. b. brucei* and *T. b. gambiense*. This is especially useful in view of the close phenotypic similarity of these two subspecies.

RFLP genotyping with some restriction enzymes segregated the 13 stocks into three groups: in group 1, the RFLP patterns of 5 out of the 6 *T. b. brucei* stocks isolated from different animal species in different regions of Africa were highly similar. The patterns of three *T. b. brucei* isolates (WaTat, H3 and STIB215) showed much closer similarity to each other than

to other isolates. In addition, the RFLP pattern of isolate 1902 isolated from a Waterbuck (in Kenya) was different from those of other animal isolates. This observation suggests that the isolate might be a unique strain within the East African sub-region. In the second group, the RFLP profiles of 6 out of the 7 *T. b. gambiense* isolated from humans from different countries and geographic regions of East, West and Central Africa gave a within-group homogeneity (Fig. 2). Thus the intraspecific variations observed when compared to the overall levels of variation indicate that the 'animal' and 'human' genotype might be conserved across different geographical regions.

The third group consisted of the TH2 isolate, a Group 2 (virulent) T. b. gambiense strain whose RFLP profiles were different from those of the other two groups, but more closely-related to those of T. b. brucei (Fig. 2). The TH2 isolate shared most of the bands common to the ‘human’ and animal isolates and had only one unique AluI fragment of about 440 bp (Fig. 2). Although the frequency of genetic exchange between T. b. brucei and T. b. gambiense may be variable or low, our data suggests that the virulent Group T. b. gambiense TH2 isolate might be a unique human-infective strain resulting from natural genetic exchange within the T. brucei subspecies. This agrees with the suggestion of Gibson and Stevens (1999) that this isolate might be a hybrid. In strict taxonomic terms therefore, a re-classification of the isolate as a separate subspecies is needed. The polymorphisms detected in the T. brucei isolates may be valuable markers for epidemiological studies. The RFLP data showed that the isolates having different outbreak history and isolated over decades apart (Table 1) have identical genomic fingerprints. This observation gives support to their epidemiological relatedness and suggests low plasticity of the analysed non-transcribed spacer region of rRNA gene. In line with generally held view, this may suggest the clonal expansion of one parental strain that had adapted to different climatic or bio-environments. Apparently, RFLP and genetic information derived from ITS1, 5.8S and ITS2 region can be used to differentiate these closely-related organisms.

The nucleotide sequence of the ITS1-5.8S-ITS2 allowed the assignment of an isolate with certainty to one of the two subspecies. Ribosomal gene clusters are located on different chromosomes; therefore, each unique sequence represents an allele of the genomic region. Overall, there was a high degree of sequence conservation among the isolates, but interesting variations occur both within and between the subspecies, depending on whether the isolate was human infective or non-human infective (Fig. 3). The general organisation of the entire region was not strictly typical in comparison with what has been observed in most of the other eukaryotic organisms. Repeated sequence elements were absent in two of the 8 sequenced PCR products. Generally, there is a high degree of sequence conservation between the repeats, but variation occurs both within and between subspecies (Fig. 4).

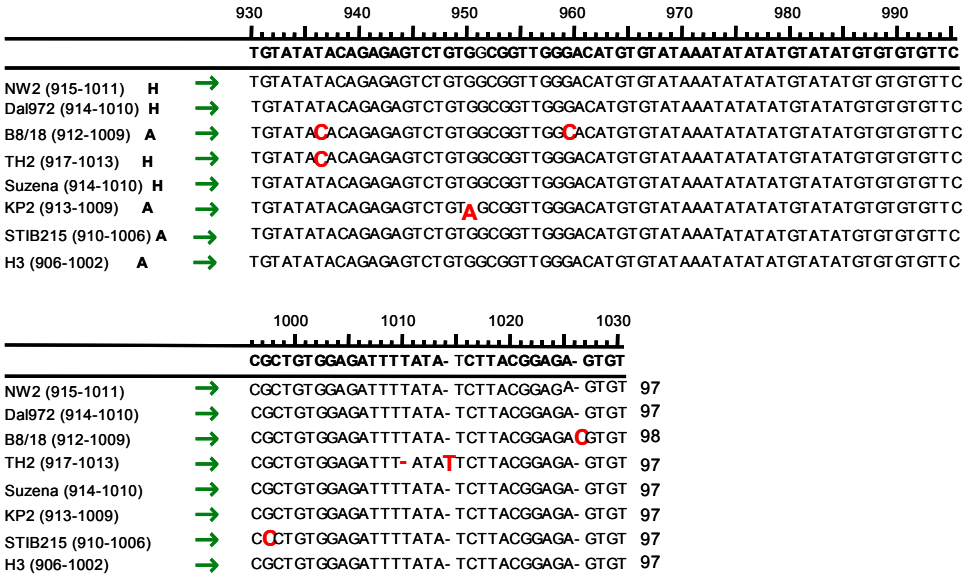


Fig. 4. Comparison of the sequences of the repeat element in the ITS2 of 4 animal and 4 human isolates: *T. b. brucei* - H3, STIB215, B8/18 and KP2; *T. b. gambiense* - Dal972, Suzena, NW2 and TH2. Animal-infective isolates are depicted as 'A' and human-infective isolates as 'H'. Base mismatches are shown in bold. The numbers to the right of the figure indicate length of each repeat. Consensus sequence is shown on the top.

The identical nature of the repeat elements among the isolates suggests that this repeating unit may represent an original sequence, which might also have an important function. However, the function, if any, of this repeat remains to be determined. Repeat units have been identified in non-transcribed spacer regions of nematodes (Kane and Rollinson, 1994), *Drosophila melanogaster* (Kohorn and Rae, 1982) and also in Xenopus (Moss, 1983), where they have been shown to influence the levels of transcription (Bushy and Reeder, 1983; Grimaldi and Nocera, 1988). To the best of our knowledge, there has not been any report of such repeat elements in trypanosomes. Region V contains 72 bp of 28S rRNA sequence. The length of region I - V for the isolates correlates well with the approximately 1300 bp length of the PCR fragments amplified with primers ED1-F and ED1-R flanking this region. However, intraspecific variation of the rRNA gene complex is believed to be far less than interspecific variation due in part to concerted evolution (Hillis and Dixon, 1991). We found that differences in the sequences of the non-transcribed spacer region between the two subspecies far exceeded variation within a subspecies. The results, although preliminary due to the relatively small number of isolates examined, indicate that the *T. b. brucei* and *T. b. gambiense* subspecies can be distinguished by differences in the size and number of DNA fragments after digestion of their ITS/5.8S region with specific restriction endonucleases. The presence of virtually identical long spacers among the isolates suggests that the approach of utilising differences in spacer lengths might be of little value in further discrimination of these subspecies. Determining the differences between the two closely-related subspecies within *T. brucei* might help determine how pathogenicity evolved.

Cluster analysis and dendograms of the ribosomal repeat unit (Fig. 5A) shows a more homogeneous *T. b. gambiense* clade, which descended from an older and more heterogeneous *T. b. brucei* line. Branches leading to the *T. b. brucei* isolates were relatively much longer, suggesting a faster rate of sequence evolution in the *T. b. brucei* subspecies. Furthermore, it suggests that the common ancestor was an animal pathogen, and that human pathogenicity might have evolved secondarily.

The phylogram, which depicts the differences in the ITS1, 5.8S and ITS2 regions between these isolates, shows the heterogeneity between the subspecies.

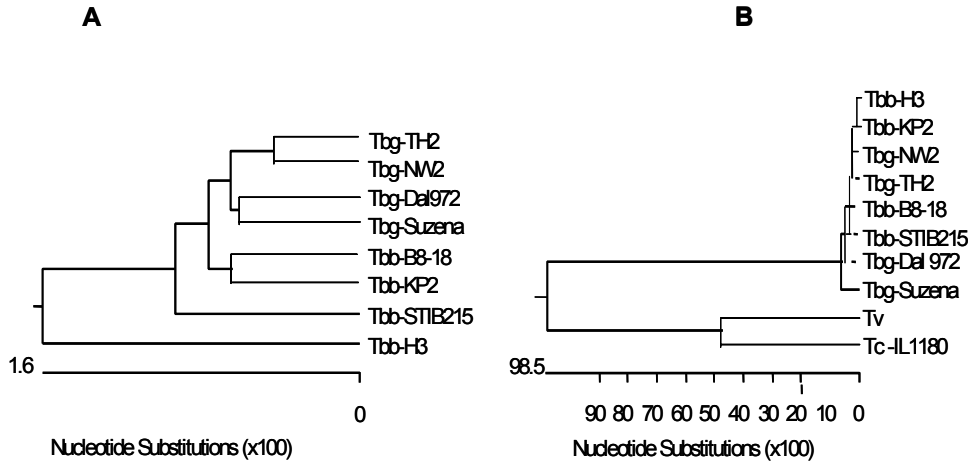


Fig. 5. Phylogenetic analysis of the ITS1-5.8S-ITS2 rDNA sequences showing the relationship between 4 *T. b. brucei* (H3, STIB215, B8/18 and KP2) and 4 *T. b. gambiense* (Dal972, Suzena, NW2 and TH2) isolates (Fig. 5A). In Fig. 5B, the phylogram is re-presented to include two outgroup species - *T. congolense* (Savannah-type IL1180) and *T. vivax*, which gives an inter-species yardstick with which to measure intra-species divergence. Sequences of the outgroup species were retrieved from the GenBank, with accession numbers U22315 and U22316, respectively.

To measure intra-species divergence, two outgroup species, *Trypanosoma congolense* and *Trypanosoma vivax*, were compared by nearest-neighbour joining method. The result showed that a neighbour-joining tree of the same region was clustered less tightly in these outgroups species, than with *T. b. brucei* and *T. b. gambiense* (Fig. 5B). Further phylogenetic analysis should concentrate on non-rDNA as they are likely to yield more phylogenetically meaningful data. It would be interesting to evaluate how these results will correlate in *T. b. rhodesiense* typing. However, it should be noted that *T. b. gambiense* is a more ancient human pathogen, supported in part by the observation that (like *T. b. brucei* in animals) it causes a chronic disease in man. Genetic background of the infected individual plays a major role as well in the outcome of infection. We conclude that this direct PCR-based typing approach may offer a simple and rapid typing technique for routine *T. brucei* diagnostic and epidemiological studies.

2.7. Acknowledgements

We express our gratitude to Dr Wendy Gibson for the gift of DNA samples. Drs M.F.W. te Pas and Philippe Büscher, and Filip Claes are gratefully acknowledged for their constructive comments on the manuscript. The Institute for Animal Science and Health (ID-Lelystad), Lelystad, The Netherlands, funded this work under internally funded *iSEO* project number 640.47293.00.

References

Aldard, R. D., S. C. Barker, D. Blair, and T. H. Cribbs. 1993. Comparison of second internal transcribed spacer (ribosomal DNA) from populations and species of *Fasciolidae* (Digenea). Int. J. Parasitol. **23:**423-425.

Altschul, S. F., W. Gish, W. Miller, E. W. Meyers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.

Bacot, C. M., and R. H. Reeves. 1991. Novel tRNA gene organisation in the 16S-23S intergenic spacer of the *Streptococcus pneumoniae* rRNA gene cluster. J. Bacteriol. **173:**4234-4236.

Betchart, B., R. Wyler, and L. Jenni. 1983. Characterisation of *Trypanozoon* stocks by isoelectric focus isoenzyme analysis. Acta Tropica **40:**25-28.

Biteau, N., F. Bringuad, W. Gibson, P. Truc, and T. Baltz. 2000. Characterisation of *Trypanozoon* isolates using repeated coding sequence and microsatellite markers. Mol. Biochem. Parasitol. **105:**185-201.

Brun, R., and L. Jenni. 1987. Human serum resistance of metacyclic forms of *Trypanosoma brucei brucei*, *T. b. rhodosiense* and *T. b. gambiense*. Parasitol. Res. **73:**218-223.

Busby, S., and R. H. Reeder. 1983. Spacer sequences regulate transcription of ribosomal gene plasmids injected into *Xenopus* embryos. Cell **34:**989-996.

Donelson, J. E., and W. T. Artama. 1998. Diagnosis of *T. evansi* by polymerase chain reaction (PCR). J. Protozool. Res. **8:**204-213.

Gibson, W. C., and B. T. Wellde. 1985. Characterisation of *Trypanozoon* stocks from the South Nyanza sleeping sickness focus in Western Kenya. Trans. R. Soc. Trop. Med. Hyg. **79:**236-237.

Gibson, W., and J. Stevens. 1999. Genetic exchange in Trypanosomatidae. Adv Parasitol. **43:**1-45.

Gibson, W., P. Borst and F. Fase-Fowler. 1985. Further analysis of intraspecific variation in *Trypanosoma brucei* using restriction site polymorphisms in the maxi-circle of kinetoplast DNA. Mol. Biochem. Parasitol. **15:**21-36.

Gibson, W.C. 1986. Will the real *Trypansoma b. gambiense* please stand up. Parasitol. Today **2:**255-257.

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

Gibson, W. C., and V. H. Mizen. 1997. Heritability of the trait for human infectivity in genetic crosses of *Trypanosoma brucei* spp. Trans. R. Soc. Trop. Med. Hyg. **91:**236-237.

Gibson, W. C., K. A. Osinga, P. A. M. Michels, and P. Borst. 1985. Trypanosomes of the subgenus *Trypanozoon* are diploid for housekeeping genes. Mol. Biochem. Parasitol. **16:**231-242.

Gibson, W. C., T. F. de C. Marshall, and D. G. Godfrey. 1980. Numerical analysis of enzyme polymorphism: a new approach to epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon.* Adv. Parasitol. **18:**175-246.

Grimaldi, G., and P. P. Di Nocera. 1988. Multiple repeated units in *Drosophilia melanogaster* ribosomal DNA spacer stimulate rRNA precursor transcription. Proc. Natl. Acad. Sci. USA **85:**5502-5506.

Hide, G., P. Cattand, D. Le Ray, D. J. Barry, and A. Tait. 1990. The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. Mol. Biochem. Parasitol. **39:**213-225.

Hide, G., A. Tilley, S. C. Welburn, I. Maudlin, and A. Tait. 2000. *Trypanosoma brucei*: identification of trypanosomes with genotypic similarity to human infective isolates in tsetse isolated from a region free of human sleeping sickness. Exp. Parasitol. **96:**67-74.

Hide, G., S. C. Welburn, A. Tait, and I. Maudlin. 1994. Epidemiological relationships of *Trypanosoma brucei* stocks from South-east Uganda: evidence for different population structures in human infective and non-infective isolates. Parasitology **109:**95-111.

Hide, G., S. D. Angus, P. H. Holmes, I. Maudlin, and S. C. Welburn. 1998. *Trypanosoma brucei*: comparison of circulating strains in an endemic and epidemic area of a sleeping sickness focus. Exp. Parasitol. **89:**21-29

Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1991. Clustal V: Improved software for multiple sequence alignment. Computers and Application Biosciences **8:**189-191.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quart. Rev. Biol. **66:**411-446.

Jenni, L., S. Marti, E. Pays, and M. Steinert.1986. Hybrid formation between African trypanosomes during cyclical transmission. Nature **322:**173-175.

Kane, R. A., and D. Rollinson. 1994. Repetitive sequences in the ribosomal DNA internal transcribed spacer of *Schistosoma heamatobium, S. intercalatum* and *S. mattheei*. Mol. Biochem. Parasitol. **63:**153-156.

Kassai, T. 1988. Standardised nomenclature of animal parasitic diseases (SNOAPAD). Vet. Parasitol. **29:**299-326.

Kohorn, B. D., and P. M. M. Rae. 1982. Non-transcribed spacer sequences promote in vitro transcription of *Drosophilia* ribosomal DNA. Nucleic Acids Res. **10:**6879-6886.

Kukla, B. A., P. A. O. Majiwa, C. J. Young, S. K. Moloo, and O. K. Ole-Mioyoi. 1987. Use of species-specific DNA probes for the detection and identification of trypanosome infections in tsetse flies. Parasitology **95:**1-26.

Majiwa, P. A. O., R. Thatthi, S. K. Moloo, J. H. Nyeko, L. H. Otieno, and S. Maloo. 1994. Detection of trypanosomes infections in the saliva of tsetse flies and buffy-coat samples from antigenic but aparasitemic cattle. Parasitology **10:**313-322.

Masake, R. A., P. A. O. Majiwa, S. K. Moloo, J. M. Makau, J. T. Njuguna, M. Maina, J. Kabata, O. K. Ole-Mioyoi, and V. M. Nantulya. 1997. Sensitive and specific detection *Trypanosoma vivax* using the polymerase chain reaction. Exp. Parasitol. **85:**193-205.

Mathieu-Daude, F., A. Bicart-See, M. F. Bossenno, S. F. Breniere, and M. Tibayrenc. 1994. Identification of *Trypanosoma brucei gambiense* group I by a specific kinetoplast DNA probe. Am. J. Med. Hyg. **50:**13-19.

Mathieu-Daude, F., and M. Tibayrenc. 1994. Isozyme variability of *Trypanosoma brucei* s.l.: genetic, taxonomic and epidemiological significance. Exp. Parasitol. **78:**1-19.

Mathieu-Daude, F., J. Stevens, J. Welsh, M. Tibayrenc, and M. McClelland. 1995. Genetic diversity and population structure of *Trypanosoma brucei*: clonality versus sexuality. Mol. Biochem. Parasitol. **72:**89-101.

Morgan, E. A., T. Ikemura, and M. Normura. 1977. Identification of tRNA genes in individual ribosomal RNA transcription units of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **74:**2710-2714.

Morlais, I., P. Grebaut, J. M. Bodo, and G. Cuny. 1998. Characterisation of trypanosome infections by polymerase chain reaction (PCR) amplification in wild tsetse flies in Cameroon. Parasitology **116:**547-554.

Moss, T. 1983. A transcriptional function for the repetitive ribosomal spacer in *Xenopus laevis*. Nature **302:**223-228.

Njagu, Z., S. Mihok, E. Kokwaro, and D. Verloo. 1999. Isolation of *Trypanosoma brucei* from the monitor lizard (*Varanus niloticus*) in an endemic focus of Rhodesian sleeping sickness in Kenya. Acta Tropica **72 :**137-148.

Paindovoine, P., E. Pays, M. Laurent, Y. Geltmeyer, D. Le Ray, D. Mehlitz, and M. Steinert. 1986. The use of DNA hybridisation and numerical taxonomy in determining relationships between *T. brucei* stocks and subspecies. Parasitology **92:**31-50.

Paindovoine, P., F. Zampetti-Bosseler, H. Couquelet, E. Pays, and M. Steinert. 1989. Different allele frequencies in *Trypanosoma brucei brucei* and *T. b. gambiense* populations. Mol. Biochem. Parasitol. **32:**61-72.

Richner, D., J. Schweizer, B. Betschart, and L. Jenni. 1989. Characterisation of West African *Trypanosoma* (*Trypanozoon*) *brucei* isolates from man and animals using isoenzyme analysis and DNA hybridisation. Parasitol. Res. **76:**80-85.

Rickman, L.R., and J. Robson. 1970. The testing of proven *Trypanosoma brucei* and *T. rhodesiense* strains by the blood incubation infectivity test. Bull. World Health Organisation **42:**911-916.

Tait, A., E. A. Babiker, and D. Le Ray. 1984. Enzyme variation in *Trypanosoma brucei* spp. I.: Evidence for subspeciation of *T. b. gambiense*. Parasitology **89:**311-326.

Uilenberg, G. 1998. African Animal Trypanosomes. In: A field guide for the diagnosis, treatment and prevention of African animal trypanosomosis, p 34.

Van der Ploeg, L. H. T., A. Bernards, F. A. M. Rijswijk, and P. Borst. 1982. Characterisation of the DNA duplication-transposition that controls the expression of two genes for variant surface glycoproteins in *Trypanosoma brucei*. Nucleic Acids. Res. **10:**593-609.

Waitumbi, J. N., and N. B. Murphy. 1993. Inter- and intra-species differentiation of trypanosomes by genomic fingerprinting with arbitrary primers. Mol. Biochem. Parasitol. **58:**501-517.

Young, R. A., and J. A. Stietz. 1978. Complimentary sequences 1700 nucleotides apart from a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. Proc. Natl. Acad. Sci. USA **75:**3593-3597.

Young, R. A., R. Maklis, and J. A. Steitz. 1979. Sequence of the 16S-23S spacer region in two ribosomal RNA operons of *Esherichia coli*. J. Biol. Chem. **254:**3264-3271.

CHAPTER 3

Molecular variation of *Trypanosoma brucei subspecies* as revealed by AFLP fingerprinting

Eddy Chukwura Agbo [1], Phelix A.O. Majiwa [3], Eric J.H.M. Claassen [2] and Marinus F.W. te Pas [1]

[1] Division of Animal Sciences, Section for Animal Genomics, Institute for Animal Science and Health (ID-Lelystad), Edelhertweg 15, 8200 AB Lelystad, The Netherlands
[2] Department of Immunology, Erasmus University Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands
[3] International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya

Published in *Parasitology* 2001; **124:**349-358

3.1. Summary

Genetic analysis of *Trypanosoma spp* depends on the detection of variation between strains. We have used the amplified fragment length polymorphism (AFLP) technique to develop a convenient and reliable method for genetic characterisation of *Trypanosome* (sub)species. AFLP accesses multiple independent sites within the genome and would allow a better definition of the relatedness of different *Trypanosome* (sub)species. Nine isolates (three from each *T. brucei* subspecies) were tested with 40 AFLP primer combinations to identify the most appropriate pairs of restriction endonucleases and selective primers. Primers based on the recognition sequences of *EcoR*I and *Bgl*II were chosen and used to analyse thirty-one *T. brucei* isolates. Similarity levels calculated with the Pearson correlation coefficient ranged from 15 to 98%, and clusters were determined using the unweighted pair-group method using arithmetic averages (UPGMA). At the intraspecific level, AFLP fingerprints were grouped by numerical analysis in two main clusters, allowing a clear separation of *T. b. gambiense* (cluster I) from *T. b. brucei* and *T. b. rhodesiense* isolates (cluster II). Interspecies evaluation of this customised approach produced heterogeneous AFLP patterns, with unique genetic markers, except for *T. evansi* and *T. equiperdum*, which showed identical patterns and clustered together.

3.2. Introduction

Trypanosoma brucei is a unicellular, protozoan parasite that is transmitted by tsetse flies. It consists of three subspecies - *T. b. brucei, T. b. gambiense* and *T. b. rhodesiense* that are indistinguishable by conventional morphological, biochemical and antigenic criteria but differ by their geographical distribution and host specificity (Gibson et al., 1980; Gibson, 1985). The disease, trypanosomosis (Kassai, 1988), is generally an acute to chronic, debilitating disease. *T. b. brucei*, the cause of 'nagana' in cattle

does not cause disease in humans because this subspecies is lysed by normal human serum. *T. b. gambiense* and *T. b. rhodesiense* are resistant to normal human serum, although *T. b. rhodesiense* can revert to a human serum sensitive phenotype (Hawking, 1973; Hawking, 1977). There is a resurgence of sleeping sickness and increasing incidence of drug resistance (Murray et al., 2000), thus the need for high-resolution analytical tools that will facilitate the typing of *T. brucei* isolates on routine basis increases. Moreover, the ability to differentiate between human-infective and human non-infective isolates has important implications for studies of the transmission dynamics and anthropo-zoonotic potential of this parasite.

As an initial step towards an appropriate strategy for management the disease, the genetic diversity of the parasite's populations must be assessed. Genetic markers have been used to study the diversity of trypanosome populations. These have included isoenzyme studies (Gibson et al., 1980, Tait et al., 1984; 1985, Richner et al., 1989; Godfrey et al., 1990; Stevens and Godfrey, 1992), randomly amplified polymorphic DNA data (Mathieu-Daude et al., 1995; Kukla et al., 1987; Waitumbi and Murphy, 1993), microsatellite and minisatellite DNA analysis (Donelson and Artama, 1998; Morlais et al., 1998; MacLeod et al., 2000; Biteau et al., 2000) and restriction fragment length polymorphism, RFLP (Hide et al., 1994; Agbo et al., 2001). However, the identification of the characteristics specific to a subspecies is not yet supported by the discriminatory power of conventional diagnostic methods. Furthermore, many of the genetic typing assays have drawbacks in that they may require a relatively large amount of high-quality DNA or, as with RAPD technique, may be difficult to reproduce between laboratories (Jones et al., 1997; Perez et al., 1998). Many markers, especially isoenzymes, can evolve too fast for use as epidemiological markers, so that results can be misinterpreted (Hide, 1999). Furthermore, in the majority of these methods, only a very limited part of the genome is covered through highly specific molecular targeting of one or more repetitive DNA elements. To avoid possible bias arising from

the use of such methods, the population structure should be inferred from neutral markers that distribute randomly throughout the genome.

We have used the amplified fragment length polymorphism (AFLP) technique, a PCR-based fingerprinting method, to investigate the genetic diversity of *T. brucei* isolates, and tested the technique for trypanosome genotyping in general. The technique has already been frequently used for linkage analysis in plant genetics, and has also been applied for high resolution fingerprinting of plant, bacterial and fungal genomes (Vos et al., 1995; Folkerstma et al., 1996; Lin et al., 1996). Moreover, it has been applied to the analysis of population diversity in potato cysts and root-knot nematodes (Folkerstma et al., 1996; Semblat et al., 2000). There are two main advantages (Dijkshoorn et al., 1996, Janssen et al., 1996, Lin et al., 1996) of this technique in characterising microbial populations**:** there is an extensive coverage of the genome under study (Janssen et al., 1996) and the complexity of the AFLP fingerprint can be advantageously managed by adding selective primers during PCR amplifications (Vos et al., 1995), and by varying the choice of restriction enzyme pairs. These features make the technique better suited for revealing higher levels of genetic variation than isoenzyme and RFLP analysis. In addition, the reproducibility is reported to be high (Vos et al., 1985), while the requirement for only small amounts of DNA makes it highly suitable for genotyping isolates on a large scale or for isolation of binary specific markers. A recent review by Masiga et al. (2000) highlighted the value of this technique to parasite genetic studies. The objective of this study was to ascertain the usefulness of the AFLP technique in assessing the genetic diversity of trypanosome strains at the species or subspecies level, and as a tool to study potential relationships of various trypanosome (sub)species.

3.3. Materials and Methods

3.3.1. Trypanosome materials

The 31 isolates used were directly derived from different hosts between 1968 and 1993, and have been characterised on the basis of their response to normal human serum, as well as on host and geographical origins (Gibson, W.C., Personal Communication). The geographical origin, year of isolation and source of strain are described in Table 1. Genomic DNA was isolated from the purified pellets by standard method (Van der Ploeg et al., 1982).

3.3.2. AFLP reactions

AFLP markers were assayed as previously described by Vos *et. al*. (1995), with the following modifications**:** 50 - 300 ng DNA was digested in a total volume of 20 µl using 10 U each of eight combinations of enzymes (*EcoRI/MseI, BglII/HpaII, HindIII/HpaII, EcoRI/HindIII, EcoRI/BglII, EcoRI/BamHI, BglII/HindIII* and *BglII/BamHI*). All restriction enzymes were from New England Biolabs or Roche Molecular Biochemicals. The digestion mixture was incubated at 37ºC for 4 h followed by the addition of 10 X ligation buffer [20 mM Tris-HCl, 60 mM KCl, 1 mM EDTA, 5 mM dithiothreitol, 50 % glycerol (v/v), pH 7.5 (4ºC), DNA ligase (1 U)] and 10 pmol (or 100 pmol for 4-bp recognising endonucleases) of respective adapters (Table 2) and incubation at 15ºC for 16 h.

In a series of preliminary experiments, PCR amplification conditions were optimised (data not shown). Based on these results, the final optimised reaction conditions for pre-selective PCR included 2.5 mM dNTP's, 2 mM $MgCl_2$, 1X PCR buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, 1.5 mM $MgCl_2$, 0.01 % (w/v) gelatin), 5 pM of each pre-selective primer (Table 1), 1.25 U AmpliTaq® DNA polymerase (Perkin-Elmer, Maarssen, The Netherlands) and four microlitres of 1:2-diluted ligation product. The optimised thermal cycling conditions were 2 min at 95°C, followed by 20 cycles of 95ºC for 30 s, 56ºC for 30 s and 72ºC for 1 min. The PCR product

was diluted 20-fold and 4 µl of the dilution was used for the selective amplification reaction, using primer combinations with a selective nucleotide at the 3' end of one or both primers. The cycling conditions were as for the pre-selective amplification, except for a final extension at 60ºC for 30 min. The final products were diluted four-fold, and 1 µl, together with a Genescan-500 internal standard (PE Applied Biosystems, Maarssen, The Netherlands), was analysed on a 7.3% denaturing sequencing gel by using an ABI 373A automated DNA sequencer. Gels were routinely prepared by using the ABI protocols and were electrophoresed for 5 h. Forty primer combinations were evaluated using all 9 possible +1/+0 or +1/+1 primer combinations for each pair of enzymes (with selective nucleotide only on 3' end of the upstream primer, or on both primers, respectively) (Table 2). The *T. brucei* isolates were analysed using three selected restriction endonuclease combinations, with the set of selective primers, viz: *Eco*+A/*Bgl*+0, *Eco*+A/*Bgl*+T and *Eco*+0/*Hind*+A.

3.3.3. Data Analysis

After electrophoresis, AFLP data were collected with the GenScan software (PE Applied Biosystems, Maarssen, The Netherlands). Densitometric values were transferred to the GelCampar v4.1 software (Applied Maths, Kortrijk, Belgium), and gels were normalised by using the internal size standard that was added to each lane. Levels of similarity between banding patterns were calculated with the Pearson correlation product-moment coefficient (Pearson, 1926). The unweighted pair group method using average linkage (UPGMA) was used to cluster the patterns (Vauterin and Vauterin, 1992).

3.4. Results

3.4.1. Selection of AFLP enzymes and primers

AFLP genetic markers were assessed for their usefulness in characterising molecular diversity among *T. brucei* subspecies. Nine *T. brucei* isolates were first analysed to determine the optimal AFLP conditions. Restriction digestion and adapter ligation were performed in alternate steps, followed by pre-selective and selective PCR amplification. Nine sets of restriction enzyme combinations were tested (Table 3). The banding patterns obtained using *EcoR*I and *Mse*I restriction enzymes (with A + T-rich recognition sequences) for example, contained more bands than fingerprints generated with *BamH*I and *Hpa*II (with G + C-rich recognition sequences).

Overall, fingerprints with the pair of 6-bp/4-bp recognising endonuclease combinations *(EcoR*I/*Mse*I, *Bgl*II/*Hpa*II and *Hind*III/*Hpa*II*)* gave unevenly distributed, non-scorable banding patterns or monomorphic profiles of mostly small-sized (<150 bp) fragments (Table 3). Thus, we did not find the sets of 6-bp/4-bp recognising endonuclease combinations suitable for useful AFLP fingerprinting of *T. brucei*. Two sets of restriction enzyme combinations (*EcoR*I/*Bgl*II and *EcoR*I/*Hind*III) were selected based on their reproducibility, even distribution of bands along the gel, number of polymorphic bands and polymorphism detected within and between subspecies. The *EcoR*I/*Bgl*II combination (with A selective nucleotide at the 3' end of *EcoR*I primer) gave the best polymorphic, evenly distributed profiles, and was chosen for further analysis. To evaluate for the reproducibility of our results, the AFLP analyses procedure was repeated three times with the 9 test isolates, starting from DNA digestion.

Table 1. Trypanosome isolates used in this study

Isolate	Origin	Year	Host
Trypanosoma brucei brucei :			
1. Lump266	Kiboko, Kenya	1969	Fly, G. pallidipes
2. PTAG130	Daloa, Ivory Coast	1985	Pig
3. KP2, clone 7	Kouassi-Perita, Ivory Coast	1982	Fly (G. palpalis)
4. B8/18	Nsukka, Nigeria	1962	Pig
5. SW3/87	Democratic Rep. of Congo	1987	Pig
6. LVBG3N	Lambwe valley, Kenya	1980	Cow
7. M249	Matuga, Kenya	1981	Sheep
8. H3	Luangwa valley, Zambia	1974	Lion
9. STIB215	Serengeti,Tanzania	1971	Lion
10. 1902 (sindo)	Kenya	1971	Waterbuck
11. AnTat2/2	Nigeria	1970	Fly, G. morsitans
12. AnTat17/1	Democratic Rep. of Congo	1978	Sheep
13. J10	Zambia	1973	N.A.
T. b. gambiense, group 1:			
14. ITMAP1843	Bouenza, Congo	1975	Human
15. A006	Fontem, Cameroon	1988	Human
16. Font-1	Fontem, Cameroon	1993	Human
17. PT16	Côte d'Ivoire	1992	Human
18. PT312	Côte d'Ivoire	1992	Human
19. Mabia	Bouenza, Congo	1989	Human
20. NW2	Uganda	1992	Human
21. Suzena	Yambio, Sudan	1982	Human
22. 1829 (Aljo)	Badundu, D.R. Congo	1970	Human
23. 1898	Democratic Rep. of Congo	1974	Human
24. Dal972	Daloa, Ivory Coast	1978	Human
T. b. gambiense, group 2:			
25. TB26	Bouenza, Congo	1983	Pig
Trypanosoma brucei rhodesiense :			
26. TRPZ320	Zambia	1983	Fly, G. pallidipes
27. Gambella II	Illubabor, Ethiipia	1968	Human
28. 058	Luangwa valley, Zambia	1975	Human
29. LVH56	Lambwe valley	1978	Human
30. LVH108	Lambwe valley	1980	Human
31. AnTat25.1*	Rwanda	1971	Human
T. evansi/ T. equiperdum:			
AnTat3.1 (T. ev)	South America	1969	Capybara
RoTat1.2 (T. ev)	Indonesia	1982	Water buffalo
AnTat4.1 (T.eq)	N.A.	unknown	N.A.
STIB818 (T. eq)	Beijing, China	1979	Horse
T. congolense:			
C49 (savannah)			
Gam2 (savannah)	The Gambia	1977	Cow
IL3900 (riverine/forest)	Burkina Faso	1982	Dog
ANR3 (riverine/forest)	The Gambia	1988	Fly
K45.1 (Kilifi)	Kenya	1982	Cow
WG5 (Kilifi)	Kenya	1980	Goat
T. simiae:			
TsØ2	Kenya	N.A.	Bushbuck
Ken2	The Gambia	1988	Fly

* Isolate was adapted-sensitive to normal human serum

Table 2. Adapters and PCR primer core sequences used in this study.

	Adapter	Primer core sequence
*EcoR*I	5'-CTC GTA GAC TGC GTA CC 3'-CAT CTG ACG CAT GGT TAA	5'-**6-FAM-** GAC TGC GTA CCA ATT C
*Hind*III	5'-CTC GTA GAC TGC GTA CC 3'-CTG ACG CAT GGT CGA	5'-**6-FAM-** GAC TGC GTA CCA GCT T
*Bgl*I	5'-CGG ACT AGA GTA CAC TGT C 3'-C TGA TCT CAT GTG ACA GCT AG	5'-**6-FAM-** GAG TAC ACT GTC GAT CT
*Mfe*I	5'-AAT TCC CAA GAG CTC TCC AGT AC 3'-G GTT CTC GAG AGG TCA TGA T	5'-**6-FAM-** GAG AGC TCT TGG AAT TG
*Xba*I	5'-GTC GTA GAC TGC GTA CG 3'-CTG ACG CAT GCG ATC	5'-**6-FAM-** GAC TGC GTA CGC TAG A
*Mse*I	5'-GAC GAT GAG TCC TGA G 3'-CTA CTC AGG ACT CAT	5'-GAT GAG TCC TGA GTA A
*Hpa*II (C/CGG)	5'-GAC GAT GAG TCC TGA T 3'-CTA CTC AGG ACT AGC	5'-GAT GAG TCC TGA TCG G

3.4.2. AFLP analysis of Trypanosoma brucei

The 31 *T. brucei* stocks comprising 13 animal-infective and 18 human-infective isolates were subjected to AFLP analysis. The AFLP profiles were found to be strikingly similar, especially for the *T. b. gambiense* isolates and SW3/87 (Fig. 1 A, lanes 19 to 31), in spite of their distant geographical origins and time of isolation. After normalisation of data, the similarity levels between individual fingerprints, calculated by Pearson correlation product-moment coefficient, ranged from 60 - 96%. The fingerprint pattern showed a relatively high degree of heterogeneity of the *T. b. brucei* and *T. b. rhodesiense* subspecies, while the *T. b. gambiense* isolates showed rather homogeneous patterns (Fig. 1A).

Table 3. Primer combinations that were evaluated by AFLP.
Fingerprints were scored (+ to ++++), based on relative total number of bands and uniformity of size distribution from 35 - 500 bp. The combinations shown in bold were selected for further analyses. 'O' is AFLP primer without any selective nucleotide; 'A' or 'T' indicate primer with the respective selective nucleotide.

Enzyme		*BamH*I			*EcoR*I			*Hind*III			*Hpa*II		*Mse*I				
	Base	O	A	T	O	A	T	O	A	T	T	TC	A	G	T	C	AT
. *Bgl*II	O	-	+	+	-	++++	++	-	**++**	+	+	+	-	-	-	-	-
	A	-	+	+	+	+++	+	+	+	+	**++**	+	-	-	-	-	-
	T	-	+	+	+	+++	+	+	+	+	**++**	+	-	-	-	-	-
. *Eco*RI	O	-	+	+				-	**+++**	**++**	-	-	-*	-*	-*	-*	-*
	A	-	+	+				-	**+**	+	-	-	-*	-*	-*	-*	-*
	T	-	+	+				-	+	+	-	-	-*	-*	-*	-*	-*
. *Hind*III	O	-	-	-							-	-	-	-	-	-	-
	A	-	-	-							**++**	+	-	-	-	-	-
	T	-	-	-							**++**	+	-	-	-	-	-
. *BamH*I	O				-	+	+	-	-	-	-	-	-	-	-	-	-
	A				+	+	+	-	-	-	-	-	-	-	-	-	-
	T				+	+	+	-	-	-	-	-	-	-	-	-	-

* Very weak bands.
-* Not scored because most bands were relatively very small (<150 bp) and faint.
- Not studied.

3.4.3 Cluster analysis

The use of Pearson coefficient (Pearson, 1926) for the calculation of inter-strain correlations allowed a rapid classification of analysed isolates into 2 main clusters, I and II (Fig. 1 A). The *T. b. gambiense* isolates (Cluster I) were maximally separated from *T. b. brucei* and *T. b. rhodesiense*, both residing in Cluster II. Cluster I isolates were further subdivided into sub-

clusters, denoted as P and Q, while cluster II has R and S subclusters (Fig. 1 A). Numerical analysis of the AFLP profiles showed that the *T. b. gambiense* cluster was joined with the *T. b. brucei/T. b. rhodesiense* cluster at a Pearson correlation level of 60%. Similarity level among the *T. b. gambiense* isolates (Cluster I) ranged from 92 to 96%, indicating a highly identical set of isolates. Despite the overall homogeneity of this cluster, our numerical data showed a further subdivision in which *T. b. gambiense* isolates from West and Central Africa were homogeneously distributed into 2 subclusters, designated as P and Q (Fig. 1 A). The first is composed of two Congo stocks from the same focus isolated from humans over a 14-year period, and all the five isolates from West Africa (Côte d'Ivoire and Cameroon). The second subcluster contains two Democratic Republic (D.R) of Congo stocks, isolated 14 years apart, and one isolate each from Congo, Sudan and Uganda. Interestingly, visual inspection of the normalised band patterns revealed two *T. b. gambiense*-specific marker bands, M1 and M2 (Fig. 1 A). One *'T. b. brucei'* stock (SW3/87), a pig isolate from Congo, was 'outgrouped' into cluster I. This implies that this particular isolate is closer to *T. b. gambiense* than it is to *T. b. brucei*/*T. b. rhodesiense* in Cluster II, which raises the issue of whether it is indeed a *T. b. brucei*. Given that it has the *T. b. gambiense*-specific bands M1 and M2, one is led to the conclusion that this isolate might in fact be a *T. b. gambiense*.

Two main clusters result comprising of isolates in the *Trypanozoon* family (*T. brucei, T. evansi, T. equiperdum*) (Cluster I), and those belonging to *Nannomonas* (*T. congolense*) and *Pycnomonas* (*T. simiae*) subgenera. Symbols P, Q and R classify species belonging to subgenera *Nannomonas*, *Pycnomonas* and *Trypanozoon*, respectively. Within Cluster II, the *T. b. brucei* and *T. b. rhodesiense* isolates were subdivided into two different sub-clusters, designated R and S. The Cluster II dendogram showed longer branch lengths of sub-clusters, with a similarity level of 60 to 80%. This indicates higher diversity among the isolates, compared to the 92 to 96% similarity level of Cluster I isolates.

Overall, highly identical fingerprints were obtained from the 12 *T. b. gambiense* isolates, while *T. b. brucei and T. b. rhodesiense* isolates showed more genetic diversity than *T. b. gambiense*.

3.4.4. Interspecies analysis

To evaluate the applicability of the technique for interspecies analysis, AFLP data obtained from two independent isolates each of *T. congolense (kilifi, riverine/forest* and *savannah), T. evansi* and *T. equiperdum* were similarly analysed. The fingerprint patterns were compared to those of randomly selected *T. brucei* isolates (Fig. 1 B). The dendogram of the 19 populations, based on AFLP analysis using the UPMGA program, resulted in 2 main clusters, in which the *Trypanozoon* (*Trypanosoma brucei ssp, T. evansi and T. equiperdum*) isolates were maximally separated from the *Nannomonas* (*T. congolense*) and *Pycnomonas* (*T. simiae*) at a similarity level of 15%. The similarity level between individual fingerprints of each pair of (sub)species, calculated by Pearson coefficient, ranged from 15 to 95%, indicating an overall high level of genetic heterogeneity among the subgenera (Fig. 1 B). The *Trypanozoon* isolates (Cluster I) were further divided into two subclusters comprising (i) the *T. b. gambiense* (PT16 and Mabia), and SW3/87, a *'T. b. brucei'* which had been included to evaluate if its clustering position would also be different in an interspecies analysis. Both are distinct from other *Trypanozoon* isolates; subcluster II comprised all the other members of *Trypanozoon* subgenus (*T. b. brucei*, *T. b. rhodesiense*, *T. evansi* and *T. equiperdum*). The Kenyan *T. b. brucei* isolate from cow (LVBG3N) is out-grouped from other members of this subcluster, showing a more distant relationship. The *T. b. rhodesiense* isolates (Gambella-2 and 058) are more closely-related to the *T. evansi*/*T.equiperdum* clade, both of which are separated at a similarity level of 73%. Interestingly, KP2, clone 7, a tsetse fly isolate from Côte d'Ivoire is on the *same T. evansi/T. equiperdum* clade. Overall, the two subclusters show a genetic similarity level of about 68%, which is a significant level of diversity within the *Trypanozoon* subgenus.

Within Cluster II, heterogeneous fingerprint patterns were obtained, except for *T. evansi* and *T. equiperdum*, which shared highly similar AFLP patterns and a high level of genetic homogeneity of 95 - 98% (Fig. 1 B). Furthermore, in the interspecies analysis, there was a higher degree of homogeneity between the fingerprint patterns of each of the analysed pair of isolates than in the intraspecific analysis (Fig. 1 B). For instance, the fingerprint patterns of IL3900 and ANR3, both *T. congolense* (riverine/forest) isolates, were quite similar (Fig. 1 B, lanes 1 and 2). Taken together, the data indicate that there exists remarkable diversity within the genomes of *Trypanozoon* isolates.

Fig. 1. (A) Numerical analysis of normalised AFLP bands patterns generated from *T. b. brucei*, *B* (n = 13), *T. b. gambiense*, *G* (n = 12) and *T. b. rhodesiense*, *R* (n = 6), using primer combination *Eco*RI+A/*Bgl*II+0. The dendogram was constructed using the unweighted pair-group method using arithmetic averages, UPGMA. Correlation levels were expressed as percentage values using the Pearson coefficient. M1 and M2 denote subspecies-specific AFLP marker bands differentiating *T. b. brucei* from *T. b. gambiense*/*T. b. rhodesiense*. (B) Comparative dendogram based on AFLP band patterns from interspecies analysis of *Trypanosoma congolense* - Tc/kil (kilifi), Tc/WA (riverine/forest) and Tc/sav (savannah), *T. simiae* (Ts), *T. evansi* (Tev) and *T. equiperdum* (Teq) strains. Their fingerprints were compared to the patterns generated from two isolates each of *T. b. gambiense* (TB26 and Mabia) and *T. b. rhodesiense* (Gambella-2 and 058), and three *T. b. brucei* isolates (KP2, clone 7, LVBG3N and SW3/87). Similarities between pairs were calculated by Pearson correlation coefficient. Fingerprints were generated with *Eco*+A/*Bgl*+0 primer combinations. Isolates grouped in P, Q and R belong to the subgenus *Nannomonas*, *Pycnomonas* and *Trypanozoon*, respectively.

Fig. 1 A

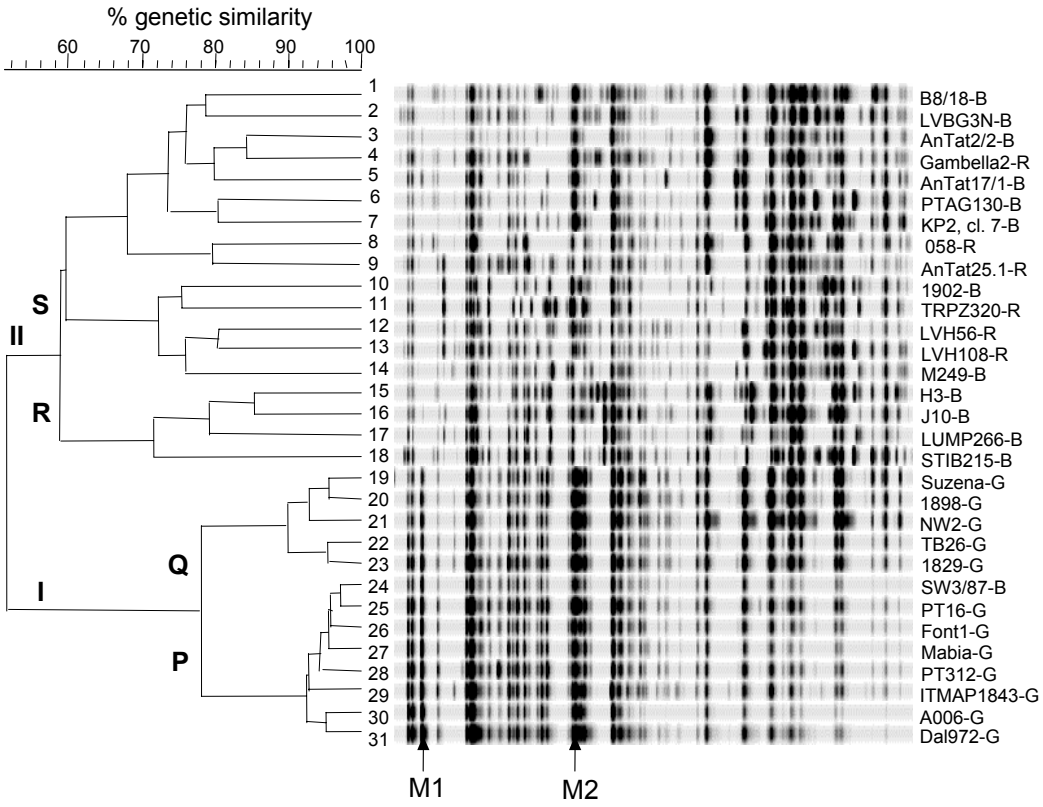
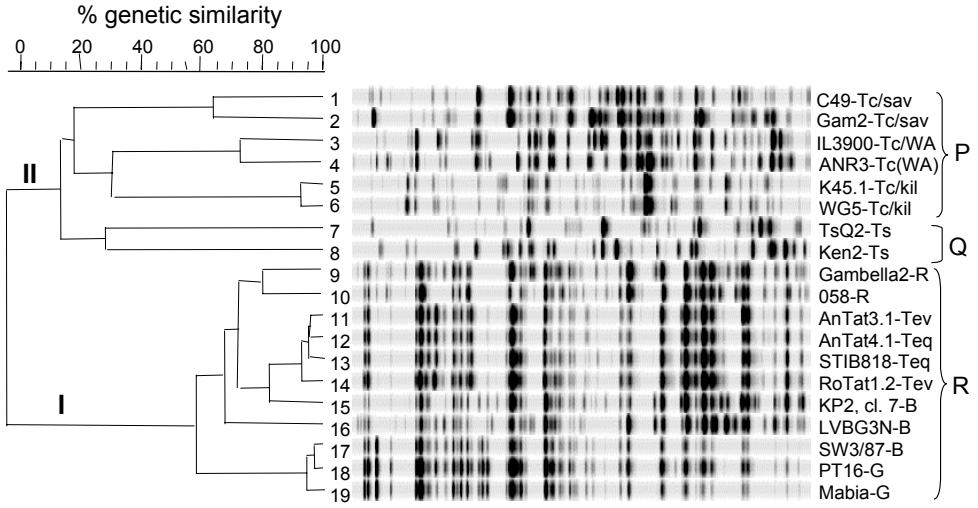

% genetic similarity
60
70
80
90
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
B8/18-B
LVBG3N-B
AnTat2/2-B
Gambella2-R
AnTat17/1-B
PTAG130-B
KP2, cl. 7-B
058-R
AnTat25.1-R
1902-B
TRPZ320-R
LVH56-R
LVH108-R
M249-B
H3-B
J10-B
LUMP266-B
STIB215-B
Suzena-G
1898-G
NW2-G
TB26-G
1829-G
SW3/87-B
PT16-G
Font1-G
Mabia-G
PT312-G
ITMAP1843-G
A006-G
Dal972-G
S
II
R
Q
I
P
M1
M2

Fig. 1 B


% genetic similarity
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
C49-Tc/sav
Gam2-Tc/sav
IL3900-Tc/WA
ANR3-Tc(WA)
K45.1-Tc/kil
WG5-Tc/kil
TsQ2-Ts
Ken2-Ts
Gambella2-R
058-R
AnTat3.1-Tev
AnTat4.1-Teq
STIB818-Teq
RoTat1.2-Tev
KP2, cl. 7-B
LVBG3N-B
SW3/87-B
PT16-G
Mabia-G
P
Q
R
II
I

3.5. Discussion

Accurate measurement of genetic relationships requires a marker system that provides an unbiased estimate of total genomic variance and is sufficiently abundant to minimise errors due to sampling variance (Spooner et al., 1996). We looked for a genotyping method suitable for trypanosome (sub)species for several reasons. The limitations of previously described methods based on single genetic locus have been highlighted above. Karyotypes of *Trypanosoma brucei* isolates, for example, are known to be highly plastic (Gibson, 1991), with unusual chromosomal size polymorphisms between isolates (Melville et al., 1998). Therefore, any genotyping technique based on a single genetic locus may provide inadequate results. The AFLP method is an extremely useful and reliable technique for detecting polymorphisms and its reproducibility is reported to be very high (Vos et al., 1995). The appropriate choice of restriction enzymes and, the number and base composition of selective bases, determine the usefulness and applicability of this technique in genetic diversity studies. In theory, each 6-bp recognising endonuclease cleaves every 4^6 (1096) bp in the genome. Since trypanosome genome is relatively small, the addition of only one selective nucleotide should yield scorable banding patterns. In our hands, the use of a pair of 6-bp recognising endonucleases (*EcoR*I/*Bgl*II), with 'A' selective nucleotide at the 3' end of the *Bgl*II primer was sufficient to generate useful fingerprints for trypanosome (sub)species genotyping. We propose this combination as the standard set with which to characterise trypanosomes to the (sub)species level.

In this study, we identified optimal AFLP conditions for *Trypanosome (sub)species* fingerprinting. The restriction endonucleases that were used to genotype other species did not produce satisfactory AFLP patterns: trypanosome genomic DNA was not adequately digested with *BamH*I, *Hha*I digestion resulted in monomorphic fragments, while *EcoR*I and *Mse*I combination resulted in too many small fragments. Restriction

endonucleases *EcoR*I combined with another 6-bp-cutting endonuclease *Bgl*II resulted in reproducible, evenly distributed and the most informative band patterns. The resolution of different selective primers tested varied, but *EcoR*I + A selective nucleotide combined with *Bgl*II primer with no selective base produced the best patterns. Genetically related strains produced homologous patterns and grouped together, whereas unrelated strains separated on the dendogram.

In certain eukaryotic systems especially plants, polysaccharides are often co-precipitated with DNA (Rether et al., 1993), and extraction steps with organic solvents (such as phenol, chloroform) may fail to strip all sugars completely. With the high glycoprotein content in trypanosome surface coat (the Variable Surface Glycoproteins, VSG) this may lead to partial or poor DNA digestion, which is an important step in the AFLP technique. To improve the chances of complete enzyme restriction, we opted for restriction digestion of genomic DNA followed by ligation in two separate steps. This produced better and more consistent results compared to digestion and ligation in one reaction (data not shown). To obtain reproducible fingerprints, a standardised protocol for AFLP analysis and computer-based analysis was required. Since a pre-amplification step could preferentially enrich heterosite fragments following adapter ligation (Vos et al., 1995), we chose the two-step PCR amplification procedure: a pre-selective reaction (without any selective nucleotide on primer), followed by selective amplification reaction.

As in previous studies with RAPD and single-locus analyses (Mathieu-Daude et al., 1995; Hide et al., 1998; Biteau et al., 2000; MacLeod et al., 2000; Agbo et al., 2001), our data correlated the close genetic relatedness among the three *T. brucei* subspecies. In our analysis, therefore, we aimed to group the isolates rather than distinguish them, crucial for bio-diversity surveys or epidemiological research. The Pearson correlation coefficient (Pearson, 1926) suggest that individual *T. b. brucei* and *T. b. rhodesiense* isolates from various foci may be quite distinct while the dendogram data

indicate that isolates from the same region are similar (Fig. 1 A). Some of the *T. b. brucei* and *T. b. gambiense* isolates used in this study have also been examined, using RFLP, for variations in their ITS/5.8S rDNA region (Agbo et al., 2001), who proposed two distinct evolutionary lines of descent. Interestingly, in these current studies, each of the main clusters of isolates revealed by numerical analysis of AFLP variations corresponded to the respective evolutionary line of the organism. It should be noted that AFLP is independent of ITS/5.8S rDNA sequence analysis because the nucleotide positions analysed by AFLP are distributed over the whole genome. Therefore, the two *T. b. gambiense*-specific AFLP fragments that we detected (Fig. 1 A) appear to additional reliable and specific markers for the characterisation of *T. b. gambiense* stocks. Clustering analysis showed that variations were more pronounced among *T. b. brucei/T. b. rhodesiense* subspecies (cluster II), which suggests a slightly different level of genetic heterogeneity between the two lines.

For *T. brucei* genotyping, our data indicate that human-infective *T. b. gambiense* form a distinct homogeneous group separate from *T. b. brucei* and *T. b. rhodesiense* isolates, which cluster together. This agrees with earlier reports that *T. b. gambiense* is a homogeneous group of isolates with high levels of similarity (Biteau et al., 2000). The low level of molecular variation among the *T. b. gambiense* isolates may be explained by two hypothesis: (i) the introduction of human-infective trypanosomes to West and Central Africa (the ecozone for *T. b. gambiense*) was more recent and, most probably, only a few strains were introduced, and; (ii) the cumulative effects adopted for the control of the disease. The control of Gambian sleeping sickness relies on case identification, principally using the card agglutination test (Magnus et al., 1978), and chemotherapy of a largely asymptomatic human reservoir. On the other hand, *T. b. rhodesiense*, the cause of Rhodesian sleeping sickness is zoonotic, with reservoirs in wild animals and domestic livestock (Welburn et al., 2001). Combined with the acuteness of the disease, its control demands a more aggressive approach. These have included large-scale tsetse control operations, aerial

spraying and aggressive medical interventions. These measures may have accumulated to induce more mutational pressure on *T. b. rhodesiense*, resulting in an increased level of genetic diversity within the subspecies.

As an interspecies yardstick with which to measure intra-specific divergence among *T. brucei* subspecies, two isolates each from *Nannomonas* (*T. congolense*) and *Pycnomonas* (*T. simiae*) subgenera were similarly analysed, and their fingerprints compared to *Trypanozoon* (*T. brucei*) patterns. As shown in Fig. 1 B, the different isolates within each subgenus have similar but distinct fingerprint patterns, except for *T. evansi* and *T. equiperdum*, whose fingerprint patterns were closely identical. This obvious close relationship between the two 'species' is in agreement with earlier reports (Gibson et al., 1980; Lun et al., 1992; Zhang et al., 1994; Brun et al., 1998; Biteau et al., 2000). However, a clear correlation was found between AFLP polymorphisms and trypanosome species. The specific bands present in each of the analysed species allowed us to separate the strains effectively. On the basis of the pronounced differences in their dendograms, different subtypes of *T. congolense* (kilifi, riverine/forest and savannah) could be considered as separate subspecies. Sequence information based on rRNA genes has been used to characterise and to infer the phylogenetic relationship among *Trypanozoon*, *Pycnomonas* and *Nannomonas* trypanosome parasites (Stevens et al., 1999; Urakawa and Majiwa, 2001; Gibson et al., 2001). Our data and the previous reports have corroborated the close genetic relatedness between *T. b. brucei* and *T. b. rhodesiense*. In addition, our results did show the existence of micro-heterogeneity among isolates of these subspecies. Overall, our data show that there does exist more genetic diversity within and between the subgenera than was previously reported.

The AFLP method has a higher multiplex ratio (= number of loci simultaneously analysed per experiment) than single sequence repeats (SSRs). As it simultaneously accesses multiple independent sites within the entire genome, it provides a more valuable tool for overall evaluation of the

phylogenetic relatedness of trypanosome strains. Our study has established the AFLP conditions and primer combinations that permit the assessment of genetic diversity of *T. brucei* subspecies. It further demonstrated that the technique could be applied in trypanosome diversity studies, with potential use for the isolation of intra- or inter-specific genetic markers. The results also show that human infective isolates derived from different ecozones may indeed be genetically separate groups. Furthermore, they correlate the close phenotypic relatedness between *T. b. brucei* and *T. b. rhodesiense*. In addition, our data suggest that variation between human-infective *T. brucei* subgenus is beyond only geographical origin, since the *T. b. rhodesiense* seem to be genetically more diverse, while the *T. b. gambiense* are more clonal or genetically more stable over time. By this technique, the two subspecies can be unambiguously classified by whole-genome fingerprinting. Through the evaluation of large number of clearly defined field samples, such AFLP fingerprinting may facilitate the identification of polymorphisms linked to parasite factors of host tropism, and contribute to the understanding of host-parasite interactions at the molecular level. Finally, our data present some further markers for defining *T. b. gambiense*, while the detected polymorphisms of trypanosome species may be valuable tools for epidemiological studies.

3.6. Acknowledgements

We thank Drs. Philippe Büscher and Magda Radwanska for their constructive comments on the manuscript. The Institute for Animal Science and Health (ID-Lelystad), Lelystad, The Netherlands, is kindly acknowledged for funding this study under internally funded *iSEO* project number 640.47293.00.

References

Agbo, E. C., P. A. O. Majiwa, E. H. J. M. Claassen, and M. H. Roos. 2001. Measure of molecular diversity within the *Trypanosoma brucei* subspecies - *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterisation. Exp. Parasitol. **99:**123-131

Biteau, N., F. Bringaud, W. Gibson, P. Truc, and T. Baltz. 2000. Characterisation of *Trypanozoon* isolates using repeated coding sequence and microsatellite markers. Mol. Biochem. Parasitol. **105:**185-201.

Brun, R., H. Hecker, and Z. R. Lun. 1998. *Trypanosoma evansi* and *Trypanosoma equiperdum***:** distribution, biology, treatment and phylogenetic relationship (a review). Vet. Parasitol. **79:**95-107.

Dijkshoorn, L., H. Aucken, P. Gerner-Smidt, P. Janssen, M. E. Kaufmann, J. Garaizer, J. Ursing, and T. L. Pitt. 1996. Comparison of outbreak and non-outbreak *Actinobacter baumanni* strains by genotypic and phenotypic methods. J. Clin. Microbiol. **34:**1519-1525.

Donelson, J. E., and W. T. Artama. 1998. Diagnosis of *T. evansi* by polymerase chain reaction (PCR). J. Protozool. Res. **8:**204-213.

Folkertsma, R.T., R. van der Voort, J. N. A. M. de Groot, K. E. de Groot, P. M. van Zandvoort, F. J. Gommers, J. Helder, and J. Bakker. 1996. Gene pool similarities of potato cyst nematode populations assessed by AFLP analysis. Mol. Plant-Microbe Interactions **9:**47-54.

Gibson, W. C., and J. K. Gashumba. 1983. Isoenzyme characterisation of some *Trypanozoon* stocks from a recent trypanosomiasis epidemic in Uganda. Trans. R. Soc. Trop. Med. Hyg. **77:**114-118.

Gibson, W. C., T. F. de C. Marshall, and D. G. Godfrey. 1980. Numerical analysis of enzyme polymorphism**:** a new approach to epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. Adv. Parasitol. **18:**175-246.

Gibson, W.C., and B. T. Wellde. 1985. Characterisation of some Trypanosome stocks from the South Nyanza sleeping sickness focus in Western Kenya. Trans. R. Soc. Trop. Med. Hyg. **79:**671-676.

Godfrey, D. G. 1982. Diversity within *Trypanosoma congolense*. In: *Perspectives in Trypanosomiasis Research*. Baker, J. R. (ed.). Wiley, London, pp 37-46.

Godfrey, D.G., R. D. Baker, L. R. Rickman, and D. Mehlitz. 1990. The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. Adv. Parasitol. **29:**1-74.

Hawking, F. 1973. The differentiation of *Trypanosoma rhodesiense* from *T. brucei* by means of human serum. Trans. R. Soc. Trop. Med. Hyg. **67:**517-527.

Hawking, F. 1977. The resistance to human plasma of *Trypanosoma brucei*, *T. rhodesiense* and *T. gambiense*. In: Analysis of the composition of trypanosome strains. Trans. R. Soc. Trop. Med. Hyg. **70:**504-512.

Hide, G., S. C. Welburn, A. Tait, and I. Maudlin. 1994. Epidemiological relationships of *Trypanosoma brucei* stocks from South-east Uganda: evidence for different population structures in human infective and non-infective isolates. Parasitology **109:**95-111.

Hide, G., S. D. Angus, P. H. Holmes, I. Maudlin, and S. C. Welburn. 1998. *Trypanosoma brucei*: comparison of circulating strains in an endemic and an epidemic area of a sleeping sickness focus. Exp. Parasitol. **89:**21-29.

Janssen, P., R. Coopman, G. Huys, J. Swings, M. Bleeker, P. Vos. M. Zabeau, and K. Kersters. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. Microbiology **142:**1881-1893.

Jones, C. J., K. J. Eedwards, S. Castaglione, M. O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, and A. Daly et al. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breeding **3:**381-390.

Kassai, T. 1988. Standardised nomenclature of animal parasitic disease (SNOAPAD). Vet. Parasitol. **29:**299-326.

Kukla, B. A., P. A. O. Majiwa, C. J. Young, S. K. Moloo, and O. K. Ole-Mioyoi. 1987. Use of species-specific DNA probes for the detection and identification of trypanosome infections in tsetse flies. Parasitology **95:**1-26.

Lin, J., J. Kuo, and J. Ma. 1996. A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. Nucleic Acids Res. **24:**3649-3650.

Lun, Z. R., R. Brun, and W. C. Gibson. 1992. Kinetoplast DNA and molecular karyotypes of *Trypanosoma evansi* and *T. equiperdum* from China. Mol. Biochem. Parasitol. **50:**189-196.

MacLeod, A., A. Tweedie, S. C. Welburn, I. Maudlin, C. M. R. Turner, and A. Tait. 2000. Minisatellite marker analysis of *Trypanosoma brucei*: Reconciliation of clonal, panmictic and epidemic population genetic structures. Proc. Natl. Acad. Sci. USA **97:**13442-13447.

Magnus, E., T. Vervoort, and N. Van Meirvenne. 1978. A card agglutination test with stained trypanosomes (C.A.T.T) for the serological diagnosis of *T. b. gambiense* trypanosomiasis. Ann. Soc. Belge Med. Trop. **58:**169-176.

Majiwa, P. A. O., M. Maina, J. N. Waitumbi, S. Mihok, and E. Zwygarth. 1993. *Trypanosoma (Nannomonas) congolense*: molecular characterisation of a new genotype from Tsavo, Kenya. Parasitology **106:**151-162.

Masiga, D. K., A. Tait, and C. M. R. Turner. 2000. Amplified restriction fragment length polymorphism in parasite genetics. Parasitol. Today **16:**350-353.

Mathieu-Daude, E. F., J. Stevens, J. Welsh, M. Tibayrenc, and M. McClelland. 1995. Genetic diversity and population structure of *Trypanosoma brucei*: clonality versus sexuality. Mol. Biochem. Parasitol. **72:**89-101.

Morlais, I., P. Grebaut, J. M. Bodo, and G. Cuny. 1998. Characterisation of trypanosome infections by polymerase chain reaction (PCR) amplification in wild tsetse flies in Cameroon. Parasitology **116:**547-554.

Murray, H. W., J. Pepin, T. B. Nutman, S. L. Hoffman, and A. A. Mahmoud. 2000. Tropical Medicine. British Medical J. **320:**490-494.

Pearson, K. 1926. On the coefficient of racial likeness. Biometrika **18:**105-117.

Perez, T., J. Albomoz, and A. Dominguez. 1998. An evaluation of RAPD fragment reproducibility and nature. Mol. Ecol. **7:**1347-1357.

Rether, B., G. Delmas, and J. Laoued. 1993. Isolation of polysaccharide-free DNA from plants. Plant Mol. Biol. Rep. **11:**333-337.

Richner, D., J. Schweizer, B. Betschart, and L. Jenni. 1989. Characterisation of West African *Trypanosoma* (*Trypanozoon*) *brucei* isolates from man and animals using isoenzyme analysis and DNA hybridisation. Parasitol. Res. **76:**80-85.

Semblat, J-P., M. Bongiovanni, E. Wajnberg, A. Dalmasso, P. Abad, and P. Castagnone-Sereno. 2000. Virulence and molecular diversity of parthenogenetic root-knot nematodes, *Meloidogyne spp*. Heredity **84:**81-89.

Spooner, D. M., J. Tivang, J. Nienhuis, J. T. Miller, D. S. Douches, and M. A. Contreras. 1996. Comparison of four molecular markers in measuring relationships among the wild potato relatives Solanum section *Etuberosum (subgenus Potatoe*). Theor. Appl. Genet. **92:**532-540.

Stevens, J. R., and D. G. Godfrey. 1992. Numerical taxonomy of *Trypanozoon* based on polymorphisms in a reduced range of enzymes. Parasitology **104:**75-86.

Tait, A., E. A. Babiker, and D. Le Ray. 1984. Enzyme variation in *Trypanosoma brucei* spp. I.: Evidence for sub-speciation of *T. b. gambiense*. Parasitology **89:**311-326.

Tait, A., J. D. Barry, R. Wink, A. Sanderson, and J. S. Crowe. 1985. Enzyme variation of *T. brucei ssp*. II. Evidence for *T. b. rhodesiense* being a set of variants of *T. b. brucei*. Parasitology **90:**89-100.

Vauterin L. A., and P. Vauterin. 1992. Computer-aided objective comparison of electrophoresis patterns for grouping and identification of microorganisms. Eur. Microbiol. J. **1:**37-41.

Vos, P., M. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Horners, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:**4407-4414.

Waitumbi, J. N., and N. B. Murphy. 1993. Inter- and intra-species differentiation of trypanosomes by genomic fingerprinting with arbitrary primers. Mol. Biochem. Parasitol. **58:**501-517.

Welburn, S. C., E. M. F. Vre, P. G. Coleman, M. Odiit, and I. Maudlin. 2001. Sleeping sickness: a tale of two diseases. Trends Parasitol. **17:**19-24.

Young, C. J., and D. G. Godfrey. 1983. Enzyme polymorphism and distribution of *Trypanosoma congolense* isolates. Annals Trop. Med. Parasitol. **77:**467-481.

Zhang, Z.Q., and T. Baltz. 1994. Identification of *Trypanosoma evansi*, *Trypanosoma. equiperdum* and *Trypanosoma brucei brucei* using repetitive DNA probes. Vet. Parasitol. **53:**197-208.

CHAPTER 4

Multiplex-endonuclease genotyping approach (MEGA): a tool for the fine-scale detection of unlinked polymorphic DNA markers[¶]

Eddy Chukwura Agbo [1], Birgitta Duim [2], Phelix A.O. Majiwa [3], Philippe Büscher [4], Eric Claassen [5], and Marinus F.W. te Pas [1]

[1] Division of Animal Sciences, Institute for Animal Science and Health, (ID-Lelystad), Edelhertweg 15, 8200 AB Lelystad, The Netherlands
[2] Department of Medical Microbiology, Academic Medical Centre, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands
[3] International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya
[4] Institute of Tropical Medicine, Department of Parasitology, Nationalestraat 155, B-2000 Antwerp, Belgium
[5] Department of Immunology, Erasmus University Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Chromosoma, In press[¶]

[¶] Patent pending

4.1. Abstract

Restriction enzyme-detectable polymorphisms have been used for assessing genetic differences and generating informative genetic markers. The most detailed fingerprinting analyses have been obtained using the AFLP technique, which accesses subsets of polymorphisms at one or two restriction sites. To combine increased discriminatory power with stringency of PCR amplification, it would be beneficial to access additional independent restriction sites per analysis, and to amplify subsets of DNA restriction fragments with only one pair of oligonucleotide primers. We have now developed a unique approach that permits the simultaneous use of four or more endonucleases in combination with one pair of adapters/primers, and applied it to genotype 21 trypanosome populations to subspecies level. The approach takes advantage of the fact that some endonucleases create cohesive ends that are compatible with the overhang sites created by other endonucleases. We demonstrate the greater resolution of identifiable polymorphic fragments over conventional ligation-mediated restriction analysis method, and discuss the value of the approach as a tool for fine genetic mapping of *Trypanosoma brucei*. Finally, we propose utility of the method for fine characterisation and for identifying co-dominant genetic markers in a variety of other taxa.

4.2. Introduction

A variety of fingerprinting methods have been developed for accurate assessment of genetic diversity, and to address questions regarding population structure and genetic relatedness among prokaryotic and eukaryotic organisms (see Mueller and Wolfenbarger 1999 for review). The ideal genome fingerprinting method should provide an unbiased estimate of the total genome variance in the population and must be sufficiently sensitive to detect minimal genetic differences and minimise errors due to sampling variance. This is particularly relevant for the comparative analysis

of genomes of closely related organisms where the relatively limited genetic variation between subspecies can only be revealed by the enhanced resolution power of the technique employed.

Trypanosome species are unicellular, protozoan parasites that cause debilitating disease, collectively called trypanosomosis (Kassai 1988), with fatal outcome in man and animals. *Trypanosoma brucei* consists of three subspecies that are indistinguishable by conventional morphological and antigenic criteria, but differ in their geographical distribution and host specificity (Gibson et al., 1980). For epidemiological identification and tracking of the subspecies strains, *T. b. gambiense* can be distinguished from *T. b. brucei* and *T. b. rhodesiense* (Hide et al., 1998; Agbo et al., 2002). However, the distinction between *T. b. brucei* and *T. b. rhodesiense* has always been more difficult to determine (Godfrey et al., 1990; Hide 1999; Agbo et al., 2002). Resistance to normal human serum (equivalent to human infectivity) is an important trait that distinguishes *T. b. gambiense* and *T. b. rhodesiense* (which are resistant) from *T. b. brucei* strains (which are sensitive). We found some association between different genetic marker loci and resistance in molecular signatures of rDNA regions of human- and nonhuman-infective *T. brucei* isolates (Agbo et al., 2001).

Members of *the T. brucei* complex possess a remarkably high degree of genetic identity and are rich in repetitive DNA, a feature that has also been exploited for molecular typing (Hide et al., 1990; Hide et al., 1994). However, this approach is cumbersome and technically demanding. A recently reported serum-resistance-associated (SRA) gene probe (De Greef et al., 1989) adapted as a PCR-based typing system (Welburn et al., 2001; Gibson et al., 2002) offers hope for a rapid tool for identifying human-infective *T. b. rhodesiense* trypanosomes. However, since parasites derived from hosts other than humans potentially contain human- as well as animal-infective strains, the need for high-resolution distinguishing analytical tools that will facilitate the definitive typing of all *T. brucei* subspecies on a routine basis persists.

The conventional AFLP method (Vos et al., 1995) has shown most value for interspecies genotyping and, to a limited extent, for delineating different sets of clonally descended individuals in several plants and fungi (Mueller and Wolfenbarger 1999). A review by Masiga et al. (2000) highlighted possible value for the conventional AFLP in parasite genetic studies but was recently reported to offer low resolution in the genotyping of trypanosome subspecies (Agbo et al., 2002), and previously in closely related strains in other taxa (Lindstedt et al., 2000). Simons et al. (1997) pioneered early attempts at the use of additional endonucleases to increase the discriminatory power of AFLP but reported only slight improvement in fingerprint patterns compared to the standard protocol. Lindstedt and co-workers (2000) generated fragments of DNA with a set of three endonucleases followed by adapter ligation and their amplification using a cognate set of three primers. Such multiplex PCR with three primers is bound to compromise or diminish stringency and robustness of the method. More recently, a triple-endonuclease/two-adapter AFLP approach was shown to be rapid and robust (Van Der Wurff et al., 2000). However, the approach is limited by the fact that only the restriction fragments resulting from two endonucleases not interspersed by a specific restriction site of the third contribute to the detectable polymorphism. In addition, from a conceptual point of view, it can further be argued that since the modification entailed a reduction in the number of potentially amplifiable fragments, this would limit its value as a finer fingerprinting approach over the conventional two-endonuclease method.

We hypothesised that to obtain additional polymorphisms in the genetic analysis of complex and closely-related subspecies, it would indeed be necessary (i) to simultaneously access multiple independent restriction enzyme sites within the genome and (ii) to robustly detect all the resulting heterosite restriction fragments. On this basis, the use of four or more endonucleases to digest genomic DNA would, in principle, provide discrimination at extra sites within the genome, while one pair of adapters and cognate primers would allow for stringent PCR. We describe here the

features of a multiplex-endonuclease/two-adapter genotyping approach that combines increased resolution power with stringency in PCR. We demonstrate that more amplified restriction fragments are generated than with the conventional AFLP method (involving the use of two endonucleases with a pair of adapters) in the analysis of genetic differences among trypanosome isolates. This approach considerably enlarged the number of restriction fragment polymorphisms that was detected in trypanosome genotyping.

4.3. Materials and Methods

4.3.1. Four-endonuclease, two-adapter genetic fingerprinting

Sets of four endonucleases in combination two adapters were used to evaluate the value of the multiplex-endonuclease, two-adapter method as a useful tool for finer DNA amplification fingerprinting. To identify the best set of endonucleases, we first evaluated the approach using seven cloned 'test' *Trypanosoma brucei* isolates belonging to the three subspecies of the parasite - *T. b brucei, T. b. gambiense* and *T. b. rhodesiense* (Table 1). Genomic DNA, isolated essentially as previously described (Heath, 1997), was digested using restriction enzyme combinations**:** (i) *Bgl*II*, Bcl*I*, EcoR*I and *Mun*I; (ii) *Bgl*II*, Bcl*I, *Acs*I and *Mun*I; (iii) *Bcl*I, *Xho*II, *Acs*I and *Mun*I, and; (iv) *Bgl*II, *Xho*II, *EcoR*I and *Acs*I, respectively. All enzymes were purchased from Roche Molecular Biochemicals (Almere, The Netherlands) or Westburg (Leusden, The Netherlands). Since complete digestion of DNA is an important step in the restriction analysis technique, 100 ng genomic DNA was digested for 4 h with 10 U of each endonuclease, essentially in two successive double digestion reactions, with an intermediate 2-propanol precipitation step. The digests were precipitated and reconstituted in 10 µl distilled water. Ten µl of a buffer containing 660 mM Tris HCl, 50 mM $MgCl_2$, 10 mM Dithiothreitol, 10mM ATP, pH7.5, and 20 pM of each *Bgl*II and *Mun*I adapters (Table 2) were added. One µl (400U) of high concentration T4 DNA ligase (New England Biolabs) was added and the

mixture incubated for 2 h at 25ºC. Pre-selective amplification was performed in a total volume of 20 µl containing 1 U of Taq polymerase (Roche Molecular Biochemicals, Almere, The Netherlands), 4 µl of 1:1-diluted ligation product, 2 µl of 10X PCR buffer (100mM Tris HCl pH 9.0, 50 mM KCl, 1% triton X-100, 0.1% w/v gelatin), 2.5 mM $MgCl_2$, 200 µM of each dNTP and 5 pM of each *Bgl*II and *Mun*I primers. The reaction mix was incubated for 2 min at 95ºC, and subjected to 20 cycles of PCR (30 s at 95ºC, 30 s at 56ºC and 2 min at 72ºC). Four µl of 1:20-diluted pre-selective products were used as template for selective reaction in a *Mun*I-0/*Bgl*II-A selective primer combination (with only 'A' selective nucleotide in the *Bgl*II primer). The *Mun*I-0 primer (without any selective nucleotide) was fluorescently labelled with FAM (i.e. 6-carboxyfluorescein). The PCR program was essentially the same as for pre-selective amplification, except that the last cycling step was followed by 30-min incubation at 60ºC. The final products were diluted 1:1 with TE, and Genescan-500 internal lane standard (PE Applied Biosystems) was added. One µl of the mix was resolved in a 7.3% denaturing sequencing gel using a model ABI 373A automated DNA sequencer. Gels were routinely prepared by using ABI protocols and electrophoresed for 5 h. To assess the reproducibility of the approach, two sets of genomic DNA were isolated at separate occasions from the 'test' isolates, and processed according to the described protocol.

4.3.2. Data analysis

The gel patterns were collected with the GenScan software (PE Applied Biosystems) and sample files were transferred to the GelCompar v4.1 software (Applied Maths, Kortrijk, Belgium). Band positions were inferred from the signal peaks in the sample files, and used to establish the total number of bands in each lane. By computer-assisted analysis, samples files were directly scored for presence/absence of signal peaks in the four-endonuclease and conventional approaches. Gels were normalised by using the internal standard that was added to each lane. A coefficient of racial likeness was used to create a similarity matrix. Clustering of the

patterns was performed with the unweighted pair group method using average linkage (UPGMA) (Vauterin and Vauterin 1992).

Table 1 Trypanosome populations used in these studies

Isolate	Source	Year of isolation
Trypanosoma brucei brucei		
AnTat2/2*	Tsetse fly (*Glossina morsitans*); Nigeria	1970
AnTat17/1*	Sheep; Democratic Rep. of Congo	1978
J10*	Hyena; Zambia	1973
T. brucei gambiense		
LiTat1.3*	Human; Cote de Ivoire	1952
PT16*	Human; Cote de Ivoire	1992
T. brucei rhodesiense		
STIB848*	Human; Uganda	1990
AnTat12.1* [+]	Human; Uganda	1971
T. evansi		
RoTat1.2	Water buffalo; Indonesia	1982
AnTat3.1	Capybara; South America	1969
T. equiperdum		
STIB818	Horse; China	1979
AnTat4.1	Unknown; South America	N/A
T. congolense		
C49 (savannah)	Cow, Kenya	1966
Gam2 (savannah)	Cow; The Gambia	1977
IL3900 (riverine/forest)	Dog; Burkina Faso	1980
ANR3 (riverine/forest)	Fly; The Gambia	1988
K45.1 (kilifi)	Cow; Kenya	1982
WG5 (kilifi)	Goat; Kenya	1980
T. simiae		
Ken2	Fly; The Gambia	1988
TsØ2	Bushbuck; Kenya	N/A

* 'Test' samples; [+] Adapted sensitive to normal human serum; N/A - data not available

Table 2. The complementary sets of *Bgl*Il and *Mun*I adapters and PCR primers

Endonuclease	Adapter	Primer core sequence
*Bgl*II	5'-CGG ACT AGA GTA CAC TGT C	5'-GAG TAC ACT GTC GAT CT
(A/GATCT)	3'-C TGA TCT CAT GTG ACA GCT AG	
*Mun*I	5'-AAT TC CAA GAG CTC TCC AGT AC	5'-6-FAM-GAG AGC TCT TGG AAT TG
(C/AATTG)	3'- G GTT CTC GAG AGG TCA TGA T	

4.3.3. Computer predictive modelling

A Monte Carlo simulation program (in QuickBasic language) was used to calculate an estimate of the occurrence of the recognition enzyme sequences. The program randomly generates a genome of adequate large size, to calculate the frequency of the recognition sequences of a set of endonucleases, with diverse combinations and permutations. Increasing the size of the genome and the number of repetitions was used to enhance the precision of the estimate.

4.4. Results and Discussion

4.4.1. Principle of the multiplex-endonuclease, 2-adapter genetic fingerprinting approach

The approach evaluated using four-endonucleases in combination with two adapters was found to be a robust and reproducible technique capable of detecting additional amplifiable fragments to increase the number of detectable polymorphisms between genomes. Sets of four different 6 bp-cutting endonucleases that simultaneously access independent sites were used to digest genomic DNA. On average, the restriction site for a '6-bp

cutter' enzyme occurs every 4096 (4^6) bp, assuming a random distribution of the four bases in DNA. Assuming a random distribution of four different 6-bp-recognising ('rare cutting') enzymes (hypothetically A, B, C and D combination), the possible number of combinations, P of their respective digestion frequency in the genome is (P_{AB} + P_{CD}). If the restriction enzyme cleavage sites are independent of one another (i.e. no shared nucleotides at similar positions), the cleavage site of each endonuclease would be expected to occur approximately every 1024 bp against 2048 bp for a pair of 6-bp cutters. In practice, this will not be the case as will be discussed later.

The four 'rare cutter' enzymes used (*Bgl*II-*Bcl*I/*EcoR*I-*Mun*I) generate sets of fragments with 16 possible restriction ends out of which 10 different sets are potentially amplifiable (Fig. 1). The endonucleases were chosen such that for each pair, the cohesive ends created by one are compatible with the overhang sites created by the other restriction enzyme. On this basis, only one pair of adapters, *Bgl*II and *Mun*I (Table 2) was ligated to allow the amplification of the fragments using a pair of cognate primers in PCR. In this approach, *Bgl*II adapter also ligated to the overhang sites created by *Bcl*I or *Xho*II, while *Mun*I adapter also ligated to *EcoR*I, *Acs*I or *Apo*I sites. The short adapter ligation step ensured a substantial reduction in the amount of time required for preparing the DNA fingerprints. With the *Bgl*II*, Bcl*I*, EcoR*I *and Mun*I combination, following adapter ligation reaction, 5 groups of products result (Fig. 1, A - E). The four 'rare cutter' enzymes used (*Bgl*II-*Bcl*I/*EcoR*I-*Mun*I) generate sets of fragments with 16 possible restriction ends out of which 10 different sets are potentially amplifiable (Fig. 1).

The endonucleases were chosen such that for each pair, the cohesive ends created by one are compatible with the overhang sites created by the other restriction enzyme. On this basis, only one pair of adapters, *Bgl*II and *Mun*I (Table 2) was ligated to allow the amplification of the fragments using a pair of cognate primers in PCR. In this approach, *Bgl*II adapter also

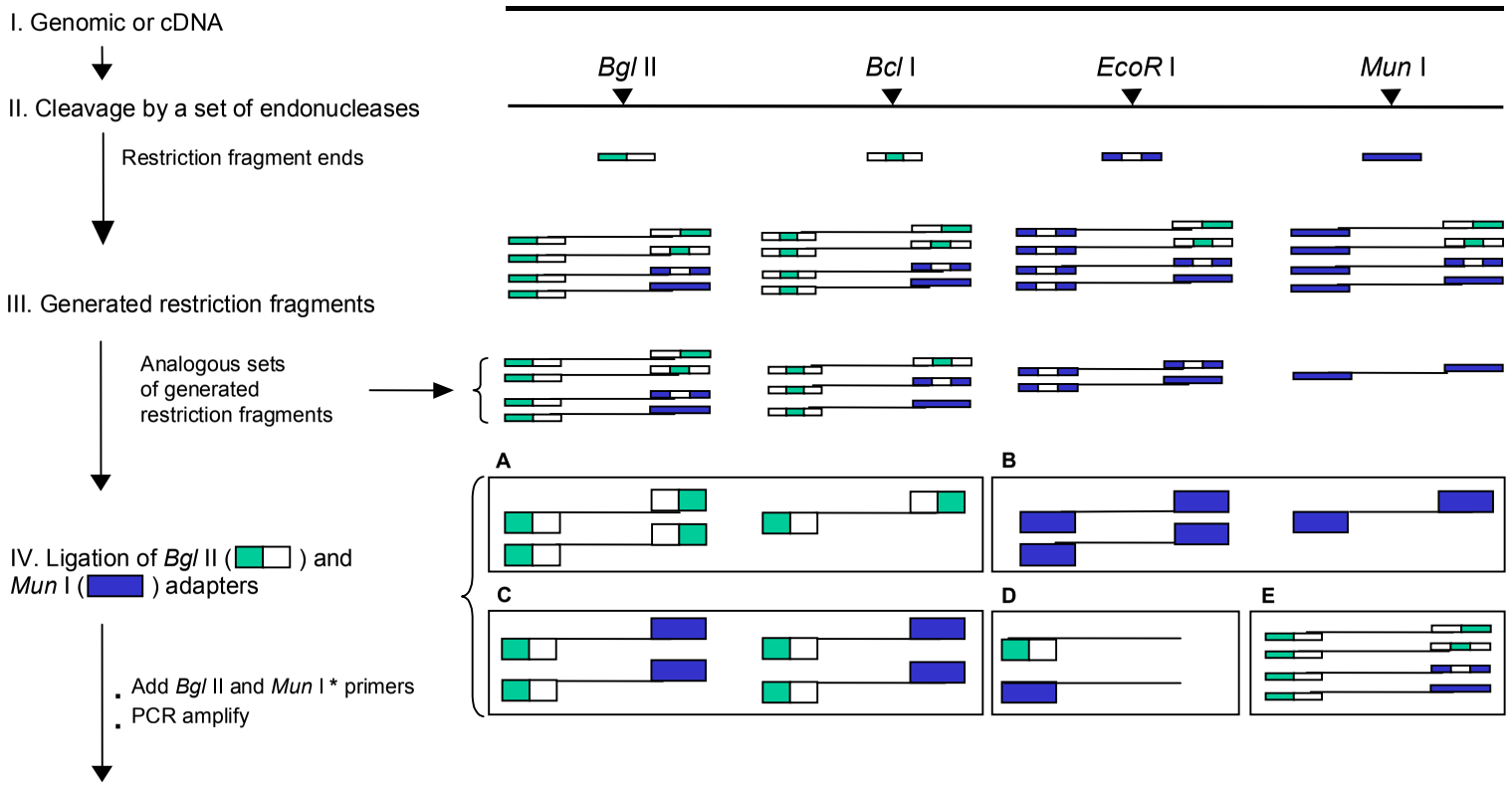
ligated to the overhang sites created by *Bcl*I or *Xho*II, while *Mun*I adapter also ligated to *EcoR*I, *Acs*I or *Apo*I sites. The short adapter ligation step ensured a substantial reduction in the amount of time required for preparing the DNA fingerprints. With the *Bgl*II*, Bcl*I*, EcoR*I *and Mun*I combination, following adapter ligation reaction, 5 groups of products result (Fig. 1, A - E). These consist of sets of fragments with (i) *Bgl*II or *Mun*I adapters at both ends (A, B), (ii) *Bgl*II and *Mun*I adapters at the ends of each fragment (heterosite) (C), (iii) *Bgl*II or *Mun*I adapter at only one end of fragment (D) and, (iv) no adapter (E). The PCR was then used to obtain exponential amplification of the heterosite products (with different adapters at each end). In contrast, DNA fragments with only one adapter undergo linear amplification and are rapidly competed out in PCR. The amplification of DNA with the same adapter at each end (Fig. 1, A and B) is suppressed because self-annealing of inverted repeat adapters inhibits binding of PCR primers. The use of only one set of adapters and primers diminishes competition during PCR, permitting robust and stringent reaction conditions. Repeat of the experiments with samples derived at two separate times from the 'test' isolates gave identical AFLP profiles, with the same bands and same polymorphisms which indicates high reproducibility of the method.

On the other hand, endonucleases that generate blunt end DNA fragments may be used in a method of the approach described here, in combination with a pair of adapters blunt at one end and cohesive on the other. Such adapters may be generated by the action of restriction endonucleases on double-stranded ribonucleotides, or generated by the annealing of two specifically designed single-stranded oligonucleotides that differ in length and that anneal such as to produce a double-stranded ribonucleotide having one blunt end and one overhanging or recessive end.

Fig. 1. Schematic illustration of the fingerprinting method based on four-endonucleases with two-adapter. Digestion of genomic DNA with four different restriction enzymes generated 16 different restriction products (III). Solid lines represent restriction enzyme fragments of DNA. Boxes at the ends of DNA fragments represent the respective restriction ends or cognate adapter. Following ligation reaction, 6 different groups of products result (IV): sets carrying *Bgl*II or *Mun*I adapters at both ends (A, B); sets with *Bgl*II and *Mun*I adapters at the ends of each fragment (heterosite) (C); sets with *Bgl*II or *Mun*I adapter at only one end of fragment (D), and, sets that did not ligate to any adapter (E). For selective PCR amplification, *Mun*I-0/*Bgl*-A selective primer combination was used, with the *Mun*I primer fluorescently labelled to enable the detection of fragments. Note that irrespective of which of the two cognate primers is labelled, the same subset of representation products, comprising the heterosite fragments, is exponentially amplified and detected (IV, C).

For Figure 1, see facing page.

Fig. 1



- **A, B** - No amplification due to primer dimerisation
- **C** - Exponential amplification
- **D** - Linear amplification
- **E** - No amplification due to absence of primer annealing site

* FAM (i.e. 6-carboxyfluorescein)-labelled PCR primer

4.4.2. Comparison of the four- and two-endonuclease data

The efficiency of the two approaches for generating additional polymorphisms for finer genotyping was assessed by directly comparing their fingerprint patterns using trypanosome DNA (Fig. 2). Here, we show that for all the (sub)species analysed, the four-endonuclease approach revealed remarkably more restriction fragments and extra polymorphisms compared to the two-enzyme method, indicating that additional restriction sites were indeed accessed. However, the increase in the number of fingerprint fragments was not linear when compared to the two-endonuclease approach with *Bgl*II and *EcoR*I. In principle, the pair of endonucleases, *Bgl*II (A/GATCT) and *Bcl*I (T/GATCA) when used together to digest the genome would cleave 6-bp palindromic 5'-GATC- sequences if an 'A' or 'T' is present at the 5' end, respectively. Assuming both restriction enzyme sites to be equally distributed in the genome and that all sites are cleaved, this would imply twice as much cleavage as with only one of the two enzymes, with a cutting frequency of one every 2048 ($4^6/2$) bp in the genome.

In reality, this is highly unlikely due to several factors. First, the recognition sequences of the endonucleases may overlap at certain positions, in which case only one of the pair can cleave its site. Taking this into account in the simulation studies, a pair of endonucleases that share the same nucleotides at two or four positions (e.g. *Bgl*II/*Bcl*I and *Bgl*II/*EcoR*I), were predicted to give approximately the same mean cutting frequency, 2269 bp and 2275 bp, respectively. However, in a *Bgl*II-*Bcl*I-*Mun*I-*EcoR*I digestion, a site for any of the four endonucleases is predicted to occur every 1501 bp. Second, there are differential effects of methylation on enzyme sensitivity to substrate. For example, while both *Bgl*II and *Bcl*I recognition sequences completely overlap Dam methylase site GATC, methylation does not block cleavage with *Bgl*II, but it does block cleavage with *Bcl*I, so that restriction may not occur at all *Bcl*I recognition sites. Similarly, *EcoR*I recognition site (G/AATTC) is insensitive to methylation, while the cleavage of *Mun*I (C/AATTG) site is completely blocked by Dam-methylation.

Thirdly, a common limitation of ligation-mediated restriction analyses is that the actual number of amplified fragments is less than that of theoretically amplifiable fragments. This phenomenon would be expanded in the four-endonuclease approach since more restriction fragments are generated than with two endonucleases. Finally, liberal assumptions have been made that all four restriction enzyme recognition sequences are randomly and equally distributed in the genome, and that all restriction sites are cleaved.

4.4.3. Trypanosome species: intra- and inter-species analyses of genetic differences

We previously evaluated the conventional AFLP approach as a tool for generating useful markers for finer characterisation of trypanosome isolates to subspecies level (Agbo et al., 2002). This approach was limited by the fact that only marked size differences based on allelic restriction fragments from two endonucleases could be assessed and, as a result, only few polymorphic markers were detected. An important issue that arises in the context of using genetic markers for classifying individuals relates to the number of fragments and markers needed to provide adequate fingerprint or clustering. The more restriction sites are accessed in the genome, the more informative the resulting representation fragments in finding polymorphisms. Such fine genetic differences may be used for correlating the underlying structural patterns of diversity in a population of interest and to obtain more precise clusters of genotypes. This unique multiplex-endonuclease method presents a tool for expanding the number of derivable fingerprint fragments, in combination with a pair of adapters and cognate primers to ensure stringency in PCR amplification. For the inter- and intra-species analyses of trypanosome isolates (Table 1), the *Bgl*II-*Bcl*I/*EcoR*I-*Mun*I combination was selected on the basis of its reproducibility, even distribution of bands along the gel and number of polymorphic bands detected (Fig. 2, A-F).

Fig. 2. Demonstration of the higher resolution power of the approach. Fingerprints resulting from intra- and inter-species analyses of trypanosome isolates demonstrate greater resolution of the four-endonuclease approach. A, C and E represent dendograms from the samples analysed with the four-endonuclease approach (using *Bgl*II-*Bcl*I/*EcoR*I-*Mun*I). These are directly compared with profiles derived from samples processed with the two endonucleases (*Bgl*II and *EcoR*I), viz B, D and F, respectively. A pair of adapters and primers (*Bgl*II and *Mun*I) was used, respectively to ligate and PCR-amplify the resulting digestion products. The total number of bands from each tested sample is outlined in bold. **(See facing page)**

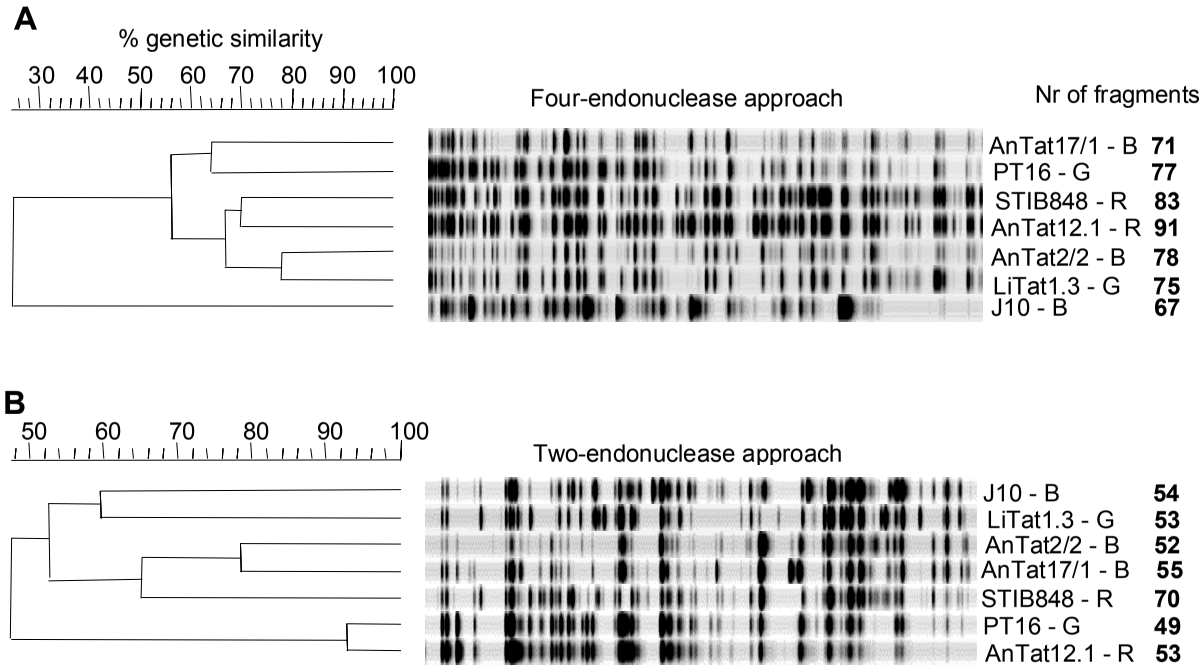

A
% genetic similarity
30 40 50 60 70 80 90 100
Four-endonuclease approach
Nr of fragments
AnTat17/1 - B 71
PT16 - G 77
STIB848 - R 83
AnTat12.1 - R 91
AnTat2/2 - B 78
LiTat1.3 - G 75
J10 - B 67
B
50 60 70 80 90 100
Two-endonuclease approach
J10 - B 54
LiTat1.3 - G 53
AnTat2/2 - B 52
AnTat17/1 - B 55
STIB848 - R 70
PT16 - G 49
AnTat12.1 - R 53

Fig. 2

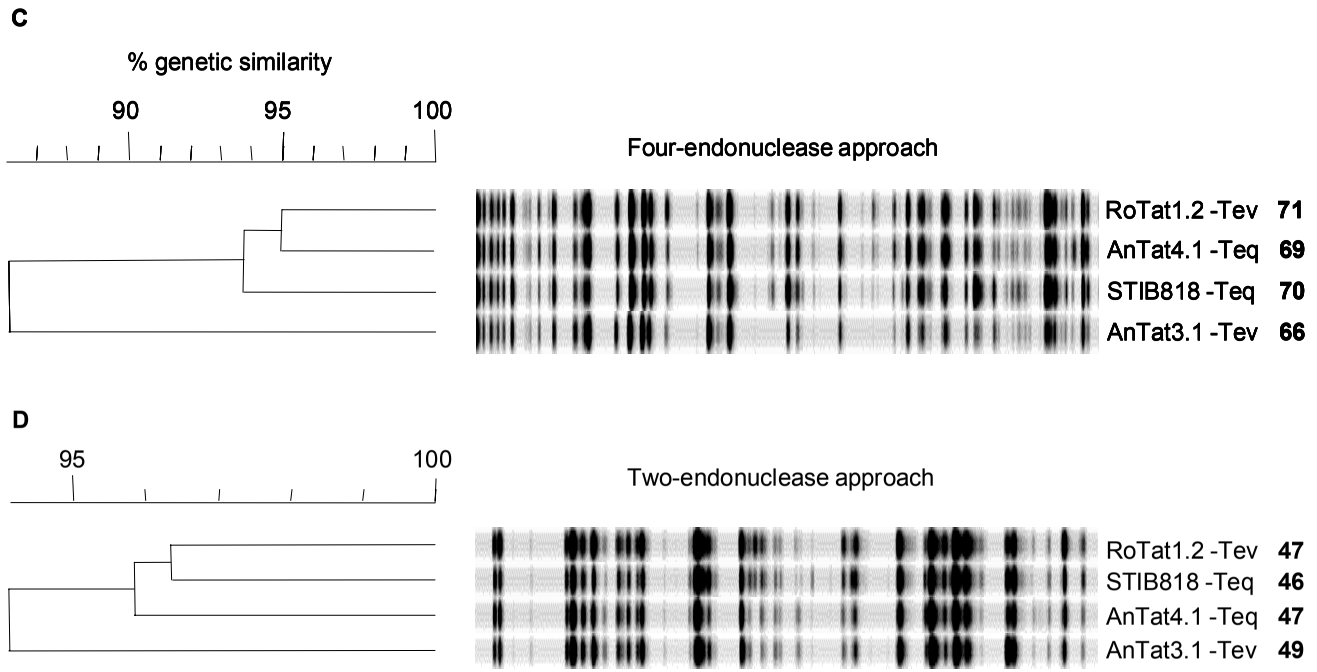
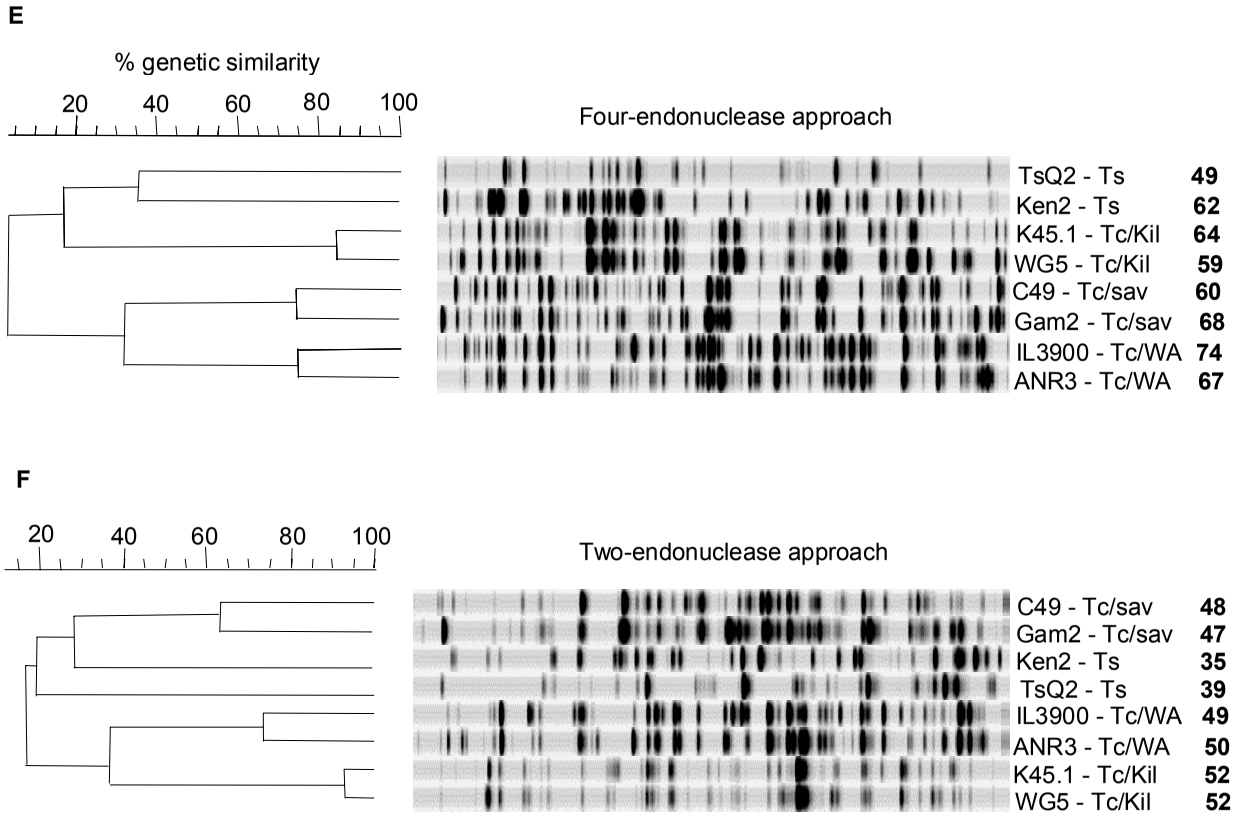

C
% genetic similarity
90
95
100
Four-endonuclease approach
RoTat1.2 -Tev 71
AnTat4.1 -Teq 69
STIB818 -Teq 70
AnTat3.1 -Tev 66
D
95
100
Two-endonuclease approach
RoTat1.2 -Tev 47
STIB818 -Teq 46
AnTat4.1 -Teq 47
AnTat3.1 -Tev 49

Fig. 2


E
% genetic similarity
20
40
60
80
100
Four-endonuclease approach
TsQ2 - Ts 49
Ken2 - Ts 62
K45.1 - Tc/Kil 64
WG5 - Tc/Kil 59
C49 - Tc/sav 60
Gam2 - Tc/sav 68
IL3900 - Tc/WA 74
ANR3 - Tc/WA 67
F
20
40
60
80
100
Two-endonuclease approach
C49 - Tc/sav 48
Gam2 - Tc/sav 47
Ken2 - Ts 35
TsQ2 - Ts 39
IL3900 - Tc/WA 49
ANR3 - Tc/WA 50
K45.1 - Tc/Kil 52
WG5 - Tc/Kil 52

The intra-specific analyses of *Trypanosoma brucei* subspecies with four endonucleases (A) resulted in more restriction fragments and extra polymorphisms compared to the two-endonuclease method (B). Similarly, with the closely related *T. evansi* and *T. equiperdum* species (C and D), a higher number of fragments resulted with the four-endonuclease approach, but less so with the number of identifiable polymorphisms between the species. This underscores the close genetic relatedness of the two species, being probably pathotypes of the same parental strain. The greater discriminatory power of the four-endonuclease approach was best demonstrated in the inter-species analyses of *T. congolense* and *T. simiae* isolates, in which far more polymorphic fragments result (E) than with the two-endonuclease approach (F). These results propose that additional sites within the genome were indeed accessed in the four-endonuclease method. When, for instance, the same region of the genome is accessed in both approaches, there is more coverage with the multi-endonuclease method so that more polymorphic restriction fragments are accessed and detected than with two or three endonucleases.

Nearly all eukaryotes contain non-coding repeat sequences that are under minimal evolutionary pressure. Therefore, since DNA polymorphisms may be concentrated at these regions, they will be highly informative for DNA fingerprinting, genetic mapping and population structure studies. It is thus possible that DNA polymorphisms can be overlooked when only a limited number of endonucleases have been tested with the conventional AFLP. This might explain the success of the four-endonuclease approach over the two-endonuclease conventional method and underline the greater inter-species genetic difference (Fig. 2 E, F) when compared to differences within the members of the *Trypanozoon* subgenus (*T. brucei*, *T. evansi* and *T. equiperdum*). The approach described here offers a useful tool for finer genome characterisation of *T. brucei* subspecies, and could be easily applied for the genotyping of a variety of taxa.

There is currently a great deal of debate regarding the feasibility of identifying genetic associations with complex *T. brucei* human serum response phenotype. It is imperative that the molecular tool employed is a genome-wide approach that accesses multiple independent sites to permit the resolution of fine genetic differences. Large population studies and/or detailed pedigree analysis can then be used to determine if observed anonymous markers either directly contribute to, or are associated with human serum response traits.

For species like trypanosomes that have highly plastic genome composition or size, the more completely the genome is accessed, the greater the chance of uncovering similarities and/or differences in coding and non-coding regions. Using this approach, the complexity of genetic fingerprints can be advantageously managed for finer genetic analysis by increasing the number and/or varying the choice of restriction enzyme pairs. These features make the approach better suited for revealing higher levels of genetic variation than with previous methods (Vos et al., 1995; Lindstedt et al., 2000; Simons et al., 1997; Van der Wurff et al., 2000). For micro-restriction mapping of *T. brucei* subspecies, additional endonucleases in varying combinations and numbers can now be utilised to directly access more genomic regions for genetic markers that may be associated with human resistance/sensitivity.

4.5. Acknowledgements

The authors thank Peter Houba and Frans Gerbens for helpful discussions and comments on the manuscript. We are grateful to Willem Buist for the computer simulation studies and statistical analysis. The Institute for Animal Science and Health (ID-Lelystad) Lelystad, The Netherlands is acknowledged for funding this research, under internally-funded project number: *iSEO* 640.47293.00.

References

Agbo, E. E. C., P. A. O. Majiwa, E. J. H. M. Claassen, and M. H. Roos. 2001. Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterisation. Exp. Parasitol. **99:**123-131.

Agbo, E. C., P. A. O. Majiwa, H. J. H. M. Claassen, and M. F. W. te Pas. 2002. Molecular variation of *Trypanosoma brucei* subspecies as revealed by AFLP fingerprinting. Parasitology **124:**349-358.

De Greef C, G. Imbrechts, G. Matthyssens, N. Van Meirvenne, and R. Hamers. 1989 A gene expressed only in serum-resistant variants of *Trypanosome brucei rhodesiense*. Mol. Biochem. Parasitol. **36:**169-176.

Gibson, W. C., T. F. de C. Marshall, and D. G. Godfrey. 1980. Numerical analysis of enzyme polymorphism**:** a new approach to epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. Adv. Parasitol. **18:**175-246.

Gibson, W., T. Backhouse, and A. Griffiths. 2002. The human serum resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infec. Genet. Evol. **1:**207-214.

Godfrey, D. G., R. D. Baker, L. R. Rickman, and D. Mehlitz. 1990. The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. Adv. Parasitol. **29:**1-74.

Heath, S. 1997. Molecular Techniques in Analytical Parasitology. In: Anlystical Parasitology (ed. Rogan, M.T.), pp 67-68.

Hide, G., P. Cattand, D. Le Ray, J. D. Barry, and A. Tait. 1990. The identification of *Trypanosoma brucei* using repetitive DNA sequences. Mol. Biochem. Parasitol. **39:**213-226.

Hide, G., S. C. Welburn, A. Tait, and I. Maudlin. 1994. Epidemiological relationships of *Trypanosoma brucei* stocks from South East Uganda: evidence for different populations structures in human infective and non-human infective isolates. Parasitology **109:**95-111.

Hide, G., S. Angus, P. H. Holmes, I. Maudlin, and S. C. Welburn. 1998. Comparison of *Trypanosoma brucei* strains circulating in an endemic and an epidemic area of a sleeping sickness focus. Exp. Parasitol. 89**:**21-29.

Hide, G. 1999. History of sleeping sickness in East Africa. Clin. Microbiol. Rev. **12:** 112-125.

Kassai, K. 1988. Standardised nomenclature of animal parasitic disease (SNOAPAD). Vet. Parasitol. **29:**299-326.

Lindstedt, B-A,. E. Heir, T. Vardund, and G. Kapperud. 2000. A variation of the amplified fragment length polymorphism (AFLP) technique using three restriction endonucleases, and assessment of the enzyme combination *Bgl*II-*Mfe*I for AFLP analysis of *Salmonella enterica* subsp. *enterica* isolates. FEMS Microbiol. Lett. **189:**19-24.

Masiga, D. K., A. Tait, and C. M. R. Turner. 2000. Amplified restriction fragment length polymorphism in parasite genetics. Parasitol. Today **16:**350-353.

Mueller, U. G., and L. L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting - a review. Trends Ecol. Evol. **14:**389-394.

Simons, G., T. van der Lee, P. Diergaarde, R. van Daelen, J. Groenendijk, A. Frijters, R. Buschges, K. Hollricher, S. Topsch, P. Schulze-Lefert, F. Salamini, M. Zabeau, and P. Vos. 1997. AFLP-based fine mapping of the Mlo gene to a 30-kb DNA segment of the barley genome. Genomics **44:**61-70.

Van der Wurff A. W., Y. L. Chan, N. M. van Straalen, and J. Schouten. 2000. TE-AFLP: combining rapidity and robustness in DNA fingerprinting. Nucleic Acids Res. **28:**E105.

Vauterin, L. A., and P. Vauterin. 1992. Computer-aided objective comparison of electrophoresis patterns for grouping and identification of microorganisms. Eur. Microbiol. J. **1:**37-41.

Vos, P., M. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:**4407-4414.

Welburn, S. C., K. Picozzi, E. M. Fevre, P. G. Coleman, M. Odiit, M. Carrington, and I. Maudlin. 2001. Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. Lancet **358:**2017-2019.

CHAPTER 5

Population genetic structure and cladistic analysis of *Trypanosoma brucei* isolates

Eddy Chukwura Agbo [1], Peter-Henning Clausen [2], Philippe Büscher [3], Phelix A.O. Majiwa [4], Eric Claassen [5], and Marinus F.W. te Pas [1]

[1] Division of Animal Sciences, Institute for Animal Science and Health (ID-Lelystad), Edelhertweg 15, 8200 AB Lelystad, The Netherlands
[2] Freie Universitaet Berlin, Institute for Parasitology and International Animal Health, Koenigsweg 67, D-14163 Berlin, Germany
[3] Institute of Tropical Medicine, Department of Parasitology, Nationalestraat 155, B-2000 Antwerp, Belgium
[4] International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya
[5] Department of Immunology, Erasmus University Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Submitted for publication

5.1 Abstract

Using a novel multilocus DNA marker analysis method, we studied the population genetic structure of *T. brucei* stocks and derived clones isolated from animal and *rhodesiense* sleeping sickness patients during a national sleeping sickness control program in Mukono district, Uganda. We then performed a cladistic analysis to trace relationships and evolution, using stocks and clones recovered from geographically and temporally matched hosts, including interstrain comparisons with *T. b. gambiense* stocks and clones. Our results show that while there was close genetic relatedness among parasite populations from the same geographical region, microheterogeneities exist between different stocks. Data are presented that indicate that not every human sleeping sickness focus may be associated with a particular human-infective trypanosome strain responsible for long-term stability of the reference focus. We provide evidence of genetic sub-structuring among type 1 *T. b. gambiense* stocks, which has potentially important implications for molecular epidemiology of *T. brucei*.

5.2. Introduction

The species *Trypanosoma brucei* consists of *T. b. brucei* (one of the causes of nagana in cattle), *T. b. rhodesiense*, and *T. b. gambiense* (the cause of human sleeping sickness in East and Central-West Africa, respectively). *Trypanosoma b. brucei* is one of the trypanosomes that have severely limited livestock productivity in over a third of sub-Saharan Africa. It does not cause disease in humans because this subspecies is lysed by normal human serum. The disease, collectively called trypanosomosis, has re-emerged as a major health threat in sub-saharan Africa, with a conservative estimate of about 500,000 current human infections (WHO, 2001).

One of the fundamental tools that is required for a full understanding of the epidemiology of trypanosomosis caused by *Trypanosoma brucei* is a reliable method for identifying and tracking strains. The complexity of the transmission cycles involved and the comparative dynamics of the transmission through parts of these cycles mean that systems for generating robust and reliable genetic markers for epidemiological studies are needed. It is now well established that *T. b. rhodesiense* and *T. b. brucei* genomes are more polymorphic than those of *T. b. gambiense* (Hide, 1999; Agbo et al., 2002). However, knowledge of the extent of genetic diversity of local parasite populations and the role of this diversity in the interactions of the parasite with both the animal and human hosts in related and distant populations remain poorly understood. The epidemiological analysis undertaken on stocks isolated during a human sleeping sickness epidemic in Tororo district of Uganda (Hide et al., 1994), as well as other studies MacLeod et al., 1999; MacLeod et al., 2001a; 2001b) provided unique opportunities to address these issues, because a number of parameters that may provide insight into population structure of natural populations of trypanosomes, were investigated.

In East Africa, certain geographically-related foci are characterised by periods of long-term endemicity interspersed with short epidemic episodes. To better understand the causes of these episodes, the molecular epidemiology and population structure of *T. brucei* within Busoga focus have been studied (MacLeod et al., 2000; Enyaru et al., 1993; Enyaru et al., 1997; Degen et al., 1995). A recently developed PCR system based on the serum-resistance-associated (SRA) gene (De Greef et al., 1989) has been shown to be specific for the identification of *T. b. rhodesiense* strains (Welburn et al., 2001; Gibson et al., 2002). However, the absence of the SRA gene from other *T. brucei* subspecies, even in the so-called *rhodesiense*-like, virulent or type 2 *T. b. gambiense* (Gibson, 1986) underscores the need for additional genetic markers.

We undertook a population-genetic study to evaluate the population structure of parasite stocks isolated during endemic and epidemic periods within Mukono district in Busoga focus, using a recently described multilocus fine genotyping marker system (Agbo et al., In press). An attractive feature of the approach is that multiple independent restriction enzyme-based polymorphisms can be genotyped in a single reaction and scored in a single lane of a gel on an automated sequencer. In this analysis, we looked for evidence of the uniqueness of circulating genotypes. Furthermore, we traced stock relationships and evolution by analysing the extent of genetic polymorphisms among human-infective stocks on the one hand, and between human- and animal infective stocks on the other hand, from geographically and temporally matched populations, within the same and different foci. Our data suggest that this approach may offer a valuable tool for fine-scale epidemiological investigations of trypanosomosis and diseases caused by other agents.

5.3. Materials and Methods

5.3.1. Trypanosome stocks and clones

The trypanosome populations listed in Table 1 were originally isolated between 1990-1992 from pigs (19), cattle (5) and *rhodesiense* sleeping sickness patients (4) during an evaluation of the National Sleeping Sickness Control Programmme (NSSCP) in Bulutwe, Mukono district, South-Eastern Uganda, a *rhodesiense* sleeping sickness endemic area. At that time 0.7% of villagers, 33.5% of cattle and 52.8% of domestic pigs harbored trypanosome infections (Nowak et al., 1992). Human serum response properties of the parasites were evaluated earlier (Von Dobschuetz, 2002; Mangeni, unpublished data) using the Blood Incubation Infectivity Test (BIIT) (Rickman and Robson, 1980) and the Human Serum Resistance Test (HSRT) (Jenni and Brun, 1982).

A collection of stocks and clones consisting of *T. b. brucei* (7), *T. b. gambiense* (10) and *T. b. rhodesiense* (8) derived during epidemic episodes from related and disparate locations (Table 2) was also analysed to compare the genotypic properties of the various populations, separated in space and time. Parasite cloned populations were generated following limited passage of cryostabilates in laboratory rodents, according to published protocols (Hawking, 1976; Brun et al., 1981). Genomic DNA from all samples was extracted according to Heath (1997). The DNA samples were screened by PCR for the presence of the serum-resistance-associated (SRA) and *T. b. gambiense*-specific glycoprotein (TgsGP) genes, exactly as described by Gibson et al. (2002) and Radwanska et al. (2002), respectively.

5.3.2. Multiplex-endonuclease analysis

Multiplex genetic fingerprint patterns were generated for each sample according to the principle described by Agbo et al. (In press), which permits the simultaneous assessment of multiple independent polymorphic sites per genotyping analysis and ensures PCR stringency through the use of only one pair of adapters and primers. Briefly, 200-300 ng of genomic DNA was digested for 4 h with 10 U of each restriction enzyme, in combinations *Bgl*II-*Bcl*I-*EcoR*I-*Mfe*I or *Bgl*II-*Bcl*I-*Xho*II-*EcoR*I-*Mfe*I-*Acs*I, respectively. The digests were precipitated and reconstituted in 10 µl distilled water. Ten µl of a buffer containing 660 mM Tris HCl, 50 mM $MgCl_2$, 10 mM dithiothreitol, 10mM ATP, pH7.5, and 20 pM of each adapter - *Bgl*II ($^{5'-}$CGGACTAGAGTACACTGTC; $^{5'-}$GATCGACAGTGTACTCTAGTC) and *Mun*I ($^{5'-}$AATTCCAAGAGCTCTCCAGTAC; $^{5'-}$AGTACTGGAGAGCTCTTG) - were added. One µl (400U) of high concentration T4 DNA ligase (New England Biolabs) was added and the mixture incubated for 2 h at 25ºC. Pre-selective amplification was performed in a total volume of 20 µl containing 1 U of Taq polymerase (Roche Molecular Biochemicals, Almere, The Netherlands), 4 µl of 1:1-diluted ligation product, 2 µl of 10X PCR buffer (100mM Tris HCl pH 9.0, 50 mM KCl, 1% triton X-100, 0.1% w/v gelatin), 2.5 mM $MgCl_2$, 200 µM of each

dNTP and 5 pM of each primer - *Bgl*II ($^{5'-}$GAGTACACTGTCGATCT) and *Mun*I ($^{5'-}$GAGAGCTCTTGGAATTG). The reaction mixture was incubated for 2 min at 95ºC, and subjected to 20 cycles of PCR (30 s at 95ºC, 30 s at 56ºC and 2 min at 72ºC). Four µl of 1:20-diluted pre-selective products were used as template for selective primer combinations *Bgl*II-0*/Mfe*I-A (with zero and an 'A' selective nucleotide in the *Bgl*II and *Mfe*I primers, respectively), *Bgl*II-0*/Mfe*I-AA and *Bgl*II-0*/Mfe*I-AT. The PCR program, electrophoresis of selective sample products, and data collection and analysis were as previously described (Agbo et al., 2002).

Table 1. Origins of trypanosome populations isolated during and endemic period and identity according to BIIT, HSRT, SRA-PCR and TgsGP19-PCR.

[a] Blood Incubation Infectivity Test (Rickman and Robson, 1970); °Tietjen; personal communication; * Mangeni (unpublished data); n.d.: not done.

[b] Human Serum Resistance Test (Jenni and Brun, 1982); ** von Dobschuetz (2002); The HSRT was performed twice per isolate, except for BOT/BU 623/7 and BOT/BU 602/7 where it was tested three and seven times, respectively (Rickman and Robson, 1970). The isolates were defined as resistant "Res" if they showed continuous growth in the presence of human serum for at least 10 days and sensitive "Sens" if they were lysed within 3 days. Isolates which showed non-continuous growth but remained alive in the presence of human serum for at least 3 days were classified as sub resistant "Subres" (Von Debschuetz, 2002).

[c] Serum-resistance associated (SRA) gene (present + / not present -), as determined using primers defined by Gibson et al (2002).

[d] *T. b. gambiense*-specific glycoprotein (TgsGP) gene (present + / not present -), established using primers defined by Radwanska et al. (2002).

Stocks/Clones	Origin	Isolation date	Host	BIIT[a]	HSRT[b]	SRA[c]	TgsGP[d]
1. SUS/BU 83/6	Bulutwe	Mar. 1992	Pig	Sens°	Sens**	-	-
2. SUS/BU 83/7	Bulutwe	Apr. 1992	Pig	Sens°	Sens**	-	-
3. SUS/BU 83/9	Bulutwe	Jul. 1992	Pig	n.d.	Sens**	-	-
4. SUS/BU 83/9 Cl.1	Bulutwe	Jul. 1992	Pig	Subres*	Sens**	-	-
5. SUS/BU 83/9 Cl.2	Bulutwe	Jul. 1992	Pig	Sens*	Sens**	-	-
6. SUS/BU 83/9 Cl.4	Bulutwe	Jul. 1992	Pig	Subres*	Sens**	-	-
7. SUS/BU 83/9 Cl.5	Bulutwe	Jul. 1992	Pig	Subres*	Sens**	-	-
8. SUS/BU 132/2	Bulutwe	Mar. 1991	Pig	n.d.	Sens**	-	-
9. SUS/BU 132/4	Bulutwe	Aug. 1991	Pig	n.d.	Subres**	-	-
10. SUS/BU 139/2	Bulutwe	Mar. 1991	Pig	Res*	Subres**	-	-
11. SUS/BU 169/4	Bulutwe	Aug. 1991	Pig	Sens°	Subres**	-	-
12. SUS/BU 319/7	Bulutwe	Apr. 1992	Pig	Subres* /sens°	Sens**	-	-
13. SUS/BU 319/7 Cl.1	Bulutwe	Apr. 1992	Pig	Sens*	Sens**	-	-
14. SUS/BU 319/7 Cl.3	Bulutwe	Apr. 1992	Pig	Subres*	Sens**	-	-
15. SUS/BU 319/9	Bulutwe	Jul. 1992	Pig	Sens°	Sens**	-	-
16. SUS/BU 347/7	Bulutwe	Apr. 1992	Pig	Res°	Res**	-	-
17. SUS/BU 373/7	Bulutwe	Apr. 1992	Pig	Res°	Res**	-	-
18. SUS/BU 561/3	Bulutwe	Jun. 1991	Pig	Subres* /sens°	Sens**	-	-
19. SUS/BU 932/7	Bulutwe	Apr. 1992	Pig	Subres*	Subres**	-	-
20. BOT/BU 483/2	Bulutwe	Mar. 1991	Cattle	Subres*	Sens**	-	-
21. BOT/BU 492/2	Bulutwe	Mar. 1991	Cattle	n.d.	Sens**	-	-
22. BOT/BU 602/7	Bulutwe	Apr. 1992	Cattle	Subres* /res°	Res**	-	-
23. BOT/BU 623/7	Bulutwe	Apr. 1992	Cattle	Res°	Subres**	-	-
24. BOT/BU 1845/7	Bulutwe	Apr. 1992	Cattle	Res°	Subres**	-	-
25. HOM/BU H1	Bulutwe	Nov. 1990	Man	n.d.	Res**	+	-
26. HOM/BU H2	Bulutwe	Nov. 1990	Man	n.d.	Res**	+	-
27. HOM/BU H5	Bulutwe	Apr. 1991	Man	n.d.	Res**	+	-
28. HOM/IG 2602	Kapyanga	Feb. 1990	Man	Res°	Res**	+	-

5.3.3. Phylotyping by multiplex-endonuclease analysis

Genomic fingerprint patterns were generated in parallel for 13 reference populations (i.e. 5 *T. b. brucei*, 6 *T. b. gambiense* and 2 *T. b. rhodesiense*). Only fragments ranging from 35-500 bp were analysed from which a schematic representation of the fingerprint patterns was constructed. From these profiles, identified on the basis of their intensity and individuality, a numerical matrix of observations based on the presence (1) and absence (0) of bands was built. The data were compared using Pearson correlation product moment coefficient (Pearson, 1926), which determined the proportion of mismatched bands between samples. Based on the similarity matrix a dendrogram was generated using the unweighted pair-group method using arithmetic averages (UPGMA).

5.4. Results

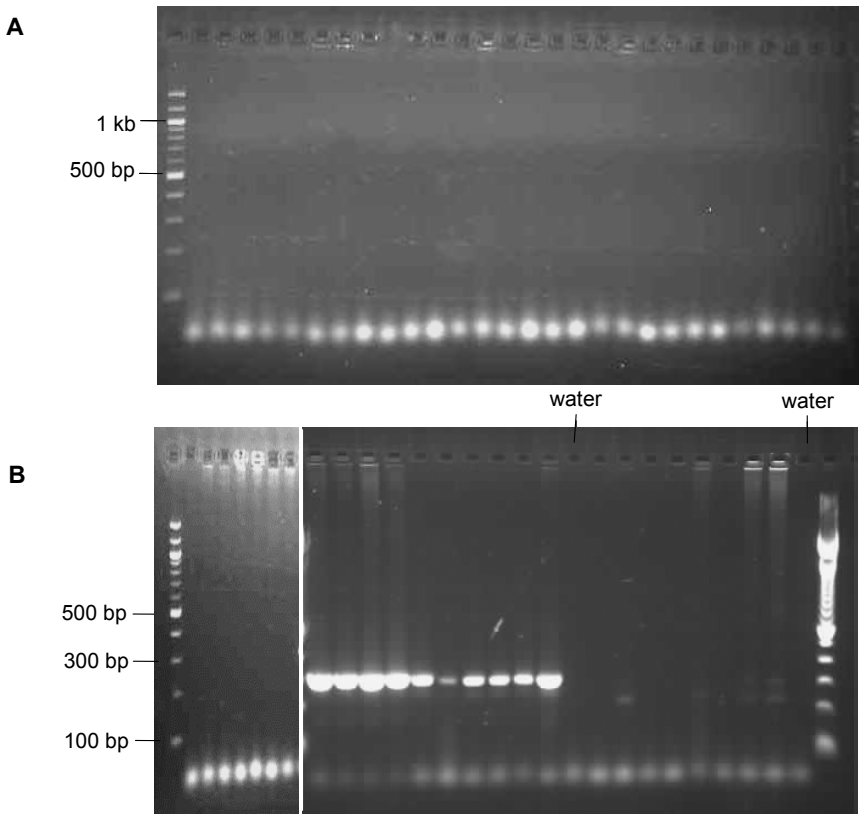
5.4.1. Trypanosome identity

The results of the BIIT and HSRT (summarised in Table 1) permitted an early classification of the stocks or clones derived during the endemicity survey in Mukono district. Trypanosomes positive (resistant) in one of these tests are considered putative *T. b. rhodesiense*. All trypanosome DNAs were further screened by PCR for a 1.2-kb SRA fragment, scored as present or absent (+ or -) (Table 1), and for a 308-bp TgsGP product (Fig. 1). The SRA gene fragment was amplified only from endemic trypanosome populations of human origin, which are considered genuine *T. b. rhodesiense*. All the samples identified as ‘sub-resistant’ in the HSRT were negative for the SRA-specific PCR product. The specific TgsGP19 fragment was generated from only the *T. b. gambiense* populations.

Table 2. Origin and identity of trypanosome populations derived during epidemic outbreaks.

Species Subspecies	Trypanosome stocks and clones	Origin	Isolation year	Original host
1. *T. brucei brucei*	AnTat1.8	Uganda	1966	Bushbuck
2. *T. brucei brucei*	AnTat2.2	Nigeria	1970	Tsetse
3. *T. brucei brucei*	AnTat5.2	Gambia	1975	Bovine
4. *T. brucei brucei*	AnTat17.1	D.R. Congo	1978	Sheep
5. *T. brucei brucei*	Ketri2494	Kenya	1980	Tsetse
6. *T. brucei brucei*	J10	Zambia	1973	Hyena
7. *T. brucei brucei*	TSW196	Côte d'Ivoire	1978	Pig
8. *T. brucei gambiense*	AnTat9.1	Cameroon	1976	Man
9. *T. brucei gambiense*	LiTat1.3	Côte d'Ivoire	1952	Man
10. *T. brucei gambiense*	AnTat11.17	D.R. Congo	1974	Man
11. *T. brucei gambiense*	AnTat22.1	Congo/Brazzville	1975	Man
12. *T. brucei gambiense*	JUA	Cameroon	1979	Man
13. *T. brucei gambiense*	BAGE	D.R. Congo	1995	Man
14. *T. brucei gambiense*	NABE	D.R. Congo	1995	Man
15. *T. brucei gambiense*	PAKWE	D.R. Congo	1995	Man
16. *T. brucei gambiense*	SEKA	D.R. Congo	1995	Man
17. *T. brucei gambiense*	PT312	Côte d'Ivoire	1992	Man
18. *T .brucei rhodesiense*	0404	Rwanda	1970	Man
19. *T. brucei rhodesiense*	STIB847	Uganda (Busoga)	1990	Man
20. *T. brucei rhodesiense*	STIB848	Uganda (Busoga)	1990	Man
21. *T. brucei rhodesiense*	STIB849	Uganda (Busoga)	1991	Man
22. *T. brucei rhodesiense*	STIB850	Uganda (Busoga)	1990	Man
23. *T. brucei rhodesiense*	STIB851	Uganda (Tororo)	1990	Man
24. *T. brucei rhodesiense*	STIB882	Uganda	1993	Man
25. *T. brucei rhodesiense*	STIB883	Uganda	1994	Man

Fig. 1. *T. b. gambiense*-specific TgsGP19-PCR analysis.
DNA marker (M) is a 100-bp ladder. Sample series A, lanes 1-28 correspond to DNA from the trypanosomes listed in Table 1 (1-28), while lanes labelled 29-53 of series B correspond to samples 1-25 listed in Table 2.



5.4.2. Complexity of T. brucei populations during an endemic period

According to the composite *Bgl*II-*Bcl*I-*EcoR*I-*Mfe*I (four-endonuclease) analysis pattern for the samples from Mukono district (Table 1), using *Bgl*II-0/*Mfe*I-A primer combination, the collection consisted of 3 phylogenetic Clusters A, B and C (Fig. 2). The pattern analysis of the control samples (data not shown), as well as human serum response trait and presence of SRA gene product, revealed that each cluster consisted of both *T. b. brucei* and *T. b. rhodesiense*. Also, the genetic relatedness and thus the distribution of the samples in the dendogram did not seem to correlate with serum response properties, and samples with the same serum response trait are not necessarily more closely-related than stocks and clones with a different trait. For instance, human serum resistant SUS/BU 347/7 and SUS/BU 373/7 isolated from pig (by definition putative *T. b. rhodesiense*) share a genetic similarity level of 96%. On the other hand, *T. b. rhodesiense* HOM/BU H1 and HOM/BU H2 from different human hosts, which share 96.5% genetic similarity are 93% similar to HOM/BU H5 (Cluster B). Furthermore, HOM/IG 2602 from human in another village is distantly outplaced from other *T. b. rhodesiense* (to Cluster C). The similarity levels between different populations, as determined by numerical analysis of fingerprint patterns, were shown to span a continuous range of values between 79-98% (Fig. 2), with a dendogram (cophenetic) correlation of 88.4%. However, within each cluster, a highly similar genotype pattern was obtained. Overall, genotype of *T. b. rhodesiense* stocks from man share specific bands, which are different from other stocks (see Boxes in Figures 2 and 3).

Since the identity of populations is highly dependent upon the resolution power of the molecular tool employed, the samples were further processed using a combination of six endonucleases and a pair of adapters. The amplified representation fragments and generated fingerprint data were similarly analysed as described for four-endonuclease procedure (Fig. 3). The approach consistently generated additional restriction fragments to permit for finer genome analysis, however, selection at both fragment ends

seemed necessary for generating discrete fragments for further analysis. Numerical analysis of 6-endonuclease fingerprint of the Mukono samples revealed two distinct clusters (Fig. 3). Cluster I comprised of a group of 7 *T. b. brucei* stocks and clones (which share genetic similarity of more than 90%) and 4 *T. b. rhodesiense* stocks which are 77.8%, 78.5, 80% and 85% related to other *T. b. brucei* stocks. Cluster II sub-divided into two subclusters comprising 3 *T. b. rhodesiense* stocks of human origin (A) and 13 *T. b. brucei or T. b. rhodesiense* stocks and clones (B), which share a correlation coefficient of 75.5%. Both clusters share an overall similarity that ranged between 73.5-98%, with a dendogram (cophenetic) correlation of 95% (Fig. 3). Overall, comparing the dendograms (Figures 2 and 3), analysis with 6 endonucleases revealed less genetic similarity between stocks or clones than with 4 enzymes.

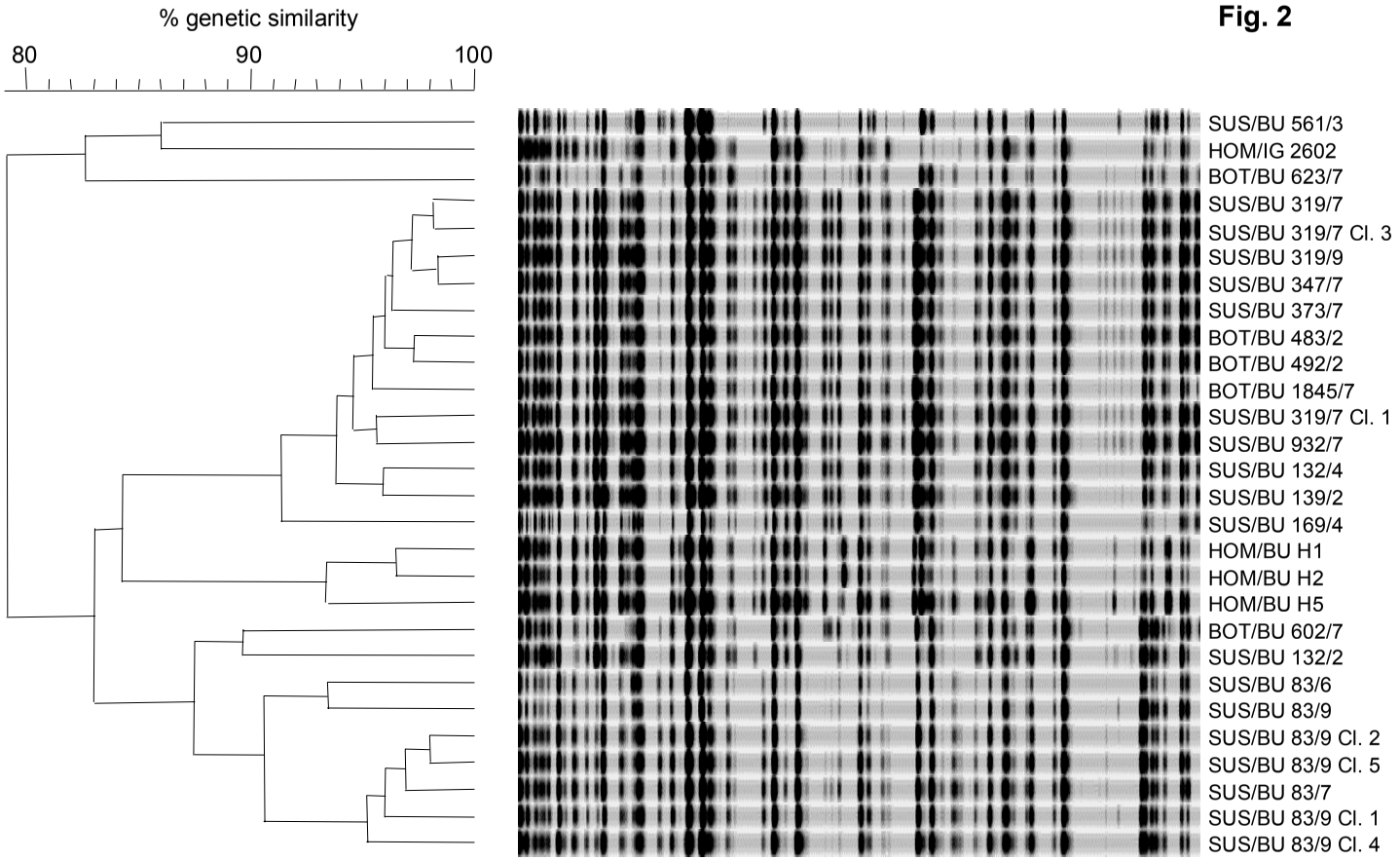
5.4.3. Cladistic analysis

Genetic relatedness of the samples was evaluated on populations isolated during periods of low circulating parasitism (endemic) and upsurge in disease incidence (epidemic). Fingerprint patterns of the stocks isolated in Mukono district, Uganda between 1990-1992 were compared to *T. b. rhodesiense* strains isolated during 1990-1994 epidemic episodes in the same Busoga focus. Relationships between the sets of populations inferred by numerical analysis of fingerprint data were expressed as percentage values of the Pearson correlation product moment coefficient. From the dendogram (Fig. 4), the two groups comprising endemic and epidemic stocks and clones were clearly separated and showed a percent genetic similarity coefficient of 78%. On the basis of percent genetic similarity, the dendogram was stratified into three groups. Window I represents a group of populations with 79-89% genetic similarity and considered to be ‘closely-related’. Fingerprints in Window II comprise of stocks or clones that share 90-95% genetic similarity (‘highly-related’), while Window III displays stocks or clones with more than 95% genetic similarity (considered ‘identical’). On the basis of this classification, it can be seen that *T. brucei* populations circulating in Mukono district during the endemicity survey were closely-

related (at 82.8% for most samples, and in the range 79-82.8% for all samples). However, these are slightly different from those isolated during the epidemic within the same Busoga focus (which are 77.8% related).

Fig. 2. Dendogram from four-endonuclease analysis of 28 *Trypanosoma brucei* populations isolated during endemic period (1990-1992), based on similarity relationships, according to composite genomic patterns. Cluster A corresponded with *T. b. brucei* populations, whereas Clusters B and C each consisted of *T. b. brucei* and *T. b. rhodesiense* populations.

Fig. 2

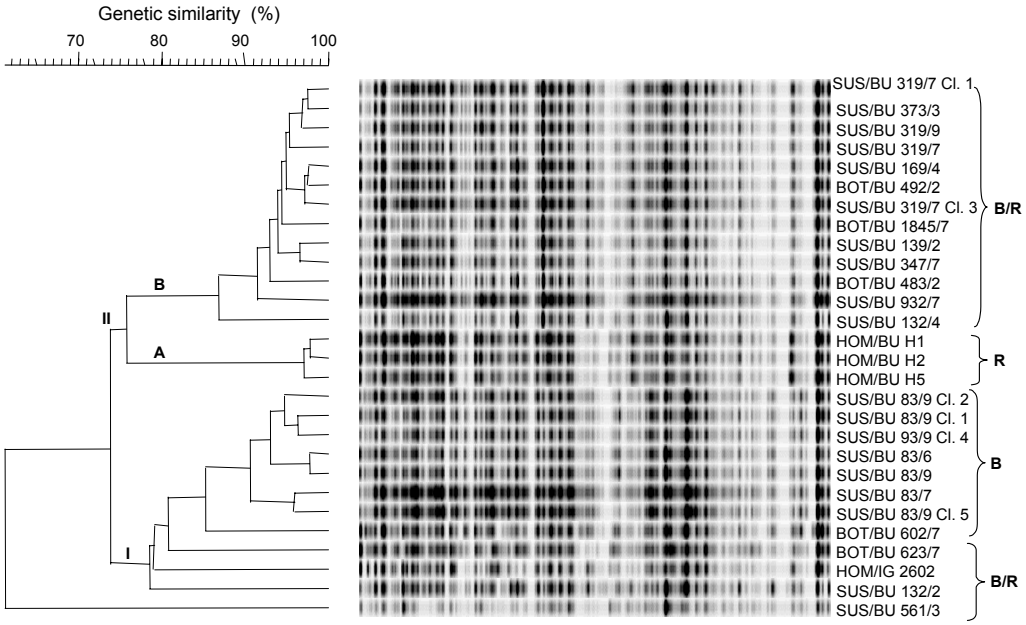

% genetic similarity
80
90
100
SUS/BU 561/3
HOM/IG 2602
BOT/BU 623/7
SUS/BU 319/7
SUS/BU 319/7 Cl. 3
SUS/BU 319/9
SUS/BU 347/7
SUS/BU 373/7
BOT/BU 483/2
BOT/BU 492/2
BOT/BU 1845/7
SUS/BU 319/7 Cl. 1
SUS/BU 932/7
SUS/BU 132/4
SUS/BU 139/2
SUS/BU 169/4
HOM/BU H1
HOM/BU H2
HOM/BU H5
BOT/BU 602/7
SUS/BU 132/2
SUS/BU 83/6
SUS/BU 83/9
SUS/BU 83/9 Cl. 2
SUS/BU 83/9 Cl. 5
SUS/BU 83/7
SUS/BU 83/9 Cl. 1
SUS/BU 83/9 Cl. 4

Although several of the stocks isolated during the endemic survey were identical (clonal), there were major genetic differences within the Mukono *T. b. rhodesiense* and *T. b. brucei* stocks, respectively. The genotype profiles indicate that most endemic stocks were genetically separated at a similarity coefficient of 82.8%, and 79% for the stocks - SUS/BU561/3, HOM/IG 2602 and BOT/BU 623/7 (Fig. 4). The analysis was then extended to include the genotype of *T. b. brucei* stocks circulating during various epidemic periods in disparate geographical regions. From the dendogram (Fig. 5), three broad groups result in which the samples isolated during the endemic survey are clustered together but share genetic similarity of only 79%. As expected, the *T. b. rhodesiense* stocks isolated during the 1990-1992 epidemic in Busoga were more genetically-related to the *T. b. rhodesiense* stocks from the endemicity survey in Mukono (also within Busoga focus) than the disparate epidemic period *T. b. brucei* populations. A genetic correlation coefficient of 77.8% indicates that the genetic composition of *T. b. brucei* populations isolated during the epidemic period were remarkably different from those of other *T. b. brucei* isolated from the same focus during endemic periods.

When the Mukono samples were compared to all disparate stocks from epidemic episodes, the dendogram correlation showed essentially two clusters (Fig. 6). Cluster P comprises of disparate *T. b. brucei* and *T. b. gambiense* populations derived outside of Busoga focus during epidemic periods, while populations from Busoga focus form Cluster Q. Each cluster is subgrouped into two subclusters differentiated on the basis of the subspecies identity of the samples. Subclusters P1 and P2 comprise of disparate *T. b. gambiense* and *T. b. brucei* stocks and clones, respectively. While the samples comprising the former are genotypically not homogeneous, the latter are most heterogeneous. In addition, the *T. b. gambiense* stocks and clones from Cameroon, Congo and Democratic Republic of Congo are substructured into 2 groups (including LiTat 1.3 from Côte d'Ivoire) that share a genetic relatedness of 86.8%. This may also suggest that different genotypes were responsible for the 1995 outbreaks in

Democratic Republic (D.R) of Congo. On the other hand, Ketri 2494, isolated in Kenya (1980) from tsetse fly, grouped within Cluster Q. It is interesting that Ketri 2494 (from Kenya) is sub-clustered in Q-1, while AnTat 1.8 also a *T. b. brucei* isolated from Uganda (a closely-related geographical area) did not fall within this group (Fig. 6). This indicates that there is close genetic relatedness of Ketri 2494 to the epidemic *T. b. rhodesiense* stocks from Uganda. While the samples within cluster P share a coefficient genetic relatedness of only 76.5%, the subgroups comprising cluster Q share genetic similarity of 82.5%. Populations within both clusters share only 74% genetic relatedness.

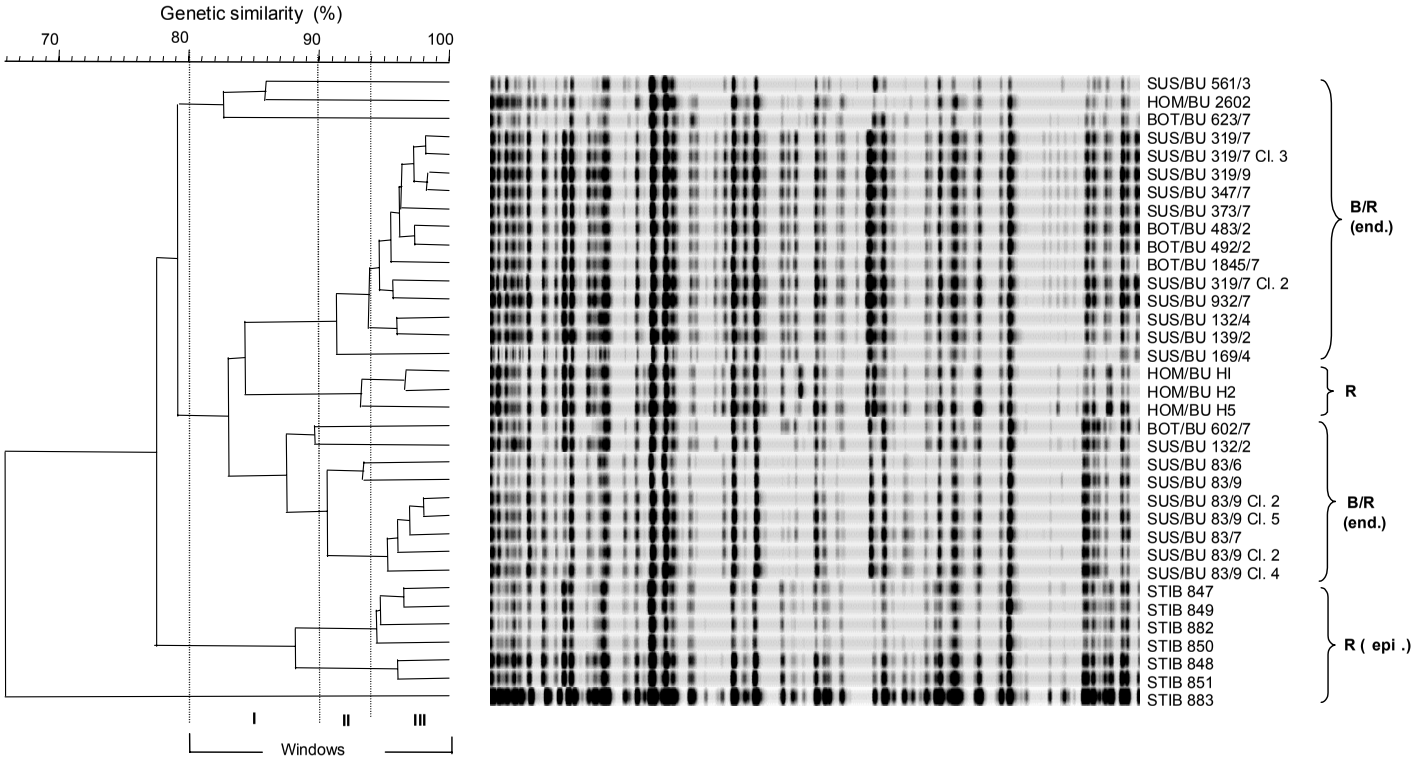
Fig. 3. Relationship between 28 *T. brucei* stocks or clones derived during endemic period survey between 1990-1992, inferred by numerical analysis of fingerprint data generated by six-endonuclease analysis. Subspecies identification 'B' and 'R' represent *T. b. brucei* and *T. b. rhodesiense*, respectively.



As expected, intra-subspecies relatedness of the disparate *T. b. brucei* stocks derived during epidemic episodes was less than for those derived during the endemic survey in Mukono district. As in previous clustering correlations (Figs. 4 and 5), STIB 883 and now PT312 were classified as outliers in the dendogram. Their correlations to other populations were 64.8% and 59%, respectively, indicating distant genetic relatedness. In addition, it shows that the STIB 883 was completely different from other stocks circulating during the epidemic in Busoga focus and suggests that new parasite strains that upset host-parasite balance may have an important role in initiating an epidemic episode.

Fig. 4. Assessment of relationship between stocks isolated during endemic and epidemic periods in Busoga focus, Uganda. The extent relatedness of *T. b. brucei* (B) and *T. b. rhodesiense* (R) from epidemic (epi) and endemic (end) periods are grouped in Windows I, II and III. (See facing page).

Fig. 4

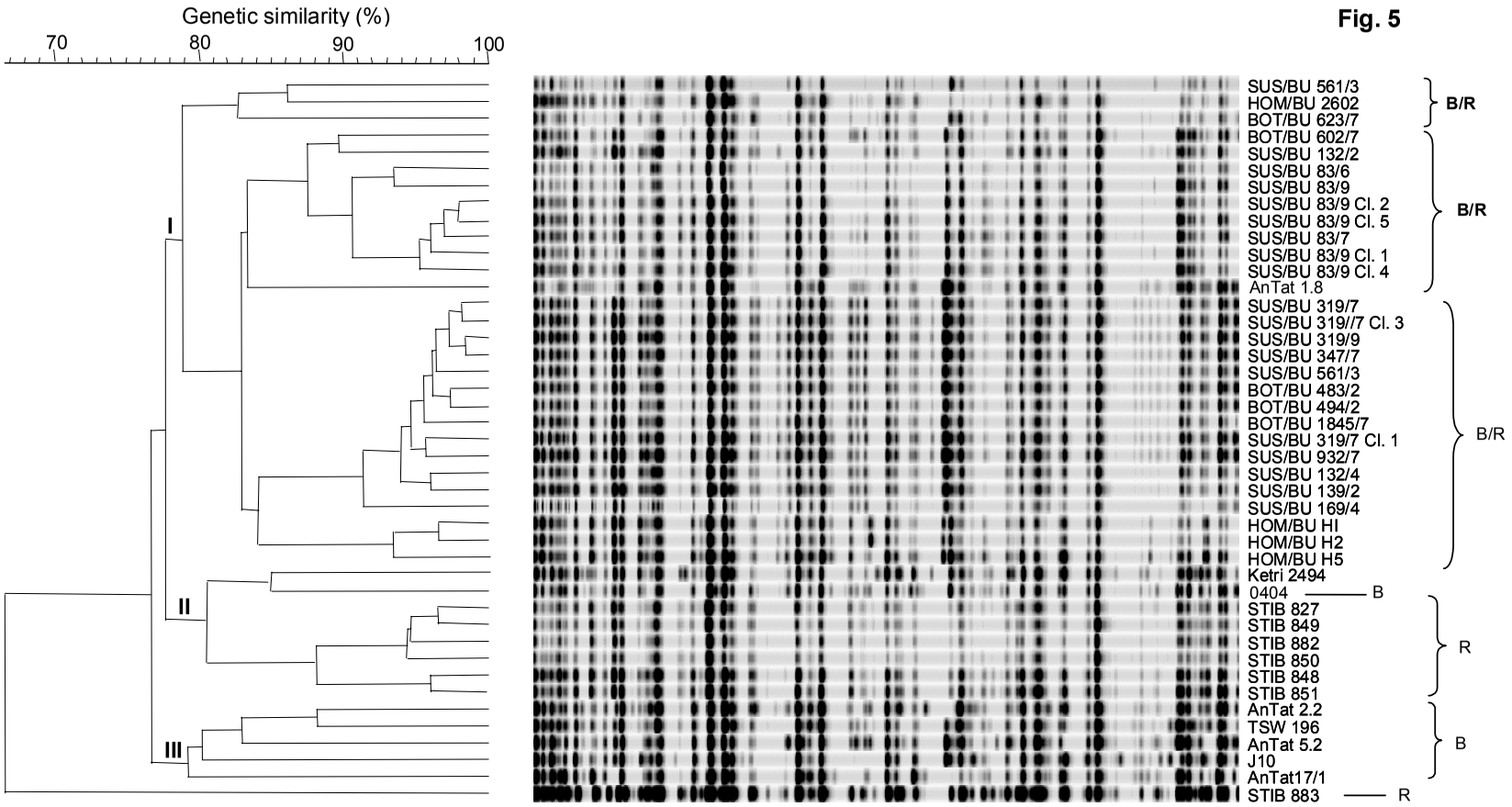

Genetic similarity (%)
70
80
90
100
I
II
III
Windows
SUS/BU 561/3
HOM/BU 2602
BOT/BU 623/7
SUS/BU 319/7
SUS/BU 319/7 Cl. 3
SUS/BU 319/9
SUS/BU 347/7
SUS/BU 373/7
BOT/BU 483/2
BOT/BU 492/2
BOT/BU 1845/7
SUS/BU 319/7 Cl. 2
SUS/BU 932/7
SUS/BU 132/4
SUS/BU 139/2
SUS/BU 169/4
HOM/BU HI
HOM/BU H2
HOM/BU H5
BOT/BU 602/7
SUS/BU 132/2
SUS/BU 83/6
SUS/BU 83/9
SUS/BU 83/9 Cl. 2
SUS/BU 83/9 Cl. 5
SUS/BU 83/7
SUS/BU 83/9 Cl. 2
SUS/BU 83/9 Cl. 4
STIB 847
STIB 849
STIB 882
STIB 850
STIB 848
STIB 851
STIB 883
B/R (end.)
R
B/R (end.)
R (epi .)

5.5. Discussion

Using restriction fragment length polymorphisms-based analysis, Hide et al. (1998) found little genetic differences between *T. b. brucei* populations isolated during an endemic period (Busia, Kenya, 1993-1994) and populations isolated during an epidemic period (Tororo, Kenya, 1988-1990), both within Busoga focus. Further, no major differences were found between the genotypes of circulating populations in the two regions. In our study, we expanded the number of genomic sites accessed per analysis using a multiplex set of four or six different endonucleases to genotype samples from the same focus in order to evaluate the complexity of parasite populations in an area in which trypanosomosis is holoendemic. In general, there was congruence between clustering based on fingerprint type and parasite subspecies identification of these stocks isolated during an endemic survey (Mukono, Uganda, 1990-1992) and during an epidemic period (Busoga, Uganda, 1990-1994), but remarkable differences were also revealed. As expected, the average genetic correlation coefficient between samples in the dendogram was less when six endonucleases were used for the analysis than when four were used. In contrast to Hide et al. (1998), we found microheterogeneities between genomes of populations from both regions.

Fig. 5. Dendogram based on fingerprints from four-endonuclease analysis of *T. b. brucei* (B) and *T. b. rhodesiense* (R) populations isolated during endemic and epidemic periods in disparate geographical regions.

Fig. 5


Genetic similarity (%)
70
80
90
100
I
II
III
SUS/BU 561/3
HOM/BU 2602
BOT/BU 623/7
B/R
BOT/BU 602/7
SUS/BU 132/2
SUS/BU 83/6
SUS/BU 83/9
SUS/BU 83/9 Cl. 2
SUS/BU 83/9 Cl. 5
SUS/BU 83/7
SUS/BU 83/9 Cl. 1
SUS/BU 83/9 Cl. 4
AnTat 1.8
B/R
SUS/BU 319/7
SUS/BU 319//7 Cl. 3
SUS/BU 319/9
SUS/BU 347/7
SUS/BU 561/3
BOT/BU 483/2
BOT/BU 494/2
BOT/BU 1845/7
SUS/BU 319/7 Cl. 1
SUS/BU 932/7
SUS/BU 132/4
SUS/BU 139/2
SUS/BU 169/4
HOM/BU HI
HOM/BU H2
HOM/BU H5
Ketri 2494
B/R
0404
B
STIB 827
STIB 849
STIB 882
STIB 850
STIB 848
STIB 851
R
AnTat 2.2
TSW 196
AnTat 5.2
J10
AnTat17/1
B
STIB 883
R

In Mukono district at the time of sampling, 0.7% of villagers, 33.5% of cattle and 52.8% of pigs were harbouring trypanosome infection (Nowak et al., 1992), and it was remarkable to find putative *T. b. rhodesiense* in the pigs, as well as in the cattle (Fig. 2). Therefore, besides the acknowledged role of cattle as reservoir hosts for *T. b. rhodesiense* (Welburn et al., 2001; Fevre et al., 2001), pigs may also constitute an important consideration in targeting human disease control strategies, such as molecular epidemiology and chemotherapy.

The importance of identifying human-infective strains in the field to determine the relevance of animal reservoirs and the prevalence of human-infective trypanosomes has previously been highlighted (Hide, 1999). The use of isoenzyme analysis on parasite populations to characterise stocks of *T. brucei* permitted early attempts to study genetic diversity and population dynamics of field strains (Godfrey and Kilgour, 1976; Gibson et al., 1980; Tait, 1980). Further isoenzyme studies on populations isolated from the same locations at the same time clearly implicated cattle as potential reservoirs (Gibson and Gashumba, 1930; Gibson and Wellde, 1985). Its application to distinguish populations of *T. b. brucei* and *T. b. rhodesiense* failed to establish unequivocal criteria for distinction (Tait et al., 1985; Godfrey et al., 1990). This may also have been so because of the reliance on human response property as a basis for strain characterisation. The SRA-PCR offered a good tool for correlating serum response properties of samples to BIIT and HSRT results. However, it is interesting to note that 2 populations isolated from pig (SUS/BU 347, SUS/BU 373/7) and 1 isolated from cattle (BOT/BU 602/7) shown to be resistant to lysis to human serum in both BIIT and HSRT are negative for the SRA gene fragment detected by PCR. This is suggestive of possible variations in the sequence of the SRA gene leading to failure to amplify the target during PCR. It may also indicate that BIIT and HSRT assays are not sufficient to unequivocally prove infectivity to humans (i.e. genuine *T. b. rhodesiense*) or underscore that the SRA gene is not the only genetic factor conferring resistance to normal

human serum. The epidemiological significance of such differences among trypanosome populations has been previously highlighted (Gibson et al., 2002). As shown by our data, the use of human infectivity as a basis for the characterisation of *T. brucei* populations especially those from East Africa needs to be fully complemented by other fine-scale analysis method, like that described in this paper. Using this multi-locus approach, we show that finer genetic differences exist among *T. brucei* stocks derived from the same focus during endemic and epidemic periods, which could be complemented with phenotypic traits for the characterisation of strains.

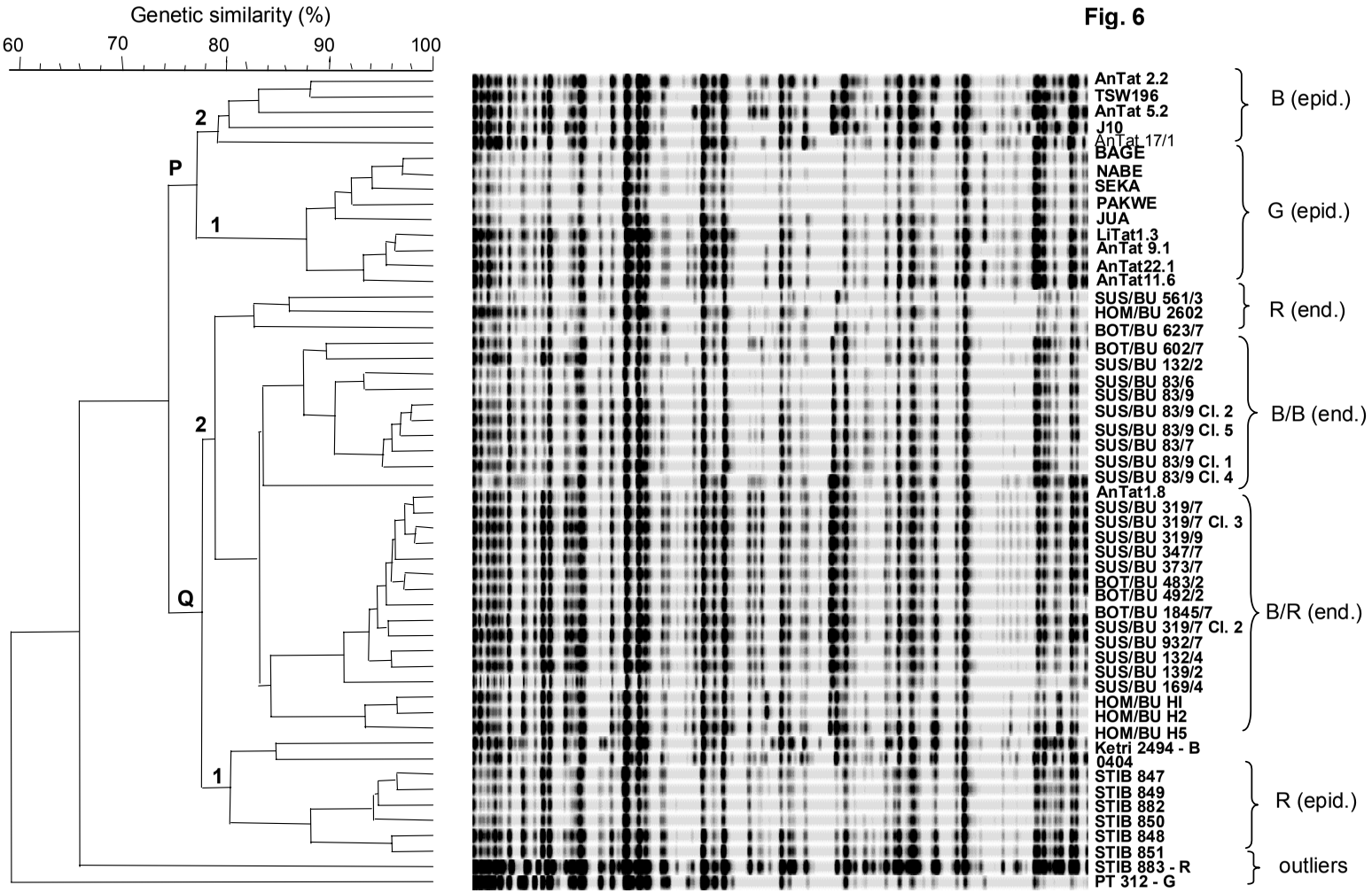
Why human and animal trypanosomosis in East Africa occur in cyclical periods of epidemics, interspersed with long periods of low-level endemicity in only a small number of foci, is unclear. It would be desirable to study the nature of these mechanisms in the overall context of possible inter-subspecies genetic interactions with strains from related and distant epidemiological areas and time. Therefore, in a pilot study, we correlated genotypes of stocks isolated during an endemic survey from related geographical regions and those of disparate stocks from epidemic episodes in distant regions. Our data show that *T. brucei* stocks isolated during 1990-1994 epidemic in Busoga focus were closely related to stocks isolated during a 1990-1992 endemicity survey in Mukono district of Uganda. However, our data further suggest that both groups are sufficiently heterogeneous to exclude clonality. As expected, the heterogeneity coefficient between stocks increased in space and time, and inter-stock differences exist on the basis of geographical origin as well as subspecies identity. When we compared genotype profiles of the 'type 1' *T. b. gambiense* populations, according to the classification of Gibson (1986), the dendogram correlations gave evidence of genetic sub-structuring within the subspecies (Fig. 6). The 'type 2' stocks have mostly been isolated from Côte d'Ivoire and, together with the clustering pattern of the fine-scale genotyping data, might explain the identity of PT312 as belonging to 'type 2'. This is in contrast to previous isoenzyme study which classified the

same stock as 'type 1' (Truc et al., 1997). In addition, STIB 883 and PT312 were classified as outliers in the dendogram (Fig. 6) and correlate at correlate at 64.8% and 59%, respectively, in relation to other populations, indicating distant genetic relatedness to other isolates. This also shows that STIB 883 was completely different from other stocks isolated during the epidemic in Busoga focus. Such observation indicates that new parasite strains that upset host-parasite balance may have an important role in initiating an epidemic episode. The introduction of such new strains and subsequent outbreaks of *T. b. rhodesiense* might have coincided with transhumance movement of cattle.

To summarise, we studied within- and between-stock variations using a fine genotyping method to develop an understanding of the levels of variation within and between geographical foci and to describe genetic relatedness of epidemiologically-defined *T. brucei* populations. Genome composition of the stocks seemed sufficiently stable in space and time, in agreement with the findings of Godfrey and Kilgour (1976). However, significant genetic differences exist among stocks even from within a disease focus, while *T. b. brucei* and *T. b. rhodesiense* appear to be sufficiently genetically separated to merit their retention as separate subspecies. Taken together, the data indicate that considerable genetic differences exist among *T. brucei* populations within the same disease focus. These differences are more evident in cladistic analysis of stocks separated in space and time of isolation.

Fig. 6. Fingerprint types and relationship between *T. b. brucei* (B), *T. b. rhodesiense* (R) and *T. b. gambiense* (G) populations from epidemic (epid) or endemic (end) periods in Busoga focus and other disparate locations. The two main clusters P and Q are further sub-clustered.

Fig. 6


Genetic similarity (%)
60
70
80
90
100
P
Q
2
1
2
1
AnTat 2.2
TSW196
AnTat 5.2
J10
AnTat 17/1
BAGE
NABE
SEKA
PAKWE
JUA
LiTat1.3
AnTat 9.1
AnTat22.1
AnTat11.6
SUS/BU 561/3
HOM/BU 2602
BOT/BU 623/7
BOT/BU 602/7
SUS/BU 132/2
SUS/BU 83/6
SUS/BU 83/9
SUS/BU 83/9 Cl. 2
SUS/BU 83/9 Cl. 5
SUS/BU 83/7
SUS/BU 83/9 Cl. 1
SUS/BU 83/9 Cl. 4
AnTat1.8
SUS/BU 319/7
SUS/BU 319/7 Cl. 3
SUS/BU 319/9
SUS/BU 347/7
SUS/BU 373/7
BOT/BU 483/2
BOT/BU 492/2
BOT/BU 1845/7
SUS/BU 319/7 Cl. 2
SUS/BU 932/7
SUS/BU 132/4
SUS/BU 139/2
SUS/BU 169/4
HOM/BU HI
HOM/BU H2
HOM/BU H5
Ketri 2494 - B
0404
STIB 847
STIB 849
STIB 882
STIB 850
STIB 848
STIB 851
STIB 883 - R
PT 312 - G
B (epid.)
G (epid.)
R (end.)
B/B (end.)
B/R (end.)
R (epid.)
outliers

5.6. Acknowledgements

We acknowledge the Institute for Animal Science and Health (ID-Lelystad), The Netherlands for funding the research of ECA (project number *iSEO* 640.47293.00).

References

Agbo, E. C., P. A. O. Majiwa, H. J. H. M. Claassen, and M. F. W. te Pas. 2002. Molecular variation of *Trypanosoma brucei* subspecies as revealed by AFLP fingerprinting. Parasitology **124:**349-358.

Brun, R., L. Jenni, M. Schönenberger, and K-F. Schell. 1981. In vitro cultivation of bloodstream forms of *Trypanosoma brucei*, *T. rhodesiense* and *T. gambiense*. J. Protozool. **28:**470-479.

De Greef, C., G. Imbrechts, G. Matthyssens, N. Van Meirvenne, and R. A. Hamers. 1989. A gene expressed only in serum-resistant variants of *Trypanosome brucei rhodesiense*. Mol. Biochem. Parasitol. **36:**169-176.

Degen, R., H. Pospichal, J. C. K. Enyaru, and L. Jenni. 1995. Sexual compatibility among *Trypanosoma brucei* isolates from an epidemic area in South-eastern Uganda. Parasitol. Res. **8:**253-257.

Enyaru, J. C. K., E. Matovu, M. Odiit, L. A. Okedi, A. J. J. Rwendeire, and J. R. Stevens. 1993. Isoenzyme comparison of *Trypanozoon* isolates from two sleeping sickness areas of South-eastern Uganda. Acta Tropica **55:**97-115.

Enyaru, J. C. K., E. Matovu, M. Odiit, L. A. Okedi, A. J. J. Rwendeire, and J. R. Stevens. 1997. Genetic diversity in *Trypanosoma* (*Trypanozoon*) *brucei* isolates from mainland and Lake Victoria island populations in South-eastern Uganda**:** Epidemiological and control implications. Annals Trop. Med. Parasitol. **91:**107-113.

Fevre, E. M., P. G. Coleman, M. Odiit, J. W. Magona, S. C. Welburn, and M. E. Woolhouse. 2001. The origins of a new *Trypanosoma brucei*

rhodesiense sleeping sickness outbreak in eastern Uganda. Lancet **358:**625-628.

Gibson, W. C. 1986. Will the real *Trypanosoma b. gambiense* please stand up. Parasitol. Today **2:**255-257.

Gibson, W. C., and B. T. Wellde. 1985. Characterisation of *Trypanozoon* stocks from South Nyaza sleeping sickness focus in Western Kenya. Trans. R. Soc. Trop. Med. Hyg. **79:**671-676.

Gibson, W. C., and J. K. Gashumba. 1983. Isoenzyme characterisation of some *Trypanozoon* stocks from a recent trypanosomiasis epidemic in Uganda. Trans. R. Soc. Trop. Med. Hyg. **77:**114-118.

Gibson, W. C., T. F. de C. Marshall, and D. G. Godfrey. 1980. Numerical analysis of enzyme polymorphism**:** a new approach to epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. Adv. Parasitol. **18:**175-246.

Gibson, W., T. Backhouse, and A. Griffiths. 2002. The human serum-resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infec. Genet. Evol. **1:**207-214.

Godfrey, D. G., and V. Kilgour. 1976. Enzyme electrophoresis in characterising the causative organism of Gambian trypanosomiasis. Trans. R. Soc. Trop. Med. Hyg. **70:**219-224

Godfrey, D. G., R. D. Baker, L. R. Rickman, and D. Mehlitz. 1990. The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. Adv. Parasitol. **29:**1-74.

Hawking, F. 1976. The resistance to human plasma of *Trypanosoma brucei*, *Trypanosoma rhodesiense* and *Trypanosoma gambiense*. I. Analysis of the composition of trypanosome strains. Trans. R. Soc. Trop. Med. Hyg. **70:**504-512.

Heath, S. 1997. Molecular Techniques in Analytical Parasitology. *In* Analytical Parasitology (ed. Rogan, M.T.), pp 67-68.

Hide, G. 1999. History of sleeping sickness in East Africa. Clin. Microbiol. Rev. **12:** 112-125.

Hide, G., S. Angus, P. H. Holmes, I. Maudlin, and S. C. Welburn. 1998. Comparison of *Trypanosoma brucei* strains circulating in an endemic and an epidemic area of a sleeping sickness focus. Exp. Parasitol. **89:**21-29.

Hide, G., S. C. Welburn, A. Tait, and I. Maudlin. 1994. Epidemiological relationships of *Trypanosoma brucei* stocks from South East Uganda**:** evidence for different populations structures in human infective and non-human infective isolates. Parasitology **109:**95-111.

Jenni, L., and R. Brun. 1982. An in vitro test for human serum resistance of *Trypanosoma* (*Trypanozoon*) *brucei*. Acta Tropica **39:**281-284.

Nowak, F., D. Kakaire, U. Tietjen, L. Hoffmann, B. Katabazi, and D. Mehlitz. 1992. *Glossina f. fuscipes* as a vector of human and animal trypanosomiasis in South-Eastern Uganda. Determination of seasonal fly density, host preference and infection rates. Zbl. Bakt. Hyg. **325:**60-61.

Pearson, K. 1926. On the coefficient of racial likeness. Biometrika **18:**105-117.

Radwanska, M., F. Claes, S. Magez, E. Magnus, A. Perez-Morga, E. Pays, and P. Büscher. 2002. Novel primer sequences for polymerase chain reaction-based detection of *Trypanosoma brucei gambiense*. Am. J. Trop. Med. Hyg. **67:**289-295.

Rickman, L. R., and J. Robson. 1970. The testing of proven *Trypanosoma brucei* and *T. rhodesiense* strains by the blood incubation infectivity test. Bull. World Health Organisation **42:**911-916.

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

Tait, A., J. D. Barry, R. Wink, A. Sanderson, and J. S. Crowe. 1985. Enzyme variation in *Trypanosoma brucei* ssp. II. Evidence for *T. b. rhodesiense* being a subset of variants of *T. b. brucei*. Parasitology **90:**89-100.

Von Dobschuetz, S. 2002 Molekularbiologische Untersuchungen zur Identifizierung potentiell humaninfektiöser *Trypanosoma (Trypanozoon) brucei*- Isolate aus Süd-Ost Uganda. *Doctoral Dissertation*, Journal-Nr. 2577, Freie Universität Berlin, p. **205**.

Welburn, S. C., K. Picozzi, E. M. Fevre, P. G. Coleman, M. Odiit, M. Carrington, and I. Maudlin. 2002. Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. Lancet **358:**2017-2019.

World Health Organisation publications, Geneva. 2001. *In* African trypanosomiasis, Fact Sheet, number **259**.

CHAPTER 6

Correlating multilocus molecular markers, differential DNA methylation analysis and human serum responsiveness for defining *Trypanosoma brucei rhodesiense* genotypes

Eddy Chukwura Agbo [1], Eddy Magnus [3], Philippe Büscher [3], Phelix A.O. Majiwa [4], Eric Claassen [2] and Marinus F.W. te Pas [1]

[1] Division of Animal Sciences, Section for Animal Genomics, Institute for Animal Science and Health, ID-Lelystad Edelhertweg 15, 8200 AB Lelystad, The Netherlands
[2] Department of Immunology, Erasmus University Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands
[3] Institute of Tropical Medicine, Department of Parasitology, Nationalestraat 155, B-2000 Antwerp, Belgium
[4] International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya

Submitted for publication

6.1. Abstract

Detailed genotyping analyses of *Trypanosoma brucei* can be used to study genetic and phenotypic diversity, in order to identify biological loci associated with intrinsic parasite factors. Variations in multilocus genotypes based on neutral polymorphisms should be positively correlated with variation in virulence and tropism, provided the markers are finely distributed and stable enough for such association to exist. We applied marker-assisted fine-scale multilocus genotyping analysis as a predictive tool for characterising ANTAR series of antigenic variants, to determine the correlation between genotype and human serum resistance (HSR) phenotypic trait. Then we evaluated for differential DNA methylation based on uniqueness of the phenotypic and genetic constitution of genotypes, using the ANTAR series and two unique isovat clones that spontaneous lost the ability to revert to human serum resistance. The classification strategy of combining the genotype information with the human serum response trait produced well-differentiated groups of genotypes, indicating significant dominant genetic differences among the parasite clones examined. The techniques applied to measure genetic relatedness demonstrated ability of the multilocus genome analysis data to correlate genotypes on the basis of clonal identity of strains, but incongruent on the basis of phenotypic attributes. The lack of correlation between methylation-dependent matrices and HSR suggest that epigenetic regulation of the trait is independent of DNA methylation.

6.2. Introduction

The tsetse-borne parasites of *Trypanosoma brucei* cause acute to chronic disease in humans and animals, are virtually identical by morphological and developmental criteria, but can be distinguished based on their ability to resist lysis by human serum (equivalent to human infectivity). *Trypanosoma b. brucei*, one of the causes of 'nagana' in cattle does not infect humans

because it is lysed by normal human serum. *Trypanosoma b. gambiense* and *T. b. rhodesiense*, on the other hand, are resistant to lysis by normal human serum (Hawking, 1973). Two important complications arise from this method of characterisation. First, in East Africa, isolates from tsetse and domestic animals may consist of a mixture of the human infective *T. b. rhodesiense* and non-human infective *T. b. brucei*. Therefore, parasites derived from hosts other than humans potentially contain both subspecies. Secondly, *T. b. gambiense* and *T. b. rhodesiense* appear to differ in their mechanism of resistance to normal human serum, *T. b. gambiense* being constitutively resistant to lysis, whereas in *T. b. rhodesiense* this mechanism is reversible (Hawking, 1977).

There have been great advances in elucidating the molecular basis of trypanosome killing by human serum or parasite resistance to lysis. Human haptoglobin-related protein has been identified as the human toxin responsible for *T. b. brucei* killing (Rifkin, 1976), corroborated by subsequent studies (Rifkin et al., 1994; Tomlinson and Raper, 1996; Raper et al., 2001). Other recent studies have shown that a serum-resistance-associated (SRA) gene transcript, present only in resistant *T. b. rhodesiense*, is responsible for human serum killing of the parasites (De Greef et al., 1989; Xong et al., 1998). In this case, an important factor in human infectivity was whether or not the SRA was expressed. The gene is present in all strains tested so far (Welburn et al., 2001; Gibson et al., 2002), but absent in *T. b. gambiense* - a subspecies that is constitutively resistant to lysis by human serum. It has been suggested that variable expression of SRA might account for the periodic loss of human infectivity in *T. b. rhodesiense* strains. It is therefore, conceivable that the complex process of human serum resistance (HSR) may transcend the activity of one gene fragment.

The identification of new genome-based tools for the study of genetic associations of human serum resistance in trypanosomes will have an important input in the overall control strategies of trypanosomosis. In other

taxa, a common genome-based approach to study such associations between different loci and resistance has been to search for polymorphisms between independent sensitive and resistant populations (Cowen et al., 1999; Huys et al., 2000; Grigg et al., 2001). Although these approaches can indicate that different populations differ at a particular locus, they are not specific for resistance phenotype unless the resistant population has been stably derived from the susceptible population. Genome scan methods have been successfully applied to characterise multigene traits relevant for animal breeding (Malek et al., 2001a, 2001b). In trypanosomes, other molecular characterisation studies were less successful in revealing specific resistance markers (Borst and Grivell, 1971; Gibson et al., 1980; Tait et al., 1985; Godfrey et al., 1990; Hide et al., 1991; MacLeod et al., 2000; Agbo et al., 2002).

The goals of this study were twofold: to identify and categorise individual human serum sensitive and resistant *T. b. rhodesiense* clones on the basis of their molecular signatures, and to evaluate for possible correlation of differential DNA methylation patterns in establishing these traits. Therefore, a marker-enrichment strategy involving sets of hexaplex endonuclease-based genotyping was used to study the correlation between human serum resistant phenotypes and genotypes of two different *T. b. rhodesiense* strains. These consisted of (1) sets of ANTAR antigenic variants with reversible serum response traits and, (2) a pair of 'mutant' strains that spontaneously became constitutively human serum sensitive. Finally, we evaluated the correlation of differential DNA methylation and genotypes to specific serum response properties.

6.3. Materials and Methods

6.3.1. Parasite characterisation and DNA isolation

Two sets of *T. b. rhodesiense* clones and strains were used. The first consisted of a pair each of AnTat 12.1 and AnTat 25.1 isovat clones,

derived from sleeping sickness patients in Rwanda (1971). Each isovat couple comprised of a non-rodent-adapted serum resistant (NA/R) clone, which after multiple short passages (>40) and subcloning in mice, gave rise to a rodent-adapted and serum sensitive (AD/S) clone with the same VAT (Fig. 1). The second set consisted of STIB 848 and STIB 884 isolated from sleeping sickness patients in Uganda in 1990 and 1993, respectively (J. Enyaru, pers. comm.). These strains now appear to be constitutively sensitive to normal human serum, as resistant populations could no longer be generated in recent repeated attempts by the authors (unpublished data). The stable phenotype of the STIB strains presents them as a reliable model for fine-scale genomics marker-based characterisation studies.

The procedures for parasite cloning, serotyping and selection of VATs were as previously described (Van Meirvenne et al., 1975). Testing for human serum resistance or sensitivity and selecting human serum resistant VATs (i.e. HSIIT, human serum infectivity inhibition test) were as previously described (Van Meirvenne et al., 1976). Trypanosomes were grown in mice and rats and pure trypanosome sediments were obtained after separation from the blood, according to Lanham and Godfrey (1970). Genomic DNA was isolated from trypanosome pellets stored at -80ºC, according to standard protocols and further characterised using primers previously described for *T. b. rhodesiense*-specific PCR (Welburn et al., 2001).

6.3.2. Multiplex endonuclease genotyping analysis (MEGA)

We explored interspecific diversity between subgroups of the 4 ANTAR clones using an extension of the recently developed MEGA technique (Agbo et al., In press). Briefly, about 200-400 ng of genomic DNA obtained from each of the clones was digested respectively, with 10U each of (*Bgl*II, *Bcl*I, *Xho*II, *EcoR*I, *Acs*I and *Mfe*I) endonuclease combination. Restriction enzymes were purchased from Roche Molecular Biochemicals (Almere, The Netherlands) or New England Biolabs (Westburg, Leusden, The Netherlands). The restriction digestions were performed in a final volume of 20 µl of appropriate buffer, according to the supplier's recommendations.

Depending on the optimal temperature of enzyme activity, reactions were incubated at 37°C or 50°C, for 6 h. When buffers were compatible, enzyme reactions were performed in double digestions. Ligation of 50 pm each of *Bgl*II and *Mfe*I adapters (Table 1) to the digested fragments was performed as previously described (Sambrook et al., 1989). The amplification cascade involved two reaction series. For the pre-selective amplification, 4 µl of double-diluted ligation products were incubated with 20 pm each of *Bgl*II and *Mfe*I primers (Table 1), 200 µM dNTPs, PCR buffer and 1.25 U Takara exonuclease *Taq* polymerase, in a final volume of 20 µl.

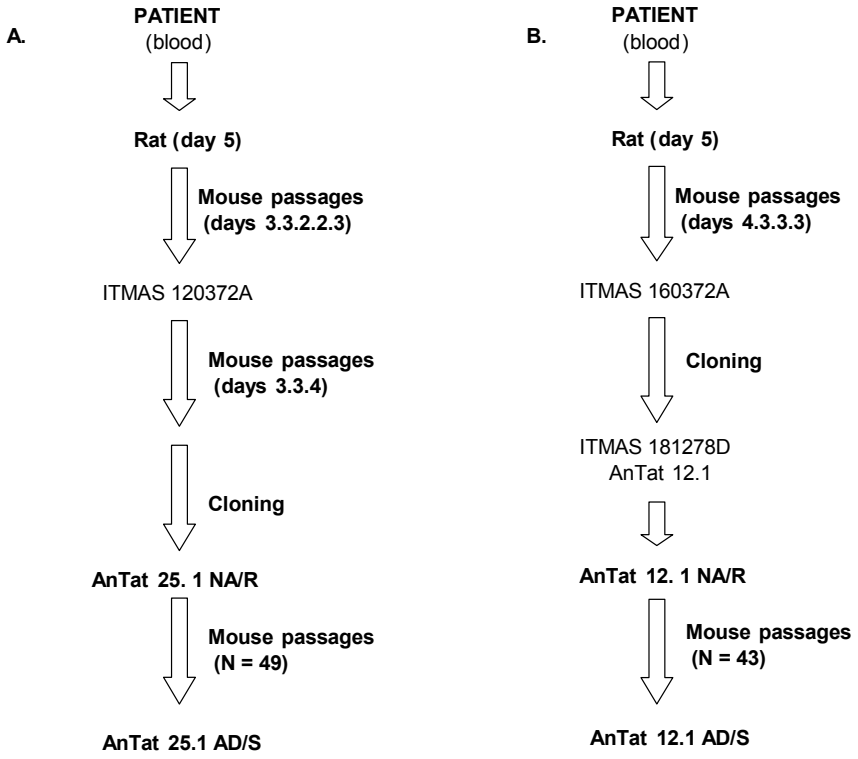


Fig. 1. Simplified pedigree of ANTAR Variable Antigenic Types (VATs) - AnTat 25.1 and AnTat 12.1 - derived from two sleeping sickness patients in Rwanda (1971). 'NA/R' and 'AD/S' are 'Non-adapted resistant' and 'Adapted sensitive' clones, respectively.

The mixture was first incubated at 72°C for 5 min and then 95°C for 30 s, followed by 20 cycles consisting of 95°C for 30 s, 56°C for 30 s and 72°C for 2 min. The pre-selective amplification products were 1:20-diluted with double-distilled water. For the selective amplification phase, 4 µl of the diluted samples was mixed with PCR buffer, 20 pm each of *Bgl*II+0 and fluorescently-labelled *Mfe*I+A, *Mfe*I+AA and *Mfe*I+AT primers, respectively. Then 200 µM dNTPs and 1.25 U Takara exonuclease *Taq* polymerase were added. The 20 µl final reaction volume was incubated at 95°C for 2 min, followed by 20 reaction cycles as for the pre-selective phase, and a final elongation phase of 10 min at 72°C. Selective amplification products were electrophoresed in 7% polyacrylamide sequencing gels, using ABI 373 sequencer (Applied Biosystems).

6.3.3. Differential methylation analysis

Methylation sensitivity data of the restriction enzymes used and the relevant references have been published at [http://rebase.neb.com/cgi-bin/mslist]. Analysis of the differential patterns of DNA methylation from the clones and the mutant non-revertant *T. b. rhodesiense* strains involved a modification of the MEGA approach. Three sets of methylation-sensitive restriction enzymes *Fba*I-*Mfl*I-*Mun*I-*Acs*I, *Fba*I-*Mfl*I-*Xba*I-*Spe*I and *Fba*I-*Mfl*I-*BstB*I-*Cla*I were used to digest 200-400 ng genomic DNA for 8 h, to ensure complete digestion. The hexacutters *Bcl*I, and *Bgl*II and *Xho*II in the original protocol were replaced with isoschizomers *Fba*I and *Mfl*I, respectively. Each endonuclease per set of two enzymes creates cohesive ends that are compatible with the overhang sites created by the other. *Fba*I and *Mfl*I do not cut when the recognition sequence is Dam methylated whereas *Bgl*II and *Xho*II are not impaired. Similarly, *Mun*I (isoschizomer of *Mfe*I) and *Acs*I are inactive in internal Dam methylation, but *Acs*I cuts hemimethylated external adenosine or guanidine, although at a lower rate. *Xba*I (recognising TCTAGA) and *Spe*I (ACTAGT) are blocked by Dam- and m5C-methylation, but cut the hemimethylated sequence, also at a lower rate than the unmethylated one. *BstB*I (recognising TTCGAA) and *Cla*I (ATCGAT) create compatible cohesive 3’ overhang ends and are blocked

by m5C-methylation. Subsequent procedures for the MEGA were as described above. On the basis of its reproducibility, even distribution of bands along the gel and number of polymorphic bands detected, the *Fba*I-*Mfl*I-*Mun*I-*Acs*I combination was selected. Details of the sequences and combinations of the adapters and primers used are given in Table 1.

Endonuclease combinations	Adapters	Pre-selective PCR primer sets	Selective primer sets: differential methylation
*Fba*I-*Mfl*I-*Mun*I-*Acs*I	*Bgl*II/*Mfe*I	*Bgl*II+0/*Mfe*I+0	*Bgl*II+A/*Mfe*I+AA* *Bgl*II+A/*Mfe*I+AG* *Bgl*II+T/*Mfe*I+AC* *Bgl*II+T/*Mfe*I+AT*

Table 1. Combinations of primers and adapters used for analyses. Equimolar amounts of cognate oligonucleotides were heated at 99°C for 10 min and placed on the bench to cool to room to permit annealing of adapters: *Bgl*II ($^{5'-}$CGG ACT AGA GTA CAC TGT C; $^{5'-}$GAT CGA CAG TGT ACT CTA GTC); *Mfe*I ($^{5'-}$AAT TCC AAG AGC TCT CCA GTA C; $^{5'-}$AGT ACT GGA GAG CTC TTG) and *Xba*I ($^{5'-}$GTC GTA GAC TGC GTA CG; $^{5'-}$CTA GCG TAC GCA GTC). Core primer sequences (without selective nucleotide, i.e. primer+0) were**:** *Bgl*II ($^{5'-}$GAG TAC ACT GTC GAT CT); *Mfe*I ($^{5'-}$GAG AGC TCT TGG AAT TG) and *Xba*I ($^{5'-}$GAC TGC GTA CGC TAG A). Symbol * indicates FAM (6-carboxyfluorescein)-labeled primer. The endonucleases are completely blocked by Dam DNA methylation.

6.3.4. Data processing

Following conversion of gel files, normalisation of sample files and background subtraction, inter-strain polymorphic amplified fragments were scored for the presence (1) or absence (0) of specific fragments depicted by signal peaks. Internal size marker loaded with every sample permitted definitive determinations of fragment sizes. Genetic similarities between fingerprint pairs were calculated using the Pearson product-moment coefficient, and matrices were clustered using the minimum variance within groups proposed by Ward (1963).

6.4. Results and Discussion

6.4.1. Strain identification and distribution of genotypes

The reversibility of the human serum-resistance phenotype in the AnTat 12 and AnTat 25 strains enabled the development of isogenic clones for comparative genotype analysis. In the *T. b. rhodesiense* ETat series of antigenic variants (Van Meirvenne et al., 1976), this character was associated with antigenic variation, and cloned isovat populations exist in both the normal human serum sensitive and resistant forms. By differential hybridisation of the sensitive and resistant ETat clones, De Greef and colleagues (1989) identified a transcript of the serum-resistance-associated (SRA) gene, present only in the resistant clones.

In this study, we first verified strain identity by using previously described primers (Welburn et al., 2001) to amplify the *T. b. rhodesiense*-characteristic 743 bp SRA product in all ANTAR and STIB samples (Fig. 2). This fragment spans through a 378 bp deletion that encodes the central region of an SRA-related VSG N-terminal domain (Welburn et al., 2001). It has been shown that for the SRA-dependent parasite gene fragment, the important factor in human serum resistance phenotype was whether or not the gene was expressed, and that not all *T. b. rhodesiense* express the transcript (Xong et al., 1998). This might explain the positive amplification

of the gene in the serum sensitive STIB strains. The ubiquity of the SRA gene (Welburn et al., 20001; Gibson et al., 2002) and its association with human infectivity is beyond doubt. However, other parasite genetic factors are thought to be involved, partly because SRA is absent in *T. b. gambiense*. Even within the *T. b. rhodesiense*, the gene is not expressed in all strains (Rifkin et al., 1994), which raises question on the role of epigenetic regulatory processes in the trait.

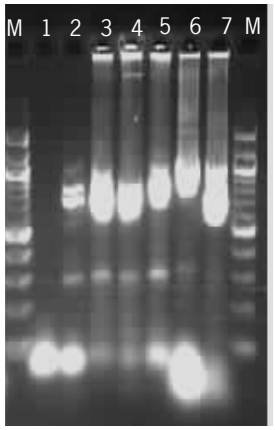


Fig. 2. Specific PCR amplification products based on serum-resistance-associated (SRA) gene (De Greef et al., 1989). Product set was generated using primer combination B537 and B538 (Welburn et al., 2001).
Lane 1 - neg. control (water); 2 - AnTat12.1 NA/R; 3 - AnTat12.1 AD/S; 4 - AnTat25.1 NA/R; 5 - AnTat25.1 AD/S; 6 - STIB 884; 7 - STIB 848; M - 100 bp marker

Few previous studies have been undertaken to directly relate the *T. brucei* trait of human infectivity to the molecular signatures of different strains (Hide et al., 1991; Hide et al., 1994; Agbo et al., 2001). In other eukaryotic organisms, genome scan methods have permitted the identification of genomic regions that control phenotypic traits (Malek et al., 2001a, 2001b). In this study, we genotyped 4 ANTAR clones using restriction fragments generated from 6-endonuclease digestions of genomic DNA and amplified with *Bgl*II+0 and *Mfe*I+A*, *Mfe*I+AA* or *Mfe*I+AT* primer combinations (Table 1). The representative genotypic profiles obtained using optimised protocols (Fig. 3) were reproducible in repeat experiments, confirming

(among other parameters) the complete digestion of genomic DNA. In principle, given the enhanced efficiency of detection of amplifiable polymorphisms with the multiplex-endonuclease genotyping approach (Agbo et al., In press), the genotype patterns would be expected to strongly correlate with the phenotypic characters of the analysed test clones. However, on the combined dendogram (Fig. 3), there was incongruence between human serum response properties and genotypic profiles, suggesting limited role for dominance-dependent polymorphisms.

The clustering of fingerprints was performed using the minimum variance within groups proposed by Ward (1963), which combines in each step, the two clusters whose fusion yields the least increase in the sum of squared distances from each individual to the centroid of the cluster to which it belongs. Therefore, the variance between groups is maximised while the variance within groups is minimised. The lack of correlation between clones determined by molecular profiles and human serum resistance suggests limited role for human infectivity on phenetic relationships.

So far, the analysis has involved human serum sensitive or resistant trypanosome clones that have been reversibly adapted to specific phenotypes. It would also be informative to analyse the constitutively resistant or sensitive strains, and to compare the clones derived from other strains of the same subspecies. In this way, it might be possible to test for correlation of human infectivity and molecular patterns, and to determine whether these co-evolve in parasite populations.

6.4.2. MEGA and differential DNA methylation patterns

Sets of methylation-sensitive endonucleases were used to search for DNA methylation patterns in genomes of the clones derived from ANTAR clones, and from STIB strains that became constitutively human serum sensitive (Fig. 4). The restriction endonucleases *Fba*I and *Mfl*I are expected to cut less often than other hexacutters, such as *Bgl*II and *Xho*II, because of their sensitivity to Dam methylated residues. It should be pointed out that the

differences in banding patterns obtained in this analysis are not restriction fragment length polymorphisms between normal human serum sensitive and resistant trypanosomes, but represent differences in Dam and m5C methylated residues.

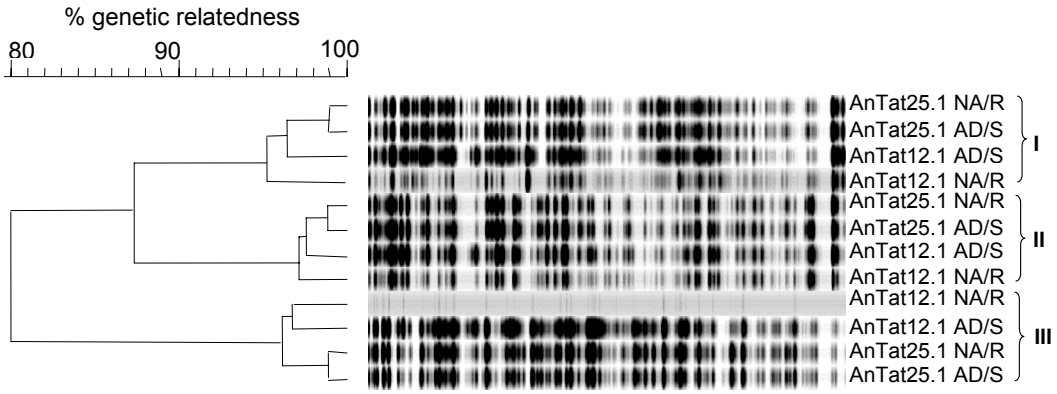


Fig. 3. Cophenetic dendogram profiles showing genotypic and human serum phenotypic correlates. The selective primer combinations used for the amplification of the 6-endonuclease digested products were *Bgl*II-0/*MfeI*-AT* [I]; *Bgl*II-0/*MfeI*-A* [II]; *Bgl*II-0/*MfeI*-AA* [III].

If differential methylation exists at the sites recognised by the restriction enzymes used, and the methylation has a role in human serum response, then the fingerprint patterns would be expected to segregate according to human serum response status of the trypanosomes. Alternatively, the lack of such a correlation would suggest a lack of differential regulation of methylation in the different strains at the analysed restriction sites.

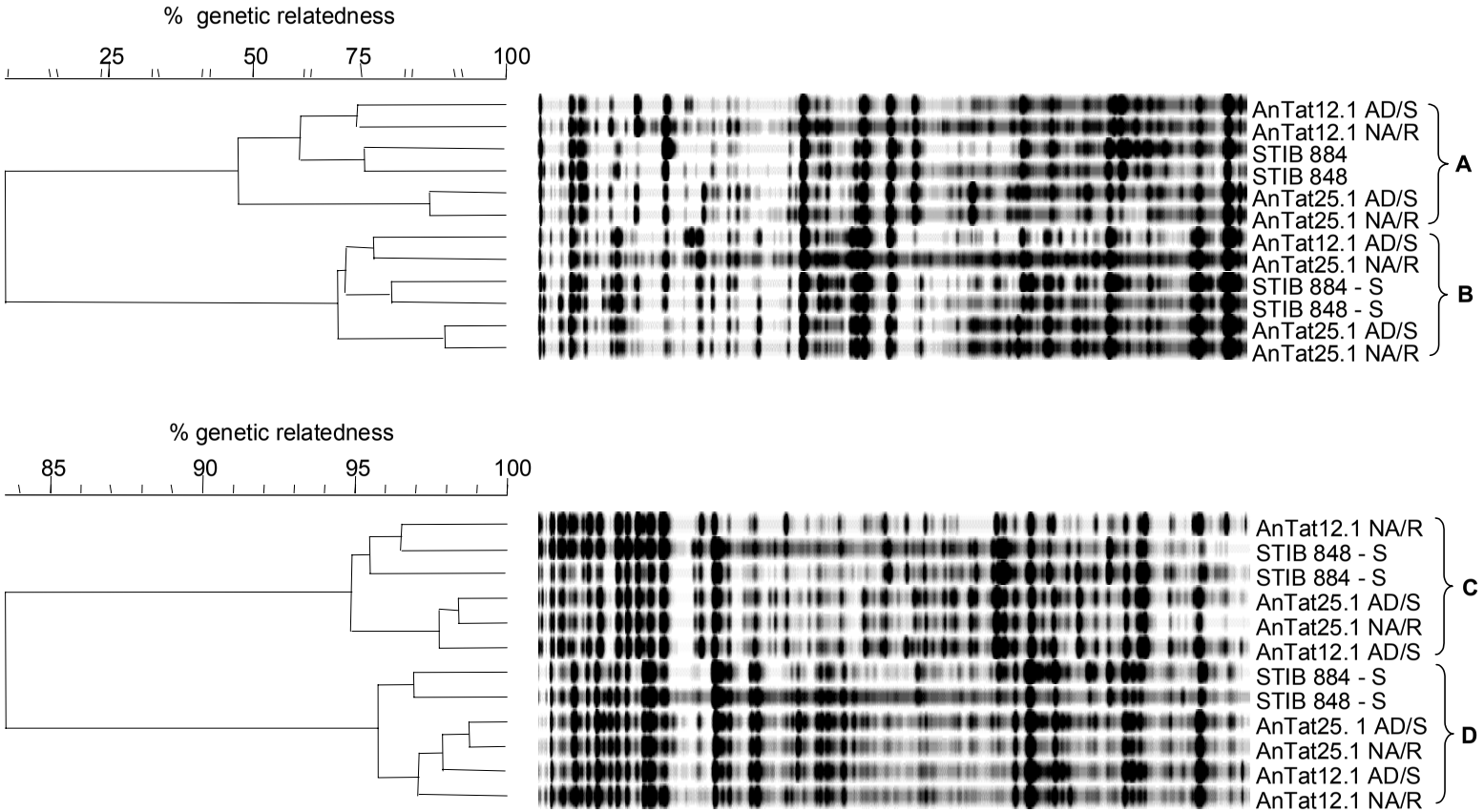
To evaluate this hypothesis, we derived similarity matrices from methylation-sensitive polymorphisms of 3 sets of endonucleases (see Materials and Methods). No correlation was found between these matrices and human serum response trait, suggesting a lack of relationship between

the components defining genetic relatedness and those that define methylation patterns. In fact, AnTat 12.1 NA/R and AD/S, and AnTat25.1 NA/R and AD/S, that are clonally related, respectively, were not grouped on the basis of the analysis of methylation-dependent polymorphisms. However, the STIB strains that are less genetically-related displayed more similar patterns of methylation-dependent polymorphisms (Fig. 4). These results indicate that there is a high degree of heterogeneity in the methylation state of specific methylation-sensitive sites among strains. Alternatively, it is possible that the two sets of strains may be different in other ways possibly unconnected with resistance to human serum. It should be pointed out that although this approach can be used to correlate methylation-dependent epigenetic regulation of human serum responsiveness at the genomic level, the relationship between presence or absence of an amplified fragment and its methylated or non-methylated state cannot be established by this analysis.

There is currently a great deal of debate regarding the feasibility of identifying genetic associations with the human serum response (HSR) trait in *T. brucei*. However, much effort needs to be invested to identify tools with the greatest potential value for increasing the level of fine resolution. In this study, we have established that genotypic markers and HSR phenotypic properties are not directly correlated among *T. b. rhodesiense* clones and strains. We also presented data that propose limited role for differential methylation in HSR trait. It would be interesting to evaluate differences in DNA methylation patterns in known trypanosome genes using this MEGA approach.

Fig. 4. Dendogram from differential methylation analysis of *T. brucei* subspecies, using the MEGA approach. Pearson correlation product moment coefficient was used to determine the proportion of mismatched fragments between samples, and plotted in the dendograms following unweighted pair group method using average linkage (UPGMA) analysis. Genetic profiles are computer reconstructions and hence, underestimate the resolution of the original gel images. The *Fba*I-*Mfl*I-*Mun*I-*Acs*I restriction fragments were amplified using primer combinations *Bgl*II-A/*Mfe*I-AC* [A]; *Bgl*II-A/*Mfe*I-AT* [B], *Bgl*II-T/*Mfe*I-AA* [C], and *Bgl*II-T/*Mfe*I-AG* [D].
(See overleaf).

Fig. 4


% genetic relatedness
25
50
75
100
AnTat12.1 AD/S
AnTat12.1 NA/R
STIB 884
STIB 848
AnTat25.1 AD/S
AnTat25.1 NA/R
A
AnTat12.1 AD/S
AnTat25.1 NA/R
STIB 884 - S
STIB 848 - S
AnTat25.1 AD/S
AnTat25.1 NA/R
B
% genetic relatedness
85
90
95
100
AnTat12.1 NA/R
STIB 848 - S
STIB 884 - S
AnTat25.1 AD/S
AnTat25.1 NA/R
AnTat12.1 AD/S
C
STIB 884 - S
STIB 848 - S
AnTat25. 1 AD/S
AnTat25.1 NA/R
AnTat12.1 AD/S
AnTat12.1 NA/R
D

References

Agbo, E. C., B. Duim, P. A. O. Majiwa, P. Büscher, E. Claassen, and M. F. W. te Pas. A multiplex-endonuclease, two-adapter genotyping approach enriches for high resolution of identifiable polymorphisms and permits for amplification reaction stringency in the genotyping of trypanosome subspecies. Chromosoma, In press.

Agbo, E. C., P. A. Majiwa, E. Claassen, and M. H. Roos. 2001. Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterisation. Exp. Parasitol. **99:**123-131.

Berberof, M., D. Perez-Morga, and E. Pays. 2001. A receptor-like flagellar pocket glycoprotein specific to *Trypanosoma brucei gambiense*. Mol. Biochem. Parasitol. **113:**127-138.

Borst, P. and L. A. Grivell. 1971. Mitochondrial ribosomes. FEBS Lett. **13:**73-88.

Cowen, L. E., C. Sirjusingh, R. C. Summerbell, S. Walmsley, S. Richardson, L. M. Kohn, and J. B. Anderson. 1999. Multilocus genotypes and DNA fingerprints do not predict variation in azole resistance among clinical isolates of *Candida albicans*. Antimicrob. Agents Chemother. **43:**2930-2938.

De Greef, C., and R. Hamers 1994. The serum resistance-associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. Mol. Biochem. Parasitol. **68:**277-284.

De Greef, C., H. Imberechts, G. Matthyssens, N. Van Meirvenne, and R. Hamers. 1989. A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol. **36:**169-176.

Gibson, W., and J. Stevens 1999. Genetic exchange in *Trypanosomatidae*. Adv. Parasitol. **43:**1-45.

Gibson, W., T. Backhouse, and A. Griffiths. 2002. The human serum-resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. Infec. Genet. Evol. **1:**207-214.

Gibson, W. C., T. F. de C Marshall, and D. G. Godfrey. 1980. Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. Adv. Parasitol. **18:**175-246.

Godfrey, D. G., R. D. Baker, L. R. Rickman, and D. Mehlitz. 1990. The distribution, relationships and identification of enzymic variants within the subgenus *Trypanozoon*. Adv. Parasitol. **29:**1-74.

Grigg, M. E., J. Ganatra, J. C. Boothroyd, and T. P. Margolis. 2001. Unusual abundance of atypical strains associated with human ocular toxoplasmosis. J. Infect. Dis. **184:**633-639.

Hawking F. 1973. The differentiation of *Trypanosoma rhodesiense* from *T. brucei* by means of human serum. Trans. R. Soc. Trop. Med. Hyg. **67:**517-527.

Hawking F. 1977. The resistance to human plasma of *Trypanosoma brucei*, *T. rhodesiense* and *T. gambiense*. Analysis of the composition of trypanosome strains. Trans. R. Soc. Trop. Med. Hyg. **70:**504-512.

Hide, G. 1999. History of sleeping sickness in East Africa. Clin. Microbiol. Rev. **12:**112-125.

Hide, G., N. Buchanan, S. Welburn, I. Maudlin, J. D. Barry, and A. Tait. 1991. *Trypanosoma brucei rhodesiense*: characterisation of stocks from Zambia, Kenya, and Uganda using repetitive DNA probes. Exp. Parasitol. **72:**430-439.

Hide, G., S. C. Welburn, A. Tait, and I. Maudlin. 1994. Epidemiological relationships of *Trypanosoma brucei* stocks from South-east Uganda: evidence for different population structures in human infective and non-human infective isolates. Parasitol. **109:**95-111.

Huys, G., L. Rigouts, K. Chemlal, F. Portaels, and J. Swings. 2000. Evaluation of amplified fragment length polymorphism analysis for inter- and intraspecific differentiation of *Mycobacterium bovis*, *M. tuberculosis*, and *M. ulcerans*. J. Clin. Microbiol. **38:**3675-3680.

Lanham, S. M., and D. G. Godfrey. 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. Exp. Parasitol. **28:**521-534.

MacLeod, A., A. Tweedie, S.C. Welburn, I. Maudlin, C. M. Turner, and A. Tait. 2000. Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. Proc. Natl. Acad. Sci. USA **97:** 13442-13447.

Malek, M., J. C. Dekkers, H. K. Lee, T. J. Baas, and M. F. Rothschild. 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mamm. Genome **12:**630-636.

Malek, M., J. C. Dekkers, H. K. Lee, T. J. Baas, K. Prusa, E. Huff-Lonergan, and M. F. Rothschild. 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. Mamm. Genome **12:**637-645.

Milner, J. D., and S. L. Hajduk. 1999. Expression and localisation of serum resistance associated protein in *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol. **104:** 271-283.

Milner, J.D., S. L. Hajduk. 1999. Expression and localisation of serum resistance associated protein in *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol. **104:**271-283.

Raper, J., M. P. Portela, E. Lugli, U. Frevert, and S. Tomlinson. 2001. Trypanosome lytic factors**:** novel mediators of human innate immunity. Curr. Opin. Microbiol. **4:**402-408.

Rifkin, M. R. 1978. *Trypanosoma brucei***:** some properties of the cytotoxic reaction induced by normal human serum. Exp. Parasitol. **46:**189-206.

Rifkin, M. R., C. De Greef, A. Jiwa, F. R. Landsberger, and S. Z. Shapiro. 1994. Human serum-sensitive *Trypanosoma brucei rhodesiense***:** a comparison with serologically identical human serum-resistant clones. Mol. Biochem. Parasitol. **66:**211-220.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. Molecular cloning, a laboratory manual. Second Edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor. New York.

Tait, A., J. D. Barry, R. Wink, A. Sanderson, and J. S. Crowe. 1985. Enzyme variation in *T. brucei ssp*. II. Evidence for *T. b. rhodesiense* being a set of variants of *T. b. brucei*. Parasitololgy 90**:**89-100.

Tomlinson, S., and J. Raper. 1996. The lysis of *Trypanosoma brucei brucei* by human serum. Nat. Biotechnol. **14:**717-721.

Van Meirvenne, N., D. Le Ray, M. Wery, and D. Afchain. 1975. Serological and parasitological diagnosis of sleeping sickness. Ann. Soc. Belg. Med. Trop. **55:**545-549.

Van Meirvenne, N., E. Magnus, and P. G. Janssens. 1976. The effect of normal human serum on trypanosomes of distinct antigenic type (ETat 1 to 12) isolated from a strain of *Trypanosoma brucei rhodesiense*. Ann. Soc. Belg. Med. Trop. **56:**55-63.

Ward, J. 1963. Hierachical grouping to optimise an objective function. J. Am. Stat. Assoc. **58:**236-244.

Welburn, S. C., K. Picozzi, E. M. Fevre, P. G. Coleman, M. Odiit, M. Carrington, and I. Maudlin. 2001. Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. Lancet **358:**2017-2019.

Xong, H. V., L. Vanhamme, M. Chamekh, C. E. Chimfwembe, J. Van Den Abbeele, A. Pays, N. Van Meirvenne, R. Hamers, P. De Baetselier, and E. Pays. 1998. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. Cell **95:**839-846.

CHAPTER 7

How does *Trypanosoma equiperdum* fit into the *Trypanozoon* group? A cluster analysis by RAPD and Multiplex-endonuclease genotyping approach

Claes, F [1, 5]., E. C. Agbo [2], M. Radwanska [1], M. F. W. te Pas [2], T. Baltz [3], D. T. de Waal [4], B. M. Goddeeris [5], E. Claassen [6], and P. Büscher [1]

[1] Prince Leopold Institute of Tropical Medicine, Department of Parasitology, Nationalestraat 155, Antwerpen, Belgium

[2] Institute for Animal Science and Health (ID-Lelystad), Division of Animal Sciences, Edelhertweg 15, 8200 AB Lelystad, The Netherlands

[3] Laboratoire de Parasitologie Moleculaire, Université Victor Ségalen Bordeaux II, 146, Rue Léo Saignat, 33076 Bordeaux, France

[4] Parasitology Division, Onderstepoort Veterinary Institute, Private Bag X05, Onderstepoort 0110, South Africa

[5] Faculty of Agriculture and Applied Biological Sciences, K. U. Leuven, Department of Animal Science, Kasteelpark Arenberg 30, 3000 Leuven, Belgium.

[6] Department of Immunology, Erasmus University Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Parasitology, In press

7.1. Summary

The pathogenic trypanosomes *T. equiperdum, T. evansi* as well as *T. brucei* are morphologically identical. In horses, these parasites are considered to cause respectively dourine, surra and nagana. Previous molecular attempts to differentiate these species were not successful for *T. evansi* and *T. equiperdum*; only *T. b. brucei* could be differentiated to a certain extent. In this study we analysed ten *T. equiperdum*, eight *T. evansi* and four *T. b. brucei* using Random Amplified Polymorphic DNA (RAPD) and the fine-scale multiplex-endonuclease genotyping approach (MEGA). The results obtained confirm the homogeneity of the *T. evansi* group tested. The *T. b. brucei* clustered out in a heterogenous group. For *T. equiperdum* the situation is more complex**:** eight out of ten *T. equiperdum* clustered together with the *T. evansi* group, while two *T. equiperdum* strains were more related to *T. b. brucei*. Hence, two hypotheses can be formulated**:** (1) only two *T. equiperdum* strains are genuine *T. equiperdum* causing dourine; all other *T. equiperdum* strains actually are *T. evansi* causing surra or (2) *T. equiperdum* does not exist at all. In that case, different clinical outcome of horse infections with *T. evansi* or *T. b. brucei* is primarily related to the host immune response.

7.2. Introduction

Dourine, Surra and Nagana are all lethal diseases in horses caused by *Trypanosoma equiperdum*, *T. evansi* and *T. b. brucei*, respectively (Office International des Epizooties, OIE list B). They are all members of the *Trypanozoon* subgenus and have morphologically identical bloodstream forms. *Trypanosoma equiperdum* and *T. evansi* are transmitted respectively by sexual contact and by blood sucking flies explaining their worldwide distribution, while the dependence on tsetse flies as the vector limits *T. b. brucei* to sub-Saharan Africa (Stephen, 1986).

Most research on the genome of pathogenic Salivarian trypanosomes is performed on *T. b. brucei.* The genomes of *T. equiperdum* and *T. evansi* have not been thoroughly studied and most investigations focus on the sequence of variable surface glycoproteins (VSGs) (Baltz et al., 1986; Roth et al., 1986; Urakawa et al., 2001), on expression sites (Florent, Raibaud and Eisen, 1991), and on the kinetoplast DNA (kDNA) (Riou and Saucier, 1979; Frasch et al., 1980; Borst, Fase-Fowler and Gibson, 1987; Masiga and Gibson, 1990; Ou, Giroud and Baltz, 1991; Lun, Brun and Gibson, 1992).

Despite numerous attempts, researchers have not been able to differentiate *T. equiperdum* from *T. evansi* consistently, neither at the serological, nor at the molecular level (Baltz, PhD thesis, Université de Bordeaux II, 1982; Hide et al., 1990; Lun et al., 1992; Lun, Brun and Gibson, 1992; Biteau et al., 2000). Previous studies performed in our laboratory further underlines the close relationship between both species (Claes et al., 2002). Only two of the ten putative *T. equiperdum* strains, the BoTat 1.1 (Morocco) and the Onderstepoort Veterinary Institute (OVI) strain (South Africa), seem to differ from the rest of the *T. equiperdum* strains in Variable Antigen Repertoire. All other *T. equiperdum* have the same characteristics as *T. evansi* strains. In the present study, we examined the characteristics of several *T. equiperdum, T. evansi* and *T. b. brucei* populations with two molecular techniques, Random Amplified Polymorphic DNA (RAPD) and a recently developed multiplex-endonuclease genotyping analysis (MEGA) (Agbo et al., In press).

7.3. Materials and Methods

7.3.1. Trypanosome populations

A collection of four *T. b. brucei*, eight *T. evansi* and ten *T. equiperdum* populations, derived from strains isolated all over the world, was used in this study (Table 1). All populations were kept as cryostabilates in liquid

nitrogen. For the *T. equiperdum* strains, the history is mostly unknown. Only the OVI strain from South Africa, was well documented.

7.3.2. Preparation of trypanosome DNA

Bloodstream form trypanosomes were expanded in mice and rats and were purified from the blood by DEAE chromatography (Lanham and Godfrey, 1970), followed by repeated centrifugation (3 times 20 min., 2000g) and sediment washes with Phosphate Buffered Saline Glucose (PSG) (38mM $Na_2HPO_4.2H_20$, 2mM $NaHPO_4$, 80mM glucose). Finally, trypanosome pellets were subsequently stored at -80°C. Twenty µl of trypanosome pellets (approximately 2 x 10^7 cells) were resuspended in 200 µl of Phosphate Buffered Saline (PBS) and the trypanosome DNA was extracted using the QIAamp DNA mini kit (Westburg, Leusden, The Netherlands), resulting in pure DNA in 200 µl of Milli-Q water. The typical yield of DNA extracted from a 20 µl pellet was 150 ng/µl or 30 µg of total DNA. The extracts obtained were diluted in Milli-Q water to a standard concentration of 50 ng/µl and stored at -20°C.

7.3.3. Random Amplified Polymorphic DNA (RAPD)

Ten µl of extracted DNA (50 ng/µl) were mixed with 40 µl of a PCR-mix containing**:** 0.5 U Taq DNA recombinant polymerase (Promega, UK), PCR buffer (Promega, UK), 3.0 mM $MgCl_2$ (Promega, UK), 200 µM of each of the four dNTPs (Roche, Mannheim, Germany) and 0.5 µM of the oligonucleotide 10-mer (Gibco BRL, UK). The different oligonucleotides used were (in 5'-3' direction)**:** RAPD 606 CGG TCG GCC A (Ventura et al., 2001) and RAPD ILO 525 CGG ACG TCG C (Waitumbi and Murphy, 1993). Amplifications were performed in a Biometra® Trio-block thermocycler. Cycling conditions were as follows: denaturation for 4 min. at 94ºC, followed by 40 amplification cycles of 2 min. denaturation at 94ºC, 2 min. primer-template annealing at 40ºC and 2 min. polymerisation at 72ºC. A final elongation step was carried out for 5 min. at 72°C. Twenty µl of the PCR product and ten µl of a 3 kb size marker (MBI Fermentas, Germany) were subjected to electrophoresis in a 2 % agarose gel (90 min at 100V).

Gels were stained with ethidium bromide (0.5µg/ml) (Sigma, USA) and analysed on an Imagemaster Video Detection System (Pharmacia, UK).

Table 1. Trypanosome populations used in these studies

Species	Clone/strain	ITMAS	Origin	Year	Host
T. b. brucei	AnTat 2.2	100297B	Nigeria	1970	Tsetse fly
T. b. brucei	AnTat 5.2	220197	The Gambia	1975	Cattle
T. b. brucei	AnTat 17.1	210596	R. D. Congo	1978	Sheep
T. b. brucei	KETRI 2494	270881	Kenya	1980	Tsetse fly
T. evansi	AnTat 3.1	070799	South America	1969	Capybara
T. evansi	RoTat 1.2	020298	Indonesia	1982	Water buffalo
T. evansi	Merzouga 56	120399D	Morocco	1998	Camel
T. evansi	Zagora I.17	040399B	Morroco	1997	Camel
T. evansi	KETRI 2480	110297	Kenya	1980	Camel
T. evansi	CAN 86 K	140799B	Brazil	1986	Dog
T. evansi	Stock Colombia	150799	Colombia	1973	Horse
T. evansi	Stock Vietnam	101298	Vietnam	1998	Water buffalo
T. equiperdum	AnTat 4.1	210983A	unknown	unknown	unknown
T. equiperdum	Alfort	241199A	unknown	unknown	unknown
T. equiperdum	SVP	241199B	unknown	unknown	unknown
T. equiperdum	ATCC 30019	020301	France	1903 ?	Horse
T. equiperdum	ATCC 30023	280201	France	1903 ?	Horse
T. equiperdum	STIB 818	010999	P. R. China	1979	Horse
T. equiperdum	American	220101	unknown	unknown	unknown
T. equiperdum	Canadian	290101	unknown	unknown	unknown
T. equiperdum	OVI	241199C	South Africa	1975	Horse
T. equiperdum	BoTat 1.1	240982A	Morocco	1924	Horse

7.3.4. Multiplex-endonuclease fingerprinting method

A fine-scale genotyping approach involving multiplex endonucleases in combination with a pair of cognate adapters was used according to Agbo et al (In press). Briefly, 100-250 ng genomic DNA were digested for 4 h using 10 U each of *Bgl*II, *Bcl*I, *Acs*I and *Mun*I endonucleases in two successive double digestion reactions. The final digestion products were precipitated and reconstituted in 10 µl distilled water. Ten µl of a buffer containing 660 mM Tris HCl, 50 mM $MgCl_2$, 10 mM Dithiothreitol, 10mM ATP, pH7.5, and 20 pM of each *Bgl*II ($^{5'-}$CGGACTAGAGTACACTGTC; $^{5'-}$GATCGACAGTGTACTCTAGTC) and
*Mun*I ($^{5'-}$AATTCCAAGAGCTCTCCAGTAC; $^{5'-}$AGTACTGGAGAGCTCTTG) adapters were added. The *Bgl*II adapter also ligated to the overhang sites created by *Bcl*I, while *Mun*I adapter also ligated to the *Acs*I site. One µl (400U) of T4 DNA ligase (New England Biolabs) was added and the mixture incubated for 2 h at 25ºC. Pre-selective amplification was performed in a total volume of 20 µl containing 4 µl of 1:1-diluted ligation product, 1 U of Taq polymerase (Roche Molecular Biochemicals, Almere, The Netherlands), 10X PCR buffer (100mM Tris HCl pH 9.0, 50 mM KCl, 1% triton X-100, 0.1% w/v gelatin), 2.5 mM $MgCl_2$, 200 µM of each dNTP and 5 pM of each *Bgl*II ($^{5'-}$GAGTACACTGTCGATCT) and *Mun*I ($^{5'-}$GAGAGCTCTTGGAATTG) primers. The reaction mix was incubated for 2 min at 95ºC, and subjected to 20 cycles of PCR (30 s at 95ºC, 30 s at 56ºC and 2 min at 72ºC). Four µl of 1:20-diluted pre-selective products were used as template for selective reaction with *Mun*-0/*Bgl*-A selective primer combination (in which the *Mun* primer was fluorescently labelled). The PCR program was essentially the same as for pre-selective amplification, except that the last cycling step was followed by 30-min incubation at 60ºC. The final products were diluted 1:1 with TE, and Genescan-500 internal lane standard (PE Applied Biosystems) was added. One µl of the mix was resolved in a 7.3% denaturing sequencing gel using a model ABI 373A automated DNA sequencer.

Gels were routinely prepared by using ABI protocols and electrophoresed for 5 h. Gel patterns were collected with GenScan software (PE Applied Biosystems) and sample files were transferred to GelCompar II software (Applied Maths, Kortrijk, Belgium).

7.3.5. Cluster analysis

The GelCompar II program was used for cluster analysis of RAPD and MEGA fingerpint profiles by Unweighted Pair Group Method with Arithmetic Mean (UPGMA) based on the Dice coefficient. With the obtained data matrices, Wagner Parsimony analysis was performed on bootstrapped data using the Seqboot, Mix and Consense programs from the PHYLIP software (Felsenstein, 1989).

Fig. 1. UPGMA Cluster analysis based on the RAPD results with primer 606.

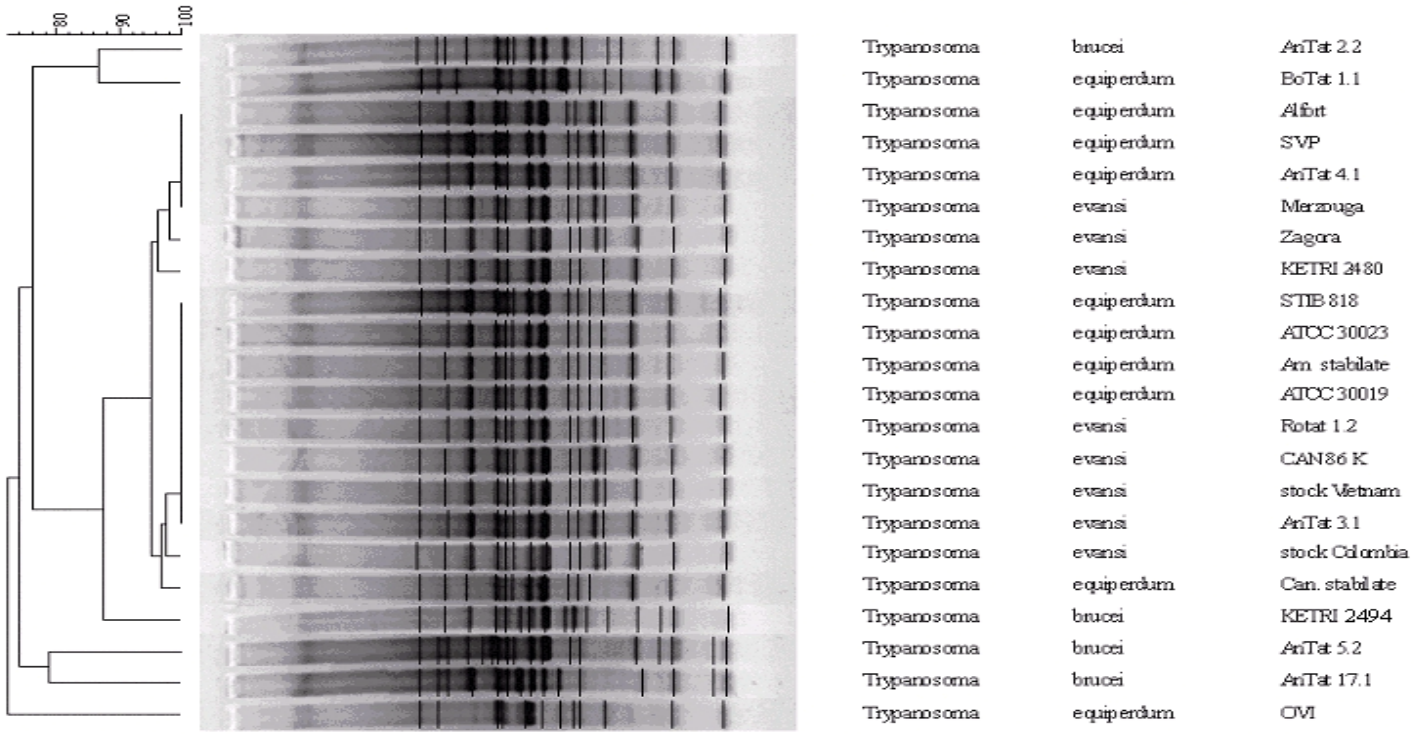
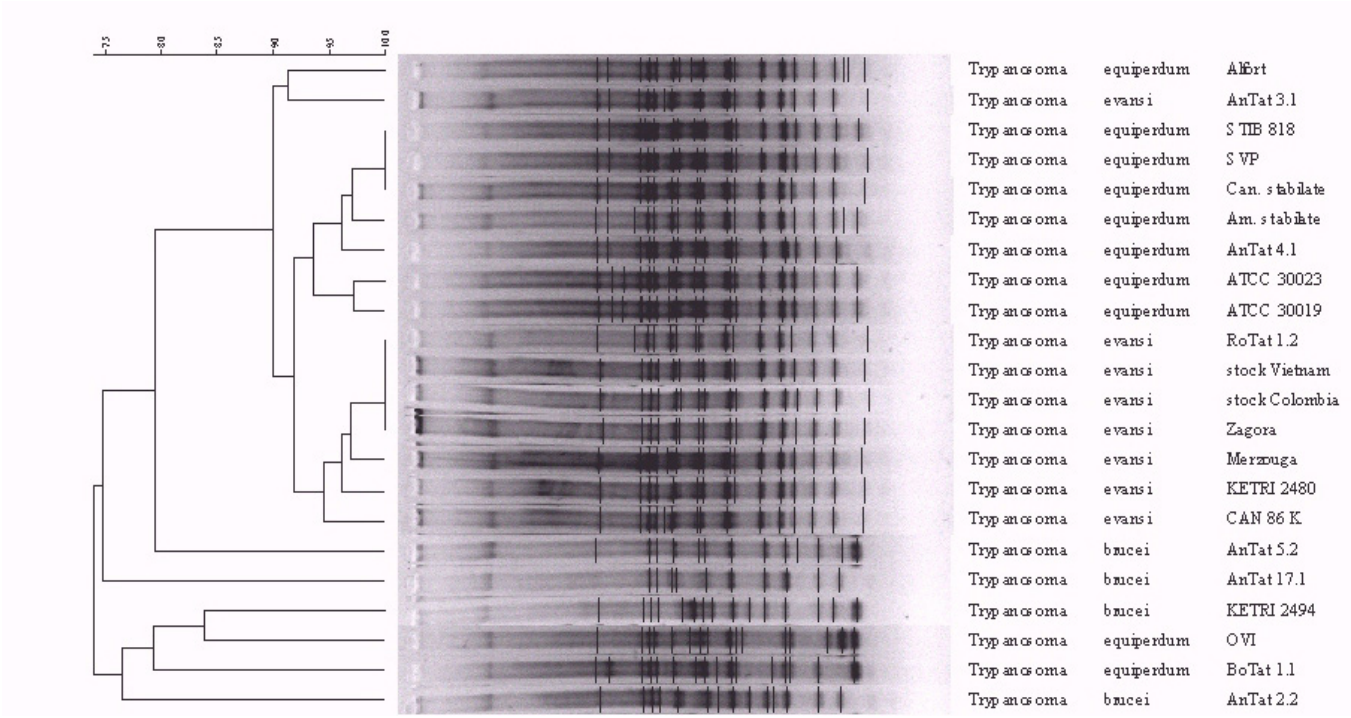


Fig. 2. UPGMA Cluster analysis based on the RAPD results with primer ILO 525.



7.4. Results

RAPD and MEGA reactions were performed with the same set of samples on different days and by different persons. The DNA banding patterns obtained confirmed the repeatability of both techniques in our laboratory (data not shown). Dendrograms from the RAPD results, analysed by pairwise fragment comparison using the Dice coefficient and by data clustering using UPGMA, are shown in Figures 1 and 2. In RAPD 606, all *T. evansi* strains cluster out in one homogenous group with a 95-100 % similarity level. Also, in this cluster eight out of the ten tested *T. equiperdum* strains are found. All *T. b. brucei* and the two remaining *T. equiperdum* isolates (BoTat 1.1 and OVI) cluster out in a more hetereogenous way (with 72-88% similarity coefficiënt). The similarity level between these two *T. equiperdum* strains is 75% and they have respectively a 76 and 74% similarity coefficient with the *T. evansi/T. equiperdum* cluster. *T. equiperdum* BoTat 1.1 shares the highest similarity with *T. b. brucei* AnTat 2.2, while the OVI strain is distinct from the rest of the group. In RAPD ILO 525, the *T. evansi* strains are grouped in one cluster with 90-100% similarity. This cluster harbours the same eight out of ten *T. equiperdum* strain. *T. b. brucei* forms a more heterogenous groups (Dice coefficients ranging from 74 to 83%) including the *T. equiperdum* Botat 1.1 and OVI. With this RAPD *T. equiperdum* BoTat 1.1 relates most to *T. equiperdum* OVI and *T. b. brucei* KETRI 2494 and AnTat 2.2; OVI is highly similar to KETRI 2494.

In the UPGMA clustering data obtained from the MEGA branches of the homology tree are longer, indicating the higher resolution power of this technique (Fig. 3). All *T. evansi* are grouped in one cluster with a similarity of 85-95 %, together with the same eight *T. equiperdum* strains. Also with this technique the *T. b. brucei* group appeared as a heterogenous cluster, including the BoTat 1.1 and OVI strains. Based on the MEGA data, the level of similarity of these two latter strains was calculated at 74 %. In this analysis, OVI seems closely related to *T. b. brucei* KETRI 2494, while BoTat 1.1 shares more homology with *T. b. brucei* AnTat 2.2.

In the RAPD and multiplex-endonuclease genotyping analyses, the clustering of the strains did not seem to be correlated to their respective geographical origin, original host species or the year of isolation. In RAPD 606, *T. evansi* and *T. equiperdum* from different regions and hosts (RoTat 1.2, AnTat 3.1, CAN 86K, stock Vietnam, STIB 818, American stabilate, ATCC 30019 and ATCC 30023) gave a 100% similarity coefficiënt (Fig. 1). On the other hand, with RAPD ILO 525 (Fig. 2), *T. evansi* stocks from different origins (RoTat 1.2, stock Colombia, Zagora I.17) showed exactly the same pattern. When mixed parsimony analysis was performed on bootstrapped data from both RAPD's and the modified AFLP, the homogenous *T. evansi*/*T. equiperdum* cluster differed from the more heterogenous group with a 80% and 100% probability coëfficient, respectively for the MEGA and both RAPD's (data not shown).

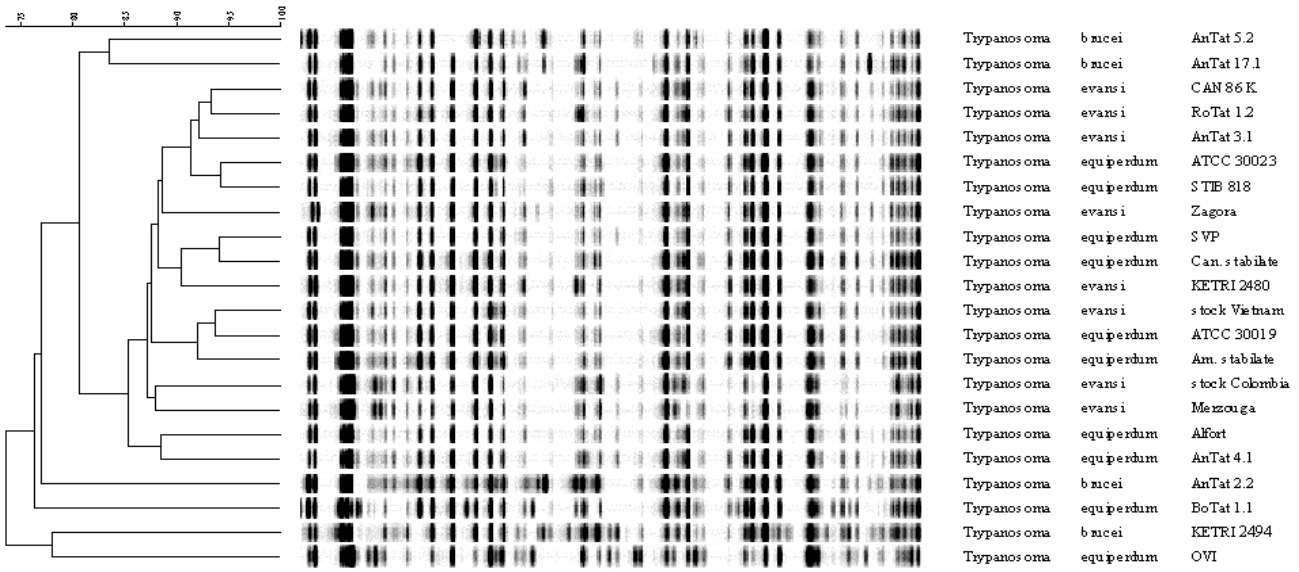
7.5. Discussion

Comparison of the RAPD 606 results in the present study with those from Ventura et al. (2001), reveals a similar close genetic relationship between *T. evansi* populations from different origins, and approximately the same distance between *T. equiperdum* BoTat 1.1 and the *T. evansi* cluster (76% similarity versus 60%, respectively). With RAPD ILO 525, Waitumbi and Murphy (1993) were able to divide the *Trypanozoon* subgenus into three groups: (1) *T. b. brucei* and *T. b. rhodesiense*, (2) *T. b. gambiense* and (3) *T. evansi.* No *T. equiperdum* was included in their analysis.

Other previous characterisation studies mainly focused on the *T. brucei* subspecies or on *T. evansi* and only few *T. equiperdum* were included. Hide et al. (1990), analysed 42 *T. brucei* by repetitive DNA probes, together with only one *T. equiperdum* and one *T. evansi*. A separate *T. b. gambiense* type I cluster was found while *T. b. brucei* and *T. b. rhodesiense* were more heterogenous. The *T. equiperdum* and *T. evansi* appeared to have a dissimilarity level of 56% with the *T. brucei* group and a dissimilarity

level of 45% between each other. Unfortunately from their paper, it is not clear which *T. equiperdum* and *T. evansi* strains were used. By both kDNA and isoenzyme analysis, Lun et al. (1992a, 1992b) could not find differences which would distinguish 12 stocks of *T. evansi* from one *T. equiperdum* (STIB 818). Agbo et al. (2002) included two *T. evansi* (AnTat 3.1 and RoTat 1.2) and two *T. equiperdum* (AnTat 4.1 and STIB 818) in their AFLP analysis of *Trypanosoma spp.*, again without conclusive results on the differentiation between *T. evansi* and *T. equiperdum*. Using microsatellite markers on three *T. equiperdum* (BoTat 1.1, STIB 818 and a South African strain), Biteau et al. (2000) observed heterogenous patterns amongst them and conclude that previous interpretation of the close relationship of *T. evansi* and *T. equiperdum* by isoenzyme and RFLP analysis might have been simplisitic. Only Zhang and Baltz (1994) found some differences between *T. equiperdum* and *T. evansi* stocks using repetitive DNA probes. BoTat 1.1 and a South African strain were separated from the *T. evansi* group. They were more similar to *T. b. brucei* than to the *T. evansi* cluster which contained a third *T. equiperdum* (STIB 818). Zhang and Baltz concluded that this outlier *T. equiperdum* STIB 818 could reflect the limit of sensitivity of the used RFLP technique or could be due to the misclassification of this strain. To our knowledge, the South African strain is the Onderstepoort Veterinary Institute (OVI) strain (T. De Waal, personal communication). Taken together, our results correspond well with the present studies that were based on a larger collection of *T. evansi* and *T. equiperdum* strains. Irrespective of the DNA amplification method, two major groups can be formed: one homogenous group including all *T. evansi* and most of the *T. equiperdum* strains and one heterogenous group including all *T. b. brucei* and two *T. equiperdum*, the BoTat 1.1 clone and the OVI strain. Previous serological and molecular studies on the same collection yielded similar results**:** all hitherto tested *T. evansi* share the presence and expression of the RoTat 1.2 VSG gene, while for the screened populations of *T. equiperdum* only BoTat 1.1 and OVI, as well as all tested *T. b. brucei*, do not express nor contain this VSG gene (Claes et al., unpublished observations).

Fig. 3. UPGMA Cluster analysis based on the modified AFLP results.



Combining these data, two hypotheses can be formulated. Firstly, BoTat 1.1 and the OVI strain are the only genuine *T. equiperdum* while all other *T. equiperdum* actually are misclassified *T. evansi,* thus extending the view of Zhang and Baltz (1994). Indeed, in experimental infections with the OVI strain by Barrowman (1976), clinical signs of dourine signs were observed in the infected horses, while in experimental infections with the American and Canadian stabilates, which in our study are both found in the *T. evansi* cluster, only general signs of trypanosomiasis were observed (Hagebock et al., 1993). Unfortunately, for most *T. equiperdum* strains, including BoTat 1.1, similar clinical experiments have not been performed. The question whether BoTat 1.1 and OVI are "real" *T. equiperdum* strains could be solved by following the clinical outcome of horses experimentally infected with both strains and by comparing the result with infections with *T. evansi*-like *T. equiperdum* strains. However, one should keep in mind that strains that have undergone multiple passages in laboratory animals might have lost or changed their pathogenicity and virulence. Alternatively, specific serological or molecular markers could be identified which can differentiate *T. equiperdum* from *T. b. brucei.* In the absence of a conclusive *T. b. brucei* marker this remains a challenging issue. An alternative hypothesis is that the species *T. equiperdum* actually does not exist but that dourine is caused by particular strains of *T. evansi* and/or *T. b. brucei.* Then the clinical outcome of the infection would merely depend on the virulence or the tissue tropism of the infective strains or the immunological response of the individual host to the trypanosome infection.

Indeed, in horses acute, subacute and chronic forms of nagana (*T. b. brucei*) have been described, sometimes with clinical signs such as oedema of prepuce and legs, and sporadically the appearance of urticarial plaques. Also for surra (*T. evansi*) in horses, both acute and chronic infections have been mentioned. Here also, oedema is evident as plaques on the ventral surface of the body or as swelling of the sheath or prepuce (Stephen, 1986). All together, some clinical signs of nagana and surra are shared with dourine, i.e. urticarial plaques and genital swellings.

Therefore, it might be that differential diagnosis based only on clinical signs is not conclusive for the infecting trypanosome species and certain chronic cases of nagana or surra might have been considered as dourine or *vice versa*. This enigma would be solved if one considers dourine as the chronic form of both diseases. Both hypotheses should be checked against the other biological characteristics of the three trypanosome species. Until now, *T. b. brucei* is considered to be only cyclically-transmitted by tsetse flies, while *T. evansi* and *T. equiperdum* are transmitted mechanically and sexually, respectively. However, we have no evidence to exclude sexual or mechanical transmission of *T. b. brucei*. The transmission route could even be linked to host specificity and virulence or tissue tropism, and for *T. evansi,* sexual transmission might also occur. Hagebock et al. (1993) were indeed able to infect horses by urogenital inoculation with the American and Canadian *T. equiperdum* strains, which in our study clustered together with *T. evansi*. Thus, if these strains are considered to be *T. evansi*, a first proof of sexual transmission of *T. evansi* in horses has been shown. Nevertheless, to prove the possibility of sexual transmission of *T. b. brucei* and *T. evansi*, more experimental infections with both species should be conducted. Based on the available clinical, serological and molecular data, we propose that there is not sufficient evidence for the existence of *T. equiperdum* as a separate species. To clarify the confusion about *T. equiperdum* taxonomy, we propose to isolate new trypanosome strains from well-defined dourine, surra and nagana cases in horses, to analyse them with the most performant serological and molecular techniques and to study their pathogenicity and transmission routes in horses.

7.6. Acknowledgements

This study received financial support from the Institute for the Promotion of Innovation by Science and Technology in Flanders, Belgium (IWT). *Trypanosoma spp.* stabilates were kindly provided by: Dr. Reto Brun, Swiss Tropical Institute Basel, Switzerland; Dr. Peter-Henning Clausen, Free University Berlin, Germany; Dr. Joyce Hagebock† and Dr. David Kinker, National Veterinary Services Laboratories, United States Department of Agriculture, USA; Dr. Lun, Zhongshan University, P.R. China; and Dr. Zablotsky, All-Russian Research Institute for Experimental Veterinary Medicine (VIEV), Russia. We also want to thank Dr. Rigouts (ITM, Mycobaterium laboratory) for the use of the Gelcompar II program.

References

Agbo, E. C., P. A. Majiwa, Claassen H. J., and M. F. te Pas. 2002. Molecular variation of *Trypanosoma brucei* subspecies as revealed by AFLP fingerprinting. Parasitology **124:**349-358.

Baltz, T., C. Giroud, D. Baltz, C. Roth, A. Raibaud, and H. Eisen. 1986. Stable expression of two variable surface glycoproteins by cloned *Trypanosoma equiperdum*. Nature **319:**602-604.

Barrowman, P. R. 1976. Experimental intraspinal *Trypanosoma equiperdum* infection in a horse. Onderstepoort J. Vet. Res. **43:**201-202.

Biteau, N., F. Bringaud, W. Gibson, P. Truc, and T. Baltz. 2000. Characterisation of *Trypanozoon* isolates using a repeated coding sequence and microsatellite markers. Mol. Biochem. Parasitol. **105:**185-201.

Borst, P., F. Fase-Fowler, and W. C. Gibson. 1987. Kinetoplast DNA of *Trypanosoma evansi*. Mol. Biochem. Parasitol. **23:**31-38.

Claes, F., D. Verloo, D. T. de Waal, T. Urakawa, P. Majiwa, B. M. Goddeeris, and P. Büscher. 2002. Expression of RoTat 1.2 cross-reactive Variable Antigenic Type in *Trypanosoma evansi* and *T. equiperdum*. Annals New York Acad. Sci. **969:**174-179.

Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics **5:**164-166.

Florent, I. C., A. Raibaud, and H. Eisen. 1991. A family of genes related to a new expression site-associated gene in *Trypanosoma equiperdum*. Mol. Cell Biol. **11:** 2180-2188.

Frasch, A. C. C., S. L. Hajduk, J. H. J. Hoeijmakers, P. Borst, F. Brunel, and J. Davidson. 1980. The kinetoplast DNA of *Trypanosoma equiperdum.* Biochim. Biophys. Acta **607:**397-410.

Hagebock, J. M., L. Chieves, W. M. Frerichs, and C. D. Miller. 1993. Evaluation of agar gel immunodiffusion and indirect fluorescent antibody assays as supplemental tests for dourine in equids. Am. J. Vet. Res. **54:**1201-1208.

Hide, G., P. Cattand, D. Le Ray, J. D. Barry, and A. Tait. 1990. The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. Mol. Biochem. Parasitol. **39:**213-226.

Lanham, S. M., and D. G. Godfrey. 1970. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. Exp. Parasitol. **28:**521-534.

Lun, Z-R., R. Allingham, R. Brun, and S. M. Lanham. 1992. The isoenzyme characteristics of *Trypanosoma evansi* and *Trypanosoma equiperdum* isolated from domestic stocks in China. Ann. Trop. Med. Parasitol. **86:**333-340.

Lun, Z-R., R. Brun, and W. Gibson. 1992. Kinetoplast DNA and molecular karyotypes of *Trypanosoma evansi* and *Trypanosoma equiperdum* from China. Mol. Biochem. Parasitol. **50:**189-196.

Masiga, D. K., and W. C. Gibson. 1990. Specific probes for *Trypanosoma (Trypanozoon) evansi* based on kinetoplast DNA minicircles. Mol. Biochem. Parasitol. **40:**279-284.

Ou, Y. C., C. Giroud, and T. Baltz. 1991. Kinetoplast DNA analysis of four *Trypanosoma evansi* strains. Mol. Biochem. Parasitol. **46:**97-102.

Riou, G. F., and J. M. Saucier. 1979. Characterisation of the molecular components in kinetoplast-mitochondrial DNA of *Trypanosoma equiperdum*. J. Cell Biol. **82:**248-263.

Roth, C. W., S. Longacre, A. Raibaud, T. Baltz, and H. Eisen. 1986. The use of incomplete genes for the construction of a *Trypanosoma equiperdum* variant surface glycoprotein gene. EMBO J. **5:**1065-1070.

Stephen, L. E. 1986. *Trypanosomiasis: A veterinary perspective*. Pergamon press, Oxford.

Urakawa, T., D. Verloo, L. Moens, P. Büscher, and P. A. O. Majiwa. 2001. *Trypanosoma evansi:* cloning and expression in *Spodoptera fugiperda* insect cells of the diagnostic antigen RoTat 1.2. Exp. Parasitol. **99:**181-189.

Ventura, R. M., G. F. Takeda, R. A. M. S. Silva, V. L. B. Nunes, G. A. Buck, M. M. G. Teixeira. 2001. Genetic relatedness among *Trypanosoma evansi* stocks by random amplification of polymorphic DNA and evaluation of a synapomorphic DNA fragment for species-specific diagnosis. Int. J. Parasitol. **32:**53-63.

Waitumbi, J. N., and N. B. Murphy. 1993. Inter- and intra-species differentiation of trypanosomes by genomic fingerprinting with arbitrary primers. Mol. Biochem. Parasitol. **58:**181-186.

Zhang, Z. Q., and T. Baltz. 1994. Identification of *Trypanosoma evansi, Trypanosoma equiperdum* and *Trypanosoma brucei brucei* using repetitive DNA probes. Vet. Parasitol. **53:**197-208.

CHAPTER 8

General Discussion

8.1. Control of African trypanosomosis

Control of tsetse-transmitted trypanosomosis in humans, livestock and wildlife has involved, to varying degrees, interventions directed at the vector (tsetse fly), the infectious organism (trypanosomes), the disease (trypanosomosis), reservoirs (wild and domestic animals) and the hosts (humans and livestock). The ideal control will comprise of an effective vaccine, complemented by a durable therapeutic agent, and a robust and rapid diagnostic tool for molecular epidemiology. No vaccine has yet reached clinical stage of development, and peculiarities of the parasite suggest that classical approaches to vaccine development will be unsuccessful (Chapter 1). The use of trypanocidal drugs to control the disease is well established and represents the most widely used approach. However, increasing trends in drug resistance (reviewed by Anene et al., 2001) represent the most serious setback in the use of chemotherapeutic agents, and motivate search for new agents and drug targets. In addition, the development of a universal molecular diagnostic system, an important complementary tool for molecular epidemiology, is currently being actively pursued. The possibilities of discovering newer effective drugs and the development of a vaccine will constitute a potential breakthrough in the quest for reducing trypanosomosis. Current advances in genome analysis are offering new tools required for identifying, developing and validating candidate targets.

The thesis began with a general review of our current understanding of African trypanosome genetics and the available tools for molecular

diagnosis and therapy. It linked the impact of genomics-based technologies to their potential impact on developing new tools for specific diagnostics and the identification of new drug targets, and the prospects for a vaccine (Chapter 1).

8.2. The importance of genetic characterisation

Genetic characterisation of trypanosomes allows their identification at the strain (epidemiological tracking) and (sub)species (taxonomic) levels. A correlation of such genotypic information may also permit the study of the impact of genetic diversity on specific biomedical parameters such as virulence or resistance to drugs. The accurate identification of *T. brucei* strains has been a challenging problem in the epidemiology of African trypanosomosis, both human and animal. Genotypic variation of the ribosomal RNA (rRNA) genes was used to elucidate the phylogenetic relatedness of *Trypanosoma brucei brucei* and *T. b. gambiense* with the aim to derive species-specific DNA probes (Chapter 2). Restriction fragment length polymorphisms (RFLP)-based variations in the rRNA gene region helped us to show that *T. b. gambiense* may indeed be a host variant of *T. b. brucei*. According to the molecular signatures, the subspecies were clearly separated in different clades in the phylogenetic analysis. This approach can be used to rapidly determine the taxonomic position of unidentified species or to give an objective and invariable criterion to characterise *T. brucei*, for example, to develop a molecular phylogeny of these subgroups in relation to their human serum response phenotypes. Similar DNA fingerprinting based on single-locus RFLP analysis has allowed a determination of identity of genotypes (Paindovoine et al., 1989; Hide et al., 2000). However, genotypic analysis in these approaches tends to be laborious if many loci need to be sampled.

8.3. Assessing genetic diversity using random molecular markers

DNA polymorphisms can provide a set of markers that has to be sufficiently numerous and spaced across the genome to be informative. Such markers have the additional advantages that they can all be typed by the same technique, and their chromosomal location can be found easily by physical mapping. The polymorphism information content, which measures how informative a marker is, is increased when more polymorphisms are detected in an analysis. We applied the Amplified Fragment Length Polymorphisms (AFLP) method (Vos et al., 1995) to genotype isolates of the three *T. brucei* subspecies (Chapter 3) for several reasons. First, the approach of RFLP analysis of the rRNA gene region detects only a single genomic locus in which recombination is known to occur in other taxa. Trypanosome genome is known to be highly plastic and reflect a high rate of recombination and point mutation, and any genotype based on a single locus may provide inadequate results. AFLP markers are PCR-based RFLP analysis that can screen over 50 independent loci per PCR. This provides the opportunity of detecting a large number of genome-wide molecular markers that can be characterised to generate information on their arrangement in the genome. Also, AFLP markers are useful for assessing genetic differences among populations, individuals, strains or clonally-related species. For a wide range of taxa, AFLP markers have been used to uncover cryptic genetic variation of strains, or related species (Janssen et al., 1996; Heun et al., 1997; Ajmone-Marsan et al., 1997).

AFLP-based genomic differences can be detected if they resulted in acquisition or loss of a restriction fragment, a marked size difference between allelic restriction fragments in two DNA samples, or as mutations at primer extension sites. Such size differences could be the result of insertion or deletion events within a restriction fragment. Nonetheless, with the conventional AFLP method only limited genotypic differences were

observed among trypanosome subspecies, due to close genetic relatedness at the intraspecific level (Chapter 3). Therefore, in resolving extremely low genetic variability, e.g. subspecific or intraspecific differences, or a confirmation of a lack of differences for clonality of individuals, a genetic fingerprinting approach will have to meet a few general criteria. It must generate multiple independent markers, provide adequate resolution of genetic differences, be reliable (repeatable) and require no prior sequence information.

Using multiple independently-accessing restriction enzymes, the number of polymorphic sites that can be exploited in a single experiment by PCR-based RFLPs were expanded and the resulting heterosite fragments detected in a robust way (Chapter 4). Amplified restriction products, or amplicons, were size-fractionated by polyacrylamide gel electrophoresis, but capillary electrophoresis can now be used to maximise fragment separation and sensitivity of the method. This multiplex endonuclease genotyping approach (MEGA) provides the opportunity of saturating and detecting unlinked markers at much higher density within the genome than the conventional AFLP. The cloning of polymorphic DNA markers and their subsequent conversion into sequence tagged site (STS) assays could permit quick and reliable analysis. Restriction endonuclease-based markers suffer from their general dominant nature and the limited genetic interpretation of the alleles and loci involved. However, detailed pedigree information allows the identification of co-dominant markers, which permit precise estimation of allelic frequencies and more powerful population genetic studies. The development of a genetic linkage map for *T. brucei* based on conventional AFLP is currently being pursued (Tait et al., 2002) and will make it possible to perform classicial genetic studies to search for genes that determine specific traits. Besides the assessment of genetic diversity, the MEGA approach is especially suited to enrich for high resolution of *T. brucei* genetic map and as a resource in marker-enrichment strategies for target loci, based on specific intervals of restriction enzyme recognition sites.

The key questions in trypanosome population genetics are concerned with gene flow within and between populations (MacLeod et al., 2001). Several marker-based systems have been used to study the population structures of trypanosome populations (Paindovoine et al., 1989; Hide et al., 1994), and the plausible and limiting attributes of each system have been highlighted (MacLeod et al., 2001). Genomic fingerprinting may become problematic because the high variability of genetic markers reduces similarity between distant taxa to the level of chance. Thus, the usefulness of multiplex genotyping markers for systematics rests more in the rapid grouping of closely-related populations, which is crucial for epidemiological studies. In Chapter 5, the MEGA approach was used to study the population genetic structure of *T. brucei* subspecies and to perform a cladistic analysis of these trypanosomes from geographically and temporally matched hosts in order to determine stock relationships and evolution. This analysis groups organisms that share derived characters more closely than those that do not. Important information on the genetic background of individual stocks and populations were derived from genotype relationships that were observed between marker profiles.

A major advantage of fine-scale genetic fingerpinting is that it gives a direct access to genotypic information rather than requiring inference from the phenotype. However, a major challenge is determining the contribution of sequence variants (both within and between trypanosome species) to phenotypes and diseases. The analysis of inheritance of specific trypanosome phenotypes has been undertaken only rarely (Gibson and Mizen, 1997), although sufficient pedigree clones are currently available (Tait et al., 2002). Fine genotyping of human serum-resistant and sensitive *T. b. rhodesiense* clones characterised on the basis of maximum and minimum variances between and within groups revealed clear genomic differences (Chapter 6). Multilocus genotyping based on methylation-sensitive endonuclease polymorphisms and a classification strategy of combining genotype information with human serum response (HSR) trait

produced compact well-differentiated groups of genotypes (Chapter 6). These results indicate that genotype-HSR phenotype correlation exists among *T. brucei* clones. However, some of these genetic variations may be silent, synonymous polymorphisms and thus not accompanied by significant variations in the proteome. Still, it is conceivable that these variations can give rise to differences in *T. brucei* phenotype, which might influence the outcome of infection. Furthermore, epigenetic status of genomic DNA and its variation in disease is well-known, recently reviewed by Fraga and Esteller (2002). For example, the role of DNA methylation in modulating cellular activities in eukaryotes, such as differential gene expression, is well-established (Busslinger et al., 1983; Razin and Cedar, 1991; Razin and Cedar, 1994). An adapter-mediated AFLP method with two endonucleases has been used to correlate differential methylation pattern differences and fungal growth properties (Reyna-Lopez et al., 1997). Genotyping based on multilocus methylation-sensitive endonuclease suggest limited role for differential DNA methylation patterns in trypanosome HSR traits.

A most valuable approach for establishing linkage between DNA markers and specific trypanosome traits would be to determine the functions of known genes in relation to the traits of interest. Such traits should include parasite resistance/sensitivity to normal human serum, drug resistance, virulence, infectivity for the tsetse midgut and ability to produce salivary gland infection. The assessment of DNA markers for specific genetic traits is technologically feasible. Unfortunately, our knowledge of potential candidate genes for the traits is still very limited, with the serum-resistance-associated (SRA) gene being the best characterised (de Greef et al., 1989). It should be emphasised that not all *T. b. rhodesiense* appear to express the SRA mRNA (Rifkin et al., 1994). Furthermore, the SRA gene is present only in the *T. b. rhodesiense* and not in the *T. b. gambiense*, which is also resistant to normal human serum. This implies that serum resistance in human-infective trypanosomes may be regulated by more than one gene and/or genome-wide accumulations of polymorphisms.

The analysis of the molecular basis of human serum resistance in trypanosomes provides an opportunity to define genotype-phenotype variations. The most informative way to define gene effects associated with specific traits is by genetic mapping, which requires the construction of a detailed genetic map. In trypanosomes, the process of genetic recombination is unclear, and gametes have not been demonstrated. As a result, establishing populations with segregating genetic loci is impossible. For this reason, DNA markers are evolving as useful tools for genetic mapping. Genetic markers are Mendelian characters, which if sufficiently polymorphic, give a reasonable chance that a randomly selected isolate will be heterozygous. For linkage analysis we need informative meiosis, which has not been observed in trypanosomes. However, when a polymorphic marker is used on a large population, linkage analysis may give statistically significant data for a specific trait. Therefore, the use of random markers (e.g. MEGA, AFLP or RFLP) offers ready tools for diversity assessment in order to rapidly discover new candidate genes. However, it should be pointed out that the association between a marker and the phenotypic expression tagged by that marker will depend on the age of the mutation causing the effect, the frequency of recombination and presence of recombinational hotspots, as well as the population structure. This is especially relevant for *T. brucei* genome, which is highly plastic. A dense map of random markers developed with the help of a fine-scale tool, especially the MEGA approach described in Chapter 4 of this thesis, would be valuable to rapidly pick up association with a focal trait.

The plasticity of trypanosome genomes could promote divergence of the population and the development of subclones, which presumably enhances adaptation to host niches. We investigated the genotypic and phenotypic characteristics of *T. b. rhodesiense* clones and strains using the MEGA to exploit an expanded assessment of more independent methylation-sensitive genomic sites. The aim was a detailed analysis of the initial methylation-dependent regulative factors of the parasites for better

understanding of the possible correlation with phenotypic characteristics and development of host niches. The contribution of DNA methylation was shown to be limited in regulating human serum response traits (Chapter 6).

The genus *Trypanozoon* is currently separated into three species, namely *T. brucei*, *T. evansi* and *T. equiperdum*, based on host niches, development cycles and transmission characteristics. However, the validity of this classification system, specifically the separation of *T. evansi* and *T. equiperdum*, is questionable. *Trypanosoma evansi* and *T. equiperdum* have various overlapping properties, and definitive proof that only *T. equiperdum* causes dourine in horses has not been achieved. We applied multiplex molecular markers in a phenetic examination of relationships between *T. evansi* and *T. equiperdum*. This correlation coefficient measured the agreement between the similarity values implied by the dendogram and those of the original similarity matrix. Finally, we used molecular markers to clarify whether their genetic differences justify their distinction into separate species. These genotype analyses allowed us to formulate two hypothesis**:** first, that host immune responses modulate the clinical outcomes of horse infections with *T. evansi*, *T. brucei* and *T. equiperdum*, in which case *T. equiperdum* is not distinct from *T. evansi* and thus does not exist as a distinct species; second, that if *T. equiperdum* truly exists, the genuineness of eight of the ten identified strains is questionable. Our conclusions favour the first hypothesis (Chapter 7).

8.4. Perspectives on potential and limits of neutral markers, and future directions

Genetic diversity within *Trypanosoma species* can be assessed using variation in neutral markers (including AFLP and MEGA), gene-targeted markers (e.g. Single Nucleotide Polymorphisms, SNP's) or gene expression (Expressed Sequence Tags, EST's). The initial investment in marker development is generally higher for the gene-targeted approach than for the neutral markers. However, it should be emphasised that some

neutral markers, such as microsatellites, also need to be developed specifically for the species under study. This is circumvented in the universality of the AFLP and MEGA approaches. EST profiling requires a cDNA library on a microarray against which fragments from different genotypes can be hybridised. For trypanosomes, this is currently being developed *de novo* for the *T. brucei* (El-Sayed et al., 2000). Once developed, neutral markers as well as functional markers can often be converted to a (semi-)high throughput system for characterisation of the trypanosomes.

The genome coverage of a specific DNA marker system indicates whether a large part of the relevant variation is covered. For random markers, this is debatable. With respect to gene-targeted markers, SNP markers rely on the availability of sequence data of candidate genes. As long as our knowledge of candidate genes is limited, genome coverage will remain low. For EST's, the coverage can be very high but this depends on the specific developmental stages of the parasite used in synthesising the cDNA library. Genotypes that differ in phenotype under the tested conditions will be picked up, regardless of how many genes affect the phenotype or where the genes are located in the genome. However, it could be that different species at the same developmental stage have different genotypes.

Neutral DNA markers give the most valuable information on historical and demographic processes such as the evolutionary history and population structure (Sunnucks, 2000). This will be especially valuable if one could statistically test whether the divergence among populations is different from profiles of genotypes that are based on target genes. For distribution of allelic differences, it could be valuable to have information about variation in specific genes, for example, using microsatellites (Biteau et al., 2001; MacLeod et al., 1999, 2000, 2001). However, for the assessment of genetic basis of trypanosome complex phenotypic traits, it seems risky to replace random markers that cover a large part of the genome with markers

targeted in a small subset of genes. For a trait that is controlled by genome-wide accumulations of single nucleotide or sequence polymorphisms rather than a loss of particular gene fragment, assessment of diversity by random markers could be preferable.

References

Ajmone-Marsan, P., A. Valentini, M. Cassandro, G. Vecchiotti-Antaldi, G. Bertoni, and M. Kuiper. 1997. AFLP markers for DNA fingerprinting in cattle. Anim. Genet. **28:**418-426.

Anene, B. M., D. N. Onah, and Y. Nawa. 2001. Drug resistance in pathogenic African trypanosomes**:** what hope for the future? Vet. Parasitol. **96:**83-100.

Biteau, N., F. Bringaud, W. Gibson, P. Truc, and T. Baltz. 2000. Characterisation of *Trypanozoon* isolates using a repeated coding sequence and microsatellite markers. Mol. Biochem. Parasitol. **105:**185-201.

Busslinger, M., J. Hurst, and R. A. Flavell. 1983. DNA methylation and the regulation of globin gene expression. Cell **34:**197-206.

De Greef, C., G. Imbrechts, G. Matthyssens, N. Van Meirvenne, and R. A. Hamers. 1989. A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. Mol. Biochem. Parasitol. **36:**169-176.

El-Sayed, N. M., P. Hegde, J. Quackenbush, S. E. Melville, and J. E. Donelson. 2000. The African trypanosome genome. Int. J. Parasitol. **30:**329-345.

Fraga, M. F., and M. Esteller. 2002. DNA methylation**:** a profile of methods and applications. Biotechniques **33:**632, 634, 636-649.

Gibson, W. C., and W. H. Mizen. 1997. Heritability of the trait for human infectivity in genetic crosses of *T. brucei* spp. Trans. R. Soc. Trop. Med. Hyg. **91:**236-237.

Hide, G., A. Tilley, S. C. Welburn, I. Maudlin and A. Tait. 2000. *Trypanosoma brucei*: identification of trypanosomes with genotypic

similarity to human infective isolates in tsetse isolated from a region free of human sleeping sickness. Exp. Parasitol. **96:**67-74.

Janssen, P., R. Coopman, G. Huys, J. Swings, M. Bleeker, P. Vos. M. Zabeau, and K. Kersters. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. Microbiology **142:**1881-1893.

MacLeod, A., A. Tait, and C. M. Turner. 2001. The population genetics of *Trypanosoma brucei* and the origin of human infectivity. Philos. Trans. R. Soc. Lond. B. **356:**1035-1044.

MacLeod, A., A. Tweedie, S. C. Welburn, I. Maudlin, C. M. R. Turner, and A. Tait. 2000. Minisatellite marker analysis of *Trypanosoma brucei*: Reconciliation of clonal, panmictic and epidemic population genetic structures. Proc. Natl. Acad. Sci. USA **97:**13442-13447.

MacLeod, A., C. M. Turner, and A. Tait. 1999. A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. Mol. Biochem. Parasitol. **102:**237-248.

Paindovoine, P., F. Zampetti-Bosseler, H. Couquelet, E. Pays, and M. Steinert. 1989. Different allele frequencies in *Trypanosoma brucei brucei* and *T. b. gambiense* populations. Mol. Biochem. Parasitol. **32:**61-72.

Razin, A., and H. Cedar. 1991. DNA methylation and gene expression. Microbiol. Rev. **55:**451-458.

Razin, A., and H. Cedar. 1994. DNA methylation and genomic imprinting. Cell **77:**473-476.

Reyna-Lopez, G. E., J. Simpson, and J. Ruiz-Herrera. 1997. Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Mol. Gen. Genet. **253:**703-710.

Rifkin, M. R., C. De Greef, A. Jiwa, F. R. Landsberger, and S. Z. Shapiro. 1994. Human serum-sensitive *Trypanosoma brucei rhodesiense*: a comparison with serologically identical human serum-resistant clones. Mol. Biochem. Parasitol. **66:**211-220.

Sunnucks P. 2000. Efficient genetic markers for population biology. Tends Ecol. Evol. **15:**199-203.

Tait, A, D. Masiga, J. Ouma, A. MacLeod, J. Asse, S. Melville, G. Lindegard, A. McIntosh, and M. Turner. 2002. Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits. Philos. Trans. R. Soc. Lond. B. **357:**89-99.

Vos, P., M. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1996. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:**4407-4414.

Summary

The advances in genomics technologies and genome analysis methods that offer new leads for accelerating discovery of putative targets for developing overall control tools are reviewed in Chapter 1. In Chapter 2, a PCR typing method based on restriction fragment length polymorphism analysis of the internal transcribed sequence (ITS) rDNA region was used to reveal distinct fingerprinting patterns that characterise human- and animal-derived *Trypanosoma brucei gambiense* and *T. b. brucei* isolates. Although these results also highlighted doubts about the uniformity of *T. brucei* subspecies, the limitation of such a typing technique that is based on a single genetic locus is obvious. As a result, the studies were extended to include all *T. brucei* subspecies in a more global amplified fragment length polymorphism (AFLP) genotyping (Chapter 3). This approach permitted an unbiased estimate of the total genome variance and revealed closer genetic relatedness between, and higher variability within, *T. b. brucei and T. b. rhodesiense* subspecies, compared to *T. b. gambiense* strains. However, it was clear from these studies that a finer-scale genotyping tool with enhanced resolution power was required. Chapter 4 describes such an advanced tool, a multiplex-endonuclease genotyping analysis (MEGA) approach that simultaneously accesses multiple independent restriction enzyme-based polymorphisms within the genome. It offered a robust and detailed genotyping tool and was, therefore, used to study the population genetic structure of *T. brucei* isolates, for epidemiological and cladistic analysis (Chapter 5).

The MEGA approach envisages the application of genotyping to identify genetic profiles that are associated with specific (parasite) traits. Therefore, in Chapter 6, genotypes were constructed and correlated with human serum response traits of *T. brucei rhodesiense* clones and strains, to

further provide a general method for measuring differential phenotypes and an objective assessment of such differences. Also, such fine-scale approach can be used to rapidly enrich for identifiable polymorphisms in a set of known DNA sequences known to be associated with a phenotypic trait of interest. Using sets of four endonucleases selected on the basis of the concept defined by the MEGA approach, the role of differential DNA methylation patterns in the human serum response properties of trypanosomes was evaluated (Chapter 6), which proved to be insignificant. Furthermore, we clarified the genetic relationships between *T. equiperdum* and other *Trypanozoon* species (Chapter 7). In Chapter 8, a general discussion of the data is presented.

In summary, three main applications of molecular marker systems in trypanosomes were described in this thesis. These involve genomic studies for (1) generating sensitive tools for molecular typing of strains, (2) elucidating taxonomy of *Trypanozoon*, and (3) the analysis of the relationships between genetic variations and their consequent functional effects that may enhance our understanding of important traits. These applications have permitted fine-scale genotypic characterisation of the parasites, and offered a template for phenotypic correlations of the genotype data.

Samenvatting

Hoofdstuk 1 bestaat uit een beschouwing van de vooruitgangen in genetische technologieën en genoom analyse methoden en de daar uit voortkomende indicaties om het ontdekken van mogelijk nieuwe aangrijpingspunten te bespoedigen. Allereerst werd een PCR (polymerase chain reaction) typeringstechniek gebruikt (hoofdstuk 2). Deze techniek is gebaseerd op Restrictie Fragment Lengte Polymorfisme (RFLP) analyse van een intern getranscribeerde rDNA sequentie regio (ITS). Deze techniek onthulde verschillende 'fingerprint' patronen die specifiek voor humaan -of diergerelateerde *Trypanosoma brucei gambiense* en *T. b. brucei* isolaten zijn. Ondanks dat deze techniek twijfels aan de uniformiteit van *T. brucei* subspecies versterkte, is het limiterend oplossend vermogen van een techniek gebaseerd op één enkel genetisch locus overduidelijk. Om deze reden werd het onderzoek uitgebreid met een meer algemene genotypering techniek, amplified fragment length polymorfism (AFLP), en werden alle *T. brucei* subspecies onderzocht (hoofdstuk 3). Deze aanpak resulteerde in een onbevooroordeelde schatting van de variatie van het gehele genoom en onthulde een grotere genetische overeenkomsten binnen *T. b. brucei* en *T. b. rhodesiense* subspecies en hogere variabiliteit in vergelijking tot *T. b. gambiense* subspecies. Deze aanpak maakte echter ook duidelijk dat een gevoeligere genotypering techniek met een hoger oplossend vermogen nodig was.

Hoofdstuk 4 beschrijft zo een geavanceerde techniek, de multiplex-endonuclease genotypering analyse (MEGA). Met behulp van deze aanpak wordt tegelijkertijd gebruik gemaakt van meerdere onafhankelijke, op restrictie-enzym gebaseerde polymorfismen binnen het genoom. Het resultaat is een robuuste en gedetailleerde genotypering techniek die daarom gebruikt kon worden om de genetische structuur van de *T. brucei*

populatie te bepalen alsmede voor epidemiologische en 'cladistic' analyse (hoofdstuk 5).

De MEGA aanpak laat zien hoe genotypering gebruikt kan worden om genetische profielen te bepalen die zijn geassocieerd met specifieke (parasitaire) kenmerken. Derhalve beschrijft hoofdstuk 6 het construeren van genotypen van *T. brucei rhodesiense* klonen en stammen, het correleren van genotypen aan humaan serum reactie kenmerken. Zodoende is een algemene methode ontwikkeld om onderscheidende fenotypes te bepalen en deze verschillen objectief te beoordelen. Verder kan een gevoeligere techniek als deze gebruikt worden om snel, uit een set van bekende DNA sequenties, polymorfismen te identificeren die gekoppeld zijn aan interessante kenmerken. Door gebruik te maken van sets van 4 endonucleases geselecteerd aan de hand van het concept gedefinieerd door de MEGA aanpak, is de rol van differentiële DNA methylering patronen ten opzichte van humaan serum reactie kenmerken geëvalueerd, welke niet significant bleek. Bovendien konden de genetische verwantschappen tussen *T. equiperdum* en andere *Trypanozoon* species verklaard worden (hoofdstuk 7). In hoofdstuk 8 wordt een algemene discussie van de data gepresenteerd.

Kort samengevat wordt in dit proefschrift een moleculaire marker systeem voor trypanosomen beschreven, met drie voorname applicaties. De applicaties bevatten genomische toepassingen om (1) een gevoelige methode opzetten voor moleculaire typering van stammen, (2) het ophelderen van de *Trypanozoon* taxonomie, (3) het analyseren van de verwantschappen tussen genetische variaties en de daaruit volgende functionele effecten die mogelijk het inzicht in belangrijke kenmerken verbetert. Deze toepassingen hebben gevoelige genotypische karakterisering van de parasieten mogelijk gemaakt en een model voor fenotypische correlaties van de genotypering data opgeleverd.

ACKNOWLEDGEMENTS

A very pleasant aspect of this thesis is remembering the many colleagues and friends who assisted me in various ways. To Stephanie Vastenhouw, Frank Harders, Ruth de Vries, Astrid de Greeff, Alex Bossers and Hilde Smith, and to the many others working in Wing 19 (of ID-Lelystad), I say a big thank you for your various supports and assistance. And to Alan Rigter, thanks 'dude' for translating the Summary to dutch. I am especially indebted to Herma Buijs and Jan Priem (my 'paranymphs') who, among other things, were instrumental in organising the reception.

I also owe thanks to many other colleagues who have read drafts and contributed thoughts and material in the course of this work. Particular thanks to (our) Cluster Head (Section Animal Genomics), Professor Mari Smits (ID-Lelystad/Wageningen University), who also actively promoted and greatly facilitated the project.

I was also fortunate to have had a very wonderful supervisory team. Their comments and suggestions helped to shape the work into its final form. I thank Dr Marleen Roos (Roos ProjectConsult, The Netherlands) for her guidance in the early phase of the project and Dr Marinus te Pas (ID-Lelystad, The Netherlands) for making the 'change of baton' such a smooth event. Marinus, thank you for you were great in keeping me on the rails. I would also like to thank Dr Phelix Majiwa (ILRI Nairobi, Kenya) for providing insight and perspectives throughout the project. Many thanks also to my Promotors, Professors Eric Claassen (Erasmus University Medical Center Rotterdam, The Netherlands), Jos van Putten (Utrecht University, The Netherlands) and Philippe Büscher (Institute of Tropical Medicine, Antwerp,

Belgium) for facilitating the project in various ways. It feels so sweet to remember your various inputs. My special appreciation to Professor Claassen for fostering the outset of the project, and for which I say a very big thank you.

I am especially grateful for the support, encouragement and patience of my family and I take this opportunity to express my gratitude to my beloved wife Ebby, our beautiful daughter Chidera, my mum Anthonia, and to my brothers and lovely sister (Chizoba).

LIST OF PEER-REVIEWED PUBLICATIONS

1. **Agbo E.C.,** Duim B., Buscher P., Claassen E. and te Pas M.F.W. 2002. Multiplex-endonuclease genotyping approach (MEGA): a tool for the fine-scale detection of unlinked polymorphic DNA markers[¶]. *Chromosoma*, In press.

   [¶] **Patent pending**; Principal Inventor - Eddy C. Agbo.

2. Claes F., **Agbo E.C.,** Radwanska M., te Pas M.F.W., Baltz T., De Waal D.T., Goddeeris B.M., Claassen E. and Buscher P. 2002. How does *Trypanosoma equiperdum* fit into the *Trypanozoon* group? A cluster analysis by Random Amplified Polymorphic DNA (RAPD) and multiplex-endonuclease genotyping approach. *Parasitology,* In press.

3. **Agbo E.C.,** Majiwa P.A.O., Claassen E. and te Pas M.F.W. 2002. Molecular variation of *Trypanosoma brucei ssp* as revealed by AFLP fingerprinting. *Parasitology* **124:**349-358.

4. **Agbo E.C.,** Majiwa P.A.O., Claassen E. and Roos M.H. 2001. Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *T. b. gambiense* as revealed by genotypic characterisation. *Experimental Parasitology* **99:**123-131.

5. **Agbo E.C.,** Majiwa P.A.O., Buscher P., Claassen E. and te Pas M.F.W. 2002. Trypanosome genomics, the challenge of identifying vaccine targets and the prospects for control of trypanosomosis: a Review. Submitted for publication.

6. **Agbo E.C.,** Clausen P-H., Buscher P., Majiwa P.A.O., Claassen E. and te Pas M.F.W. 2002. Population genetic structure and cladistic analysis of *Trypanosoma brucei* isolates. Submitted for publication.

7. **Agbo E.C.,** Magnus E., Buscher P., Majiwa P.A.O., Claassen E. and te Pas M.F.W. 2002. Correlating multilocus molecular markers, differential DNA methylation analysis and human serum

responsiveness for defining *Trypanosoma brucei rhodesiense* genotypes. Submitted for publication.

8. Akinrinmade J. F. and **Agbo E.C.** 1992. A 3-year survey of Caesarean operation in sheep and goats in South-western Nigeria. *Tropical Veterinarian* 10:37-42.

POSTERS

1. **Agbo E.E.C.,** te Pas M.F.W. and Claassen E. 2001. Marker-inferred analysis as a valuable tool for subspecific characterisation of *Trypanosoma brucei*. *Proceedings of the 26th International Scientific Council for Trypanosomosis Research and Control (ISCTRC), Ouagadougou, Burkina Faso,* October 0.,k1-05, 2001, p. P240.
2. **Agbo E.E.C.,** te Pas M.F.W. and Claassen E. 2001. Isolation of binary specific PCR-based markers and their value for diagnostic applications in human and animal sleeping sickness. *Proceedings of the Netherlands Society for Medical Microbiology*, Arnhem, March 26 - 28, p. S42.

ABOUT THE AUTHOR

Edwin Chukwura Agbo was born at Mbu-Amon in Enugu State of Nigeria on 21 December, 1966, to Ephraim† and Anthonia Agbo. He attended primary school at Central School Mbu-Amon, won state scholarship to secondary school at Government College, Maiduguri, Nigeria. Thereafter, he trained as a veterinarian at the University of Ibadan, Nigeria, qualifying with the Doctor of Veterinary Medicine (DVM) degree in 1991. He received his Master of Science (MSc) degree from Wageningen University, The Netherlands in 1998. The work described in this thesis was a strategic expertise research funded by the Institute of Animal Science and Health (ID-Lelystad), Lelystad, The Netherlands, and implemented under a collaborative program with Utrecht University, The Netherlands. It is being rewarded with the award of a doctorate degree (PhD) by Utrecht University this day: March 27, 2003.