MOLECULAR SYSTEMATICS OF MALARIA VECTORS :

STUDIES BASED ON RAPD PCR AND RELATED TECHNIQUES

THESIS SUBMITTED IN ACCORDANCE WITH THE

REQUIREMENTS OF THE UNIVERSITY OF LIVERPOOL

FOR THE DEGREE OF DOCTOR IN PHILOSOPHY

BY

NAVID DINPARAST DJADID

OCTOBER 1997

VOLUME 1

بسم الله الرحمن الرحيم

ABSTRACT

This thesis describes a series of investigations of inter- and intra-specific DNA sequence variation in malaria vectors from Africa, the Middle East and Indian sub-continent. A variety of PCR-based techniques have been employed, particularly Random Amplified Polymorphic DNA (RAPD) PCR. With careful attention to the conditions of amplification, the technique was reliable and gave reproducible results.

With a total of four RAPD primers it was possible to differentiate and identify to species all six members of the *Anopheles gambiae* complex. Other primers revealed intra-specific variation in *An.gambiae* and *An.arabiensis* and may be useful in defining the population structure of these species. Phylogenetic trees based on RAPD data confirmed that *An.gambiae* and *An.arabiensis* are sister taxa. A cloned RAPD fragment from *An.bwambae* showed a high sequence homology with a cloned RAPD sequence from *An.culicifacies* and may include part of a ubiquitin gene. By testing parents and their progeny it was confirmed that the major RAPD bands are inherited in a dominant fashion.

Application of RAPD PCR to the two major species of the *An.culicifacies* complex showed that species A and B could readily be separated in samples from Iran, Pakistan, India and Sri Lanka. Some genetic differences were apparent between sub-populations of species B within Sri Lanka, but Indian species B appears to be genetically distinct from Sri Lankan species B. These differences reflect the vastly differing roles in malaria transmission of species B in the two countries. An Iranian population of mosquitoes identified morphologically to *An.culicifacies* s.l. proved to be comprised of two forms differing in their RAPD patterns. On sequencing the rDNA ITS2 region of these two forms, one was virtually identical in sequence to other species A populations while the other was highly divergent, (50% nucleotide divergence) suggesting that it may be a new species related to but outside the *An.culicifacies* complex.

An.fluviatilis, which is known to comprise three sibling species in India, is an important secondary vector of malaria in parts of India and Iran. Field collected specimens from Iran were examined by RAPD PCR, SSR PCR and sequencing of the rDNA ITS2 region. RAPD results suggested three populations but only two forms of ITS2 were seen, implying the presence of two distinct species within this taxon in Iran.

In India, *An.subpictus* comprises four sibling species of which two are known to occur in Sri Lanka. Samples from Sri Lanka examined by RAPD PCR fell into four population groups representing the four sample sites, one coastal and three inland. However, ITS2 sequence analysis was compatible with the existence of two groups corresponding to species A and B.

In contrast to many malaria vectors, *An.stephensi* is thought to be a single species, despite its enormous geographical range. RAPD readily distinguished seven laboratory colonies derived from collections in Iraq, Iran and India. Field samples from Iran and Tadjikistan were similar by RAPD but differed from one sample from Pakistan. Rural and urban collections from Iran differed in their RAPD patterns but it is unclear whether these two samples represent *mysorensis* and the type form. Tests with parents and their F1 progeny showed that RAPD could be used to determine paternity in this species, a feature that could be useful in evaluating experimental laboratory crosses.

Samples of the *An.maculipennis* complex from Iran and Russia were examined using RAPD. *An.maculipennis* s.s. from Iran could readily be distinguished from *An.sacharovi*. Furthermore, *An.beklemishevi* from Siberia could be distinguished from both northern and southern karyotypes of *An.messeae* from the same region. When the rDNA ITS2 region of the *An.maculipennis* complex samples from Iran was sequenced, three different types of sequence were obtained which, it is suspected, correspond to *An.atroparvus*, *An.messeae* and *An.sacharovi*.

This thesis demonstrates the power of RAPD PCR in elucidating the systematics of groups of insect vectors for which no prior DNA sequence data exist. It can serve equally to indicate genetic differences between populations or between individuals within populations. Once thus identified, it becomes possible to examine individuals or populations of particular interest using a battery of molecular techniques, including sequencing of various gene segments, particularly in rDNA or mtDNA, or carrying out RFLP or SSCP analysis. By such means the timeframe from initial recognition of genetically variant individuals or populations to resolving their taxonomic status can be radically shortened.

ACKNOWLEDGEMENTS

I would like to express my special thanks to Professor H. Townson for the opportunity to work alongside him, for sharing his wide knowledge and experience, for his support and understanding, and for his encouragement and invaluable criticism in all stages of this study.

I gratefully acknowledge the Ministry of Health and Medical Education, I.R.Iran for providing me with the scholarship towards the PhD study and all the staff of the Zahedan University of Medical Sciences and Public Health Department of Sistan & Baluchistan Province for their moral support and mosquito collection.

My thanks go to all those colleagues and institutes who provided me with the field collected specimens from Tanzania, Kenya, Uganda, Mozambique, South Africa, The Gambia, Ghana, Nigeria, Zimbabwe, Sudan, Burkina Faso, Iran, Iraq, Dubai, Pakistan, India, Sri Lanka and Russia.

For laboratory strains of *An. gambiae* species complex and *An.stephensi*, I am grateful to Mr Ken Sherlock , Mr Andy Steven, Dr. Rajni Vij, Professor M.W.Service and the London School of Tropical Medicine and Hygiene.

Thanks go also to the Director, Professor D.H. Molyneux and staff of the Liverpool School of Tropical Medicine especially the Department of Medical Entomology, to Professor W. W. Macdonald and Professor M.W. Service for their scientific advice, to Mrs Tara Adeniran for her assistance, and to Miss Lisa Bluett.

I would like to thank Dr H. Noyes for his always helpful discussions; Professor Bianco and his scientific group, especially Dr. Yung Wu, and Dr. M. Hughes for automated DNA sequencing.

I remember the helpful assistance of the staff in Medical Illustration.

Finally, thanks to all colleagues in different parts of the university and School for their moral support, assistance and advice.

Thanks to God

and

To all my family, especially Sedigheh, Azadeh and Mohammad for their understanding, patience and support during this time

LIST OF ABBREVIATIONS

APS	ammonium persulphate
ATP	adenosine triphosphate
BEPS	BEPS hybridization buffer
bp	base pairs
BSA	bovine serum albumin
°C	degrees centigrade
cDNA	complementary DNA
Ci	curies
cpm	counts per minute
ddH_2O	double distilled water
DNA	deoxyribonucleic acid
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanine triphosphate
dTTP	deoxythymidine triphosphate
E. coli	Escherichia coli
EDTA	ethylenediamine tetra-acetic-acid
ELISA	enzyme linked immunosorbent assay
EtBr	ethidium bromide
g	gram
mg	milligram
IFAT	immunofluorescent antibody test
Ig	immunoglobulin
kbp	kilobase pairs
kDa	kilodaltons
kDNA	kinetoplast DNA
L	litre
ml	millilitre
M	molar concentration
mA	miliamperes
mRNA	messenger ribonucleic acid
mV	millivolts
NCP	nitrocellulose paper
ng	nanogram
OD	optical density
OTUs	operational taxonomic units
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
pg	Picogram
Poly(A)$^+$	polyadenylated ribonucleic acid

RAPD	random amplified polymorphic DNA
RNA	ribonucleic acid
RNase	ribonuclease
RPM	revolutions per minute
rRNA	ribosomal RNA
RT	room temperature
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
SSR	simple sequence repeat
TBS	tris buffered saline
TEMED	N,N,N,N -tetramethylethylenediamine
TX-100/114	triton X-100/114
u	unit
μg	microgram
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indoyl-B-D-galactopyranoside

GLOSSARY

Biological species concept: A concept of species category stressing reproductive isolation, and the possession of a genetic programme effecting such isolation.

Classification: The delimitation, ordering, and ranking of taxa.

Identification: The determination of the taxonomic identity of an individual.

Nomenclature: A system of names.

Phylogeny: The study of the history of the lines of the evolution in a group of organisms; the origin and evolution of higher taxa.

Sibling species: Pairs or groups of closely related species which are reproductively isolated but morphologically identical or nearly so.

Species: Group of organisms capable of exchanging genetic material with one another and incapable, by reason of their genetic constitution, of exchanging such material with any other group of organisms. The limits of species are indicated by the comparative study of morphological and other characters.

Species: Groups of actually (or potentially) interbreeding natural populations which are reproductively isolated from other such groups.

Species complex: Group of closely related organisms, the exact specific status of which is uncertain although they resemble some well-recognized type species.

Systematics: The science dealing with the diversity of organisms.

Taxonomy: The theory and practice of classifying organisms.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

Exactly one hundred years ago, Ross (1897) found some peculiar pigmented cells in two mosquitoes fed on a malaria patient, and since then mosquitoes have been a focus of scientific study. Following the incrimination of anophelines as vectors of malaria (Ross 1895, 1898, 1900), it was soon realised that not all species in the genus *Anopheles* are vectors and hence control could be focused on those species that were important in malaria transmission (Watson, 1915; Swellengrebel and De Buck 1938). However, the recognition in southern Europe of " anophelism without malaria" subsequently led to the discovery of the *Anopheles maculipennis* complex (Hackett, 1937; Missirolli, 1939; Bates, 1939, 1940).

It has since been apparent that many of the malaria vectors, in each of the continents where malaria occurs, had members of sibling species complexes. That is belong to a group of morphologically very similar organisms which nevertheless constitute good biological species and more importantly, which show important differences in their ecology, biology, and behaviour , including their role in malaria transmission.

As a part of comparative studies on applied biology and molecular systematics, and in order to evaluate the validity of the biological versus the morphological species

concept, this project has aimed to develop PCR-based methods, mainly Random Amplified Polymorphic DNA (RAPD), for defining the principles of molecular genetic variations in populations of malaria vectors species complexes by access to random primed parts of mosquito genome and its novel applied features.

This aim is well established and is in concordance with the views of systematicians as described by Lane & Crosskey (1993) that, one of the basic stages in systematics is recognition of biologically operational units (species, subspecies, etc.) which are the basic units of any classification. However, the species problem is the oldest problem in biology (Dobzhansky, 1951, 1976) and diseases transmitted by insects, and particularly by mosquitoes, remain some of the most important health problems in the world, as illustrated in last decades by the failure to eliminate malaria from much of the developing world, especially tropical Africa.

Malaria is the most important insect-transmitted parasitic disease and the control of its vector is based on a number of chemical, biological, environmental and genetic techniques. It has mainly relied on the elimination of breeding sites and the application of chemical agents (insecticide) against larvae and adults, bednets and repellents and to a lesser extent, on biological control using *Bacillus thurengiensis* and other related organisms. The only approved and available treatment in humans is still based on chemotherapy and prophylaxis by quinine, chloroquine, mefloquine and other drugs.

But, all these strategies are faced with problems such as 1) insecticide, antimalarial drugs and public resistance which have been developed in vectors, parasites and local people in endemic areas due to long term application of insecticides and different generation of antimalarial drugs, 2) differences in the biology of the major and secondary malaria vectors, some of those related to human movement, cultivation and urbanization which led to resurgence of malaria (e.g. in Madagascar, several countries in Southeast Asia, and Brazil), 3) bio-diversity in four morphologically distinct, cosmopolitan species of *Plasmodium*, especially in *P. falciparum* and its subsequent host-parasite interaction, 4) the high cost of developing and registering new insecticides and awareness of pollution from insecticide residues, 5) changes in policies and administration bodies both in World Health Organization (WHO) and affected countries which disrupted or reoriented the current advised control measures.

Regarding these problems, there is no doubt that we desperately need alternative measures and new tools or at least should improve the efficiency of those currently applicable techniques in control of the disease. At this stage of the development, it seems that the most important contribution of molecular research on mosquitoes will probably be in the areas of molecular systematics and evolutionary biology. This is because of the applied importance of data on the extent of genetic variations and the forces that tend to produce local genetic differentiation or genetic homogeneity and the balance between these phenomena in a defined species population.

Answering these kind of basic questions which are pre requisite in developing control measures, needs new tools and more effective means in molecular systematics and much greater understanding of the basic molecular biology of mosquito vectors.

However, the results of chromosomal analyses suggest that *Anopheles* is primitive within the family Culicidae. Evidence from *Anopheles* supports the extensive divergence from other mosquitoes, including significantly smaller chromosomes and lower nuclear DNA content (Rao and Rai, 1990), and their unique possession of dimorphic sex chromosomes and long-period interspersion of repetitive sequences in the genome (Black and Rai, 1988). It has a genome of approximately 260 megabases (Besansky & Powel, 1992), which is about 50% larger than the genome of *Drosophila melanogaster* (Ashburner, 1989), and like *Drosophila*, also has a long period interspersion of repeat sequences (Black & Rai, 1988)

The total number of known species in the genus *Anopheles* is currently around 458 (Harbach, 1994), of these, perhaps 40 species are of practical importance in the transmission of malaria. For example, in *An. gambiae* as a well studied species complex, *Anopheles gambiae s.s.* and *An. arabiensis* are the two most effective malaria vectors in most of sub-Saharan Africa and also an important vector of *Wuchereria bancrofti* and several African arboviruses, while *An. quadriannulatus*, another member in this species complex, although found sympatric with *An. arabiensis* seems to have a minor role in the transmission of malaria.

Nowadays in applied molecular systematics it has been realised that DNA-based procedures have many advantages over two other routine methods, species-specific allozymes and polytene chromosome inversions, which has been widely used for identifying the species of individual specimens from species complexes. For example life stage, sex, type of preservation of specimens and amount of DNA are not considered any more as the limiting factors in different stages of these procedures. Although these methods vary in their susceptibility to detect new molecular species differing from the starting template but PCR is exceptionally resistant to in vitro evolution, whereas methods such as QB replicase and 3SR are much less robust (Bull & Pease, 1995).

DNA-based identification techniques mainly are based on diagnostic PCR fragments produced either by the use of random PCR primers (Williams et al. 1990; Welsh & McClelland 1990), or by amplifying DNA with primers based on known species-specific sequences or DNA probe using species-specific repeat sequences (Collins and Paskewitz, 1996).

Another feature of this project is the sibling species concept which has been defined as pairs or groups of closely related species that are morphologically indistinguishable (isomorphic) but reproductively isolated, and which frequently live in the same area (i.e. are sympatric)(Lane & Crosskey 1993), and as Coluzzi (1970) mentioned, the biological difference among sympatric sibling species are no less important than those characterized morphologically differentiated species (Coluzzi,

1970). When active mosquito control became possible with the advent of insecticides, it was essential to detect vector and non-vector species.

However, in field conditions where control measures are under practice, these biological differences and geographical variation in species complexes together with their enormous reproductive capacity and genetic flexibility led to rapid development of insecticide resistance and adapting to human environments (Besansky & Collins, 1992;; Coluzzi, 1984, 1992; Coluzzi *et al.*, 1985, 1977, 1979).

As phenotypic variation is reflective of genotypic and environmental variation, there are some explanations for these variations but the fact that the nuclear genomes of eukaryotes are subject to a continual turnover through unequal exchange, gene conversion and DNA transposition is a basic point to follow. These processes operate both within and between chromosomes (Dover, 1982, 1986). Furthermore, as Young (1979) described, some families of sequences in a wide range of plant and animal species act as 'fluid' because of the continually changing nature of their composition, abundance and position.

On the other hand, eukaryotic genomes contain substantial numbers of multiple-copy families of genes and non-coding sequences, and it has been apparent for some years that many families exhibit unexpected sequence homogeneity within and between individuals of a species. while a family that is shared between several species, often reveals substantial heterogeneity between the species (Dover, 1982).

Seppa & Pamilo (1994), considered the mutation, genetic drift and locally differing selection pressures as evolutionary factors leading to major genetic differences, and ultimately to speciation of isolated populations.

Apart from these factors, genetic differences may arise because of the founder effect. If local populations are founded by a small number of individuals, through bottlenecks, they may differ from their source populations because of chance (e.g. Nei & Chesser 1983; Nei 1987). On the other hand, Slatkin (1985, 1987) believed that dispersal of individuals prior to reproduction may lead to gene flow, which is a powerful factor opposing reproductive isolation. Generally a very small numbers of migrants is enough to prevent substantial differentiation among populations (1977, 1981). Also gene flow can be restricted because of sub-division of populations, where individuals move mainly within physically or otherwise separated sub-populations, or because of isolation by distance in a continuous population (Seppa & Pamilo, 1994).

Gene flow may either constrain evolution by preventing adaption to local conditions or promote evolution by spreading new genes and combinations of genes throughout a species range.

So "genetic drift" could be defined as the unpredictable change in gene frequency due to finite population size, and "gene flow" is the change due to movement of gametes, individuals, or groups of individuals from one place to another (Slatkin 1977).

How much gene flow is sufficient to prevent genetic variation and speciation?. The balance achieved between gene flow and genetic drift provides a background against which to consider the effects of different kinds of selection. Genetic drift, like gene flow, has the same average effect on all nuclear genes. In a group of completely isolated populations, genetic drift alone would tend to fix different alleles in different local populations. Any gene flow at all among populations will prevent complete fixation but gene flow must exceed a certain level to prevent substantial genetic differentiation due to genetic drift.

Gene flow can inhibit genetic evolution by preventing natural selection and genetic drift from establishing and maintaining local genetic differences. Gene flow can also promote genetic evolution although the evidence for this is less well established.

Aims and objectives of this thesis:

Because DNA-based polymorphisms identified using RAPDs tend to originate from repetitive sequences with features such as inverted repeat structures (although they certainly can be derived from single-copy sequences as well), DNA-based methods, mainly RAPD-PCR have been applied in order to facilitate the taxonomic description of species complexes and to provide more data on the structure of malaria vector genomes and their population genetic phenomena as basic requirements for studies of the role of the various species in malaria transmission.

The general aims are:

1- To address the inter- and intra-specific variation within the most important malaria vectors from Africa, Middle East and Indian sub-continent.

2- To compare the potentials and abilities of RAPD markers and their sequences with other related PCR-based techniques such as rDNA-PCR, SSR-PCR and DNA probe.

The specific objectives are:

1- To provide some criteria and determine the optimal conditions under which amplifications are both efficient and consistent.

2- To identify sympatric species within species complexes and to generate species-specific PCR markers.

3- To analyze intra-specific heterogeneity in defined populations and identify population-specific PCR markers.

4- To analyze quantitatively mixed biosamples, analogous to the analysis of field samples of different species or other OTUs.

5- To examine the inheritance of RAPD bands by mating studies and to assess the kinship relationships.

6- To characterize amplified loci of particular interest in order to find the nature of RAPD products.

7- To create specific probes or primers.

8- To investigate the genetic relatedness of mosquito populations by constructing phylogenetic trees based on similarities and distance indices generated from independent RAPD markers.

CHAPTER TWO

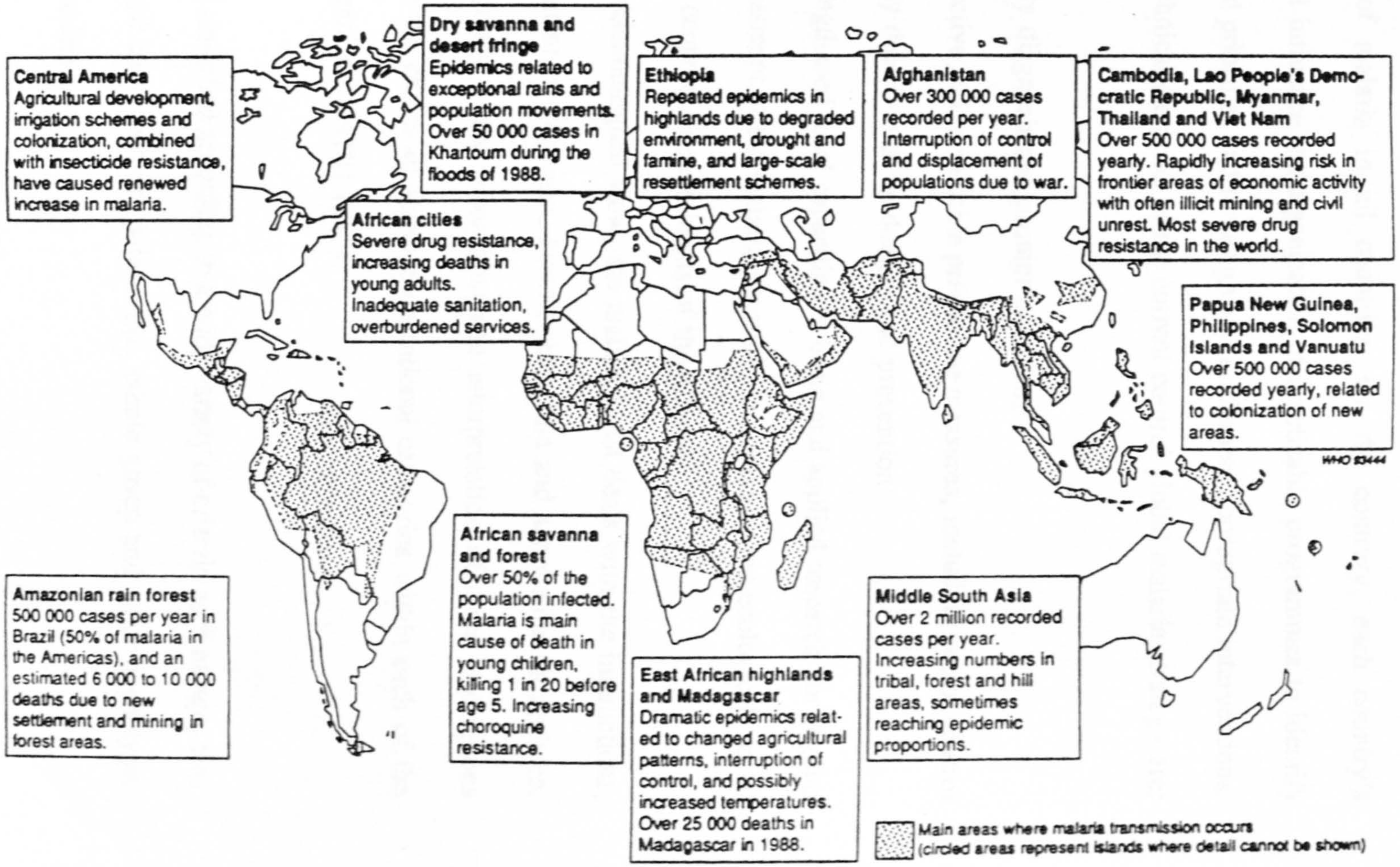
LITERATURE REVIEW

2.1 Global malaria and its control strategy

Malaria is by far the world's most important tropical parasitic disease, and kills more people than any other communicable disease except tuberculosis (WHO 1996b). Today Malaria is a public health problem in more than 90 countries, inhabited by a total of some 2400 million people, 40% of the world's population. The worldwide prevalence of the disease is estimated to be of the order of 300-500 million clinical cases each year, with 90% of all cases from sub-Saharan Africa, and two-thirds of the remainder concentrated in six countries: India, Brazil, Sri Lanka, Viet Nam, Colombia and The Solomon Islands, in decreasing order of prevalence (WHO 1996a, Fig. 2.1).

Malaria is estimated to be responsible for between 1.5 and 2.7 million deaths each year. The vast majority of these deaths occur in young children in Africa, especially in remote rural areas with poor access to health services. Other high-risk groups are women during pregnancy, and non-immune travellers, refugees, displaced persons and labourers entering endemic areas. Malaria epidemics related to political upheavals, economic difficulties and environmental problems also contribute in a most dramatic way to death and human suffering (WHO 1996a,b).

Fig. 2.1- Malaria distribution and problem areas in the world (CTD/MAL/96.1)



Since malaria varies throughout the world no single prescription can be made for the control of malaria in all countries. On the contrary, each country's circumstances will influence the organization of practicable programmes to identify local problems and priorities and to design and implement appropriate interventions. The four basic technical elements of the current control global malaria strategy are:

1 early diagnosis and prompt treatment

2 selective and sustainable preventive measures, including vector control

3 early detection of epidemics and prevention

4 strengthened local capacities in basic and applied research for regular assessment of the malaria situation, in particular the ecological, social and economic determinants of the disease.

From an epidemiological view, the malariologist deals with the interactions, in a given environment, among man, human plasmodia and anopheline mosquitoes. The degree of accuracy in the description and interpretation of these interactions depends mainly on the choice of suitable operational categories within each of the three groups of organisms (WHO 1995).

Humans are classified according to a wide variety of criteria such as age, sex, immune status, social and economic parameters, ethnic group and other phenotypic and genotypic characters.

In the case of parasites and vectors the primary task is to recognize different species within the genera *Plasmodium* and *Anopheles*. Further subdivision at the intra specific level will be only applicable once the specific level is well defined. However, Anopheline species identification should be understood as the malariologist's tool towards his actual task, namely the biological characterization and effective monitoring of vector populations in the context of epidemiological observations or control programmes (Coluzzi 1988).

Different behaviours, biologies, and vectorial status of many isomorphic *Anopheles* species make it essential to identify precisely the different species to evaluate their individual roles in malaria transmission (Service 1991). During last ten years, two DNA probe-based methods and one rDNA-IGS diagnostic PCR have been developed to identify sibling species within the *Anopheles gambiae* complex, which contains the most dangerous malaria vectors. One of those DNA-probe approach requires electrophoresis followed by southern blotting (Collins *et al.* 1987, 1988a,b,c) and hence has very limited field application, while the other uses squash-blot and non-radioactive DNA probe protocol for the field identification and is considerably simpler (Hill *et al.* 1991a, 1991b, 1992, Hill & Crampton 1994, Crampton 1994). A constraint is that at present the probe can not distinguish females of *An. gambiae s.s.* and *An. arabiensis*. The diagnostic rDNA- IGS PCR (Scott *et al.* 1993) can identify single specimens of *An.gambiae* species complex members and has been tested on specimens from a wide geographical range and caught during different seasons. However, other PCR-based techniques such as RAPD and microsatellite

proved useful and are needed to investigate any geographic or temporal intra-specific variations in DNA.

On the other hand, molecular genetics techniques can be used to create bio-engineered *Anopheles* that could be made refractory to malaria infection, infertile, or have their behaviour disrupted, for example, so they become unable to locate their hosts (Service 1991).

In summary, at present there is interest in two extreme approaches: low technology and community participation as exemplified by impregnated bednets, and high technology methods, such as the production of transgenic mosquitoes, although after about 50 years from malaria eradication programme and several decades of research on malaria vaccine development and vector molecular genetics studies, malaria situation has not changed and risks have not changed, only the drugs have changed and our understanding of the nature of the infection has advanced.

There is usually a wide gap between theoretical simplicity and elegance of a control strategy generated at the desk or laboratory bench and its successful introduction into the field. Sophisticated and intellectually stimulating control ideas should not be advocated if common sense indicates they are too complicated to be successfully integrated into a disease control programme operating in a tropical, and often developing country (WHO 1995).

2.2 Anopheles species complexes

Among sexually reproducing organisms, a species is defined primarily by reproductive isolation, since it makes irreversible the process of genetic divergence initiated by extrinsic mechanisms of isolation (Dobzhansky 1951; Mayr 1963, 1991; White 1978).

A direct correlation is generally recognized between the degree of a genetical difference and the possible expression of such a difference at the morphological level. However, the available data in different groups of organisms show that morphological divergence may be acquired either more rapidly or more slowly than reproductive isolation. In this process, the first occurrence is likely to produce subspecies, the second, sibling species which were defined by Mayr (1963) as morphological similar or identical natural populations that are reproductively isolated.

The failure of the morphologists to recognize sibling species of *Anopheles* can mean and has meant in the past, failure to distinguish an epidemiologically important species from an unimportant one. Moreover, the study of host preference, infection rate, resting habits, and biting cycle, as well as the assessment of control measures, may be seriously misleading if carried out on a 'morphological species' which is in fact a mixture of two or three species. It is now clear, for instance, that many of the supposed changes in behaviour recorded in malaria vectors under insecticide pressure were due to the different survival of sympatric sibling species (Davidson, 1956a,b, 1962 and 1964a, 1964b, Davidson & Jackson 1962).

Early in the present century it was recognized that within the species group of some *Anopheles* there are forms with biological and morphological differences. The observation of 'Anophelism without malaria' at the end of the first world war and the detection of 'long-winged' and 'short-winged' forms of *Anopheles maculipennis* in the Netherlands and their association with fresh and brackish water of their breeding places and further studies on the basis of their egg markings and behaviour patterns and reproductive compatibility (DeBuck *et al.* 1934) showed that within *An. maculipennis s.l.* there were at least five sibling species. It is now known that in the Palaearctic regions there are ten species within the *An. maculipennis* complex, with another five Nearctic species (White 1978).

2.3 Population Genetics of Malaria Vectors

Population genetics is the study of genetic variation in natural populations and the factors that influence this variation. Population genetics is a critical area of study in arthropod vector biology; thus understanding genetic variation in vector populations provides a foundation for understanding the role of the vector in disease epidemiology. The factors that shape genetic variation in vector species also influence variation in vector capacity and vector competence traits.

Ideally, a population genetic study of a vector should help define:

1 the major factors that influence genetic variation in its natural populations;

2 criteria to measure and analyse this genetic variation;

3 the relevance of this genetic variation to vector-borne disease

epidemiology and vector control.

By far the most extensive amount of population genetic data has been gathered through the use of isozyme electrophoresis. Recent methods for analyzing genetic variation in natural populations of vectors focus directly on the use of DNA markers.

2.4 Polymorphic DNA Markers

Polymorphic DNA markers are based on variation in DNA sequence. There are five basic types of useful polymorphisms; RFLP, Restriction Fragment Length Polymorphism; VNTR, Variable Number Tandem Repeat polymorphism; RAPD, Randomly Amplified Polymorphic DNA; SSR, Simple Sequence Repeat (microsatellite) polymorphism; and SSCP, Single Strand Conformation Polymorphism (Rafalski & Tingey 1993). These markers have not only been used extensively in genome mapping but they have also proved extremely useful in population studies. In this section I provide an introductory overview of the methodological basis for studies of these polymorphisms. Those markers that have been applied to disease vectors are then reviewed in depth in the subsequent section.

RFLP and VNTR polymorphisms are detected by Southern blot hybridization, in which a DNA probe is hybridized to a membrane containing genomic DNA that has been restriction digested, electrophoresed, and transferred to that membrane (Southern 1975). RFLP indicate changes in fragment length that result from the loss (or gain) of a restriction site. VNTR polymorphism is the result of different alleles carrying a different copy number of a short core repeat (10 bp to >100 bp). RAPD

and SSR polymorphisms rely on the use of the Polymerase Chain Reaction (PCR), in which a DNA fragment can be amplified manyfold by repeated denaturation and DNA replication with a pair of primers flanking the DNA target (Saiki *et al.* 1988).

RAPD-PCR amplification uses an arbitrary primer (often 10 bp) and genomic DNA as the template, seen on subsequent electrophoresis, representing loci in which the primer sequence is found in inverted orientation twice, separated by other sequences not exceeding the length (approximately 3 kb) that can be amplified by PCR; some of these bands are polymorphic.

SSR, or microsatellite polymorphism, results from variations in the number of tandemly arrayed di-, tri, or even tetra-nucleotide repeats. PCR amplification of microsatellite markers requires the use of a pair of unique, i.e. locus specific primers that flank the repeat arrays. SSCP results from a variation in sequence that leads to the conformational change in the single-stranded DNA. Although SSCP can detect a single base change in sequence that is undetectable by almost other methods, less routine electrophoretic techniques, such as non-denaturing gradient gel, are required for genotype determination making it intrinsically more difficult to study.

So far only RFLP, RAPD, RAPD-SSCP, and microsatellite markers have been used as genetic markers in disease vectors. RAPD and microsatellite markers are particularly important because they are produced by PCR and hence only require minute amounts of sample DNA; thus, they permit the scoring of single individuals for a very large number of markers.

2.4.1 RFLP

Restriction enzyme digestion of genomic DNA from different individuals, followed by southern blot hybridization with cloned probes of known or unknown sequence, often yields detectable polymorphisms in the size of DNA fragments produced, or RFLPs. The individual to individual variation in the observed hybridization patterns is caused by underlying differences in the target DNA sequence, such as base substitution that create or abolish restriction sites, or DNA insertions or deletions in the region between the restriction sites. RFLPs are dispersed widely throughout the genome, and they are often inherited as co-dominant, Mendelian markers. Any cloned DNA fragment can potentially detect polymorphism and serve as an RFLP marker; thus, significant regions of the genome can be saturated with RFLP markers. RFLP markers have been used for map construction for both plants and animals.

RFLP analysis is limited because it can only monitor a portion of the genome containing the restriction sites under examination. In cases in which a comparison between 2 closely related individuals is required, RFLP may not provide sufficient markers that are discriminating (Beckmann 1988). Another drawback to the use of RFLP analysis with mosquitoes is that only limited quantities of DNA can be isolated from single mosquito (0.5-3 μg), and most of this genomic DNA is required for the blot, as compared to 5 ng (or less) of DNA required for PCR-based polymorphism analysis. In addition, the time required to complete an experiment and the reliance on radioactively labelled probes for greatest sensitivity are also drawbacks, especially

when multiple assays are required from each sample.

Some RFLP-based markers of short length could be converted to PCR-based assays when DNA sequences external to the 2 pertinent restriction sites are known. The genomic DNA is used as template for the PCR amplification with two opposing oligonucleotide primers that flank the RFLP region. After electrophoresis and band isolation, the purified band can be cut using restriction endonucleases and analyzed by gel electrophoresis (Akopyanz *et al.* 1992). RFLP markers from *Ae. aegypti* have been used to examine genetic diversity among 10 laboratory populations of *Ae. aegypti* and 9 populations representing 4 *Cx. pipiens* subspecies (Severson *et al.* 1994). Many of the *Ae. aegypti* markers cross hybridize under conditions of high stringency with other genera of the subfamily Culicinae and to a lesser extent with a member of the Anophelinae (Severson, 1993).

In contrast, examination of mitochondrial DNA (mtDNA) RFLPs in 17 populations of *Ae. albopictus* revealed extremely low level of variation; over 99% of all fragments identified were shared in all populations (Kambahampati & Rai 1991a). When used to examine the relationships of 7 species of the *Ae. scutellaris* subgroup and 4 of the *Ae. albopictus* subgroup, restriction fragment analysis of mtDNA revealed a great deal of polymorphism among the species. These data, however, yielded conflicting results in phylogenetic analysis when compared with the available morphology and allozyme data (Kambahampati & Rai 1991b).

2.4.2 Microsatellites (Simple Sequence Repeats)

The term microsatellite has resulted from an interesting historical progression. Originally, the term satellite DNA was applied to nucleic acid material recovered at specific densities in cesium chloride gradients spun to equilibrium (Kit 1961). If a main band of genomic DNA was recovered at a density of, say, 1.683 in a cesium chloride gradient, often a minor auxiliary or satellite band was observed at a slightly different density such as 1.705 (Haymer 1994). In most cases, these sequences were localized to heterochromatic or centromeric regions of chromosomes (Pardue & Gall 1970).

Microsatellite or simple sequence repeat (SSR) DNA consists of tandemly repeated copies of a core sequence such as $(dA)_n$, $(dG\text{-}dT)_n$, $(dC\text{-}dA\text{-}dC)_n$, or $(dG\text{-}dA\text{-}dT\text{-}dA)_n$ (Tautz & Renz 1984). For example, long contiguous stretches of CA dinucleotides, also referred to as poly (CA) repeats, have been discovered in regions of DNA (Weissenbach *et al.* 1992). Other repeating units such as poly (CT) (Senior & Heun 1993), and combinations such as poly (CA*GT) have also been found (Weber & May 1989).

Microsatellites are found widely dispersed throughout most of the genome in many eukaryotes and can occur as frequently as once every 10 kb (Tautz 1989). Usually they are located outside of the open-reading frame or transcribed regions, but studies in human show that different microsatellite repeats occur with different frequencies in introns, exons or the intergenic regions. Extensive length

polymorphism occurs in microsatellite sequences, permitting their use for relationship studies of individuals within and between populations, and as a source of polymorphic DNA markers for genome mapping and linkage studies (Tautz 1989). The length polymorphisms which originate from unequal crossing-over, unequal sister chromatid exchange, or slipped-strand mispairing during replication or repair (Traut *et al.* 1992). Such copy number variation are the usual basis of allelic variation at microsatellite loci but insertions and deletions in flanking or interruption sequences can lead to non-canonical allelic variation resulting the production of null alleles due to changes in the annealing site of primers used to amplify microsatellite loci.

Repeats such as $(dG\text{-}dT)_n$ have been found in all eukaryotic organisms, except yeast. In *An. gambiae*, conservative estimates suggest the presence of about 10,000 (dG-dT)n microsatellites in the haploid genome. In *Ae. aegypti*, the search for SSRs has not been as fruitful. For example, approximately 10^4 recombinant cosmids were screened with a $(dG\text{-}dT)_{15}$ probe, and microsatellites were detected at frequency of about 1 in 300 colonies screened. Since the average insert size was >30 kb, it appears that $(dG\text{-}dT)_n$ microsatellites are relatively infrequent in the *Ae. aegypti* genome. For comparison, in similar studies on the human genome yields, approximately 1 in 2 colonies examined are positive (Litt & Luty 1989). The *Ae. aegypti* results have been duplicated by others, but 15 SSR markers have been identified and several have been found to exhibit Mendelian segregation. These findings emphasize the need to tailor the search for microsatellite sequences to repeats that occur frequently in the particular vector species. SSR markers distributed along

the 3 Aedes chromosomes will be helpful in reconciling the linkage map with the physical map.

The function of microsatellite sequences is not known. Microsatellites such as $(dG\text{-}dT)_n$, however, have been found to be close to sites of DNA rearrangement (Boehm *et al.* 1989). Furthermore, DNA fragments containing $(dG\text{-}dT)_n$ sequences might form Z-type DNA structure (Haniford & Pulleyblank 1983; Nordheim & Rich 1983). The expansion or contraction of certain trinucleotide repeats has been implicated in several human genetic disorders (Fu *et al.* 1991).

2.4.3 VNTR, or Minisatellite

Minisatellites are hypervariable, more complex regions of DNA, and are found dispersed at numerous sites in the genome. they exhibit polymorphism due to a variable number of tandem repeats (VNTR) of a short 10-15-bp core sequence (Jeffreys *et al.* 1985a,b). A hybridization probe consisting of the core sequence can detect many polymorphic minisatellite loci simultaneously to provide an individual organism's fingerprint; hence the synonymous term DNA fingerprinting.

The use of multilocus minisatellite probes has become a standard tool for the establishment of associations or exclusions in criminal cases and for the determination of family relationships in paternity disputes (Jeffreys *et al.* 1991a). The same minisatellite loci used for human diagnostics were also found to be equally valuable in studying house sparrow demographics (Wetton *et al.* 1987). Burke (1989, 1991)

used DNA fingerprinting , multilocus and single locus minisatellite analysis in population biological studies. Blanchetot (1989) reported a tandemly repeated element from the house fly that created individual fingerprints when probed against genomic DNA. This element also cross hybridized with *Ae. aegypti* DNA.

2.4.4 rDNA Cistrons

In most disease vectors, the rDNA is usually found at one cytogenetic location, for example, near the centromere of the X chromosome in *An. gambiae* and *An. arabiensis* and in chromosome I in *Ae. aegypti* and *Culex quinquefasciatus* (Kumar & Rai 1990; Marchi & Pili 1994). However rDNA can be found at two positions in *An. quadriannulatus*, *An. melas*, and *An. merus* and in *An. quadrimanulatus*, it is present in the Y chromosome as well as the X. In *Ae. triseriatus*, rDNA can be found in chromosomes I and III (Clements 1992).

As in many other eukaryotes, the rDNA is arranged as head-to-tail tandem repeats and is present in 100-1000 copies in most vectors. There are about 250 copies in *An. gambiae* (Collins *et al.* 1989) and about 500 in *Ae. aegypti* and *Ae. albopictus*; the average length of the *Ae. aegypti* repeat unit is about 9.0 kb (Gale & Crampton 1989; Kumar & Rai 1990). The number of repeats increases to about 1200 in adult *Ae. aegypti* (Park & Fallon 1990).

A typical rDNA repeat consists of an intergenic spacer (IGS), followed by the transcribed region, which is post-transcriptionally processed to form the 18S, 5.8S,

28Sa, and 28SB rRNAs. Preceding the 18S coding sequence is the external transcribed spacer (EST), also known as the intergenic spacers (IGS) and surrounding the 5.8S rDNA are 2 internal transcribed sequences (ITS1 and ITS2)).

Genetic variation within and between mosquito populations has been examined by studying rDNA sequence variation. Black *et al.* (1989) found extensive and continuous variation in the non-transcribed spacer regions (IGS, ITS) of *Ae. albopictus* with little evidence of conservation within each of 17 populations from around the world. Although IGS sequences in *An. gambiae* complex show both intra- and inter-species variation, resulting in part from the presence of small repeat sequences, the IGS sequences near the coding units show a species-specific restriction pattern. Species diagnostic IGS probes or primers have been developed to distinguish the morphologically similar taxa of the *An. gambiae* complex (Taylor et al. 1993; Scott *et al.*, 1993).

2.4.5 Mitochondrial DNA (mtDNA)

Mitochondrial DNA (mtDNA) offers another molecular marker for population genetics. MtDNA is primarily maternally inherited. Unlike nuclear markers, there is no recombination among genomes of different maternal clones. Studies of intra-specific variation in mtDNA have shown that one or more maternal clones often characterize local populations. The presence of these clones outside of their typical ranges provides information on gene flow and dispersal. The number of maternal

clones in populations provides information on the number of females founding new populations, and thus evidence for population bottlenecks. Because of this sensitivity, it is possible to map distributional and founding events of population through their matrilineality. This makes it possible to trace maternal lineage and to determine phylogenies of existing populations independent of founder effects and selection or adaptation. Such evidence can not be obtained as easily from studies of genomic systems experiencing recombination or biparental inheritance. mtDNA evolves at a rate approximately 10 times faster than nuclear DNA, and therefore, intra-specific variation is frequently detectable.

To date mtDNA has not proven to be very useful in understanding the population structure of vectors. Kamphampti and Rai (1991a) examined RFLP variation among 17 worldwide populations of *Ae. albopictus* and found 99% of fragments were shared among the populations. Mitchell et al. (1992) found similarly low variability in mtDNA RFLPs in the four sibling species of the *An. quadrimaculatus* species group. This low variability suggests either the populations were too recent in origin to have accumulated differences or that the mtDNA in culicidae evolve slowly.

2.4.6 Transposable Elements

Another important category of interspersed, moderately repetitive DNA is the transposable elements. Transposable elements are mobile genetic entities now known to be ubiquitous components of genomes. These are DNA sequences capable of

moving from one place to another within the genome while retaining a copy at the original location; consequently, they appear in varying locations in different strains. The distribution patterns of DNA-mediated elements are extremely patchy and the principle cause appears to be the horizontal transfer of elements between host lineages. RNA-mediated elements appear to undergo much longer periods of vertical evolution within host lineages and evidence for their horizontal transfer remains scant (Robertson & Lampe 1995).

Much of the current knowledge on transposable elements comes from research on *Drosophila*, in which these elements are responsible for many spontaneous mutations and chromosomal rearrangements. The proportion of a genome that consists of transposable elements varies from species to species but averages about 10% (Finnegan 1992). There are two main types of transposable elements: transposons, which transpose directly from DNA to DNA; and retrotransposons, which transpose by an RNA intermediate and have structural features suggestive of such a transposition mechanism (Boeke & Corces 1989). Transposable elements are of great interest because they could be used to deliver foreign genes into an organism, to tag genes or promoter elements in mutagenesis experiments, or to spread a gene through a disease vector population.

P elements in *D. melanogaster* are an example of transposons. A complete *P* element is 2.9 kb long and has 31 bp inverted terminal repeats. Intact *P* elements occur at a frequency of about 50 copies per genome and encode a transposes enzyme,

which catalyses its excision and transposition. Smaller elements resulting from internal deletions are incapable of transposing by themselves but can transpose when supplied with a functional transposase in *trans* by an intact element (Eggleston 1991). *P* and a related transposon, *hobo* element (McGinnis *et al.* 1983; Streck *et al.* 1986), have been used to develop reliable germline transformation procedures in *D. melanogaster* (Spradling & Rubbin 1983). Transformation has been accomplished with a third element, *Minos*, which was derived from *D. hydei* but absent from *D. melanogaster* and which is related in sequence to the *Tc1* element of *C.* elements (Franz & Savakis 1991; Franz *et al.* 1994). *P* and *hobo* are present in some *D. melanogaster* strains but absent from others, and at least for *P*, strong evidence exists that it has spread in *D. melanogaster* populations across the world in a very brief period, approximately 50 years (Kidwell 1983).

Another transposon is *mariner*, which was identified originally as a insertion element in the white gene of *D. melanogaster* (Jacobson *et al.* 1986). A *mariner*-like element has also been identified from the lepidopteran *Hyalophora cecropia* (Lidholm 1991). It is a middle repetitive element with about 30-70 copies in *D. melanogaster* and about 100 in *H. cecropia*. The element is about 1,3 kb long and has inverted repeats (28 and 38 bp in *D. mauritiana* and *H. cecropia*, respectively) at the 2 ends.

Mariner-like elements have been found in ten other insects, including 6 different orders (Robertson 1993; Robertson & MacLeod 1993). Full-length clones were recovered from all of the species examined, including the mosquito *An.*

gambiae. Interestingly the majority of the mariner elements in *An. gambiae* shared high sequence similarity (91% at the nucleotide level) with those from the horn fly. In light of the more than 200 Myr divergence between these two organisms, it has been suggested that a recent horizontal transfer of this subfamily of elements might have occurred.

Retrotransposons consists of 2 categories, the long terminal repeat (LTR) containing elements, which are similar in structure and function to retroviral proviruses of vertebrates, and the non-LTR retrotransposons (Xiong & Eickbush 1990). LTR elements range from 5-8.5 kb in length and have long, direct terminal repeats and short, imperfect inverted repeats at the terminal ends. The copia-like elements of Drosophila are an example of the LTR retrotransposons. The number of copia elements varies from 20-60 copies per genome. Examples of the non LTR retrotransposons include the mammalian *LINE-1* elements and *Drosophila I* factor, *F* elements and *Jockey* (Berg & Howe 1989), and *R1* and *R2* elements of *Bombyx mori* (Burke *et al.* 1993). They all lack inverted repeats (non-long terminal repeat elements) but have a poly(A) or (A)-rich tail at the 3' end. In each case, the full length element is about 4-6 kb, consisting of 2 open-reading frames. The first ORF contains cysteine-rich domains and the second shares homology with many reverse transcriptases.

Two classes of non-LTR elements have been described on the basis of target site insertion. The random elements appear to be distributed promiscuously

throughout the host genome; a second group, the site-posons, insert into specific DNA sequences (Aksoy 1991) and are exemplified by the *R1* and *R2* elements, which insert at the same location of the 28S rRNA gene in many insects (Burke et al. 1993; Xiong & Eickbush 1990). Besansky (1990a) described the *T1* family of retrotransposons from the *An. gambiae* complex. Full length *T1* elements are 4.7 kb long, including 2 overlapping open-reading frames, one of which possesses structural motifs characteristic of reverse transcriptases. At the 3'end is an (A)-rich tail consisting of tandem repeats of TGAAA. *T1* occurs in approximately 100 dispersed copies belonging to 2 subtypes, *T1a* and *T1B*. Although both subtypes are present in all 5 sibling species examined, *T1a* seems to be absent in *An. merus* (Besansky 1990b).

Two site- specific retrotransposon families, *RT1* and *RT2*, have also been identified from *An. gambiae* complex. Both compete for an insertion site in the coding region of the 28S rRNA gene (Besansky *et al.* 1992), which is 634 bp 3' from the typical site of *R1* insertions in other insects but are also present outside of the rDNA repeat from a few to about 100 copies. Among the sibling species of the complex, *An. merus* seems to be devoid of *RT1* and *RT2* whereas *An. quadriannulatus* lacks *RT1*.

Mouches et al. (1991) described a highly repetitive element, *Juan* from *Cx. pipiens quinquefasciatus* and *Cx. tarsalis*. Sequencing of three different copies revealed similarities to the *Drosophila I*, *Jocky*, and *F* elements. Full length (4.7 kb)

Juan-A elements were shown to be widely dispersed in about 200 copies throughout the genomes of *Ae. aegypti*, *Ae. albopictus*, *Ae. polynesiensis*, *Cx. tarsalis*, *An. gambiae*, and an agricultural pest, *Ceratitis capitata* (Mouches *et al.* 1991, 1992).

Identification and characterization of additional transposable elements in mosquitoes will be an important first step in any future genetic control strategies designed to interrupt vector-born disease cycles. A key goal is to identify potentially exploitable transposable elements that can be used for the development of transformation vectors capable of directing the integration of recombinant genes into the mosquito genome.

2.5 RAPD-PCR

2.5.1 RAPD and other complementary methods

This section will be presented in more details, in order to provide a review the different aspects of this latest offspring of PCR, which its application and development are the most important part of the current study.

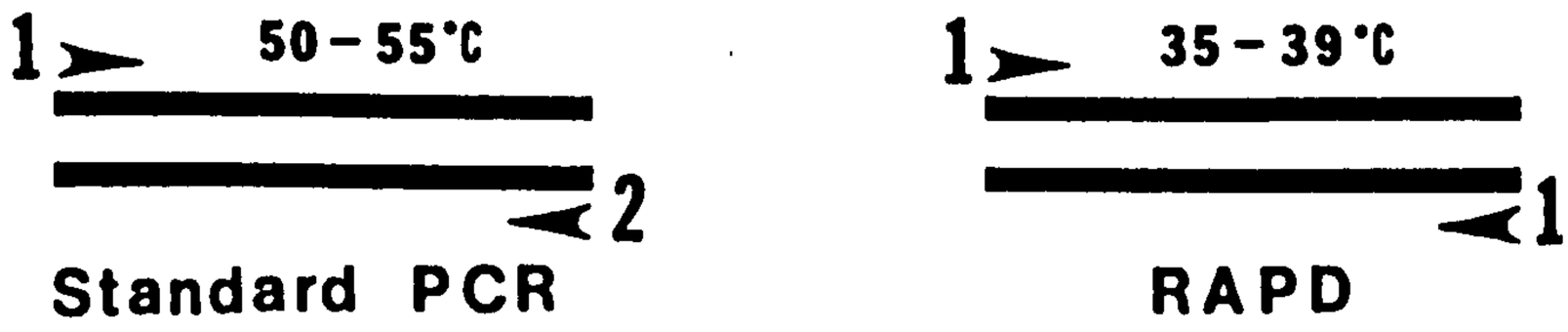
Williams *et al.* (1990), described a new DNA polymorphism assay based on the amplification of random DNA segments with typically single 10-base primers of arbitrary nucleotide sequence called RAPD-PCR (Fig. 2.2). These polymorphisms, simply detected as DNA segments which amplify from one parent but not the other, are inherited in a Mendelian fashion and can be used to construct genetic maps in a variety of species. The technique works regardless of genome size or organization as demonstrated by studies involving bacteria, fungi, plants, invertebrates, and vertebrates (Williams *et al.* 1990, 1993). Increased numbers of polymorphisms were reported through the use of multiple 10-base primers in a single reaction (Klein-Lankhorst *et al.* 1991). While the use of primers with more than 10-base primers has been found useful in some studies (Adamson et al. 1993; Baruffi et al. 1995).

The results of Williams *et al.* (1990) also showed that RAPD assay may in some instances detect single base changes in genomic DNA. However it does not mean that all amplifications are the result of perfect pairing between the primer and the DNA template. Other sources of polymorphisms may include deletions of a priming site, insertions that render priming sites too distant to support amplification,

or insertion that change the size of a DNA segment without preventing its amplification. Some ambiguous polymorphisms may result from poor discrimination by a primer between alternative priming sites of slightly different nucleotide sequences. These authors described nearly all RAPD markers as dominant and it seems that it is not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies) with a dominant RAPD marker. Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are detected only rarely (about 5% of RAPD markers). They suggested that RAPD markers are well suited for genetic mapping, for plant and animal breeding applications, and for DNA fingerprinting, with particular utility for studies of population genetics. RAPD markers can also provide an efficient assay for polymorphisms, which should allow rapid identification and isolation of chromosome-specific DNA fragments. Hybrid cell lines or genetic stocks carrying deletions or additions of large chromosomal segments could be screened relative to appropriate controls, to identify the region of the genome carrying the deletion or addition (Williams *et al.* 1993).

Genetic mapping using RAPD markers has several advantages over other methods: (i) A universal set of primers can be used for genomic analysis in a wide variety of species (ii) no preliminary work, such as isolation of cloned DNA probes, preparation of filters for hybridizations, or nucleotide sequencing is required (iii) each RAPD marker is equivalent of a Sequence Tagged Site (STS) as described by Olson et al. (1989) which can greatly simplify information transfer in collaborative research

Fig. 2.2- Use of two primers of common length 18-25 bases in standard PCR versus the use of a single primer of common length 9-11 bases in RAPD reactions.

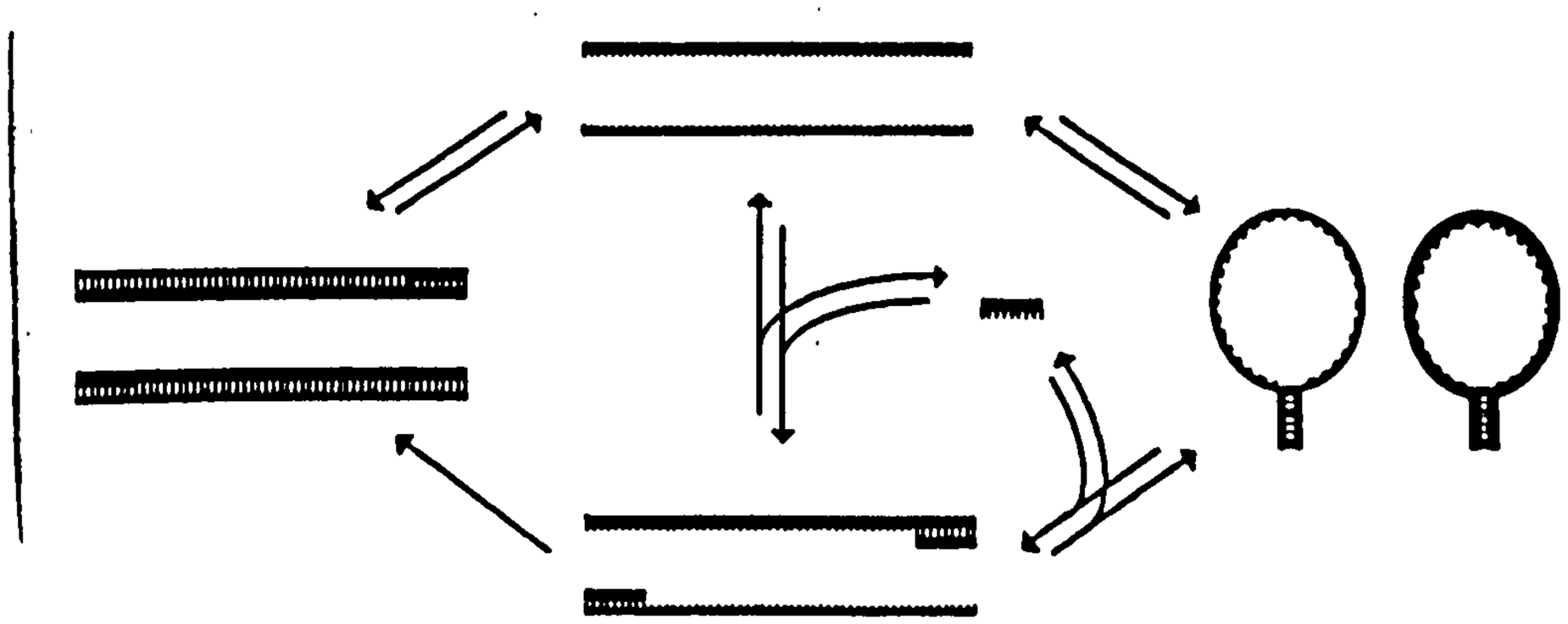


programs. Perhaps the most significant advantage to this method is that the determination of genotype can be automated. Genetic maps consisting of RAPD markers can be obtained more efficiently, and with greater marker density, than by RFLP or targeted PCR-based methods.

Two early modification of detecting RAPD makers have been described as Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) and DNA Amplification Fingerprinting (DAF). Welsh & McCleland (1990) described a method to identify species and strains that is simple and fast, and can be applied to any species for which DNA can be prepared. This method which they called arbitrarily primed polymerase chain reaction (AP-PCR) 'fingerprinting' has the further merit of requiring little knowledge of biochemistry or molecular biology of the species being studied, and it is useful in breeding programs, genetic mapping, population genetics or epidemiology.

As a first application, and to demonstrate the ease and utility of this method, they tested strains from five species of *Staphylococcus*, where their relationships had been determined by DNA-DNA hybridization, eleven strains of the human pathogen *Streptococcus pyogenes*, and three varieties of the rice, an agriculturally important plant. Welsh & McCleland (1990), were able to generate specific patterns of PCR products for each strain and these were generally quite similar within species. However, they concluded that it is possible to extend the AP-PCR method to other species, and using similar conditions and various primers, they have generated

Fig. 2.3 - A model of interaction between molecular species formed during DNA amplification with a single, arbitrary oligonucleotide primer. The first-round amplification products are initially single-stranded and have palindromic termini which allows the formation of hairpin loops. In subsequent rounds of amplification, the products can be in the form of template-template and primer-template duplexes as well as single-strand and hairpin loops. The different species produced tend to establish an equilibrium, while enzyme anchoring and primer extension transform the relatively rare primer-template duplexes into accumulating amplification products (after Caetano-Anolles *et al*.1992).



discrete fingerprints from genomes 50,000 to 3,000,000,000 base pairs in size, including the genomes of viruses, humans and plants.

They also investigated the parameters that affected the production of reproducible, species and strain-specific, AP-PCR fingerprints, including the concentration of salts,primer annealing temperature, template concentration, primer length, and primer sequence. Fig. 2.3 represent a model of interaction between molecular species formed during DNA amplification with a single, arbitrary primer (Caetano-Anolles et al., 1992).

AP-PCR dose not require a particular set of primers, instead, this method uses primers chosen without regard to the sequence of the genome to be fingerprinted. Thus, AP-PCR requires no prior knowledge of the molecular biology of the organisms to be investigated. Each primer gives a different pattern of AP-PCR products, each with the potential of detecting polymorphisms between strains. Thus, the data produced allows the differentiation of even closely related strains of the same species, which in this respect it is similar to isozyme studies.

The second modification or complementary method (DAF), demonstrated in 1991 by Caetano-Anolles, Bassam & Gresshof (1991a,b), uses short random primers of 5-8 bp and visualizes the relatively greater number of amplification products by polyacrylamide gel electrophoresis and silver staining. McIntosh et al. (1996) reported the identification of insect cell lines from Lepidoptera, Diptera, Coleoptera and

Fig. 2.4:- Inter-SSR PCR: A single primer targeting a (CA)n repeat, anchored either at the 3'(light arrows) or at the 5' end (dark arrows) of the repeat, is used to amplify genomic sequence flanked by two inversely oriented (CA)n elements.

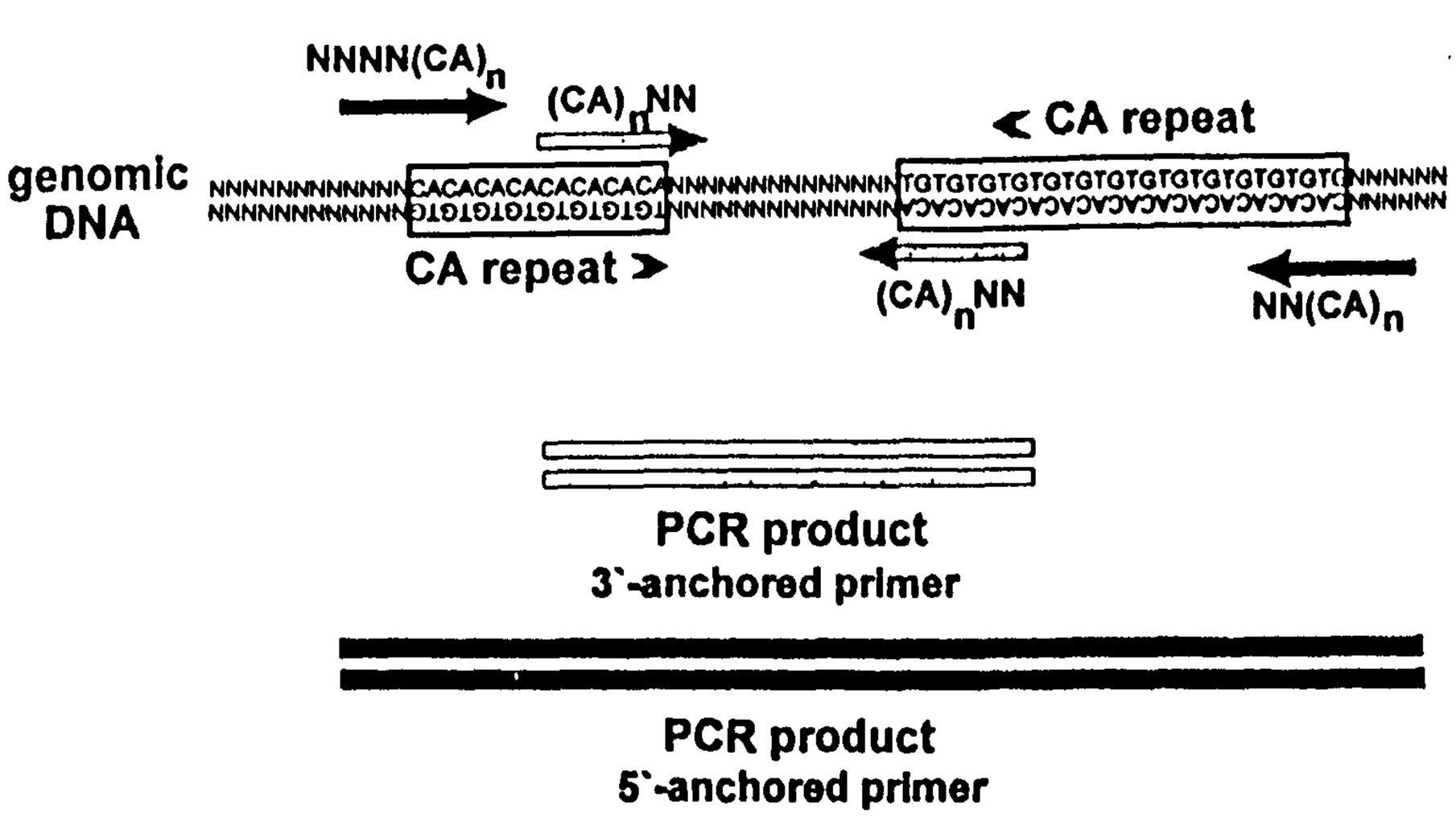
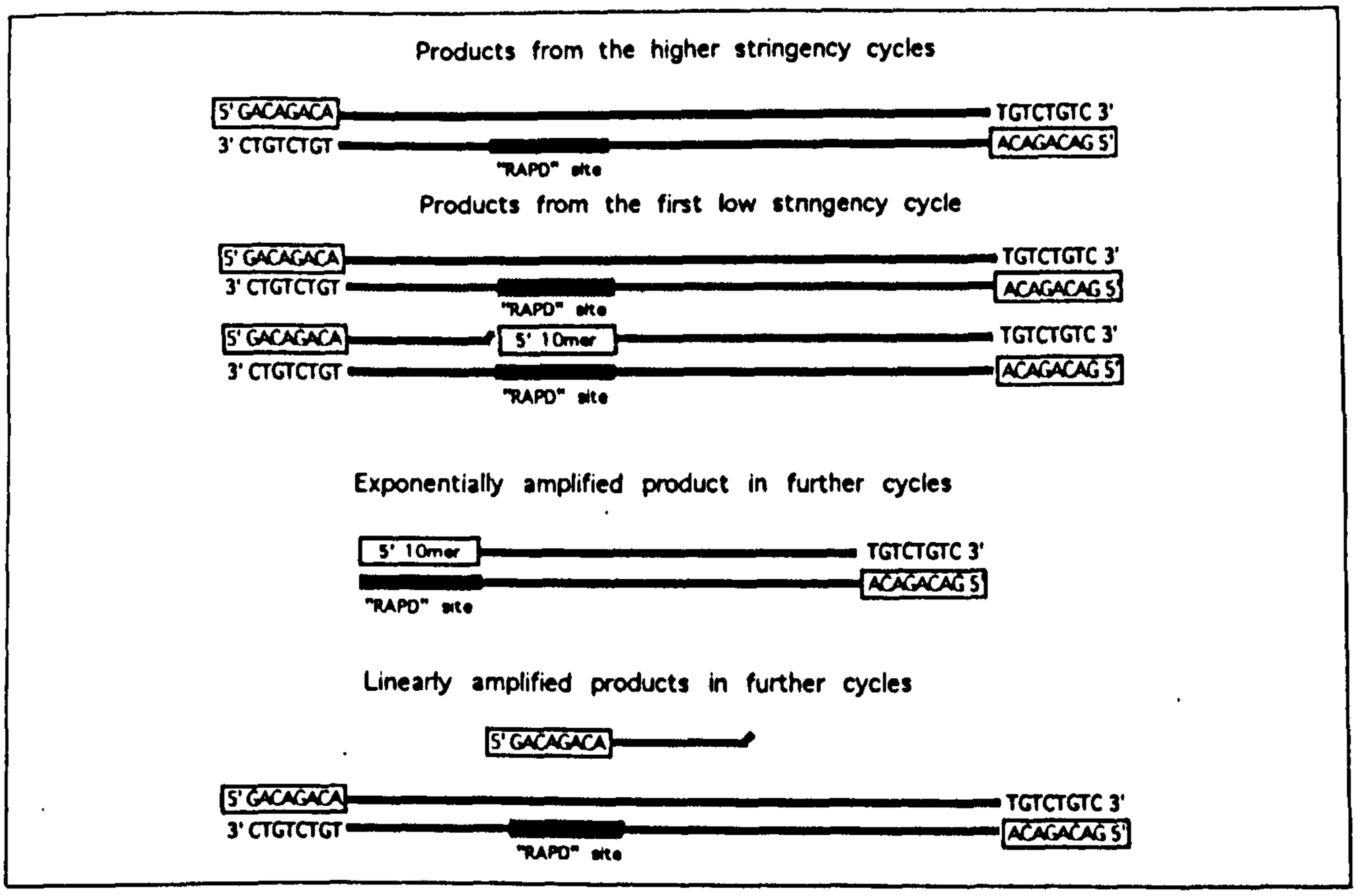


Fig.2.5 - Diagram of the Double Stringency PCR (Matioli & de Brito)



Homoptera by DAF.

In 1995, three other modifications of RAPD, based on the application of both RAPD and microsatellite primers have been described by Matioli and de Brito, Richardson *et al.*, and Cifarelli *et al.*. It is obvious that inter-SSR PCR, using a single 3' or 5' anchored repeat primer, amplify genomic sequences flanked by two inversely oriented elements (Fig. 2.4)(Zietkewicz *et al.* 1994) But, Matioli and de Brito (1995), used two different annealing temperature in a single reaction mixture. During the first cycles, a single microsatellite oligonucleotide is used as primer to amplify DNA between two microsatellite region and then the annealing temperature and consequent stringency is lowered and in further cycles a RAPD primer exponentially amplifies parts of the previously amplified fragments (Fig. 2.5). In contrast, Cifarelli *et al.* (1995) developed Random Amplified Hybridization microsatellites (RAHM), that is based on prior RAPD amplification of target genome, transferring amplified DNA onto Hybond-N+ filter, and then hybridization of filter to a probe carrying simple sequence repeats (SSR) labelled with Digoxigenin-ddUTP. Richardson *et al.* (1995) reported Random Amplified Microsatellite Polymorphisms (RAMPO) method, which is following the same procedure as described by Cifarelli *et al.* (1995) except that the later technique uses a 32P labelled microsatellite-complementary oligonucleotide probe.

2.5.2 RAPD Application

The rapidly growing interest in RAPD technology, reflected in many published applications in molecular ecology, taxonomy and identification, analysis of interspecific gene flow and hybrid speciation, assess kinship relationships, analysis mixed genome samples, genom mapping, and create specific probes in plants and other organisms (Carlson *et al.* 1991; Martin,Williams & Tanksley 1991; Hadrys *et al.* 1992; Welsh *et al.* 1992; Plomion *et al.* 1995).

In determination of taxonomic identity, for any given primer, RAPD amplification products can be broadly classified into two groups: variable (polymorphic) or constant (non-polymorphic). These definitions are relative for a given operational taxonomic unit (OTU). For instance, consider a RAPD analysis of several individuals within a species, and several species within a given genus. constant fragments diagnostic for a genus may be identified (genus-specific marker) (Perring et al. 1993), as well as fragments which are polymorphic between species within the genus (species-specific marker: Arnold, Buckner & Robinson 1991; Black *et al* 1992; Kamphampati *et al.* 1992; Adamson *et al.* 1993; Roherdanz *et al.* 1993; Gawel & Bartlett 1993; Wilkerson *et al.* 1993 & 1995, Garner & Slavicek 1996). Both types of product can be exploited for establishing systematic relationships.

RAPD also provided markers within subspecies- specific (Crowhurst *et al.* 1991; Ballinger-Crabtree *et al.* 1992) and strain-specific (Welsh & McClelland 1991a, Welsh *et al.* 1991b; Chapco *et al.* 1992; Baruffi *et al.* 1995, Guo & Johnson 1995),

biotype-specific (Gawel & Bartlett), serotype-specific (Mazuvier *et al.* 1992), cultivar-specific (Hu & Quiros 1991; Graham *et al.* 1994)), Clone-specific (Morgan *et al* 1993; Wilde *et al.* 1992; Schierwater 1993; Van den ven & McNicol 1995), Clonal " individual"-specific (Smith *et al.* 1994), or which can be used for parasitoid detection (Black *et al.* 1992), and RAPD alleles also have been used as fingerprint markers in estimating the number of full sibling families at an oviposition site (Apostol *et al.* 1993, 1994, 1996). Thus RAPD products can be generated that serve as diagnostic molecular characters at different taxonomic levels.

For the analysis of interspecific gene flow and hybrid speciation, RAPD specific diagnostic markers at individual level can be applied to parentage analysis (Gasperi *et al.* 1991; Scott *et al.* 1992; Hunt & Page 1992); at the population or species level RAPD may be used to detect hybrid populations or species (Hadrys *et al.* 1993, 1992). The detection of hybrids relies upon the identification of diagnostic RAPD markers for the parental genotypes under investigation.

Using fragments that are polymorphic among individuals, RAPD analysis may be used to assess paternity and kinship relationships in large offspring samples. A common problem in behavioral ecology is to determine the actual father from a number of potential fathers. The earliest application of RAPD fingerprints to paternity analysis resolved the question of paternity in an unknown mating system of the dragonfly (Hadrys 1991). For statistical analysis the number of polymorphic markers can be increased by increasing the number of diagnostic primers, and conventional

band-sharing coefficients can be applied (Lynch 1990, 1991; Burke *et al.* 1989, 1991; Keane *et al.* 1991).

In principle RAPD markers can formally be treated as Mendelian alleles, and for parentage analysis, analytical approaches may be developed which are based on knowledge of allelic frequencies, e.g. as used in statistical analysis of single-locus fingerprint profiles. In practice, however, allelic frequencies of scorable dominant RAPD markers in diploid organisms might be difficult to estimate and markers revealed by the same primer may be linked (Williams et al. 1990; Carlson et al. 1991). The occurrence of non-parental bands in offspring from known primate pedigrees has raised concerns in parentage determinations when single individuals are analyzed (Riedy, Hamilton & Aquadro 1992). In contrast, the synthetic offspring (PCR products amplified by mixed DNA from both parents in a single PCR reaction) is a complete representation of both parental genomes and will match the profile of a large sample of offspring, since in both the synthetic and the real offspring clutch, allele combinations that may cause amplification artifacts are represented in equal amounts. The degree of mismatch between synthetic offspring and actual offspring clutch is indicative of mixed paternity, which can be measured by quantitative determination of mixed genome samples (Riedy, Hamilton & Aquadro 1992).

Quantitative estimates of the relative proportion of different genomes in mixed DNA samples may be approached using RAPD-PCR. In many polygamous mating systems,especially in insects, sperm competition and mixed paternity may occur.

Using DNA from a dragonfly, Hadrys et al. (1992), reconstituted mixed-genome samples experimentally by varying the relative proportion of DNA from two male individuals over two orders of magnitude. This DNA was amplified and the relative amounts of individual male-diagnostic amplification products quantified by densitometry. The relative intensity of diagnostic bands from two different individuals varied predictably with the relative DNA concentrations of each genome in the reaction, and the DNA concentration of a given genome could be estimated from band intensities as long as the relative amount of this genome was at least 20% (Michelmore et al. 1991).

RAPD markers can also be eluted from the gel, reamplified, radiolabelled and serve as a probe in southern analyses. In order to improve the specificity of these probes, after reamplification , cloning and sequencing, a partial consensus sequence could be available as a probe. These kind of probes may be used to exclude the possibility of co-migration of fragments of different sequence but similar size. Other applications include generation of probes for taxonomic analysis or the quantitative estimation of the presence of a certain genome in a mixed sample by Southern analysis (Michelmore et al. 1991; Jeffreys et al 1991a). RAPD probes have also been used to detect RFLPs (Martin et al. 1991; Klein-Lankhorst et al. 1991).

There are several potential further applications for RAPD such as 1) sex determination 2) generation of specific PCR primers for anonymous genomes 3) quantitative analysis of mixed biosamples and 4) Phylogeny. However, RAPD

fingerprinting is not without problems, difficulties and limitations e.g.:

2.5.3 Reproducibility of RAPD

Different factors affect the reproducibility of RAPD which will be discuss below:

1) The size of the primer:

Primer size will determine the degree of specificity in genome scanning. It may be expected that primers of short length will amplify an unreasonable large number of sequences and that larger primers will amplify too few sequences to be routinely informative. Beyond a certain primer size (e.g. 15 mer)) increasing primer length may also increase non-specific primer annealing, consequently increasing the probability of random non-reproducible amplification patterns (Black, 1993). All studies using standard RAPD conditions have found 10-mer primers to be an efficacious size (Haymer 1994a, 1994b). A G+C content of the primer similar to the GC content of the analyzed genome will maximize the frequency of binding sites and hence amplification products (Black 1992, Haymer 1994b).

2) Sensitivity to reaction conditions

Slight changes in RAPD-PCR condition may affect the reproducibility of amplification products (Williams *et al.* 1990; Arnold *et al.* 1991; Carlson *et al.* 1991; Klein-Lankhorst *et al.* 1991). RAPD is sensitive to:

a) shape of temperature profile : Although some thermal cyclers allow the profile to be adjusted, the shape of temperature profile is a property of the thermal cycler and must be standardized. Cycling parameters, specially annealing

temperatures and ramp times (the time taken to lower or raise the temperature between two steps), also affect banding patterns. Roherdanz & Flanders (1993) reported variation due to ramp times and magnesium concentrations when working with aphid parasitoids. Ellsworth *et al.* (1993) demonstrated artifacts arising from variation in the above parameters when analyzing DNA from a single grouse. Variation arising through differences in cycling parameters is of greatest concern because a large assortment of thermal cyclers are used by different laboratories and even assuming that temperatures can be standardized, ramp times will inevitably vary.

b) type of Taq polymerase

c) Mg^{2+}concentration

d) template DNA concentration

Only strictly standardized reaction conditions will guarantee reproducible amplification products. Furthermore, the optimal concentration of template DNA per reaction may vary substantially from typical conditions (25 ng per reaction) depending on the primer-template combination used (Carlson et al 1991). However, Ellsworth *et al.* (1993), in working with birds, suggested that standard amounts of template DNA should be used in all RAPD analyses. This is not feasible with most arthropods, especially most mites and small endoparasitic wasps. An alternative solution, used by Black (1993) was to use fixed volumes of DNA extracted from individuals but to increase the primer concentrations to 1 μM (1 pM/μl) or higher. Ellsworth *et al.* (1993) suggested a similar solution using primer concentrations of 1.6-6.4μM.

Black et al. (1992) found a change in the number of bands when comparing RAPD patterns among daughter aphids from a single mother. The change was apparently due to degradation of template DNA during freezing and thawing on consecutive days. This instability was due to the manner in which the DNA was extracted and procedures that yield more stable template have been reported (Ballinger-Crabtree *et al.* 1992)

3) Non-reproducible amplification products

Some RAPD fragments may be ambiguous and not easy to score (Williams *et al.* 1990). These unclear and non-reproducible fragments, which may drive from heteroduplex formation between related amplification products (or other secondary structure artifacts, which can prevent normal amplification patterns) are not useful as genetic markers. However if the RAPD amplification is repeated two or more times, the majority of markers are clearly reproducible and scorable. Sometimes amplification products are found even in the absence of template DNA in the reaction (Innis *et al.* 1990; Klein-Lankhorst *et al.* 1991). However in all reported cases those ghost bands disappear if the template DNA under investigation is added to the reaction. Ghost bands may be a consequence of RAPD amplification of residual bacterial DNA in the supplied DNA polymerase (which is of bacterial origin)

2.5.4 RAPD-PCR analysis

Because of the large number of polymorphic loci revealed within individuals by arbitrarily primed PCR, the data can be used to address intra-specific questions concerning the genetic relatedness of individuals as well as populations. There are various statistical techniques for analysis of RAPD-PCR markers in sibling analysis, population genetics and species diagnostics. The starting point for all of these analyses is a data set containing the RAPD profile of each individual. These data sets typically contain the available genotype information for each individual encoded on a separate line with the name of individual followed by a 1 for the presence of a specific band or a 0 for the absence of the band (Black 1993; Welsh *et al.* 1992, 1991a). in such an analysis only consistently well-amplified bands are used and variation in intensity is ignored. Morgan *et al.* (1993) scored all visible bands and weighted the shared presence of bands. Individual bands were scored as present or absent (1 or 0) for each isolate and the inverse of Jaccard's similarity coefficient (S_j, as mentioned by Sneath, 1957, Sneath & Sokal 1973) was used, as follows:

$$S_j = 1 - a / a+u = a / a+b+c$$

Where a=number of shared bands, b=presence of band, c=absence of band, u= number of mismatches (u=b+c). Values of S_j range from 0 for identical profiles to 1 for profiles with no shared presence of bands. They applied UPGMA (Sneath & Sokal 1973) clustering strategy, using a modified Taxan 2^R (statistical analysis computer package)(Burr, 1970, cited in Sneath & Sokal 1973).

In genetic fingerprinting (comparison of genetic markers among individuals),

pair-wise comparison measure is required. A number of different authors (e.g. Chapco et al. 1992) have used Nei & Li's (1974, 1978, 1979) index:

$S = 2N_{AB}/(N_A + N_B)$

To measure genetic similarity among individuals, where N_{AB} is the number of bands two individuals A and B share in common and N_A and N_B are the total number of bands scored in A and B respectively. S is used when working with VNTR (Variable Number of Tandem Repeats) polymorphisms commonly employed in genetic fingerprinting. However, RAPD-PCR polymorphisms segregate independently, suggesting that they arise from separate loci and are therefore different from VNTR polymorphisms that arise through slippage replication and crossing over within a single genetic locus and do not segregate independently.

Black (1993) used a different approach in which he measured the similarity of individuals by scoring both the shared presence and the shared absence of a band. He estimated the proportion of matches (M) using the formula: $M = N_{AB}/N_T$

Where N_{AB} is the total number of matches in individuals A and B (i.e. both bands absent or present) and N_T is the total number of bands scored in the overall study. Unlike for Nei & Li's S, the denominator (N_T) of M is fixed, and the absence of a band is scored because it represents the recessive phenotype at a locus. An M value of 1 indicates that two individuals have identical patterns; a value of zero indicates that two individuals share no band in common. As with VNTR markers, we also assume that fragments that co-migrate arise from identical alleles. However, we

assume that the absence of a band in two individuals arose from an identical mutation. Both these assumptions are probably valid for local populations but the identity for absence of a band is questionable when comparing distant populations where, given the wide variety of mutations that can eliminate amplification, two recessive alleles are probably not identical in state. The assumption that absence of a band arises from an identical mutation is almost certainly invalid above the species level, whereas co-migrating bands shared by related species may represent identical alleles.

M values are used to construct a symmetrical distance matrix and this matrix is collapsed to construct a dendrogram using UPGMA with a FORTRAN program RAPDPLOT. Application of these techniques with RAPD markers for fingerprinting studies have recently been described (Apostol *et al.* 1993)

Molecular taxonomy and systematics are important areas in which RAPD-PCR may be applied. For these studies Nei & Li's S is usually used because, as Welsh *et al.* (1992) pointed out, many of the arguments applied to the use of RFLPs in phylogenetic analyses (e.g. gain or loss of a restriction site through point mutations) also apply to arbitrarily amplified loci. Studies have used both distance and parsimony methods in phylogenetic analysis computer packages (e.g. PAUP (Swofford 1993) and PHYLIP (Swofford 1996)) to analyze RAPD data sets.

Kambhampati *et al.* (1992) and Fang *et al.* (1993) used cluster analysis of RAPD markers to test whether the derived phylogeny was congruent with those

derived by morphology, isoenzyme and mitochondrial DNA (mtDNA). The results revealed that species belonging to the same subgroup did not cluster together. This suggests that shared presence of bands among species do not reveal 'true' phylogenic relationships among those taxa.

Co-migrating bands may have no sequence similarity. Furthermore, homologous regions might not be amplified in both of the species to be compared and, even if they are, but may would not be co-migrating might vary greatly in size. homology could be tested by extracting bands, labelling them and attempting to hybridize them to amplified DNA in other species (Quiros *et al.* 1995). If cross hybridizing bands are identical in size, then this could legitimately be scored as a shared character. However if they differ in size, then the number of mutational events that could lead to size differences render any estimation of genetic distances based on size difference tenuous. Sequencing would be required.

While arbitrarily primed PCR reactions will probably not be useful in phylogeny reconstruction, the patterns they reveal can be species-specific (Kambhampati *et al.* 1991). In the cited studies, when RAPD patterns of 'unknown' individuals were analyzed alongside patterns from identified specimens, 'unknown' species were correctly placed into clusters with conspecific. This suggests that cluster analysis of RAPD patterns can potentially be used as 'bar-codes' for species identification. RAPDs undoubtedly find use in molecular taxonomy.

For population genetic studies, the simplest analytical procedure involves making the assumption that genotypes at a locus are in Hardy-Weinberg equilibrium and then calculating the frequency of recessive individuals. An assumption of this method is that recessive alleles are identical in state. Dominant allele frequencies are one minus the recessive frequency. Once these have been calculated, genetic distances can be computed. Furthermore Wright's F_{ST}, a standardized measure of genetic distance among populations, can be computed as can a hierarchical analysis of populations (Wright 1978; Apostol *et al.* 1996). Program like RAPDBIOS will convert a RAPD data set into a DATYP=3 dataset for analysis with conventional population genetic tools such as BIOSYS-1 (Swofford & Selander, 1981, 1989). This approach was used in separating populations and subspecies of *Aedes aegypti* (Ballinger-Crabtree *et al.* 1992).

Apostol *et al.* (1993) estimated allele frequencies by making four assumptions in a study of the breeding structure of *Ae. aegypti* based on RAPD-PCR markers. The first assumption was that RAPD products segregate as dominant alleles in a mendelian fashion. Mendelian segregation was observed in an earlier genetic fingerprinting study of *Ae. aegypti* (Apostol *et al.* 1993) and they used these markers to construct an *Ae. aegypti* linkage map. The second assumption was: genotype frequencies at RAPD loci are in Hardy-Weinberg proportions. The third, that alleles in a homozygous recessive individual are *identical in state* (*iis*) (i.e. that they arose from identical mutations) among and within individuals. The fourth, that dominant, amplified alleles are also *iis*. They found that the average genic heterozygosity among RAPD loci (H=0.354)

was over twice that among 11 allozyme loci in an earlier survey in Puerto Rico (H=0.163; Wallis *et al.* 1984) or in a survey of 23 allozyme loci in *Ae. aegypti* populations world-wide (H=0.152; Tabachnick *et al.* 1978).

In a particularly useful study, Clark & Lanigan (1993), considered whether the degree of similarity of the banding patterns could be used to estimate nucleotide diversity and nucleotide divergence. They concluded that with haploid data, fragments generated by RAPD-PCR can be treated in a fashion very similar to that for restriction -fragment data. However, amplification of diploid samples, requires consideration of the fact that the presence of a band is dominant to absence of the band. After describing a method for estimating the expected number of RAPD bands and nucleotide divergence on the basis of diploid samples, they summarize the restrictions and criteria that must be met when RAPD data are used for estimating population genetic parameters. Their criteria for estimating nucleotide divergence of closely related taxa, are :

1 Primer selection must not be biased in favour of those that reveal the most polymorphism. Commercially available primers tend to have a G+C content of 60%-80%, which may result in an overestimate of human nucleotide diversity, because of the high degree of polymorphism at CPG sites.

2 All polymorphic and monomorphic bands must be carefully scored. If some bands are not scored, then there must be no bias in scoring monomorphic bands versus polymorphic bands.

3 Polymorphic bands must be shown to behave as Mendelian factors.

4 Allelism of bands and homology of bands of the same size in different species must be ascertained by Southern blotting or segregation analysis (Which is not practical).

Even if one can identify bands that segregate as good Mendelian markers, DNA preparations of low quality may result in higher rates of mispriming, making it impossible to get an accurate count of monomorphic bands (Williams *et al.* 1990). For the purposes of estimation, the monomorphic bands should represent unvaried sites, which is not the case for PCR artifacts (Clark & Lanigan 1993). Later in 1994, Lynch and Milligan, presented estimators for several population genetic parameters (gene and genotype frequencies, within and between-population heterozygosities, degree of inbreeding and population subdivision, and degree of individual relatedness) along with expressions for their sampling variances.

2.5.5 Other considerations

The most common DNA fingerprint technologies differ substantially in 1) complexity of technological procedures, 2) required amount of DNA, 3) sequence information needed for a genome being scanned, 4) analytical power of assigning genotype relatedness, 5) expense in term of labour and money, 6) broadness of applications. In this context, RAPD fingerprinting seems likely to have wide potential for applications in molecular ecology, and requires the least in technology, labour and expenses.

The cost of producing one individual DNA fingerprint by southern hybridization can be very substantial (Weatherhead & Montgomerie 1991); in contrast, the average expense for one RAPD fingerprint can be as low as $2.00. On the other hand RAPD markers are not competitive with analysis using sequence information or single locus probe fingerprint technologies. However RAPDs are detected more easily than RFLPs and can be competitive with RFLPs in analysis of genomes with high level of heterozygosity (Williams *et al.* 1990; Carlson *et al.* 1991; Hu & Quiros 1991).

2.6 The *An. gambiae* species complex

2.6.1 Ecological and biological characters

In each result chapter, the relevant species will be discussed but because of the importance of species concepts, especially those sibling species in this study, in this section, the *An. gambiae* species complex, and its identification techniques have been reviewed.

Anopheles gambiae, the most important malaria vector in Africa has been studied more than any other species complex within malaria vectors. Until 1956, *An. gambiae* was considered a single species with some dark or light varieties breeding in fresh and salt water (Ribbands 1944; Murihead-Thompson 1948, 1951; Davidson 1956). A vast amount of research has now shown that it is complex of at least six sibling species exhibiting differences in behaviour and each with inversions or other micromorphological features of the polytene chromosomes (White 1974). Crosses between the members of these sibling species produce sterile hybrid males but all hybrid females mosquitoes are fertile, although in certain crosses few or no females are produced (Davidson 1966, 1967; Davidson & White 1972; Davidson & Hunt 1973).

Mosquitoes of the *Anopheles gambiae* species complex include the major vectors of human malaria in tropical Africa and in some regions of Africa are the primary vectors of bancroftian filariasis and some arboviruses. Four out of six are

freshwater breeding species: *An. gambiae* Giles *s.s.*, *An. arabiensis* Patton, *An. quadriannulatus* Theobald, and *An. bwambae* White. Two species are usually restricted to salt water breeding sites: *An. melas* Theobald and *An. merus* Donitz. *An. gambiae* and *An. arabiensis* are the most abundant and widespread of the *An. gambiae* species and are responsible for the majority of malaria transmission. Both species co-exist widely over much of the continent, but two trends are apparent. The first is the evident absence of *gambiae s.s.* from all areas to the north-east of Uganda and the Kenya highlands, whereas in West Africa its range extends well into the Sahel in Mauritania. The second is the general absence of *arabiensis* from the forest belt and more humid areas of West Africa (Coluzzi *et al.* 1979).

However their relative abundance may vary considerably over time and by locality (Service 1993). This is probably due to climatic variations, but may also involve differences in ecology between the two species. Therefore assessing mosquito populations and their ecology on a local basis, should not be generalized to other areas.

The infectivity of *An. gambiae* s.s. and *An. arabiensis* for malaria sporozoites is usually several percent (4.1, 2.2%) of the specimens examined (Mbogo *et al.* 1993). Mbogo 1993, demonstrated a high degree of human-feeding by *An. gambiae* s.l. (94.4%) on the Kenya coast by ELISA, and Robert *et al.* (1995) working in Cameroon found 25% of total (1030) collected mosquitoes were *An. gambiae*.

The use of residual insecticides in houses and impregnated bednets are used for control of these vectors. Karch (1993) studied mosquito nets impregnated against malaria in Zaire, where malaria is stable and *An. gambiae* was the main and nearly exclusive vector of malaria. He reported 94% reduction in *An. gambiae* density in the village protected by impregnated bed nets and a 98% decrease in infective bites/man/year. In these areas the longevity of the vector expressed by its expectation of life decreased from 11.02 days to 3.64 days and malaria prevalence decreased by 50%, five months after the beginning of the experiment.

He found that treated mosquito nets are efficient when used on a large scale and not on an individual level. Such a large scale use, acting on longevity and infectivity of vectors, always induced a decrease in malaria transmission by more than 90% .

Treated bed nets had no significant effect on the overall parasite rate, showing that malaria transmission was not stopped. But it was usually found that there was a significantly smaller number of children with parasitemia higher than a critical threshold, a sensitive parameter of malaria morbidity. In *An. gambiae* complex, *An. arabiensis* resistance to dieldrin (Davidson 1956) and DDT in Sudan and Senegal were reported first and then *An. gambiae* resistance to dieldrin (Davidson 1962) and DDT resistance in Burkina Faso and Togo were reported.

An. quadriannulatus appears to have a very distinct distribution although doubtless other localised populations in the equatorial zone remain to be discovered.

An. quadriannulatus is very zoophilic and exophilic and therefore not thought to be a vector of human malaria. In several control programmes, the failure to eradicate An. gambiae by home-spraying actually represented simply a failure to eradicate the exophilic and zoophilic *An. quadriannulatus* sibling species (Brink 1958).

An. bwambae distribution is restricted to an area in Uganda, where it breeds in mineral water swamps it is a local vector but a very minor vector. *An. melas* is sympatric with *An. gambiae* s.s. and *An. arabiensis* on the west coast of Africa where it breeds in salt water mangrove swamps. Similarly, An. merus is sympatric with the two common fresh water species on the east coast of Africa, its distribution may also extend inland along river during the dry season and has been reported from Swaziland (Patterson 1963) and Madagascar and Mauritius.

An. melas and *An. merus* easily breed in saltwater with up to two times salinities of seawater and also in freshwater. Both species have a physiological capacity to sustain *Plasmodium falciparum* similar to that of *An. gambiae* and *An. arabiensis*, However in natural populations they do not develop such high infectivity rates.

2.6.2 Identification of species of the *An. gambiae* complex

The six species of the *Anopheles gambiae* complex differ widely in their ability to transmit malaria (Gillies & Coetzee 1987). Surely it is because of the differences in their house resting behaviour and breeding site preference we need to

distinguish. If all the species had identical behaviour but differed in their physiological capacity to support parasites we would ignore the fact that these are 6 species when it come to contrast. Several methods have been described for species identification, including ecological evidence, coupled with slight morphological distinctions, crossing experiment, cytogenetic analysis (Coluzzi, 1964, 1966, 1968, 1970, 1972; Coluzzi and Sabatini 1967, 1968, 1969; Coluzzi *et al*, 1973, 1975, 1979; Hunt 1972, 1973, 1987, Hunt & Coetzee 1995) and isoenzyme diagnosis (Mahon *et al.* 1976; Miles 1978, 1979; Marchand *et al.* 1985; Hunt & Coetzze 1986a, 1986b; Collins *et al* 1988b), and cuticular hydrocarbons analysis (Carlson & Service 1980; Philips *et al.* 1987, 1988).

Mensural and Morphometric characters

Morphometrics is the measurement and analysis of form. The form can be virtually anything: A lake basin, sand grain, cellular organelle, ape skull, or the form of a phenomenon. Needless to say, morphometrics is not a coherent discipline. It is practised in partial isolation in such fields as anthropology, cytology, entomology, geology, nematology, and paleobiology. Its unity as a discipline is derived from the use of common statistical methods and certain special techniques that have broad application (Daly 1985). An example of a morphometric method in *Anopheles*, using a simple statistical approach is the use of the number of antennal coeloconic sensilla and/or the palpal index for distinguishing most females of salt-water *An. melas* and *An. merus* from the sympatric fresh-water species of *An. gambiae* complex (Coluzzi 1964). In Tanzania a combination of both characters was needed to separate *An.*

merus from *An. gambiae* and even then it only provided 94% segregation, and *An. gambiae* could not be reliably distinguished from either species (White 1979).

Gillies and Coetzze (1987) point out that while morphological characters have be examined in an attempt to identify species-specific patterns, most of them include characters that are only locally valid, such as hind leg banding patterns, which enable separation of southern African populations of *An. gambiae, An. arabiensis* from *An. merus* and *An. quadriannulatus*. No morphological characters that can be universally applied have yet been found.

Chromosomal differentiation or Cytotaxonomy

Polytene chromosomes are giant chromosomes that may be 200-300 times bigger than mitotic chromosomes (WHO, 1975). Each chromosome consist of spiralised DNA helical filaments known as chromonemata. In Diptera, polytene chromosomes can be found in different organs and different life stages, for example in the adult ovarian nurse cells of some mosquito species (Coluzzi, 1966, 1968), larval salivary glands of some mosquitoes and simuliids (Coluzzi and Kitzmiller, 1975; Procunier and Post, 1986), malpighian tables of mosquitoes and phlebotomines, rectum and even the pulvilli of some insects. However some insects do not appear to have readable polytene chromosomes, including many culicine species, some anophelines and phlebotomine sandflies. In all the *Anopheles* species studied so far, six chromosomes are found in the mitotic karyotype (Coluzzi, 1988). When they polytenise, the centromeres fuse and the homologous bands synapse. Thus, only one

arm is seen (n=6, but 3 chromosomes are visible). If there are any deletions, inversions or other such alternations, the altered region of the arm will not synapse. This leads to splitting of the chromosome which will be synapsed either side of the asynapsis.

Polytene chromosomes can therefore be characterised by their banding patterns and cytogenetically useful maps can be produced for the purpose of identifying sibling species in a complex on the basis of differences in the banding patterns. However, in some cases, for example two members of *An. maculipennis* species complex, *An. labranchiae* and *An. atroparvus* are homosequential species and have identical banding patterns .

The polytene chromosome differences found among the species of the *gambiae* complex consist essentially of paracentric inversions which lead to in changes of band sequence. On the basis of the available bionomical and morphological data, one would be inclined to regard *arabiensis-gambiae* and *merus-melas* as the most closely related pairs of sibling species in the complex. This is contrary to the chromosomal evidence which indicate the phylogenetic relationships shown below (Coluzzi *et al.* 1979).

In the *An. gambiae* complex, chromosome X which represents 11% of the total polytene complement length has five out of ten fixed inversions, while chromosome 3, being 37% of the complement has only one fixed inversion. The expected numbers of fixed inversions is found only on chromosome 2. Chromosome

2R, representing less than 30% of the polytene complement, carries more than 60% of the polymorphic inversions.

Separation of the different members of the complex is now primarily based on differences in banding patterns of polytene chromosomes, following the initial studies by Coluzzi and Sabatini (1967, 1969) on the chromosomes found in the salivary glands of fourth stage larvae with fixed inversion differences between the five species, *gambiae, arabiensis, quadriannulatus, melas* and *merus*. The discovery of polytene chromosomes in the ovarian nurse cells of half gravid females (Coluzzi 1968) provided an adequate tool for the identification of all members of the complex (Green 1971; Hunt 1972).

Crossing Studies or Reproductive Isolation

All 30 possible crosses between the six sibling species have been achieved in the laboratory, results and references are given by Davidson (1964a, 1964b) and Davidson & Hunt (1973). Hybrids almost , always exhibit 'hybrid vigour' in their viability, but only two of the 30 crosses appears to give fertile hybrid progeny, namely those between *gambiae* male and *quadriannulatus* female and between species D (*bwambae*) male and *quadriannulatus* female.

All the other crosses, including the reciprocal of those just mentioned, give sterile F1 hybrid males. The degree of sterility varies with the parental species involved and the hybrid females are always fertile and in some cases show a higher

fertility than the parent females.

Estimates of the frequency of hybridization in nature have been obtained of wild specimens collected in areas of sympatrism, since all hybrid chromosomal complements are easily recognized. However, the available data revealed an overall frequency of hybridization lower than 0.1%, indicating the existence of efficient precopulatory mechanisms of reproductive isolation.

Electrophoresis

Another method for identifying members of the *An. gambiae* Giles group of mosquitoes is by visualizing allozymes using horizontal starch or vertical acrylamide gel electrophoresis. This method depends on using the relative mobilities of enzyme molecules to recognize the alleles at enzyme loci (Mahon et al. 1976; Miles 1978, 1979). Differences in allele frequencies, or even the absence of certain alleles may be used to characterise populations or even species.

Reliable identification by diagnostic electromorphs requires correlation of the enzyme electromorph frequencies to a *priori* groups. This should be defined for each species occurring in an area, either chromosomally or by their crossing characteristics with known members of the group. Mahon et al. 1976; Miles 1978,1979; Marchand and Mnzova 1985; Hunt & Coetzee 1986a, 1986b and others carried out different diagnostic allozyme studies for routine identification of the *An.gambiae* complex species either laboratory or natural population, which were scored for the enzyme

protein structural loci coding for AdK, Est 1, Eat 2, Est 3, GoT, a Gpd, HK, Idh, Ldh, Lap-2, Me, Odh, Pgm-1, Pgm-2, 6-Pgd, Phi, Sod, etc. Some of them lead to a biochemical key for routine identification of adults of the *An. gambiae* complex (Miles 1979, Marchand & Mnzava 1985).

Cuticular Hydrocarbons Analysis

It is one of the biochemical techniques used to identify the adults of sibling species (and larvae and pupae of most insects). It involves determining species-specific differences in the hydrocarbons contained in the wax layer of insect cuticle. Carlson and Service (1979) were the first to use the method to identify sibling species of Diptera. They were able to distinguish adults of the *An.gambiae* complex (*An. gambiae* s.s. from *An. arabiensis*) by differences in their cuticular hydrocarbons.

The technique at that time was unrefined both statistically and chromatographically, but none the less illustrated that species-specific differences between sibling species did not exist.

The application and advantages of this technique are :

1 the identification of larvae and pupae

2 analysis of pinned dry specimens and identification of the origin of laboratory colonies

3 combination with other techniques (e.g. isoenzyme, DNA-probe and immunological techniques

4 its ability to use dead dry specimens

However this technique requires sophisticated and relatively expensive laboratory equipment as well as biochemical expertise and these requirements clearly obviate its use as a field method.

DNA-based methods

DNA - Probe

New methods based on species-specific DNA sequences have been tested in recent years. DNA-probes have great potential for the detection of pathogens and the identification of cryptic species of mosquitoes (Cockburn 1989,1990; Collins *et al.* 1988a, 1988b, 1988c). One of these, the DNA dot blot method (Gale & Crampton 1987) using two different DNA probes produced good specific separation of *An. gambiae* s.s., *An. arabiensis*, *An. melas*, and *An. merus*, except that unfortunately one of the cloned sequences is sex linked so these probes can only be used to distinguish the males of *An. gambiae* s.s. and *An. arabiensis*. DNA-probe detection is roughly comparable to ELISA in difficulty, however it has not yet became a common tool in applied entomology. Non-radioactive DNA labelling and detection kits (similar to ELISA kits) are commercially available, but non-specific signals with crude insect homogenates makes these kits of limited use.

However Hill *et al.* (1991a, 1991b, 1992) and Hill & Crampton (1994) have developed Deoxyribonucleic acid (DNA) probes, for the identification of specimens in a number of important vector species including *Anopheles gambiae*. They introduced a simplified, non-radioactive DNA-probe protocol for the field

identification of insect vectors. They have used alkaline phosphatase-labelled oligonucleotide probes which is based on highly repetitive sequences from the non-coding region of the genome. The method has been simplified in numerous ways to make it as field-friendly as possible. it is currently being evaluated for possible future use as a field tool.

rDNA

Ribosomal DNA (rDNA) is the source of species-diagnostic restriction fragment length polymorphism (RFLP) (Collins *et al* 1987, 1988a) that have been tested against bench-mark cytological techniques (Collins *et al* 1988b) as well as allozyme analysis (Collins *et al.* 1988c). RFLP analysis is highly accurate and can be used regardless of life stage (except first instar) (Copeland *et al.* 1992) or sex of the mosquito tested, but it relies on radioactively DNA probes. A new promising method, in terms of practical usefulness, was introduced by Paskewitz and Collins 1990, based on amplifying DNA fragments of species specific length using the polymerase chain reaction (PCR).

Scott et al. (1993 and Cornel & Collins (1996), extended the rDNA-PCR approach to include the five most widespread members of the complex and can now identify individuals of any life stage, gender, or species hybrids with only a small part of a specimen. Recently this approach has been extended to the sixth species *An. bwambae* (Harbach *et al.* 1997). They suggest that this technique could be readily extended to other *Anopheles* species as well as other important complexes of cryptic

species.

The internal transcribed spacer has been successfully exploited for this assay that distinguishes the sibling species *An. freeborni* and *An. hermsi* (Porter & Collins 1991) as well as the members of the *An. quadrimaculatus* and *An. punctipennis* complexes (Cornel *et al.* 1996). In the *An. gambiae* complex, the rDNA internal transcribed spacers are not sufficiently different among the complex members to be a source of species-specific primers (Collins & Paskewitz 1996).

In summary, the use of rDNA for the identification of diagnostic sequences for cryptic *Anopheles* species complexes is a valuable technique. The easily accessible ITS region have been shown to contain diagnostic differences in most species complexes thus far examined (Townson & Adeniran, in press). For groups where the ITS regions provide an insufficient level of variation, the IGS region, if short enough, can be accessed by PCR with 18S and 28S primers (Collins & Paskewitz 1996). Differences identified in the IGS and ITS may also be useful for oligonucleotide hybridization to squash or dot blots (Collins & Paskewits 1996). However, every species-diagnostic assay requires a detailed picture of the degree of intra-specific variation relative to interspecific variation (e.g. Tabachnick & Black, 1994), and although intra-specific variation in the rDNA spacer sequences of *Anopheles* species has generally been found to be low relative to interspecific variation, the same may not be true for other insects (Collins & Paskewits 1996).

RAPD-PCR

The usefulness of random amplified polymorphic DNA (RAPD) was examined as a potential tool to differentiate cryptic mosquito species. It proved to be a quick, effective means of finding genetic markers to separate two laboratory populations of *An. gambiae* and *An. arabiensis* (Wilkerson et al 1993). In an initial screening of fifty-seven RAPD primers, 377 bands were produced, 295 of which differed between the two species. Based on criteria of interpretability, simplicity and reproducibility, thirteen primers were chosen for further screening using DNA from thirty individuals of each species of which produced diagnostic bands in two species.

Favia et al. (1994a) used RAPD markers for genome analysis in the malaria vector *An. gambiae*. Using more than eighty different commercially available primers they identified more than six different DNA segments that were differentially amplified in different strains of *An.gambiae* s.s. and *An.arabiensis*. An estimate of the cytogenetic position of these markers was provided by hybridization to chromosomal division specific dot-blot filters. They suggested that RAPD markers can be genetically mapped relative to other genes or rearrangements (Dimopoulos *et al.* 1996). They also identified markers of their potential use in the identification of different mosquito strains.

In another study, Favia et al. (1994b) applied PCR amplification using random primers to distinguish between incipient species of *An.gambiae*. Individuals belonging to three chromosomally characterized West African forms of this mosquito, which are

important epidemiologically as they differ in vectorial capacity, were sampled both from laboratory stocks and from wild populations collected in three localities. Primers derived from tRNA gene consensus sequences and the decameric RAPD primers, allowed for the unambiguous classification of the mosquitoes, providing a tool for rapid and efficient diagnosis, which previously relied on cytological examination of polytene chromosomes.

Recently, RAPD markers have been integrated in the genetic and cytogenetic maps of the malaria vector mosquito, *An. gambiae* (Dimopoulos et al. 1996). Fifteen of these markers were mapped by recombination, relative to microsatellite markers that had been mapped previously. Thirty-four gel purified RAPD bands were cloned and sequenced, generating sequence tagged sites (STSs) that can be used as entry points to the An. gambiae genome. They localized thirty-one of these STSs on polytene chromosomes through their unique hybridization signal in *in situ* hybridization experiments. Five STSs map close to the breakpoints of polymorphic yinversions, which are notable features of the *Anopheles* genome.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study area

Mosquitoes have been provided from 11 countries in Africa and 7 countries in the Middle East and Indian sub-continent (Fig. 3.1A, 3.1B countries highlighted in map of Africa and Asia). Table 3.1 summrized the origin, sample size and preservation methods of mosquito specimens used in this study.

3.2 Materials

Biological materials

3.2.1 Mosquitoes

An. gambiae species complex :

In this study air-dried, alcohol or Carnoy's preserved, and pinned dried specimens from different parts of Africa have been used. Here is the list of countries which their field collected specimens or laboratory strains of *An. gambiae* species complex members have been examined in specific and intra-specific level. Fig.3.1A shows the origin of *An. gambiae* species complex used in this study (see also Table 4.1, 4.1B).

An. gambiae s.s. :	Tanzania, Kenya, Uganda, Nigeria, Ghana, Gambia
An. arabiensis :	Zimbabwe, Tanzania, Uganda, Kenya, Sudan, Ghana, Burkina Faso, Mozambique
An. merus :	South Africa, Mozambique

Fig. 3.1A- Origin of *An. gambiae* species complex used in this study (red circles) from East and West Africa (see also Ch. 4 for more details).

Fig. 3.1B-Origin of malaria vectors used in this study from Middle East and Indian subcontinent: An. culicifacies (Brown circles), An. stephensi (purple circles), An. subpictus (blue circles), An. fluviatilis (green circles), An. maculipennis complex (red circles)

Table 3.1- Origin, sample size and preservation method of mosquito specimens have been used in this study. More details in chapter 3 (text) and related result chapters.

SPECIES	ORIGIN	SAMPLE SIZE	PRESERVATION METHOD
An. gambiae s.s.	Tanzania (KWA, ZANDS, KILL)	271	Fresh, Dead, Alcohol, Carnoy's
	Kenya	55	Dead, pyrethrum spray collection
	Nigeria (16CSS)	110	Fresh, Dead, Alcohol, Carnoy's
	Gambia (GAG)	30	Fresh, Dead, Alcohol, Carnoy's
	Uganda	35	Dead, Alcohol, Carnoy's
	Ghana	23	Dead, Alcohol, Carnoy's
An. arabiensis	Tanzania	19	Dead
	Kenya	40	Dead, pyrethrum spray
	Uganda	29	Dead, Alcohol, Carnoy's
	Ghana	20	Dead, Alcohol, Carnoy's
	Mozambique	7	Dead
	Zimbabwe (KGB)	63	Fresh, Dead, Alcohol, Carnoy's
	Sudan (SENN)	71	Fresh, Dead, Alcohol, Carnoy's
	Burkina Faso	25	Dead, Alcohol
An. merus	Mozambique	22	Dead
	South Africa	20	Dead, Alcohol, Carnoy's
An. melas	Gambia	25	Dead, Alcohol, Carnoy's

An. quadriannulatus	Mozambique	5	Dead
	South Africa	20	Dead, Alcohol, Carnoy's
An, bwambae	Uganda	55	Dead,(10 pinned)
An. culicifacies	Iran		
	Zabol	35	Dead, pyrethrum spray, Alcohol
	Zahedan	32	Dead, pyrethrum spray, Alcohol
	Khash	30	Dead, pyrethrum spray, Alcohol
	Saravan	35	Dead, pyrethrum spray, Alcohol
	Iranshahr	39	Dead, pyrethrum spray, Alcohol
	Nikshahr	40	Dead, pyrethrum spray, Alcohol
	Chah-Bahar	45	Dead, pyrethrum spray, Alcohol
	Pakistan		
	Peshavar	47	Dead, pyrethrum spray
	Lahore	42	Dead, pyrethrum spray
	India (New Delhi)	5	Pinned
	Sri Lanka		
	Putalam	100	Carnoy's
	Kataragama	100	Carnoy's
An. stephensi	Iran		
	Kazeroun	25 + 6	Pinned & pyrethrum spray
	TEH	43	Fresh, Dead, Alcohol, Carnoy's
	BAN	51	Fresh, Dead, Alcohol, Carnoy's

An. stephensi	Iraq (IRAQ)	44	Fresh, Dead, Alcohol, Carnoy's
	Dubai		
	APR	41	Fresh, Dead, Alcohol, Carnoy's
	LPR	49	Fresh, Dead, Alcohol, Carnoy's
	India		
	BEECH	35	Fresh, Dead, Alcohol, Carnoy's
	IND-S	75	Fresh, Dead, Alcohol, Carnoy's
	Tadjikestan	10	Dead, Alcohol
	Pakistan		
	Peshawar	44	pyrethrum spray
	Lahore	39	pyrethrum spray
An. fluviatilis	Iran		
	Minab	15+1	Dead, Pinned
	Iranshahr	13	Dead
	Kahnouj	12	Dead
	Kazeroon	5	pyrethrum spray
An. subpictus	Sri Lanka		
	Inland (3 collection sites)	20-30/site	Dead, Alcohol
	Coastal	25	Dead, Alcohol
An. maculipennis	Iran		
	Anzali	15	Dead, Pinned
	Ramsar	15	Dead, Pinned

An. maculipennis	Iran		
	Babol	10	Dead, Pinned
	Sari	10	Dead, Pinned
	FSU (Teguldet, Tomsk, Siberia)	19	Alcohol
An. sacharovi	Iran (Sari)	10	Pinned
An. hyrcanus	Iran (Babol)	5	Pinned
An. pulcherrimus	Iran (Ahwaz)	5	Pinned
An. dthali	Iran (Kazeroun)	5	Pinned
An. superpictus	Iran (Kazeroun)	5	Pinned
An. claviger	Iran (Dash-E-Argan)	5	Pinned
An. turkhudi	Iran (Iranshahr, Kazeroun)	5 + 2	Pinned

An. melas : Gambia

An. quadriannulatis : South Africa

An. bwambae : Uganda

An. gambiae: **KWA**

Originating from Kwale, 35 km north of Tanga, Tanzania in East Africa. Pooled eggs were sent from the field by Dr. Francis Bushrod. This strain is susceptible to DDT, dieldrin, organo-phosphate and carbamate insecticides. It was colonised at the London School of Hygiene and Tropical Medicine in 1975, before being established by Rajni Vij in Liverpool in October 1991.

An. gambiae: **16CSS**

The original strain of *An. gambiae* s.s. was the LAGOS strain. Eggs were obtained from wild-caught adults from Lagos, Nigeria in West Africa. The 16CSS strain was derived from this strain in 1974. It is susceptible to all insecticides. Colonisation at the London School of Hygiene and Tropical Medicine occurred in 1951 and its maintenance in Liverpool was undertaken by Dr Rajni Vij in October 1991. This strain was selected as homozygous for collarless (c) by Mason (1967).

An. gambiae: **GAG**

Collected from Georgetown, The Gambia, established at SAIMR in 1980s.

An. gambiae: Savannah and Bissau

Dr. Virgilio Do Rosario from Center for Malaria and Tropical Diseases (CMTD), Universidade Nova de Lisboa, Lisbon, Portugal kindly provided these two strains (forms).

An. arabiensis: **SENN**

This strain was re-introduced to the London school of Hygiene and Tropical Medicine in 1986. It was derived from a colony kept at Sennar, Gezira area of Sudan. The colony was established in Liverpool by Dr Rajni Vij in October 1991.

An. arabiensis:**KGB**

Originally obtained from Harare, Zimbabwe. It was derived from material caught at Kanyamba, Zambesi Valley. This strain is susceptible to DDT and dieldrin. It was colonised at the London School of Hygiene and Tropical Medicine in 1975, and established by Dr Rajni Vij in Liverpool in October 1991.

An. arabiensis: **MA**

Collected from Maputo, Mozambique and established at SAIMR around 1992.

An. arabiensis: **BF**

Collected in Burkina Faso by Dr Beard in 1992

An. merus: **MAF**

Originating from saline springs near Shingwedzi in the Kruger Park. These were reared in 25% saline water and established at the South African Institute for Medical Research (SAIMR) by Drs. M. Coetzee and R. Hunt. Eggs were brought back to Liverpool by Rajni Vij in May 1992 and a colony was established.

An. merus: **ZAM**

Originally from Makanini's Drift in Zululand, South Africa collected by Dr. R. Hunt in 1978. These were reared at South African institute for Medical Research and established in Liverpool by Dr Rajni Vij in 1992 and again in 1994.

An. quadriannulatus: **SHINQUAD (QUAD)**

Caught at Shinwedzi near the Kruger Park in South Africa by Dr. M. Coetzee in 1992. The eggs were brought back to Liverpool where a colony was established by Dr Rajni Vij in 1992 and again in 1994.

An. bwambae:

Adult females were collected in January 1994 by Mr Ambrose Onapa by pyrethrum spray collection from houses in the village of Sempaya, some 300 m from the main breeding sites in the sulphurous, alkaline and highly saline Buranga Hot Springs, Bwamba County, Uganda.

An. culicifacies species complex:

Mosquitoes were either air-dried, alcohol or Carnoy's preserved. Pinned specimens (collected in 1978) of *An. culicifacies* provided by Department of Medical Entomology, Tehran School of Public Health, have been used as standard control for identification of unknown population of species A from Iran. Prof. M. Zaim, Tehran University of Medical Sciences, at present in World Health Organization (WHO/CTD), kindly sent *An. culicifacies* mixed samples from Galbagour, Angoury, Halpamcan (Bent), Zainnoddin in Baluchestan and Sistan province, Iran.

Other *An. culicifacies* samples from Baluchestan and Sistan province, Iran, provided by Mr Panjabi, Mr Hout Balaj and other staff in Medical Entomology Division, Health Department and Zahedan University of Medical Sciences, from Zabol, Zahedan, Saravan (Ibik), Khash (Hidouj, Chegar) and Nikshahr in 1994. Chahbahar samples collected from houses or shelter pits in May 1993 except samples from Bogan (1) which had been collected in 1986

2. Ouraky, Pir-sohrab, Dashte-yari
3. Kahr-hounan, Konarak
4. Safar Zehi, Pir-sohrab, Dashte-yari
5. Shahrestan, Plan, Dashte-yari
6. Chaker zehi, Plan, Dashte-yari
7. Dargaz, Bahouklat, Dashte-yari
8. Houtgate Paiin, Bahouklat, Dashte-yari

9. Alagdari Osmani, Negor, Dashte-yari

Iranshahr samples collected by Institute of Public Health Research, Iranshahr.

Air dried *An. culicifacies* species A from Peshawar and Lahore in Pakistan were received from a collection made by Dr. Mark Rowland and colleagues in 1990.

Sri Lanka samples were kindly provided by Dr. Abeyewickreme were preserved in Carnoy's solution and from Putalam, North-Western province,and Kataragama, Uva province.

An. culicifacies species B from India was air-dried and from Delhi. Fig. 3.1B (Brown circles) shows the collection sites of *An. culicifacies* in different countries.

An. stephensi:

Field samples collected by Ladonni from urban and rural area of Kazeroun (Khesht), Iran in 1983 were air dried and pinned. Pakistan specimens of *An.stephensi* have been collected by Dr. Rowland and colleagues in 1991 from Pinjab and Lahour.

Laboratory strains; Iran (BAN, TEH), Iraq (IRAQ), Dubai (LPR, APR), India (BEECH, IND-S) were established in Liverpool School of Tropical Medicine (LSTM) as follows:

TEH : a laboratory strain that is resistant to both DDT and dieldrin at the adult stage

and susceptible to all insecticides at the larval stage. The eggs of this strain provided by Tehran University of Medical Sciences and clonised at the LSTM since November 1992.

BAN : a wild strain from Bandar Abbass in Iran. It is resistant to DDT, dieldrin, and malathion at the adult stage and susceptible to all insecticides at the larval stage. The eggs of this strain provided by Tehran University of Medical Sciences and clonised at the LSTM since November 1992.

IRAQ : susceptible to pyrethoids. Obtained from London School of Tropical Medicine & Hygiene and maintained at the LSTM.

DUB-S : a wild strain based on larvae, collected from Dubai (U.A.E.), and colonised at LSTM in may 1986.

DUB-APR : a sub-strain from DUB-S, selected with permethrin at the adult stage in LSTM.

DUB-LPR : a strain with larval permethrin resistance, selected originally at LSTM from a sample collected in Dubai (U.A.E.) and colonised in 1986. Originally partly resistant but since 1992 this strain has been intermittently re-selected for permethrin resistance.

BEECH : a laboratory stock of Indian origin, from Beechham's laboratory.

IND-S : a strain susceptible to permethrin, and a standard laboratory colony originating from New Delhi in 1947.

Field materials of *An. stephensi* (established in lab) from Tadjikestan that has been used in this study probably originated from India because previous USSR is not included in zoogeographical distribution of *An. stephensi*. Fig. 3.1B (purple circles) shows the collection sites of *An. stephensi* in different countries.

An. subpictus:

Mosquitoes from coastal and Plains areas of Sri Lanka were provided by Dr. Abeyewickreme both air dried and alcohol preserved. Fig. 3.1B (blue circles) shows the collection sites of *An. subpictus* in Sri-Lanka.

An. fluviatilis:

Field samples of this species from Minab (Hormozgan province, Iranshahr (Bluchestan and Sistan province), Kahnouj (Kerman province) in East, South and South-East, Iran were sent by Prof. Zaim, Tehran University of Medical Sciences. One pinned *An. fluviatilis* also collected from Minab (Cheloo), 1984, by Yaqhubi has been used. Fig. 3.1B (green circles) shows the collection sites of *An. fluviatilis* in Iran (see also Fig. 6.2).

An. maculipennis

An. maculipennis collected from Anzali, Ramsar, Babol and Sari by Djadid in 1988-1989.

An. sacharovi from Sari by Djadid in 1989 .

Other Anophelines from Iran:

An. hyrcanus from Babol in 1987 by Yazdanpanah and Shahmohammadi.

An. pulcherrimus from Ahvaz in 1990.

An. dthali from Kazeroun in 1983 by Ladonni.

An. superpictus from Kazeroun in 1990.

An. claviger from Dashte Arjan in 1990.

An. turkhudi collected by Bashekouh from Iranshahr in 1982.

and also from Kazeroun (Kohmareh) in 17.6.1984 by Ladonni.

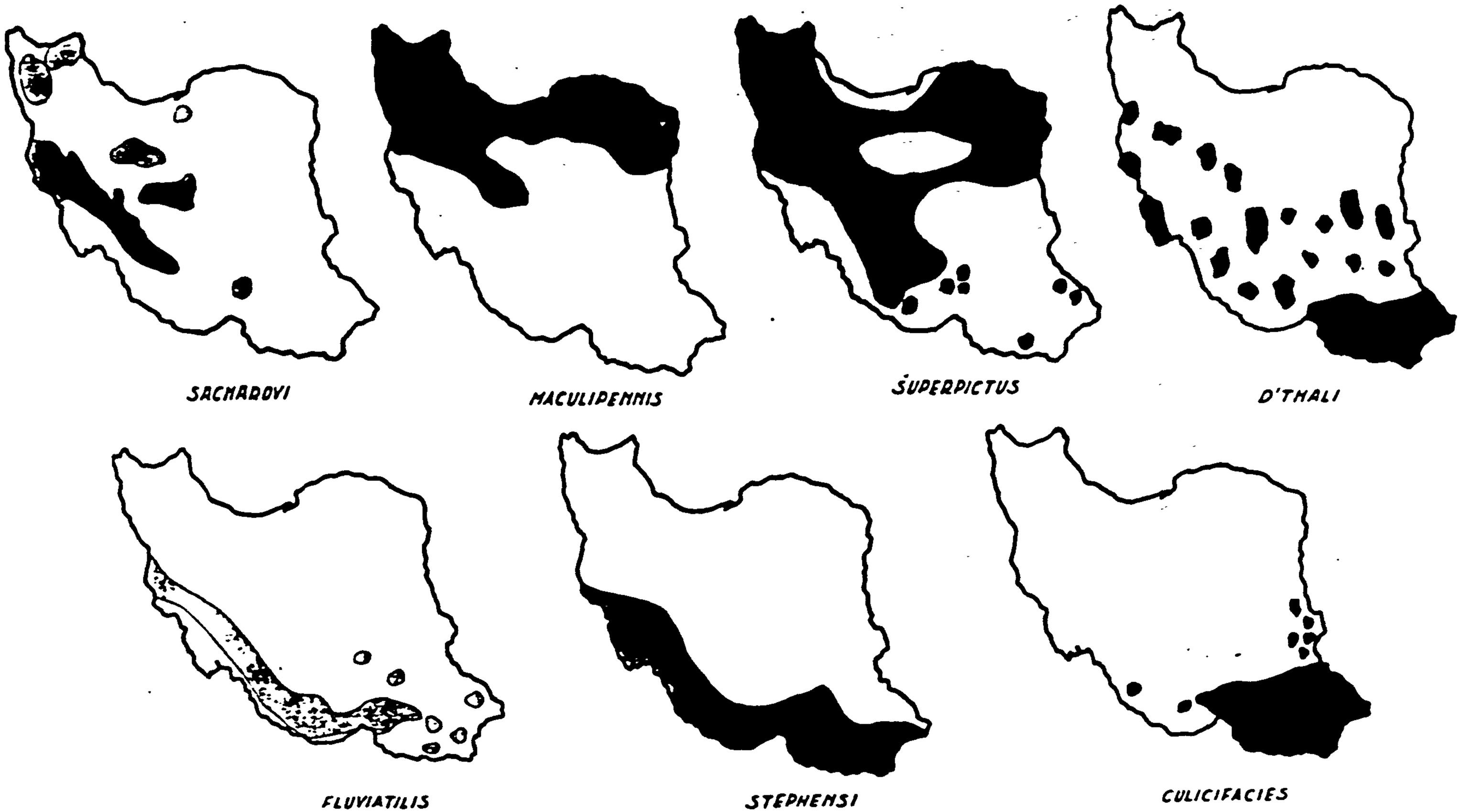
Fig. 3.1B (red circles) shows the collection sites of these An. maculipennis complex collection sites in northern Iran. Since *An. culicifacies, An. stephensi, An. fluviatilis* and *An. maculipennis* complex from Iran studied in this thesis, Fig. 3.2 shows the distribution of malaria vectors in Iran.

3.2.2 Capture techniques used in the field

Almost all field mosquitoes used in this study were in the adult stage, and were either indoor resting (Houses and animal sheds) or outdoor populations resting in natural or artificial outdoor resting shelters.

In houses, mosquitoes were caught from the walls (e.g. Northern Iran), ceilings, clothing (e.g. Bluchestan and Sistan), furniture (e.g. Kazeroun) including the underside of beds, using a mouth aspirator and a torch. Adults were also collected from hay and straw which has been stored in animal sheds.

Fig. 3.2- Distribution of malaria vectors in Iran



Another collection technique specially used in collecting *An. culicifacies* samples from houses or animal sheds was pyrethrum (usually 0.5-1%) spray sheet collection. In this the case of space spraying, mosquitoes fell on sheets in knock-down position. In order to prevent recovery of mosquitoes from the knock-down effect, 0.2% chloroform had been added to pyrethrum. Outdoor resting populations of mosquitoes in Iran have mainly collected from pit-shelters or by CDC light trap.

An. culicifacies and *An. stephensi* from Pakistan, have been collected by all night hourly catch (mouth aspirator in houses, under bed net, human bait or from inside experimental huts by aspirator or pyrethrum space spray collections.

3.2.3 Mosquito conditions and preservation techniques

Mosquitoes have been kept in different conditions and the following preservation techniques have been used and examined prior to DNA extraction.

A) Fresh materials

1 Freshly killed mosquitoes have been used directly in RAPD and conventional PCR, and dot blot.

2 Freshly killed but kept at room temperature, 4°C, or -20°C for different periods of time before extraction.

B) Dead

1 Dead, dried specimens collected from the bottom of the cages.

2 Dead, pyrethrum spray collection mosquitoes kept in tissue papers and sent to lab and

a) used straight or

b) left for few days in humid chamber before use.

3 Museum pinned samples from British Museum and Tehran University of Medical Sciences.

C) Preservation

1 Alcohol (absolute or 75%)

2 Carnoy's fixative

3 Modified Carnoy's fixative

4 Lactophenol (in larvae collection)

3.3 Mosquito Rearing

In this study, laboratory strains of *An. gambiae* species complex and *An. stephensi* have been reared, both to provide species samples and to examine intra-specific variation and in crossing experiments.

All the colonies were maintained in the insectaries at the Liverpool School of Tropical Medicine. The temperature and relative humidities were recorded from time to time, using a Mason's Wet and Dry Bulb thermometer. Temperatures were regulated by a thermostat and generally fluctuated from 27-28°C, a humidifier maintained relative humidities of 60-85%, depending on the insectary. A 12 hourly

day and night cycle was maintained by setting a time switch to the light circuit.

Mosquitoes were reared in deep trays that were 30 cm X 25 cm with a depth of 13 cm or shallower trays of 28 cm X 20 cm with a depth of 6.5 cm. Emerging mosquitoes were collected using a mouth aspirator from underneath nets that covered these trays; a procedure that was done gently to prevent damage to the fragile newly emerged adults. Allowing the adults to emerge from water of the rearing trays obviates the need for picking pupae. The deeper trays were only 1/4 filled with water to give the emerging mosquitoes enough flying room. However, fewer than 10 eggs were reared in the egg pots so as to limit loss during transfer into another tray, and the pots were placed in a cage at the first sighting of pupae, to allow the mosquitoes to emerge directly into their new environment.

An. gambiae GAG, *An. arabiensis* MAN, *An. arabiensis* ZULUVIJ, *An. quadriannulatus* SHINQUAD (QUAD), were all reared in distilled water, whereas *An. arabiensis* SENN, *An. arabiensis* KGB, *An. gambiae* 16CSS and *An. gambiae* KWA , *An. stephensi* TEH, BAN, IRQ, LPR, APR, BEECH, IND-S were reared in normal Liverpool tap water. At a later stage *An. arabiensis* SENN and *An. arabiensis* KGB were also found to do better in distilled water in that the larvae were extremely active in comparison to those in tap water, and the mortality amongst them was lower. *An. merus* MAF and *An. merus* ZAM eggs, larvae and pupae were kept in 25% saline water (standard BDH sodium chloride was used). On one occasion an

aeration pump bubbled air through the water, but there was no noticeable difference in larval survival rates.

The duration of rearing from eggs to emerging adults ranged from 7-12 days amongst all the species and strains though eclosion in *An. quadriannulatus* QUAD and *An. arabiensis* KGB took place three days after oviposition as opposed to one to two days for all the other species and strains. All mosquito larvae were fed on Tetramin fish food flakes, although application of tetramin inm a powder form was found to be more suitable than the flake form. It was important to add sufficient food for the larvae without causing the water to go scummy. Distilled water was squirted into the larval trays on a daily basis to aerate the water. The water was only changed if there were signs of cloudiness along with sluggish swimming behaviour of the larvae.

In order to avoid contamination between different species or strains, all larval trays were cleaned with hot water after each rearing cycle, when all the pupae had emerged into adults and before setting up a new batch of eggs, and then after drying, the trays were swabbed with cotton wool moistened with 70% ethanol.

Pipettes and egg pots were colour-coded for each sibling species. This method of larval rearing not only helps prevent cross contamination but also safe-guarded the larvae from microsporidian and flagellate protozoal infections. Originally, adult

mosquitos (except in cross experiments) were constantly provided with a wick of cotton wool placed in a universal tube containing saturated 10% sugar solution in tap water. Such cotton wool wicks required changing and resaturation with the sugar solution every two or three days due to fungal infections on the cotton and evaporation. The sugar feeding apparatus was placed in a square polystyrene block to prevent it from falling over in the cage.

Female mosquitoes were given guinea pig blood three times a week subject to requirement for eggs. This was done by placing an anaesthetised guinea pig over the top of the cage, usually between 1130 and 1300 hr, and allowing the mosquito to feed through the metal netting of the cage. occasionally the cage was blown to entice the mosquito to commence feeding by creating a carbon dioxide attractant.

On occasions when the number of females of any species in a cage was low it proved to be more successful to allow them to feed directly on a human hand placed in the cage, probably because by placing a hand into the cage the few mosquitoes are more likely to be aware of the presence of a host. Several attempts were made to membrane-feed adults with blood from various sources. Heparinised bovid blood, rabbit blood, and expired human blood (screened for HIV antibodies) were used. To provide a membrane through which the mosquitoes would be able to probe to obtain the blood, the skin of a five-day old chick was moistened with tap water and fitted over the open end of the glass apparatus using an elastic band. The

chamber was fitted with an inlet and outlet rubber tube, to allow water heated to 37°C to pass through one of the jackets, mimicking a warm-blooded host. The blood was removed directly from the blood transfusion bag with a 5 ml syringe and injected into the inner chamber of the membrane-feeding apparatus, and a rotating-paddle was also placed into this blood to keep it circulating. The membrane-feeding chamber was held in place on top of the cage with a boss and retort stand and left for 24 hr, after which the chamber and paddle were dismantled and immersed in a beaker containing a concentrated bleach solution for at least 24 hr before stringently cleaning them.

When attempts were made to establish *An. arabiensis* SENN and KGB, it was found that they failed to lay any eggs whilst maintained in the main insectary (28°C temperature and 77-85% relative humidity), but in another insectary (27-28 °C, and 55-60% humidity) with vitamin B supplement to the diet, the females oviposited three days after a blood-meal. This indicated that even though vitamin B played a crucial role for vitality of the mosquitoes, a lower humidity was still preferred for oviposition. *An. arabiensis* is a species which can be found in drier areas (e.g. Sudan, Ethiopia), and this may explain the failure of females to mate and lay eggs at higher humidities (77-85% RH) in the main insectary.

3.4 Enzyme Electrophoresis for Vector Identification

3.4.1 Starch gel

Starch gels were prepared with 11g of electro or potato starch in 125ml of TME pH 7.4 or TBE buffer. The mixture was heated until it became translucent, degassed before pouring and left for 16 hours at 4°C. However the gel could be used after setting (less than two hours). Single mosquitoes were homogenized by grinding in 10μl of tank buffer in a ground glass homogenizer. 4mm squares of cellulose acetate membrane were soaked in the homogenate and inserted into slots in the gel which made by a steel comb. Bromophenol blue was used as dye front. The gels were run at 150V, 90mA, at 4°C about 2-2.5 hours. Developing solutions were mixed with an equal volume of 1-2% agar and poured into a mould laid over the gel. By cutting each starch gel to two or more slices, it is possible to use more than one enzyme with each run (e.g. PGM for top and HEX for bottom layer).

3.4.2 Polyacrylamide Gel

Gel Preparation:

Reagent	Stacking Gel	Resolving Gel
0.13 M Tris pH 6.3	250 µl	-
0.375 M Tris pH 8.9	-	2.95 ml
50% Acrylamide	155 µl	5.31 ml
0.1% Triton X-100	2.5 µl	35 µl
Ammonium Persulphate 10%	12.5 µl	175 µl
TEMED	1.5 µl	20 µl
Distilled H_2O	2.1 ml	21.02 ml
Total	2.5 ml	29.51 ml
Gel Hight	~ ~ 1 cm	~ ~ 13 cm

Chamber Buffer:

0.09 M Tris = 10.9 g/l

0.08 M Boric Acid = 4.95 g/l

make 5 litres for tank buffer

Sample Preparation

1 Homogenise individual mosquitoes in 15-28 ul of trituration buffer:

0.13M Tris pH 6.3	9.9 ml
Stacking Gel Buffer	
1%Triton X-100	100 ul
10% (w/v) Sucrose	1 g

2 Centrifuge samples at 13,000 X rpm for 5 minutes.

3 Load 10-15 μl of supernatant onto gel.

4 Apply 10 μl of 1-2% bromophenol blue to one end slots as a dye front indicator.

Running the Gel:

1 Use constant current

i.e., 30 mA during the stacking gel adjust to 60 mA once the samples have entered the resolving gel.

2 Run until bromophenol blue reaches the bottom of the gel (~ 1.5 hours after entering resolving gel).

3.4.3 Detecting Enzyme Activity

Eight enzyme systems were used, the developing solutions for which are as follows:

Lactate dehydrogenase(LDH)

Developer solution:	
Lithium lactate	**300 mg**
NAD (APAD)	**7 mg**
MTT	**6 mg**
PMS	**1 mg**
0.1M Tris-HCl pH 8.5	**30 ml**

Octanol Dehydrogenase (Odh)

Developer solution:	
0.3 mM NAD	**20 mg**
0.5 mM MTT	**21.2 mg**
Tris 0.1 M pH 8	**1.21 mg**
Octanol	**1 ml**
Ethanol 0.25%	**0.25 ml**
Methanol 0.25%	**0.25 ml**
Distilled H_2O	**98.5 ml**

Place gel in stain in a dark box and incubate at 37°C in a shaking water bath. After 20 minutes add 0.16mM PMS (5mg) and incubate for an additional 20 minutes.

Destain & Fixing solutions (for 100ml) :

Glycerol 10%	10ml
Acetic acid 7.5%	7.5ml
Methanol 5%	5ml
ddH_2O	77.5ml

For a pernanent record, the gels were dried on filter paper using a slab-gel drier.

Octanol Dehydrogenase (Odh)+ Sod (Superoxide dismutase)

Developer solution:

0.3 mM NAD	25 mg
0.5 mM NBT	100ml
Octanol	0.5ml
Ethanol 0.25%	1ml

Incubate at 37°C in a shaking water bath for 2 hrs in day light. Then add 0.16mM PMS (5mg) and continue incubation in light until bands develop. Place the gel in distilled water to destain.

For making 0.05M Tris/HCl buffer pH 8.5:

solution A: Tris 0.2M (24.23 g/L)

solution B: HCl 1.0M (85.9ml conc. HCl/L ddH_2O)

To 250 ml soln. A, add 12.2 ml HCl (soln. B). This gives a pH of about 8.6.

Add a few more drops of HCl using a pH meter and make up to 1L.

SOD (Superoxide dismutase)

Developer solution:

NBT	5 mg in 1ml H_2O
0.05M Tris/HCl buffer pH 8	25ml
PMS	5mg in 1ml H_2O
Agar 2%	25ml

Expose to light for several minutes, then incubate at 37°C and the result is seen as pale zones on a dark background.

Esterase for Acrylamide gel

Developer solution:

Fast Blue RR	0.1g
a-Naphtyl acetate	0.125g
Acetone 50%	12ml
0.1M Tris/Maleate buffer pH 6.4	100ml

To make 0.1M Tris/Maleate buffer pH 6.4:

Solution A:	0.2M Tris	24.23 g/l
	0.2M Maleic Acid	23.2 g/l
Solution B:	1.0M NaOH	40g/l

Mix together 500 ml soln. A (250 ml each of Tris and Maleic acid) and 37.0 ml of soln. B. Adjust the pH to 6.4 with soln. B and make up to 1L. Dissolve the fast blue in the buffer. Dissolve a-Naphtyl acetate in 50% Acetone. Add to the fast blue, acetone mixture immediately before use. Addition of acetone causes solution to become cloudy yellow. Stand at room temperature in the dark with gentle shaking. Place the gel in 5% acetic acid destain overnight.

PGM

Developer solution:

$MgCl_2$	120mg
NADP	8mg
Glucose 1-phosphate	60mg
NBT	5 ml
Tris-HCl pH 7	5ml
PMS	~1mg
Agar 1%	10ml

Agar normally heated in microwave for about 2-3 minutes and then put in 50°C bath until use. Mix well with reaction solution before pouring on gel, and leave it in incubator for about 45 minutes to see the bands. Gels could be preserved in 10% acetic acid.

HEX (Hexokinase)

Developer solution:

$MgCl_2$	7mg
NADP	8mg
ATP	8mg
D-Glucose	30mg
NBT	5ml
Tris-HCl pH 8	5ml
Glucose 6-phosphate dehydrogenase	20μl
Agar	10ml
PMS	1mg

EST (Esterase)

Developer solution:

Fast Blue RR or Fast Red TR	50mg
0.15M phosphate buffer pH 6.5	60ml

Shake for about 1 hour in shaking bath (35-40°C) and then add 1ml of 2% 1-naphthyl acetate in acetone, inside the staining solution.

GOT (Glutamate Oxaloacetate Transaminase)

Developer solution:

a-Ketoglutaric acid	**60mg**
a-Aspartic acid	**100mg**
Fast Blue RR	**120mg**
0.2M Na_2HPO_4/HCl pH 7.3	**25ml**

Melt 25 ml already prepared frozen 2% Nobel agar. Dissolve the fast blue in a little buffer to make a paste. Add the ketoglutaric acid to the aspartic acid, then the rest of the buffer and just before use, add the agar, which has been cooled to 56°C. Pour on the gel as an overlay. Leave for exactly 45 minutes at 37°C in the dark. Remove the overlay immediately and submerge the gel in a 5% acetic acid destain.

To make 0.2M Na_2HPO_4/HCl pH 7.3; dissolve 28.39 g/l of Na_2HPO_4 in 800ml ddwater, and adjust the pH to 7.3 with conc. HCl and make up to 1L.

3.5 Polymerase Chain Reaction (PCR)

3.5.1 Genomic DNA extraction

The isolation of DNA from insects normally does not present any specific problem. Therefore, any of multiple techniques used for isolation of DNA from other organisms will usually work with insect tissue. For restriction digestion, it is necessary to remove most of the protein and RNA, although complete deproteinization is not required.

In conventional PCR, crude homogenates could be used without further purification, since the small amount of material added to the reaction dose not appear to affect the functioning of the Taq polymerase. However, for RAPD-PCR, like library construction, a further purification by phenol extraction provides more reproducible and consistent results. In total 14 different published extraction methods for isolation of DNA in a variety of organisms were tested.

Most of these methods apply an ethanol precipitation step and although there are some reports (Micheli *et al.* 1994) that ethanol precipitable contaminants in DNA extracted by standard techniques are a major, if not sole, source of variability in RAPD, this problem did not arise in such experiments. Here described are some of those methods and their modifications that have been used more frequently in this study based on reproducibility, simplicity, speed and safety.

Phenol extraction (Ballinger-Crabtree *et al.* (1993)

1 Grind mosquito to powder in liquid nitrogen with microfuge pellet pestle grinders, or directly homogenize the mosquito in lysis buffer with a glass rod (step 2).

2 resuspend in 305 μl of lysis buffer :

100 mM	Tris-HCl, pH 8.0
50 mM	NaCl
50 mM	EDTA, pH 8.0
0.15 mM	Spermine
0.5 mM	Spermidine
10%	SDS
20 mg/ml	Proteinase K (P.K.)

In order to make 10 ml of lysis buffer :

0.5 M	Tris-HCl, pH 8.0	2 ml
0.5 M	NaCl	1 ml
0.5 M	EDTA, pH 8.0	0.2 ml
15 mM	Spermine	0.1 ml
50 mM	Spermidine	0.1 ml
ddH2O		6.5 ml

Make the mix and keep in freezer. For using add

10%	SDS	30 μl
20 mg/ml	P.K.	5 μl

To each tube separately. So 305 μl is the final volume of lysis buffer in each

tube.

3 Suspensions incubate overnight at 50°C (or 90°C for 30 minutes, 65°C for 1 hour.

4 Buffered Phenol extraction : suspensions gently extracted twice with buffered phenol, pH 8.0.

- add 305 μl buffered phenol, pH 8.0
- mix very carefully on wheel = 20 rpm
- centrifuge 5 mins at 13000 rpm
- transfer the upper aqueous layer to a fresh tube
- repeat above extraction

5 DNA precipitation

- add 0.2 volumes of 10M Ammonium Acetate and 2.0 volumes of ethanol
- solution gently mixed on wheel
- centrifuge for 5 mins to pellet the DNA (preferably 10 mins at 13000 rpm)
- remove the supernatant
- wash the pellet with 200 μl of 70% ethanol

6 Air-dry the pellet for at least 1 hr or vacuum-dry for 5 min

7 Resuspend the pellet in 20 μl ddH_2O

8 Store DNA at 4°C

DNA extraction using boiling method

1 Mosquitoes homogenized over liquid nitrogen

2 Add 45 μl ddoH_2O

3 Boil for 15 mins.

4 Add further 20 μl ddH_2O

5 Centrifuge for 5 mins.

6 Supernatants store at 4°C

DNA extraction for rDNA PCR as described by Collins et al. (1987)

1 A single mosquito was homogenized in a 0.5 ml eppendorf tube using a glass rod tapered at the end, in 50 ul of buffer:

NaCl	0.08 M
Sucrose	0.16 M
EDTA	0.06 M
SDS	0.5%
Tris-HCl pH 8.6	0.1 M

For 1ml Lysis Buffer:

0.5 M	EDTA	120 μl
0.5 M	Tris-HCl	200 μl
0.5 M	Sucrose	320 μl
0.5 M	NaCl	160 μl

10%	SDS	50 μl
dd	H_2O	150 μl

add 6.25 μl of 8M potassium acetate to give 1M final conc.

2 Incubate at 65°C for 30 mins.

3 Add 6.25 μl of 8M potasium acetate to give 1M final concentration.

4 Incubate on ice for 30 mins.

5 Centrifuge 10 mins.

6 Supernatant removed to a fresh tube and 100 μl of 95% alcohol was added.

7 Centrifuge again for 10 mins.

8 Discard supernatant and wash pellet twice with cold 70% ethanol.

9 Pellet was air-dried and resuspended in TE or ddH_2O

Miniprep DNA extraction as described by

Medina-Acosta (1993)

1 Homogenize a single mosquito over liquid N_2

2 Add 150 μl lysis buffer (TELT) :

Tris-HCl pH 8	50 mM
EDTA pH 9	62.5 mM
LiCl	2.5 M
Triton X-100	4%

3 Completely suspend homogenate by sharply inverting tube 3 times

4 Leave at room temperature for 5 mins.

5 Add 150 μl of phenol-Chloroform (1:1 v/v) and mix gently for 5 mins.

6 Centrifuge for 5 mins.

7 Remove nucleic acid layer to a fresh tube.

8 Add 300 μl ethanol.

9 Gently stir for 15 secs and leave for 5 mins.

10 Collect nucleic acids by centrifugation for 10 mins.

11 Wash pellet once with 200 μl ethanol 70%.12

Leave to dry and resuspend in ddH_2O.

DNA extraction as described by Boyce *et al.* (1989)

1 Samples homogenised over liquid nitrogen.

2 Add 500 μl of 2X CTAB buffer:

0.1 M	Tris-HCl
0.02 M	EDTA
1.4 M	NaCl
0.2%	B-Mercaptoethanol
0.05	Hexadocylthriethylammonium Bromide pH 8.3

3 Homogenate gently suspended in buffer.

4 Incubate at 65°C for 60 mins.

5 Centrifuge for 15 sec.

6 Supernatant decanted and saved.

7 Pellet resuspended in 300 μl fresh CTAB.

8 Gently rehomogenized and incubated at 65°C for 10 mins.

9 This supernatant was pooled with the first and

10 Extract supernatant once with an equal volume of chloroform isoamyl alcohol (24:1).

11 Precipitated with ice cold isopropanol.

12 Air dried and resuspended in TE.

When DNA produces a smear or nothing in a PCR reaction, RNase treatment of DNA is necessary as follows:

1 0.5 μl RNase (10 mg/ml) ie, 5 μg added to each 20 μl DNA sample.

2 Incubate at 37°C for 1 hr.

3 1:1 Phenol / chloroform (24:1 isoamyl) and

4 chloroform (24:1 isoamyl) extraction

5 precipitation with 0.2 vol. 10 M Am.Act. and 2 vol. ethanol.

or after that steps 1 and 2 ;

1 Sample volume made up to 50 μl.

2 50 μl of 1:1 Phenol:Chloroform (24:1 isoamyl) added.

3 Mixed on wheel for 15 mins.

4 Centrifuge for 10 mins.

5 Equal volume of Chloroform:isoamyl added to recovered aqueous layer.

6 mixed on wheel for 20 mins.

7 centrifuged for 15 mins.

8 Add 2X volume Etoh and 0.2 volume Am Act. to aqueous layer.

9 Centrifuge until a good pellet was observed.

10 Wash with 70% alcohol.

11 Pellet left to air dry and resuspended in less volume of ddH_2O.

Based on our experience of sensitivity of RAPD to DNA quality, we applied and compared the modification of the following extraction methods; phenol extraction, Collins *et al.*, Mini-Prep or their combination. Air-dried pellets have been resuspended in 100 μl of double distilled H_2O or TE buffer and stored at 4°C.

Rapid isolation and purification of DNA for restriction digestion as described by Cockburn and Fritz (1996)

1- Put 1 mosquito into a microfuge tube with 100 μl of homogenization buffer and 25 μl of lysis buffer. adjust volume up or down for larger or smaller quantities or organisms but keep the total volume under 500 μl for a 1.5 ml tube.

Homogenization buffer: 0.1M NaCl, 0.2M sucrose, 0.01M EDTA, 0.03M Tris-HCl, pH 8.0.

Lysis buffer: 0.25M EDTA, 2.5% (w/v) SDS, 0.5M Tris-HCl, pH 9.2

2- Grind mosquitoes thoroughly because it is a critical step in extraction, also different stages and species have different contaminants which may or may not interfere with subsequent use of the DNA.

3- Incubate at 65℃ for 30 min, although the exact temperature is not important. However, it should provide appropriate temperature allowing the detergent to thoroughly solublize the protein.

4- Add 17 μl of 8M potassium acetate, vortex, spin briefly, and put on ice for at least 30 min (overnight in refrigerator is best). Using high K^+ concentration precipitates properly the detergent/protein complex.

5- Spin in a refrigerated microfuge at maximum speed for 15 min and transfer supernatant to a clean tube. Care should be taken not to disturb the rather loose pellet of protein/SDS/cuticle. It is always better to leave a little of the supernatant to get SDS into the DNA preparation.

6- Add 400 ul of 95% ethanol; spin down the nucleic acids in microfuge for 5 min at 12000g. A large amount of white flocculent precipitate should be observed in this step. More DNA can be recovered by ethanol precipitation at 4oC than other temperatures.

7- Pour off the supernatant and refill the tube with 70% ethanol. Vortex until the pellet comes off of the side of the tube. Spin for 10s. Repeat once more to remove residual SDS, EDTA, and salts from the nucleic acid pellet.

8- Air dry the pellet and dissolve in 85 ul of TE; add RNase A to 1 ug/ml. RNA

is the most likely contaminants to interfere with DNA modifying enzymes.

3.5.2 Reamplification of PCR products

DNA extraction from the gel:

The simplest way is

1 run the PCR products on a low melting agarose gel,

2 choose the fragment of interest and excise from the gel and

3 melt agarose containing the band at 65-90°C.

At this stage, 1-2 ul of this melted agarose (including DNA) can be added to the PCR mix prepared fresh in the same time. I had successful reamplification using this procedure in RAPD and conventional PCR. However, if more pure DNA is needed for reamplification or probing the next steps should be followed:

4 extract DNA from gel two times with phenol-chloroform-isoamyl acetate and chloroform-isoamyl acetate.

5 precipitate with absolute ethanol.

6 wash the pellet with 75% ethanol.

7 air dry and store in 4°C for reamplification or probing

Rapid recovery of DNA from agarose gels

In one method, (Innis *et al.* 1990), the agarose is frozen in liquid nitrogen or put in -20°C and then ultracentrifuged. The main benefit of freezing the agarose is to disrupt its structure, allowing the DNA greater opportunity to diffuse out.

But in the rapid recovery protocol, described by Vaux (1992) scissors were used to cut the tip from a 0.5 ml PCR tube and five holes were inserted in the end using a 25 gauge needle. The same needles was used to insert a single hole into a second 0.5 ul tube that had had its lid removed. A 2 mm X 2mm piece torn from a siliconized glass filter paper disc was pushed to the base of the second tube using a yellow tip, and the tubes were assembled on top of each other into a 1.5 ml screwtop microfuge tube.

When centrifuged for 5-10 min, an agarose gel slice in the top tube is disrupted as it is forced into the second tube. Here, the glass paper blocks passage of the agarose but lets the DNA pass into the lower tube. A UV light can be used to monitor progress of ethidium bromide-stained DNA. Recovery is around 75-85%, but yield depends on fragment size and amount of agarose present.

A modified procedure also has been examined as follows:

Bands excised from agarose gels were placed in a previously prepared 500μl Eppendorf tube. The bottom of the tube had been pierced with a 25swg needle and the hole overlaid with a plug of sterile glass wool. The 500 μl Eppendorf containing

the agarose plug was placed in a 1.5 ml Eppendorf and spun (5 sec) in a microfuge. The liquid collected from the 1.5 ml tube was used directly in the PCR or labelled by random priming without further purification for use as a DNA probe.

3.6 RAPD-PCR

3.6.1 Optimization:

At the begining of this study there were only a few paper published on RAPD-PCR and only one in mosquitoes (Ballinger-Crabtree et al. 1992) population. On the other hand no individual protocol will be optimal for all PCR reactions, nor will any single, simple set of variables to be optimized necessarily produce a functioning protocol for a specific case. Therefore, optimization is the first step to understanding mechanistically the relevance of most of the underlying physico-chemical principles in a specific PCR.

3.6.2 DNA concentration and quality:

DNA concentration in ng/μl is determined spectrophotometrically (assuming that 1 A_{260} corresponds to 50 μg/ml for double-stranded DNA. Total genomic DNA extracted from a single mosquito using this technique showed about 0.7- 0.9 ug.

DNA concentration and quality can also be estimated by running the DNA or PCR product alongside a molecular weight marker. The optimal amount of DNA for RAPD is 5pg to 10ng per reaction, but routinely 5ng was used. However, less DNA

produce more reliable pattern because of more diluted inhibitor or contaminant whatever their nature is.

3.6.3 $MgCl_2$:

Mg^{2+} is acting as a co-factor in the amplification reaction and its low and high concentration change the pattern and number of bands amplified and it is relevant to both the specificity and yield (Oste 1989). Higher concentrations of Mg^{2+} stabilize double-stranded DNA and prevent complete denaturation of the product at each cycle, reducing yield. Excess Mg^{2+} could also stabilize spurious annealing of primer to incorrect template sites, resulting in larger amounts of undesired products and lower specificity. Some Mg^{2+} ions will also be chelated by the dNTPs in the reaction mixture. In order to optimize Mg^{2+} concentration, 5, 2.5, 2, 1 and 1.5 mM were used in RAPD-PCR amplification mix and 2mM found to be the optimal concentration. In specific PCR, 2 or 2.5 mM $MgCl_2$ has been used.

3.6.4 Primer :

Primers used in this study have been listed in table 3.2. They were commercially available and purchased from Operon (OPA series), California, USA; University of British Columbia (UBC series), Vancouver, Canada; or synthesized to order by OLIGO Express Limited, UK. RAPD primers were mostly 10 mers but I also used universal primers such as M13 sequencing primers (17 mer) and some gene specific rDNA primers (UN from rDNA IGS). The G + C content of RAPD primers

Table 3.2- List of primers and their sequences used in this study

primer	5....sequence ....3'	GC%
AB1-01	GTTTCGCTCC	60
AB1-02	TGATCCCTGG	60
AB1-03	CATCCCCCTG	70
AB1-04	GGACTGGAGT	60
AB1-05	TGCGCCCTTC	70
AB1-06	TGCTCTGCCC	70
AB1-07	GGTGACGCAG	70
AB1-08	GTCCACACGG	70
AB1-09	TGGGGGACTC	70
AB1-10	CTGCTGGGAC	70
AB1-11	GTAGACCCGT	60
AB1-12	CCTTGACGCA	60
AB1-13	TTCCCCCGCT	70
AB1-14	TCCGCTCTGG	70
AB1-15	GGAGGGTGTT	60
AB1-16	TTTGCCCGGA	60
AB1-17	AGGGAACGAG	60
AB1-18	CCACAGCAGT	60
AB1-19	ACCCCCGAAG	70
AB1-20	GGACCCTTAC	60
UBC 301	CGGTGGCGAA	70
UBC 302	CGGCCCACGT	80
UBC 303	GCGGGAGACC	80
UBC 304	AGTCCTCGCC	70
UBC 305	GCTGGTACCC	70
UBC 306	GTCCTCGTAG	60
UBC 307	CGCATTTGCA	50
UBC 308	AGCGGCTAGG	70
UBC 309	ACATCCTGCG	60
UBC 310	GAGCCAGAAG	60
UBC 311	GGTAACCGTA	50
UBC 312	ACGGCGTCAC	70
UBC 313	ACGGCAGTGG	70
UBC 314	ACTTCCTCCA	50
UBC 315	GGTCTCCTAG	60
UBC 316	CCTCACCTGT	60
UBC 317	CTAGGGGCTG	70
UBC 318	CGGAGAGCGA	70
UBC 319	GTGGCCGCGC	90
UBC 320	CCGGCATAGA	60
UBC 321	ATCTAGGCAC	50
UBC 322	GCCGCTACTA	60
UBC 323	GACATCTCGC	60
UBC 324	ACAGGGAACG	60
UBC 325	TCTAAGCTCG	50
UBC 326	CGGATCTCTA	50
UBC 327	ATACGGCGTC	60
UBC 328	ATGGCCTTAC	50

Table 3.2- List of primers and their sequences used in this study

primer	5....sequence ....3'	GC%
UBC 329	GCGAACCTCC	70
UBC 330	GGTGGTTTCC	60
UBC 331	GCCTAGTCAC	60
UBC 332	AACGCGTAGA	50
UBC 333	GAATGCGACG	60
UBC 334	ATGGCAAAGC	50
UBC 335	TGGACCACCC	70
UBC 336	GCCACGGAGA	70
UBC 337	TCCCGAACCG	70
UBC 338	CTGTGGCGGT	70
UBC 339	CTCACTTGGG	60
UBC 340	GAGAGGCACC	70
UBC 341	CTGGGGCCGT	80
UBC 342	GAGATCCCTC	60
UBC 343	TGTTAGGCTC	50
UBC 344	TGTTAGGCAC	50
UBC 345	GCGTGACCCG	80
UBC 346	TAGGCGAACG	60
UBC 347	TTGCTTGGCG	60
UBC 348	CACGGCTGCG	80
UBC 349	GGAGCCCCCT	80
UBC 350	TGACGCGCTC	70
UBC 400	GCCCTGATAT	50
M13R/PUC	CAGGAAACAGCTATGAC	53
M13F	GTAAAACGACGGCCAGT	70
OPA1	CAGCCCCTTC	60
OPA4	AATCGGGCTG	60
OPA8	GTGACGTAGG	50
GT1-18	$(GT)_7ATCC$	44
GT2-18	$(GT)_7TGTA$	42
AA-19	$(AGT)_5GCCA$	50
(1)G	$(GT)_8$	47
(2)G	$T(GT)_8$	25
(3)G	$(GATA)_5$	67
(4)G	$(GTG)_5$	50
(5)G	$(GGAT)_4$	60
5.8S	ATCACTCGGCTCGTGGATCG	39
28S	ATGCTTAAATTTAGGGGGTAGTC	40
QD	CAGACCAAGATCGTTAGTAT	55
ME	TGACCAACCCACTCCCTTGA	60
UN	GTGTGCCCCTTCCTCGATGT	55
GA	CTGGTTTGGTCGGCACGTTT	45
AR	AAGTGTCCTTCTCCATCCTA	

varied between 50-90%.

RAPD primers were usually used in the range of 10-100 ng and universal and specific primers in 60-150 ng. However, primer excess can lead to formation of primer dimers and to excessive mispriming. Thus, for very diluted template DNA, even lower primer concentrations may give the best results.

On arrival, primers were aliquoted to small amounts and diluted to working concentration in ddH_2O or TE buffer. 20 μl of undiluted primer were kept in the freezer in case of deterioration of diluted primers. The working tube is stored frozen and then thawed for use, and re-frozen after use, until evident deterioration is seen as judged by poor results.

3.6.5 Taq Polymerase

Different thermostable DNA polymerases may amplify different RAPD products (Schierwater & Ender 1993), perhaps because both the activity and specificity of different polymerases depend on slightly different temperature and reaction conditions, which affect the outcome of possible competition reactions between the products amplified in the first and most critical cycles. In this study usually 0.5 unit per 25 μL reaction were used. Taq (Thermus aquaticus) polymerase was purchased from Promega.

3.6.6 Annealing temperature (T_a)

T_a, the temperature at which one-half of the primers are annealed to their target sequence, is roughly calculated for oligonucleotides 20 bases long or less by following the equation

$$T_a = 4(G + C) + 2(A + T)$$

where A, T, G, and C are the numbers of those bases in the oligonucleotide. The T_m of primers was calculated using PRIMER program.

In RAPD the basic annealing temperature was 36°C and in its modification 39, 40, 50 °C were used depending on the product expected and the number of bands.

In specific PCR, if the T_a is too low, non-specific DNA fragments are amplified, causing the appearance of multiple bands on agarose gels. If the T_a is too high, the yield of the desired product, and sometimes the purity is reduced due to poor annealing of primers. T_a optimal (T_a^{OPT}) calculated using the empirical formulation:

$$T_a^{OPT} = 0.3\ T_m^{primer} + 0.7\ T_m^{product} - 14.9$$

In which T_m^{primer} is the calculated T_m of the less stable primer-template pair and $T_m^{product}$ is the T_m of the PCR product (Rychlik et al. 1990).

3.6.7 Deoxynucleotides dNTPs

For RAPD 0.1 mM of each deoxynucleotide has been used. It is very important that the concentration of all deoxynucleotides are equal to prevent

misincorporation errors.

3.6.8 Enzyme Stabilizers

Some manufactures include gelatin or triton X-100 in the enzyme storage buffer. In order to stabilize the enzyme during thermal cycling, each reaction buffer contained gelatin at a final concentration of 0.01% (w/v) or Triton X-100 at a final concentration of 0.1% (v/v).

3.6.9 Amplification profile

In specific PCR, 30 cycles of amplification is recommended. However, usually 45 cycles of the three steps (denaturation, annealing, and extension) are run to obtain a RAPD pattern (Program 1). This program can last as long as 5 hours depending on the thermal cycler that is used. On the other hand a lengthy PCR program can be a limitation when large number of samples are analyzed. To overcome this problem fourteen programs were examined to determine the effect of :

1 Different annealing temperature (36, 42, 47, 50, 60 °C),

2 Variable numbers of cycles, and

3 Two and three step program.

Based on these data, I designed two other programs which seemed more useful in RAPD. Program 2 has two step annealing; 15 cycles at 50°C and 20 cycles at 36°C. The second, program (7), contain three different annealing (36, 45, 50) each step in 10 cycles.

Program 1

94	5mins.	1 cycle
94	1min.	
36	1min.	45 cycles
72	2mins.	
72	7mins	1 cycle

Program 2

94	5mins.	1 cycle
94	30 sec.	
50	30 sec.	15 cycles
72	30 sec.	
94	30 sec.	
36	30 sec.	20 cycles
72	30 sec.	
72	7 min.	1 cycle

Program 3

94	5mins.	1 cycle
94	30 sec	
45	1 min	10 cycles
72	30 sec	
94	30 sec	
36	30 sec	10 cycles
72	30 sec	

Program 7

94	5mins.	1 cycle
94	30 sec	
45	1 min	10 cycles
72	30 sec	
94	30 sec	
39	30 sec	10 cycles
72	30 sec	
94	30 sec	
35	30 sec	10 cycles
72	30 sec	
72	7 min	1 cycle

Prog 9:

94	30 sec	
41	30 sec.	30 cycles
72	30 sec	

Prog 10:

94	30 sec
50	1 min 30 cycles
72	30 sec

In **Prog 4**, 50°C has been used in 30 cycles

and **Prog 5**, 50°C has been used in 45 cycles

Program 6 has two 39 and then 50 °C annealing.

Prog 8, 45°C has been used in 40 cycles

3.6.10 DNA amplification

All RAPD reactions were performed in a total volume of 25 μl. Each mixture contained 2.5 μl of 10X reaction buffer, 2 mM $MgCl_2$, 10-100 ng of primer, 0.001% gelatin, 0.1 mM each of dATP, dTTP, dCTP, and dGTP, 0.5 unit of promega Taq polymerase, and sterile double-distilled water to 25 μl. Reactions were overlaid with 50 ul of mineral oil and amplified in a Hybaid Omnigene thermal cycler.

3.6.11 PCR products analysis by gel electrophoresis :

Briefly, following PCR, 10 μl of amplified DNA was mixed with Ficoll/orange G loading buffer and electrophoresed in polyacrylamide or agarose gel in TBE buffer containing ethidium bromide at 1.5 V/cm for 16 hours (large gel) or 5 V/cm for two hours (mini gel).

Agarose gel electrophoresis

Buffer and Agarose Gel Preparation

1X TBE :

0.089 M	Tris (121.14 MW)	10.78 g
0.089 M	Boric Acid (61.83 MW)	5.5 g
0.002 M	EDTA (292.2 MW)	0.584 g
	ddH2O	1 Lit.

Working concentration = 0.5X

Preparation of 500 mls of 0.5 X TBE Buffer needs

Tris	2.7 g
Boric Acid	1.4 g
EDTA	0.15 g
Ethidium Bromide	10 ul

Take 30 ml of the buffer prepared and add 0.36 g agarose (i.e. 1.2% agarose gel). Bring to the boil on a magnetic stirrer and allow to cool to 50°C before pouring. Large gels need 300 ml of buffer and 3.6 g agarose. 2500 ml buffer also is required for tank (large gels) and 350 ml (mini gel).

Loading buffer (20% w/v Ficoll; 100 mM EDTA; Orange G) was added to the PCR products (20% v/v) and loaded onto agarose gels in a 0.5x TBE with 200 μg/l of ethidium bromide. For analysis of RAPD products 5-15 ul of PCR product

was loaded onto a 1.2%-2.5% (based on the size of products) agarose (Appligen) gel at 1.5 V/cm for 16 hours (large gel) or 5 V/cm for two hours (mini gel). A mixture of BglI and HinfII digest of pBR328 (Boehringer Mannheim) were used as molecular weight marker for PCR product size determination (molecular weights 2176, 1766, 1230, 1033, 653, 517, 453, 394, 298, 298, 234, 234, 220, 154, 154 bp). Gels visualised under UV light were photographed with black and white Polaroid 55 film or Ilford FP4 film.

Polyacrylamide gel for PCR products

-Minigel system :

- 400-500 ml TBE buffer for tank

per gel : 6% acrylamide (made in TBE) 5 ml

10% APS 40 μl

TEMED 4 μl

6-8% polyacrylamide gels were prepared by diluting 30% 29:1 acrylamide: bisacrylamide stock solution (Severns Biochemicals) with 0.5xTBE (0.045M Tris-borate; 0.001M EDTA pH 8.2) and adding 10% ammonium persulfate (10μl ml^{-1}) and 1μl ml^{-1} of TEMED. PCR products were mixed 1:1 with loading buffer (20% Ficoll; 100mM EDTA; Orange G) and 5 μl of the mixture was loaded onto the gel.

Gels were run in the 0.5 or 1x TBE under the conditions indicated in the figures and silver stained by the method of Bassam et al. (1991). Gels were fixed in 10% acetic acid for twenty minutes, washed 3x3 minutes in ddH_2O, incubated with shaking in 0.15% $AgNO_3$, 0.15% formalin for 30 minutes, rinsed 2x5 seconds in ddH_2O, developed in 3% $NaCO_3.5H_2O$ and fixed in 10% acetic acid using sufficient volumes to cover the gel.

3-6-12 Preparation of gel for sscp products

29.4 g acrylamide

0.6 g bis 30% solution

100 mls ddH_2O

filter in a 0.4 μm filter

5% acrylamide gel used (diluted in 1x TEB). For example:

3 mls acrylamide + 15 ml TBE + 160 μl 10% APS + 20 μl TEMED

Preparation of samples

- 5 μl PCR product mixed with 10 μl Loading buffer (10 mM NaOH + 95% formamide + 0.05% B.B. + 0.05% Xylene Cyanol)

- Heated at 95 ℃ 2 min, plunged into liquid nitrogen, and thawed on ice

- Load about 5 μl onto gel, and run at 20 V/Cm for 2-2.5 hrs, using 0.5x TBE in both upper and lower chambers.

3-6-13 Silver stain

10% Acetic acetic		20 min.
ddH_2O	3x	2 min.
0.15% $AgNO_2$ + 0.15% formalin		30 min
ddH_2O		10 sec
3% NaCo3 (4 °C) + 0.15% formalin + 0.0002% sodium thiosulphate		until bands appear
10% Acetic Acid		5-10 min.

3-7 Southern Blotting

Gels for bloting were placed in denaturing solution (0.5M NaOH, 1.5M NaCl) for one hour with shaking and then changed to neutralising solution (NaCl 3M, Tris-HCl pH 5.6) for 2x30 minutes. The blot consisted of a pad of paper towels overlaid with two sheets of Whatmans 3MM and soaked with 20xSSC (Standard Saline Citrate- 1xSSC is 0.15M NaCl, 0.015M trisodium citrate), the gel was placed on this and a nylon membrane (Amersham N+) soaked in 2xSSC was laid on top. The membrane was overlaid with two sheets of Whatmans 3mM soaked in 2xSSC and one dry sheet, on top was placed a pad of dry paper towels and the whole stack was

weighted with approximately 1 Kg distributed over a glass plate. The blot was left for 16 hours at room temperature. The membrane was removed from the blot and baked at 80°C for two hours.

3-8 Dot Blots

Samples were diluted to the appropriate concentration in 6xSSC and applied to a Hybond N+ (Amersham) nylon filter through a Minifold (Schleicher & Schnell) apparatus. The filter was then transferred successively to Whatmans 3MM paper soaked with:

1) denaturing solution (0.5M NaOH, 1.5M NaCl) 5 min.

2) neutralising solution (3M NaCl, Tris-HCl pH5.6) 5 min.

3) 2xSSC 5 min. The filter was air dried and baked at 80°C for two hours.

3-9- Labelling Probes (DNA & Primers)

DNA probes were labelled with [a-^{32}P]dCTP to $>10^9$cpm by the random priming method using the Boehringer random priming kit. PCR products that were to be used as probes were separated from dNTPs by passing through a S400 spin column (Pharmacia), 20-50 ng of DNA in a 11 μl volume was boiled for ten minutes then cooled on ice. After a brief centrifugation 2 μl of random hexanucleotide mix, 5 mM each of dATP, dTTP and dGTP, 20μCi dCTP^{32}, and 2 units of Klenow enzyme were added. The mixture was briefly centrifuged and incubated for 30 minutes at 37°C.

Oligonucleotides were labelled using a terminal transferase kit (Boehringer) according to the directions of the manufacturer. Labelled probes were separated from free nucleotides by passage through a Sephedex G50 column. The activity of fractions collected from the column was measured in a scintillation counter (LKB). Fractions corresponding to the first peak of radioactivity were pooled, boiled for 5 min and cooled on ice.

3-10- Probing Southern Blots

Filters from Southern blots were soaked in 2xSSC for 10 minutes (RT) and prehybridised in BEPS hybridisation buffer (0.2ml/cm^2) (1% BSA, 1mM EDTA, 0.5M Na_2HPO_4 pH 7.4, 7%(w/v) SDS) for 16 hours at the temperature indicated in the figures. The labelled probe was added and the filter incubated for at least four hours. The filter was then washed in 3xSSC, 0.1% SDS for 20 minutes at RT. The stringency of washes was increased by reducing the salt concentration (2X and 1X SSC) and raising the temperature.

After probing, radioactive filters were autoradiographed on pre-flashed X-ray film (Fuji). The films were exposed for a suitable period at -70°C and developed for 2 minutes in Phenisol high contrast developer (Ilford), stopped for 1 minute in 5% acetic acid, fixed for 1 minute in fixer (Ilford), washed in water and air dried.

Filters that were to be reprobed were stripped in 0.1M NaOH (15min), rinsed briefly in water and neutralised in 0.5M Tris-HCl pH 7.0 (20 min) and air dried.

3-11 Cloning

Cloning of PCR products was carried out by using the Invitrogen original TA cloning kit in which the inserted PCR product is flanked on each side by EcoR I sites. Clones were sequenced by a AB1 PRISM 377 DNA sequencer (cycle sequencing reaction) using AMPLI TAQ FS and manufacture's guideline with M13 forward and reverse primers. In manual sequencing (chain termination reaction), sequenase version 2.0 T7 DNA Polymerase (USB) was used.

The original TA cloning kit with pCR™II provides a quick, one-step cloning strategy for the direct insertion of a polymerase chain reaction (PCR) products into a plasmid vector. The advantage of this kit is that it eliminates any enzymatic modifications of the PCR products and does not require the use of PCR primers which contain restriction sites. In this procedure, PCR inserts could be ligated efficiently with the vector, because Taq polymerase has a nontemplate-dependent activity which adds a single deoxyadenosine (A) to the 3' ends of PCR products and also the linearized vector supplied in this kit has single 3' deoxythymidine (T) residues.

However, thermostable polymerases containing extensive 3' to 5'exonuclease activity, such as Vent and Pfu do not leave 3' A overhangs. In this case, for adding 3' A overhangs, PCR products should be incubated with Taq at the end of PCR cycles. The map of the linearized vector, pCR™II, is shown below. the sequence of multiple cloning site is shown with a PCR product inserted by TA cloning. (Fig. 3.3)

3-11-1 Method:

1- generating PCR products that would be cloned in this stage modification of the PCR primer by phosphorylation or addition of a restriction site is not necessary.

2- The PCR product is ligated into PCR™II and then,

3- Transform into one shot (INVaF) competent cells and select colonies for analysis

4- Isolate plasmid and confirm presence and orientation of cloned insert by restriction digestion and/or sequencing

5- The correct recombinant plasmid is then purified for further subcloning or characterization.

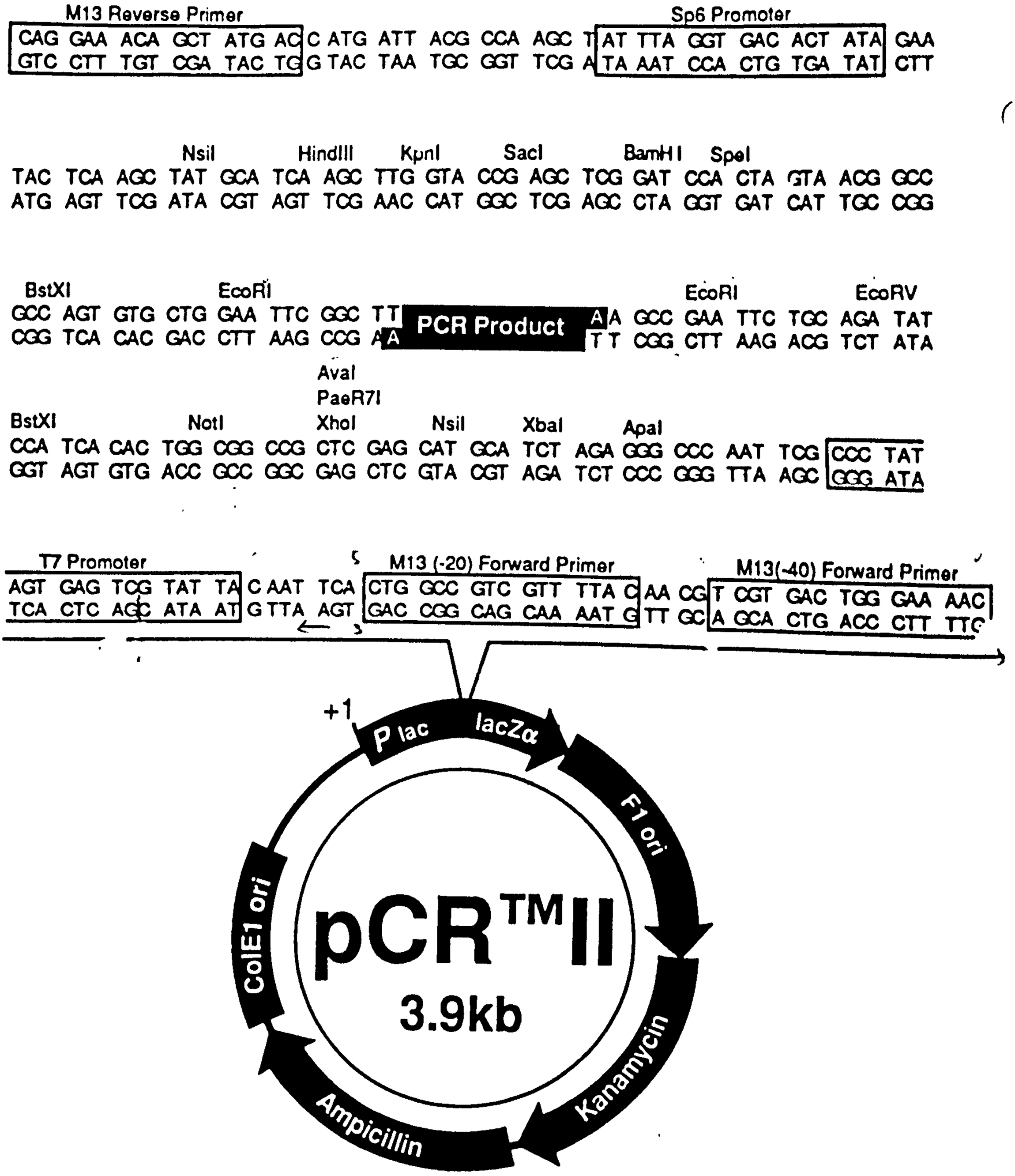
3-11-2 Amplifying the product

1- Perform the amplification reaction in a 50μl volume

2- Overlay the reaction with 70 μl of mineral oil.

3- Perform the appropriate amplification cycles with a final 7-10 minutes extension step. The final 7-10 minute extension step is necessary to make sure all DNA is double-stranded with 3' A-overhangs.

Fig. 3.3- Map of cloning vector used in this study



4- Remove 2-5 μl from each PCR sample and analyze by gel electrophoresis in a 0.8-1.5% agarose gel.

5- Quantify the amount of DNA (ng/μl) by measuring against a molecular weight marker run on the same gel.

3-11-3 Optimization of PCR

Smearing or multiple banding of product may necessiate gel purification, however, this may decrease ligation efficiency. It is recommended that PCR products should be optimized in order to avoid gel purification . A "hot start" before starting cycling program is an alternative. This is done by setting up the amplification reaction and withholding either the dNTPs or the Taq polymerase. Heat the samples to 80°C and add the missing reagent.

3-11-4 Gel purification

If the fragment of interest needs to be purified from gel (high quality agarose), all sources of nuclease contamination should be removed. For optimal ligation efficiencies, fresh PCR products were used.

3-11-5 Cloning into pCR™II

1- Use the formula below to estimate the amount of PCR product needed to ligate with 50 ng (20 fmoles) of pCR™II vector:

X ng PCR product = (Y bp PCR product)(50 ng pCR™II vector) / 3900*

* size in bp of the pCR™II vector.

2- Using the concentration determined by size marker, calculate the volume needed to give the amount determined in above formula.

3- Ligation reaction :

pCR™II vector (25ng/ul)	2μl
10X ligation buffer	1μl
PCR product (about 50-60 ng)	6 μl
T4 DNA ligase	0.6-1μl

Incubate 14-15°C overnight. Next day, if reagents are not ready to transformation, leave the ligation reaction at -20°C.

3-11-6 Transformation :

Thaw on ice DH5α competent cell (or any competent cell included with kit)

- Put 5μl ligation reaction in a 1500ul eppendorf tube
- Add 60 μl of defrosted DH5α competent cell
- Leave 30 min on ice
- Heat shock for 30 sec (2 min) in the 42°C water bath. Do not shake or mix.
- Remove the tube from water bath and place on ice for 2 min.
- Add 800 μl LB Broth
- Put in environmental shaker for 1 hour, and then return the transformed cells on ice.
- Fast spin: centrifuge for 20 sec in microfuge at maximum speed (13000 rpm)
- Pick up and through out the supernatant (about 700μl)
- Resuspend the remaining cells by gently pipetting

- Spread about 50μl of transformation vial on each two prepared LB agar plates containing 50μg/ml of ampicillin, X-Gal and 40μ of a 100mM IPTG solution. (Make sure the liquid is absorbed).

LB Broth:	Bacto Trypton	10 g
	Yeast Extract	5 g
	Sodium Chloride	10 g

Make up to 1 litre with ddH_2O and add 9 drop of 10N NaOH

LB Agar: add 15 g of agar to above recipe, after autoclave, cool to 50°C and add 50 μg/ml of ampicillin. Pour into 10cm plates. Let harden, then invert and store at 4°C.

- Invert the plates and place them in a 37°C incubator for at least 18 hours (overnight). Plates should then be shifted to 4°C for 2-3 hours to allow for proper color development.

- Pick at least 4-10 white colonies and place into 2 ml LB $^{+AMP}$ and leave it overnight.

* For RAPD inserts in this study (330-630 bp), about 50- 100 colonies per plate were obtained, most of those (>70%) were white.

3-11-7 Plasmid isolation

- Take about 1500 μl of transformed cells (after pipetting) and put in 1500ul eppendorf tube.

- Centrifuge for 2 min

- throw away the supernatant and resuspend the in 100ul of GTE solution with pipetting.

GTE solution:	Glucose	0.9 g

Tris-HCl (1M) pH 8 2.5 ml (from 25mM)

EDTA 0.5M 2ml (from 10mM)

make up to 100 ml with ddH_2O, mix on stirrer, and pass through filter (Genman, Acrodisc 32, supor 0.2 μM).

- Add the following reagents to each tube:

200μl of equal volumes of SDS(10%)+ 0.4M NaOH

150μl 3M KOAC pH 4.8

475 μl 7.5M NaOAC

- centrifuge 10 min at 14000g

- Remove supernatant and add 650μl isopropyl alcohol to it

-Leave at -20°C for 30 min

- Centrifuge for 20 min at 13ooo rpm

- Wash the pellet with 70% alcohol and/or remove supernatant and add 200ul 70% alcohol and centrifuge for 5 min.

- Dry the pellet, add 50μl TE and run on 0.8% agarose gel

5ul isolated plasmid + 5μl H_2O + 2.5μl of RNase Dye*

*RNase Dye: 40μg/ml RNase in bromophose blue or other dye.

3.11.8 Restriction endonuclease digestion of PCR products, isolated plasmids and total genomic DNA

For restriction fragment analysis on agarose gels, 10 ul of PCR product was diluted 1:1 with 2x restriction enzyme buffer and digested for 16 hours with one unit of enzyme under the conditions recommended by the

enzyme supplier (Boheringer). When products were to be analyzed on polyacrylamide gels, 2-3 μl of PCR product was first run on a 1.2% agarose gel and the concentration was estimated visually by comparison with size markers. 100 ng of PCR product was made up to 20 μl final volume with water, 2μl of 10x enzyme buffer and 1 unit of enzyme and incubated for 16 hours (overnight) at 37°C.

In order to determine the right size of insert in an isolated plasmid, 10μl of DNA, 2μl buffer, 0.5μl of each enzyme, and 7μl ddH_2O were mixed and incubated at least for 2 hours and then run on a 0.8% agarose gel. The same procedure is for total genomic DNA digestion.

3-12 Sequencing of PCR products

A 1μl aliquot of PCR product to be sequenced was checked for purity by agarose gel electrophoresis. The DNA concentration was estimated by comparison with size markers. The PCR product was purified from primer, buffer and other reagents using an S400 spin column (Pharmacia). The purified PCR product was then ethanol precipitated and resuspended in 10μl ddH_2O or TE buffer to a final concentration of 20- 80ng ul^{-1}. The sequence of *An. culicifacies* RAPD- AB11 band was determined by manual sequencing and also on a AB1 377DNA sequencing system (Applied Biosystems) using the PRISM Ready Reaction DyeDeoxy Terminator cycle sequencing kit.

3-13 SSR-PCR :

Simple sequence repeat primers used in this study were

GT1-18 $(GT)_7ATCC$

GT2-18 $(GT)_7TGTA$

AA-19 $(AGT)_5GCCA$

which have synthesized to order by Oligo-Express, UK. Other primers $(GT)_8$, $T(GT)_8$, $(GATA)_5$, $(GTG)_5$, $(GGAT)_4$, were kindly provided by Dept. of Genetics and Microbiology, Liverpool university:.

Amplification condition :

1.5 mM $MgCl_2$, 1 unit Taq, 0.01% Gelatin, 0.01% Triton-X, 2% Formamide, 1μM primer (ie 150 ng) per tube.

The following Programme has been used with SSR primers. However, when a mix of RAPD and SSR primers was used, the program was changed to a two step annealing with 39 and 50°C.

94	7 mins.	1 cycle
94	30 sec.	
52	45 sec.	27 cycles
72	2 mins.	
72	7 mins	cycle

3-14 rDNA-ITS2 PCR

In order to compare RAPD data with other PCR based techniques I applied rDNA-ITS2, and IGS PCR. ITS2 examined with *An.culicifacies*, *An. fluviatilis*, *An.maculipennis* and *An. subpictus* populations collected from Iran and Sri-Lanka.

Primers :

Genomic DNA from the above species of mosquitoes was amplified by PCR using a set of primers based on the published nucleotide sequence of *Cx. tritaeniorhynchus* 5.8S rDNA (Shimada and Sasaki 1991) and *Anopheles hermsi* 28S rDNA (Porter & Collins 1991).

5.8S 5'- ATCACTCGGCTCGTGGATCG - 3'

28S 5'- ATGCTTAAATTTAGGGGGTAGTC -3'

50 ng of each primer used

program :

95	5 mins.	1 cycle
94	30 sec.	
50	1 min.	30 cycles
72	30 sec.	

3-15 rDNA-IGS PCR

DNA was extracted from individual mosquitoes as described by Collins et al (1987) and PCR carried out using primers published in Scott *et al* (1993); 12.25ng of primer UN (GTGTGCCCCTTCCTCGATGT), 6.25ng of primer GA (CTGGTTTGGTCGGCACGTTT), 18.75ng of primer AR (AAGTGTCCTTCTCCATCCTA), 25ng of primer QD (CAGACCAAGATCGTTAGTAT), and 25ng of primer ME (TGACCAACCCACTCCCTTGA). These primers are designed based on the intergenic spacer region of An. gambiae species complex. These primers were applied with the members of this complex and also other species to confirm the identification of field materials.

Prog. 10 has been used in this diagnostic PCR.

3-16 Crossing experiment:

This study was carried out in order to find the extent and ability of RAPD primers to identify specific markers in parents and progeny of crosses between male IND-S and female LPR, and male IND-S X female RED eye strains of *An. stephensi*. Force mating was tried, but the results of this experiment are based on natural mating of these strains. In each cage 1 male and at least 10 female were left until after the second blood meal (3 days interval) and then females were separated to small cages or pots to lay eggs. Larvae, pupae, and adults of F1 of these crosses have been examined by RAPD and SSR primers (Ch. 8).

3-17 Data analysis:

RAPD data have been analysed by PHYLIP program (Felsenstein, 1993).

Sequence data analysis carried out using Clustal V Multiple Sequence Alignments (Higgins & Sharp 1988), DNA star and GCG (1994) programes. Details of these programs will be discussed in detail in the relevant chapters.

CHAPTER FOUR

MOLECULAR VARIATION IN THE ANOPHELES GAMBIAE COMPLEX DETECTABLE WITH RAPD-PCR

4.1 ABSTRACT

Random amplified polymorphic DNA (RAPD)-PCR is a powerful tool for fingerprinting anonymous genomes. Because previous RAPD studies of the *An. gambiae* complex were based largely on laboratory colonies, we have analyzed the taxonomic significance of RAPD markers with respect to field material from different parts of Africa.

After studying 75 arbitrary primers, 15 primers have been found to be useful in terms of consistency and reproducibility of amplification of molecular markers. Two 10-mers and two 17-mers were identified that produce species-specific PCR products in all six formally named members of the *An. gambiae* complex : *An. gambiae s.s., An. arabiensis, An. merus, An. melas, An. quadriannulatus, and An. bwambae.*

Eleven other primers showed variation within populations of five of the species, namely *An. gambiae, An. arabiensis, An. merus* and *An. bwambae.* Some of these primers produce single, double or multiple bands that are common to two or more species. Such sequences may be of use in the study of conserved regions in the

genomes.

Phylogenetic trees for the *An. gambiae* complex constructed on the basis of the products of all fifteen RAPD primers were in concordance with trees based on mtDNA, and rDNA, revealing that *An. gambiae* and *An. arabiensis* are sister taxa.

The generation of reproducible DNA fingerprints from dried specimens of *An. bwambae* demonstrated another powerful feature of RAPD-PCR.

After cloning and sequencing, a RAPD fragment in *An. bwambae* has been shown to have more than 96% homology with another characterized RAPD sequence in *An. culicifacies* and ubiquitin-related genes in other organisms.

By following the progeny of RAPD-typed adults, some non-parental bands and life-stage specific RAPD patterns were identified.

Reproducibility of the RAPD products was enhanced by using a) less DNA (5pg to 10ng per mosquito), b) a phenol extraction stage during DNA purification. c) Taq polymerase from reliable sources (without contaminating bacterial DNA) and d) a two step program with two different annealing temperatures.

4-2 INTRODUCTION

4.2.1 Background

It was a study of the mode of inheritance of insecticide (dieldrin) resistance in what was then known as *An. gambiae* that led to the discovery that this "species" was in fact a complex of sibling species (Davidson 1962). Mosquitoes of the *An. gambiae* species complex include the major vectors of human malaria in tropical Africa . Their population genetics have been studied more than any other anopheline species complex. These studies have analyzed chromosome inversions (Coluzzi 1968 ; Coluzzi & Sabatini 1967, 1968, 1969; Coluzzi *et al.* 1977; Coluzzi *et al.* 1985; Garcia *et al.* 1996; Toure et al. 1996, 1997), isoenzymes (Mahon *et al.* 1976; Miles 1978, 1979; Marchand & Manzova 1985; Hunt & Coetzee 1986a, 1986b; Collins *et al.* 1988b), cuticular hydrocarbons (Carlson & Service 1979, 1980; Phillips *et al.* 1988; Milligan *et al.* 1993), repetitive sequences as DNA probes (Gale & Crampton 1987; Collins *et al.* 1987; Collins *et al.* 1988a, 1988c; Hill *et al.* 1991a, 1991b; Hill & Crampton 1994), rDNA (Collins *et al.* 1987, 1989; Paskewitz & Collins 1990; Taylor *et al.* 1993; Scott *et al.* 1993; Townson & Onapa 1994; Cornel & Collins 1996, Collins & Paskewitz, 1996), *in situ* hybridization to polytene chromosomes (Kumar & Collins 1994) and recently described RAPD-PCR (Williams *et al.* 1990; Welsh & McClelland 1990; Wilkerson *et al.* 1993; Favia *et al.* 1994a and 1994b) and microsatellites (Zheng *et al.* 1993a, 1993b; Lanzaro *et al.* 1995, Lehmann *et al.*, 1996b, 1997).

4.2.2 Aims and objectives

The aim of this study was to determine the extent of inter- and intra-specific genetic variation in species of the *An. gambiae* complex using the technique of RAPD-PCR. Emphasis was put on simple and reproducible procedures that could be employed with field-collected specimens stored in a variety of ways.

4.3 MATERIALS AND METHODS

4.3.1 Source of mosquitoes

The six formally named species of the *An. gambiae* complex : *An. gambiae*, *An. arabiensis*, *An. merus*, *An. melas*, *An. quadriannulatus*, and *An. bwambae* mostly field and a few laboratory samples originating in different parts of Africa (Fig. 4.1 and Table 4.1).

4.3.2 Template preparation, PCR primers and Programs

Template preparation

A comparison between the following extraction methods described by Ballinger Crabtree *et al.* (1992), Collins *et al.* (1988a), Medina-Acosta & Cross (1993), Tassanakajon *et al.* (1993), Snounou *et al.* (1993a, 1993b) revealed that those which included a phenol extraction were most suitable for RAPD PCR.

PCR primers

M13 sequencing primers (M13F and M13R 17-mers), OPA series (four 10-mers) AB01 series (twenty 10-mers), and fifty one 10-mers from the UBC series. Table 4.2 lists those primers for which results have been presented in this part of the study.

PCR Programs

Seventeen programmes were examined based on their different annealing

Fig. 4.1- Map of Africa show the origin of mosquitoes have been used in this study.

Table 4.1A - Origin of the six members of *An. gambiae* species complex from different parts of Africa. Using a wide rang of field material along with few laboratory strains, provided more data on inter- and intra- specific genetic variation of these important malaria vectors.

	Origin	Field	Lab.
An. gambiae s.s.	Tanzania	+	KWA,KIL,ZANDS
	Kenya	+	
	Uganda	+	
	Ghana	+	
	Nigeria		16CSS,REF-MA
	Gambia		GAG
An. arabiensis	Tanzania	+	
	Kenya	+	
	Uganda	+	
	Ghana	+	
	Mozambique	+	
	Zimbabwe		KGB
	Sudan		SENN
	Burkina Faso		BF
An. merus	Mozambique	+	
	South Africa	+	
An. melas	Gambia	+	
An. quadriannulatus	South Africa	+	
	Mozambique	+	
An. bwambae	Uganda	+	

Table 4.2 - selected primers and their sequences have been used in this study

primer	5....sequence ....3'	GC%
M13	GTAAAACGACGGCCAGT	53
M13R	CAGGAAACAGCTATGAC	47
OPA1	CAGGCCCTTC	70
OPA8	GTGACGTAGG	60
AB1-01	GTTTCGCTCC	60
AB1-04	GGACTGGAGT	60
AB1-19	ACCCCCGAAG	70
UBC-302	CGGCCCACGT	80
UBC-303	GCGGGAGACC	80
UBC-305	GCTGGTACCC	70
UBC-306	GTCCTCGTAG	60
UBC-308	AGCGGCTAGG	70
UBC-309	ACATCCTGCG	60
UBC-353	TGGGCTCGCT	70
UBC-400	GCCCTGATAT	50

temperature (36, 38, 39, 42, 47, 50, 60 °C), variable numbers of cycles (20, 30, 40, 45), and two/three step programmes (e.g. a two step programm with an annealing temperature of 36 and 50°C). These programmes are described in chapter 3 and are identified as Prog 1, Prog 2, Prog 3, etc.

4.3.3 Cloning and Sequencing

The cloning of PCR products was carried out using the Invitrogen original TA cloning kit in which the inserted PCR product is flanked on each side by EcoR I sites (see chapter 3). Prior to sequencing, clones and PCR products were purified by passage through a S400 spin column (Pharmacia), ethanol precipitated and resuspended in a volume of ddwater based on DNA concentration estimated on a gel by comparison with size markers. Sequences were determined by the cycle sequencing reaction in a ABI PRISM 377 DNA sequencer using AMPLITAQ FS and the manufacture's guidelines for M13 forward and reverse primers. In manual sequencing based on the chain termination reaction, Sequenase version 2.0 T7 DNA Polymerase (USB) was used.

4.3.4 Data Analysis

RAPD data were compared using the following similarity and distance coefficients calculated for six members of *An. gambiae* species complex :

Jaccard similarity (Southwood, 1978) : $S_S = A / B+C$

Jaccard similarity (Cibulskis *et al.* 1986): $S_J = A / A+B+C$

Dice similarity (Beverly *et al.* 1987): $S_D = 2A / 2A+B+C$

Similarity of Nei & Li (Nei and Li 1979): $S_N = 2N_{ab} / Na + N_b$

Simple Matching coefficient (Sneath & Sokal, 1973): $S^M = A+D / A+B+C+D$

Montpellier distance (Tibyrenc et al. 1993a, 1993b): $S_{Jm} = 1- [A / A+2(B+C)]$

where A is the number of shared bands in two individuals, while B and C are the total number of bands scored in each individual. NAB is the number of bands two individuals A and B share in common, NA and NB are the total number of bands scored in A and B respectively and D is the number of bands absent from both individuals (see also section 2.5.4).

The MIX program in PHYLIP (Felsenstein, 1993) using Wagner and Camin-Sokal parsimony methods (1965) has been applied to RAPD and SSR data in the specific format of 1 for presence and 0 for absence of a band (e.g. 1001110101110). GCG (version 7.0, 1994), DNA Star, and Clustal V (Higgins & Sharp 1988) packages have been used to enter, align and analyze the RAPD sequences and other homologous sequences.

4.4 RESULTS

4.4.1 Optimization:

Different DNA concentrations and volumes of PCR mix have been examined in order to find the effects of these parameters on quantitative and qualitative changes in PCR amplification. The results revealed that 5pg to 10ng of phenol extracted DNA produce sharper and more reproducible bands (Fig. 4.2A). With respect to the volume of PCR mix, there is no discernable difference between 25 and 50 μl PCR mix products, but the brightness of bands decreases with lower volumes (Fig. 4.2B). In other experiments, 1 unit of Taq polymerase produced clearer patterns than lesser amounts (data not shown).

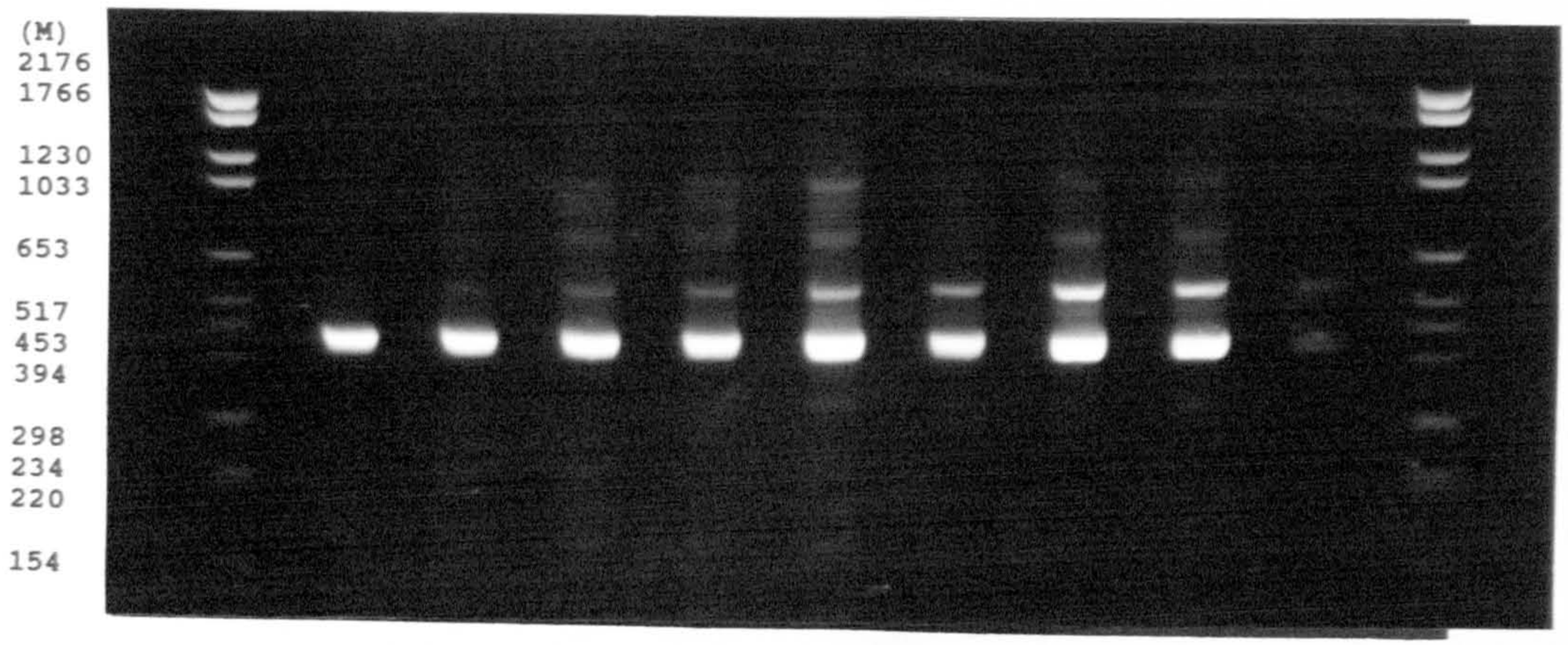
4.4.2 Species-specific markers

Isoenzymes:

The application of the octanol dehydrogenase (ODH) method to differentiate *An. gambiae* and *An. arabiensis* has been reported before (Collins *et al.* 1988). Polyacrylamide gel electrophoresis and the ODH enzymatic reaction was optimized for the identification and separation of laboratory stocks of *An. gambiae* and *An. arabiensis*, with *Aedes aegypti* included for comparison. The results showed three octanol dehydrogenase electromorphs corresponding to the 100, 98 and 95 allelic forms. 100 for *An. gambiae*, two 100 and 98 allelic form for *An. arabiensis* and 95 for *Aedes aegypti* (Fig. 4.3). PGM, SOD, and HEX also have been used with laboratory strains of *An. gambiae* and *An. arabiensis*, and were able to differentiate

Fig. 4.2- Optimization of RAPD condition

A: different DNA concentrations from left to right; 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 5, 10, 25ng.



B: different volums of PCR mix from left to right; 50, 25, 20, 15, 10, 5 ul.

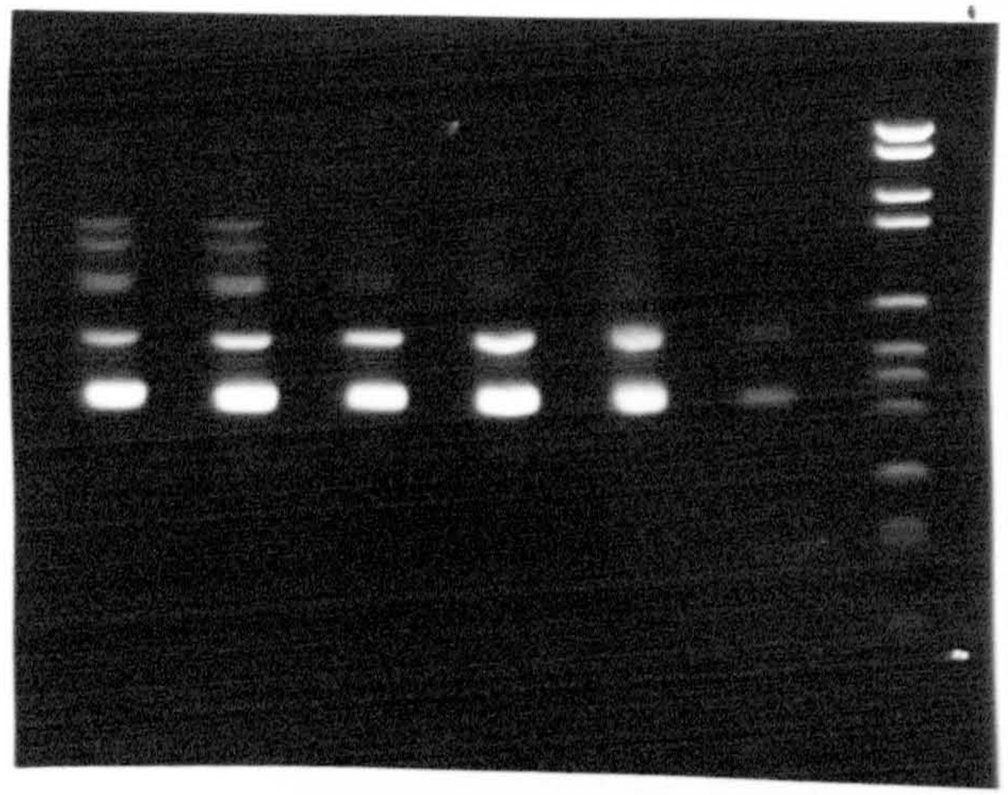
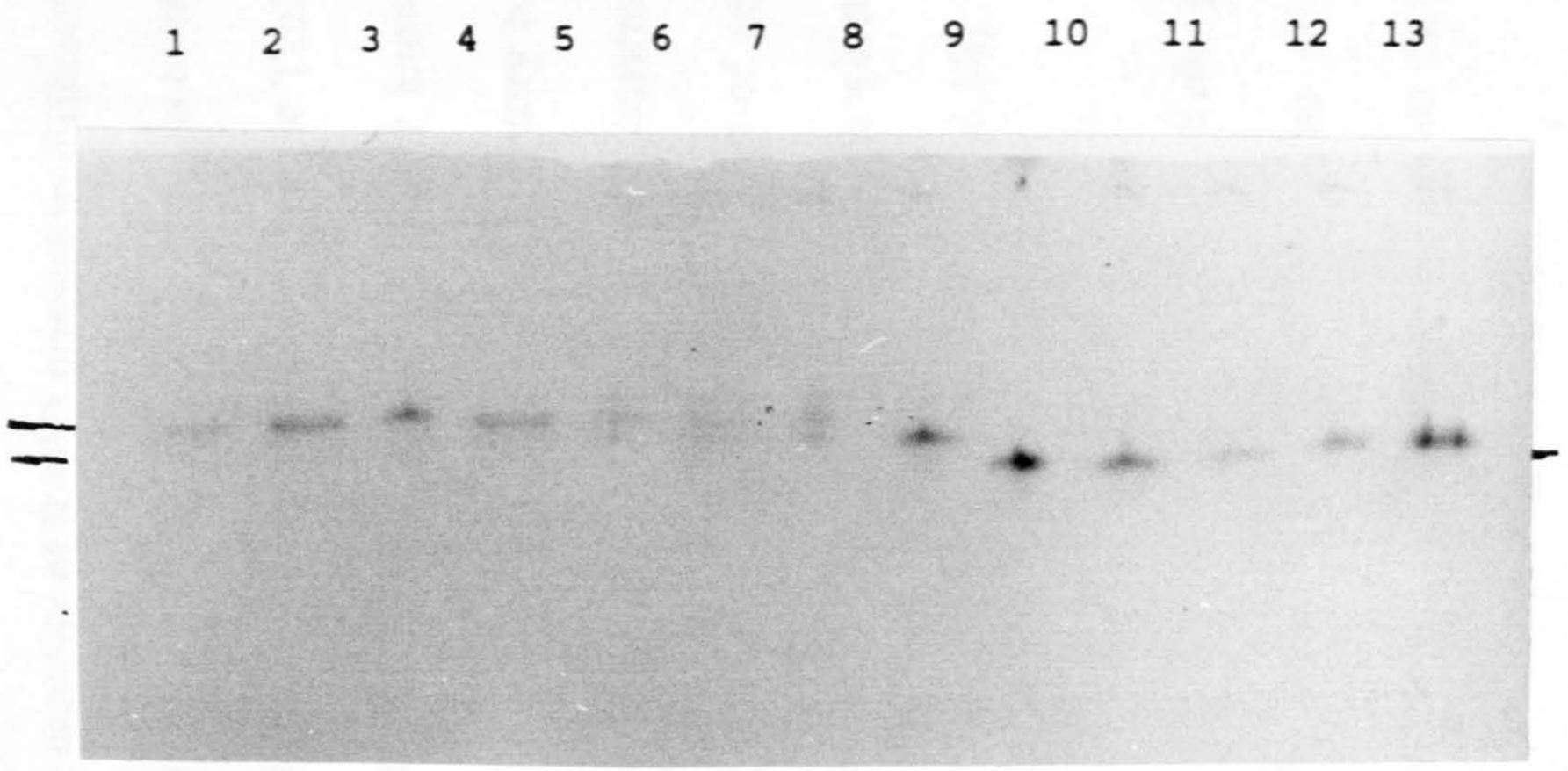


Fig. 4.3- odh enzyme activity in *Aedes aegypti* (1-5), *An. arabiensis* (6-9), and *An. gambiae* (10-13)



two species.

RAPD-PCR

In evaluating the capacity of RAPD primers to differentiate species within the *An. gambiae* complex, the aim has been to test individuals of each species from more than one geographical area within sub-saharan Africa (see Table 4.1B), although this was not possible with *An. melas* (*An. bwambae* is only known from one locality). Initial evaluation of primers is often simplest with laboratory colonies but in all cases results with laboratory colonies have subsequently been confirmed with field samples. In the figures which drawn, representative photographs of gels are shown. Unless otherwise stated 'species-specific RAPD bands' are those which are seen consistently in all specimens from each locality sampled (see Table 4.1B)

A total of 75 RAPD primers were tested with the *An. gambiae* complex. Of these, 15 primers were found to be useful in terms of the consistency and reproducibility of their amplification products. The results of RAPD PCR with the most useful of these primers are described below.

Two 10-mer (UBC-303 and UBC-306) and two 17-mers (M13F and M13R) were identified that produced species-specific PCR products in all six formally named members of the An. gambiae complex.

M13F

M13F in a program with 36°C annealing and 45 cycles (Prog 1), differentiated

all six members of the complex with at least 9 bands in each species (Fig. 4.4): The *An. bwambae* individuals were from field in Uganda, *An. merus* and *An. quadriannulatus* from South Africa, *An. arabiensis* from Sudan, Kenya and Uganda. *An. gambiae* specimens originated from Tanzania and Kenya and *An. melas* from S. Africa (see Table 4.1B).

M13R

M13R with higher annealing (39°C) temperature in 40 cycles program (Prog 2), amplified species-specific patterns, usually with a single or two main bands for all six members of *An. gambiae* complex (Fig. 4.5). A 310bp in *An. bwambae* from Uganda, 280bp in *An. gambiae* from Kenya and Nigeria, or three bands of 340, 287, and 190bp in *An. arabiensis* originating from Zimbabwe and Burkina Faso. In other members, *An. quadriannulatus* and *An. merus* provided from S. Africa and Mozambique, and finally *An. melas* specimens were from the Gambia. Origin of other specimens used with this primer listed in Table 1B.

UBC-303

In UBC series, primer UBC-303 showed the following species-specific bands: *An. gambiae* from Ghana and Tanzania with two 300 and 385bp, *An. arabiensis* from Ghana and Tanzania showed two 440 and 500bp, *An. merus* from South Africa have 517 and 610bp fragments, *An. melas* from Gambia (455, 517bp), *An. quadriannulatus* from South Africa with three 200, 368 and 505bp, and finally *An. bwambae* from Uganda with a single 250bp which always appears as a relatively faint band in different individuals examined (Fig. 4.6, Table 4.1B).

Table 4.1B- Origin and number of specimens have been used to generate species-species markers with RAPD primers

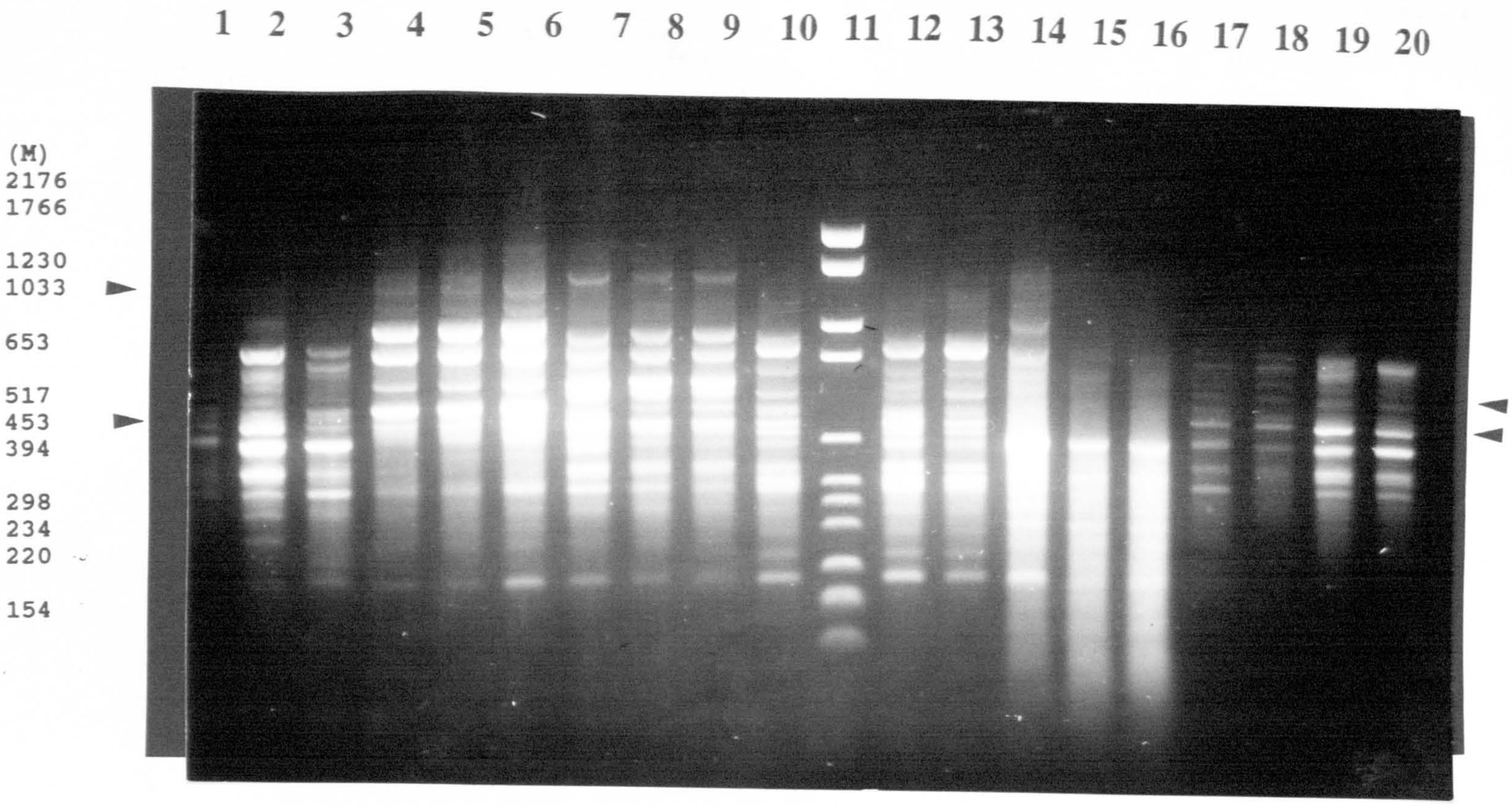
	gambiae n	arabiensis n	merus n	melas n	quadriannulatus n	bwambae n
M13F	Tanzania (L) 19 KWA 10 ZANDS 5 KILL 4 Kenya (F) 15 Nigeria (L) 18 16CSS Gambia (L) 7 GAG Uganda (F) 9	Sudan (L) 19 SENN Kenya (F) 17 Uganda (F) 15 Ghana (F) 8 Burkina Faso (L) 5 BF	South Africa (F) 20 Mozambique (F) 19	Gambia (F) 18	South Africa (F) 19 Mozambique (F) 7	Uganda (F) 33
M13R	Kenya (F) 17 Nigeria (L) 25 16CSS Ghana (F) 5 Uganda (F) 11	Zimbabwe (L) 18 KGB Burkina Faso (L)5 BF Kenya (F) 13 Uganda (F) 11	South Africa (F) 14 Mozambique (F) 19	Gambia (F) 19	South Africa (F) 17 Mozambique (F) 5	Uganda (F) 22 Museum 9
UBC-303	Tanzania (F) 7 Ghana (F) 11 Uganda (F) 7 Gambia (L) 11 GAG Kenya (F) 7	Ghana (F) 7 Tanzania (F) 10 Zimbabwe (F) 5 Sudan (L) 6 SENN Uganda (F) 5	South Africa (F) 17	Gambia (F) 13	South Africa (F) 14	Uganda (F) 15
UBC-306	Tanzania (F&L) (F) 5 16CSS 11 Ghana (F) 5 Kenya (F) 7 Uganda (F) 8	Ghana (F) 7 Tanzania (F) 7 Sudan (L) 8 Uganda (F) 5	South Africa (F) 17	Gambia (F) 14	South Africa (F) 17	Uganda (F) 18

(F): field (L): lab. stock Museum: dried specimen Numeral= no. of mosquitoes examind

Legend to Fig. 4.4

M13f in Prog 1 differentiated all six members of the complex with at least 9 bands in each species. *An. bwambae* individuals are from field in Uganda (lanes 1-3), *An. merus* (4-6) is from South Africa, *An. quadriannulatus* (7-9) from S. Africa, *An. arabiensis* (lane 10 lab. stock from Sudan, while the other two are field collections from Kenya (12) and Uganda (13). In *An. gambiae*, the first (14) is KWA originated from Tanzania and other two (15-16) from Kenya and *An. melas* (17-20) are from S. Africa.

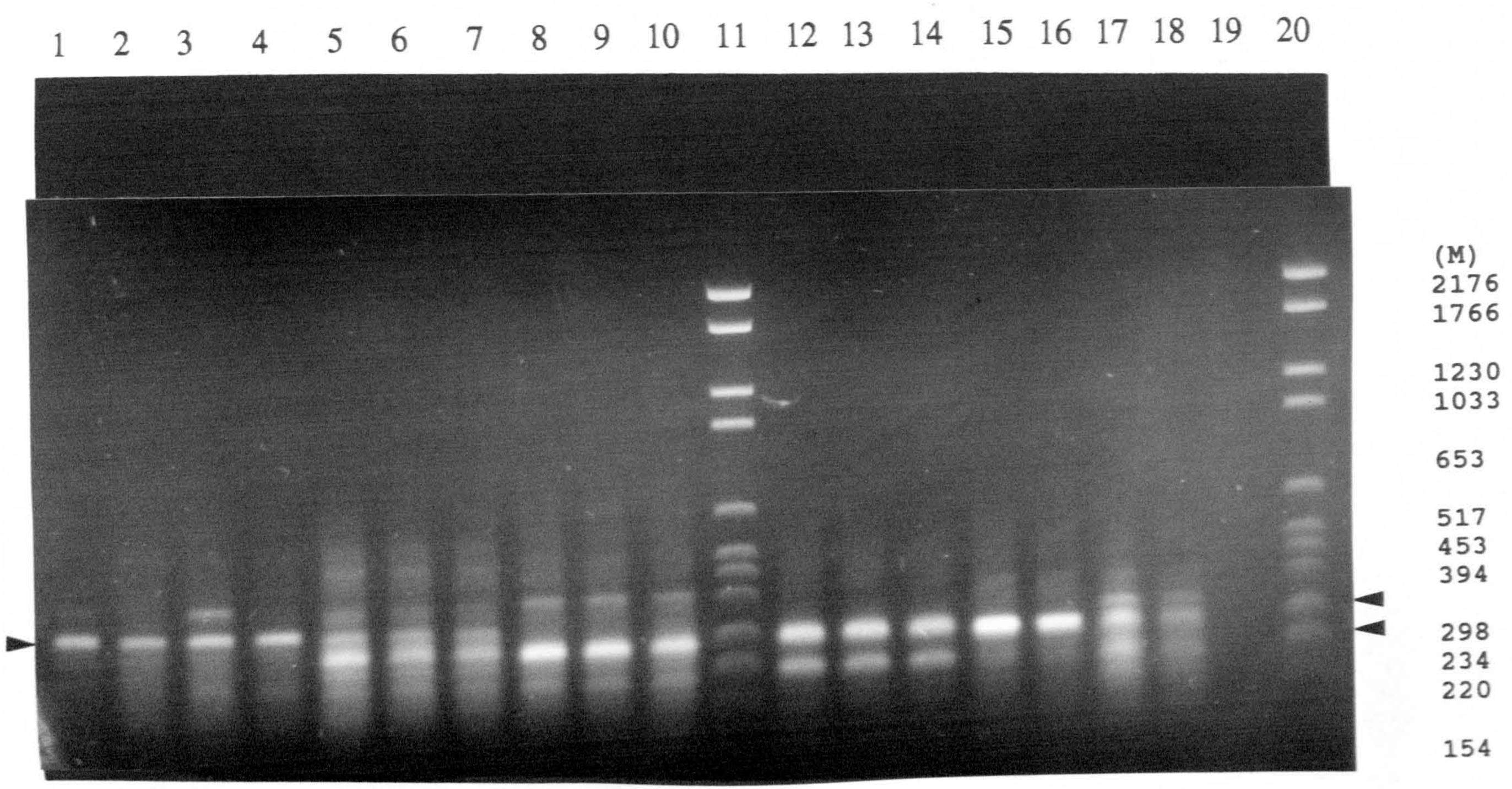
Fig. 4.4- RAPD-(Prog 1)-primer M13F products in *An. gambiae* species complex: *An. bwambae* (1-3), *An. merus* (4-6), *An. quadriannulatus* (7-9), *An. arabiensis* (10, 12-13), *An. gambiae* (14-16), *An. melas* (17-20); molecular marker (11).



Legend to Fig.4.5

M13R with higher annealing (39°C- Prog 2), amplified species-specific patterns, usually with a single or two main bands for all six members of *An. gambiae* complex Λ 310bp in *An. bwambae* from Uganda, dried (1,2) and alcohol preserved (3,4), a 280bp in *An. gambiae* from field (Kenya, 15) and Nigeria (lab, 16), or three bands of 340, 287, and 190bp in *An. arabiensis* originated from Zimbabwe (17) and Burkina Faso (18). In other members, *An. quadriannulatus* provided from field in S. Africa (5) and Mozambique (6-7), *An. merus* from S. Africa (8) and Mozambique (9-10) and finally *An. melas* specimens were from the Gambia. Origin of other specimens used with this primer listed in Table 1B.

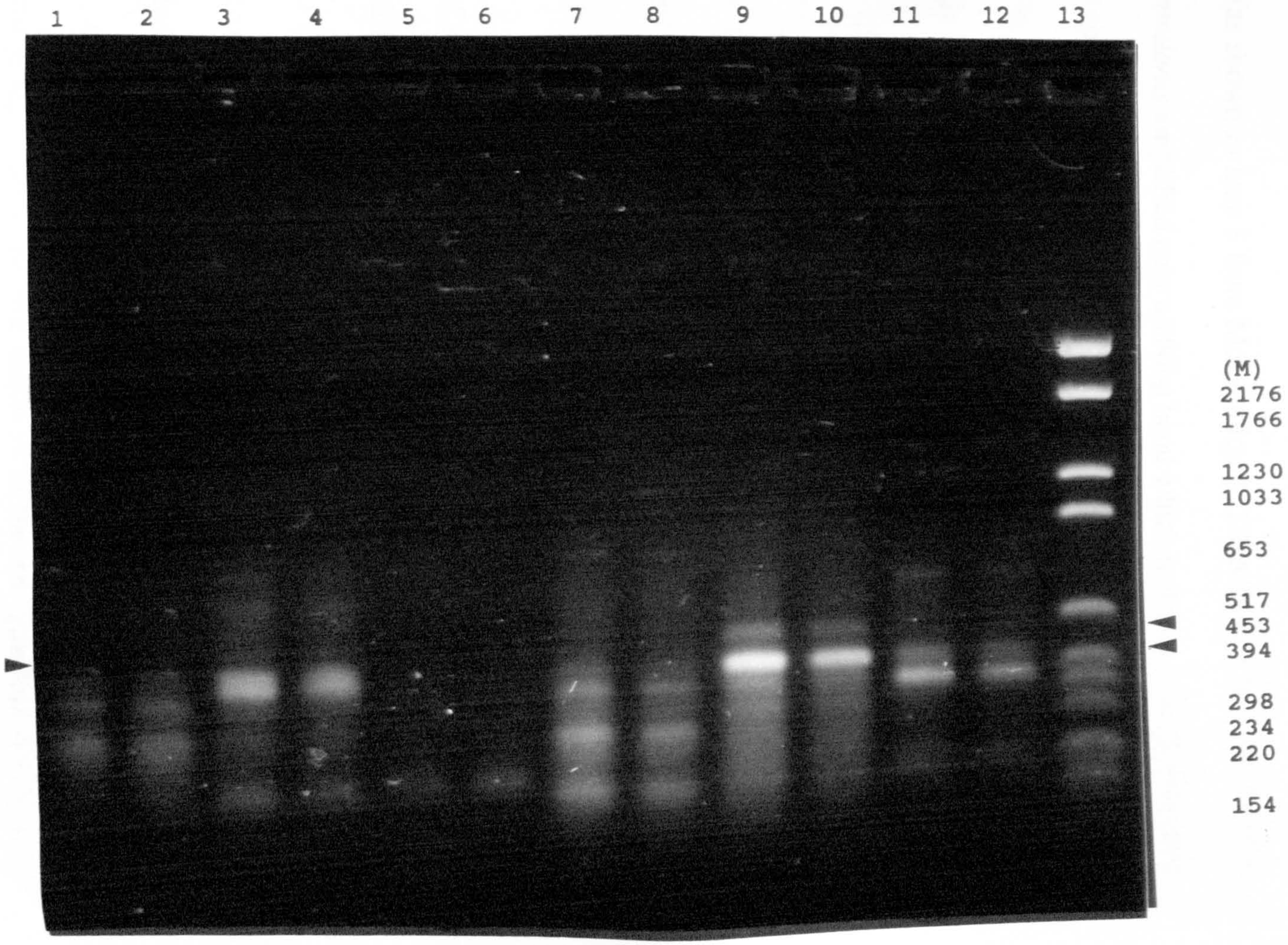
Fig. 4.5- RAPD-1 M13R (anealing 39°C, 40 cycles) products in *An. gambiae* species complex: *An. bwambae* (1-4), *An. quadriannulatus* (5-7), *An. merus* (8-10), *An. melas* (12-14), *An. gambiae* (15-16), *An. arabiensis* (17-18),; Marker -ve (19). RAPD pattern in each species is Species-specific and differentiate



Legend to Fig. 4.6

UBC-303 showed the following species-specific bands: *An. gambiae* from Ghana (1) and Tanzania (2) with two 300 and 385bp, *An. arabiensis* from Ghana (3) and Tanzania (4) showed two 440 and 500bp, *An. merus* from South Africa (9-10) have 517 and 610bp fragments, *An. melas* (11-12) from Gambia (455, 517bp), *An. quadriannulatus* (7-8) from South Africa with three 200, 368 and 505bp, and finally *An. bwambae* from Uganda with a single 250bp which always appears as a relatively faint band in different individuals examined (see also Table 1B).

Fig. 4.6- RAPD primer UBC303 products in *An. gambiae* species complex: *An. gambiae* (1-2), *An. arabiensis* (3-4), *An. bwambae* (5-6), *An. quadriannulatus* (7-8), *An. merus* (9-10), *An. melas* (11-12); Marker(13)



UBC-306

Another series of specific markers are provided by primer UBC-306 (Fig. 4.7). The results revealed a 1060bp in *An. gambiae* from Ghana and Kenya, and two 590 and 760bp bands in *An. bwambae* from Uganda. However, in some amplifications (as shown in lane 8 from Mozambique), DNA from a few individuals of *An. quadriannulatus* amplified only a 760bp band which it shares with all members of complex except *An. gambiae s.s.* and do not show the 230bp specific fragment normally seen in *An. quadriannulatus*. In this case a re-extraction of the DNA of those individuals recovered the species-species band (Fig. 4.7, Table 4.1B).

There are other RAPD primers that produce one or more common bands in some members of the complex and additional species-specific bands. The PCR products of some of these primers are therefore diagnostic for some of the species.

AB1

50 ng of AB1 primer in a two-step annealing program (Prog 3) produced a major strong 420 bp band in *An. gambiae* from Uganda (1) and Gambia(2), *An. quadriannulatus* (3) from South Africa, and *An. arabiensis* from Kenya (6), Burkina Faso (8) and Mozambique (9) while in *An.merus* (4-5) from South Africa this main band is 380bp and *An. bwambae* (10-11) collected in Uganda has both 420 and 380bp fragments (Fig. 4.8). Hence under these amplification conditions, primer AB1 can differentiate three groups among the five members of the *An. gambiae* complex, 1) *An. gambiae*, *An. quadriannulatus*, and *An. arabiensis*, 2) *An.merus* and 3) *An.*

Legend to Fig. 4.7

UBC-306 revealed a 1060bp in *An. gambiae* from Ghana (12-13) and Kenya (13-14), or two 590 and 760bp bands in *An. bwambae* (4-5) from Uganda. However, in some amplifications (as shown in lane 8 from Mozambique), DNA from few individuals in *An. quadriannulatus* amplify only a 760bp band which it shares with all members of complex except *An. gambiae s.s.* and do not show the 230bp specific fragment while the other two (6-7) originated from S. Africa have two diagnostic bands of *An. quadriannulatus*. In *An. arabiensis* also some variation is detectable although the first three (16-18) are from Ghana and the other (19) is from Uganda (see also Table 1B).

Fig. 4.7- RAPD primer UBC306 products in *An. gambiae* species complex: *An. melas* (1-3), *An. bwambae* (4-5), *An. quadriannulatus* *An. merus* (9-10), *An. gambiae* (12-15), *An. arabiensis* (16-19); Marker(11); -ve (20).

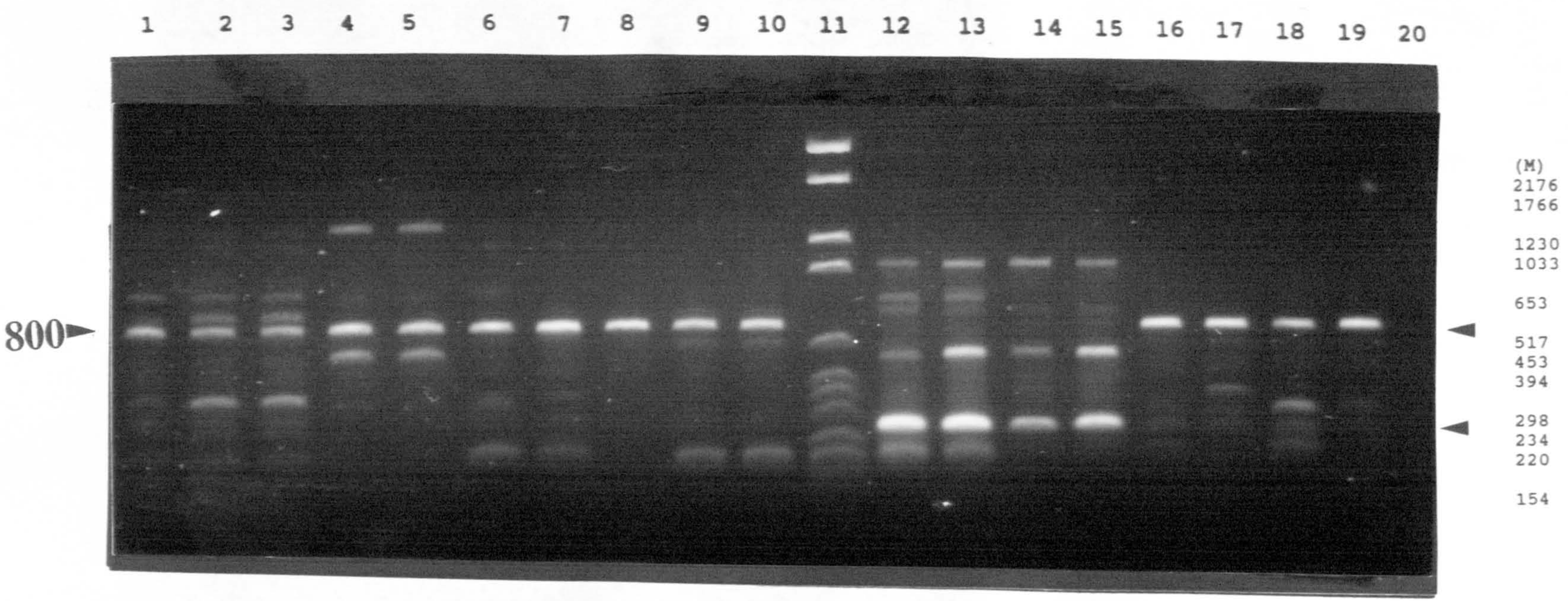
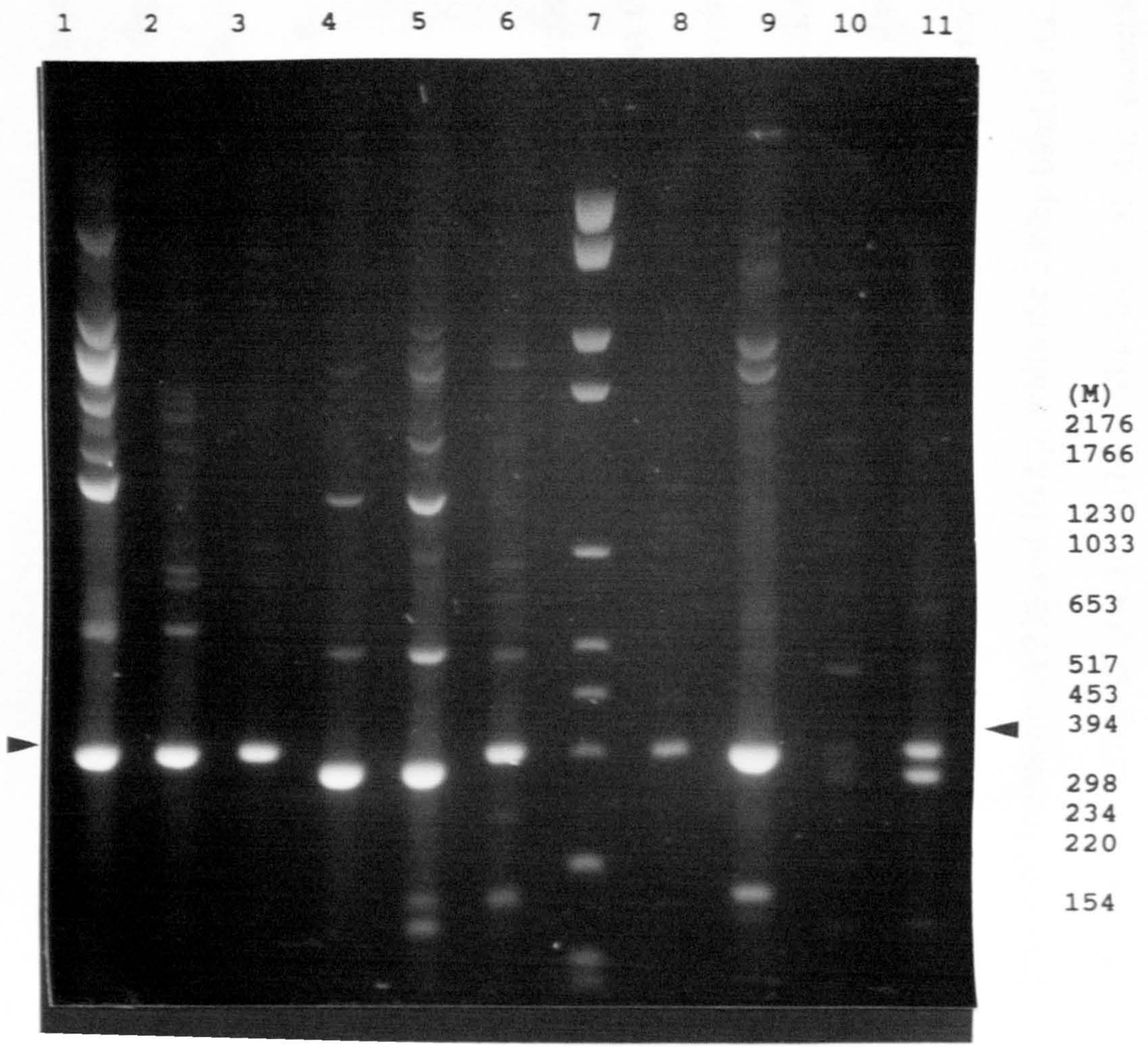


Fig. 4.8- RAPD primer AB1 products in *An. gambiae* species complex: *An. gambiae* (1-2), *An. quadriannulatus* (3), *An. merus* (4-5), *An. arabiensis* (6, 8-9), *An. bwambae* (10-11); Marker(7)



bwambae.

Based on experience with other primers and in an attempt to reduce the background and faint bands, the primer concentration was reduced to 30ng and the program changed to one step annealing with 50°C (Prog 4). Fig. 4.9, shows the amplification of AB1 primer in Prog 4, which produced a single band of 380-420bp in four members of *An. gambiae* complex and double bands in *An. bwambae* but without any other faint bands.

To see if these single bands from the different species represent homologous sequences, they were further characterized by excising them from the gels, and, after eluting and purifying the DNA, digesting it with the restriction endonucleases. When the common band seen in Fig.4.10 was digested, at least two restriction sites for DraI were seen in each species. In *An. gambiae* and *An. arabiensis*, the 420bp band was cut into fragments of 235 and 160bp; the 400bp band in *An. quadriannulatus* produced fragments of 235 and 160bp, while the 380bp band of *An. merus* fragments of 223 and 157bp (Fig.4.10). The 392bp band of *An. bwambae* was cut into fragments of 235 and 157bp (data not shown).

OPA8

In contrast to AB1 which produce one common band in some members of complex, other RAPD primers have the ability to generate complex patterns of common bands which are suggestive of a high level of sequence homology. Where there are a number of these common bands, it can be argued that they have been

Fig. 4.9- RAPD primer AB1 (30ng- 50°C) products in *An. gambiae* species complex: *An. gambiae* (1-2), *An. quadriannulatus* (3-4), *An. merus* (7-8), *An. bwambae* (9); Marker(5); -ve (10)

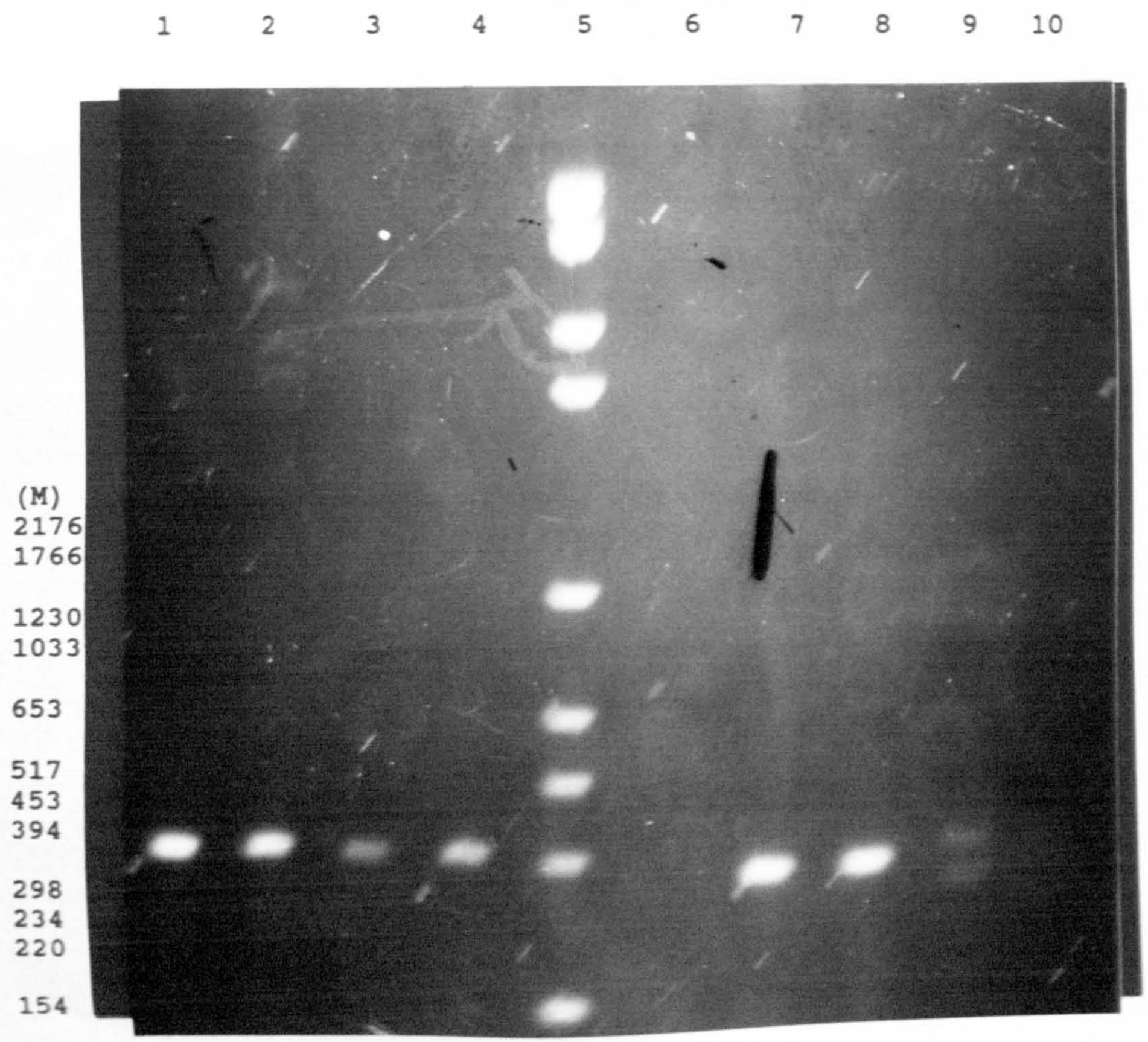
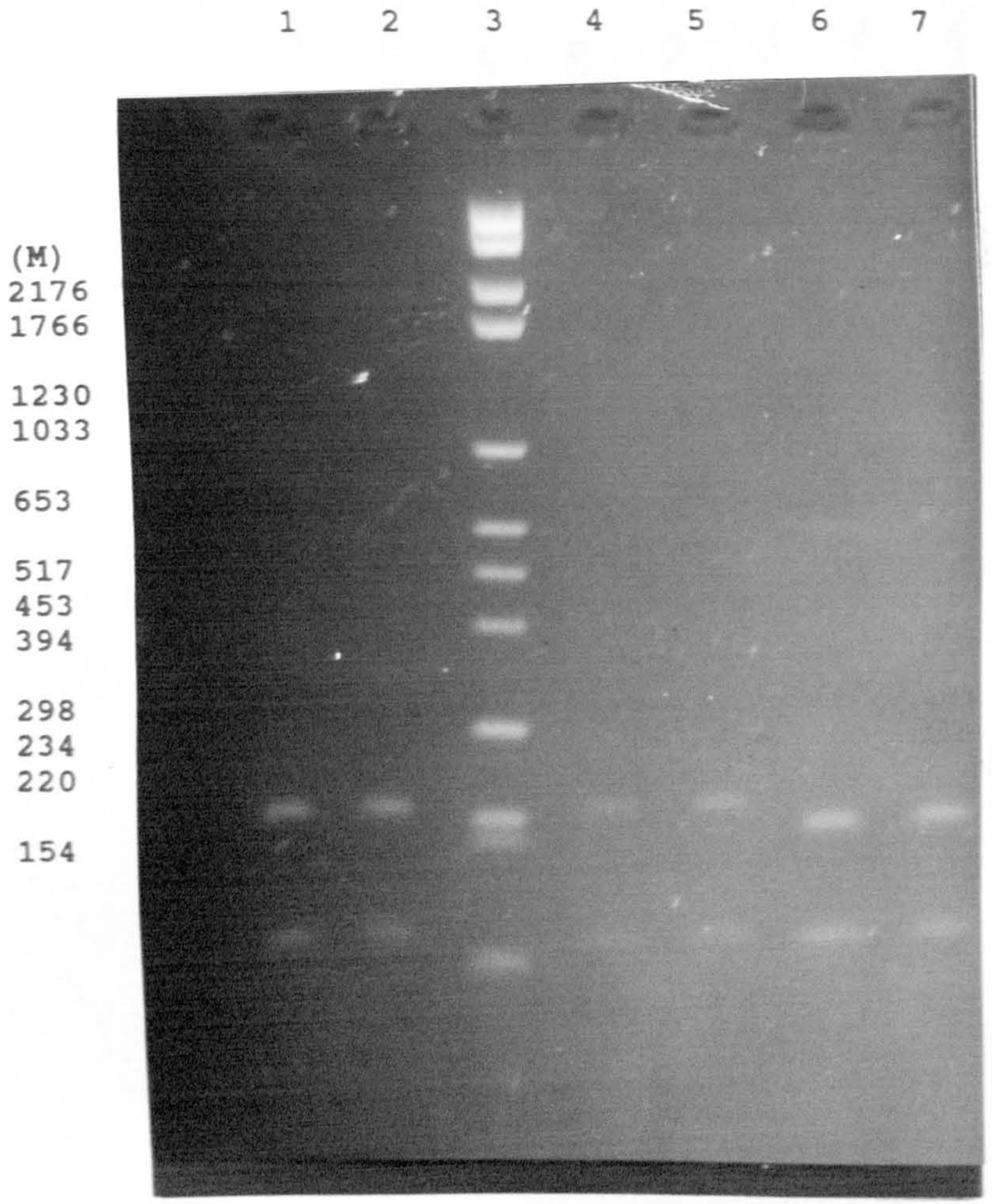


Fig. 4.10- Restriction digestion of AB1 (30ng- 50°C) products in *An. gambiae* species complex: *An. gambiae* (1-2), *An. quadriannulatus* (4-5), *An. merus* (6-7)



amplified from a highly conserved region. OPA8 is one of these primers with at least six bands common to five members of the *An. gambiae* species complex tested. This primer in Prog 1 (36°C annealing temperature), consistently amplified a common pattern in *An. arabiensis* , *An. quadriannulatus*, *An.melas*, *An. merus* (Fig. 4.11) and *An. gambiae* (not shown). However, *An. quadriannulatus* has an extra 325bp fragment, present only in this species. A 285bp fragment has been amplified from some individuals within a *An. melas* population and one of 365bp in *An. arabiensis* from Sudan (lane 7)(Fig. 4.11).

AB19

In a 3-step annealing (45, 39, 35°C- Prog 7), 50ng of primer AB19 produced two common bands (410 and 525bp) in all members of *An. gambiae* complex (Fig. 4.12). These two bands have been consistently amplified on different occasions from members of *An. gambiae* complex. PCR products run on a 6% acrylamide gel followed by silver-staining showed almost the same pattern except for an extra band in *An. quadriannulatus* and another one in *An. bwambae* (Fig. 4.13). Although primer AB19 is apparently not useful for species diagnosis since its PCR products originate from regions conserved across species, characterizing such fragments by sequencing and designing primers based on nucleotide differences found in different species could be useful for further work. The rational behind experiments in which primers for conserved regions have been used singly as RAPD primers will be discussed later in this chapter.

Legend to Fig.4.11

OPA8 in Prog 1 (36°C annealing), consistently amplified a common pattern in *An. arabiensis* (Mozambique (1-6), and Sudan (lane7), *An. quadriannulatus* (Mozambique), *An.melas* (Gambia) and *An. merus* (Mozambique). However, *An. quadriannulatus* has an extra 325bp fragment, present only in this species. A 285bp fragment has been amplified from some individuals within *An. melas* population or a 365bp in *An. arabiensis* from Sudan (lane 7).

Fig. 4.11- RAPD primer OPA8 products in *An. gambiae* species complex: *An. arabiensis* (1-7), *An. quadriannulatus* (9-11), *An. merus* (12-14), *An. melas* (16-20); Marker(8,15).

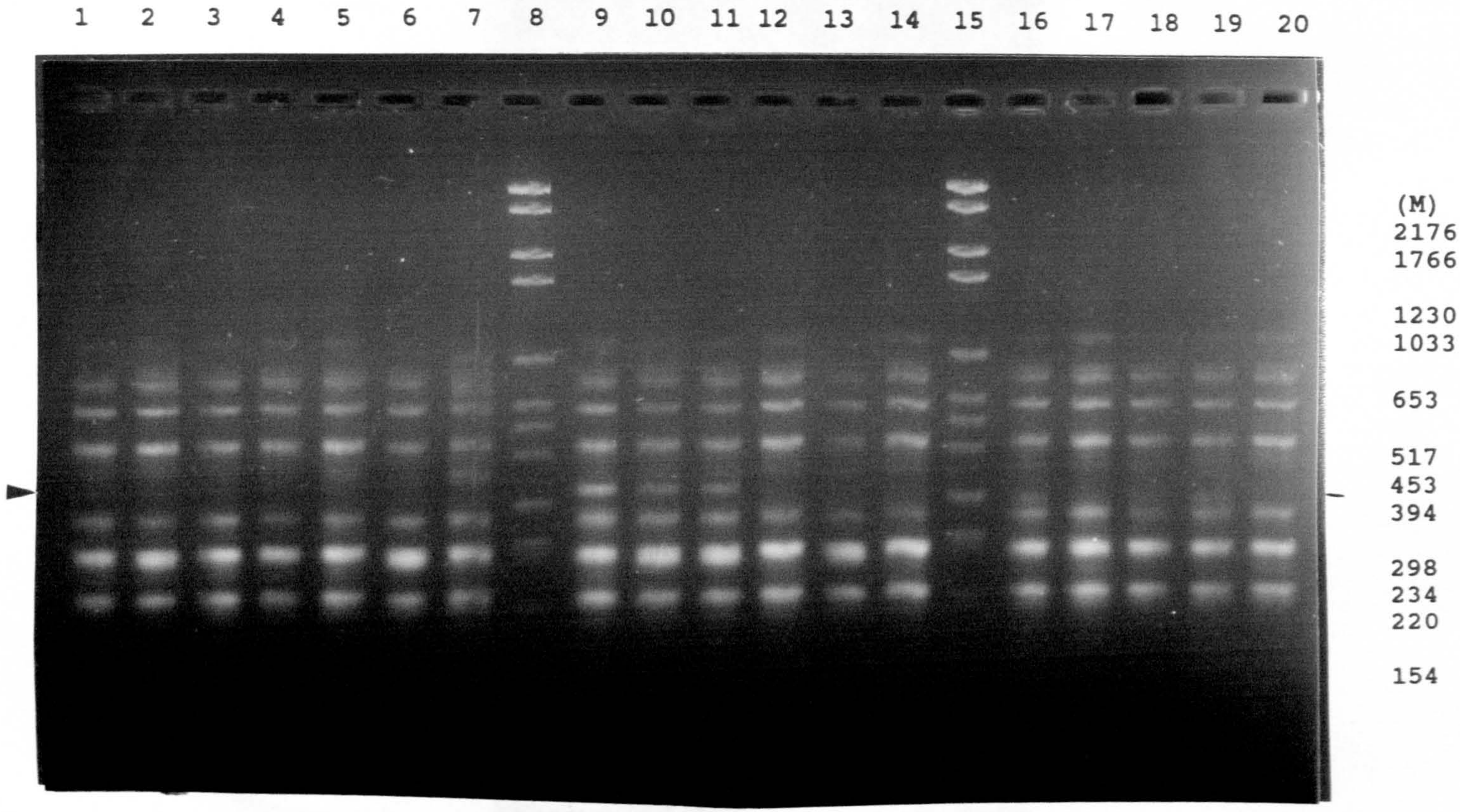


Fig. 4.12- RAPD primer AB19 (50ng- 3 steps annealing) products in all six members of *An. gambiae* complex. Amplification of only two fragments common to all species may be to a very conserved region.

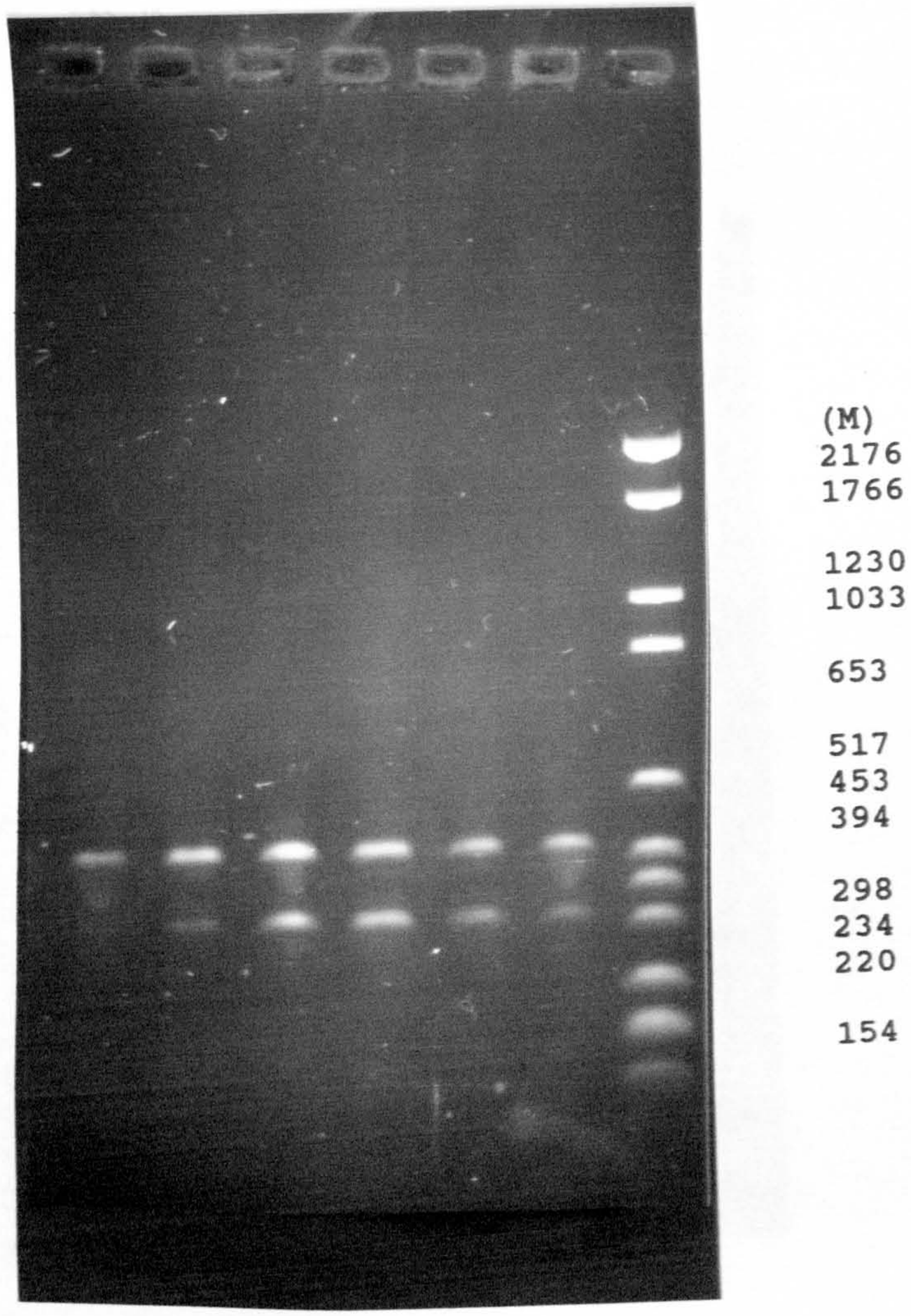
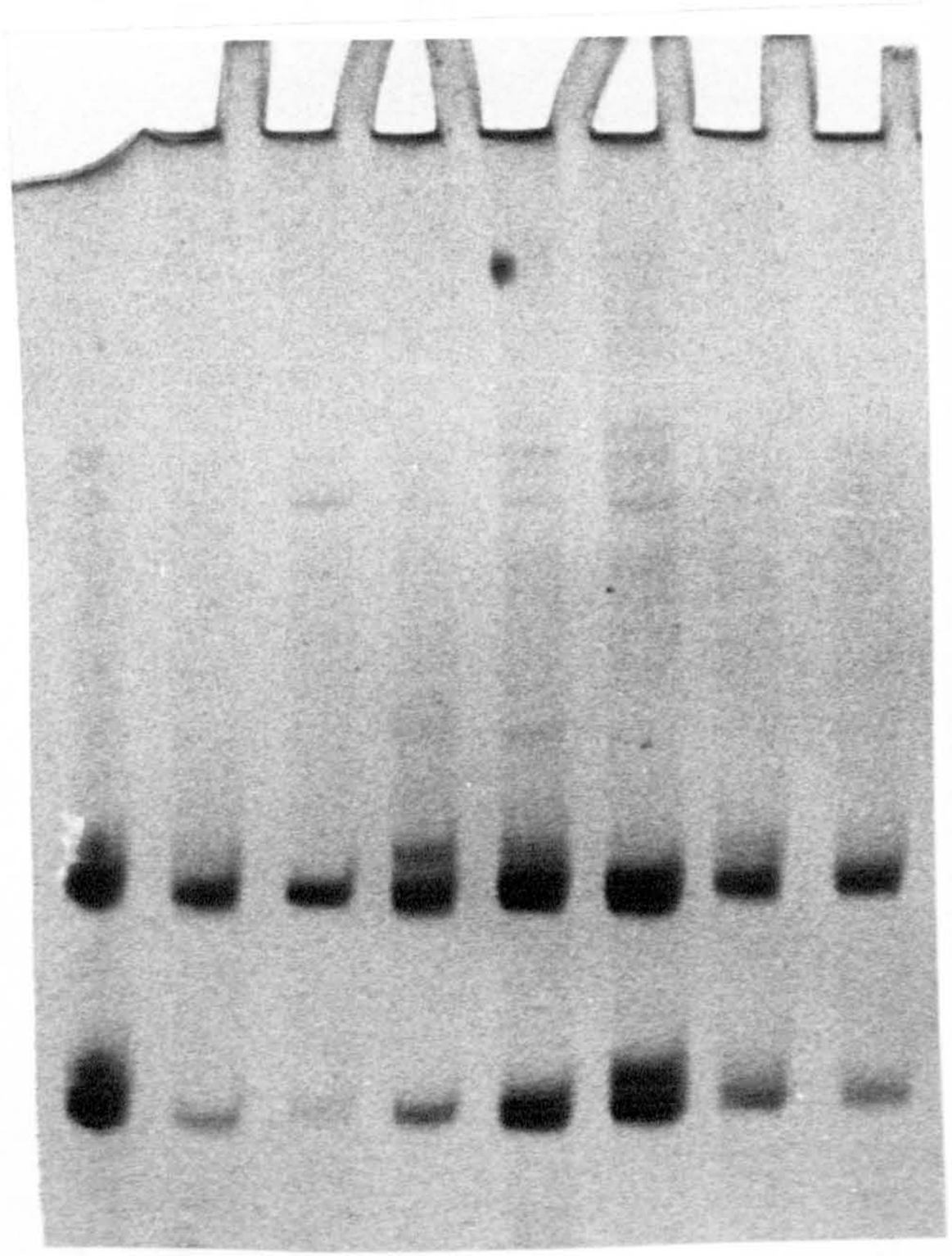


Fig. 4.13- Silver stained 6% acrylamide gel of RAPD primer AB19 (50ng- 3 steps annealing) products in six members of *An. gambiae* species complex



4.4.3 Intra-specific variation

An important character of RAPD is its capacity to define intra-specific variation, as has been shown in this project with other species complexes. In this way RAPD primers can be divided into two groups:

1- Those primers that produce species-specific markers in some species but uncover intra-specific variation within other members of the complex.

For example primer **UBC-305** produces some species-specific patterns but with a considerable diversity of bands between individuals of *An. merus* and *An. melas* (Fig. 4.14).

AB4 is another RAPD primer that in the *An. subpictus* complex identified species-specific markers (chapter 7), but in species of the *An. gambiae* complex showed considerable intra-specific variation, especially in individuals of *An. arabiensis* which cannot be distinguished from *An. merus*. However there are similarities in patterns of *An. melas*, *An. quadriannulatus* and *An. bwambae* (Fig. 4.15).

2- RAPD primers that reveal genetic variation within populations of *An. gambiae* members from Kenya, Zimbabwe and Uganda.

Fig. 4.14- RAPD primer UBC305 products in *An. gambiae* species complex: (1-3), *An. bwambae* (1-7), *An. arabiensis* (8), *An. gambiae* (10), *An. quadriannulatus* (11), *An. merus* (12-15), *An. melas* (16-18); Marker(9); -ve (19).

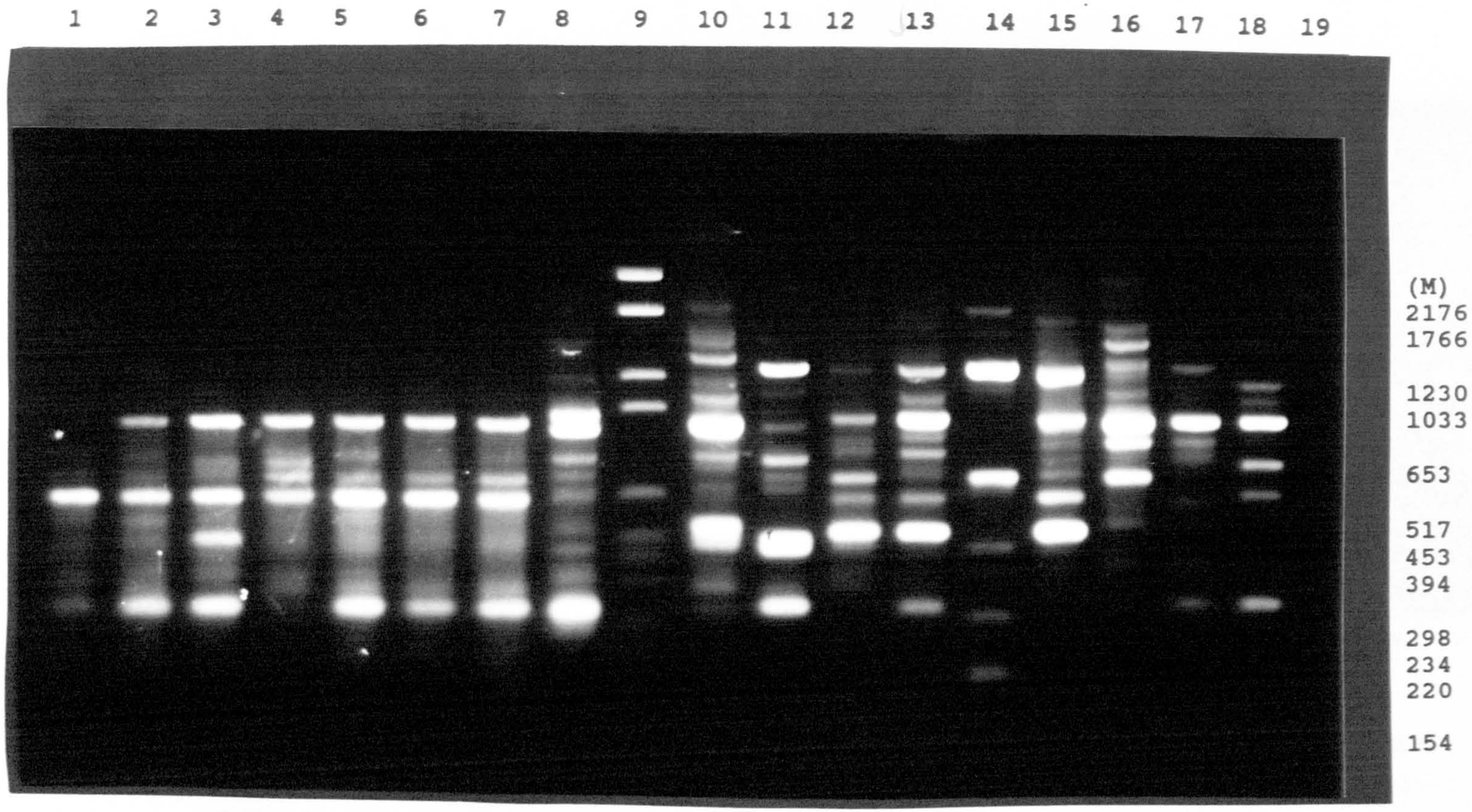
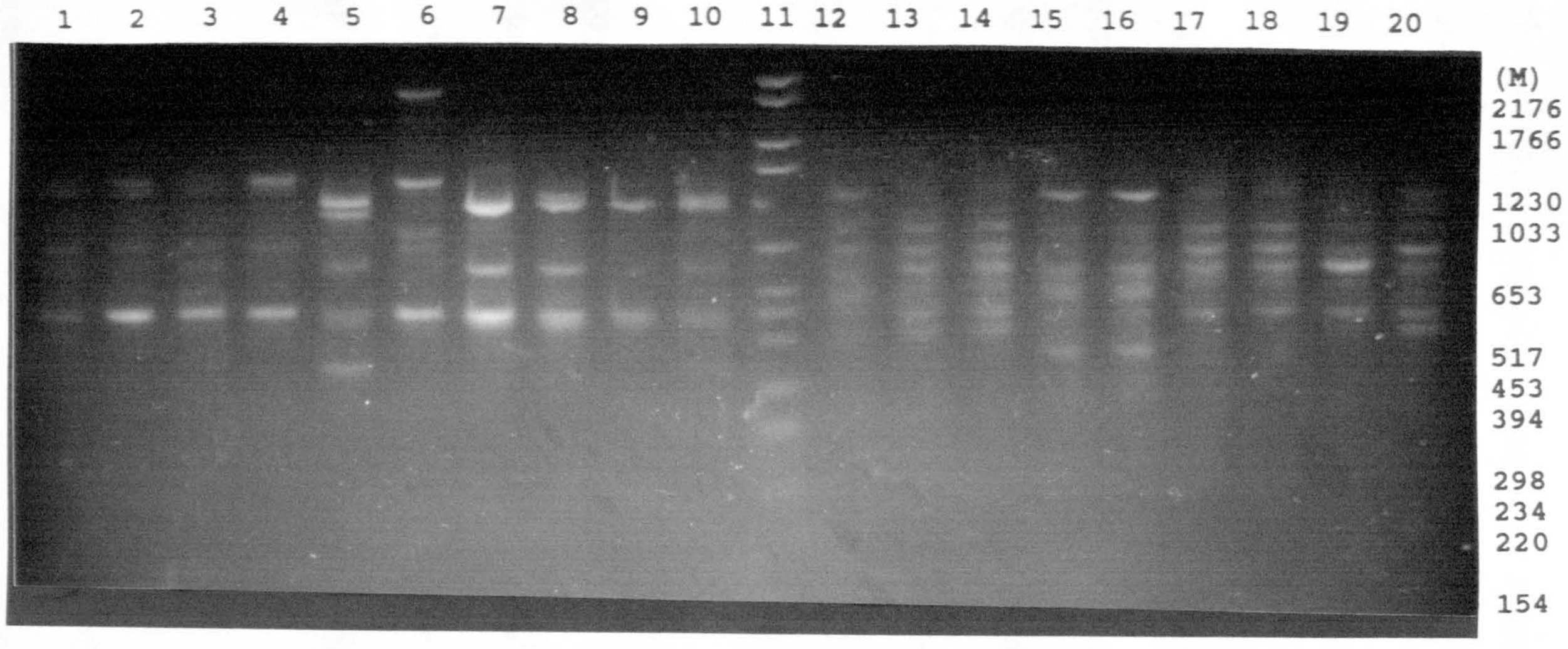


Fig. 4.15- RAPD AB4 primer products in *An. gambiae* species complex: *An. arabiensis* (1-7), *An. merus* (8-10), *An. melas* (12-14), *An. bwambae* (15-16), *An. quadriannulatus* (17-20).



Kenya:

Primers UBC-302, 308, 353, and 400 have been examined with specimens of *An. arabiensis* and *An. gambiae* from Kenya with a view to identify genetic variation within these species.

UBC-400

With primer UBC 400 both species share 280 and 450bp bands, and all *An. gambiae s.s.* specimens examined have a 345bp species-specific fragment which never amplified in *An. arabiensis*. However, of greater interest is the variation within species, especially in *An. gambiae s.s.* in which bands of 650, 620, 560 and 420bp may be present or absent (Fig. 4.16).

UBC-353 and **UBC-308**

Following their use with *An. stephensi*, two primers UBC-353 and UBC-308 have been selected to amplify the same individuals in each species. Both primers revealed variation within *gambiae* and *arabiensis* but this variation was more extensive in *An. arabiensis* specimens (Fig. 4.17, Fig. 4.18).

UBC-302

With UBC-302, DNA from the same individuals used with UBC-353 and UBC-308 from Kenya have been amplified. While *An. gambiae s.s.* specimens showed variation in the presence or absence of three main bands, amplification products in individuals of *An. arabiensis* showed extensive variation (Fig. 4.19).

Fig. 4.16- RAPD UBC-400 primer products in *An. gambiae* (left panel) and *An. arabiensis* (right panel). Although the two species have some common bands but *An. gambiae* s.s. has a 345bp fragment which never been amplified in *An. arabiensis* specimens.

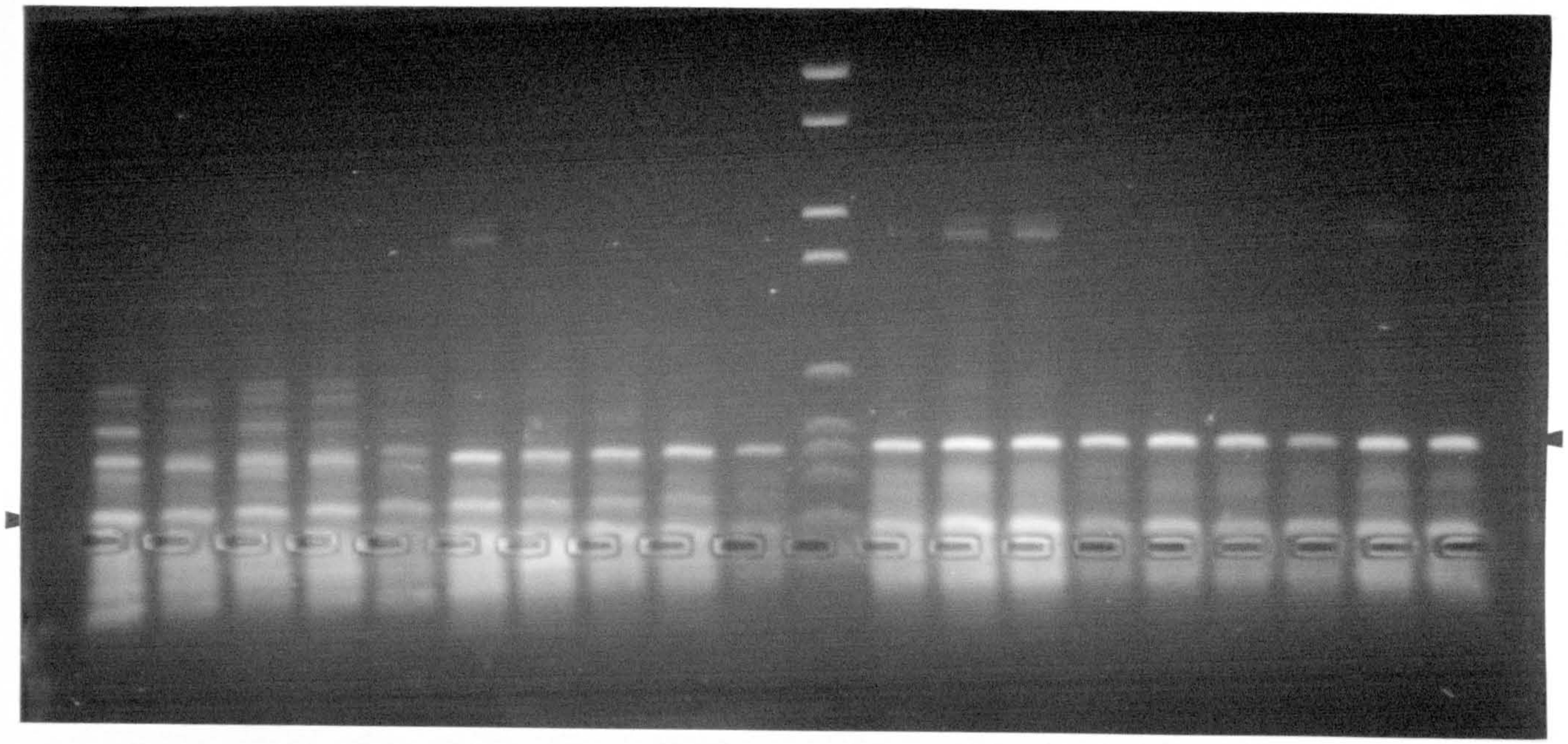


Fig. 4.17- RAPD UBC-353 (left panel) and UBC-308 (right panel) primers products in *An. gambiae s.s.* populations from Kenya which show intra-specific variation. It seems that UBC-308 products are more homogenous than UBC-353 primer in this species.

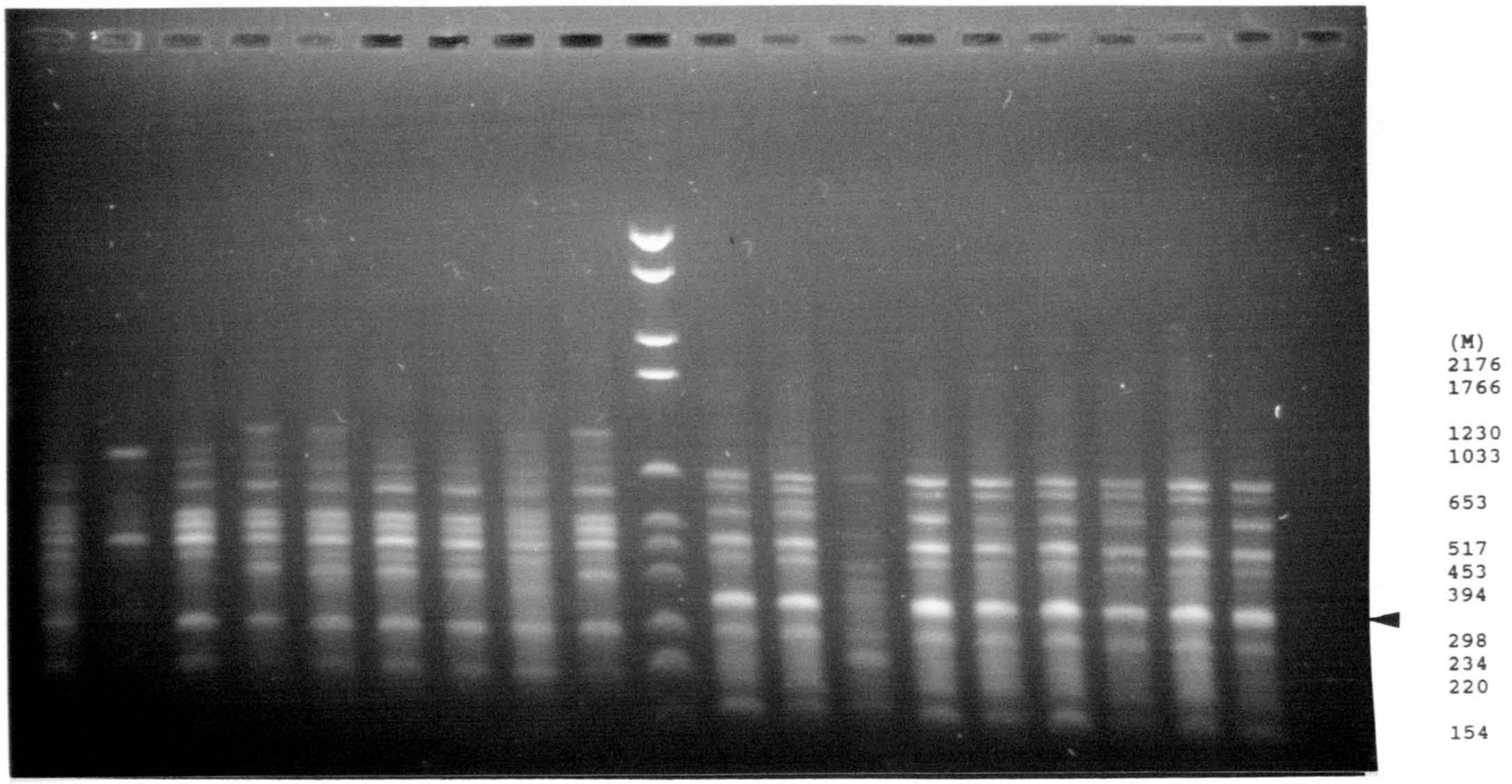
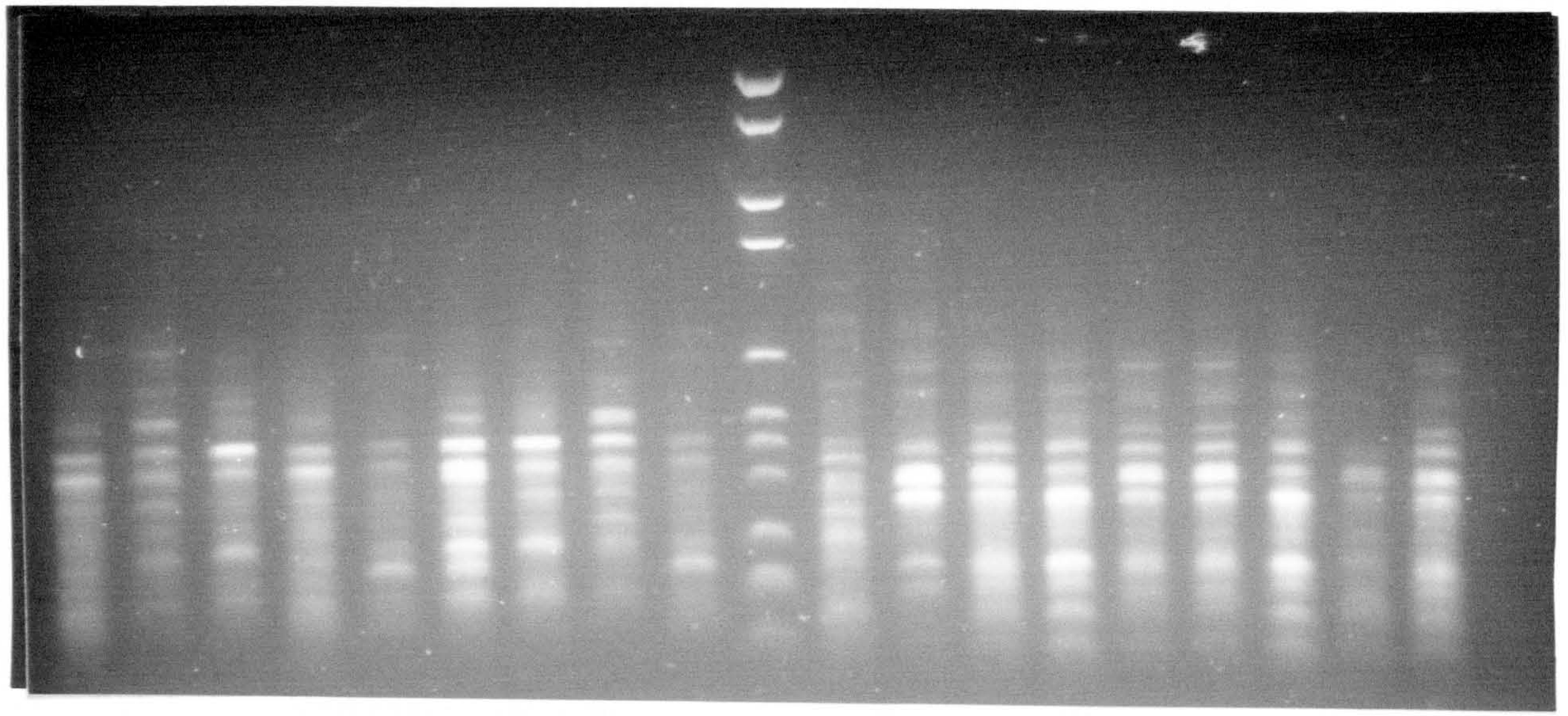


Fig. 4.18- RAPD UBC-353 (left panel) and UBC-308 (right panel) primers products in *An. arabiensis* populations from Kenya which show extensive intra-specific variation and in compare to *An. gambiae s.s.* populations.



Mozambique:

Correct identification of sympatric species in local populations of a species complex is very important in epidemiological study prior to application of any control measure. Specimens collected from Mozambique and identified as members of *An. gambiae* complex by colleagues using rDNA-IGS primers have been tested to evaluate the ability of RAPD primers to detect and distinguish mixed populations of *An. gambiae* species complex members.

AB19

In a single run, AB19 primer (70ng in 36°C annealing- Prog 1) detected the only *An. arabiensis* among the *An. merus* specimens (Fig. 4.20) by the presence of a 500bp band specific to *An. arabiensis*.

UBC-302

Following this experiment, DNA from the same specimens from Mozambique has been used for amplification with UBC-302. Again, the only *An. arabiensis* within that mixed field population has been detected, its PCR products being virtually indistinguishable from two laboratory stocks of *An. arabiensis* (Fig. 4.21).
In each of the above cases the identities of the mosquitoes were confirmed by rDNA PCR using the method of Scott *et al.* (1993).

Uganda

An. bwambae is the latest described species of the *An. gambiae* complex and there is little reported work on this mosquito. With M13 primer in a 36°C annealing

Fig. 4.19- RAPD-(Prog 1) primer UBC-302 amlification in *An. gambiae* (1-7) and *An. arabiensis* (9-16). As it found with other primers, *An. arabiensis* shows more variation within species, while individuals of *An. gambiae* are more homogenous molecular marker (8, 18); -ve control (17).

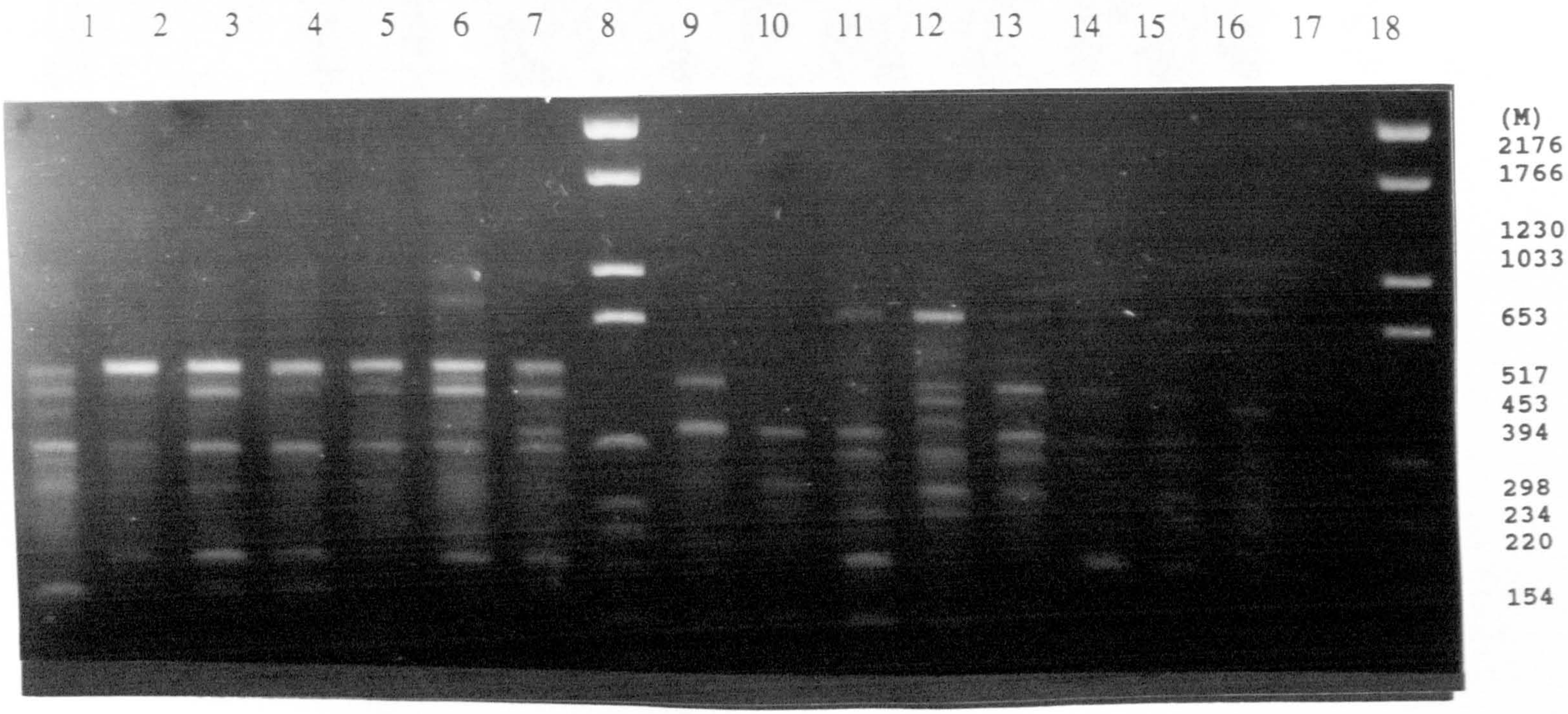


Fig. 4.20- RAPD primer AB19 (70ng- 36°C) products in a mixed population of *An. gambiae* species complex from Mozambique detected the only *An. arabiensis* (12) from 17 other *An. merus* (1-10, 13-19); Marker(11); -ve (20).

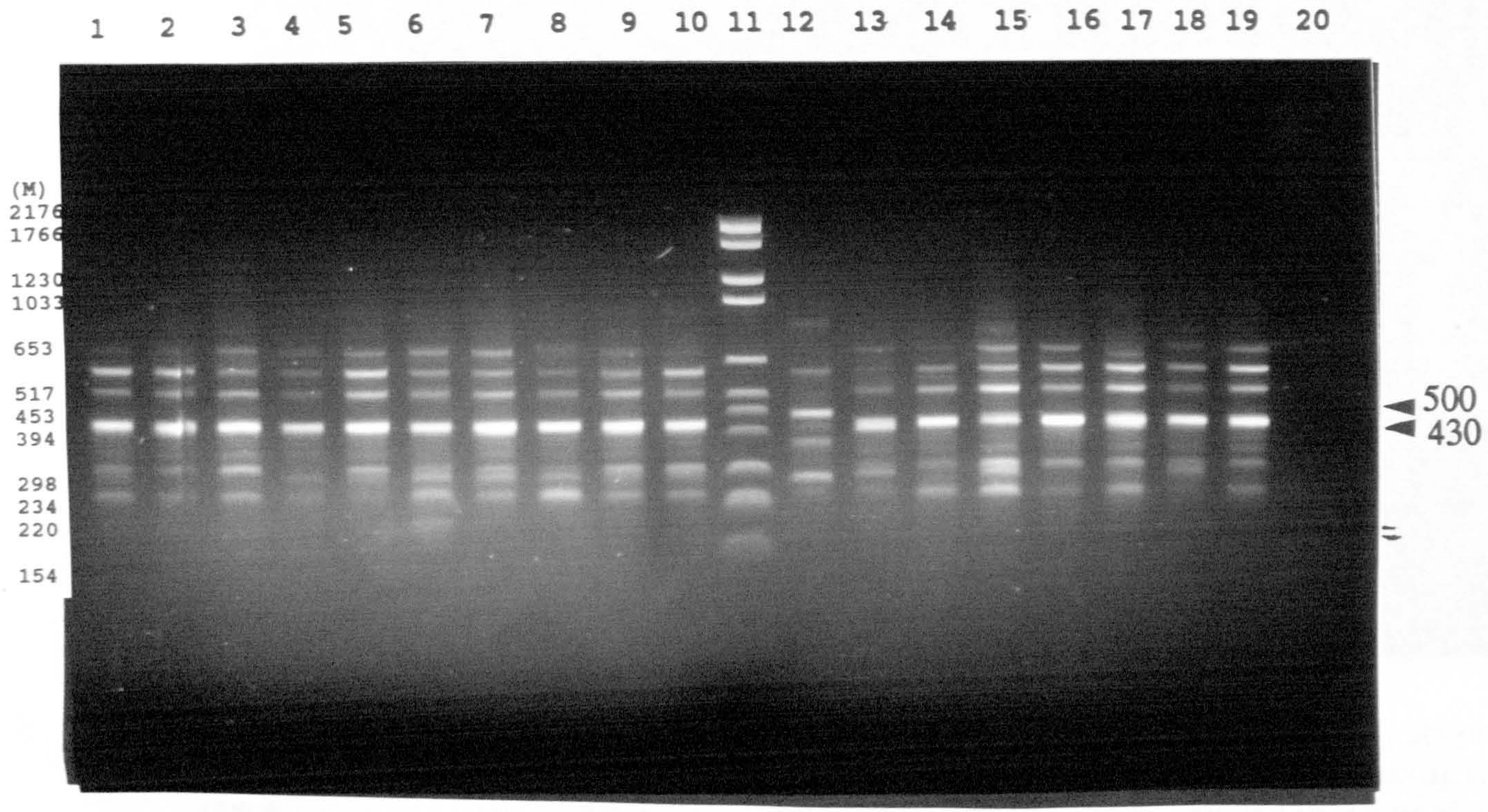
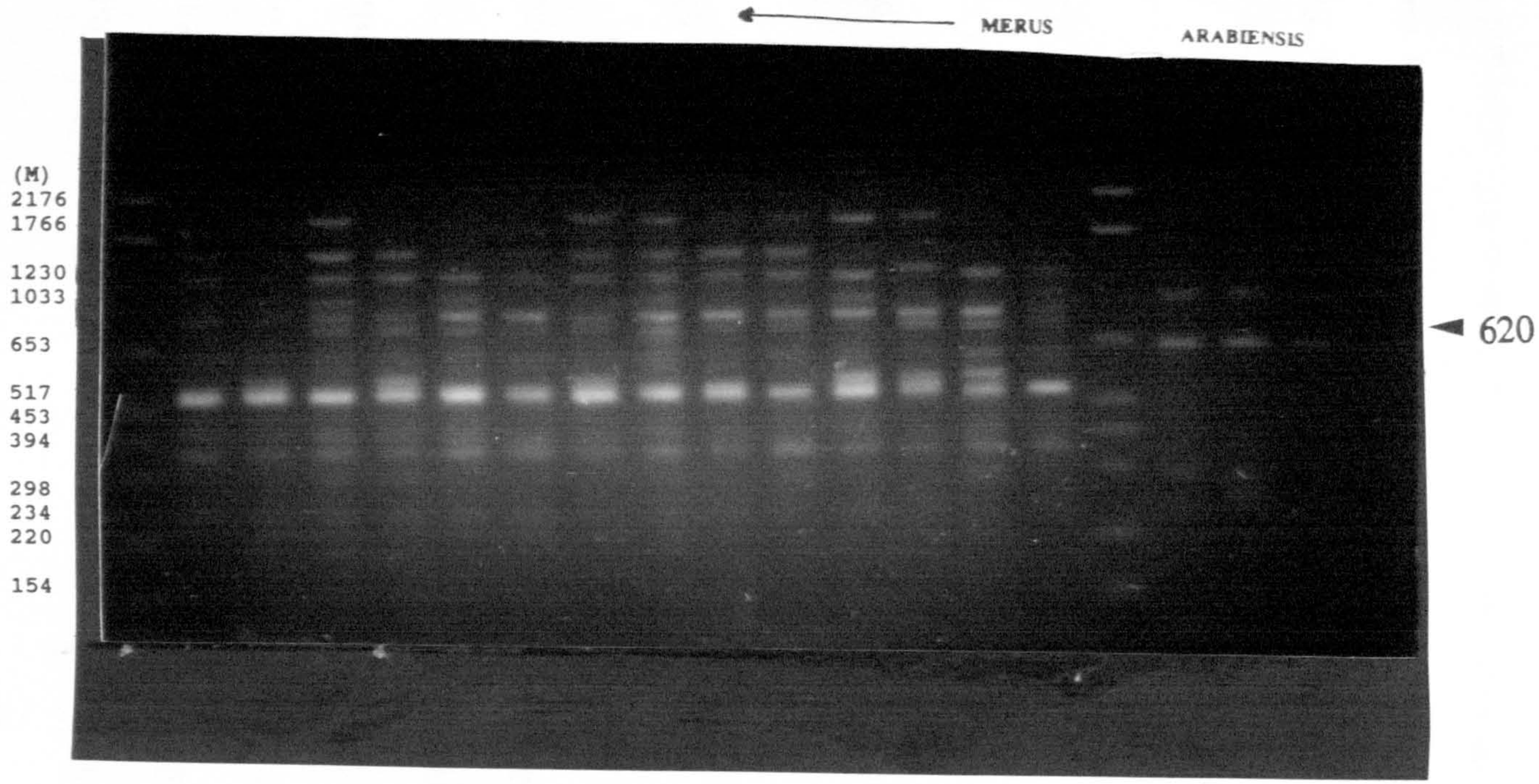


Fig. 4.21- RAPD primer UBC-302 products in a mixed population of *An. gambiae* species complex from Mozambique which detected the only *An. arabiensis* (right panel) from *An. merus* (left panel). Some intra-specific variation could be seen in *An. merus* population but the whole pattern is similar in compare to *An. arabiensis*.



program it was shown that field samples of *An. bwambae* collected by Ambrose Onapa from Uganda have a different pattern from *An. gambiae* and *An. arabiensis* (data not shown) and could be distinguished by RAPD-PCR.

RAPD identification of *An. bwambae*

M13

In order to evaluate the ability of RAPD-PCR to amplify dried specimens, M13 has been used to amplify the DNA from a series of *An. bwambae* specimens identified by an rDNA IGS-based primer developed by Townson, Besansky and Collins (Harbach et al. 1997). This primer successfully amplified all specimens (Fig. 4.22A). These were four main bands of 450, 550, 850 and 920bp. The consistency of amplification has been tested by re-amplification of one of the main bands (550bp)(Fig. 4.22B).

UBC-305

UBC-305 also produced specific PCR products (335, 610, and 950bp) with specimens of *An. bwambae* (Fig. 4.23) which distinquish it from other members of the complex.

OPA1

OPA1 also has been applied with the same specimens. Two bands of 380 and 530bp were present in all individuals (Fig 4.24), but some variation in individual patterns could be seen, for example fragments in 410, 435, and 710bp were present in some specimens, either reflecting real variation within species or perhaps failure

Fig. 4.22- RAPD M13 primer pattern in *An. bwambae* A)- dried specimens from Museum (1-6, 8-13), +ve control (field collected *An. bwambae* lane 15); marker (7); -ve (14) and B)- reampliication of one of those common bands (540bp) in a museum specimen.

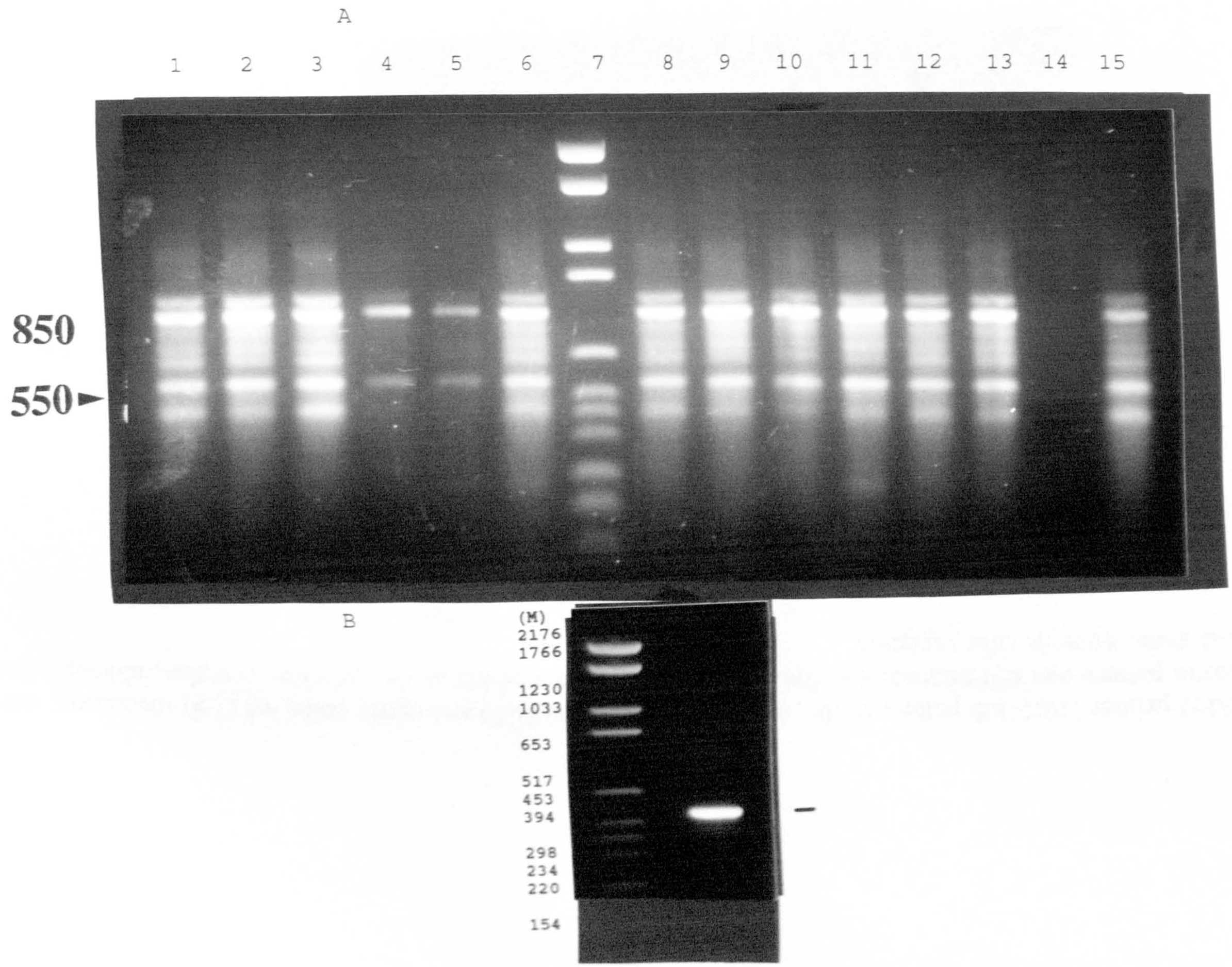


Fig. 4.23- RAPD primer UBC-305 pattern in *An. bwambae* from dried (1-5) and newly collected (7-9) specimens; marker (6). Diagnostic pattern can differentiate *An. bwambae* from other members of *An. gambiae* complex, although this primer revealed some intra-specific variation.

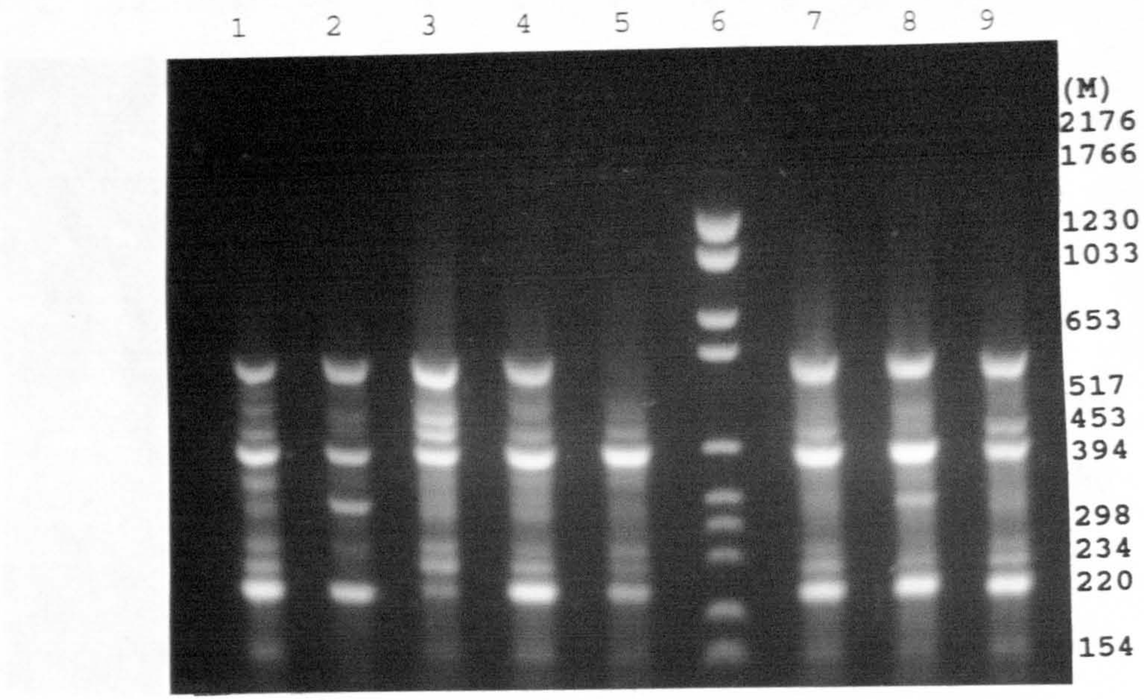
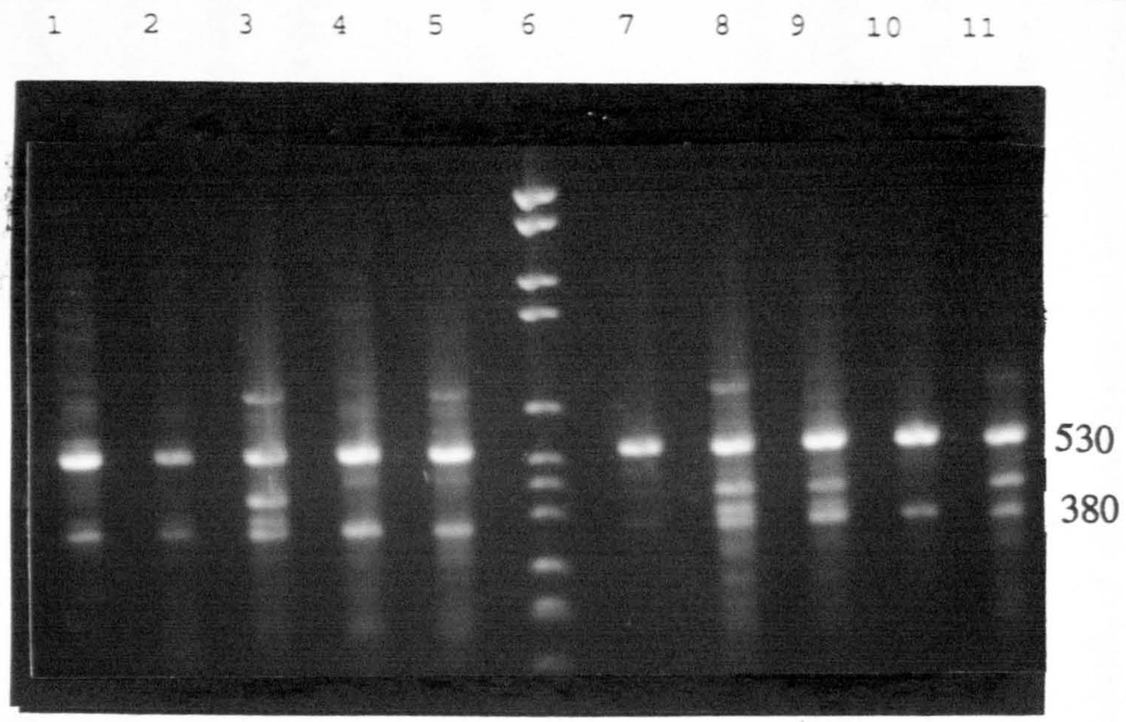


Fig. 4.24- RAPD OPA1 primer pattern in *An. bwambae* specimens from collection (1-5) and museum (7-11); marker (6). Some variation in individual pattern could be seen, although two bands of 380 and 530bp are present in all individuals.



in amplification.

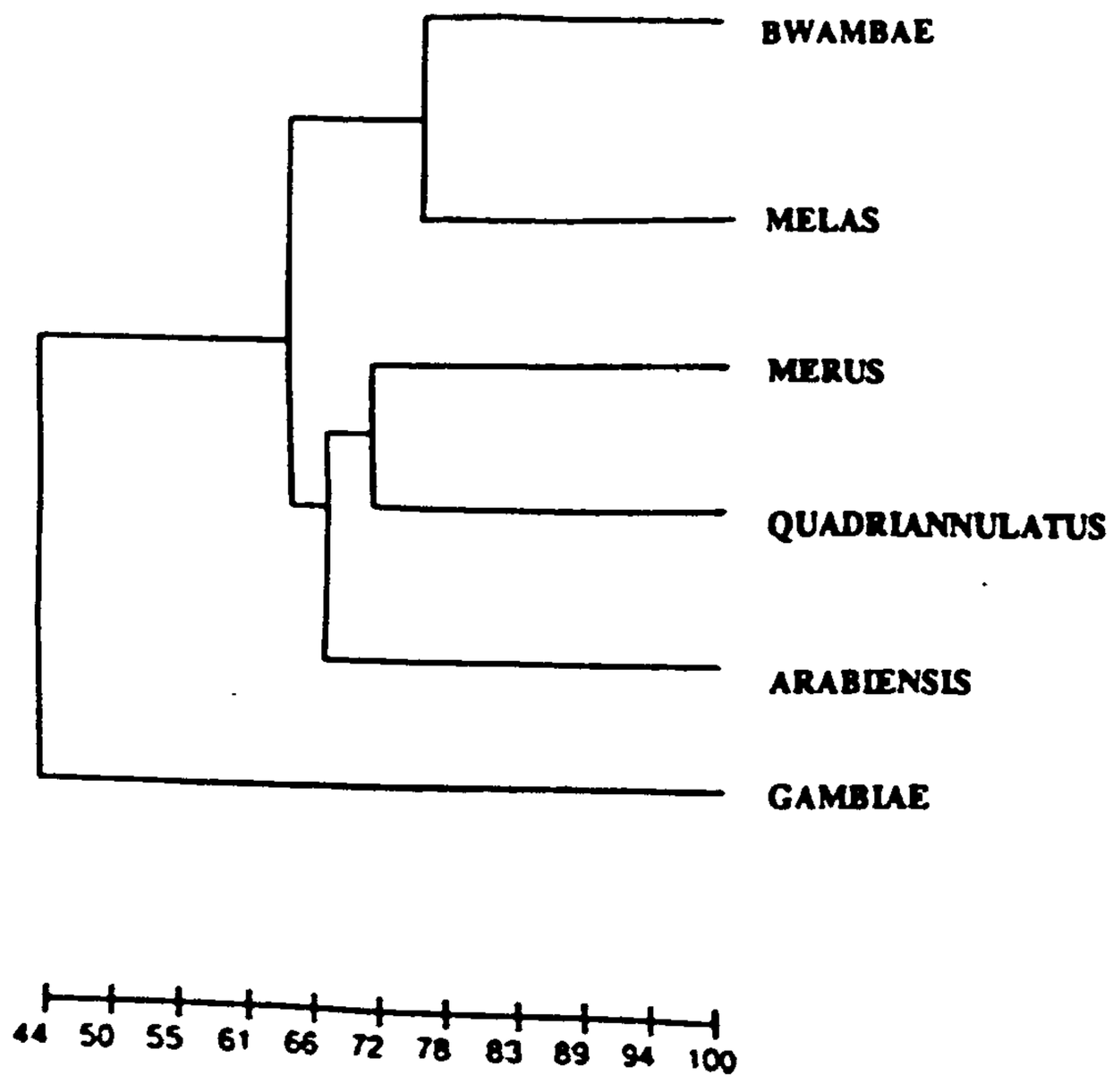
4.4.4 Distance measures and inferred phylogenies

All RAPD fragments amplified by primer M13F in different individuals and in different gels were scored and used for generating a dendrogram based on Average Linkage analysis (UPGMA) analysis. This clustered species *An. gambiae* of the complex into three main groups: 1)- *An. bwambae* and *An. melas*, 2)- *An. merus* and *An. quadriannulatus* with *An. arabiensis* and 3)- *An. gambiae* (Fig. 4.25A). The Dendrogram generated from M13R data revealed a different category within the complex, as *An. melas*, *An. gambiae* and *An. bwambae* are in one group and *An. quadriannulatus* with *An. arabiensis* in another, while *An. merus* is far from these two groups (Fig. 4.25B).

However, fragments generated by all primers were scored and used in calculation of distance matrices. Tables 4.3-4.6 summarised the distance matrices in six members of *An. gambiae* species complex using Jaccard, Dice, Nei, Montpellier distances and dismatch coefficient. Results of all indices are in concordance with each other. Using the results of all RAPD primers scored in these species on the basis of presence and absence of bands, phylogenetic trees were generated using the PHYLIP program (Fig. 4.26A, 4.26B).

Fig. 4.25- Dendrogram based on (A) M13F and (B) M13R primers fingerprints in all members of *An. gambiae* species complex

A)- M13F



B)- M13R

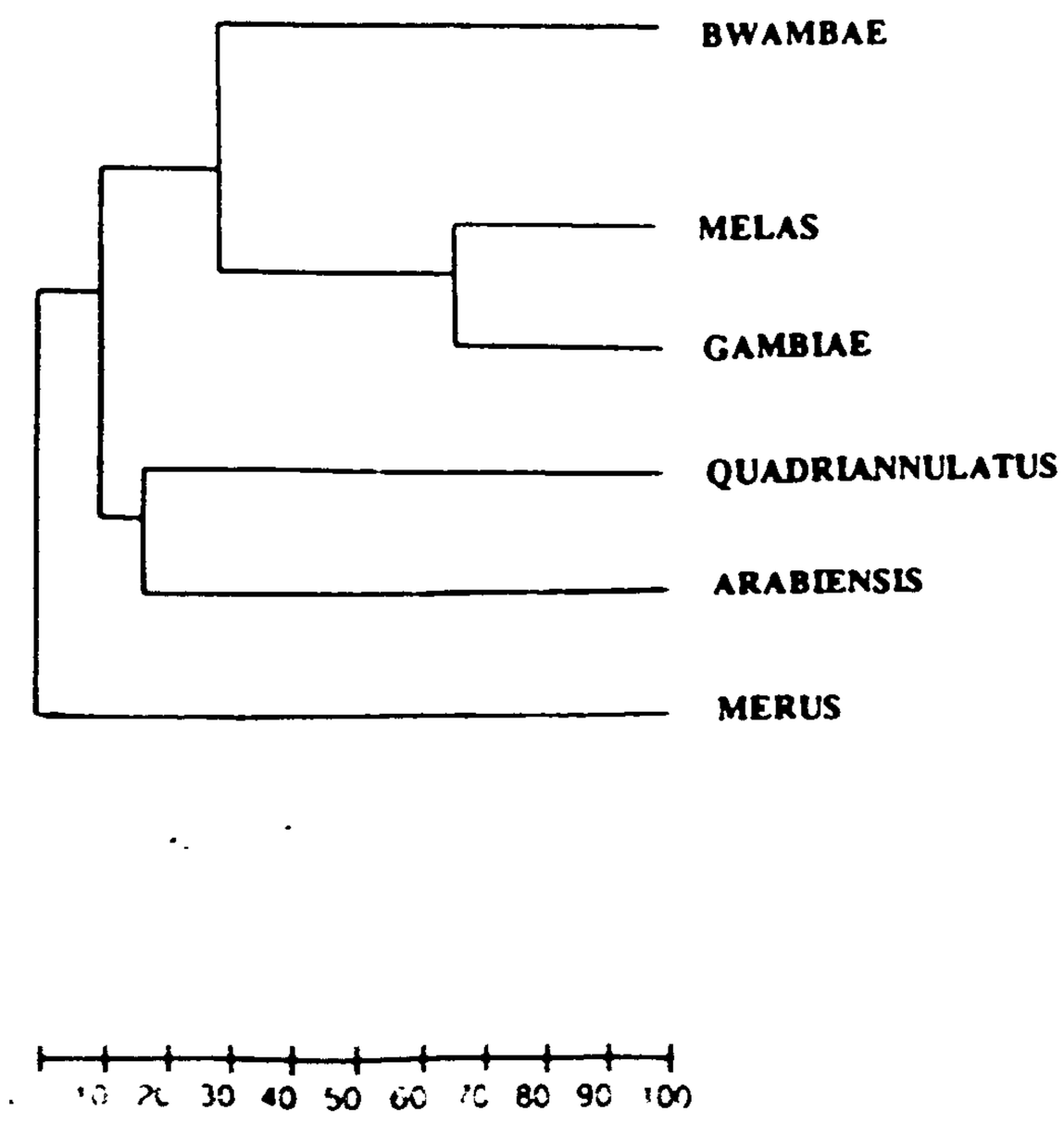


Table 4.3 - Jaccard distances (JSW = A / B+C, JCB= A / A+B+C) in all members of *An. gambiae* species complex

JSW\JCB	B	G	A	Mr	Ml	Q
B	---	0.80	0.73	0.71	0.71	0.67
G	0.86	---	0.63	0.73	0.80	0.67
A	0.78	0.68	---	0.63	0.75	0.61
Mr	0.79	0.78	0.63	---	0.58	0.42
Ml	0.81	0.85	0.78	0.98	---	0.75
Q	0.78	0.74	0.63	0.56	0.81	---

Table 4.4 - Dice distance (1-(2A / 2A+B+C)) and Nei & Li distance [1-($2N_{ab}$ / Na + N_b)] in all members of *An. gambiae* species complex

DIC\NEI	B	G	A	Mr	Ml	Q
B	---	0.66	0.57	0.56	0.54	0.49
G	0.67	---	0.48	0.57	0.66	0.49
A	0.57	0.49	---	0.43	0.60	0.43
Mr	0.56	0.58	0.46	---	0.41	0.27
Ml	0.54	0.66	0.60	0.41	---	0.59
Q	0.50	0.50	0.44	0.27	0.60	---

Table 4.5 - Simple Dismatch coefficient (1- S^M = 1-(A+D / A+B+C+D) and Montpelier distance: S_{Jm} = 1- |A / A+2(B+C)| in all members of *An. gambiae* species complex

SDM\MTP	B	G	A	Mr	Ml	Q
B	---	0.88	0.82	0.83	0.83	0.80
G	0.49	---	0.79	0.85	0.89	0.79
A	0.49	0.46	---	0.77	0.86	0.75
Mr	0.45	0.50	0.45	---	0.74	0.59
Ml	0.38	0.52	0.55	0.34	---	0.86
Q	0.37	0.52	0.42	0.23	0.48	

4.6-

Table - Distance matrices based on pair-wise comparisons of the proportion of shared, present and absent fragments in all members of *An. gambiae* species complex

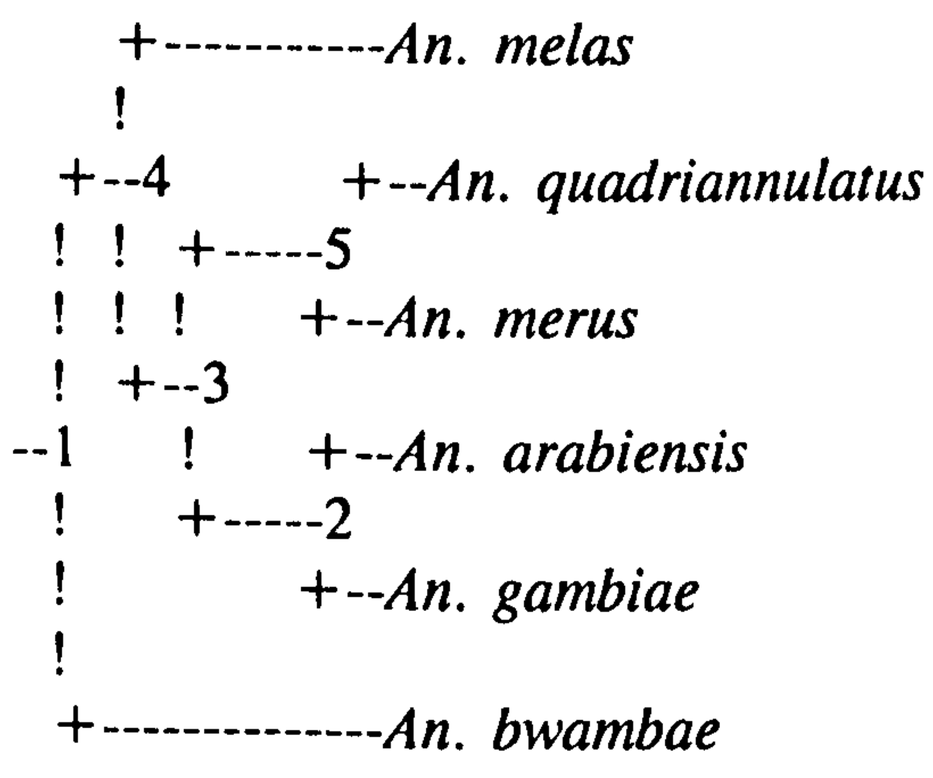
SP DIS	JACSW	JACIB	DICE	NEI	DISM	MONTP		X	X^2	$\bar{x}$	SIMIL	X_{n-}	ST.ER	XN
B–G	0.86	0.80	0.67	0.66	0.49	0.88		4.36	3.28	O.73	0.27	0 15	0 06	0 14
B–A	0.78	0.73	0.57	0.57	0.49	0.82		3.96	2.70	0.66	0.34	0.13	0.05	0 12
B–Mr	0 79	0 71	0.56	0.56	0.45	0.83		3.9	2.65	0.65	0.35	0.15	0.06	0 14
B–Ml	0.81	0.71	0.54	0.54	0.38	0.83		3.81	2.58	0.64	0.36	0.18	0.07	0.16
B–Q	0.78	0.67	0.50	0.49	0.37	0.80		3.61	2.32	0.60	0.40	0.17	0 06	0 16
G–A	0.68	0.63	0.49	0 48	0.46	0.79		3.53	2.17	0.59	0.40	0.13	0.05	0 12
G–Mr	0 78	0.73	0.58	0.57	0.50	0.85		4.01	2.78	0.67	0.33	0.14	0 05	0 13
G–Ml	0.85	0.80	0.66	0.66	0.52	0.89		4.38	3.30	0.73	0.27	0.14	0.05	0 13
G–Q	0.74	0.67	0.50	0.49	0.52	0.79		3.71	2.38	0.62	0.38	0.13	0.05	0 12
A–Mr	0.63	0.63	0.46	0.43	0.45	0.77		3.37	1.99	0.56	0.44	0.14	0.05	0.12
A–Ml	0.78	0.75	0.60	0.60	0.55	0.86		4.14	2.93	0.69	0.31	0.12	0 04	0 11
A–Q	0.63	0.61	0.44	0.43	0.42	0.75		3.28	1.89	0.55	0.45	0.14	0.05	0 12
Mr–M l	0 68	0.58	0.41	0.41	0.34	0.74		3.16	1.80	0.53	0.47	0.16	0.06	0 15
Mr–Q	0.56	0.42	0.27	0.27	0.23	0.59		2.34	1.04	0.39	0.61	0.16	0.06	0 14
Ml–Q	0 81	0.75	0 60	0.59	0.48	0.86		4.09	2.90	0.68	0.32	0.15	0 06	0 13

B *An bwambae*. G: *An gambiae*. A: *An. arabiensis*. Mr: *An. merus*. Ml: *An. melas*. Q: *An. quadriannulatus*

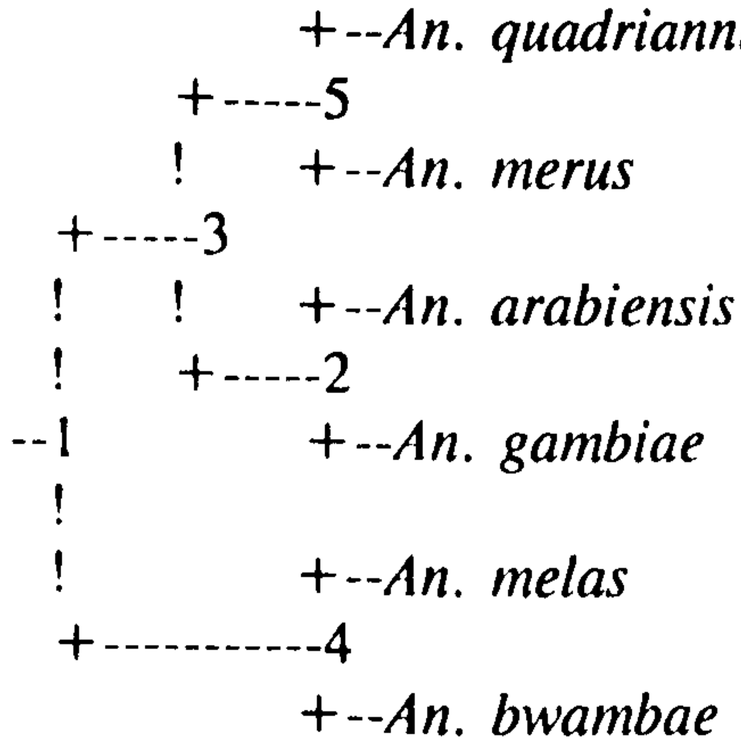
Jaccard distances (JACSW = A B+C. JACIB= A A+B+C). Dice distance (1-(2A / 2A+B+C)) and Nei & Li distance (1-(2N_{ab} Na + N_b)
Simple Dismatch coefficient (1 S^M = 1-(A+D A+B+C+D) and Montpelier distance: S_{Jm} = 1- [A A+2(B+C)]
SIMIL similarity. ST.ER: standard error

Fig. 4.26- Phylogenetic tree based on RAPD fingerprints in all members of *An. gambiae* species complex in PHYLIP program, Mixed parsimony algorithm, version 3.54c: (A) Wagner and (B) Camin-Sokal methods.

A)- Wagner parsimony method
One most parsimonious tree found:
requires a total of 110.000



B)- Camin-Sokal parsimony method
One most parsimonious tree found:
requires a total of 133.000



4.4.5 Characterization of RAPD products

Reasons for characterization of RAPD markers

Conventional PCR using specific designed primers is almost the last step in the application of a series of basic, time consuming and laborious techniques (including library construction, screening, cloning, sequencing, etc) which normally produce a single fragment amplified from a defined and known part of the genome.

In contrast, RAPD primers amplify usually a complex pattern of bands, from which, those of interest may be useful for further analysis. However, some RAPD primers can amplify single bands which offer the new sources of molecular data, available through additional steps such as sequencing, labelling, RFLP analysis,etc. since:

1- Their sequence is based on more reliable and reproducible amplifications.

2- These sequence data could be extended and applied in more than one species

3- These purified or re-amplified bands may be useful as probes in other related species.

On the other hand, those primers which produce complex patterns are not suitable for characterization, due to technical problems in separation and re-amplification of interested bands. However, these fingerprints could be used in finding similarities/distances and generating data for phylogenetic trees.

RAPD amplification using mixture of primers:

It has been found that a mixture of two RAPD primers is able to produce new amplification products not found with either of the two primers when used separately. Presumably, this in part arises from the two primers annealing respectively to the plus and minus strands of the template DNA. Although not all the products from such reactions can be readily explained, the use of RAPD primer pairs provides an additional tool for studying genetic diversity within and between species, without recourse to the synthesis of entirely new primers.

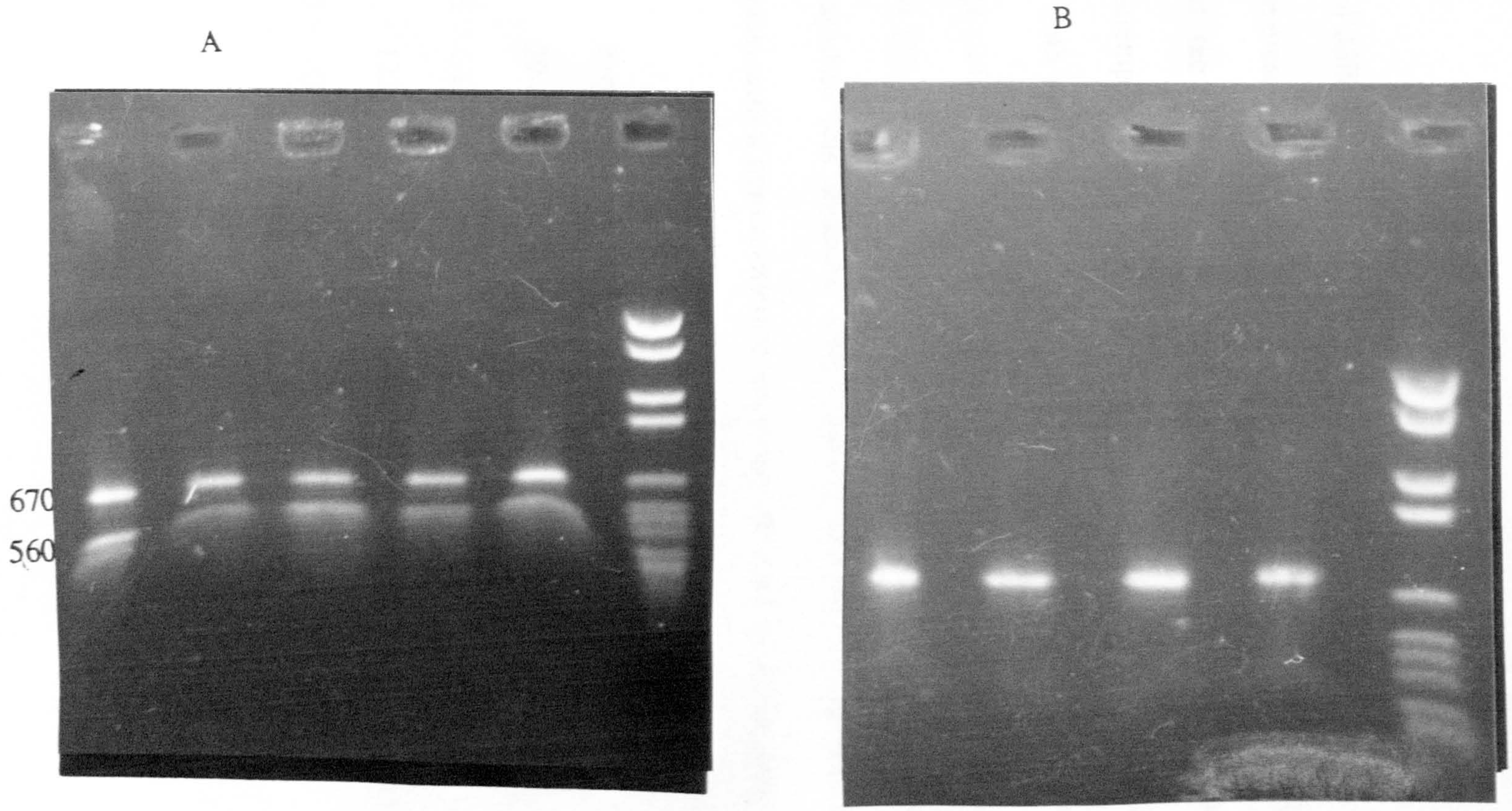
M13 F+R mixed primers

PCR products of M13F and M13R primers, specially the simple pattern of M13R proved useful in differentiation of the six members in *An. gambiae* complex. In order to find new markers, these two primers have been mixed in different amplification conditions as follows:

- A mixture of M13F+R (Prog 5) amplified a single 670bp band in all members of *An. gambiae* complex. Then this band was eluted and the DNA digested with AluI, EcoRI, and DraI restriction enzymes. AluI cut the 670bp into two bands, 560 and 110bp (this 110bp band does not appear on the gel) while DraI produced two 330 and 340bp bands (Fig. 4.27A and 4.27B). The restriction sites were common to all species of complex.

- A mixture of M13F+R in a 40 cycles program with a 50°C annealing temperature produced a 330bp fragment in *An. gambiae* species complex. However, some

Fig. 4.27- AluI (left panel) and DraI (right panel) digestion of M13 mix primers product in *An. gambiae* species complex. The 670bp fragment cut by AluI into two bands, 560 and 110bp.



quantitative differences of this band have been seen on normal agarose gel in members of complex. Acrylamide gel following silver-stain provided better resolution and revealed that this band consisted of three different size products in different species: 330bp in *An. quadriannulatus*, *An. merus*, *An. melas* and *An. bwambae*, 300bp in *An. gambiae* and double 300 and 330bp bands in *An. arabiensis* (Fig. 4.28A, 4.28B).

Characterization of 330bp common band in *An. merus*

The common presence of a 330bp band in species of the complex suggests that the same region of the genome is being amplified in these species. Automatic sequencing was attempted but failed probably because priming took place simulataneously at both ends of the fragment. Therefore the 330bp RAPD product from *An. merus* was cloned into the pCR™II cloning vector and sequenced using the Dye Deoxy method (Ch.3). Because of low resolution of the sequencing gel only 84bp of reliable sequence was obtained. This sequence showed a 77% homology (28/36 base, plus/plus) with a human esterase sequence (BLAST genbank data)

Species	**Region of homology**
An. merus	29-64

TTGGAAGAGACCAGTCAGCAGCAGTACCATGCTGGC

Homo sapiens	12388-12423

TTGGAAGAGACCCACCAGCAGGAGAGCCAGGATGGC

Fig. 4.28A- Silver stained 6% acrylamide of 330bp M13F+R (40oC- 45 cycles) product in An. gambiae (3-4), An. arabiensis (5-6), An. bwambae (7-8); marker (2,9), -ve (1)

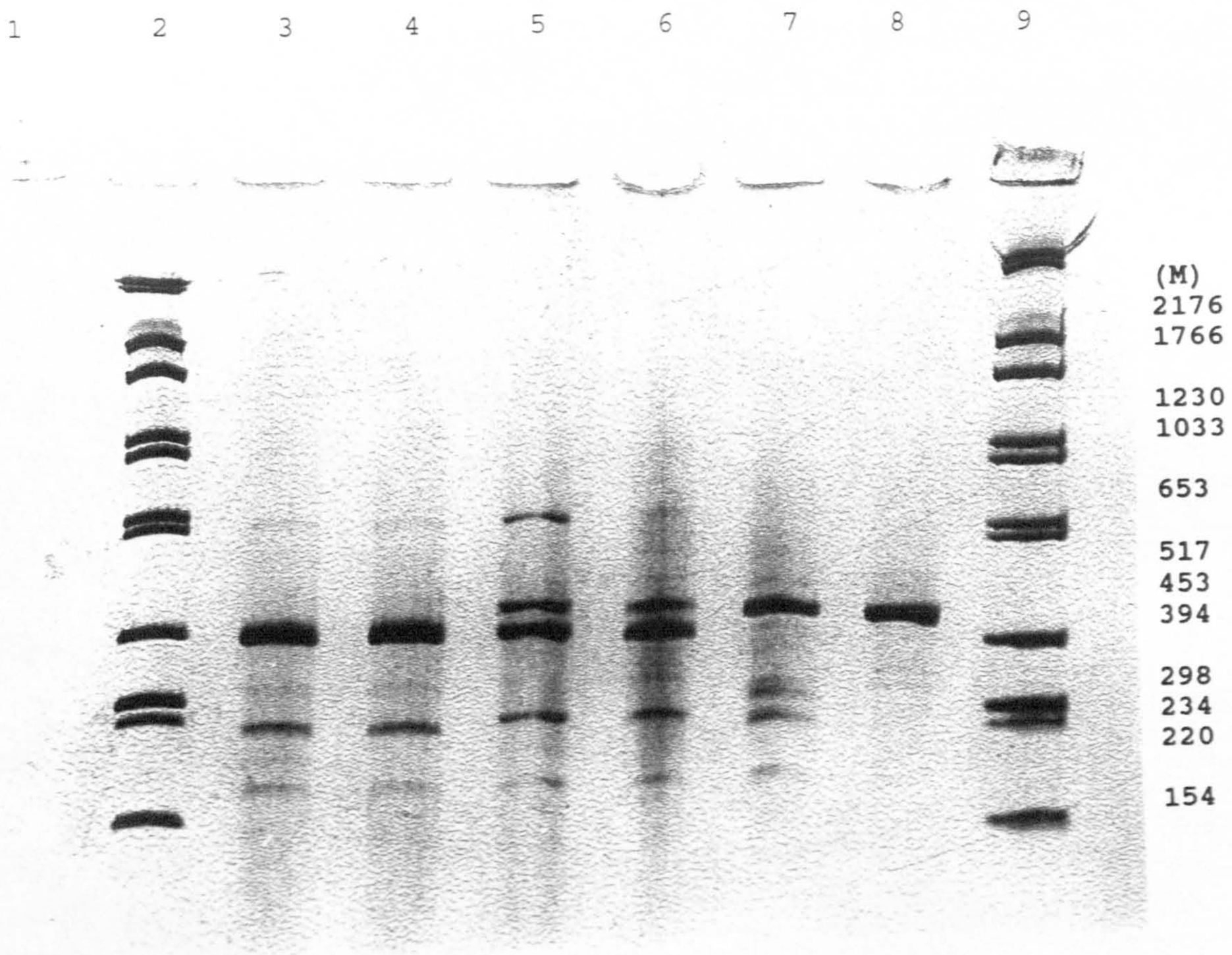
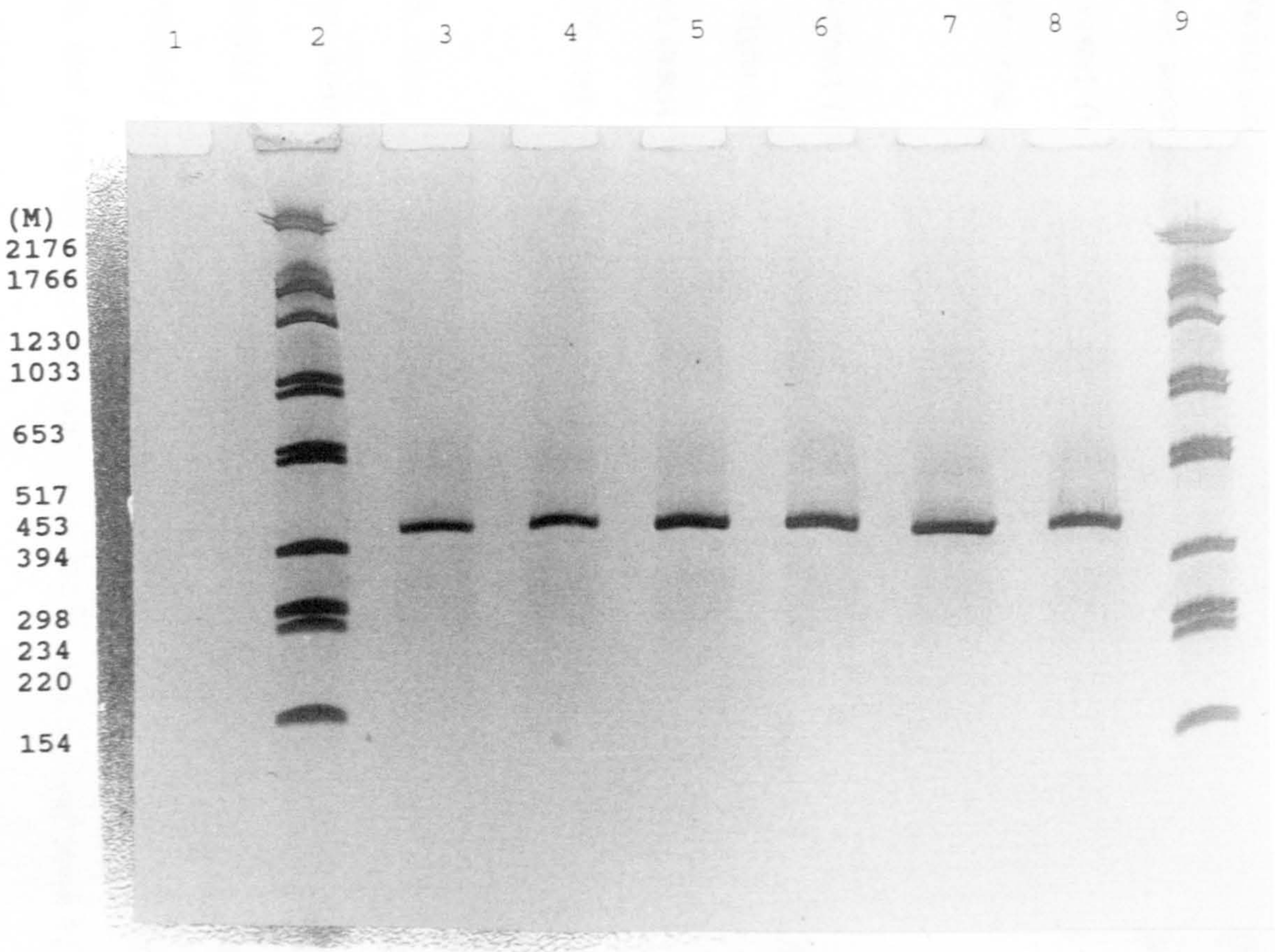


Fig. 4.28B- Silver stained 6% acrylamide of 330bp M13F+R (40oC- 45 cycles) product in An. quadriannulatus (3-4), An. merus (5-6), An. melas (7-8); marker (2,9), -ve (1)



Characterization of a RAPD fragment in *An. bwambae*

This part of the study is based on a previous result with a mix of M13 primers which showed a single band. The PCR product of a single 570bp band was excised from the gel and re-amplified (Fig. 4.29), purified by passing through the column (Pharmacia) and cloned into plasmid pCR™II (see Ch.3). Fig. 4.30 shows the plasmid which contains the insert (lanes 1,3) and the original 570bp PCR product (lane 5). Because the insert is in the EcoRI site of the plasmid, the size of insert was examined following EcoRI digestion of plasmid DNA from four clones. Surprisingly the insert size was about 420bp in two clones (Fig. 4.31, lanes 1,5) and other two did not show any insert (lanes 2,4) on 0.8% agarose gel. The reduced size of insert (420bp) may have one of three explanations:

1)- The PCR product (insert) has one or more EcoRI site

2)- It is not the right insert

3)- It is the right insert, but the original PCR product was a mixed product of two different size: 570 and 420bp.

The DNA from two clones (lanes 1,5 in Fig 31) was purified and because M13F+R primers were used to amplify the original PCR, T7 primers were used instead for automated sequencing (see Ch.3 for details -plasmid map). The trace sequence is completely clean and, starting just after T7 primer, includes 68bp of plasmid sequence, and then the insert starting with M13F and ending with M13R primer sequence, and after that 39bp of plasmid sequence. The total size of readable sequence is 607bp, of which 500 bp was the RAPD insert (Fig. 4.32, trace sequence).

Fig. 4.29- Reamplification of 600bp M13F+R mix product in An. bwambae which showed the same size of original PCR (3B and 4B) and has been used for cloning and susequent sequencing.

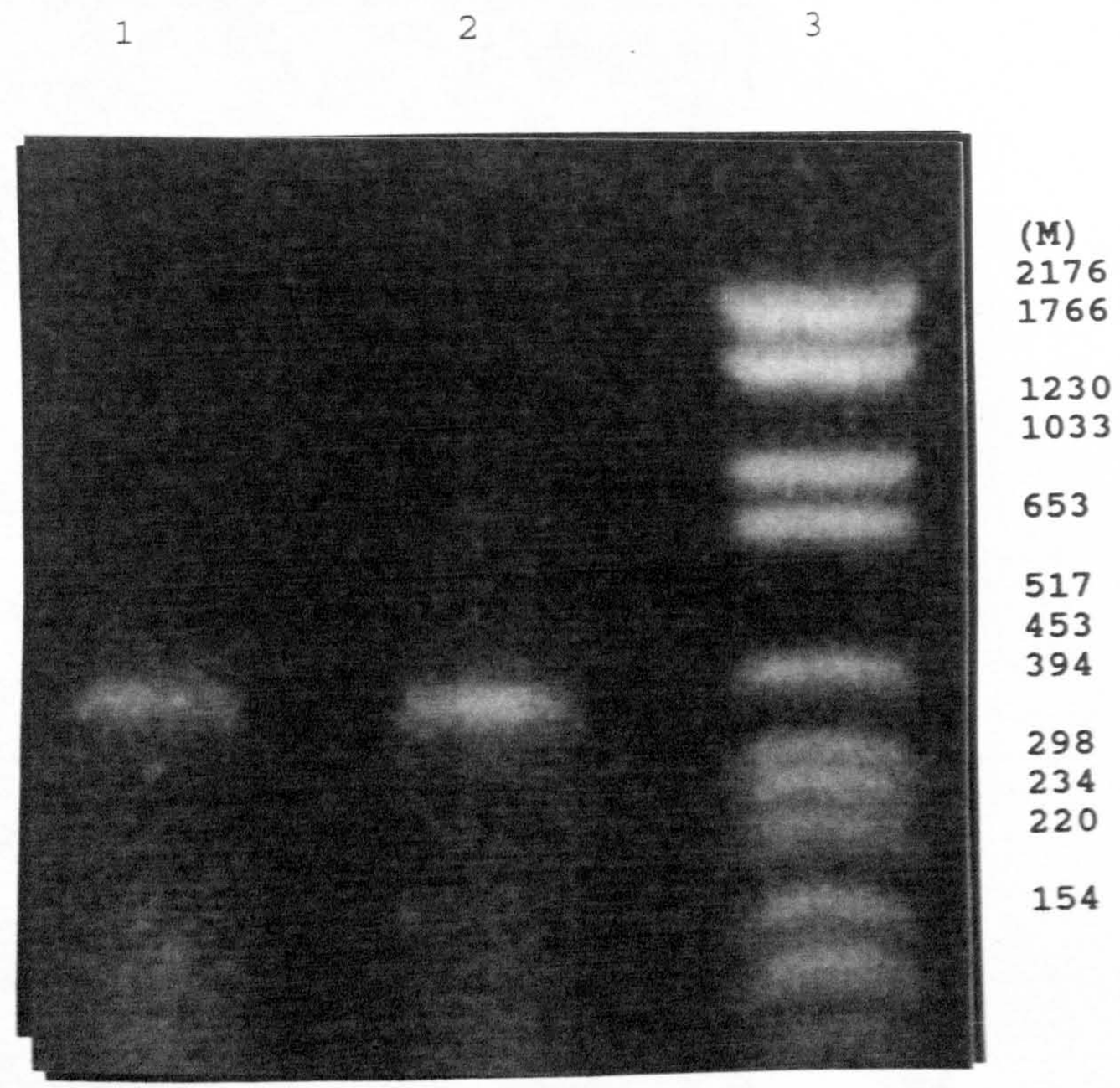


Fig. 4.30- Plasmid plus insert after cloning (1,3), original PCR product 600bp M13F+R mix product in *An. bwambae* (5); marker (2,4).

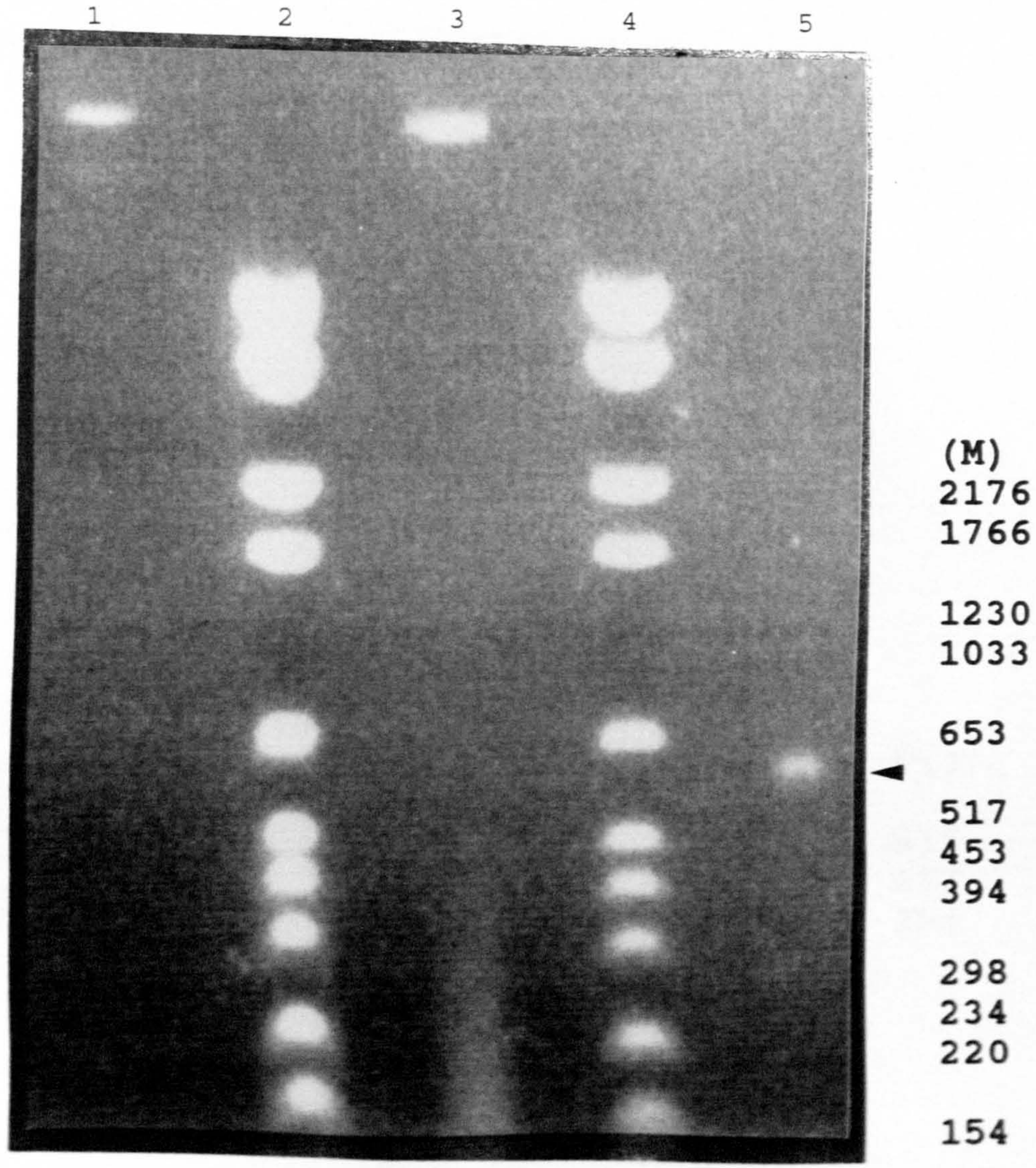
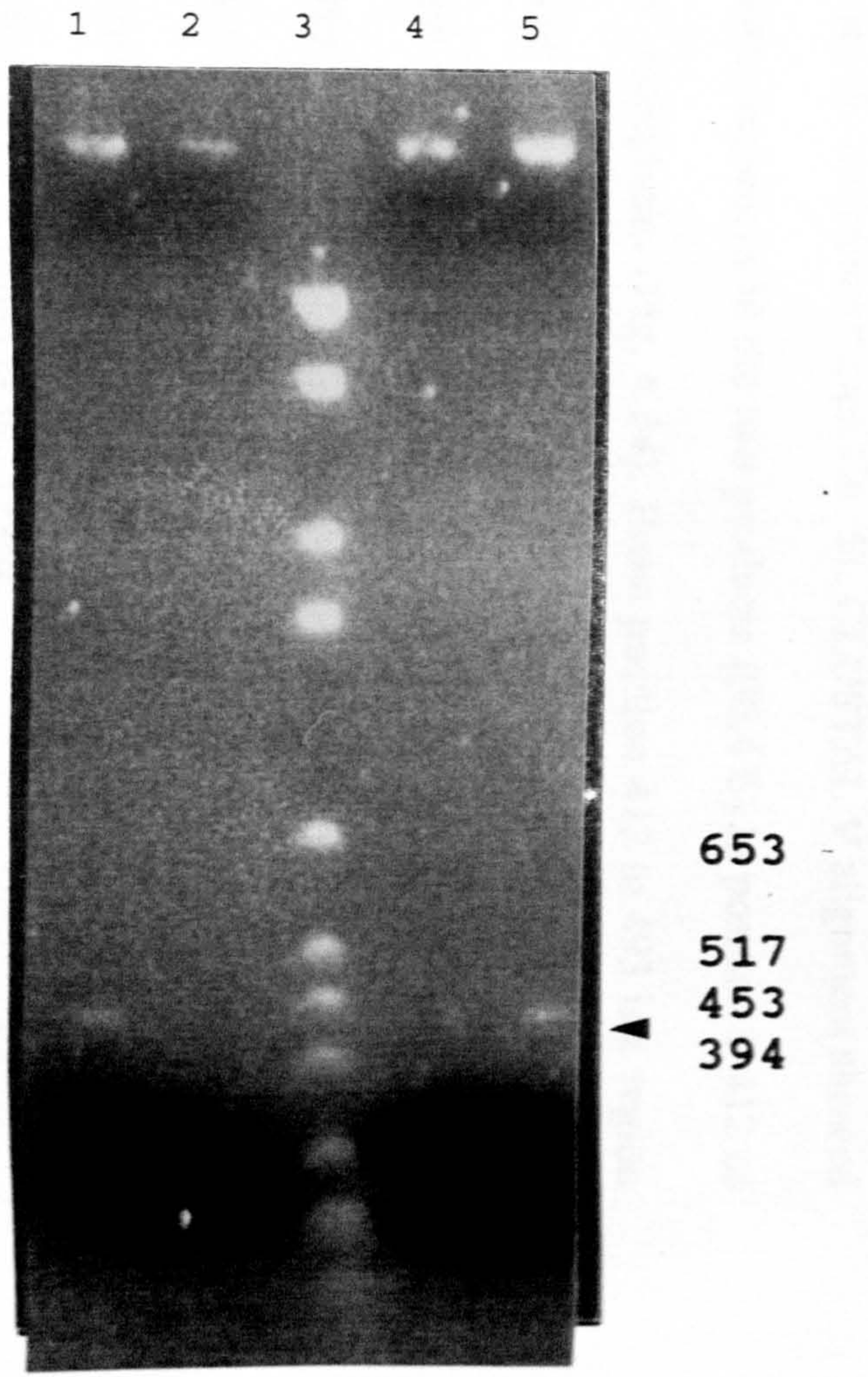


Fig. 4.31- EcoRI digestion of cloned 600bp M13F+R mix product in *An. bwambae* (1-2, 4-5); marker (3). Only lanes 1 and 4 showed that contain the insert and have been used for sequencing. Howeve, the size of the insert is about 420 (reasons in related text).



Based on this sequence an EcoRI site was found at position 418-424, and this provided the explanation for the reduced size of insert (420bp) after EcoRI digestion of the clone. Presumably the original PCR product had two EcoRI sites. Other enzymes that cut this 500bp fragment are BamHI, HindIII and SacI (Fig. 4.33)

Genbank homology between *An. bwambae* 570bp band and other sequences:

The main homology is with the sequence of 412bp common band produced by RAPD primer AB11 in *An. culicifacies* (see Ch. 5). CLUSTAL V alignment showed near identical nucleotide sequences in the two products (96.4%) in position 1-412 of *An. bwambae* and *An. culicifacies* (Fig. 4.34). From position 412 to 493 is a region showing homology with other published sequences (Fig. 4.35, Table 7). Starting at position 32 is a region showing homology with *Spisula solidissima* cyclin-specific ubiquitin carrier protein which its homology is starting from position 392. The same homology with ubiquitin-related genes has been found in the *An. culicifacies* 412bp common band in positions 351-396* in *C. elegans* TPEI, Human ubiquitin and the *An.gambiae* polyubiquitin gene (Collins *et al.* 1996).

The 96.4% homology in 412 bases of *An. bwambae* and *An. culicifacies* sequences may be explained by the fact that the AB11 primer sequence (GTAGACCCGT) is present in this fragment in both species in position 1-10 (forward) and 412-402 (reverse), even though, the band in *An. bwambae* is the amplification product of M13F+R primers.

Fig.4.32- Automated sequencing of *An. bwambae* clone. Total readable sequence is 607bp which starts just after T7 primer sequence and ends 42bp after M13R primer sequence.

1 CATTGGGCCC TAGATGCATG CTCGAGCGGC CGCCAGTGTG ATGGATACCT

51 GCAGAATTCG GCTTGTAGAC CCGTGAGTGG ATCCGGTGCG GTCGTGAATA

101 TTTATTGCCC GGCGAAGAAC AGTACGCGGA ATTGGCTTAA TCCCGTGTCA

151 CCGGACTGGC CTCACCATTA ATCATTGATC GGAGTGTAGT GAAACTGAAA

201 ATACACTTTA AAAAGTAGCT AGCGTACCGC GGCGTGTTTG TACTCGCTTG

251 CATGGGGAGC TGTAATGCAT CCCGCGGCCC ATCCACAGCA GTAAAGTCAC

301 CCGTAATGAA TGTTTAATGG CAGGAGCTCA CCAGATGGGA CACCGTTGAG

351 CTGATTGATG CTTTCTTGGG CGCGGAAGTT TGGTGTGTGT CTATTACTGG

401 CGTGTATGGA GAGTTTTGTT TGAAATGTTG TACTCGGTTG ATCTCGGTTG

451 AGTACGGTCG ATGGGACGGG TCTACAAGCC GAATTCCAGC ACTCTGGCGC

501 CGTTTCTAGT GGATCCGAGC TCGGTTCCAA GCTTGGCGTT AATCATGGTC

551 ATGCTTGTTT CCTGAAGCCG AATTCCATCA CTCTGGCGGG CCGTTACCTA

601 GTGGATC

Underlined: sequences of M13F and M13R

Fig 4.33 - BAMHI, SACI, ECORI and HINDIII are the four enzymes that cut the 500bp fragment of M13 mix primers product in *An. bwambae*

```
                    B
                    a
                    m
                    H
                    I
   GTAGACCCGTGAGTGGATCCGGTGCGGTCGTGAATATTTATTGCCCGGCGAaGAACAgTA
1  ---------+---------+---------+---------+---------+---------+ 60
   CATCTGGGCACTCACCTAGGCCACGCCAGCACTTATAAATAACGGGCCGCTtCTTGTcAT

   CGCGGAATTGGCTTAATCCCGTGTCACCGGACTGGCCTCacCATTAATCATTGATCCGAG
61 ---------+---------+---------+---------+---------+---------+ 120
   GCGCCTTAACCGAATTAGGGCACAGTGGCCTGACCGGAGtgGTAATTAGTAACTAGCCTC

   TGTAgTGAAACTGAAAATACaCTTTAAAAAGTAGCTAgCGTACCGCGGCGTGTTTGTACT
121 ---------+---------+---------+---------+---------+---------+ 180
   ACATcACTTTGACTTTTATGtGAAATTTTTCATCGATcGCATGGCGCCGCACAAACATGA

   CGCTTGCATGGGGAGCTGTaATGCATCCCGCGGCCCATCCACAGCAGTAAAGTCACCCGT
181 ---------+---------+---------+---------+---------+---------+ 240
   GCGAACGTACCCCTCGACAtTACGTAGGGCGCCGGGTAGGTGTCGTCATTTCAGTGGGCA

                        S
                        a
                        c
                        I
   aATGAATGTTTAATGGCAGGAGCTCacCAGATGGGACACCGTTGAGCTGATTGATGCTTT
241 ---------+---------+---------+---------+---------+---------+ 300
   tTACTTACAAATTACCGTCCTCGAGtgGTCTACCCTGTGGCAACTCGACTAACTACGAAA

   CTTGGGCGCGGAAGTTTGGTGTGTGTCTATtACTGGCGTGTATGGAgAGTTTTGTTTGAA
301 ---------+---------+---------+---------+---------+---------+ 360
   GAACCCGCGCCTTCAAACCACACACAGATAaTGACCGCACATACCTcTCAAAACAAACTT

                                                        E
                                                        c
                                                        o
                                                        R
                                                        I
   ATGTTGTACTCgGTTGATCTCGGTTGAGTACgGTCGATGGGACgGGTCTaCAAGCCGAAT
361 ---------+---------+---------+---------+---------+---------+ 420
   TACAACATGAGcCAACTAGAGCCAACTCATGcCAGCTACCCTGcCCAGAtGTTCGGCTTA

                                                H
                                                i
                    B                           n
                    a               S           d
                    m               a           I
                    H               c           I
                                    I       I   I
   TCCagCACtCTGGCGCCGTTTCtAGTGGATCCGAgCTCGGTtCCAAgCTTGCCgTTAATC
421 ---------+---------+---------+---------+---------+---------+ 480
   AGGtcGTGaGACCGCGGCAAAGaTCACCTAGGCTcGAGCCAaGGTTcGAACCGcAATTAG

   ATGGTCATgCTTGTTTCCtG
481 ---------+---------+ 500
   TACCAGTAcGAACAAAGGaC
```

Fig.4.34- CLUSTAL V multiple alignment of *An. bwambae* 500bp with *An. culicifacies* 412bp sequences which showed more than 96% homology.
B: *An. bwambae*, C: *An. culicifacies*.

```
B       GTAGACCCGTGAGTGGATCCGGTGCGGTCGTGAATATTTATTGCCCCGCCG
C       GTAGACCCGTGAGTGGATCCGGTGCGGTCGTGAATATTTATTGCCCCGCCG
        **************************************************

B       AAGAACAGTACGCGGAATTGGCTTAATCCCGTGTCACCGGACTGGCCTCA
C       ATGAACAGTACGCGGAATTGGCTTAATCCCGTGTCACCGGACTGGCCTCA
        * **********************************************

B       CCATTAATCATTGATCGGAGTGTAGTGAAACTGAAAATACACTTTAAAAA
C       TCATTAATCATTGATCGGAGTGTAGTGAAACCGAAAATACACTTTAAAAA
         ********************************* ********************

B       GTAGCTAGCGTACCGCGGCGTGTTTGTACTCGCTTGCATGGGGAGCTGTA
C       GTAGCTAGCGTACCGCGGCGTGTTTGTACTCGCTTGCATGGGGAGCTGTA
        **************************************************

B       ATGCATCCCGCGGCCCATCCACAGCAGTAAAGTCACCCGTAATGAATGTT
C       ATGCATCCCGCGGCCCATCCACAGCAGTAAAGTCACCCGTAATGAATGTT
        **************************************************

B       TAATGGCAGGAGCTCACCAGATGGGACACCGTTGAGCTGATTGATGCTTT
C       TAATGGCAGGAGCTCAACAGATGGGACACCGTTGAGCTGATTGATGCTTT
        ****************** **********************************

B       CTTGGGCGCGGAAGTTTGGTGTGTGTCTATTACTGGCGTGTATGGAGAGT
C       TTTGGGCGCGGAGGTTTGGTGTGTGTCGATAACTGGCGTGTATGGAGAGT
         *********** **************** ** ********************

B       TTTGTTTGAAATGTTGTACTCGGTTGATC-TCGGTTGAGTACGGTCGATG
C       TTTGTTTGAAATGTTGTACTCGGTTGATCCTCGGTTGAGTACGGCCGATG
        ********************************* *************** *****

B       GGACGGGTCTACAAGCCGAATTCCAGCACTCTGGCGCCGTTTCTAGTGGA
C       GGACGGGTCTAC----------------------------------------
        ************

B       TCCGAGCTCGGTTCCAAGCTTGGCGTTAATCATGGTCATAGCTGTTTCCT
C       ----------------------------------------------------

B       G
C       -
```

Fig4.35- Sequence of *An. bwambae* 500bp band with the regions which showed almost all homology with other published sequences (underlined and in bold) and also primers sequences (in double underlined)

AB11 primer

GTAGACCCGT GAGTGGATCC GGTGCGGTCG TGAATATTTA TTGCCCGGCG

AAGAACAGTA CGCGGAATTG GCTTAATCCC GTGTCACCGG ACTGGCCTCA

CCATTAATCA TTGATCGGAG TGTAGTGAAA CTGAAAATAC ACTTTAAAAA

GTAGCTAGCG TACCGCGGCG TGTTTGTACT CGCTTGCATG GGGAGCTGTA

ATGCATCCCG CGGCCCATCC ACAGCAGTAA AGTCACCCGT AATGAATGTT

TAATGGCAGG AGCTCACCAG ATGGGACACC GTTGAGCTGA TTGATGCTTT

CTTGGGCGCG GAAGTTTGGT GTGTGTCTAT TACTGGCGTG TATGGAGAGT

TTTGTTTGAA ATGTTGTACT CGGTTGATCT CGGTTGAGTA CGGTCGATGG

GACGGGTCTA CAAGCCGAAT TCCAGCACTC TGGCGCCGTT TCTAGTGGAT

CCGAGCTCGG TTCCAAGCTT GGCGTTAATC ATGGTCATAg CTGTTTCCTG

M13R primer

Table 4.7 summarises the results of a BLAST search for homology of 500bp of the *An. bwambae* 570bp band with other published and reported genes; these homologies varied between 94-77% . The sequence of *Anopheles albimanus* clone Que 13 showed 22.8% homology with the *An. bwambae* (500bp) and *An. culicifacvies* (410bp) sequences. Q13 is a transposon derived sequence.

4.4.6 Crossing experiment:

A cross between two strains of *An. gambiae* s.s., KWA and ZAND (reciprocal), was carried out but only a few of the F1 individuals from each cross survived. DNA from individual F1 progeny was amplified with primer UBC 309. The F1 progeny carried the main bands of both parents plus one main non-parental band (1000bp). In some larvae there was no amplification of bands which were present in pupae, adults and parents. Also, the size of one of the main bands (420bp) in female parent and adults is higher than in larvae (410bp). This band in pupae revealed dimorphism and showed both 410 and 420bp sizes (Fig. 4.36).

4.4.7 rDNA PCR

1)- rDNA with field materials:

In most cases, field specimens of the *An. gambiae* complex used in this study were identified to species using the method of Scott et al. (1993) either by myself or by those providing the specimens (see Ch.3). In the case of one specimen from Kenya rDNA PCR amplified both *An. gambiae* and *An. arabiensis* diagnostic bands (Fig. 4.37), suggesting that this individual may have been a hybrid.

Fig. 4.36- RAPd UBC-309 primer products in cross experiment of *An. gambiae s.s.* strains: F1 Female (19) of ZAND (Female) X KWA (Male) X F1 Male (20) of KWA (Female) X ZAN (Male) and their F1 progenies (larvae: 2-5, pupae: 6-10, adults: 12-18); marker (11).

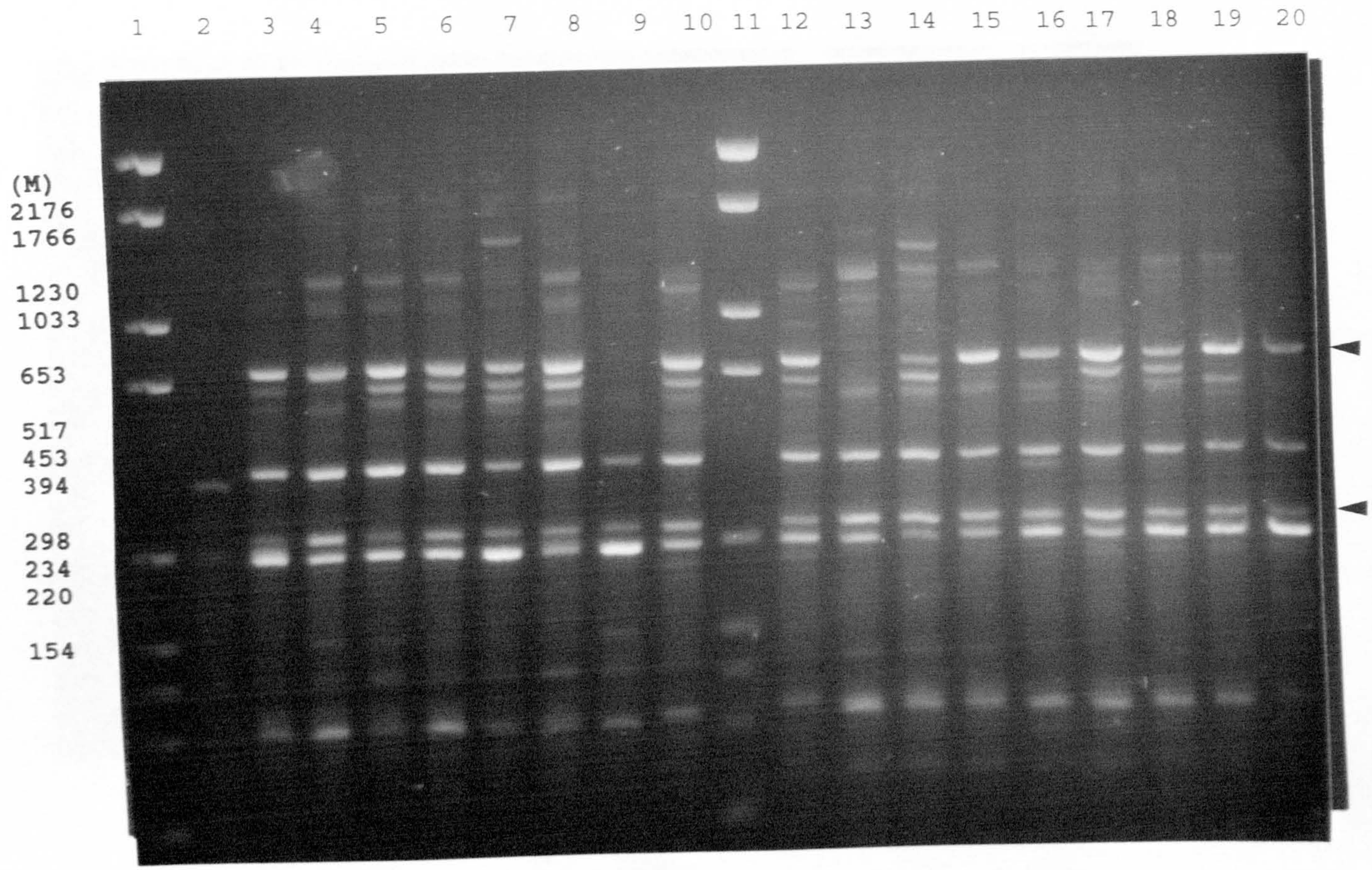
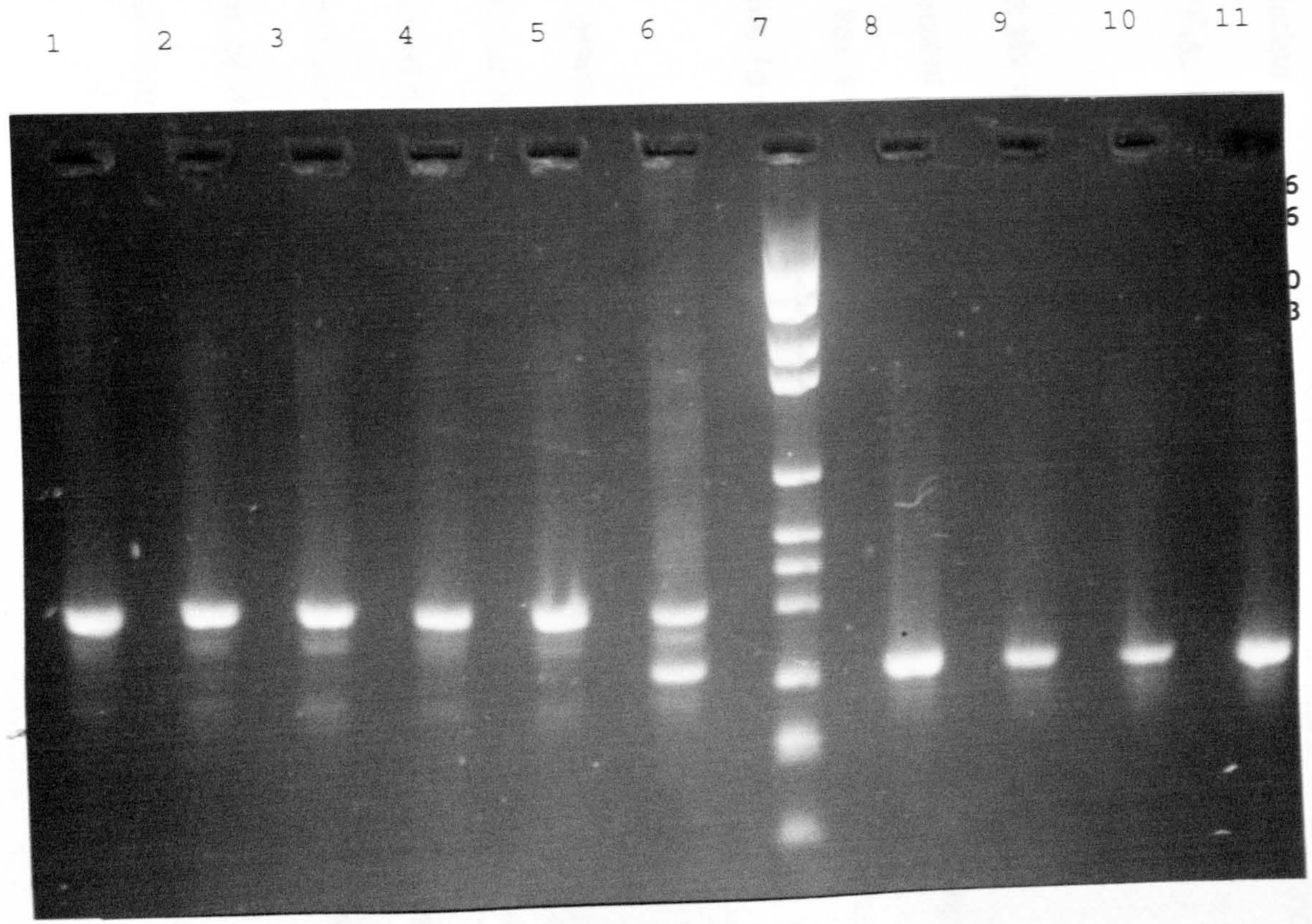


Fig. 4.37- rDNA diagnostic PCR in *An. gambiae* (1-5) and *An. arabiensis* (8-11) and hybrid pattern (6) from Kenya; marker (7). Hybrid individual shows both *An. gambiae* and *An. arabiensis* diagnostic bands.



2)- Using OPA1 primer in rDNA-PCR:

OPA1 primer has 8bp common with UN primer which is the universal forward primer in diagnostic PCR. OPA1 was used with the specific GA, AR, MR, QD primers both at the low annealing temperatures characteristic of RAPD programs (38 and 36°C annealing). Both amplifications, specially in 36°C showed that it could be used in differentiation of those three species *An. merus*, *An. melas* and *An. quadriannulatus*) (Fig. 4.38A, 4.38B).

3)- rDNA with all six members:

This experiment showed that *An. merus* and *An. melas* share a 520bp band while *An. merus* has another shared 150bp band with *An. quadriannulatus* (Fig. 4.39). *An. bwambae* has two 730 and 400bp.

4)- UN as RAPD primer:

The UN primer has been used as a RAPD primer in a 36°C annealing program (Prog 1) and produced differentiation patterns in *An. gambiae* and *An. arabiensis*, as well as detecting variation within those species (Fig. 4.40). For example ZAND and 16CSS, two strains of *An. gambiae s.s.*, differ from SENN and KGB strains of *An. arabiensis*. Furthermore strain-specific differences can be observed.

Fig. 4.38A- Diagnostic PCR based on amlification of rDNA-IGS region in three members of *An. gambiae* complex: *An. merus* (1-5), *An. melas* (6-9), *An. quadriannulatus* (10); marker (11). In a 40 cycles of 38°C program five rDNA-IGS primers OPA1,GA, AR, QD, MR) amplified a 470bp band in these three species, while *An. quadriannulatus* has a 153bp which in *An. merus* has some variation (140-153bp). *An. melas* specific band is 120bp.

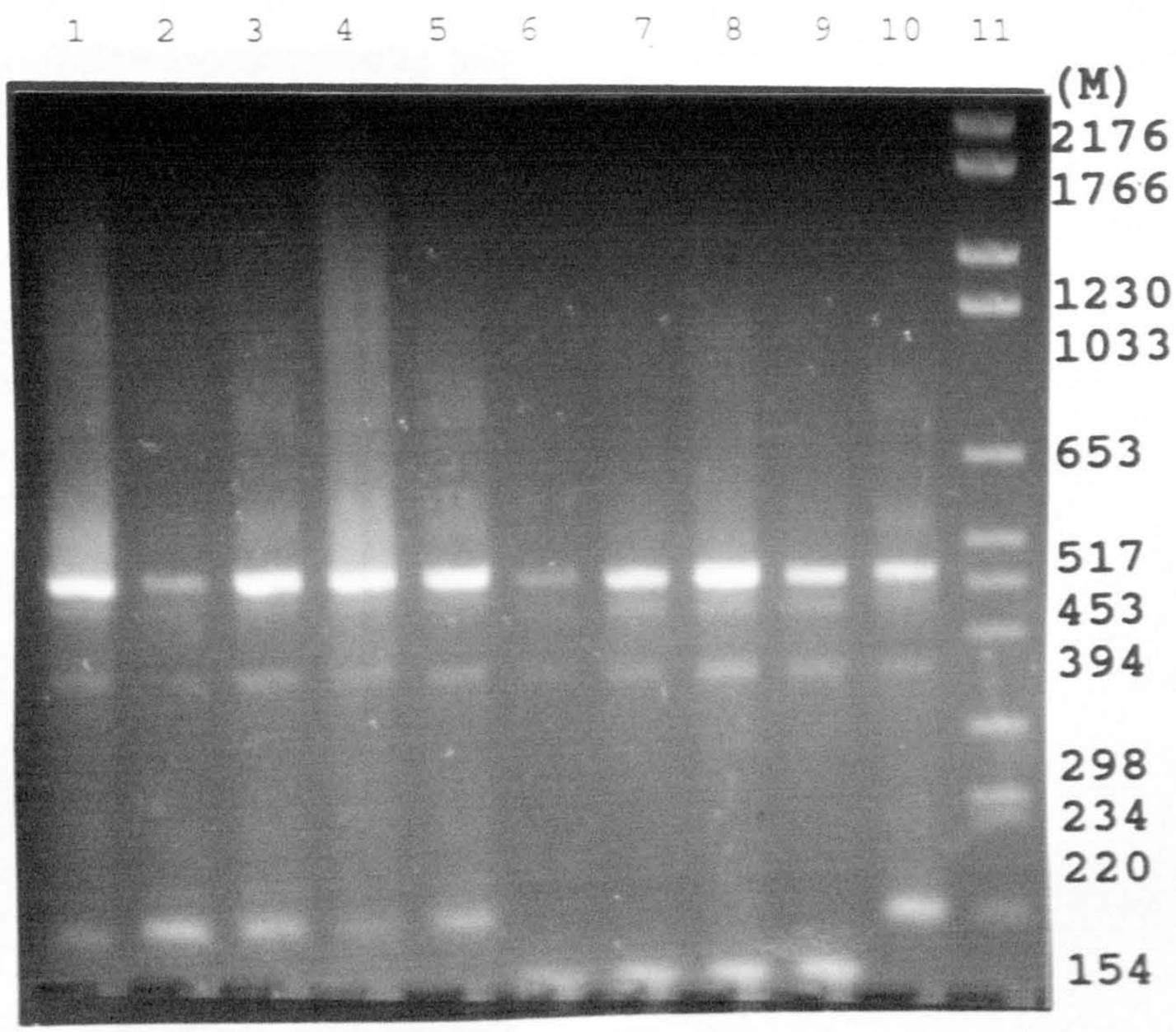


Fig. 4.38B- Diagnostic PCR based on amplification of rDNA-IGS region in three members of *An. gambiae* complex: *An. melas* (2-7), *An. merus* (8-9), *An. quadriannulatus* (10-11); marker (1). 12.25ng of primer OPA1 and 25ng from MR and QD have been used as a cocktail in a 40 cycles program with 36°C annealing.

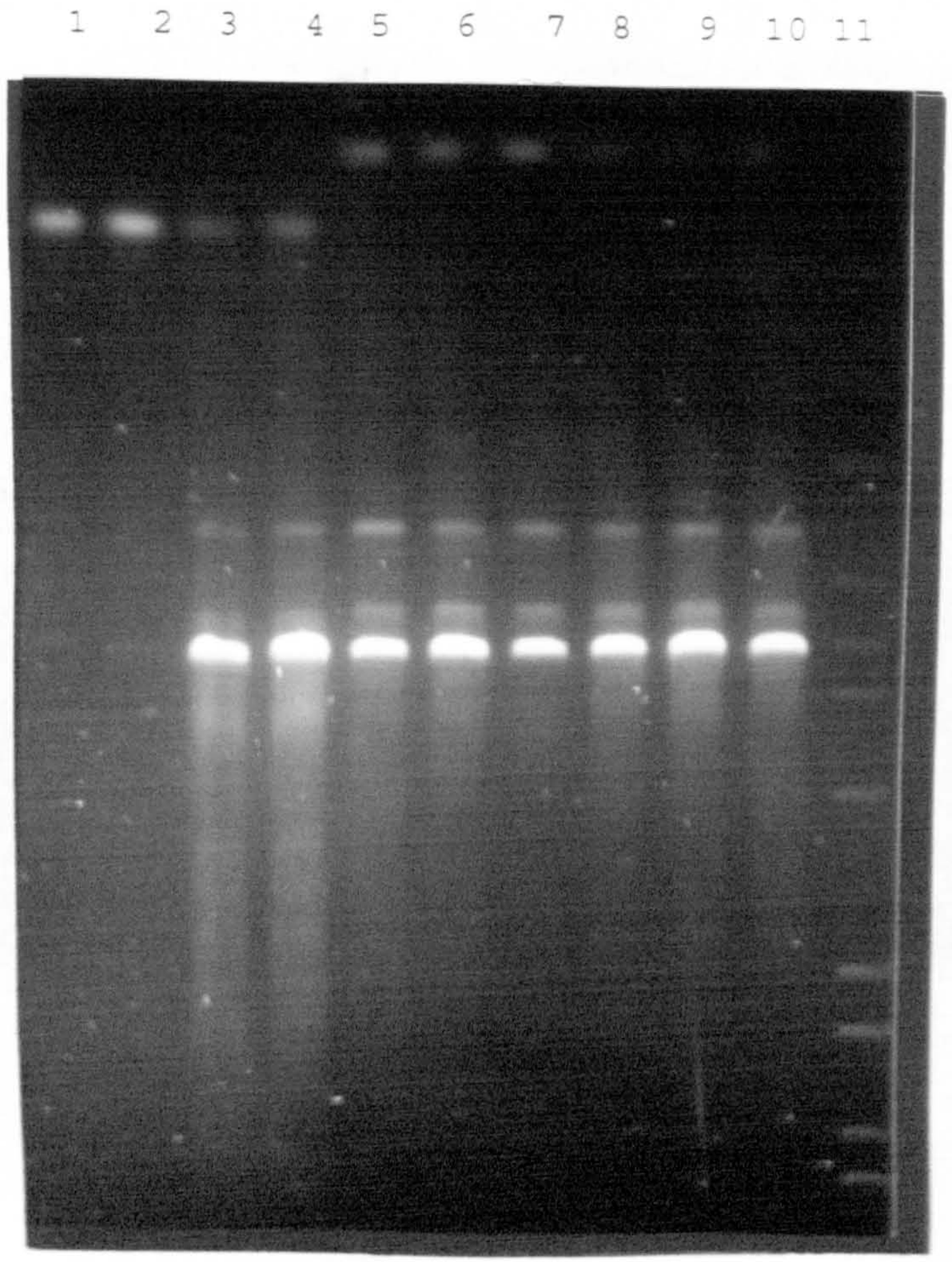


Fig. 4.39- Diagnostic patterns of rDNA-IGS specific primers (50°C-30 cycles) in all six members of *An. gambiae* complex run on 2% agarose gel: *An. bwambae* (3-4), *An. quadriannulatus* (5-8), *An. gambiae* (9), *An. arabiensis* (10), *An.merus* (12-16), *An.melas* (17-19); marker (1,11,20), -ve (2). *An. quadriannulatus* has a 153bp band share with *An.merus*, while *An.merus* and *An.melas* have a common 465bp fragment.

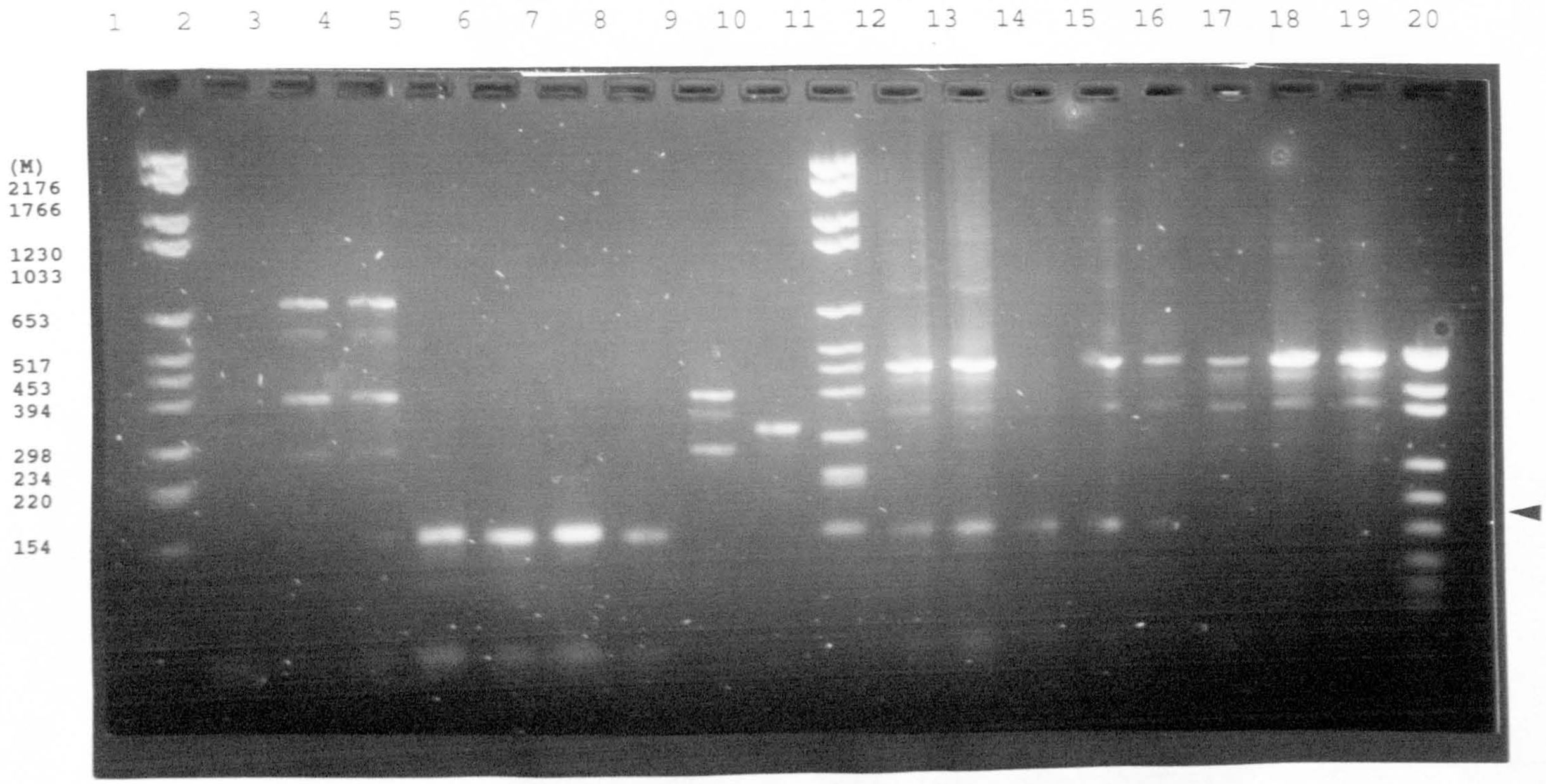
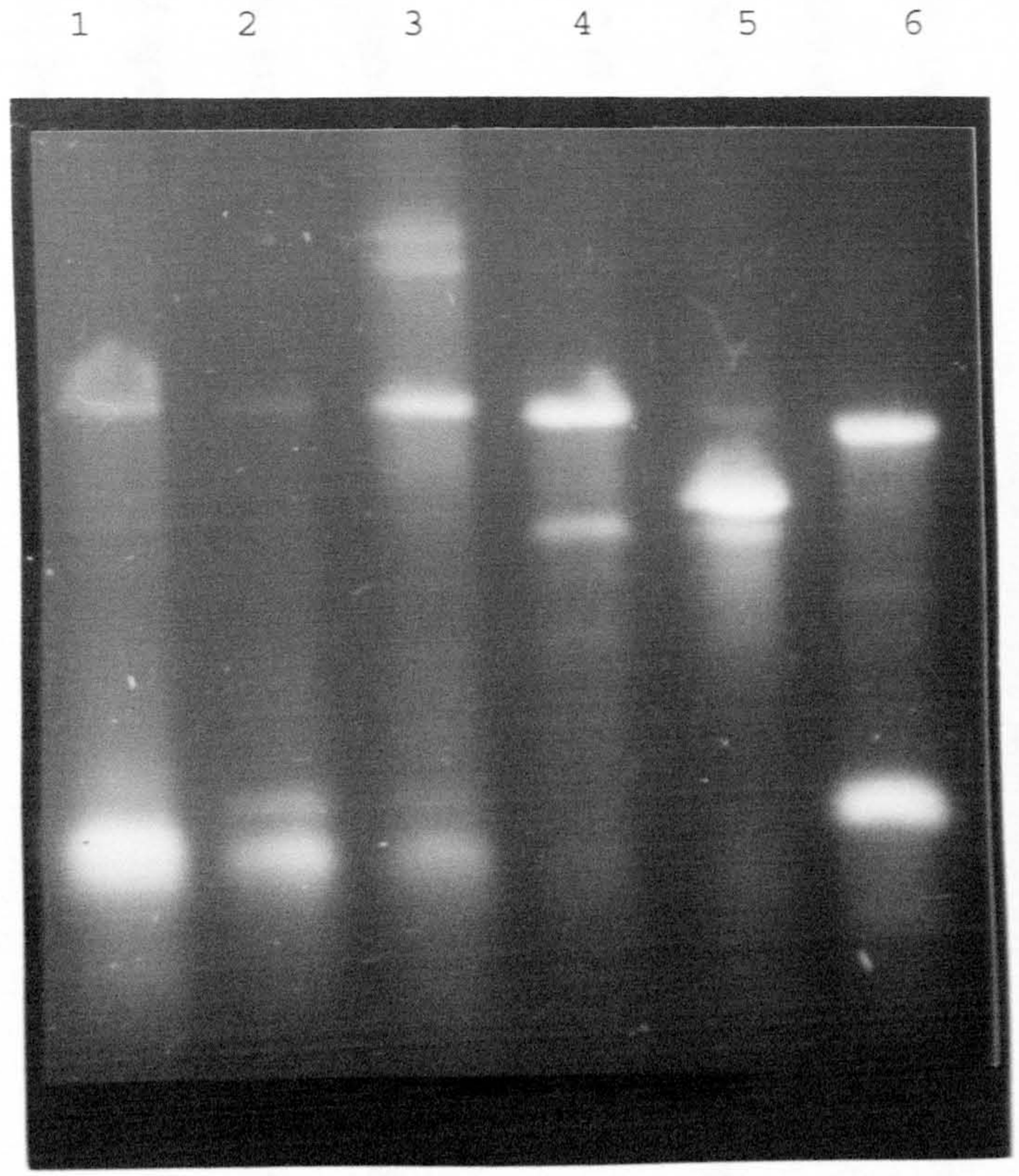


Fig. 4.40- RAPD-PCR amplification of UN (rDNA-specific forward primer) in *An. gambiae*: 16CSS (1-2), ZAND (3); *An. arabiensis*: SENN (4) , KGB (5); and *An. stephensi*: IND-S (6) as out group. Although each species has specific pattern but there are some variation between strains of same species



4.5 DISCUSSION

4.5.1 Inter and intra-specific variation

Detection of genetic variation is essential to a wide range of comparative genetic research, including *Anopheles* species complexes.

Members of the *An. gambiae* complex were at first recognized by their crossing characteristics, then latter by polytene chromosome morphology (Davidson 1962; Davidson & Jackson 1962; Davidson 1964a, 1964b; Davidson & White 1972; Davidson & Hunt 1973; Coluzzi *et al.* 1978) and a biochemical key based on enzyme electrophoresis (Mahon *et al.* 1976).

There are few published papers on RAPD in the *An. gambiae* complex, and these are solely based either on laboratory stocks of *An. gambiae* s.s. and *An. arabiensis* or on field specimens of just one or two species within the complex (Wilkerson *et al.* 1993, Favia *et al.* 1994a & 1994b). Recently Dimopoulos *et al.* (1996), reported the production of an integrated genetic map of *An. gambiae* by using RAPD and microsatellite polymorphisms.

However, the present work appears to be the first RAPD study of all six formally named species of the complex. Results have shown that RAPD can differentiate these six species, detect intra-specific variation and identify individual specimens to species in mixed populations. Coluzzi *et al.* (1979, 1985); Coluzzi (1992); Toure *et al.* (1994) described the correlation of different ecological

environment with different chromosomal inversion frequencies in *An.gambiae s.s.* in west Africa. It is generally held that inversion polymorphism is much less common in *An.gambiae* from East Africa. However, in western Kenya, RFLP on the rDNA-IGS region indicated a heterogenic population structure (McLain *et al.*, 1989) even though Petrarca & Beier (1992) in the same area, had found a rather homogenous population using karyotype analysis. Later, Lehmann *et al.* (1996b) compared populations of the Savanna karyotype from west and east Africa using allozyme and microsatellite loci and found that gene flow is only weakly restricted over a distance of about 6000 km. More recently, Lehmann *et al.* (1997) reported that the minimum area associated with a deme of *An.gambiae* in western Kenya is larger than 50 km in diameter.

RAPD analysis has proved useful in defining the population structure of the *Aedes aegypti* (Apostol *et al.* 1993). It seems likely that RAPD markers might prove useful in studies of population structure in species of the *An.gambiae* complex. These

Extensive variations has been detected in *An. arabiensis* with primers UBC-302, 308, and 353 which could be used in the population structure studies to look at correlations between RAPD varieties and the environment and ecological features of populations.

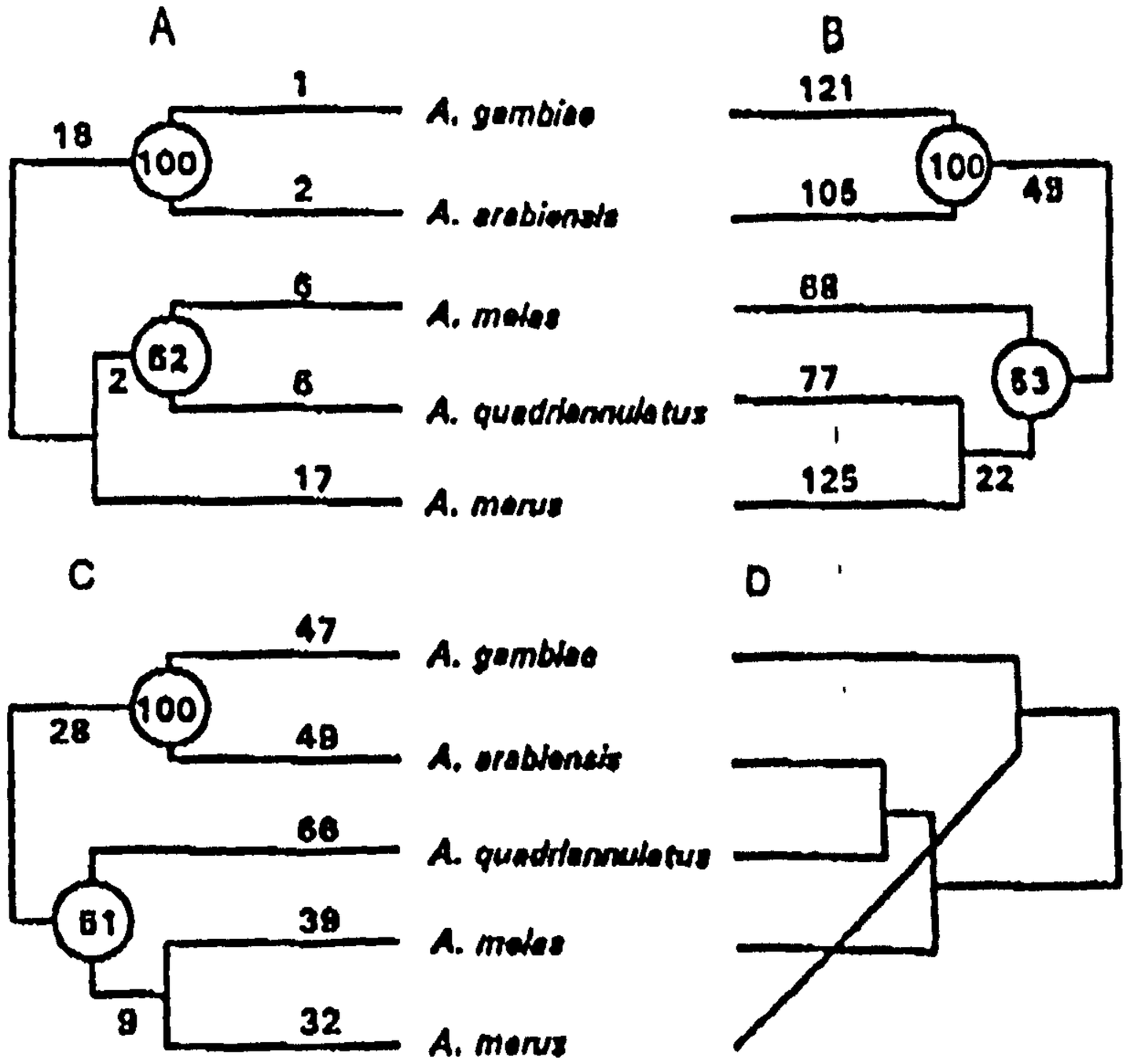
In *An. bwambae* considerable intra-specific variation was found which is perhaps surprising when we recognize that all specimens of this species came from same locality in Uganda, the only place where this species is found.

4.5.2 RAPD data analysis and phylogeny

Dendrograms based on M13F and M13R fingerprints in *An. gambiae* species complex, have shown two different trees. In comparison to other published trees, the M13F and M13R dendrograms are more similar to the tree based on chromosomal inversion (Coluzzi *et al.* 1979) than other published trees. *An.melas* and *An.bwambae* considered as are sister taxa as with M13F, but with M13R *An. quadriannulatus* and *An. arabiensis* become sister taxa as they do in the chromosomal inversion tree (Fig. 25A & 25B, Fig. 4.41)(Besansky *et al.* 1994).

Tables 4.3-4.6 summarised the distance matrices in six members of *An. gambiae* species complex using Jaccard, Dice, Nei, Montpellier distances and dismatch coefficient. Results of all indices are in concordance with each other. Overall results with RAPD primers scored in these species on the basis of presence and absence of bands were used to generate phylogenetic trees with the PHYLIP program (Fig. 4.26A, 4.26B). In this case, both the Wagner and Camin-Sokal parsimony methods showed that *An. gambiae* and *An. arabiensis* are sister taxa, agreeing with the results of Besansky et al. (1994) based on mtDNA and rDNA (Fig. 4.41). Bringing *An.melas* and *An.bwambae* together as sister taxa in the tree constructed by the Camin-Sokal parsimony method is more similar to the tree constructed based on chromosomal inversions (Fig. 4.41)(Besansky *et al.* 1994, Coluzzi *et al.* 1979). However, with RAPDs, both Wagner and Comin-Sokal methods showed *An. quadriannulatus*, *An.merus* and *An. melas* as sister taxa. It can be concluded that RAPD is a powerful technique, not only in identification of members of species complexes, but also for population genetic studies.

Fig4A1 Phylogenetic trees based on A)- mtDNA, B)- rDNA, C)- combined mtDNA and rDNA, D)- chromosomal inversions (after Besansky *et al.* 1994)



4.5.3 RAPDs characterization

The cloning and sequencing of M13F+R single product in *An. bwambae*, revealed a high degree of homology with the AB11 common band in *An. culicifacies* species A and B, and the similarity of both sequences with ubiquitin-related genes is promising. Ubiquitin is a protein which can be found in all eukaryotic organisms and has a highly conserved amino acid sequence. The ubiquitin gene structure is very complex, including sequential and repetitive polymorphisms (Wibrog *et al.* 1985, Dworkin-Rastl *et al.* 1984). This gene could be used for comparative studies of the kind shown for the ubiquitin gene of *Drosophila melanogaster* (Izquierdo *et al.* 1984, Arribas *et al.* 1986, Boge *et al.* 1994). It may prove worthwhile to develop specific primers based on this sequence or to use it as a probe to study the amount of relatedness or variation in all members of *An. gambiae* species complex.

4.5.4 Identification of An. bwambae from Uganda

The results of RAPD with *An. bwambae* specimens from the field and specially dried collections showed another powerful feature of RAPD in generating fingerprints from dried specimens which only has been reported by using rDNA diagnostic PCR (Townson *et al.* unpublished data, Harbach 1997) and non-radioactive labelled DNA probes (Vij *et al.* 1997). The successful amplification of DNA prepared from about 20 years age dried specimens of *An. culicifacies* also has been reported in this thesis.

Table 4.7- BLAST search result on homology (94-77%) of *An. bwambae* 500bp band with other reported genes in different organisms. Total lenght and region of homology can explain the significance of these data.

GENBANK NUMBER	SPECIES AND RELATED GENE	Total length	Region of homology
gb|U65165|XPU65165	*Xenopos pecki* 18S ribosomal gene	1421	412-497 / 793-918
gb|L39119|MUSFRE	*Mus musculus* (clone DE-5) mRNA	226	410-497 / 19-148
gb|U52949|SSU52949	*Spisula solidissima* cyclin-specific	1242	392-470 / 1163-1242
gb|U60877|AA60887	*Anopheles albimanus* clone Que 13 Que	758	471-411 / 5-66
gb|U60880|AAU60880	*Anopheles albimanus* clone Que 16 Que	779	471-411 / 1-62
j|D44443|MUSDIP	Mouse mRNA for dexamethasone induced	573	412-491 / 343-473
gb|U60876|AAU60876	*Anopheles albimanus* Quetzal transpos	776	471-410 / 12-74
gb|L22944|HIV1SE364A	Human immunodeficiency virus type 1	3110	470-412 / 1-60
gb|U67221|HSU67221	Human clone HS4.14 Alu-Ya5 sequence	590	471-412 / 16-76
emb|X90639|PCDNA3ZEO	Cloning vector pcDNA3ZEO DNA	3989	472-414 / 887-946

4.5.5 Cross experiment

A Crossing study with two strains of *An. gambiae* s.s. was carried out with a few primers. The result confirmed the inheritance of most fragments in a dominant fashion. Size variation in some main bands between larvae and other stages and non-parental bands may reflect the sensitivity of RAPD to genomic variation although its control mechanism needs more investigations.

4.5.6 rDNA IGS

rDNA-specific primers results with different species within a complex were in concordance with the results of RAPD primers, although some RAPD primers are able to both diagnose species and detect genetic variation within populations. UN was used as a single RAPD primer due to its similarity with primer OPA1 sequence. The results showed that gene-specific primers may be useful in detecting polymorphism in closely related species or even within species since it apparently amplify specific region of the genome.

4.6 CONCLUSION

The overall results of this study demonstrate the wide range of application of RAPD to generate valuable data in different aspects of population genetics. In comparison with other molecular and PCR-based techniques which are dependent on previously known sequences, RAPD is the method of choice where no prior sequence data is available.

It seems that, in general, co-migrating bands (found with several RAPD primers) are homologous, therefore these shared characters can be used for phylogenetic inference, as long as reproducibility can be established.

CHAPTER FIVE

RAPD, SSR AND ITS2 STUDIES ON ANOPHELES CULICIFACIES SPECIES COMPLEX FROM MIDDLE EAST AND INDIAN SUBCONTINENT

5.1 ABSTRACT

An. culicifacies s.l. has a wide distribution from Iran in the west through Indian subcontinent, Burma, Thailand and China in the east. It is a malaria vector of great importance in most parts of India and also in several neighbouring countries.

Previously ecological data, cytogenetic and isoenzyme analysis and DNA probe have been used in various countries to identify the four members of the *Anopheles culicifacies species* complex. We have used RAPD, SSR and ITS2-PCR to study genetic variation in field-collected specimens of species A and B of the *An. culicifacies* complex from the Middle East and Indian subcontinent.

16 RAPD primers consistently and reproducibly differentiated between species or populations of the complex. Strong homology in co-migrating RAPD fragments have been confirmed by Southern blot hybridization of a 10mer primer products in species A and B to a random primed labelled common band. Cloning and sequencing of this common band has shown 63-72% homology with the nucleotide sequence of the ubiquitin gene of *An. gambiae*, human and *C. elegans*.

Another Primer, AB19 produced in all specimens a single band which differed in size in samples from the north and south Sri-Lanka samples. Two $(GT)_7$ and a $(AGT)_5$ repeat primers with a 3'-anchor produced characteristic fingerprints of the two species and their populations. Phylogenetic trees based on 291 bands scored in RAPD and SSR-PCR products separated species A from B and also differentiated species B populations from the north and south of Sri-Lanka. Species B from India was most similar to the south Sri-Lanka population. Iranian populations of species A with higher intra-specific variation in comparison with species B, clustered into two groups; 1) collections from areas closer to the border of Afghanistan and Pakistan, and 2) those from southern parts of Baluchistan and Sistan.

Some specimens, identified on morphological ground as *An. culicifacies* , and originated from Baluchistan, Iran, had unusual RAPD patterns and have been tested by sequencing the ITS2 region. The size, initiation and ending sequences of the ITS2 in these specimens is similar to that of other *An. culicifacies* but overall, the nucleotide sequence differences with species A and B is about 43%. A phylogenetic tree based on ITS-2 suggests a position between *An. culicifacies* and *An. fluviatilis* although this inference should be treated with care until more data are available.

This study demonstrates the extent and importance of specific and intra-specific variation revealed by RAPD, SSR and ITS2 in a major malaria vector populations and its possible use in malaria epidemiology.

5.2 INTRODUCTION

The failure to recognize sibling species of anopheles can mean, and has meant in the past, failure to distinguish an epidemiologically important species from an unimportant one (Service, 1988a; Coluzzi, 1992; Hunt & Coetzee, 1995). Moreover, the study of host preference, infection rate, resting habits, and biting cycle, as well as the assessment of control measures, may be seriously misleading if carried out on a mixture of two or more apparently isomorphic species (Service, 1991).

Many malaria vectors in the Middle East and Indian sub-continent belong to sibling species complexes. For example *Anopheles culicifacies* Giles 1901, an efficient vector of human malaria in Iran, Afghanistan, Pakistan, India, Sri-Lanka, Bangladesh, Myanmar, Thailand, Nepal and southern China is known to comprise at least four sibling species known as A,B,C,D (Milligan *et al.*, 1986; Subbarao, 1988; Subbarao *et al.*, 1988; Vasantha *et al.*, 1991; Subbarao *et al.* 1993; Nagpal & Sharma, 1995). In India, all four sibling species occur, although species B is reported to be a non-vector (Tiwari *et al.* 1994a, 1994b). *An. culicifacies* has been regarded as predominantly zoophilic and a wide range of variation in anthropophily is reported in the literature. Since the precise species of the complex are rarely recorded, it is not known to what extent variations in anthropophily / zoophily reflect different composition of the constituent species of the complex.

The karyotype of *An. culicifacies* was first described from the ovary and testis as comprising 1 pair of short subtelomeric sex chromosomes and 2 pairs of longer autosomes, 1 submetacentric and 1 metacentric. Saifuddin *et al.* (1978) confirmed this

description, although some figures have variation in the long arm of the X-chromosome. It was in 1980 that Green & Miles reported two types of X-chromosomes in a natural population originating in the Delhi area, with apparent absence of heterozygotes and this was interpreted as evidence of two biologically distinct species. Contemporary studies on ovarian nurse cell polytene chromosomes using indoor resting collections from villages bordering Delhi lead to the recognition of two chromosomally distinct species, termed A and B (Subbarao *et al.* 1980).

Subsequently, Subbarao *et al.* (1983) described a new chromosomally distinct population which they believed to be a distinct species (species C). They found that species A and B are sympatric around Delhi, and species B and C are sympatric in Surat district (Gujarat), while the three sibling species are sympatric in Hoshangabad district (Madhya Pradesh).

Extracts of cuticular wax from laboratory colonies of adult females of the three sibling species (A, B, C) were analyzed by gas liquid chromatography. The three cytospecies were found to be significantly different in their cuticular hydrocarbon composition as shown by multivariate analysis and specimens from stocks of known identity could be allocated to the correct group with a high rate of success (Milligan *et al.* 1986).

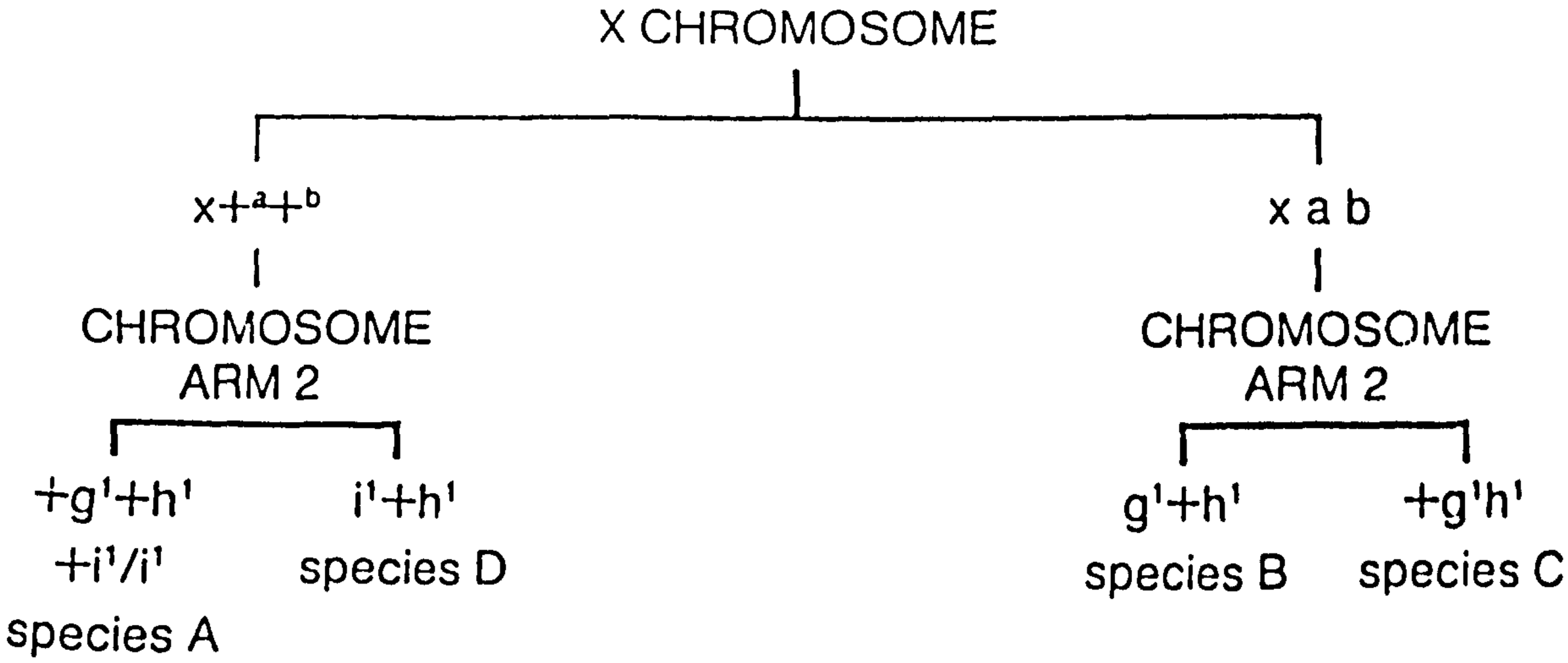
Suguna *et al.* (1989) while studying the ovarian polytene chromosomes of *An. culicifacies* from Tamil Nadu, South India, found a new inversion on chromosome 2. The presence of both inverted and standard arrangement in homozygous state but

the apparent absence of heterozygotes in the field population suggested the presence of a new species, designated as species D. Subsequently, based on cytogenetical and statistical analysis of natural populations examined over several months from different localities, Vasantha *et al.* (1991) supported the finding of species D, reported by Suguna *et al.* (1989), although it was not clear whether the data presented by Suguna *et al.* were from a single collection or pooled from several collections.

However, the polymorphic forms of species A and D are homosequential for polytene chromosome arrangement. Therefore, the polytene chromosome arrangement cannot be used to identify species A and D at an individual level. They suggested that the two species may be distinguished at the individual level through the use of diagnostic allozymes and the development of species-specific DNA probes (Vasantha *et al.* 1991).

A population of *An. culicifacies* s.l. collected from Rameshwaram Island, of Tamil Nadu, India was identified as species B on the basis of the diagnostic inversion karyotype Xab 2g super (1+h1) as observed in polytene chromosomes (Subbarao et al. 1993). Among male mitotic karyotypes made from larval neurogonial cells, two types were observed: one with an acrocentric Y-chromosome and the other with a sub-metacentric Y-chromosome, both had sub-metacentric X and metacentric autosomes. Despite these miotic karyotype differences, the Rameshwaram population is identical to species B in its genetic relationship with species A and C as determined by experimental hybridizations (sterile and fertile male hybrids, respectively) (Subbarao et al. 1993). Fig.5.1 shows the paracentric inversions in four members of

Fig. 5.1- Paracentric inversions in the identification of sibling species of An. *culicifacies*.



An. culicifacies species complex.

Adak *et al.* (1994), found that genetically controlled enzyme variation exists within and between four sibling species of the *An. culicifacies* complex of malaria vectors in India. A study on electrophoretic variation of nine enzymes revealed that the lactate dehydrogenase (Ldh) locus has fast (F) and slow (S) allozymes distinguishing species A+D from species B+C with a probability of 95% . Since *Ldh* is autosomal and expressed in all life stages, the authors think that LDH allozymes represent a useful tool for identification of species pairs in the *An. culicifacies* complex. Gunasekera *et al.* (1995) also developed a DNA probe for the identification of sibling species A of the *An. culifacies* based on isolation of three highly repetitive DNA sequences.The cloned DNA sequences were found at a higher copy number in species B and C, than in species A .

An. culicifacies s.l. in Iran has been studied mainly by Zaim and co-workers. Zaim *et al.* (1993, 1995) reported that *An. culicifacies* (probably species A)(Zaim & Javaherian 1991) is the main vector of malaria in Baluchistan, southeastern Iran. In order to determine the role of *An. culicifacies* s.l. and *An. pulcherrimus* in malaria transmission in Baluchistan, Iran, a two-site immunoradiometric assay (IRMA) was performed on the head and thorax of female mosquitoes collected during the two peak malaria transmission seasons (May and September-October 1991). Positive IRMA results revealed the 2 species as potential vectors of malaria in this highly endemic area (Zaim et al. 1993). This finding serves as the first report on natural infection of *An. pulcherrimus* in Iran and is the second on natural infection of *An. culicifacies*

since the previous report of 1959 by Manouchehri *et al.*

An. culicifacies is considered to be an important malaria vector in rural areas of Pakistan (Subbarao et al. 1980; Curtis & Rawlings 1980). Studies on anopheline mosquitoes in selected villages of Punjab Province, have incriminated *An. culicifacies* species A as the primary malaria vector. Although *An. stephensi* and *An. subpictus* showed higher immediate gut infection rates, estimations of relative abundance, age, structure and survivorship, and observation of late gut and salivary gland infection rates suggested that neither species was a major vector in these villages.

A survey of *An. culicifacies* populations detected the presence of both species A and B in several localities in Pakistan (Mahmood *et al.* 1984, cited in Zahar 1990). Mahmood & Macdonald (1985) carried out ecological studies on malaria vectors with special reference to the vectorial capacity of *An. culicifacies* species A in a rural arca in Punjab, Pakistan. Five species of *Anopheles* were recorded with *An.culicifacies* being the most abundant throughout the year. The maximum numbers of *An. culicifacies* were caught in the November-December post-monsoon period, with two smaller peaks in April and July (Zahar 1990).

In Sri Lanka which is endemic for malaria and has perennial transmission, sibling species B of *An. cuilicifacies* is regarded as an important vector (Wickramasinghe *et al.* 1991a; Abhayawardana *et al.* 1996). Studies of the population dynamics of anthropophilic mosquitoes during the northeast monsoon season in the malaria epidemic zone of Sri Lanka, showed that *An. culicifacies* is the main malaria

vector (Ramasamy *et al.* 1994).

Abhayawardana *et al.* (1996) conducted a study to examine the existence and distribution of sibling species of *An. culicifacies* in different climatic and eco-epidemiological localities throughout Sri Lanka. Their results confirmed the occurrence of sibling species B, but no evidence of other sibling species (A, C,or D) was found.

This study have been carried out in order to examine the usefulness of the RAPD-PCR technology for identifying species, sub-species, and for evaluating the genetic relatedness of unknown populations of field-collected specimens of *An. culicifacies* to known populations of this major malaria vector. It is the first report on genome analysis of anopheline field collected specimens using RAPD-PCR .

5.3 MATERIALS AND METHODS

5.3.1 Mosquitoes

At least 40 individuals were examined from each collection site, apart from a few Indian samples, for which fewer specimens were available (Fig.5.2). Field collected mosquitoes were either air-dried, preserved in 70% alcohol or Carnoy's solution. In addition to recently collected material, pinned specimens of *An. culicifacies* provided by department of medical entomology, Tehran School of Public Health, collected in 1978 have also examined. Table 1 lists the origin of mosquitoes used in this study.

5.3.2 Genomic DNA extraction

Prior to DNA extraction, mosquitoes that had been preserved in alcohol or Carnoy's solution, were rinsed twice in double distilled water and left at least one hour. In view of the sensitivity of RAPD to DNA quality, we compared the following extraction methods each with minor modification; phenol extraction (Ballinger-Crabtree (1992); the method of Collins *et al.* (1987); Mini - Prep (Medina-Acosta & Cross (1993) or their combination. In phenol extraction, the extraction time could be reduced by using 65°C for one hour or 90°C for half an hour instead of overnight incubation. Also, it was found that one step of phenol,chloroform/isoamyl alcohol and another of chloroform/isoamyl extraction yield better result than two times extraction with phenol. The Collins method has been modified by adding a phenol extraction step, and leaving supernatant with alcohol at -20oC for 20 minutes at precipitation step. Air-dried pellets have been resuspended in 100 ul of double distilled H_2O or TE buffer and stored at 4°C.

Fig. 5.2- Origin of *An. culicifacies* species A and B specimens used in this study (brown circles)

Table 5.1-
origin of *An. culicifacies* specimens (all field collections)

Iran: Baluchistan and Sistan province

Zabol(Zab), Zahedan(Zhd), Saravan (Sav), Khash (Ksh), Iranshar(I.Shar), Chahbahar(C.bah), Nikshar(N.Shr)

India: Delhi

Pakistan : Peshawar and Lahore

Sri-Lanka : Putalam, North-Western province (NSL),and Kataragama, Uva province (SSL)

5.3.3 Primers

The following primers were examined for their capacity to random prime: M13 forward (M13F) and reverse (M13R) sequencing primers, OPA series (four 10-mer primers from Operon, Almaada, Callifornia, USA), AB01 series (twenty 10-mer, Advanced Biotechnologies, Leatherhead,UK), UBC series (thirty 10-mer, University of British Columbia, Vancouver, Canada) and three SSR 3'- anchored primers (18-19 mer- OLIGO Express Limited, UK). Table 2 lists the name, sequence and GC content of the primers for which results are presented. Other primers either failed to produce a product or did so inconsistently.

rDNA internal transcribed 2 (ITS2) region amplified by using a set of primers based on the published nucleotide sequence of *Cx. tritaeniorhynchus* 5.8s rDNA (Shimada and Sasaki 1991) and *An. hermsi* 28s rDNA (Porter & Collins 1991):

primer	5.....sequence.....3'	GC%	
5.8S	ATCACTCGGCTCGTGGATCG	60	
28s	ATGCTTAAATTTAGGGGGTAGTC		39

5.3.4 DNA amplification and scoring of PCR products

All RAPD reactions were performed in a total volume of 25 ul. Each mixture contained 2.5 ul of 10X reaction buffer, 2 mM $MgCl_2$, 10-100 ng of primer, 0.001% gelatin, 0.1 mM each of dATP, dTTP, dCTP, and dGTP, 0.5 unit of promega Taq

Table 5.2 - selected primers and their sequences

Primer	5'.....sequence.....3'	GC%
M13	GTAAAACGACGGCCAGT	53
OPA1	CAGGCCCTTC	70
OPA8	GTGACGTAGG	60
AB1-01	GTTTCGCTCC	60
AB1-04	GGACTGGAGT	60
AB1-11	GTAGACCCGT	60
AB1-12	CCTTGACGCA	60
AB1-19	ACCCCCGAAG	70
UBC-301	CGGTGGCGAA	70
UBC-302	CGGCCCACGT	80
UBC-303	GCGGGAGACC	80
UBC-304	AGTCCTCGCC	70
UBC-305	GCTGGTACCC	70
UBC-306	GTCCTCGTAG	60
UBC-308	AGCGGCTAGG	70
UBC-343	TGTTAGGCTC	50
UBC344	TGTTAGGCAC	50
GT1-18	$(GT)_7ATCC$	50
GT2-18	$(GT)_7TGTA$	44
AA19	$(AGT)_5GCCA$	42

polymerase, and sterile double-distilled water to 25 ul. For amplification of the ITS2 PCR mix was prepared with the following changes; 1.5 mM $MgCl_2$, 50 ng of primer, 0.2 mM each of dATP, dTTP, dCTP, and dGTP. Reactions were overlaid with 50 ul of mineral oil and amplified in a Hybaid Omnigene thermal cycler.

In this study fourteen programmes were examined to determine the effect of: 1) - Different annealing temperature (36, 42, 47, 50, 60 °C), 2) - Variable numbers of cycles, and 3) - programs with two or three steps. The following RAPD program was selected for most of the subsequent amplifications: (1)- one cycle of 94°C for 5 min; (2)- 15 cycles of 94°C for 30 sec, 36°C for 30 sec, 72°C for 30 sec; (3)- 20 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec; (4)- one cycle of 72°C for 7 min. For amplification of ITS2 the following protocol was followed: (1)- one cycle of 94°C for 5 min; (2)- 30 cycles of 94°C for 1 min, 50°C for 1 min, 72°C for 2 min; (3)- 72°C for 7 min. Following PCR, 10 ul of amplified DNA mixed with Ficol/orange G loading buffer, electrophoresed in 6% polyacrylamide or 1.2% agarose (Appligen) in TBE buffer containing ethidium bromide at 1.5 v/cm for 16 hours (large gel) or 5 v/cm for two hours (mini gel).

A mixture of BglI and HinfII restriction enzyme digests of pBR328 (Boehringer Mannheim) were used as molecular weight marker for PCR product size determination (molecular weights 2176, 1766, 1230, 1033, 653, 517, 453, 394, 298, 298, 234, 234, 220, 154, 154 bp). Gels have visualised under UV light and photographed with black and white Polaroid 55 film or Ilford FP4 film.

5.3.5 Southern blot and Random prime labelling

Southern blot was originally a hybridization of a probe to restriction enzyme digested genomic DND. In this study, pseudo southern blot of gel fractionated PCR products were carried out using the method of Maniatis *et al.* (1982) adapted for Amersham N+ nylon membrane. DNA probes were labelled with [a-^{32}P]dCTP to > 10^9 cpm by the random-priming method of Feinberg and Vogelstein (1983) using the Boehringer random primming kit. Labelled probes were separated from free nucleotides by passage through a Sephadex G50 column. The activity of fractions collected from the column was measured in a scintillation counter (LKB).

After probing the Southern blots, the radioactive filters were autoradiographed on pre flashed X-ray film (Fuji). The films were exposed for a suitable period at -70 and developed for 2 minutes in Phenisol high contrast developer (Ilford), stopped for 1 minute in 5% acetic acid, fixed for 1 minute in fixer (Ilford), washed in water and air dried (see Ch.3 for more details).

5.3.6 Cloning and Sequencing

Cloning of PCR products was carried out using the Invitrogen TA cloning kit in which the inserted PCR product is flanked on each side by EcoR I sites. Prior to sequencing, clones and PCR products were purified by passage through an S400 spin column (Pharmacia), ethanol precipitated and resuspended in proper volume of ddwater based on DNA concentration estimated on gel by comparison with size marker. Sequences were determined by AB1 PRISM 377 DNA sequencer (cycle sequencing reaction) using AMPLI TAQ FS polymerase and manufacturer's

guidelines with M13 Forward and reverse primers. In manual sequencing (chain termination reaction), sequenase version 2.0 T7 DNA Polymerase (USB) was used.

5.3.7 Data Analysis:

All reproducible band in the range of 150-3000 were included in scoring and subsequent analysis. RAPD and SSR data were compared using the following similarity and distance coefficients calculated for all populations :

Jaccard similarity (Southwood, 1978) :

$S_s = A / B+C$

Jaccard similarity (Cibulskis *et al.* 1986):

$S_J = A / A+B+C$

Dice similarity (Beverly *et al.* 1987):

$S_D = 2A / 2A+B+C$

Similarity index of Nei & Li (Nei and Li 1979):

$S_N = 2N_{ab} / Na + N_b$

Simple Matching coefficient (Sneath & Sokal, 1973):

$S^M = A+D / A+B+C+D$

Montpelier distance (Tibyrenc et al. 1993):

$S_{Jm} = 1- [A / A+2(B+C)]$

The Mix program in Phylip (Felsenstein, 1993), using Wagner and Camin-Sokal parsimony methods (Camin and Sokal 1965) and the Consensus Tree Program version 3.54c have been applied to RAPD and SSR data in the specific format of 1 for presence and 0 for absence of a band (e.g. 1001110101110). GCG (version 7.0, 1994), DNA Star, and Clustal V (Higgins & Sharp 1988) packages were used to enter, align and help analyze the sequences.

5.4 RESULTS

5.4.1 RAPD Optimization:

In these experiments, use of small quantities (5pg-10 ng) of phenol-extracted DNA produced sharper and stronger patterns. Different concentrations of magnesium, primer, dNTPs and Taq polymerase have been examined (Fig.5.3). It seems that 2mM $MgCL_2$, 0.1 mM of dNTPs, 50-100 ng of primers (in some primers 10 ng also had successful amplification) , 0.5 unit of Taq provide optimal condition for amplification.

5.4.2 RAPD primers:

16 out of 56 RAPD primers (28.6%) tested in this study produced consistent patterns in field-collected specimens of *An. culicifacies* species complex from Iran, Pakistan, Sri Lanka, and India.

AB11

RAPD primer AB11, in a two steps annealing program (Prog 3) differentiates species A (Iran and Pakistan) and B (Sri Lanka and India) by a 620 bp band specific to species B and a 412 bp band common to both species (Fig.5.4). The results of restriction digestion by DraI of the 412bp common band from species A and B are shown in Fig.5.5. In both species two fragments were produced of 266, 146bp (Fig.5.5). For SacI, fragments of 264 and 148bp and for TaqI, 326 and 86bp were produced in both species.

Fig. 5.3- RAPD-PCR (primer 308 and 10ng DNA): optimization of $MgCl_2$: 2.5, 1, 0.5 mM, (lanes 1-3); dNTPs:0.1, 0.05, 0.025 (lanes 5-7); Primer: 100, 50, 10ng (9-11); Taq polymerase: 1, 0.5, 0.25 unit (lanes 13-15); mloecular marker (lanes 4,8,12)

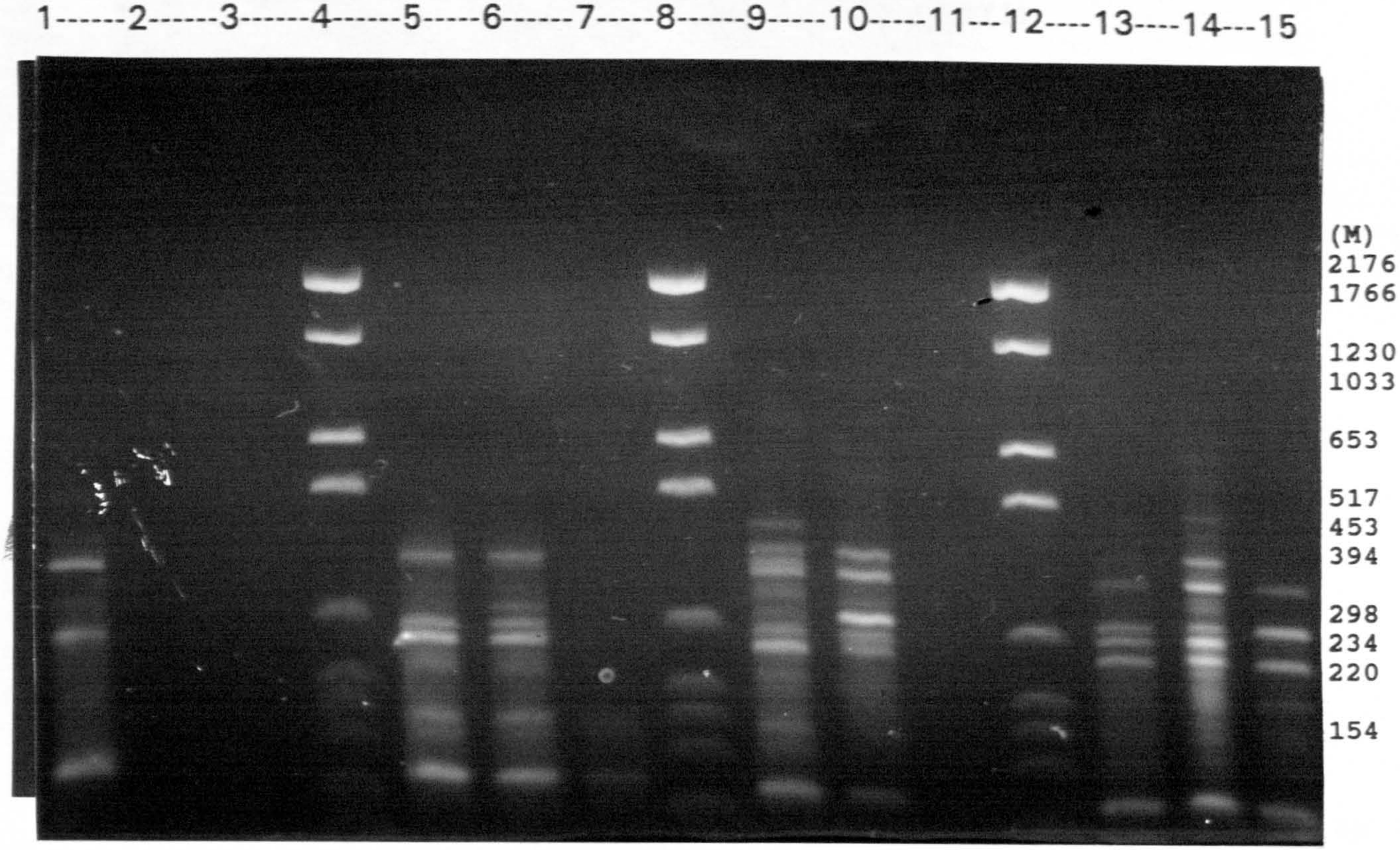


Fig. 5.4- RAPD primer AB11 products in *An. culicifacies* species B from Sri-Lanka (alcohol preserved,lanes 1-5), and A from Iran (pinned: lanes 6-7, alcohol preserved: lanes 8-9, airdried: lane 10), molecular Marker(lane 11)

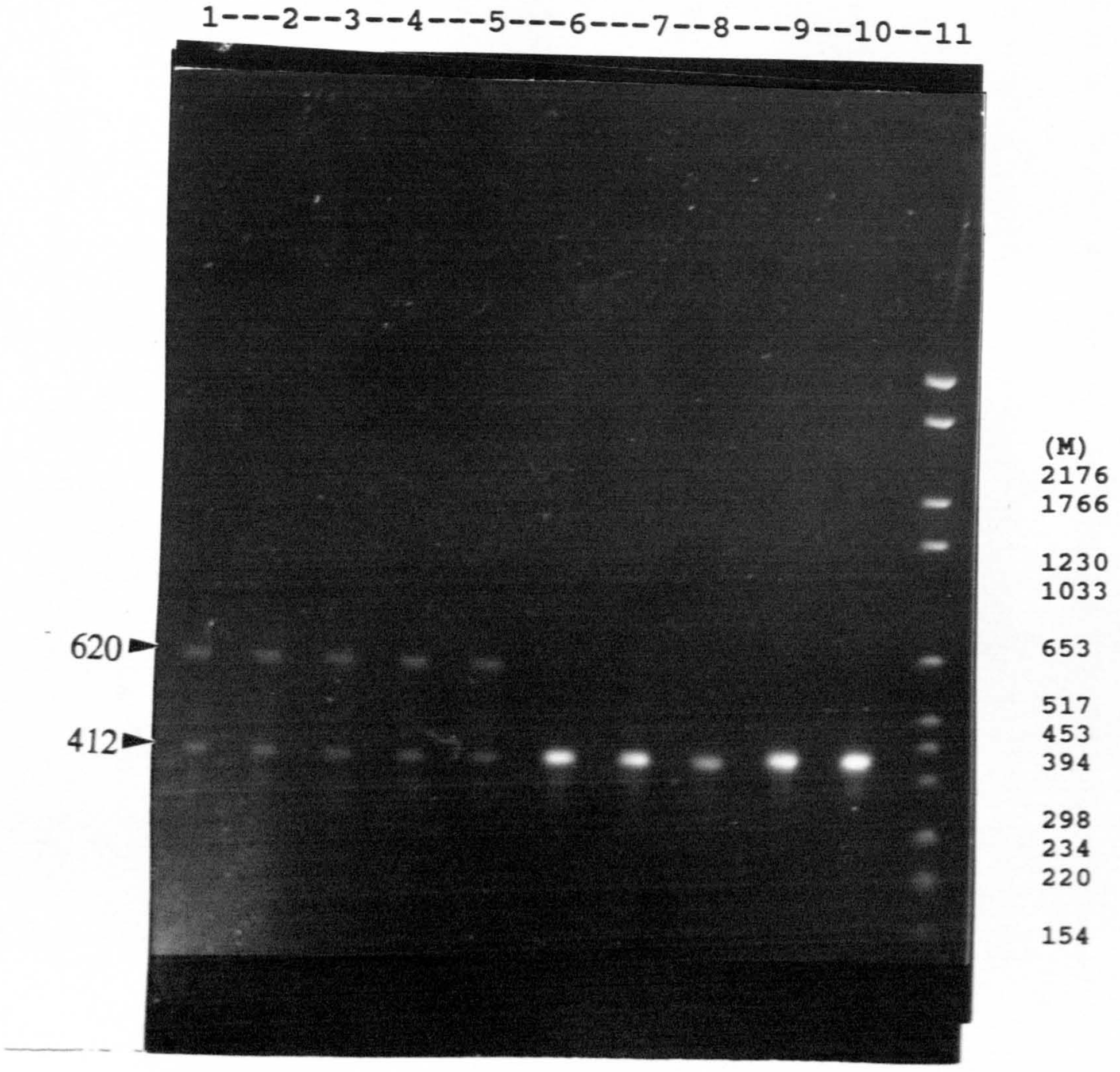
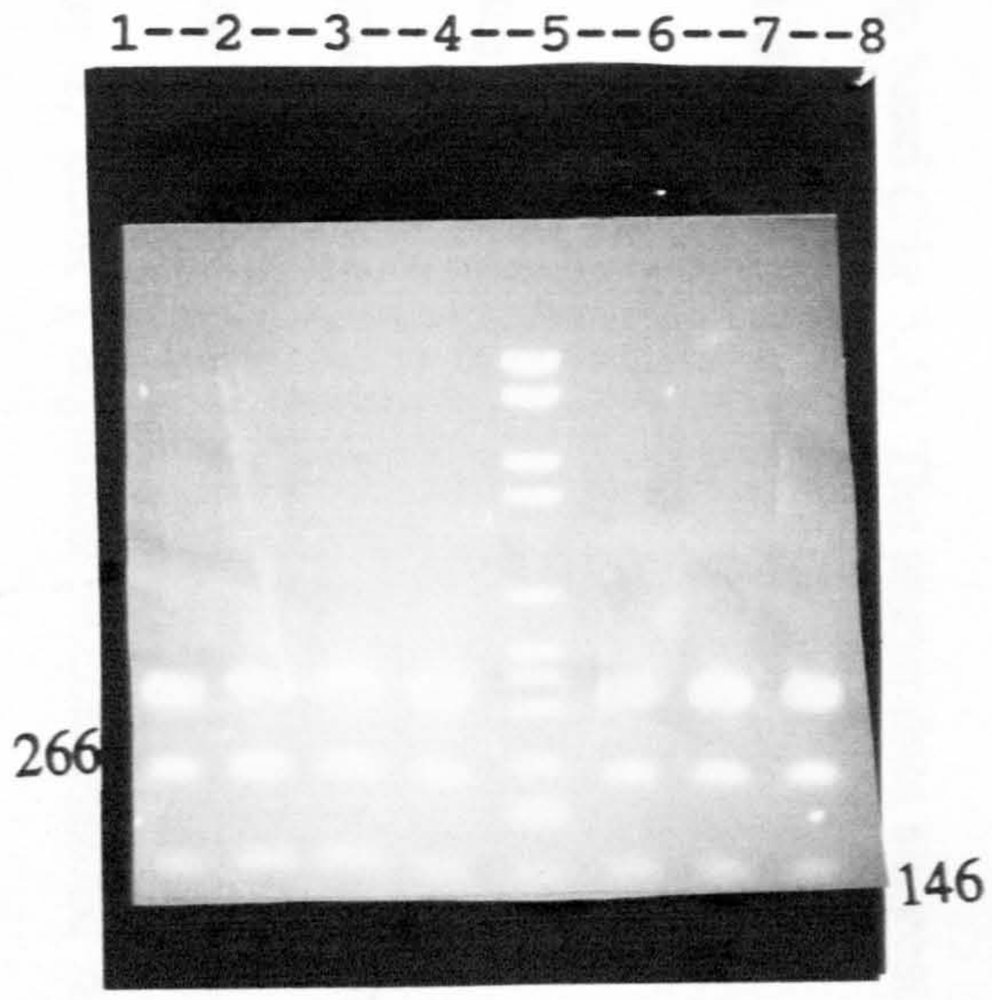


Fig. 5.5- DraI digestion of 430bp common band of AB11 product in species B (lanes 1-4) and A (lane 6-8), molecular marker (lane 5)



DNA probe

The common band of 412bp from species A was excised from the gel and after reamplification and purification by passing through column (Pharmacia), 20-50ng DNA was used fro labelling (see also Ch.3).

After random primed labelling (Boehringer random priming kit) of this common band, it was used to probe a nylon filter 'southern blot' of the AB11 RAPD PCR products. In both species A and B the probe hybridized to the 412bp common band but not the species B-specific 620bp band (Fig.5.6). The above data suggest that the 412bp band is homologous between the two species.

Cloning and sequencing

In order to characterize further the 412bp fragment, the band was excised from the gel, reamplified and cloned to permit sequencing (Fig.5.7). The size of the insert was confirmed by 1) EcoR I digestion which showed 412bp, 2) M13 F+R amplification which also revealed the expected size of sequencing region (Fig.5.8), and 3) AB11 primer amplification of cloned PCR product which amplified a single band of 412bp (Fig.5.9). The 412bp fragment then sequenced from both directions using M13 primers and both automated and manual sequencing (Fig.5.10). Gene bank data search of *An. culicifacies* common band sequence (Acc.No.....) showed 72, 64, 63% similarities in 48, 46, 36bp to tissue factor pathway inhibitor precursor (TFPI) of *Caenorhabditis elegance (C. elegance)*, GenBank Acc.No. U64857; Human ubiquitin activating enzyme E1, GenBank Acc.No. L13852; and *An. gamblae* polyubiquitin gene, GenBank Acc.No. L36067, respectively.

Fig. 5.6- RAPD primer AB11 products hybridization to ^{32}P labelled 430bp common band of species B (lanes 1-6) and species A (lanes 7-11)

1----2---3----4---5----6----7----8---9---10--11

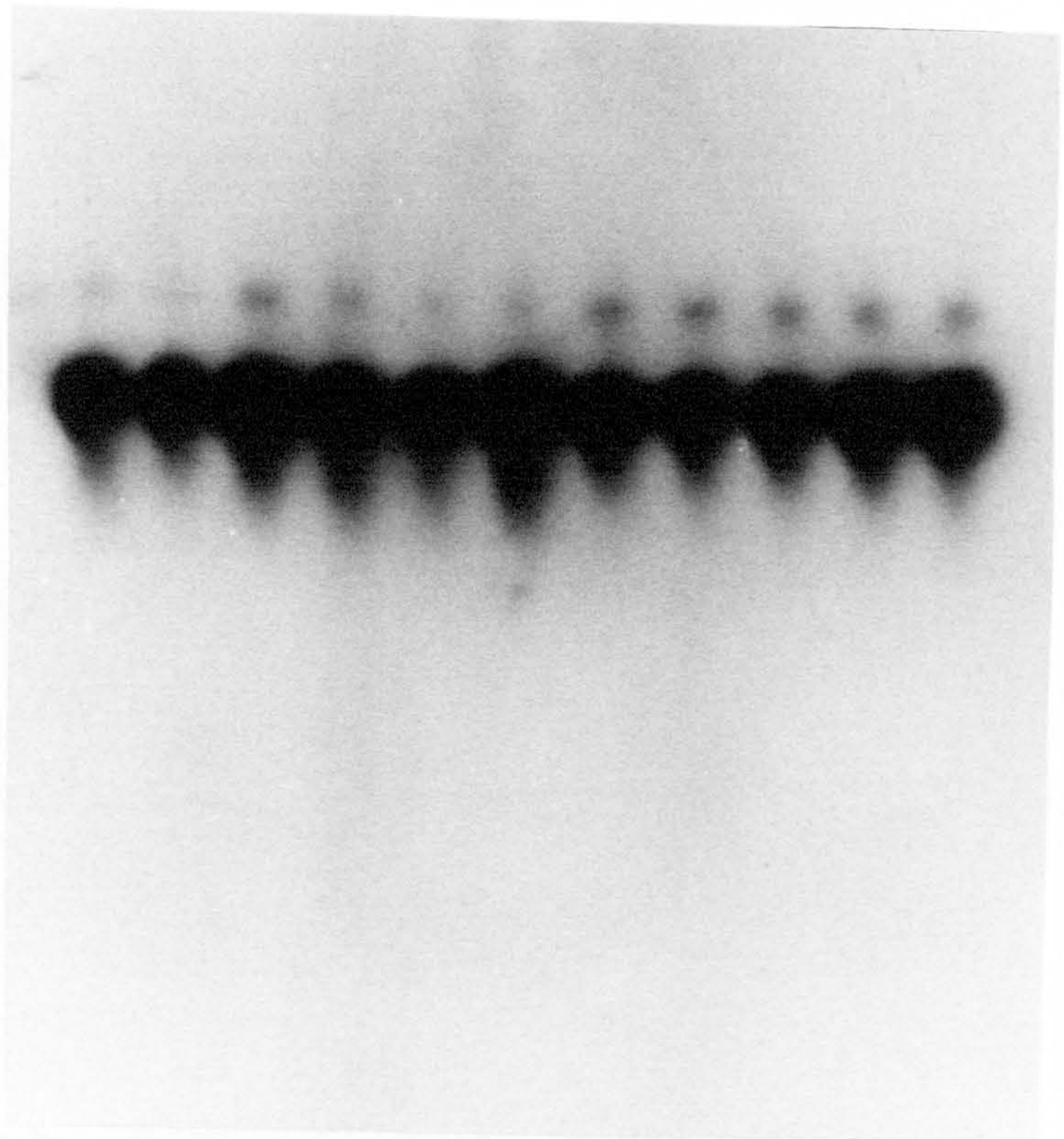


Fig. 5.7- cloned AB11 primer product (412bp) of *An. culicifacies* from Iran

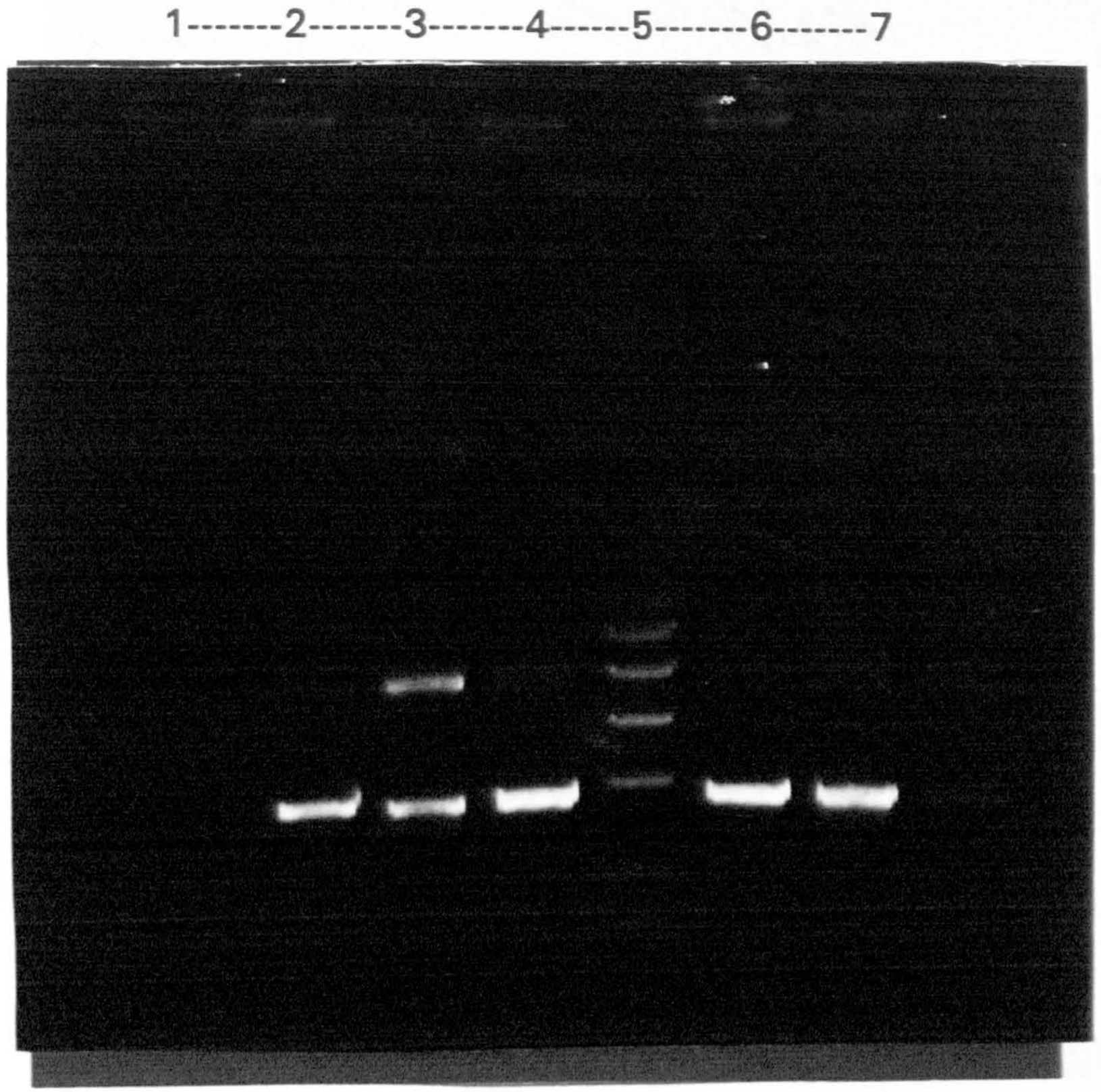


Fig. 5.8- EcoRI digestion of cloned PCR product (lanes 1-6) and M13 F+R amplification of cloned PCR product (lane 8)

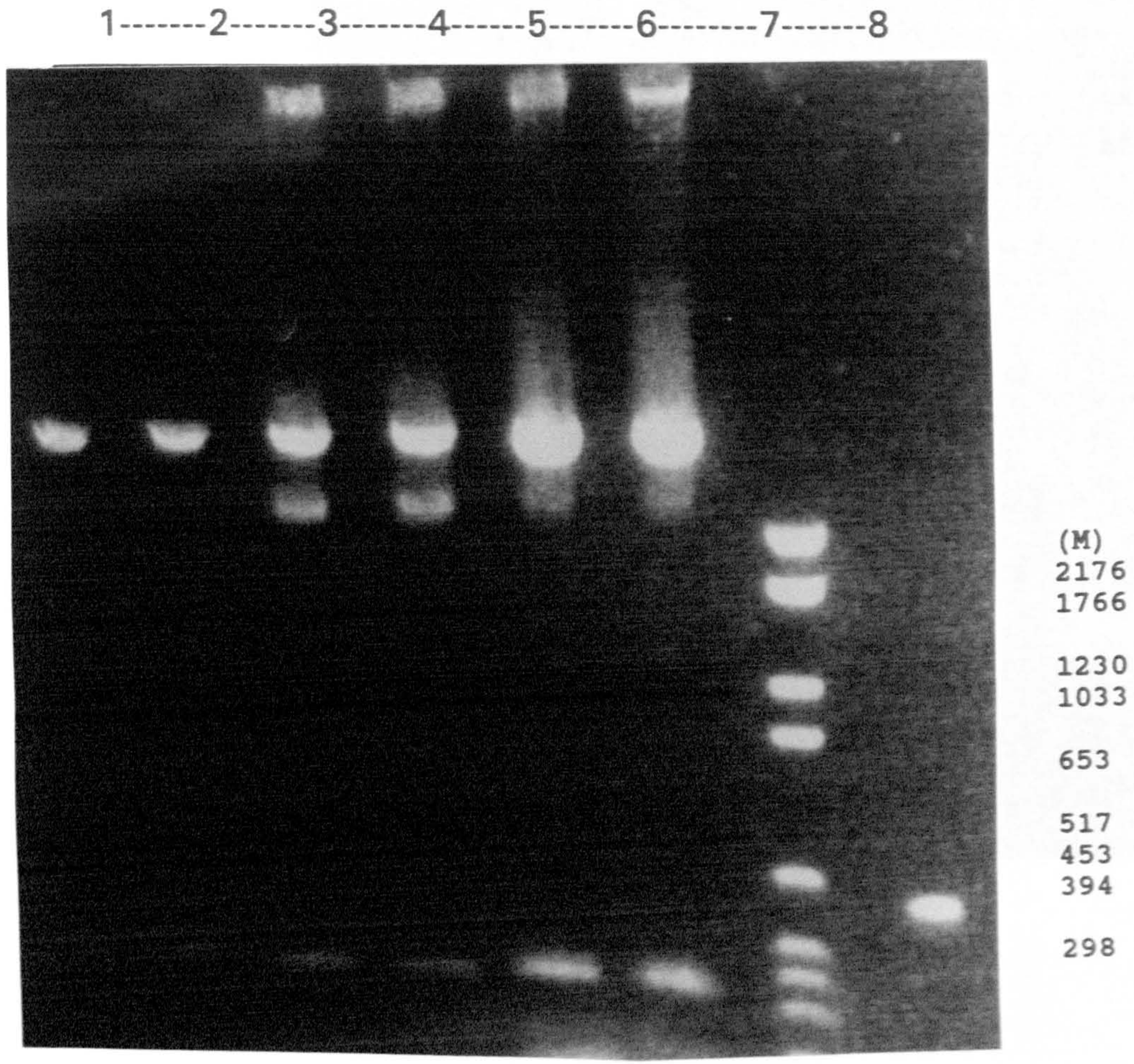
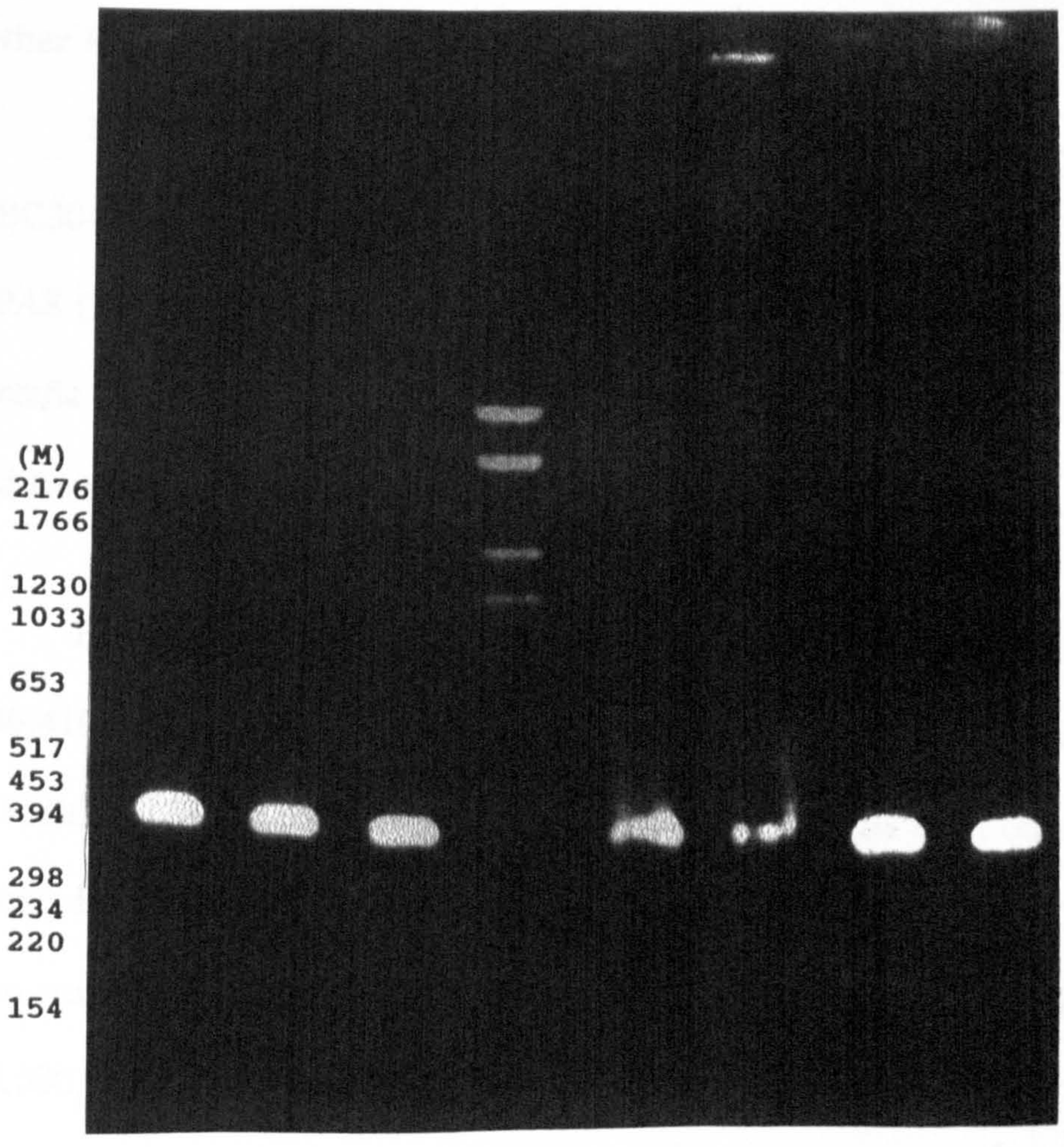


Fig. 5.9. RAPD primer AB11 amplification (412bp) of cloned PCR product



This sequence has six interrupted reading frames with a GC content of about 50% . Two nine base repeat (CTCGGTTGA) separated by a (TC) present in positions 369-388, which is also a part of GenBank data homology to ubiquitin gene. Other repeats, such as $(GT)_4$ in position 319-326 and TGA, CCCG and TTAAT have been identified which are common to the polyubiquitin gene of *An. gambiae* (Fig. 5.10).

Other RAPD primers

RAPD primers AB1 (Fig. 5.11), AB12, UBC301 (Fig. 5.12), UBC302, UBC304 (Fig. 5.13), UBC305 (Fig. 5,14), UBC306 (Fig. 5.15), UBC308, OPA1 and OPA8 (Fig. 5.16) not only differentiated species A and B, but also identified region specific patterns for *An. culicifacies* from north and south Sri-Lanka, India (Delhi), and Iran.

For example AB1 primer produces 7 bands in species A, of which four (440,410,345,265bp) are specific for this species but individuals from Saravan have an additional 207bp band which is not present in specimens from other localities. In species B, Indian samples have 4 specific fragments (540,530,375,320bp), but those from north Sri Lanka are 600,500,365,275bp, while those in south samples are 610,500,330,282,170bp (Fig. 5.11).

UBC306 primer produced a smaller number of bands; a band of 760bp in India and south Sri-Lanka, a major 710bp band in north Sri-Lanka, and species A-specific pattern (370, 600, 760bp) in Iran samples with additional 850bp and 1110bp band in Chahbahar and Zahedan samples (Fig. 5.15).

The M13F primer also differentiated species A and B, and showed intra-specific variation specially in Iranian samples from seven different collection sites

Fig. 5.10- *An. culicifacies* 412 bp common band (AB11 primer product) in species A and B. Underlined is the primer forward and reverse sequence. Double underlined show the two nine base repeat (CTCGGTTGA) separated by a TC. $(GT)_4$ in position 319-326 and TGA, CCCG and TTAAT repeats are bold.

GTAGACCCGT GAGTGGATCC GGTGCGGTCG TGAATATTTA TTGCCCGGCG
ATGAACAGTA CGCGGAATTG GCTTAATCCC GTGTCACCGG ACTGGCCTCA
TCATTAATCA TTGATCGGAG TGTAGTGAAA CCGAAAATAC ACTTTAAAAA
GTAGCTAGCG TACCGCGGCG TGTTTGTACT CGCTTGCATG GGGAGCTGTA
ATGCATCCCG CGGCCCATCC ACAGCAGTAA AGTCACCCGT AATGAATGTT
TAATGGCAGG AGCTCAACAG ATGGGACACC GTTGAGCTGA TTGATGCTTT
TTTGGGCGCG GAGGTTTGGT GTGTGTCGAT AACTGGCGTG TATGGAGAGT
TTTGTTTGAA ATGTTGTACT CGGTTGATCC TCGGTTGAGT ACGGCCGATG
GGACGGGTCT AC

Fig. 5.11- AB1 primer products in Sri-Lanka north (lanes 1-3), Sri-Lanka south (lanes 4-7), India (lanes 8-9), Iran (Saravan, lanes 11-12), Iran (Iranshahr, lanes 13-14), molecular marker (lane 10)

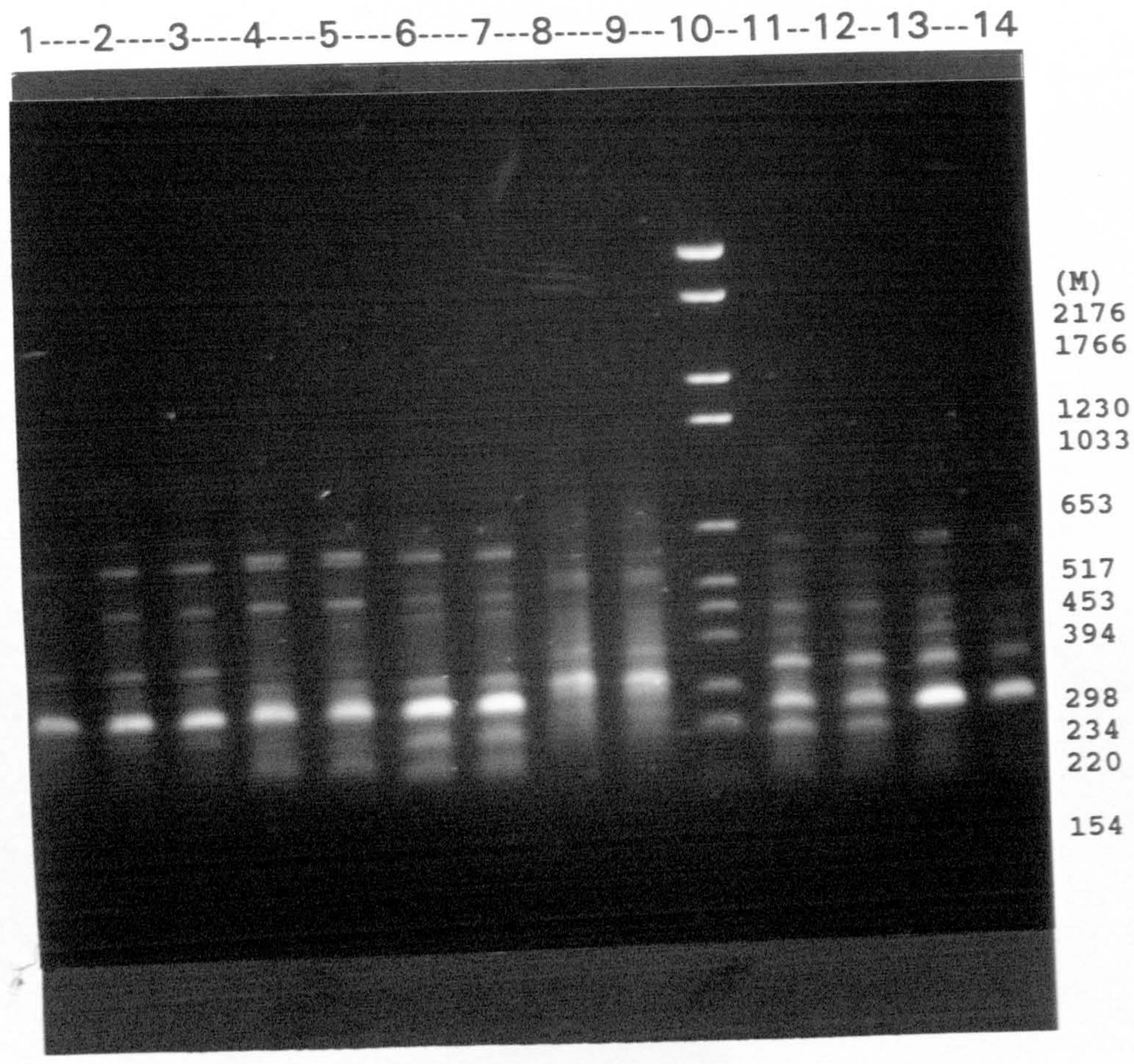


Fig. 5.12- UBC301 primer products in Iran: Chahbahar (1), Saravan (2), Nikshahr (3-4), Iranshahr (5); Sri-Lanka: north (6-8), south (9-10), molecular marker (11).380bp and 970bp species B specific and a 500bp A specific.

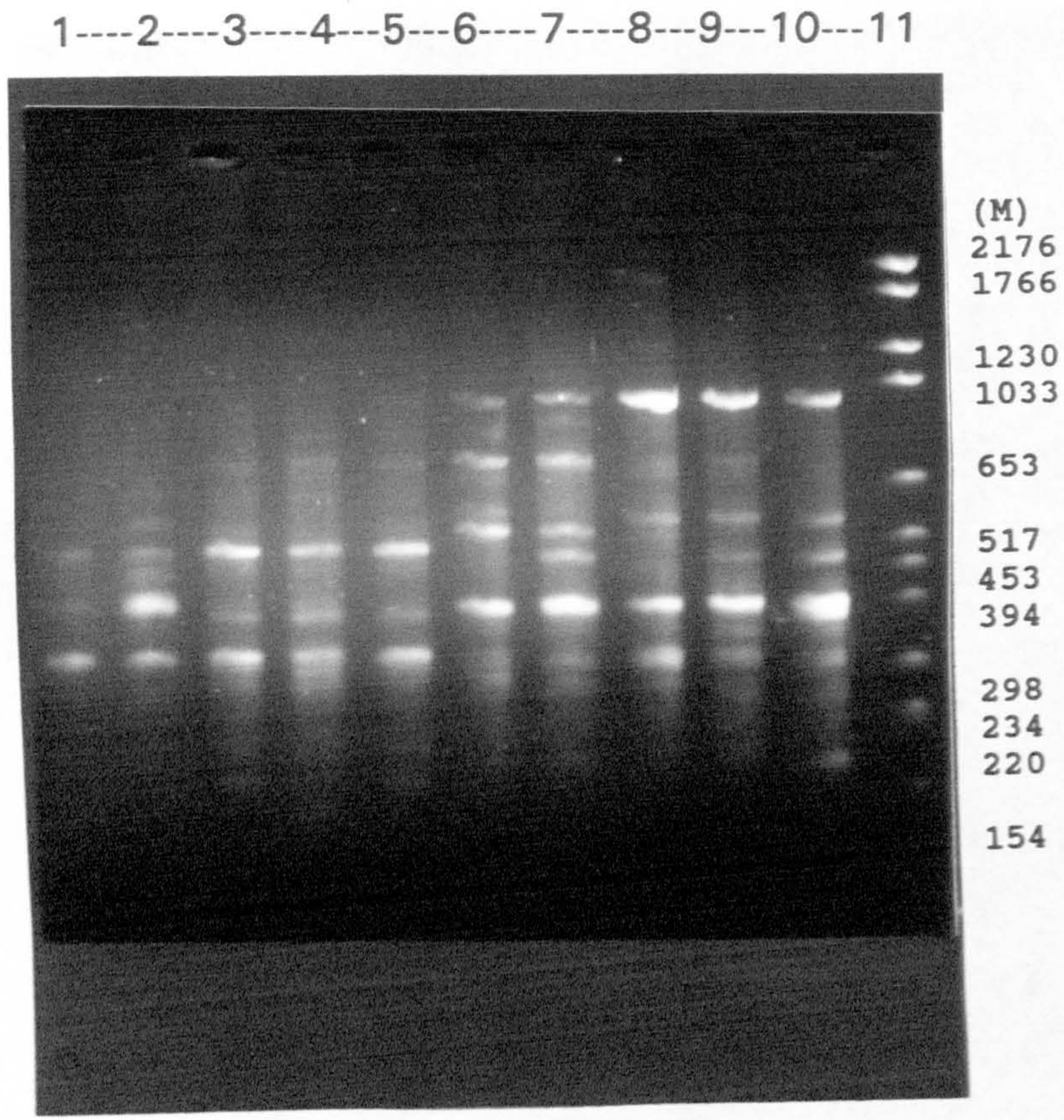


Fig. 5.13- primer UBC 304 bandding patterns in Sri-Lanka: south (1-4), north (5-8); India (9-10); Iran: Iranshahr (12-13), Saravan (14-15), molecular marker (11), -ve control (16)

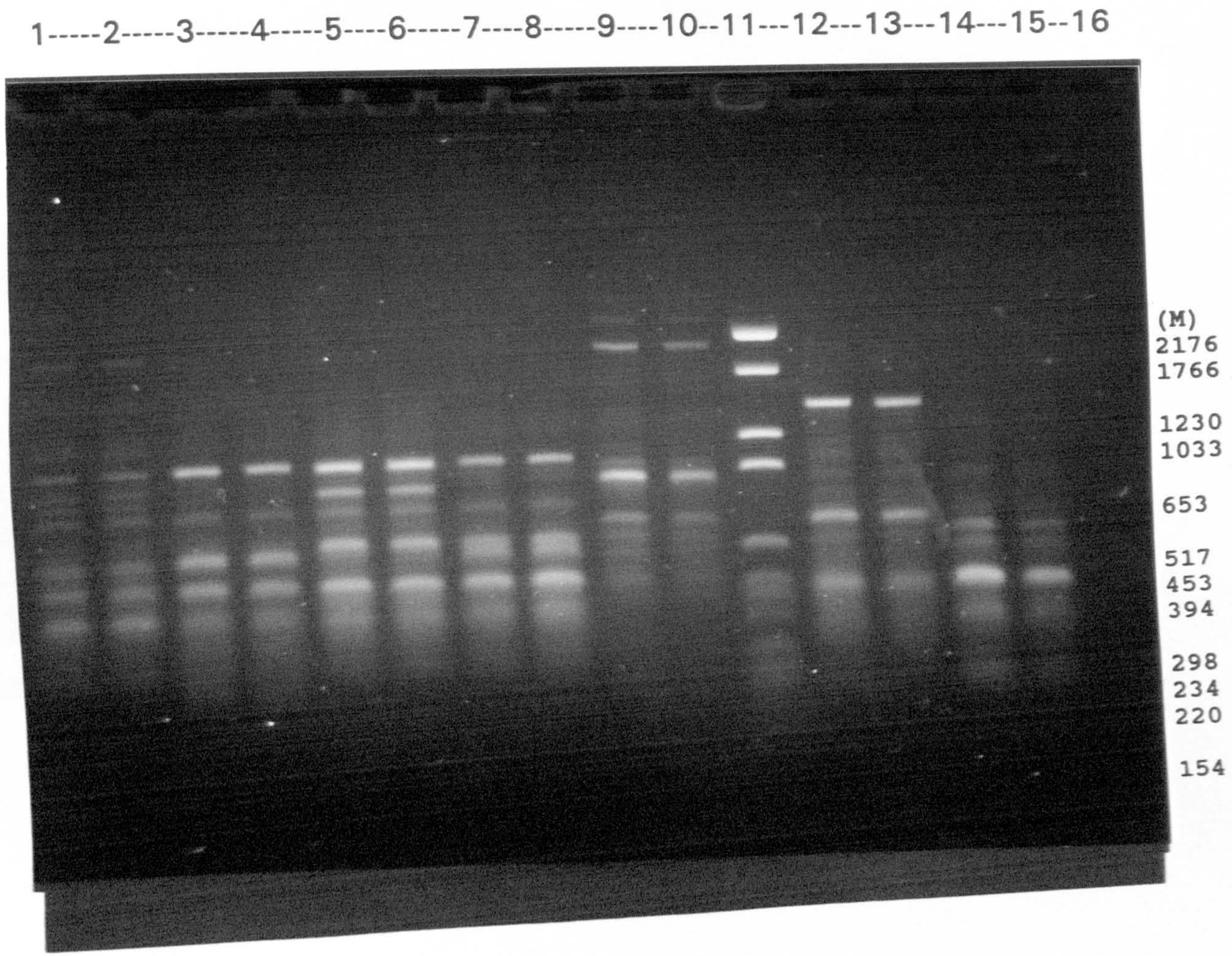


Fig. 5.14- UBC305 products in Iran: Zabol (2,3), Nikshahr (4-5); Sri-Lanka: north (6-7), south (8-9), mix DNA north & south (10); molecular marker (1). Mix DNA has both north and south pattern

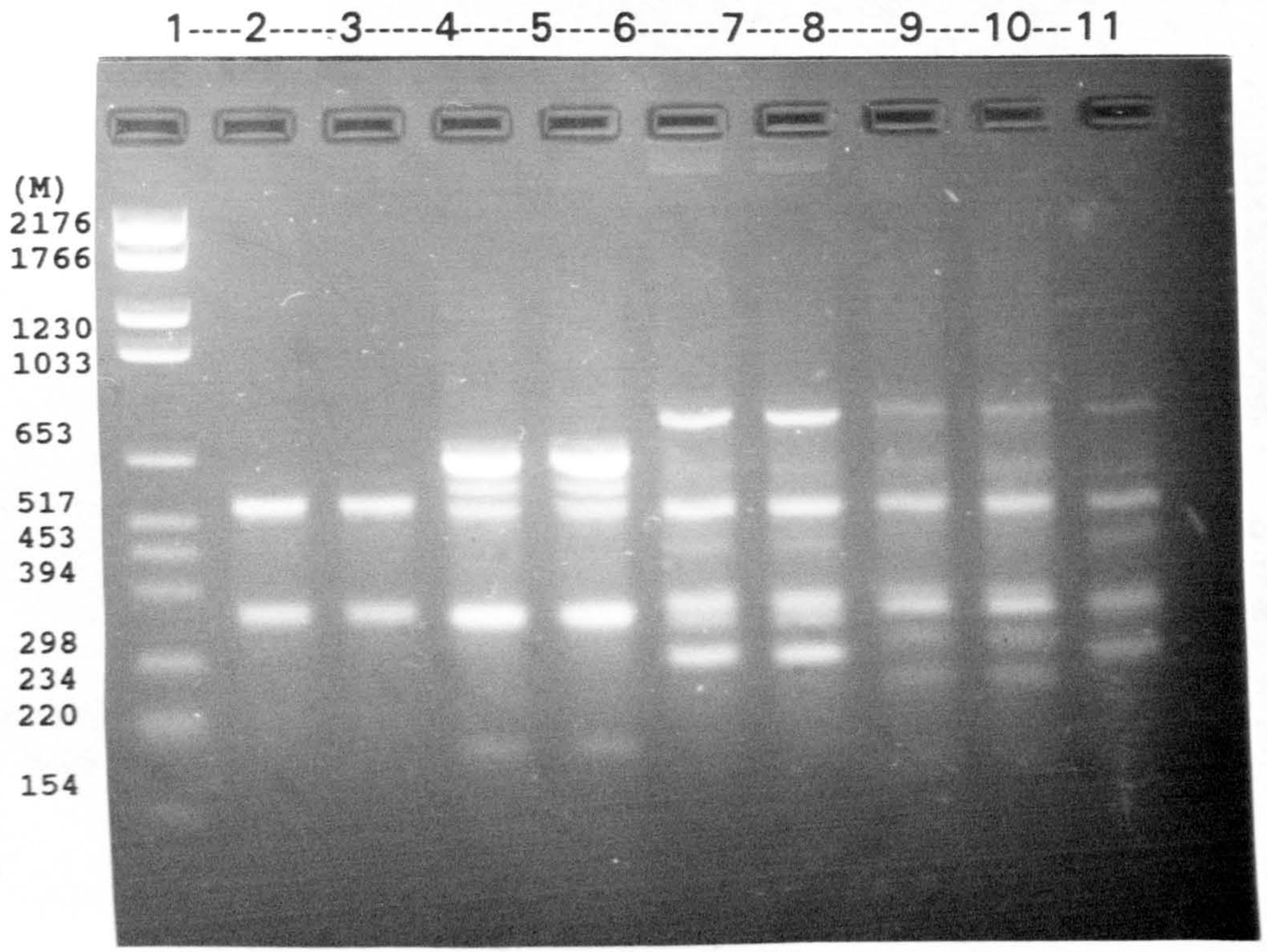


Fig. 5.15- UBC 306 products in India (2,3); Sri-Lanka: north (4,5), south (6,7), mix DNA of north and south (8); Iran: Chah-Bahar (10,11), Iranshahr (12-13); An. gambiae s.l. (15-16); molecular marker (1,9,14); -ve control (17)

* An. gambiae as outgroup has no similar band at all with culicifacies
* mix DNA of north and south has both populations pattern
* variation within Iran1 populations

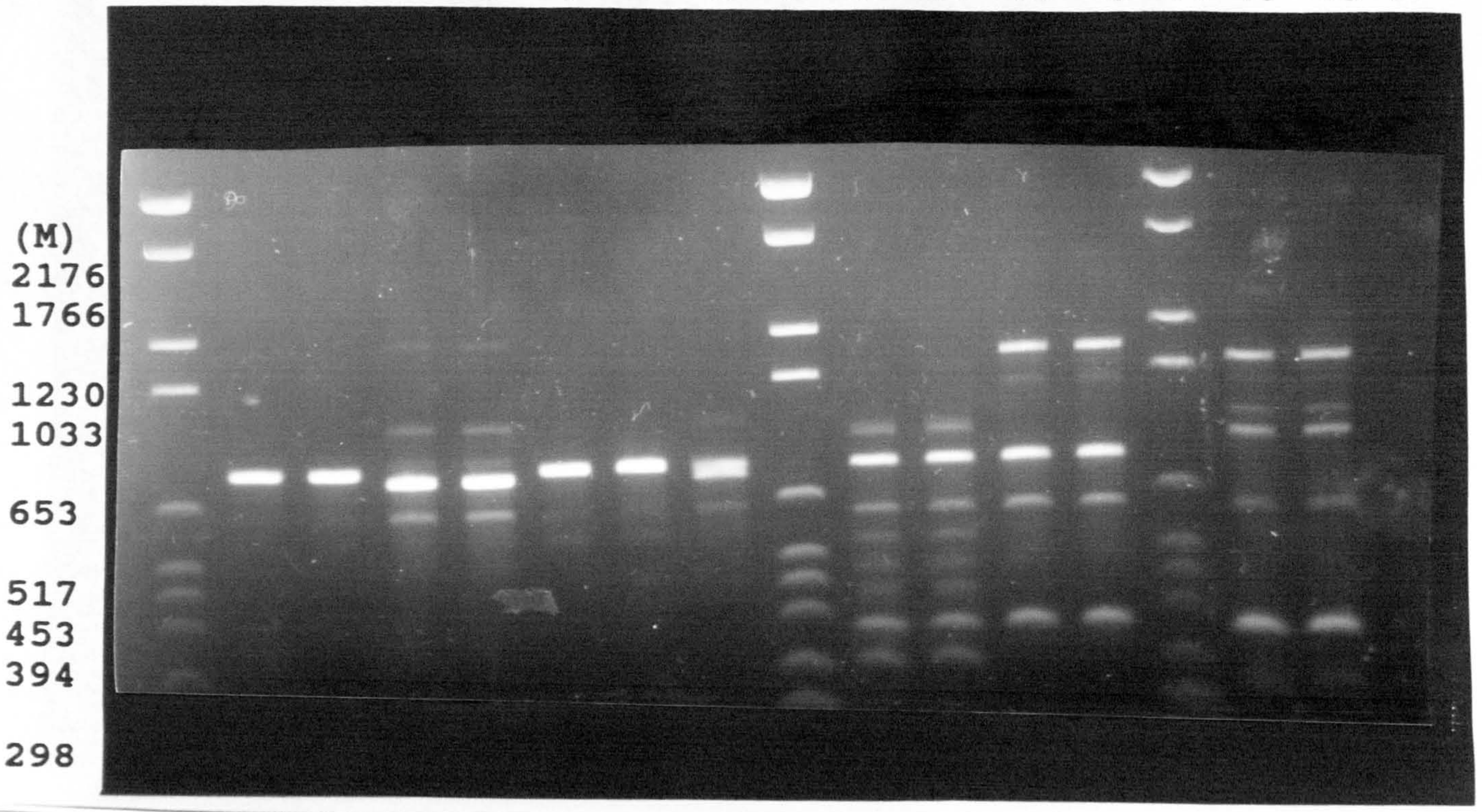
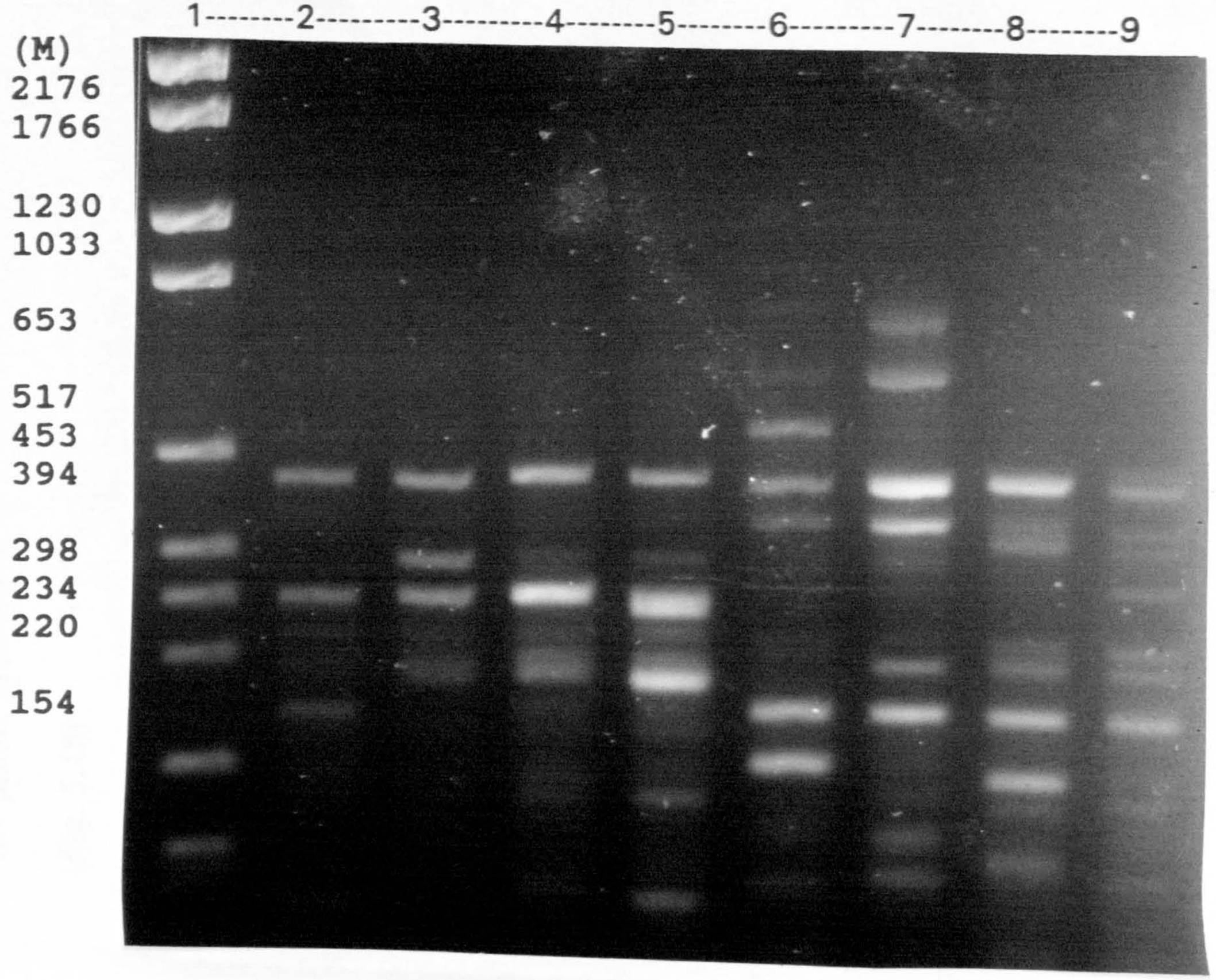


Fig. 5.16- primer OPA8 products in Iran: Saravan (2), Chah-Bahar (3), Nikshahr1 (4), Nikshahr2 (5); Sri-Lanka: north (6-7), south (8-9)); molecular marker (1)
* variation within Iran1 and Iran2 populations
* variation within Nikshahr1 and 2 populations



(Fig.5.17).

AB4 primer separated the two species but did not reveal any variation within species (Fig. 5.18).

'That RAPD PCR products may vary depending upon primer concentration' is shown clearly with primer AB19. An amplification reaction containing 20ng of primer in 25ul, in a3-step program (15 cycles of 45°, 39°, 35°C- Prog 7), produce a single band in both species, but its size in Iran and Sri Lanka (north samples) is 525bp and in Sri-Lanka (south) 565bp (Fig. 5.19). However a 25ul mix with 50ng of this primer produced three different sizes of this major band in different individuals: 525bp in Sri Lanka,north; 525 and 535 in Sri Lanka, south and India; and 525 in Saravan, 535 in Nikshahr and 555 in Zabol from Iran (Fig. 5.20). Furthermore, at the annealing temperature of 41°C this primer differentiates north and south Sri-Lanka populations (Fig.5. 21).

Primer UBC301 (Fig. 5.22) detected variation within *An. culicifacies* populations collected over a period of 15 nights (1st to 15th August 1994) from Azakhil in Pakistan. Although UBC 343 and UBC 344 differ only in a single base (T<-->A), but there is no similarity in their RAPD patterns (data not shown).

5.4.3 SSR PCR

Following PCR with $(GT)_7ATCC$ (hereafter GT1) primer, 'Fingerprints' characteristic of the two species were obtained (Fig. 5.23). Another SSR primer AA19, showed intra-specific variation within species B (Sri-Lanka, north ,south and India), although their pattern is different from species A (Fig. 5.24). GT2 repeat

Fig. 5.17- primer M13 products in Iran and Pakistsan populations of *An. culicifacies,* Iran: Zahedan (2), Khash (3), Iranshahr (4), Chahbahar (5), Nikshahr (6), Zabol (7), Saravan (8), Zahedan (9); Pakistan (11-13); molecular marker (10); -ve control (1)

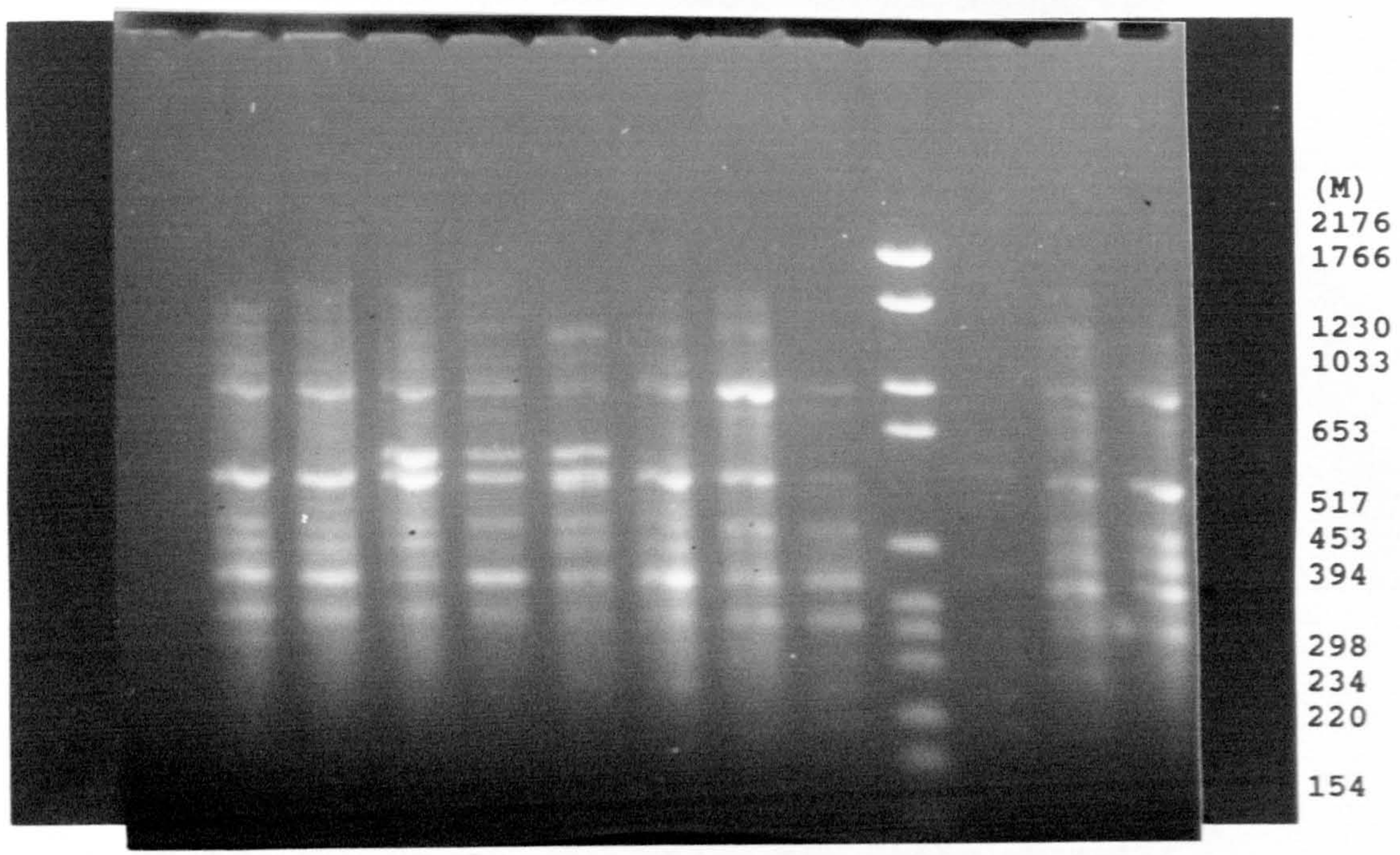


Fig. 5.18- RAPD-AB4 primer products in *An. culicifacies* species B from Sri-Lanka (1-5), and species A from Iran (7-11); molecular Marker (6)

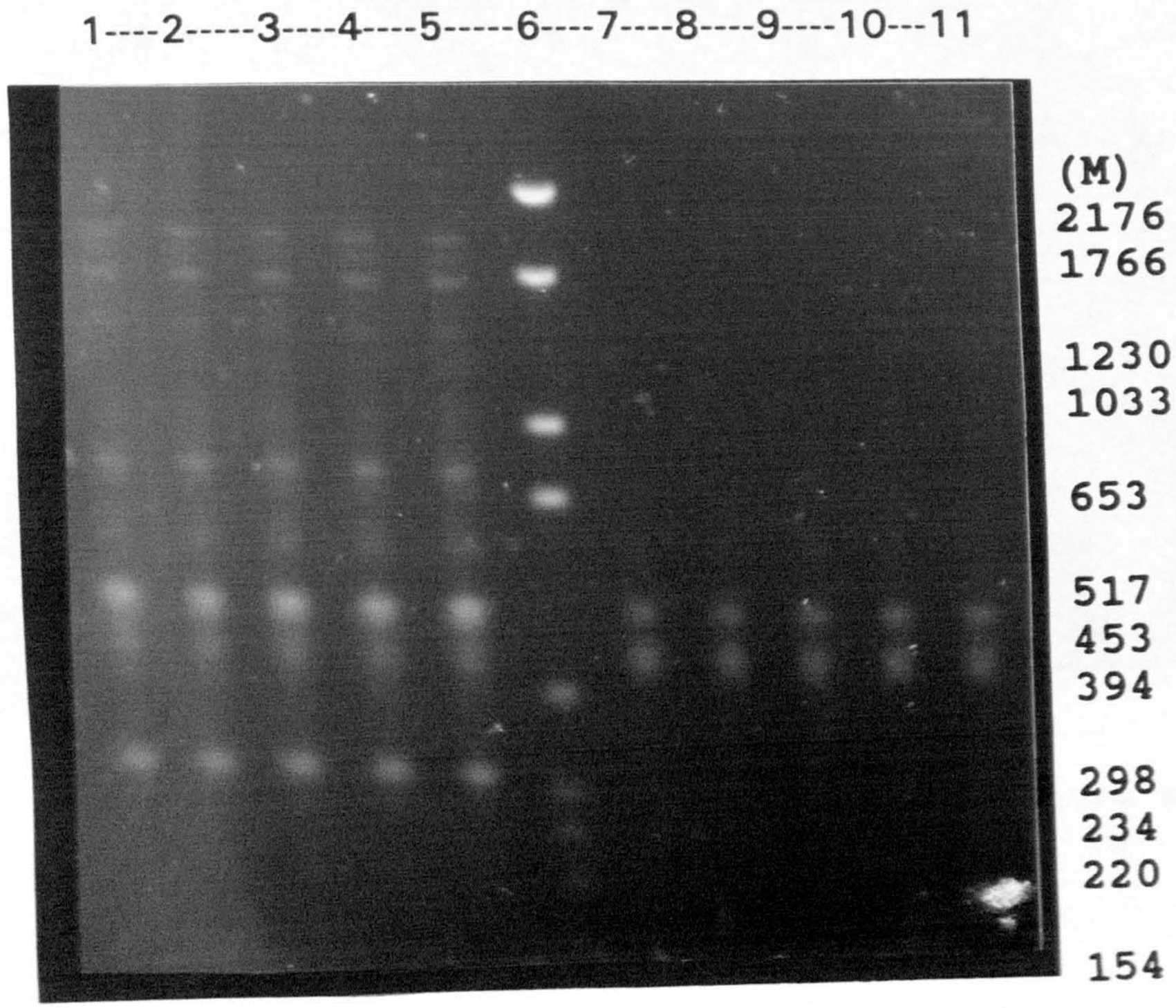


Fig. 5.19- RAPD (3 steps annealing) primer AB19(20ng) products in Sri-Lanka: south (1-3), north (5-7); Iran (9-11), molecular marker (lanes 4,8)

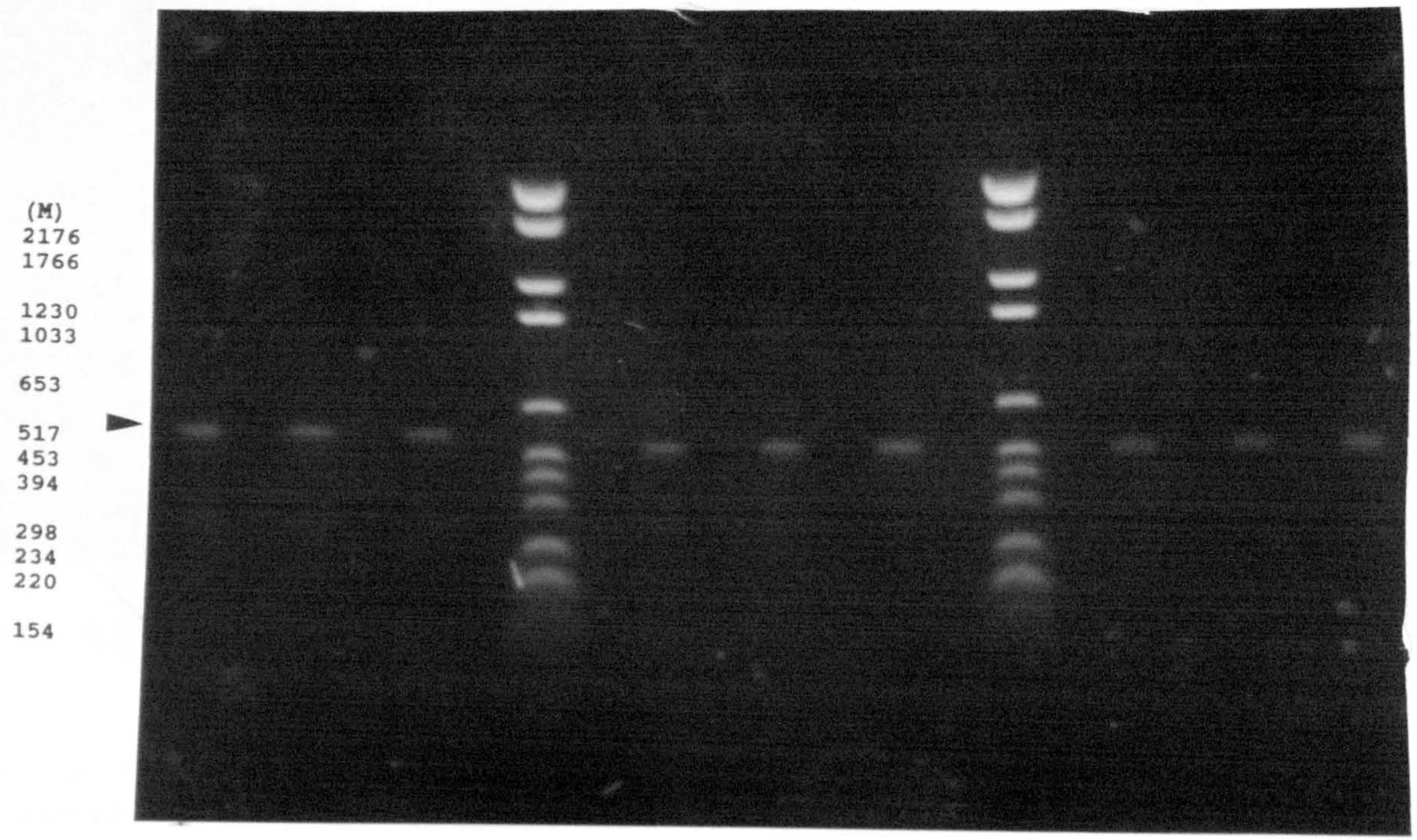


Fig. 5.20- RAPD (3 steps annealing)- AB19(50ng) products in Sri-Lanka: south (1-8), north (9-13); Iran: Saravan (15), Iranshahr (16-17), Nikshahr (18-20); molecular marker (lane 14)

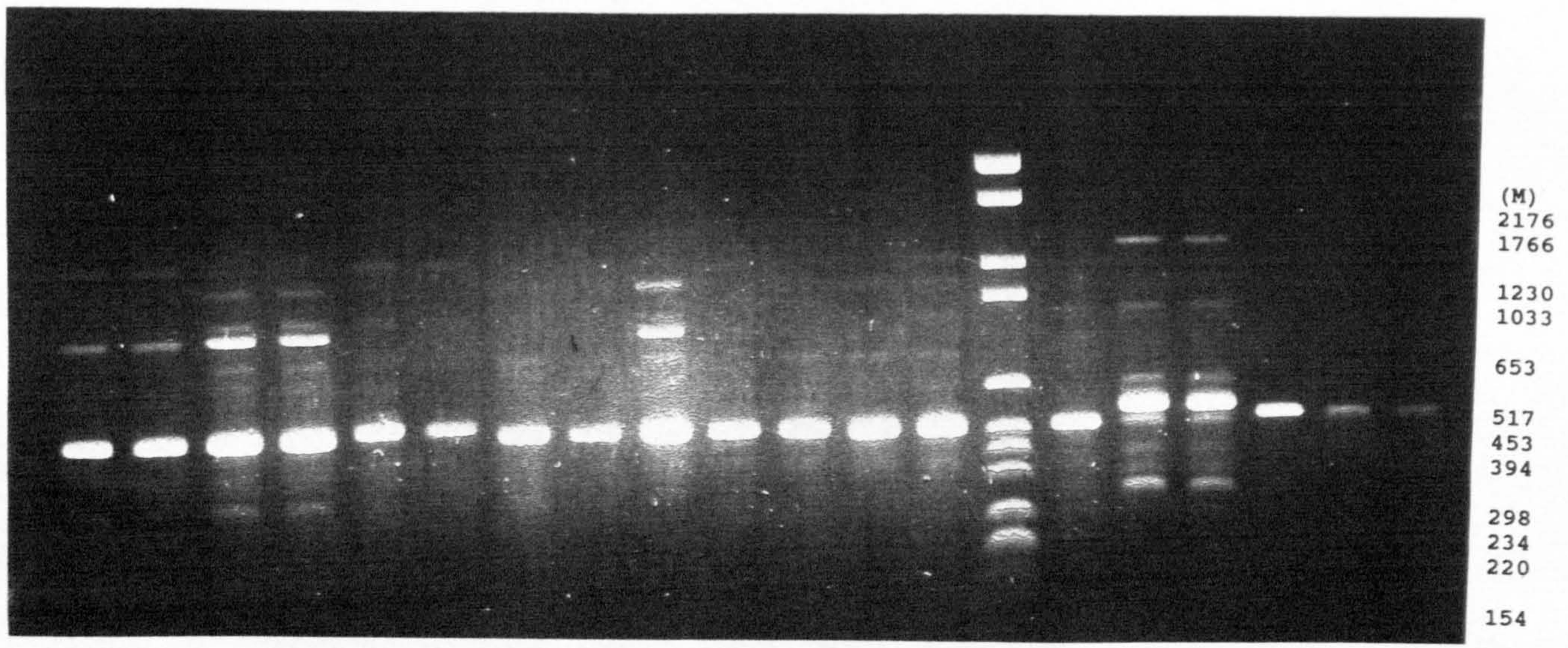


Fig. 5.21- RAPD- 41°C annealing- AB19(50ng) primer products in Sri-Lanka: south (1-5), north (7-11); molecular marker (6)

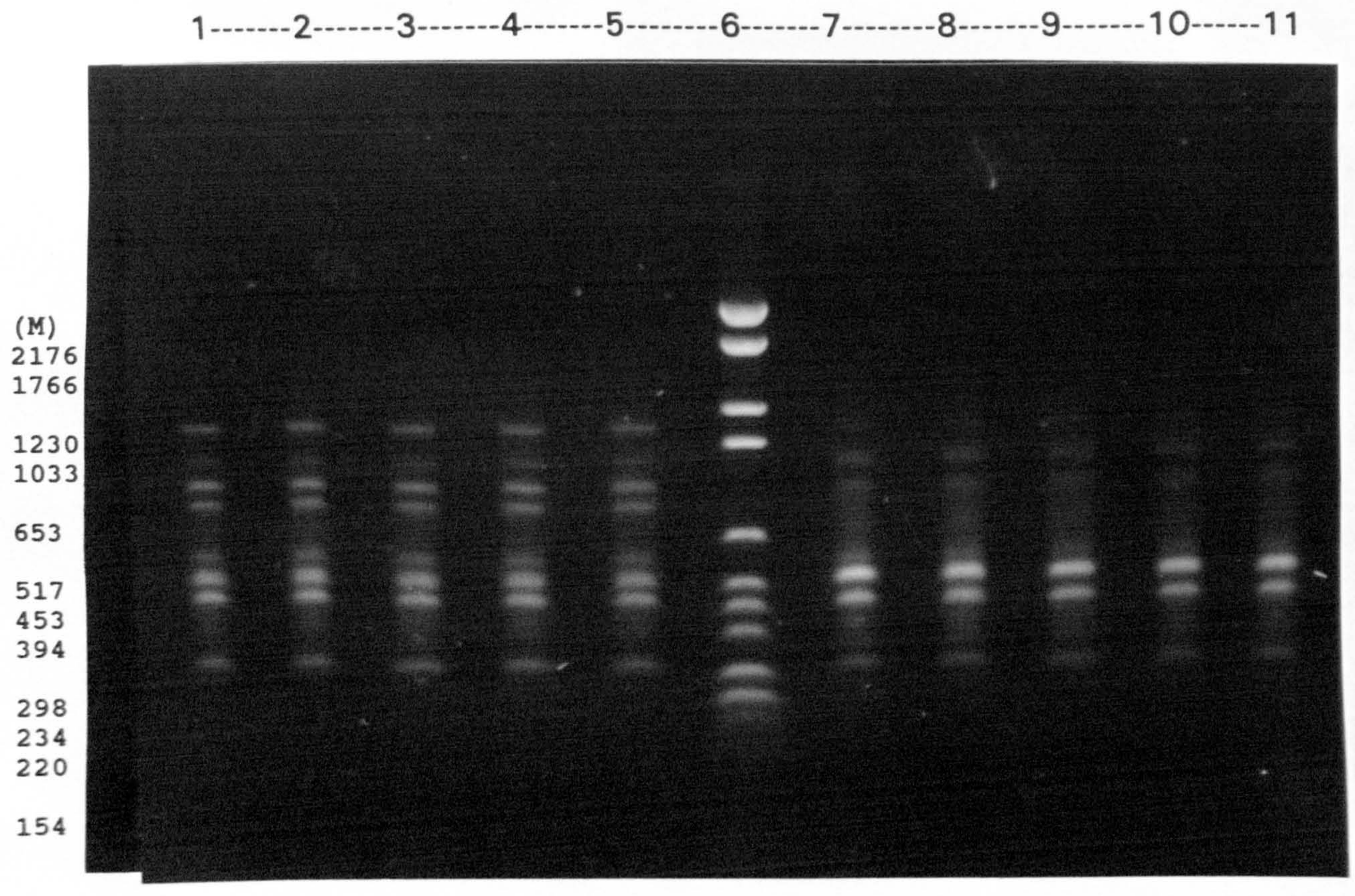


Fig. 5. 22- RAPD primer UBC301 products in *An. culicifacies* populations from Pakistan (Azakhil) collected during 15 nights from 1-15.8.94 (2-11, 13-20); molecular marker (12); -ve control (1)

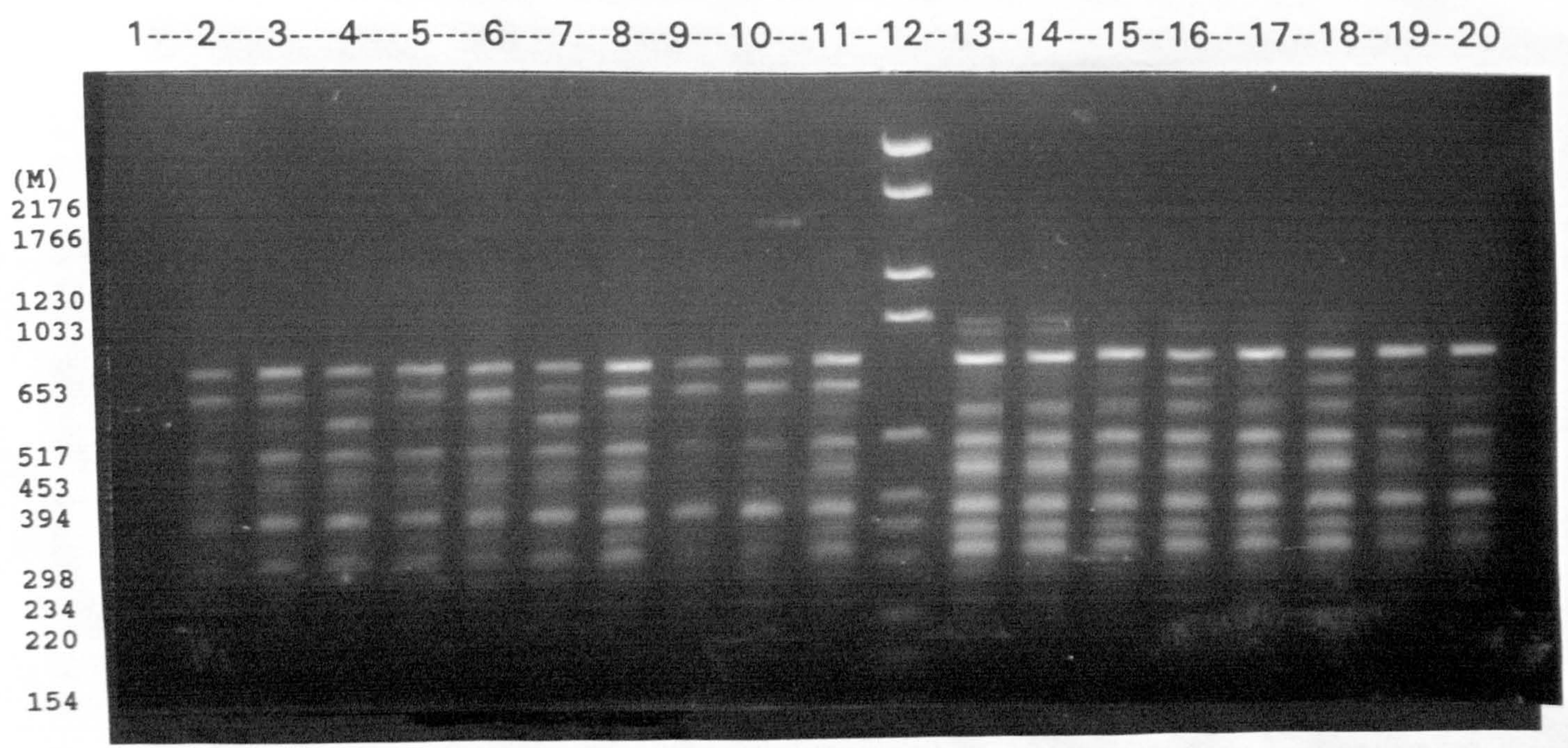


Fig. 5.23- SSR primer GT1 fingerprints in species A: Iran (1-2), Pakistan (3-4); and species B from Sri-Lanka (6-7), molecular marker (5), -ve control (no DNA, 8)

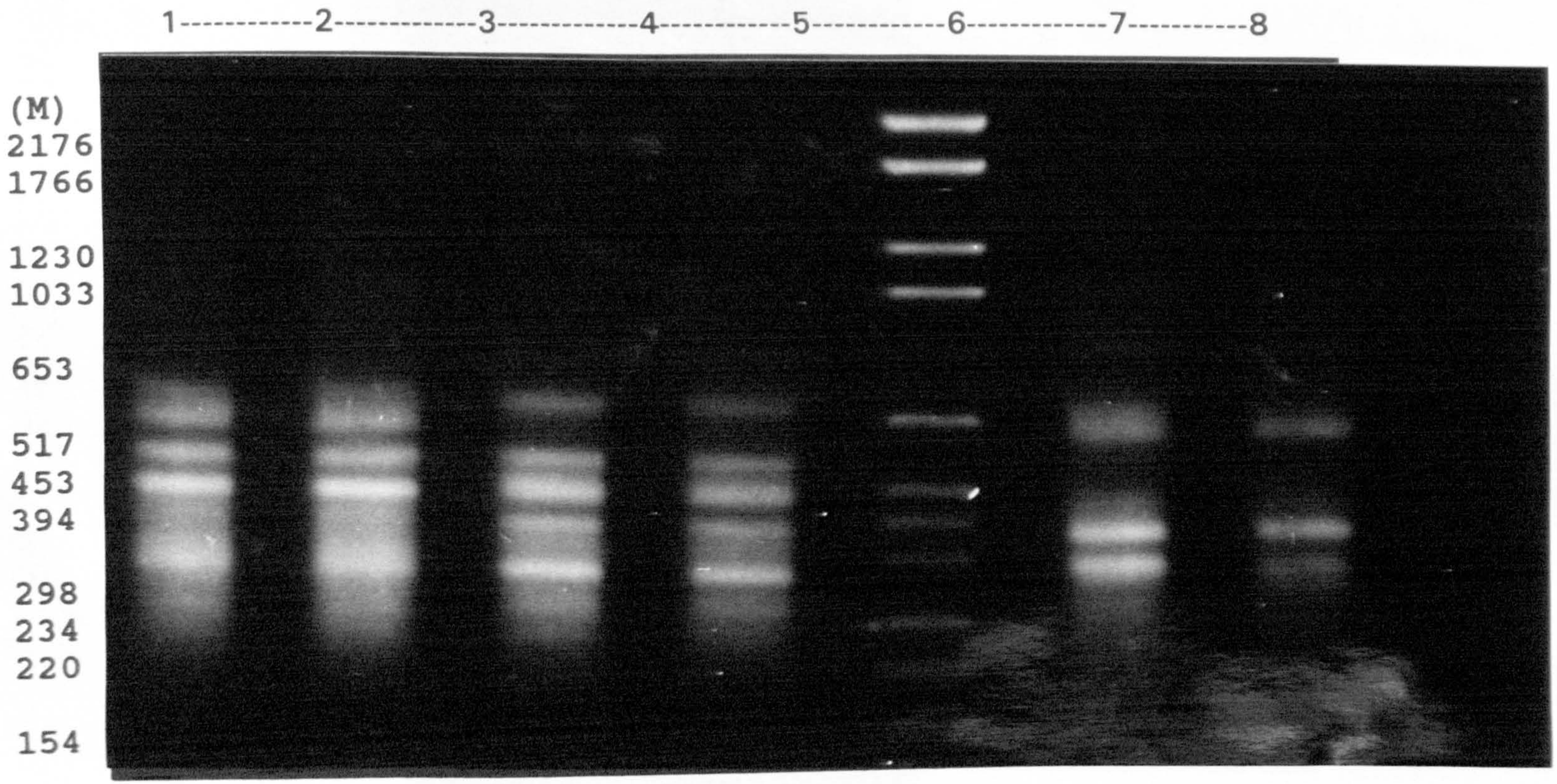
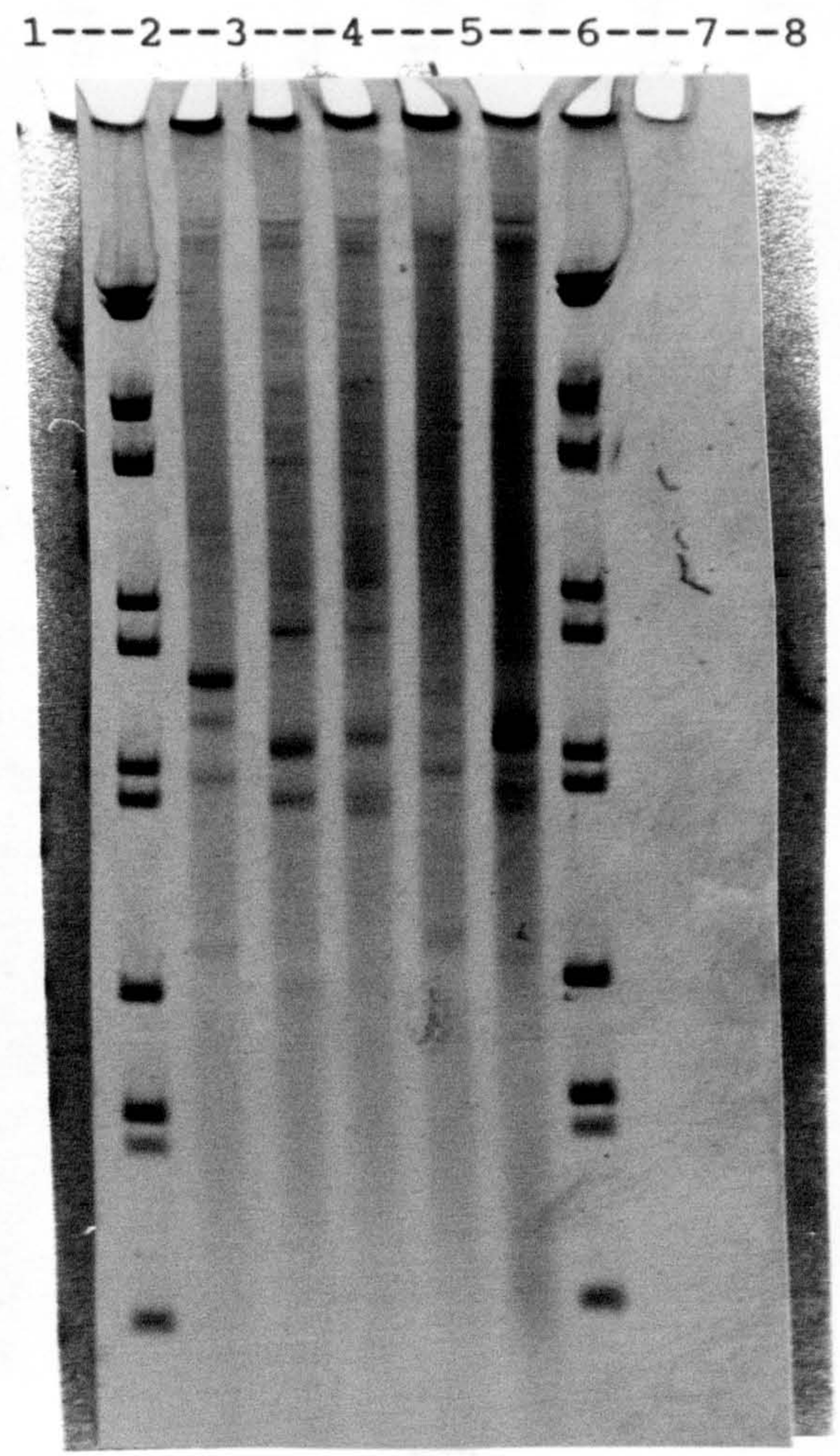


Fig. 5.24- SSR primer AA19 products in Iran(2), Sri-Lanka north (3), Sri-Lanka south (4), India (5), mixed DNA of Sri-Lanka south and north (6), molecular marker (1,7), -ve control (8)



primer also differentiate species A from B and revealed variation within two species (Fig. 5.25).

5.4.4 ITS2 PCR

Seasonal activity of *An. culicifacies* in Zainoddin village in the district of Ghasreghand (Fig. 5.26), where no anti-mosquito control measures have been undertaken, revealed that *An. culicifacies s.l.* has two peaks of activity (Fig. 5.27)(after Zaim *et al.*, un- published data), one in April-May (main peak) with rice fields as the main source of breeding, and the other in October-November when larvae are usually found in streams and stream fed pools in the absence of rice fields (Zaim et al. 1993). Mosquito collections corresponding to these two peaks and their associated differences in ecology have been tested to identify any genetic differences between these two groups. Firstly, RAPD profile with primers UBC306 and AB19 showed unusual pattern with the second peak collections in compare to other specimens from Iran.

ITS-2 region has been amplified in two individuals from each peak. The size of the amplified band, initiation and ending sequence in all specimens is identical to ITS-2 sequence of *An. culicifacies* species A collected from Iranshahr, Chahbahar and Saravan in Iran. However, the sequence differences between individuals of *An. culicifacies* collected from the second peak (9T and 5X) with first peak (4X) and is about 50% and this is the same with other mosquitoes from Iran, species A from Pakistan, and also species B from Sri-Lanka. Fig. 5.28 shows CLUSTAL V multiple sequence alignment of three specimens from Nikshahr collected in first and second peak.

Fig. 5.25- SSR primer GT2 fingerprints in species A: Iran (1-4), Pakistan (5-9); and species B: India (11-13), Sri-Lanka (14-15); unknown populations from species A (16-20); molecular marker (10).

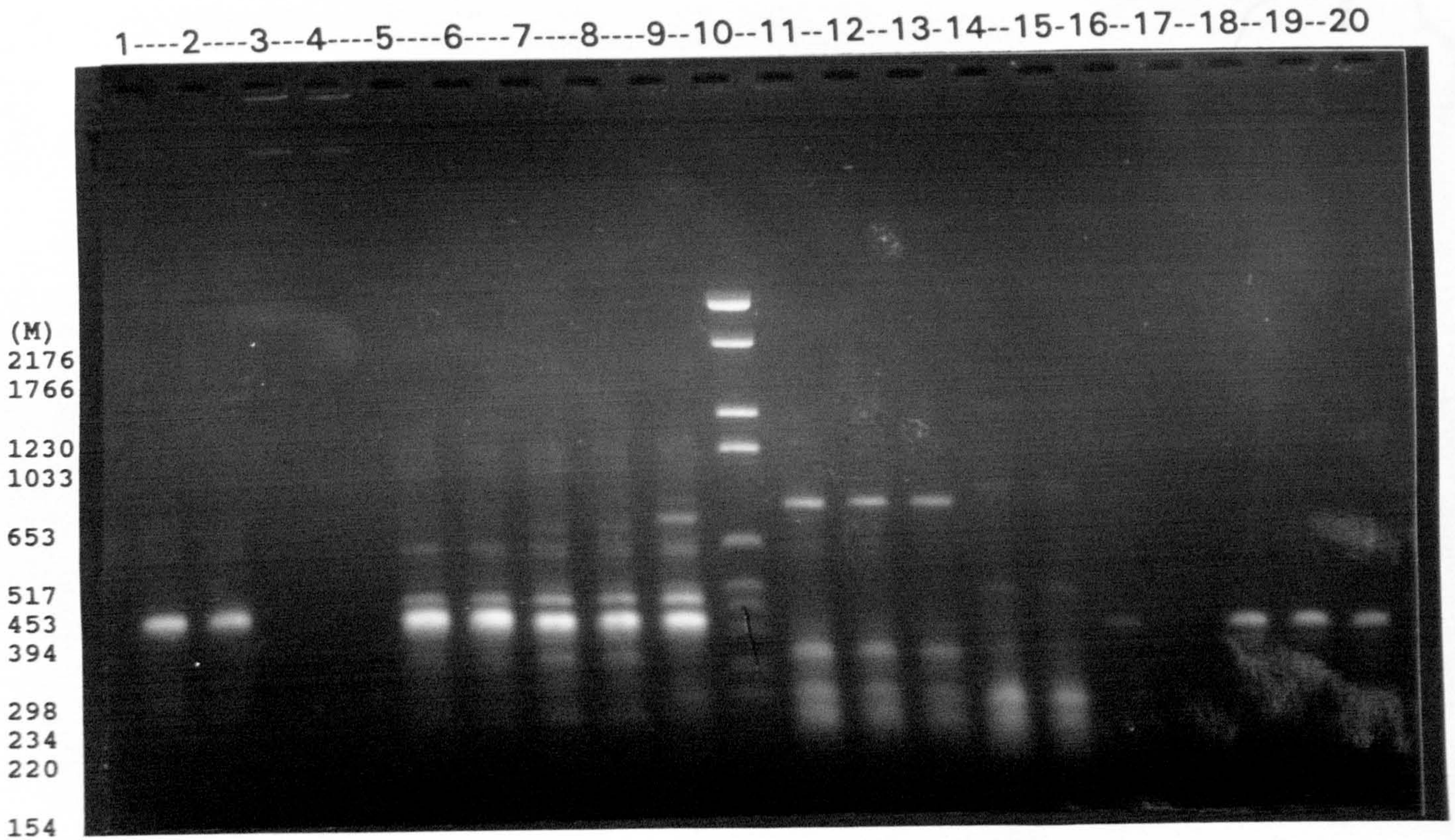


Fig. 5.26- Collection sites of Ac. culicifacies in the province of Baluchestan and Sistan, Iran (after Zaim et al. 1991)

Fig. 5.27- Mean indoor resting density (log) of *An. culicifacies* and reported malaria cases during 1990-1992 (Zaim, unpublished data). The presence of two peaks (April-May and October-November) perhaps corresponds to Iran1 and Iran2 forms in Ghasreghand, Iran.

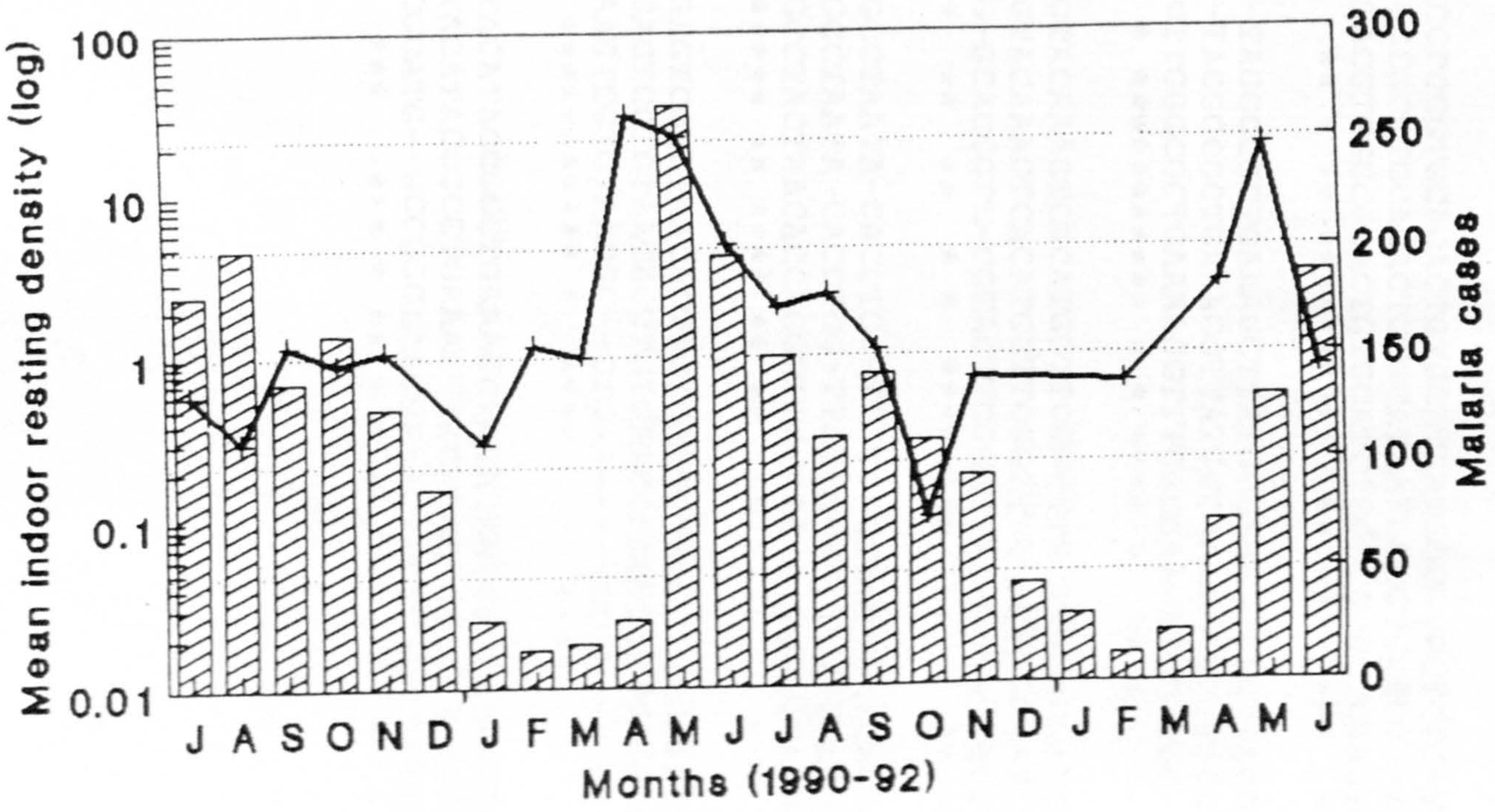


Fig. 5.28- CLUSTAL V multiple sequence alignment of three specimens from Nikshahr, Iran, 4X from first peak and other two (9T, 5X) from second peak.

```
IRCU9T      CAAATCCTTGTTACACAATAAACCTAACTACAGGGTGCCGCGTGCAG-CG
IRCU5X      CAAATCCTTGTTACACAATAAACCTAACTACAGGGTGCCGCGTGCAG-CG
IRCU4X      CAAATCCTTGTTACACAATAAACCTAACTACAGGGCGC-GCGTGCAGGCA
            ********************************** ** ******** *

IRCU9T      AACATA--CACCATGGCGAGCAGCCCGCC----------TGGTG--TC
IRCU5X      AACATA--CACCATGGCGAGCAGCCCGCC----------TGGTG--TC
IRCU4X      AGCAAAAACACCGGGACGAGCAGCCCGTCCCAACGCATAAGTGGTGGTTT
            * ** *  ****  * ********** *               *****  *

IRCU9T      GCTGGCCAACGCGCGTGCACACTGTGCATAATGGCGTGCTCGG-TCCTCC
IRCU5X      GCTGGCCAACGCGCGTGCACACTGTGCATAATGGCGTGCTCGG-TCCTCC
IRCU4X      GCAGCACTGTACGTGGGCGCACTGTGCATCTTGGCGTGCTCGAATCCCGC
            ** *  *    ** * ** **********  **********  ***  *

IRCU9T      ---GGGACT-TAGGGCGCTGAAAGGTTATGGCAAGTACTGTTACCTCAAA
IRCU5X      ---GGGACT-TAGGGCGCTGAAAGGTTATGGCAAGTACTGTCACCTCAAA
IRCU4X      AAGGGGGCTCTTGGGCGCTGAAAAGGTTTGGCGAA-ACGGGTGC---GCG
               *** ** * ********** * * **** *  ** *   *

IRCU9T      AGTACGCGTGGTACAAAGTCACATGGTTCGGCTTTGATAGACGATCTGAG
IRCU5X      AGTACGCGTGGTACAAAGTCACATGGTTCGGCTTTGATAAACGATCTGAG
IRCU4X      CGT-CAAGTCG-GCACGGTT-CGTACTTCGGCTTAG---GATGACCTGAT
             ** *  ** *  **  **  * *  ******** *    * ** ****

IRCU9T      ACCCC-GTTAGCCTAATA-CACCTGG-TTATCGCTAGTTTATCAGACCCA
IRCU5X      ACCCC-GTTAGCCTAATA-CACCTGG-TTATCGCTAGTTTATCAGACCCA
IRCU4X      TCCCCCGGCAGCCTACTAACACCAGGCTTGCCGAGATGGGTCCAGGGGTA
             **** *  ****** ** **** ** **  **  *       ***    *

IRCU9T      ATCTTCTATCGAGTCGTGTAACACTTGTGAACCCGGCCGATCGTGCTTAC
IRCU5X      ATCTTCTATCGAGTCGTGTAACACTTGTGAACCCGGCCGATCGTGCTTAC
IRCU4X      AG-TCCGGCCAAGTCGTGTAACGC--GTGA------CCGA---------C
            *  * *   * ********** *  ****       ****         *

IRCU9T      CTATGTGCAGTACATACCGCCTGAAAGCTACACAACCCCGTTAGCTAATA
IRCU5X      CTATGTGCAGTACATACCGCCTGAAAGCTACACAACCCCGTTAGCTAATA
IRCU4X      CCATACG-GGCGCATG--GCCGGGAAACCA-AC--CTTTACCACTTTACC
            * **  *  *  ***   *** * ** * * **  *       *  * *

IRCU9T      T
IRCU5X      T
IRCU4X      A
```

5.5 DISCUSSION

The study of molecular genetic variation in malaria vectors is in its infancy. In the major African vector group, the *An .gambiae* complex, several recent studies have used molecular techniques to address the problems of speciation, population structure and gene flow (Favia *et al.* 1994a, 1994b; Lanzaro *et al.* 1995; Lehman *et al.* 1996b, 1997; Djadid & Townson, in preparation), factors of considerable importance in malaria epidemiology and control. However, there is currently virtually nothing known about molecular variation at the inter- and intra-specific level in the malaria vectors of the Middle East and Indian sub-continent.

5.5.1 RAPD and SSR PCR

By using PCR with RAPD and SSR primers, we have found consistent differences between field specimens of species A and B of the *An.culicifacies* complex. Ten RAPD primers have shown intra-specific variation within these two species when populations from different geographical locations in Iran and Sri-Lanka are compared.

AB11 primer differentiated two species by species specific patterns in all examined specimens of species A and B. The result with this primer confirmed that a single RAPD primer can reliably distinguish species A and B. Also its Southern blot hybridized to probe shows sequence of 412bp band is share while that of 620bp is unique to species B. Squashed dot blot hybridization of DNA from *An. stephensi* and *An. bwambae* revealed although not strong but the same signal. Moreover, sequencing of the 412 bp band showed a high sequence homology with *An. bwambae* clone. This sequence contains two adjacent CTCGGTTGA repeat and others repeats like $(GT)_4$.

Sequence homology with genbank data showed its similarity to poly ubiquitin gene in *An.gambiae*.

AB19 primer in certain condition (three different annealing programme and less primer concentration) produced a single band but in two different size, first a 525bp which is share between species A from Iran and species B (north Sri-Lanka), and second, a 535bp in samples from south Sri-Lanka (species B) which we do not know is due to two different loci or only differences in copy number of a common part of genome or any variation happened in a specific gene that allow multiple site amplification.

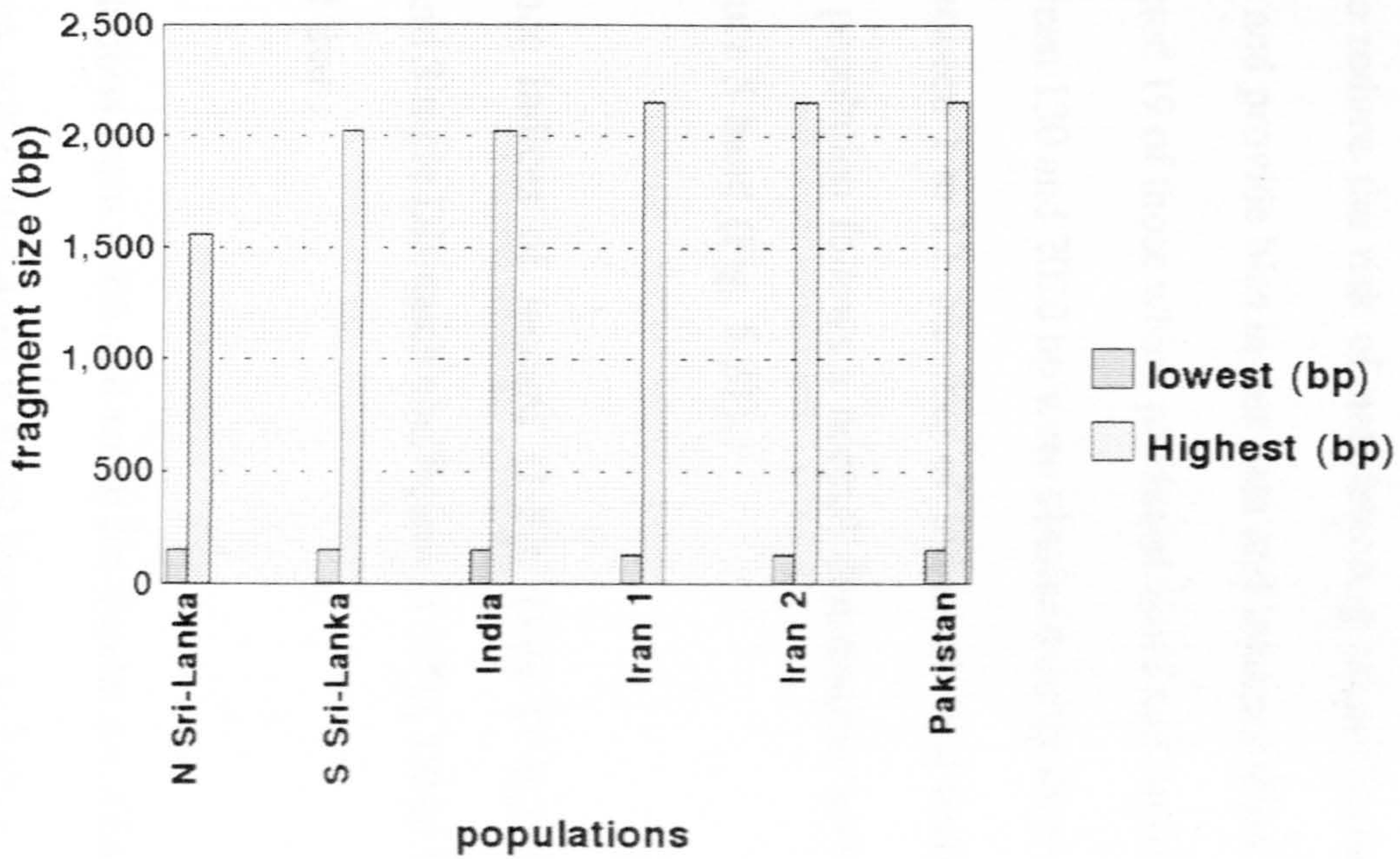
AB4 primer amplified two distinct patterns in species A and B with no intra-specific variation, hence this could be a useful species diagnostic probe.

SSR primer GT1 produced two distinct patterns for species A and B, although within species A (Iran and Pakistan) it showed a few variation. This primer also detected differences between brackish and freshwater populations of *An. subpictus* from Sri-Lanka which may represent species-specific differences (Djadid *et. al.*, in preparation). Two other repeat primers also are useful for looking to intra-specific variation in both species.

With respect to the size range in PCR products, it seems that there is no significant difference between populations of both species (Fig. 5.29). The ability of primers in producing molecular markers within populations was independent of the GC/AT ratio of the primers and the number of fragments amplified but is correlated

Fig. 5.29

An. culicifacies populations' lowest and highest size of fragments produced by RAPD primers



with range of products which increased by higher GC% till 70% (Fig. 5.30).

Varying the concentration of DNA, $MgCl_2$, Taq and primer in the PCR mixtures resulted in both quantitative differences and qualitative changes in RAPD bands and patterns. So PCR reaction mix has been optimized on using 5pg to 10ng of DNA, 2mM $MgCl_2$, 0.5 unit Taq and 20-100ng primer per 25ul reaction.

5.5.2 RAPD and SSR data analysis

In order to reduce the risk of not detecting amplification failure in samples during screening, and provide bias in our data and interpretation, we pre-screened 59 primers and selected 19 of those which produced more and reproducible bands. DNA fragments of between 130 and 2020 bp were obtained using these primers and number of bands range between 1 to 12. Number of fragments and their frequency in tested primers for each population follows a normal distribution with more frequency on primers that produce 5 band (Fig. 5.31).

The distance indices of Jaccard ($1-S_J$), Dice ($1-S_D$), Nei & Li's ($1-S_N$), Montpelier (S_{Jm}) and Simple Dismatch coefficient ($1-S^M$) (Table 5.3A-5.3F; and Figs. 5.32-5.37 showed that :

1) species B populations from north and south Sri-Lanka are different and in compare to Indian specimens, south Sri-Lanka is more similar to Indian *An. culicifacies*.

2) Species A collected from 7 different localities in south-eastern Iran at least 200 Km apart each other came to two groups;

Fig. 5.30

An. culicifacies populations' lowest and highest size of fragments produced by different GC content primers

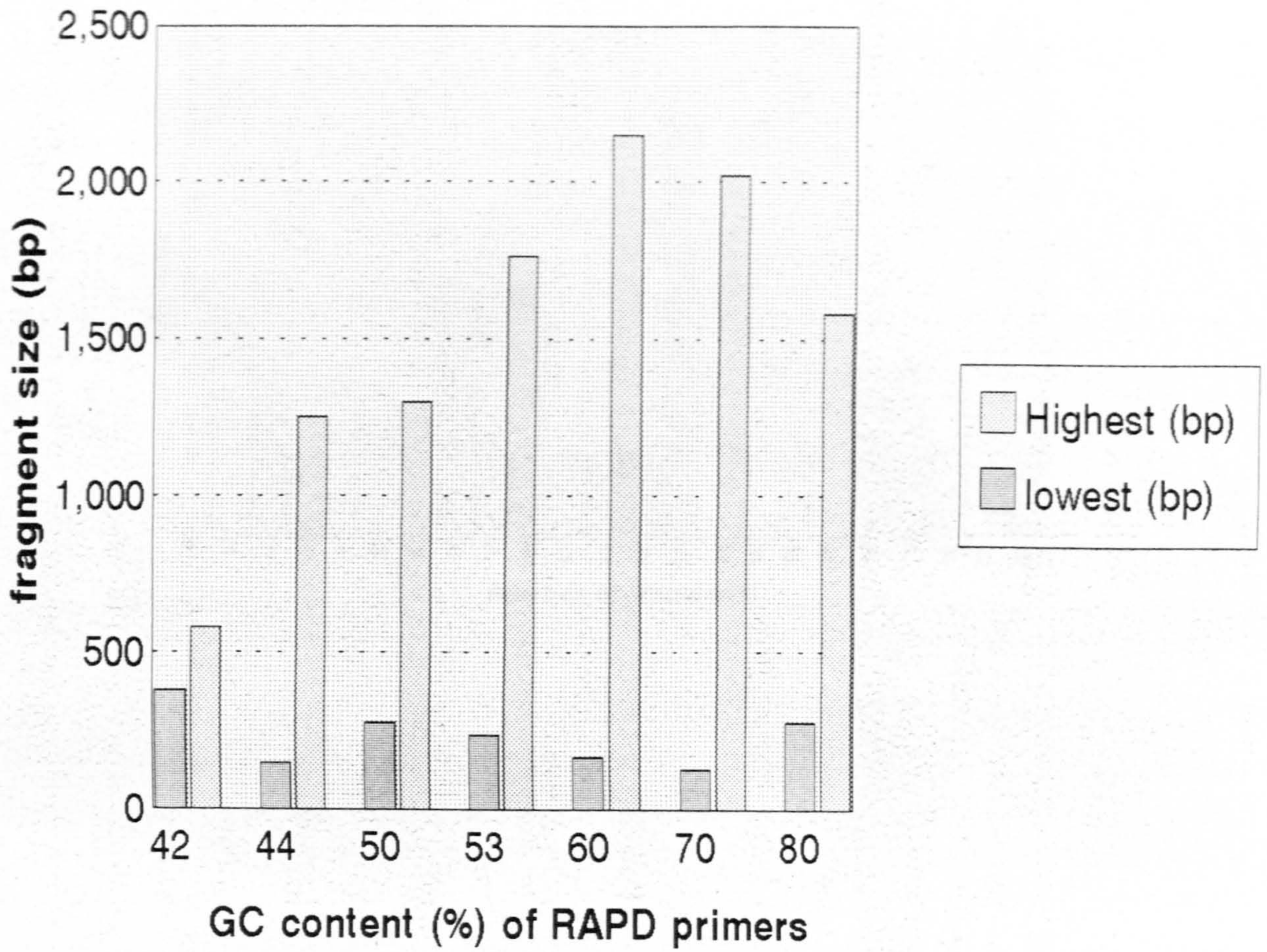


Fig. 5.31

Observed distribution of the number of fragments amplified for *An. culicifacies* populations X primers combination

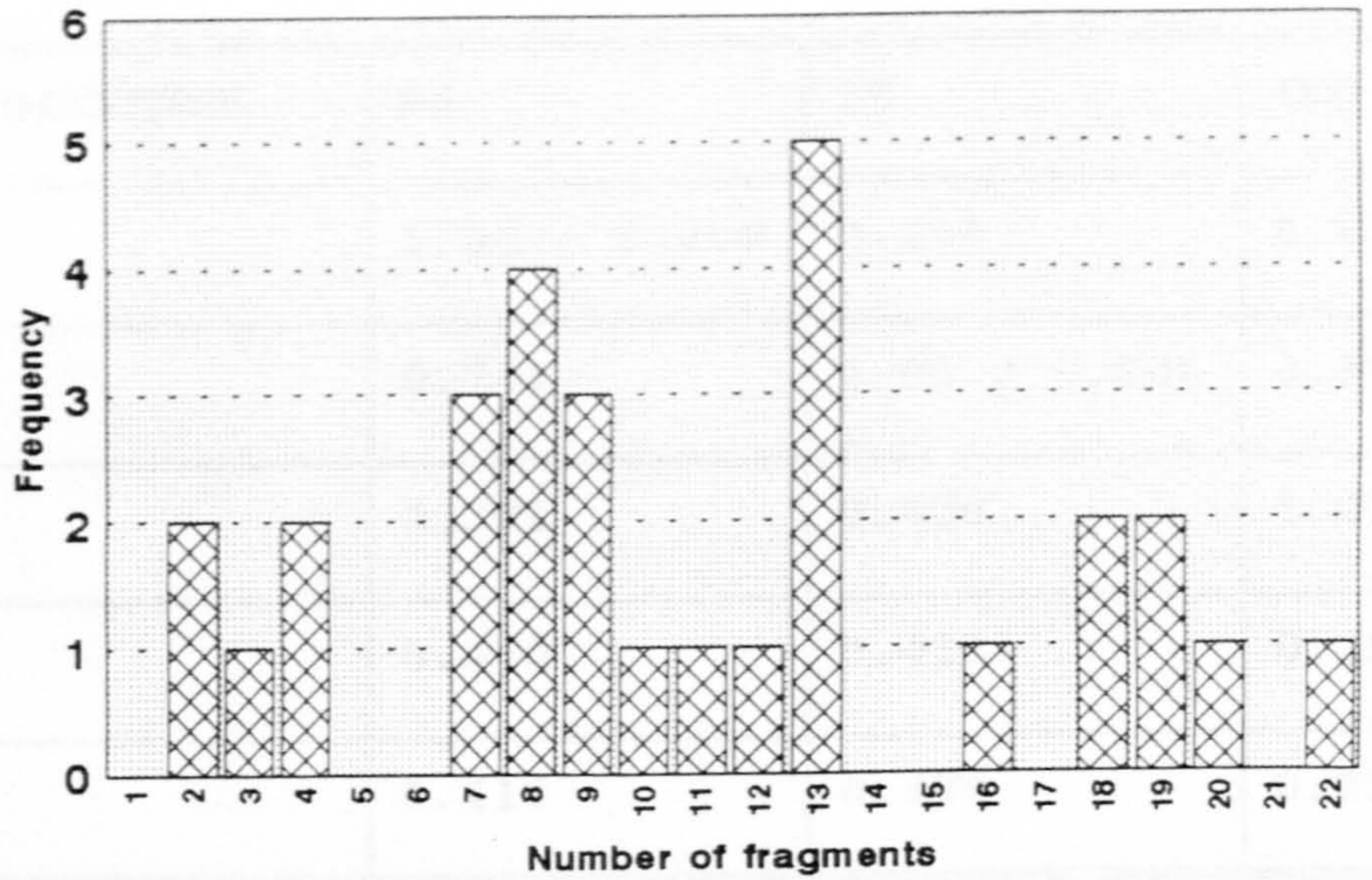


Table 5.3A

An. culicifacies populations:

Nei & Li's similarity/distance matrix:

SIMILARITY / DISTANCE	NS	SS	IND	NI	SI
NS	1.00 / 0.000	0.259	0.610	0.560	0.587
SS	0.741	1.00 / 0.000	0.580	0.590	0.616
IND	0.391	0.420	1.00 / 0.000	0.750	0.750
NI	0.440	0.412	0.250	1.00 / 0.000	0.160
SI	0.413	0.384	0.250	0.840	1.00 /0.000

NS: North Sri-Lanka, SS: South Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Table 5.3B

An. culicifacies populations:

Jaccard (S_j) similarity/distance matrix:

SIMILARITY / DISTANCE	NS	SS	IND	NI	SI
NS	1.00 / 0.000	0.539	0.773	0.720	0.740
SS	0.461	1.00 / 0.000	0.717	0.739	0.742
IND	0.227	0.283	1.00 / 0.000	0.831	0.828
NI	0.280	0.261	0.169	1.00 / 0.000	0.541
SI	0.260	0.258	0.172	0.459	1.00 /0.000

NS: North Sri-Lanka, SS: South Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Table 5.3D

An. culicifacies populations:

Dice (S_D) similarity/distance matrix:

SIMILARITY / DISTANCE	NS	SS	IND	NI	SI
NS	1.00 / 0.000	0.402	0.697	0.590	0.632
SS	0.598	1.00 / 0.000	0.544	0.612	0.623
IND	0.303	0.456	1.00 / 0.000	0.723	0.746
NI	0.410	0.388	0.277	1.00 / 0.000	0.417
SI	0.368	0.377	0.254	0.582	1.00 /0.000

NS: North Sri-Lanka, SS: South Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Table 5.3E

An. culicifacies populations

Montpelier similarity/distance matrix:

SIMILARITY / DISTANCE	NS	SS	IND	NI	SI
NS	1.00 / 0.000	0.614	0.824	0.811	0.814
SS	0.386	1.00 / 0.000	0.796	0.824	0.834
IND	0.176	0.204	1.00 / 0.000	0.847	0.792
NI	0.189	0.176	0.153	1.00 / 0.000	0.687
SI	0.189	0.166	0.208	0.313	1.00 /0.000

NS: North Sri-Lanka, SS: South Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Table 5.3F

An. culicifacies populations

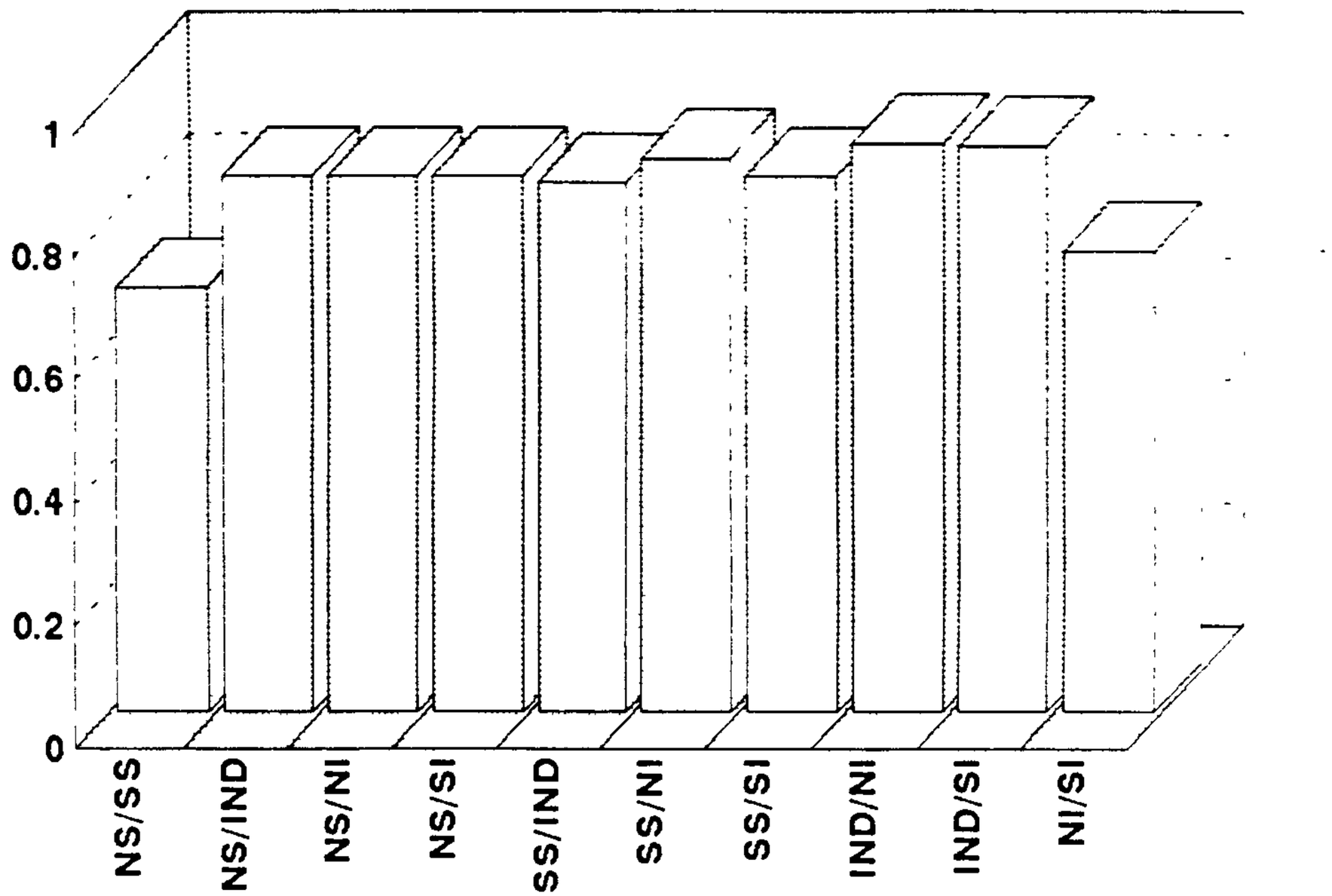
Simple match/dismatch matrix:

SIMILARITY / DISTANCE	NS	SS	IND	NI	SI
NS	1.00 / 0.000	0.240	0.395	0.508	0.562
SS	0.760	1.00 / 0.000	0.412	0.549	0.547
IND	0.605	0.588	1.00 / 0.000	0.494	0.544
NI	0.492	0.451	0.506	1.00 / 0.000	0.145
SI	0.438	0.453	0.456	0.855	1.00 /0.000

NS: North Sri-Lanka, SS: South Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Figs. 5.32

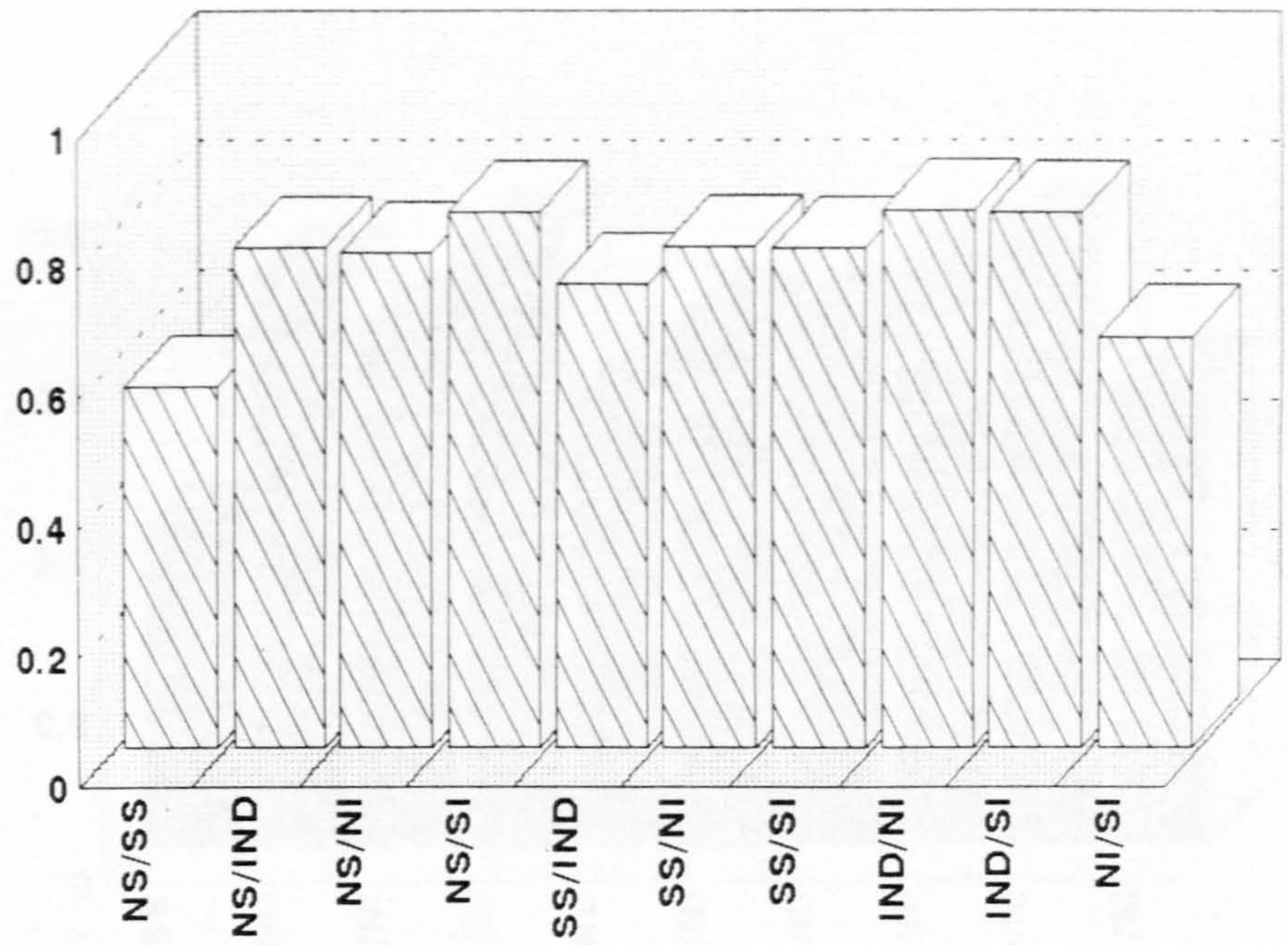
Jaccard (SW) distance in *An. culicifacies* 5 populations



NS: north Sri-Lanka, SS: south Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Fig. 5.33

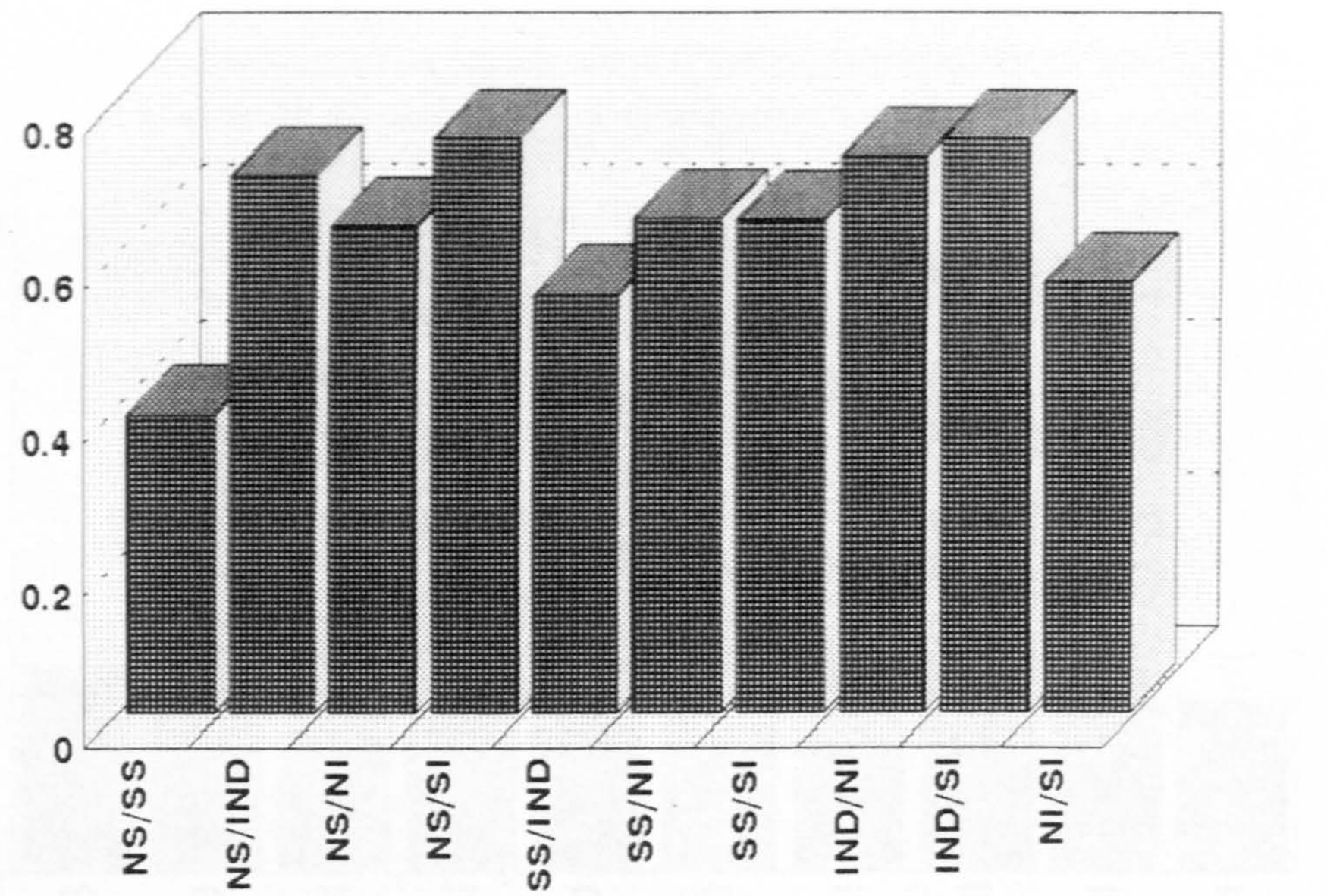
Jaccard (Sj) distance in *An. culicifacies* 5 populations



NS: north Sri-Lanka, SS south Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Fig. 5.34

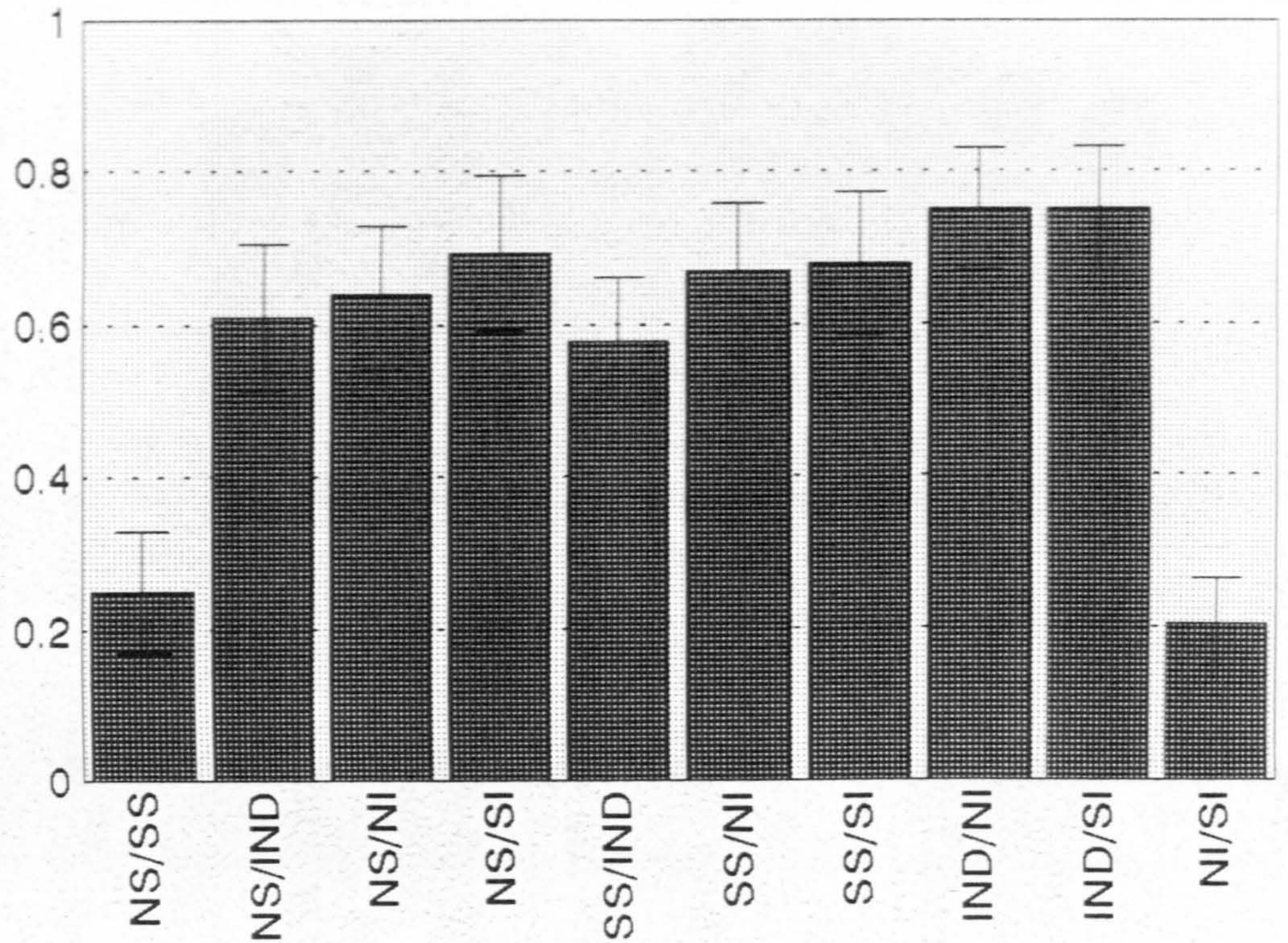
Dice distance in An. culicifacies 5 populations



NS: north Sri-Lanka, SS: south Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Fig. 5.35

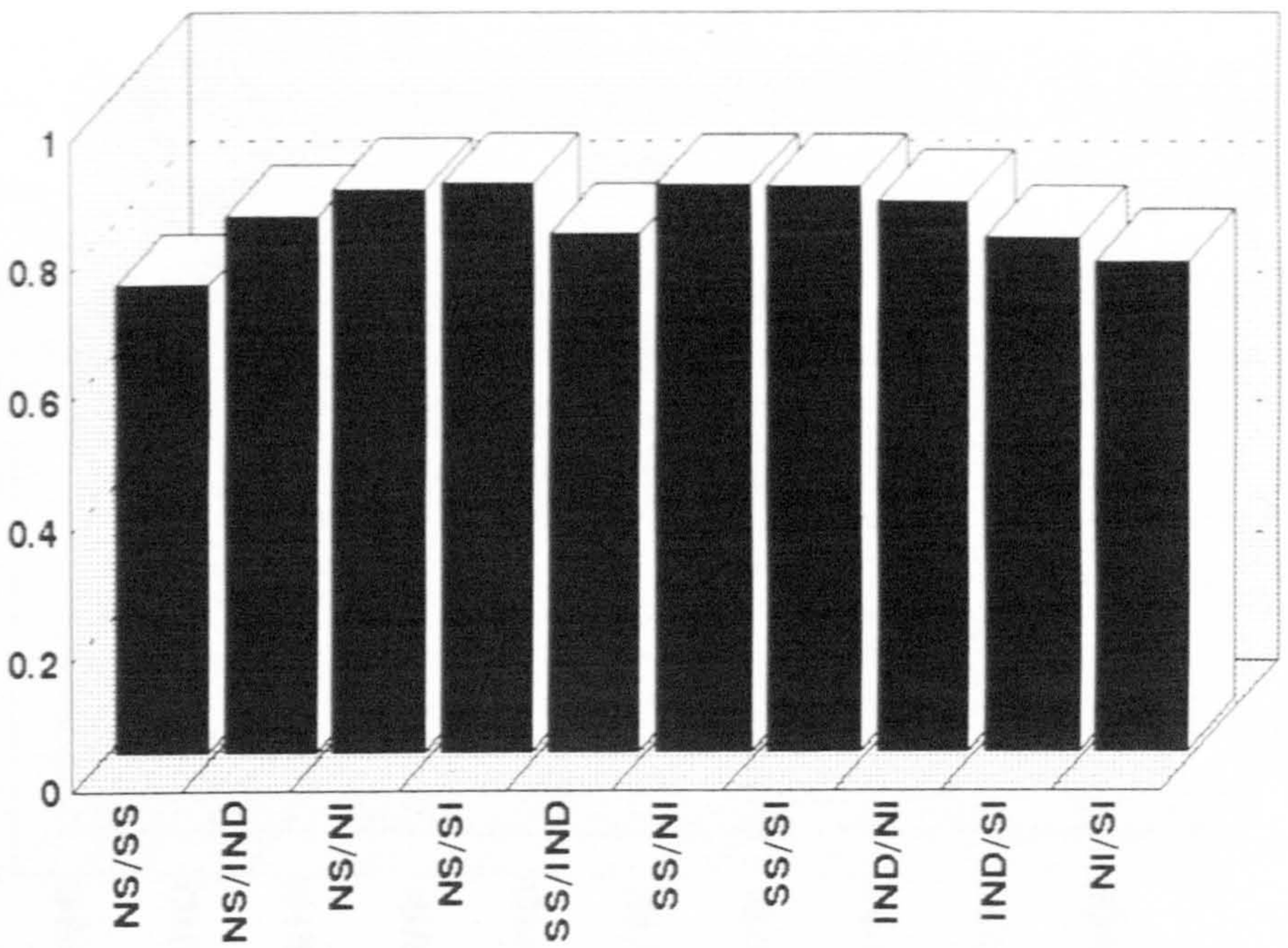
Nei & Li's distance in ***An. culicifacies*** 5 populations



NS: north Sri-Lanka, SS: south Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Fig. 5.36

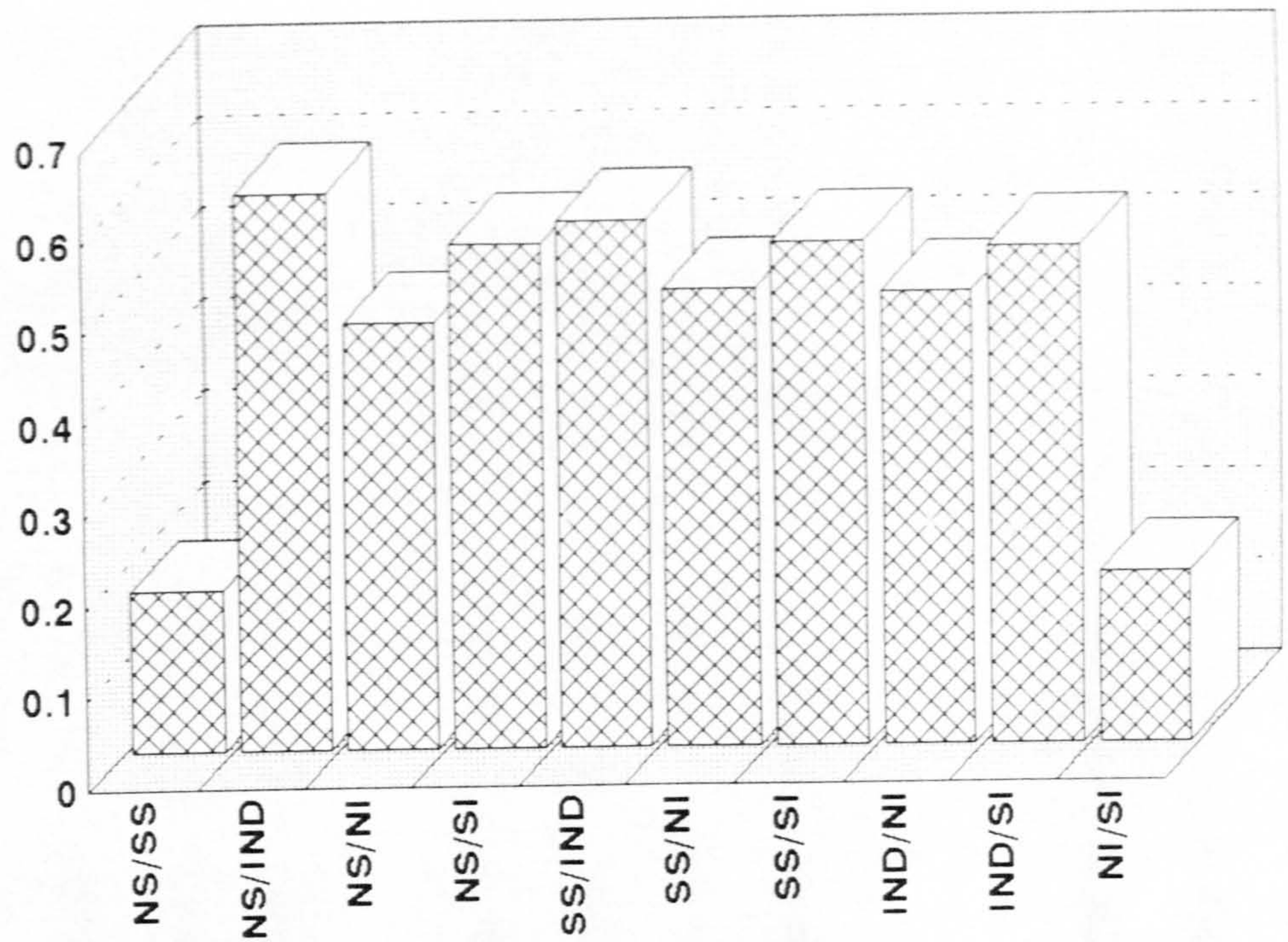
Montpelier distance in *An. culicifacies* 5 populations



NS: north Sri-Lanka, SS: south Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

Fig. 5.37

Dismatche (1-Ssm) in *An. culicifacies* 5 populations



NS: north Sri-Lanka, SS: south Sri-Lanka, IND: India, NI: Iran1, SI: Iran2

2-1) NI: collections from areas closer to the boarder of Afghanistan and Pakistan including Zabol, Zahedan, Khash and Saravan

2-2) SI: Iranshahr, Nikshahr, Chah-Bahar from southern parts of Bluchestan and closer to Kerman province. Variation between these two groups is more than Sri-Lanka north and south populations based on 10 RAPD primers product. However, by increasing the number of primers tested to 20, it showed less variation (Fig. 5.38).

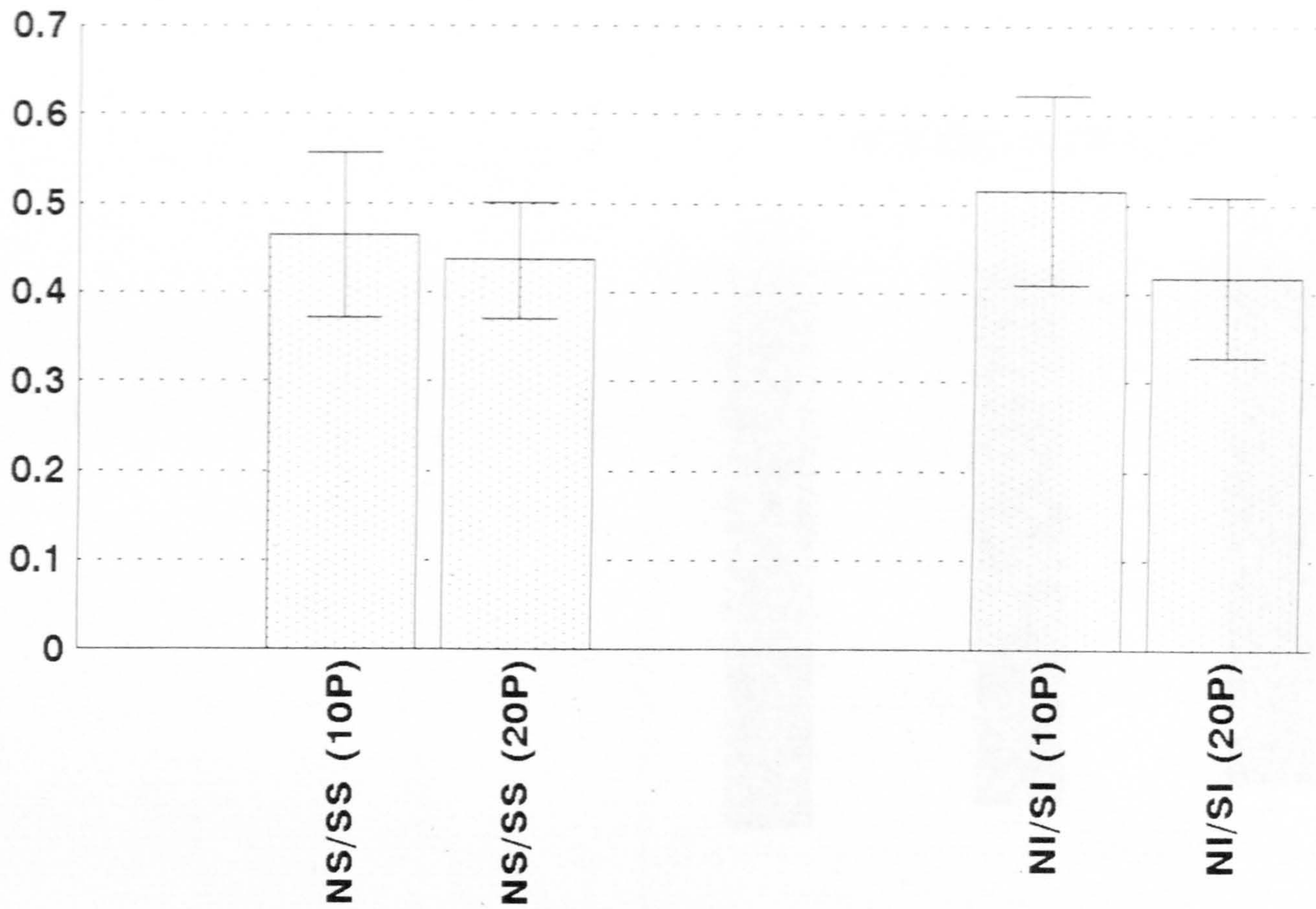
Relatedness of *An. culicifacies* from three countries based on six different distance indices (Fig. 5.39) shows that Sri-Lanka populations distance with Iran is less than Indian species B with Iran populations. Fig. 5.40 shows the most parsimonious tree for 5 population in *An. culicifacies* species A and B.

5.5.3 ITS-2 PCR in a suspected species:

The results of ITS2-PCR confirmed the presence of another group of mosquitoes with unusual pattern which we found in RAPD-PCR. Although the beginning and ending sequences and length of ITS2 in both groups are almost the same but the high degree of nucleotide difference should be considered in any interpretation. Phylogenetic tree based on sequence of these specimens (second peak collection) and *An. culicifacies* from first peak in Nikshahr , collections from other parts of Sistan & Baluchestan province, species A from Pakistan, Species B from Sri-Lanka and also *An. fluviatilis* from Iranshahr (Bluchestan, Iran) revealed that those suspected specimens take a place between other *An. culicifacies* (species A and B) and *An. fluviatilis* (Figs. 5.41A, 5.41B).

Fig. 5.38

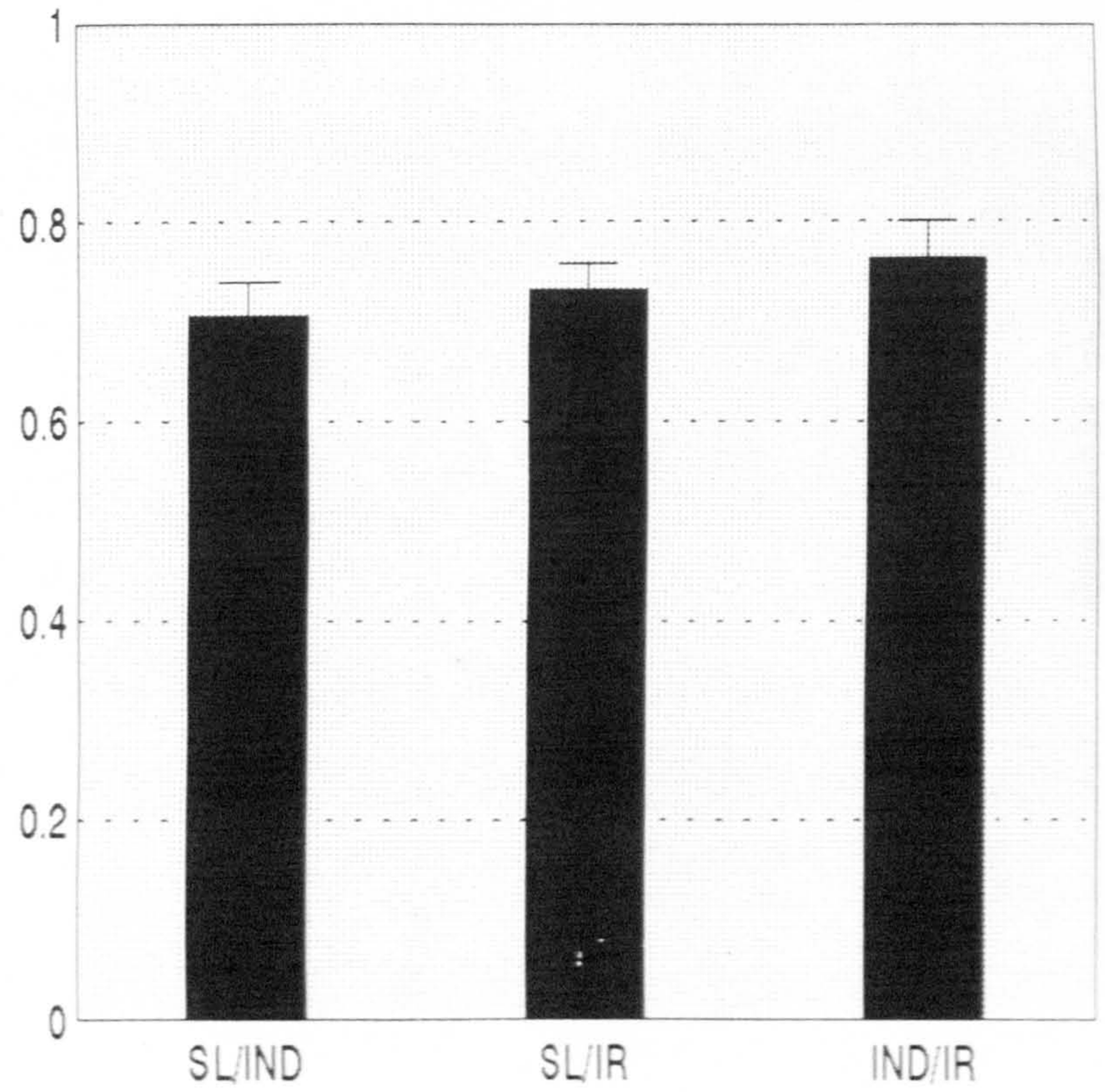
Relatedness within *An. culicifacies* species B (Sri-Lanka north and south) and species A (Iran 1 and Iran 2) populations with 10 and 20 primers' product



NS: north Sri-Lanka, SS: south Sri-Lanka, NI: Iran1, SI: Iran2
10P: 10 primers' product, 20P: 20 primers' product

Fig. 5.39

Relatedness of *An. culicifacies* from three countries based on distance indices



SL: Sri-Lanka, *IND*: India, *IR*: Iran

Fig. 5.40- Comin-Sokal parsimony method found the most parsimonious tree based on RAPD and SSR primers pattern in 5 populations of *An. culicifacies*

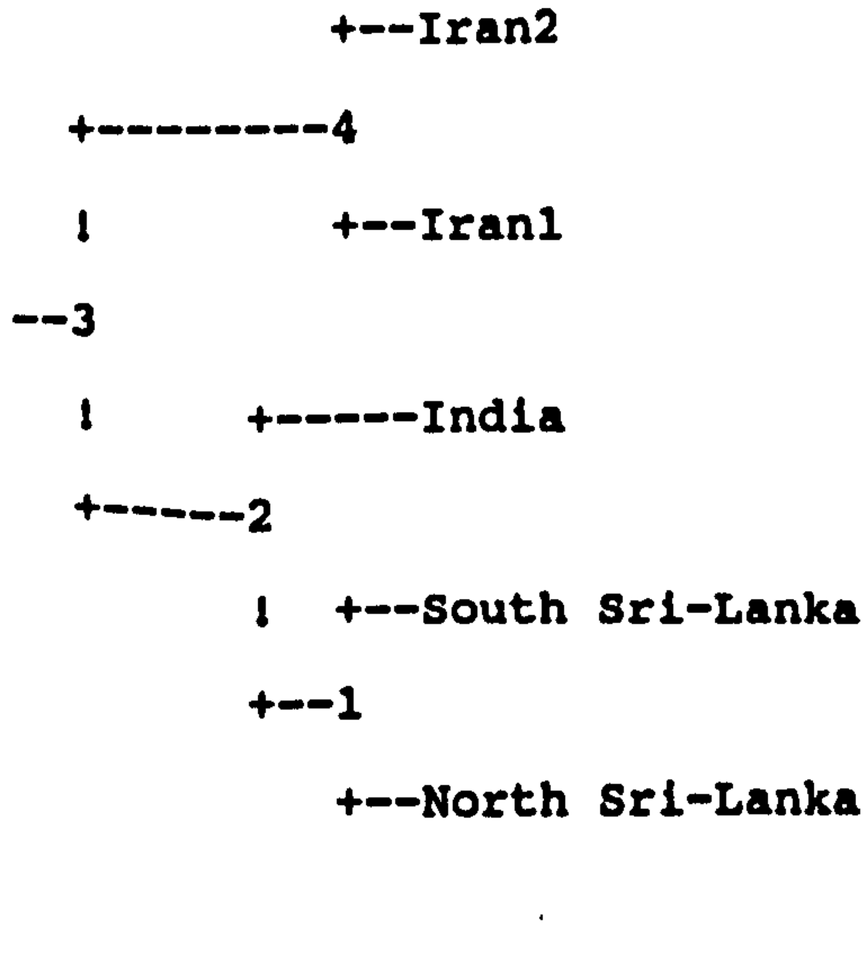
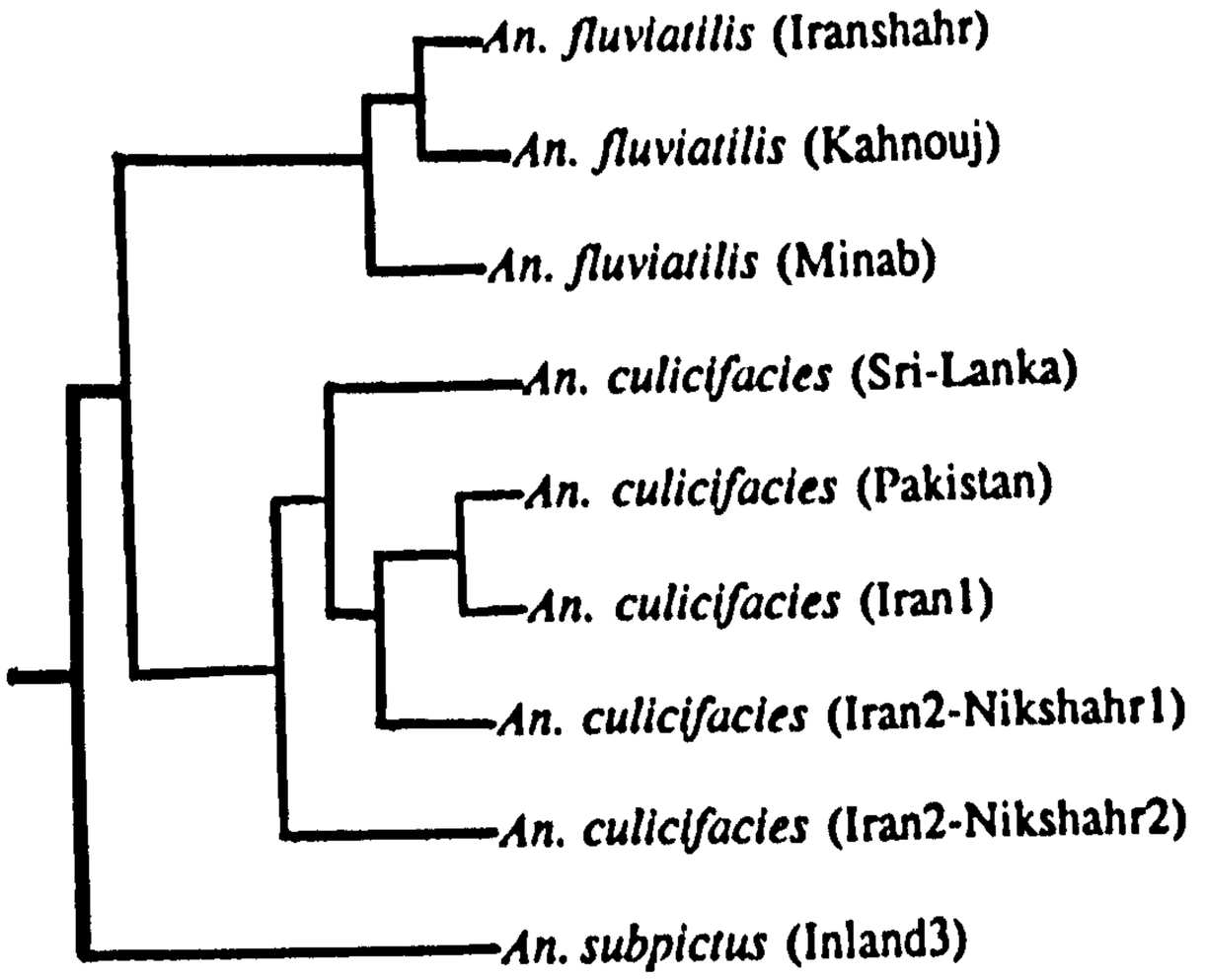
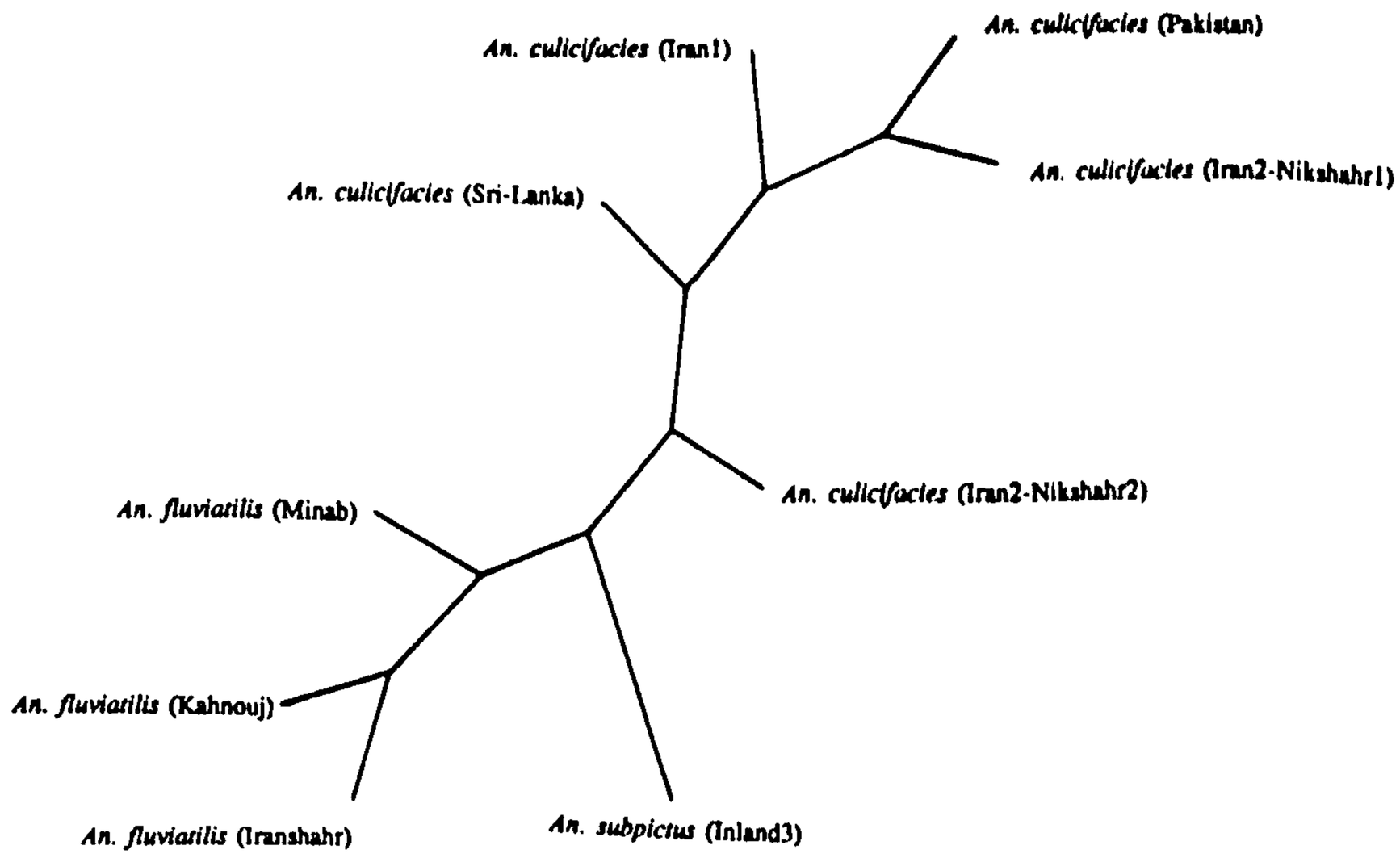


Fig.5.41- Phylogenetic tree based on part of 5.8S, 28S and the whole ITS2 sequence of An. culicifacies species A from Iran and Pakistan and species B from Sri-Lanka by A) DNAPARS and B) DRAWTREE program in PHYLIP. An. fluviatilis and An. subpictus as outgroup.

A) DNAPARS



B) DRAWTREE



5.6. Conclusion

In conclusion, there are several advantages to the use of RAPD-PCR to detect DNA polymorphisms over other molecular-based assays. Only a very small amount of DNA with no prior information on its sequence and a single primer is required for a successful amplification. From DNA extraction to visualising the products on agarose gel, it takes less than 18 hours.

It is clear that PCR-based techniques that examine RAPD and microsatellite loci have considerable potential in looking at the genetic differentiation within and between seemingly isomorphic species of anopheline mosquitoes. However,There are some reports on problems with its reproducibility among laboratories (Penner *et al.* 1993) and artifactual variation in randomly amplified polymorphic DNA banding patterns (Honeycutt *et al.*, 1993; Micheli *et al.*, 1994), but the results of this study with all field materials preserved in different conditions and a large number of individuals examined from each collection site, confirmed its reproducibility and subsequent application in epidemiological studies based on specific and intra-specific variation revealed in both species A and B.

The first step in applying RAPD-PCR is to optimize its conditions and after that by changing those conditions it is obvious getting different or unusual pattern (Figs. 5.19, 5.20) but it is the same with conventional PCR that use gene specific designed primers whereas RAPD use a 10 mer random sequence (the only random - not been chosen based on a known homologous sequence in the target genome) which seems more sensitive to un-optimized conditions.

However, in gene-specific PCR, changing even one degree of annealing temperature, using more or less DNA or low quality DNA, more concentrated primer, $MgCL_2$ and dNTPs, additional number of cycles in amplification, etc., all affect PCR products (wrong size of product; more than one band, sometimes 7-8 band; no amplification and smears).