these regions, which are generally less than 800bp. The IGS region is also an excellent candidate for detecting species-specific differences, but is of much greater size. In *An. quadrimaculatus* A, the IGS region is estimated at approximately 2-4 Kb (Collins *et al.*, 1990). The IGS region has been used for diagnostic identification of the *An. gambiae* complex (Collins *et al.*, 1987, 1988a, 1988b; Paskewitz and Collins, 1990; Scott *et al.*, 1993). Sequencing and amplifying such a large fragment is difficult and it complicates analysis, because cloning and subcloning is needed before sequencing can be initiated. Cloning and amplification of such large regions requires specialized facilities and biological reagents.

The third advantage is that the ITS region has already been used successfully in diagnostic identification of other mosquito species, such as the *An. quadrimaculatus* species complex (Cornel *et al.*, 1996; Rutledge *et al.*, 1996). All species of *Anopheles* studied so far have species-specific sequences present in the non-transcribed ITS regions that can form the basis of a diagnostic assay (Collins and Paskewitz, 1996).

3.1.2 D3 domain

The 28S gene is characterized by having variable lengths between various organisms and can range in size from 2 900 bases in prokaryotes to 5 025 bases in humans (Gonzalez *et al.*, 1985). The size variation occurs by the expansion or contraction of variable joining sequences at 10-12 specific points within the molecule. The amount of variable domains present is dependent on the species. Studies on mammals calculated 12 variable domains and these have been called divergent, D domains or variable regions (Gonzalez *et al.*, 1985). They are characterized by specific features similar to that of the internal transcribed spacer, including a high GC-content, contain short repeated sequences and are able to form self-contained double-helical structures (Gonzalez *et al.*, 1985).

The variable domains are located in the same positions in all eukaryotes. The sizes of variable domains may vary between species, for example in *Escherichia coli,* the D3 variable domain extends from position 602-655 of the 23S rRNA (53bp) (equivalent to the 28S in eukaryotes) (Michot *et al.*, 1990), while the same region in *Schistosoma* (Class: Trematoda) species is 200bp long (Littlewood and Johnston, 1995). The D3 domain is located at the 5' end of the 28S rRNA gene in *Schistosoma.* Comparison of the D3 DNA sequences between four species of *Schistosoma* species, showed that within the D3 region there were conserved sequences, as well as species-specific sequences. The degree of variation seems to vary within the D3 domain, since variation at position 901-920 showed a 5% to 25% sequence difference when compared among the four species of *Schistosoma*, while position 1070-1090 bp showed a 3% to 60% variation. The 3' end of the D3 domain shows 15% difference in one species while the sequences in the other three species are homologous. Position 930-1 000 bp showed no sequence variation between any of the four species.

These variable domains are well studied in bacteria, *Schistosoma* species and humans and are used to resolve the phylogenetic relationships between kingdoms, taxa and species. Although the 28S rRNA gene is showing a faster evolution than the 18S rRNA gene, it is not as fast as the two ITS regions (Brower and DeSalle, 1994). Littlewood (1994) used the variable domains (D1-D3) in a phylogenetic study of cupped oysters to determine the systematic relationship between the oyster taxa, which complement the deficiency of morphological characters so far recognized.

3.1.3 Cytochrome oxidase II

The cytochrome oxidase II (COII) gene on the mitochondrial DNA is also often used in phylogenetic studies of insects (Liu and Beckenbach, 1992) and the COII gene product plays a part in the e^- transport/ oxidative phosphorylation in living organisms.

The cytochrome oxidase II (COII) gene codes for the second subunit of cytochrome oxidase enzyme forming part of the cytochrome oxidase complex. The cytochrome oxidase complex accepts electrons from cytochrome c and transfers these electrons to oxygen, playing a role in the complex process of ATP synthesis. The cytochromes were discovered in 1925 as compounds that undergo rapid oxidation and reduction in living organisms as bacteria, yeast and insects (described in Alberts *et al.*, 1994.).

This region is often used in phylogenetic studies of closely related species. It is also well established that the mitochondrial genome undergoes an evolutionary process of its own (Hartl, 1991). Early studies of mammals showed that the average rate of evolution in the mitochondrial genome is five to ten times higher than that of the nuclear genome (Simon *et al.*, 1994). Powell *et al.* (1986) showed that the nucleotide substitution rate in *Drosophila* mtDNA is similar to the low copy nuclear DNA.

Mitochondrial DNA (mtDNA) like the rDNA occurs in multiple copies per cell and is relatively easy to manipulate. Mitochondrial DNA is an especially good candidate for resolving relationships among close relatives, because it does not recombine and because it usually occurs in only one haplotype per individual (Brower and DeSalle, 1994).

Sequencing of the COII region in members of the *An. gambiae* group showed that this region is 684 bp long and has a relatively low homology (88.3%) to *An. quadrimaculatus* and a lower homology (66.5%) to *Ae. mellifera* (Beard *et al.*, 1993b).

The COII region has never been used in the development of species-specific primers, but has been used to detect species differences in *Simulium* species by using hetroduplex analysis (a technique to detect mutations). This region has never been sequenced in any members of the *An. funestus* group and may be useful in the development of a diagnostic assay for the *An. funestus* group. It may also be useful in future studies of phylogenetic relationships within this species group.

3.1.4 Primer Design

The average rate of detectable mutation per gene is 10^{-6} (Freifelder, 1986), but varies depending on the region of DNA. If these defects are left uncorrected, spontaneous DNA damage would rapidly change the DNA sequence. In the course of hundreds of millions of years there have evolved efficient systems for correcting the occasional mistakes that occur during replication. However, if repair was perfect, mutations would never persist and species would no longer evolve. Species would be incapable of either speciation or adaption and therefore species-specific DNA variation is inevitable in the course of evolution.

DNA is subject to damage in many ways (Freifelder, 1986) and can be altered by either base substitutions occurring during replication, base changes due to chemical instability of the bases or the alteration due to action of other chemical and environmental agents (Freifelder, 1986). These alterations will lead to defects such as correct hydrogen bonds not forming, base

mispairs (A pairs to G or a C to a T), altered bases, single-strand or double-strand breaks and even cross linking. These species-specific DNA differences, can be used to distinguish species from one another using a PCR assay.

To develop a species-specific PCR assay, primers need to be designed to distinguish between the various species and it is fundamental that species-specific sequence variation is present. Primers can then be designed around or on either side of these regions, giving products of different sizes after amplification that allow for easy analysis by electrophoresis. The selected primer sequences will determine the size and location of the PCR product and will define the *Tm* of the region of hybridization/ annealing (Dieffenbach and Dveksler, 1995). The *Tm* temperature is the dissociation temperature of the primer/template duplex (Dieffenbach and Dveksler, 1995). *Tm* is a physical parameter shown to be important in product yields after amplification. Well designed primers can help avoid the generation of background and non-specific products and the aim of primer design is to obtain a balance between the specificity of amplification and the efficiency of amplification.

Designing primers needs the consideration of various parameters to obtain the best possible primer pair (Dieffenbach and Dveksler, 1995). The first consideration is the primer length. Oligonucleotides between 18-24 bases tend to give sequence specific amplification if the annealing temperature of the PCR is set within a few degrees of the primer *Tm*. Secondly the terminal nucleotide in the PCR primer (3' -terminal position) is essential for controlling mis-priming. Primers complementary to each other, particularly at their 3' ends will lead to the undesirable primer-dimer phenomenon. This phenomenon is the result of the amplification of the hybridized primer pairs themselves, which results in competition between primer-dimer

products and the native template. The third consideration to keep in mind is the GC content which determines the *Tm* value. Oligonucleotides 20 bases long with a 50% G+C content generally have *Tm* values between 56-62°C which is ideal.

3.2 OBJECTIVE

The aim of this part of the project was, mainly to develop a species-specific PCR assay to identify the common species of the *An. funestus* group present in South Africa; *An. vaneedeni*, *An. rivulorum*, *An. leesoni* and the malaria vector, *An. funestus*. In order for the PCR assay to be of practical use in malaria control surveillance activities, it has to be as inexpensive as possible, meaning the use of a multiplex PCR assay. A multiplex PCR assay uses various primers specific for each species, combined in one reaction mixture. A universal primer is used to function as a forward or reverse primer and recognise sequence homology in all species present. The secondary primers are species-specific and act as the reverse or forward primer respectively. The multiplex PCR assay eliminates the need to perform individual PCR reactions with only two specific primer pairs and therefore is more economical. In this work the ITS1 and D3 regions on the rDNA and the COII region on mtDNA were sequenced in the hope of finding species-specific variation for the development of a PCR assay to identify members in the *An. funestus* group.

3.3 MATERIALS AND METHODS

DNA extraction, amplification and sequencing

DNA was extracted from single mosquitoes or legs, using the standard procedure described in Collins *et al.* (1987). Individual mosquitoes or single legs were homogenized in 200µl buffer (0.08M $NaCl_2$, 0.16M sucrose, 0.06M EDTA, 0.5% SDS, 0.1M Tris-Cl pH8.6). After

incubation at 70°C for 30 min, 8µl potassium acetate (8M) was added and the samples were incubated on ice for at least 30 minutes to aid the precipitation of proteins.

Centrifugation at 16 000 rpm precipitated proteins and cell debris. Supernatant was removed and added to two volumes of 99% ethanol. DNA was precipitated by centrifugation for 5 min at 16 000rpm for 30-35 min. Salts were removed by washing the DNA pellet with one volume 70% ethanol. Ethanol was decanted and the DNA was vacuum or air dried. The DNA pellet was resuspended in 200µl 1X TE buffer. $1/20^{th}$ of the purified DNA was loaded on a 1% agarose gel to confirm the presence of DNA. DNA aliquots were stored at -20°C and at -70°C for future use. For each set of DNA extractions, a negative control was included. The negative control contained all the reagents used in extraction procedure excluding the mosquito DNA or mosquito leg. Pestles used in homogenizing the mosquito samples and in negative controls were sterilised under a UV (ultra-violet) light for approximately 30 minutes to one hour or by placing the pestles in hypo-chloric acid for 30 min followed by rinsing in distilled water. Sterilised pestles were stored in a sterile container with a screw-on lid.

Amplification

The extracted DNA was used in a 25µl PCR reaction. The PCR mix (PCR cocktail) contained the following: 2.5 µl of the 10 x reaction buffer (500 mM KCl, 100 mM Tris-Cl pH 8.3), $MgCl_2$ at a concentration specific for each primer pair, 1 µM of each primer, 200 µM of each dNTP, 2 U *Taq* DNA polymerase, overlaid by 25 µl mineral oil. Prior to amplification 0.5µl-0.75µl DNA template was added through the mineral oil. Amplification conditions were as follow: denaturation at 94°C for 30 sec, annealing (temperatures were dependant on the primers, Table 3.1) for 30 sec and extension at 72°C for 30 sec, for 30 cycles followed by a

final extension at 72°C for 10 min.

Optimisation of PCR

Optimisation of PCR was done for each set of primers as well as for each species. PCR optimisation was initially performed using a low annealing temperature of 35°C and subsequently the annealing temperatures were increased with 5°C intervals. The PCR products were loaded on a 1% agarose gel after amplification and the presence of a single specific band indicated that the PCR conditions for that reaction were optimal with respect to the annealing temperature. Different magnesium chloride ($MgCl_2$) concentrations between 0.5mM and 3mM, were tested at the same annealing temperature to optimise yield. Samples were loaded by mixing 8µl PCR product with 4µl ficoll loading dye (50% sucrose, 0.05M EDTA pH7, 0.1% bromo-phenol blue and 10% ficoll). Electrophoresis was performed at a constant voltage of 100V in 1X TAE buffer. The gel was stained using 0.03mg/100ml Ethidium bromide. Size markers (1Kb- ladder), were loaded on the agarose gel to calculate the relative size of the PCR product by using a regression line formula: $\hat{y}=bx+a$ where $b=\Sigma xy-n\bar{y}\bar{x}/\Sigma x^2-n(\bar{x}^2)$ and $a=\bar{y}-bx$. y = Log Kb (length of DNA fragment) and x = the distance migrated (cm) from the well, and n= sample size (the amount of DNA fragments compiling the size marker and used in calculation) (Sambrook, *et al.*, 1989). The remaining PCR product was used for sequencing reactions once amplification proved successful.

Sequencing

Approximately 100 ng/µl DNA was used per sequencing reaction. Sequencing was done using a Sequenase PCR Product Sequencing kit (Amersham Life Science). PCR products were enzymatically treated using Exonuclease I (10.0 units/µl) and Shrimp Alkaline (2.0 units/µl)

Phosphatase included in the kit.

Table 3.1 Primer sequences used in this study for amplification and sequencing. Optimal annealing temperatures for amplification are indicated.

Name of the primer used in PCR reaction	Primer Sequence	Annealing temperatures	Reference
ITS1 (28S)(*R)	ATG CTT AAA TTT AGG GGG TA	66°C	Porter and Collins, 1991
(5.8S)(*F)	TGT GAA CTG CAG GAC ACA TG		
COII: LEU (*F)	TCT AAT ATG GCA GAT TAG TGC A	59°C	Rosie Sharpe (pers comm 1997)
LYS (*R)	ACT TGC TTT CAG TCA TCT AAT G		
28S D3A (*F)	GAC CCG TCT TGA AAC ACG GA	62°C with 3mMMgCl (*funestus* and *vaneedeni*)	Rosie Sharpe (pers comm 1997)
D3B (*R)	TCG GAA GGA ACC AGC TAC TA	40°C and 1.5mM MgCl *An. rivulorum*	

(F)= forward primer used in sequencing, (R)= reverse primer used in sequencing, both forward and reverse primers were used in all sequencing reactions.

Exonuclease I removes residual single-stranded primers and any extraneous ssDNA (single-strand DNA) produced by PCR. Shrimp Alkaline Phosphatase removes the remaining dNTPs from the PCR mixture which could interfere with the labelling step in the sequencing process. Heating the samples to 80°C for 15 minutes inactivated both enzymes. Annealing step was conducted by using 5µl (10ng/µl) treated PCR product and 5pmol/µl of one specific primer. Treated PCR products were heated to 94°C to denature the DNA and were placed on ice after 5 minutes. Cooling on ice assisted in the annealing of primers to the DNA template. Primers

used in sequencing reaction: the forward primer (F) labelled with primer name see (Table 3.1) was used to read the positive (sense) strand and the reverse or (R) primer was used to sequence the negative (anti-sense) strand.

The labelling reaction was done according to manufacturers' suggestions using [$\propto^{35}$S]dATP (5μCi). The labelling step used all four dNTPs, one of which was ∝- labelled. Sequenase, version 2.0 enzyme used in labelling reaction possessed a lower exonuclease activity, higher processivity, higher purity, higher specific activity and better stability than other manufactured sequenase enzymes.

Termination of chain elongation is done by adding termination mix (dideoxy-nucleotides). The termination mix contained a nucleotide analog that will not support continued DNA elongation. The chain-terminating nucleotide analogs are the 2',3'-dideoxynucleotide 5'-triphosphates (ddNTPs), which lack the 3'OH group necessary for DNA chain elongation. Four separate reactions, each with a different ddNTP give complete sequence information. The radioactivity labels the chains of various lengths and can be visualized by autoradiography. During the termination step the chains are extended on an average of an additional 50-100 nucleotides. The reactions are terminated by the addition of ethylenediamine tetraacetic acid (EDTA) and formamide. The substitution of a nucleotide analog for dGTP (dITP or 7-deaza-dGTP) which forms weaker secondary structures has been successful in eliminating gel artifacts, eg. compressing of bands or separation between bands further than normal and it destabilizes the DNA secondary structure during synthesis as well as electrophoresis. 7-deaza-dGTP was used in this study where gel artifacts were visualised on autoradiograph (Mizusawa *et al.*, 1986).

Electrophoresis was done on a 6% acrylamide:bisacrylamide (19:1)/ 7M urea gel. Gels were pre-run for 1 hour to ensure a uniformly heated matrix before samples were loaded. Gels were run at a constant temperature of 50-55°C with the aid of a fan. Electrophoresis was done at a constant voltage of 2000V for 5-7hours. Using volts less then 2000V resulted in loss of resolution. A short electrophoresis time (3-5hours), gave sequencing information closer to the primer ± 200 base pairs, while long electrophoresis time (5-7 hours) gave sequence information further away from the primer. After electrophoresis, the gel was transferred onto filter paper, covered with plastic wrap, vacuum dried for 1-2 hours and exposed overnight with X-Ray film. DNA concentration in initial sequencing reaction was decreased if the bands close to the primer sequence were faint. In order to visualise sequences close to the primer, the labelling mix (5 x concentration with 7.5μM GTP, CTP, TTP) was diluted 1:20 and additional manganese buffer (0.15M sodium isocitrate, 0.1M $MnCl_2$) was added. The addition of Mn^{2+} reduces the average length of DNA synthesised in the termination step. This helps in intensifying bands corresponding to sequences close to the primer.

Sequence results were aligned using MegAlign software ((DNASTAR Inc. Madison, Wisconsin) and the sequences were compared to other sequences in GenBank (BLAST search: http://www2.igh.cnrs.fr/bin/blast-guess.cgi) to confirm the identity of the sequenced region. Appendix II provides the alignment results for all the regions sequenced. Sequencing data of the D3 domain were submitted to GenBank under the following accession numbers: AF007094 *An. funestus*, AF007095 *An. vaneedeni*, AF036036 *An. rivulorum*. Table 3.2 indicates all the species in the *An. funestus* group used for this study as well as the localities from which they originated.

Table 3.2 Species of the *An. funestus* group sequenced in this study

Species	Locality	No of families	No. of individuals
An. funestus	Madagascar	1	4
An. funestus	Bagamoya	1	2
An. funestus	Matimbwa	1	1
An. rivulorum	Komatipoort	3	5
An. vaneedeni	Komatipoort	3	5
An. vaneedeni	Thomo village	1	1
An. rivulorum	Komatipoort	1	2
An. leesoni	Thomo village	1	1

3.4 RESULTS

3.4.1 The ITS1 region

Amplification of the ITS1 region produced amplicons of an estimated size of 300 and 700bp (Figure 3.2). Two PCR products were consistently obtained under all conditions tested including different annealing temperatures (62°C to 68°C), $MgCl_2$ concentrations (0.5mM-4mM) and the addition of 10% DMSO. At 68°C no amplification was visible, while at 67°C annealing temperature the yield was considerably less and the secondary bands were already visible. The 600bp and 300bp bands were individually excised from the agarose gel, incubated in a heating block (70°C) with 1X TE buffer to aid in the elution of the DNA, and reamplified using the same primers. Reamplified products were electrophoresed and a smear was obtained on the gel. Further dilutions (1:100-1:1 000 000) of excised PCR product were made, prior to PCR, to eliminate the smear, but a single fragment was never produced. The respective

amplicons could have been cloned into plasmid vectors after they were excised from agarose gel, but this avenue of research was never investigated as other regions were available for sequencing.

3.4.2 The Cytochrome oxidase II region

The third region sequenced was COII on the mitochondrial genome, in two species, *An. funestus* from Madagascar and *An. vaneedeni* from Komatipoort. Comparison of the obtained sequence with the sequences deposited in the GenBank database, confirmed that the correct region was amplified using the primer pair Leu and Lys (Table 3.1) (Appendix II). The complete sequence of the mitochondrial genome of *An. gambiae* was published in 1993 by Beard *et al.* (1993b) and alignment between *An. funestus* and *An. gambiae* COII sequences, was difficult in certain areas. Figure 3.3 shows the alignment result between *An. funestus* and *An. vaneedeni* COII sequences, demonstrating the presence of only minor mutation differences. The estimated size of the amplified COII region was approximately 700bp (Figure 3.4). Inter-specific DNA sequence variation in the COII region between *An. funestus* and *An. vaneedeni* was only 2.7% (nucleotide differences/ total nucleotides). The calculated species-specific variation was not sufficient for the development of species-specific primers.

Although it is possible to design a primer pair with as little as one base pair difference at the 3' end of the primer, it is not preferred when the diagnostic PCR will contain multiple primer pairs. In these circumstances it is favoured to have five or more base pair differences between the different primer pairs in question (A.J. Cornel pers. comm). Using multiple primers means that the annealing temperature used are not always optimal for each primer pair, but would be the lowest *Tm* of the primer pairs in the cocktail. Thus with too few nucleotide differences

between the species-specific primers, the chance of non-specific annealing is increased, with consequent misidentification.

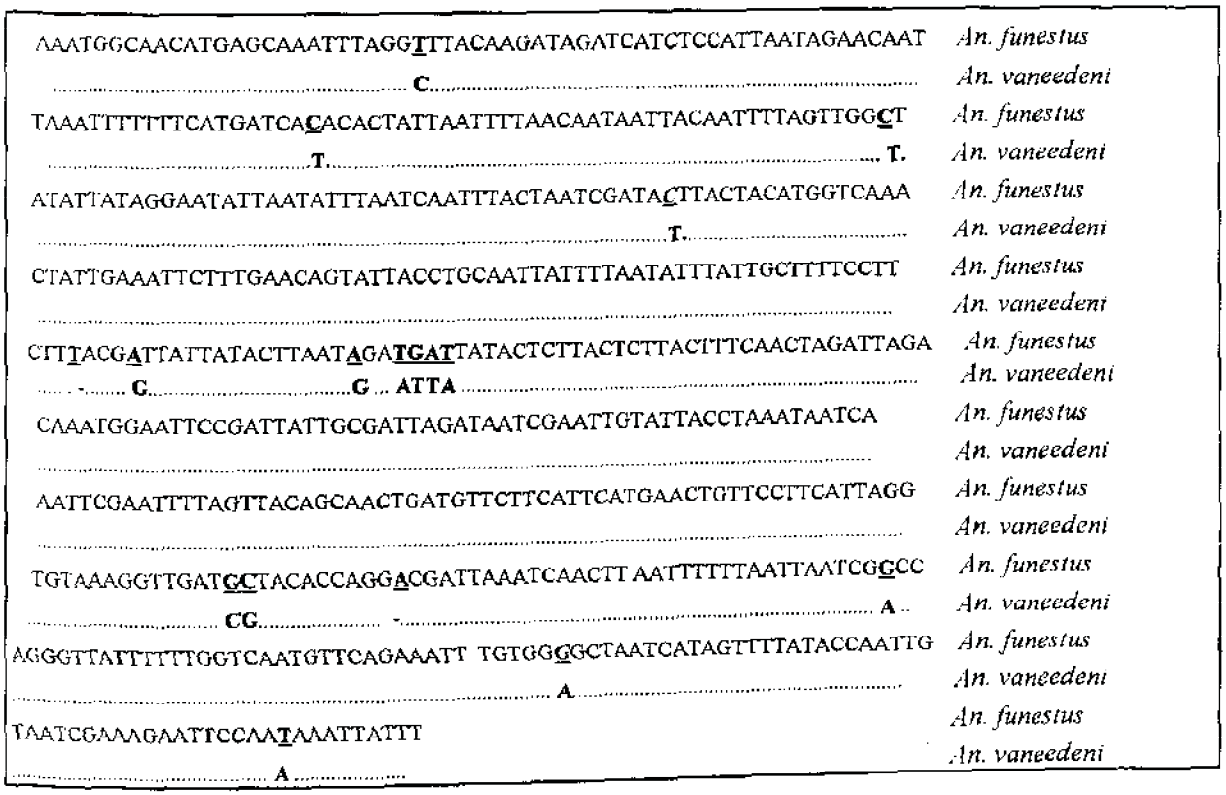

Figure 3.3 Partial COII sequence alignments between *An. funestus* and *An. vaneedeni*. Exact sequence matches are indicated by (.), deletions by (-) and sequence differences are indicated with the appropriate bold nucleotide.

3.4.3 The D3 region

The D3 domain was approximately 380bp in size (Figure 3.5). Alignment between the *An. vaneedeni, An. funestus, An. rivulorum, An. leesoni* and *An. gambiae* D3 domains is shown in Figure 3.6.

Both DNA strands can easily be sequenced using the forward primer (D3A), which gives

information on the sense strand and D3B as the reverse primer, which gives sequence information for the anti-sense strand. Sequencing both DNA strands increases the reliability of the sequence information gained. The PCR conditions were standardised for each species in order to have the best possible yield after amplification. *An. funestus* and *An. vaneedeni* DNA was amplified using conditions as follow: 3mM $MgCl_2$ with 62°C annealing temperature, while *An. rivulorum* DNA was amplified using 1.5mM $MgCl_2$ and an annealing temperature of 40°C.

Design of primers was done manually and results of possible primers are indicated in Table 3.3. It was not possible to design primers so that the amplified products could be separated on a agarose gel. Separation needs at least 30bp size difference between fragments, and even then it can be difficult to distinguish products if the gel had a low agarose concentration or if the time of electrophoresis was too short.

Table 3.3 Primer pairs to identify species in a PCR assay, D3A is used as a universal forward primer for all species in question

Species	Reverse primer position (according to positions in Figure 3.7)	Approximate length of amplified fragment using specific primer pair
An. funestus	114-134bp	123bp
An. vaneedeni	113-135bp	123bp
An. rivulorum	264-283bp	269bp
An. leesoni	58 -78bp	78bp

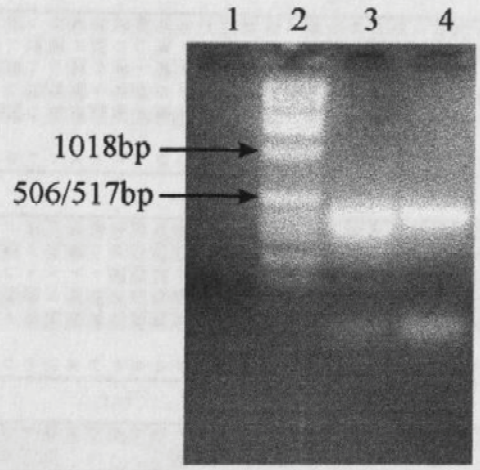

Figure 3.5 D3 amplified fragment for *An. funestus* (Lane 3) and *An. vaneedeni* (Lane 4). Lane 2: 1-Kb size marker. Lane 1: Negative control.

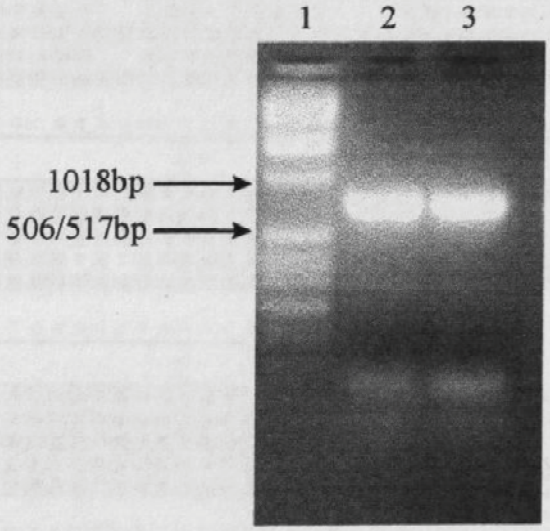

Figure 3.4 Amplified COII fragment for *An. funestus* (Lane2 and 3). Lane 1: 1-Kb size marker.

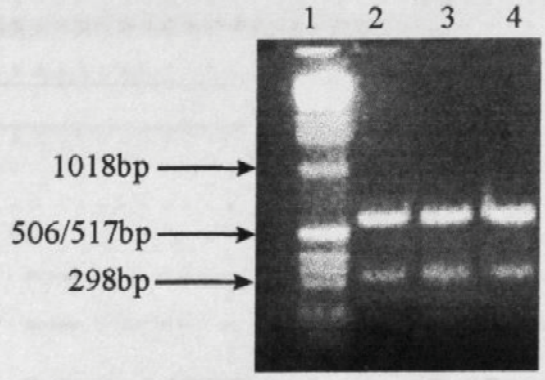

Figure 3.2 PCR products after amplification with the ITS1 primers. Lane1: 1-Kb size marker, 2,4: *An. funestus*, 3: *An. vaneedeni*

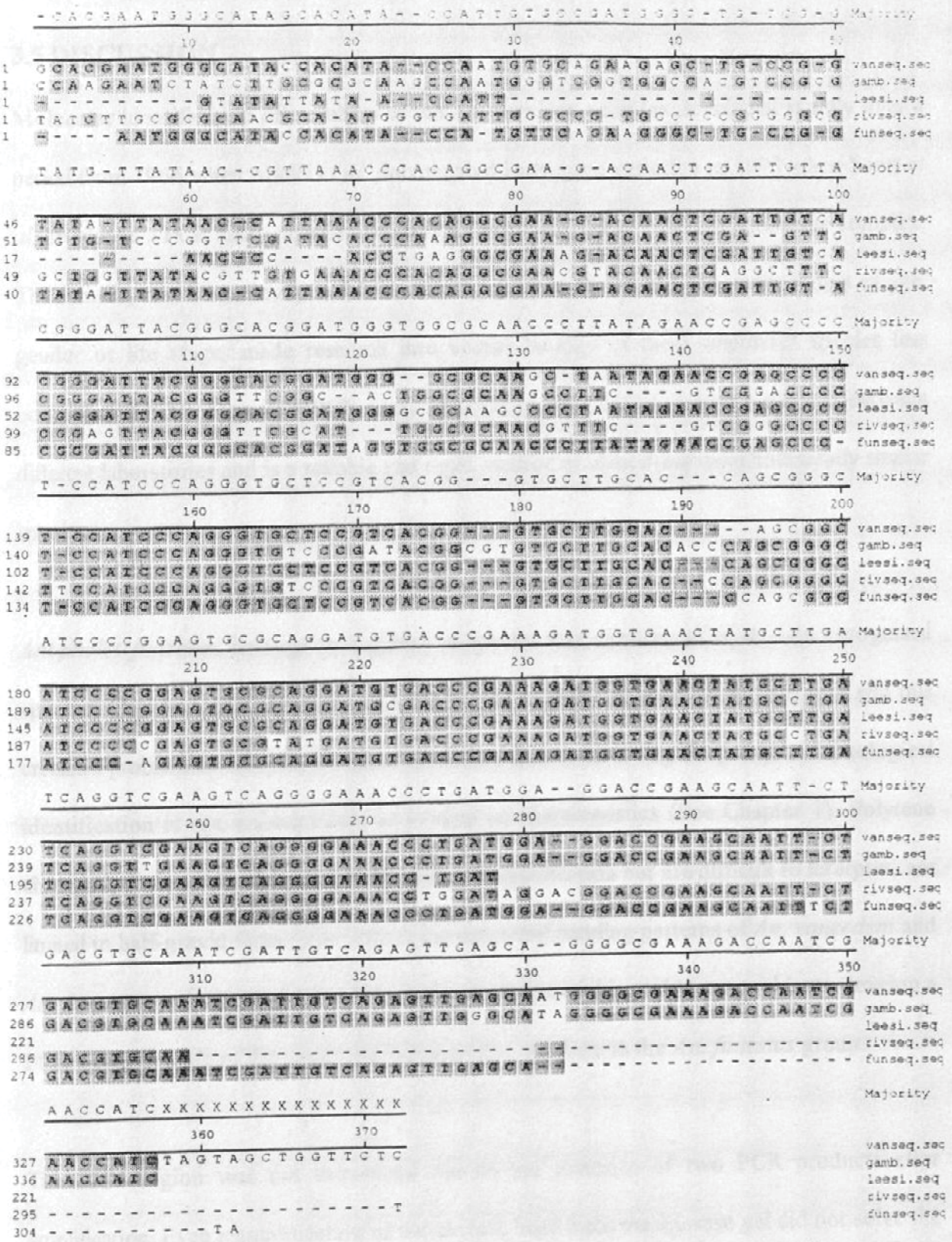

Figure 3.6 Sequence alignment of D3 fragment amplified from *An. funestus*, *An. vaneedeni*, *An. rivulorum*, *An. leesoni* and *An. gambiae*.

Figure 3.6 Sequence alignment of D3 fragment amplified from *An. funestus*, *An. vaneedeni*, *An. rivulorum*, *An. leesoni* and *An. gambiae*.

3.5 DISCUSSION

Molecular identification of members of the *An. gambiae* complex developed rapidly over a period from 1966 when Micks *et al.* documented isoenzyme variation to 1993 when Scott *et al.* developed a PCR assay capable of distinguishing five species of the *An. gambiae* complex. The actuality that it was now possible to identify members in the complex independent of age, gender or life stage, made research into vector biology of these important species less complicated. The diagnostic PCR assay of Scott *et al.* (1993) has been evaluated between different laboratories and is a reliable and rapid method of identifying morphologically similar species.

Anopheles funestus, belongs to a group of species that contains at least eight recognised species. Currently there is no accurate and rapid method for identifying these taxa and this creates a problem for surveillance and transmission studies involving the group. Morphological identification is not accurate due to overlap of characteristics (See Chapter 1). Polytene chromosomes provide an alternative means of identification but are difficult to interpret, are limited to half-gravid females and the homosequential banding patterns of *An. vaneedeni* and *An. funestus* presents a problem. The obvious solution to this problem would be to develop a PCR assay capable of distinguishing between the members in the *An. funestus* group.

The ITS1 region was not sequenced due to the presence of two PCR products after amplification. Even reamplification of the excised band from the agarose gel did not solve the problem of multiple bands. The only solution to this problem would be to clone each fragment in order to sequence each fragment individually to determine which is the ITS1 amplified product.

Once this has been established it would be necessary to excise the appropriate band after electrophoresis, and clone it each time before sequencing can be done. Sequencing another region proved to be more successful.

The COII region sequenced in *An. funestus* was AT rich (28% GC-content) and the amount of DNA variation found between *An. funestus* and *An. vaneedeni* was extremely low making this region less suitable for species primer development. The high degree of homology between *An. funestus* and *An. vaneedeni* can again be explained by the fact that they are newly evolved species and share a great deal of ancestral mtDNA sequences. This is supported by research conducted by Green (1982) where chromosome inversions were used to show that *An. funestus* and *An. vaneedeni* are sister species in the *An. funestus* group.

Beard *et al.* (1993b) found relatively low frequency change (frequency of change is the number of changes per total number of bp in gene) when comparing the COII region of *An. gambiae* and *An. quadrimaculatus*. In *An. gambiae* the COII region is 684bp long and shows only a 0.115 (11.5%) frequency of change when compared with *An. quadrimaculatus*. Comparing *An. gambiae* with *Drosophila yakuba*, a frequency change of 0.199 (19.9%) was recorded (Beard *et al.*, 1993b). Amino acid length of protein coding domains is the same between *An. gambiae*, *An. quadrimaculatus* and *D. yakuba*, but differs from *Apis mellifera* by only three amino acids. This high degree of homology between unrelated species can be explained by the fact that the region is part of a gene essential to normal cell regulation and mutations causing protein alteration might be lethal.

The D3 region showed 9% DNA differences between *An. funestus* and *An. gambiae*, while a

4.5% DNA variation was observed between *An. funestus* and *An. vaneedeni* (Figure 3.6). Other species of the *An. funestus* group were not available during the initial stages of this project, but the information gained from sequencing *An. funestus* and *An. vaneedeni* D3 regions was used for the development of an intermediate assay. A restriction fragment length polymorphism (RFLP) assay is described in detail in Chapter 4. This assay was initially used to identify *An. funestus* and *An. vaneedeni* until a better assay was developed to include other members of the *An. funestus* group.

The sequenced of *An. vaneedeni* D3 region showed intra-specific variation of an AG at position 194-195 where the AG was absent from some isolates. This meant that it was not possible to design a primer in that area. The 5' end of the region 10-145bp, showed the most variation, while the 3' end was more conserved (Figure 3.6).

Sequencing of the other species in the *An. funestus* group was performed once they were collected and identified using morphology. Table 3.3 gives possible primer pairs that would result in species-specific fragments after amplification with their relative fragment sizes amplified. These primers can however not be used together in a PCR cocktail, because of the nearly identical sizes of the expected PCR products between *An. vaneedeni* and *An. funestus*, creating the possibility of misidentification.

A solution to this problem would be to perform two PCR reactions. The one set would contain the universal primer and the reverse species-specific primers for *An. funestus*, *An. rivulorum* and *An. leesoni*. The second PCR set would contain the universal and the species-specific primer for *An. vaneedeni*. However, this method is not financially feasible in a malaria control

programme where hundreds or thousands of specimens need to be identified. The incorporation of more species into the system will mean the addition of more primer pairs as well as more PCR reactions needed for the identification of a single unknown specimen. In the case of eight different species for example, eight different primer pairs and with three or more different PCR reactions will be needed to identify one unknown, making it too expensive to be used as a routine assay. The DNA regions analysed in this study can, however, be used for future phylogenetic studies. It is obvious that for the development of a suitable PCR assay, it would be important to have as much variation as possible especially if one needs to distinguish between more than three species. It is likely that the use of the IGS region would be more appropriate for the development of such an assay.

Chapter 4

RESTRICTION FRAGMENT LENGTH POLYMORPHISM

4.1 INTRODUCTION

Many DNases are known to cut DNA in a particular position (Old and Primrose, 1993). One such group of enzymes is the restriction endonucleases or restriction enzymes (REs). REs are enzymes that recognize a specific nucleotide (nt) sequence in a DNA molecule and cleave the DNA at that specific sequence, which generates 3'OH and 5'P termini (Freifelder, 1986). Working with DNA was complicated before the 1970s, because there was no known way to cleave double strand DNA into discrete fragments. The discovery of endonucleases in *Haemophilus influenzae* in 1970 was a breakthrough for molecular DNA technology (Kelly and Smith, 1970). The discovery of restriction endonuclease (type II) in *H. influenzae* showed none of the unusual properties displayed by RE found in *Escherichia coli* (type I). This discovery made it possible to restrict a duplex DNA molecule into discrete fragments (Kelly and Smith, 1970). Today REs are purified from a wide range of different microorganisms on a large scale.

REs are grouped according to either Type I or Type II restriction enzymes. Type I enzymes recognise specific nucleotide sequences in double strand DNA and appear to cleave at random sites in one strand of DNA 1000 -5000nt from the recognition site (Sambrook *et al.*, 1989; Old and Primrose, 1993). These enzymes are therefore not very useful for gene manipulation, since

their cleavage sites are non-specific (Old and Primrose, 1993; Freifelder, 1986).

Type II REs recognise specific sequences on double strand DNA and make two single-strand breaks in each DNA strand within or near that recognition sequence. These breaks create flush or blunt ends if the break occurs at the centre of symmetry or they create cohesive ends when the cut is symmetrically placed around the line of symmetry as showed by Figure 4.1 (adapted from Freifelder, 1986).

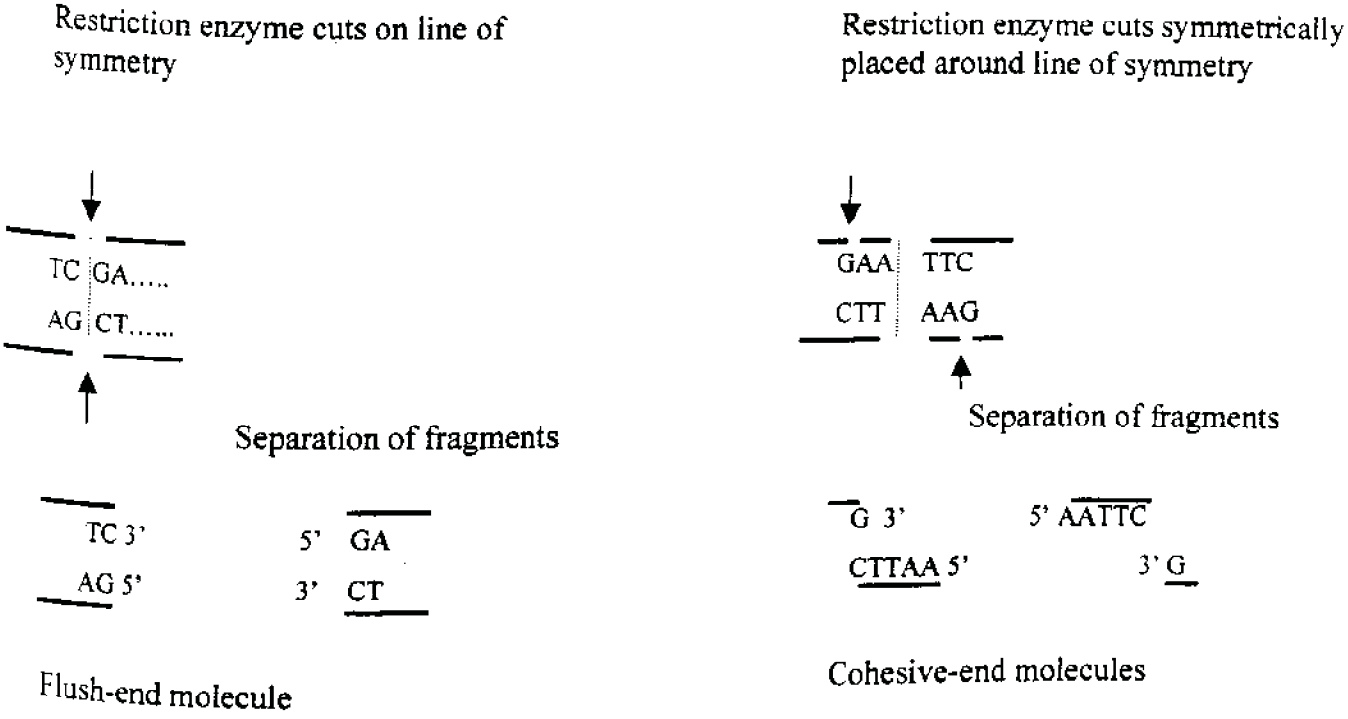

Figure 4.1 Breaks in double strand DNA molecule made by restriction enzymes. Arrow indicates the cleavage site. Dashed line is the centre of symmetry in the DNA sequence.

These enzymes are used frequently in gene manipulation since the target sites are specific and will lead to DNA fragments of specifically defined length and sequence (Old and Primrose.

1993).

Most of the REs recognise palindromic sequences (Old and Primrose, 1993). Since a particular enzyme recognises a unique sequence, the number of cuts made in the DNA from an organism is limited. This specificity of a particular RE generates a unique family of fragments from a particular DNA molecule (Freifelder, 1986). Another RE generates a different family of fragments from the same DNA molecule (Old and Primrose, 1993). DNA fragments can be visualized after digestion using electrophoresis, which separates fragments according to size.

Mutations can change these recognition sites on the DNA. These mutations might lead to the loss of a specific recognition site for an enzyme, creating different fragments and different banding patterns after electrophoresis. These differences in banding patterns created by the same enzyme, in two individuals are called restriction fragment length polymorphisms (RFLPs). RFLPs can be used to detect mutations linked to diseases, population differences or species differences (Freifelder, 1986).

RFLPs were successfully used to identify members of the *An. gambiae* complex by Collins *et al.* (1987, 1988a, 1988b) before the development of the PCR assay by Scott *et al.* (1993). Collins *et al.* (1987) used clones containing a complete rDNA cistron, and digested these with various restriction enzymes. After restriction the DNA fragments were analysed by southern blotting in order to find non-conserved regions which would indicate species-specific differences. Using two different enzymes, *EcoR* I and *Sal* I, *An. gambiae* and *An. arabiensis* could be distinguished from each other based on hybridization of the rDNA cistron to

fragments of differing sizes. The same method was used for the identification of *Culex pipiens* and *Cx. quinquefasciatus* (Severini *et al.*, 1996) where the ITS2 region was amplified and *Mbo* I used to differentiate between the two species. Black *et al.* (1989) were able to use RFLP analysis to identify variation in the nontranscribed spacer region (NTS or IGS) between different geographical populations within the species *Aedes albopictis*.

The aim of this study was to identify possible restriction enzymes capable of cleaving the DNA, creating species-specific RFLP profiles distinguishing between *An. vaneedeni* and *An. funestus*.

4.2 MATERIALS AND METHODS

DNA extraction, amplification:

DNA was extracted from single mosquitoes using the standard procedure described in Collins *et al.* (1987) and amplification of the D3 region was performed, as described in Chapter 3, Materials and Methods. Single legs were amplified by placing the leg directly into the PCR mixture (cocktail solution), centrifuging for four minutes at 16 000 rpm and then adding mineral oil. The samples were then amplified and the products stored at -20°C, after analysis by agarose gel electrophoresis and until required for RFLP analysis. A standard curve of molecular weight vs. the distances migrated from the wells was drawn, which was then used to estimate the sizes after amplification (See Chapter 3 Materials and Methods for calculation formulae).

RFLP analysis:

Sequencing results from *An. funestus* and *An. vaneedeni* were analysed using the MAPDRAW

programme from the DNASTAR suite of programmes (DNASTAR Inc. Madison, Wisconsin). This analysis was performed to find a specific RE that generated different fragment lengths after the PCR product was restricted. These different fragments were used to differentiate *An. vaneedeni* from *An. funestus*. *Hpa* II digestion was as follow: 1μl *Hpa* II (Boehringer Mannheim), 1.5μl Buffer L as recommended by suppliers, 10-15 μl PCR product in a total reaction volume of 15μL. Restriction was conducted for two hours at 37°C in the thermocycler or water bath. Digested PCR products were loaded on a 2% agarose gel for separation of fragments and visualised under UV light.

4.3 RESULTS

These results of this study have been published and a copy of the manuscript has been attached to the end of this chapter (see Acknowledgements at the beginning of this thesis for the contributions of each author). What follows is a description of the procedures followed in order to reach the results that were eventually published.

A summary of analysis from MAPDRAW and the possible enzymes that can distinguish between *An. vaneedeni* and *An. funestus* is shown in Figure 4.2, while figure 4.3 is a illustration of the restriction enzyme map of *Hpa* II on the D3 amplified product.

Anopheles rivulorum tested at a later stage, gave the same banding pattern as was visualised for *An. funestus*.

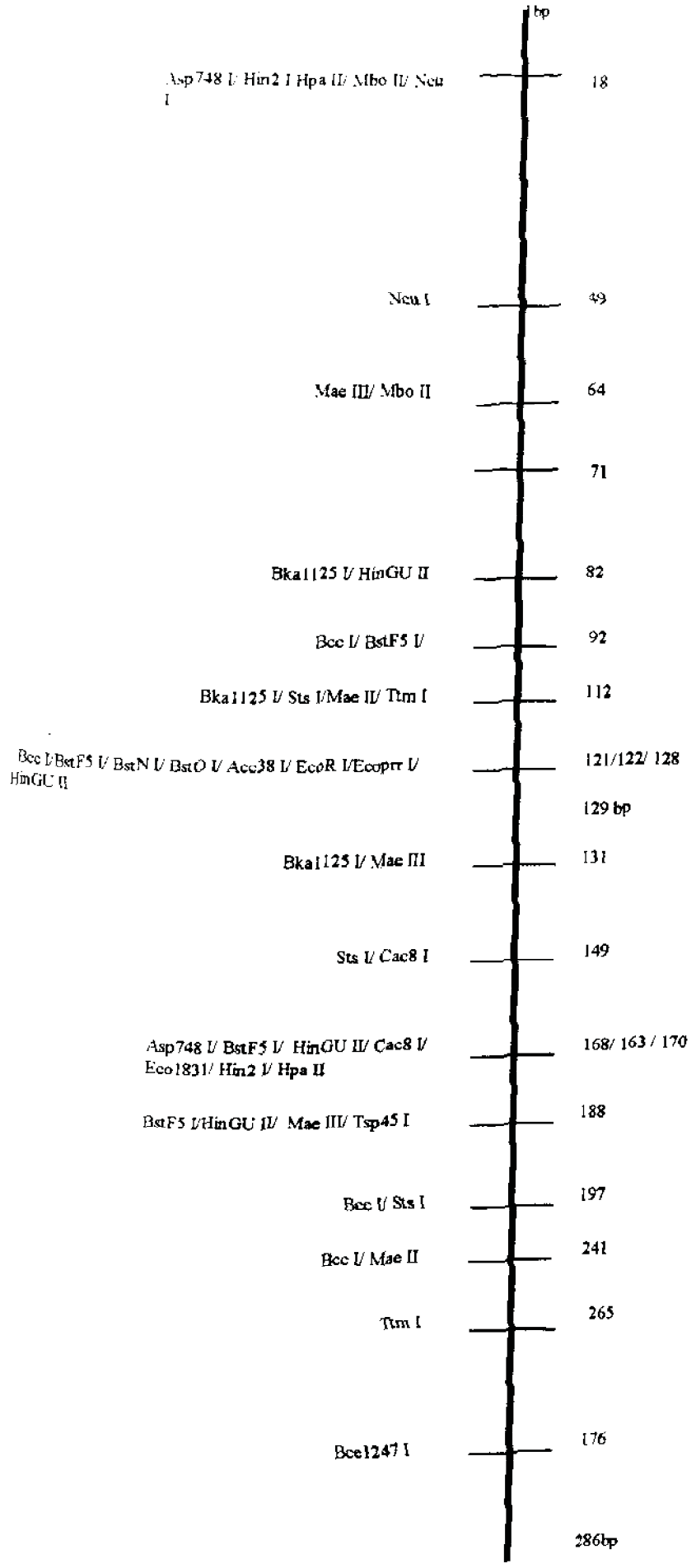

Figure 4.2a Restriction map of the amplified D3 fragment of *An. vaneedeni*

An. funestus

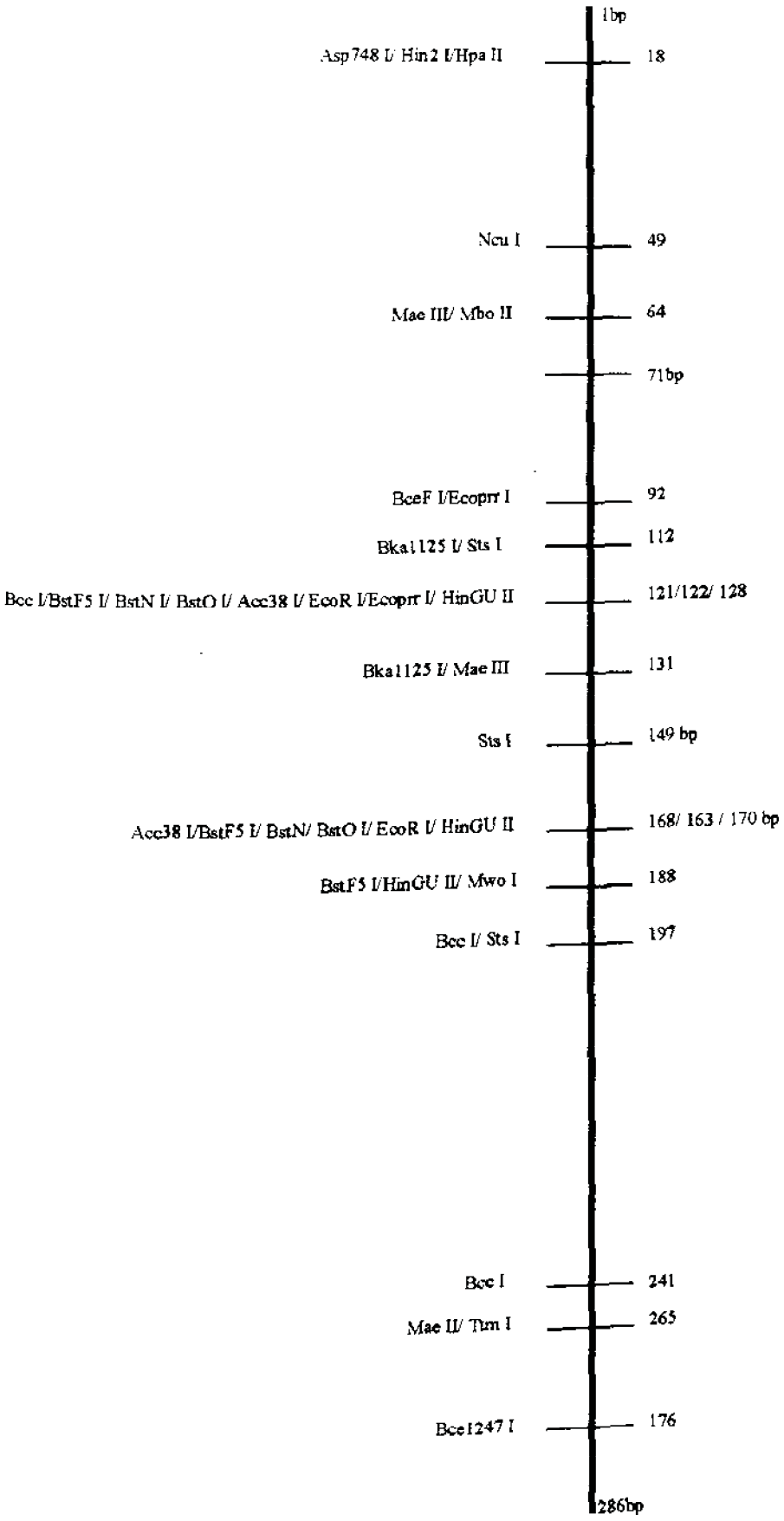

Figure 4.2b Restriction map of the amplified D3 fragment of *An. funestus*

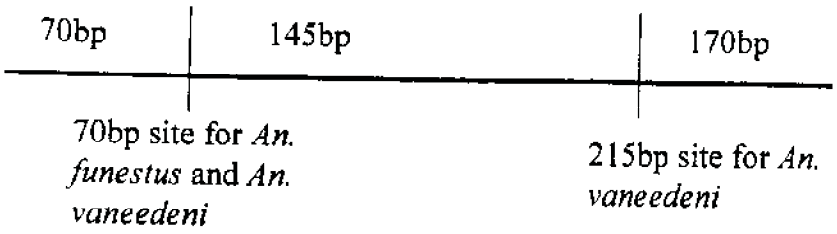

Figure 4.3: Restriction map of *Hpa* II on the D3 amplified product.

4.4 DISCUSSION

This study was initiated to distinguish *An. vaneedeni* from *An. funestus* and to develop a rapid identification method for distinguishing between *An. vaneedeni* and *An. funestus*. These two species show overlap in morphology at all life stages.

Hpa II recognises the following sequence on DNA and cleaves DNA with the particular tetra nucleotide, C↓CGG. *Anopheles rivulorum* gave the same diagnostic banding pattern after digestion as *An. funestus*. This is due to the absence of unique additional restriction sites in *An. rivulorum* compared to *An. funestus*.

Members of the *An. funestus* group are morphologically almost identical and it is likely that the same situation will also occur at the DNA level. More than one restriction enzyme will probably be needed to separate the different species and the assay would probably work on elimination process between the different restriction enzymes. De Meillon *et al.* stated in 1977: *"We are fully aware of the fact that the situation regarding the funestus group of species will not be resolved until more modern taxonomic techniques are applied"*. Twenty years later,