18S rDNA as a tool for molecular characterization of mosquito species from various regions within Limpopo.

Mzwandile Hadebe[1] and Moses Okpeku[1]

[1]University of KwaZulu-Natal

April 06, 2024

Abstract

Anopheles mosquitos are the primary vectors for the rapid spread of malaria in Limpopo, South Africa, negatively impacting all malaria elimination agenda; particularly for a country with vector control as one of the major malaria elimination strategies. The implementation of vector control strategies, however, depends on the accurate identification of mosquito vectors, which has been lacking in Limpopo as a result of earlier studies that mainly relied on morphological identification of vectors, which has numerous drawbacks that lead to misidentification of mosquitoes, thus resulting in the development and assignment of ineffective control strategies. All this together set back elimination strategies and programs. The present study therefore aimed to molecularly identify mosquitoes collected from various regions within Limpopo province. This was accomplished by examining genetic composition of mosquitos from different regions within Limpopo province, using molecular genetics techniques such as 18S rDNA-based PCR analysis, 18S rDNA sequencing, etc. This investigation was successful since 18S rDNA was able to identify every one of the 42 mosquito samples that were analysed, revealing that the samples belonged to 7 different Anopheles species and the majority of the species were molecularly demonstrated to have been misidentified morphologically. The outcomes of the molecular analysis, which involved comparing the query sequences to the sequences in GenBank, were supported by the close relationship among mosquitoes of the same species in maximum likelihood trees and the presence of high to 100% bootstrap support values, giving confidence to conclude that mosquitoes of the same species under study are genetically related. Due to a number of constraints covered in this study, morphological identification of mosquitoes is prone to misidentification of species, hence genetic characterization is recommended to be used as a confirmatory approach in the identification of the species.

18S rDNA as a tool for molecular characterization of mosquito species from various regions within Limpopo.

Mzwandile Thabani Hadebe[1] and Moses Okpeku[1]*

[1]Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Westville, Durban 4000, South Africa.

*Corresponding Author:*okpekum@ukzn.ac.za.*

MH:*hadebemzwandile160@gmail.com.*

MO: *okpekum@ukzn.ac.za.*

Abstract

Anopheles mosquitos are the primary vectors for the rapid spread of malaria in Limpopo, South Africa, negatively impacting all malaria elimination agenda; particularly for a country with vector control as one of the major malaria elimination strategies. The implementation of vector control strategies, however, depends on the accurate identification of mosquito vectors, which has been lacking in Limpopo as a result of earlier studies that mainly relied on morphological identification of vectors, which has numerous drawbacks that lead

to misidentification of mosquitoes, thus resulting in the development and assignment of ineffective control strategies. All this together set back elimination strategies and programs. The present study therefore aimed to molecularly identify mosquitoes collected from various regions within Limpopo province. This was accomplished by examining genetic composition of mosquitos from different regions within Limpopo province, using molecular genetics techniques such as 18S rDNA-based PCR analysis, 18S rDNA sequencing, etc. This investigation was successful since 18S rDNA was able to identify every one of the 42 mosquito samples that were analysed, revealing that the samples belonged to 7 different *Anopheles* species and the majority of the species were molecularly demonstrated to have been misidentified morphologically. The outcomes of the molecular analysis, which involved comparing the query sequences to the sequences in GenBank, were supported by the close relationship among mosquitoes of the same species in maximum likelihood trees and the presence of high to 100% bootstrap support values, giving confidence to conclude that mosquitoes of the same species under study are genetically related. Due to a number of constraints covered in this study, morphological identification of mosquitoes is prone to misidentification of species, hence genetic characterization is recommended to be used as a confirmatory approach in the identification of the species.

Keywords : *Anopheles* , Multiple Alignment sequencing, Phylogenetic construction, 18S rDNA, Genetic variation.

Introduction

Anopheles mosquitoes are essential vectors in the distribution of the malarial parasite, *Plasmodium* , hence monitoring mosquito populations is a crucial factor in predicting malaria risk in a geographic region of interest. The number of mosquito species is estimated to be 3500, and they are divided into 41 genera (Zheng, 2020). Anopheles, Aedes, and Culex are three of these 41 genera that are thought to be crucial in the spread of human illness caused by mosquitoes (Adugna et al., 2021). Of the 530 species of*Anopheles* , only 30 to 40 naturally spread malaria (Nicoletti et al., 2020). Various Anopheles mosquito species have been found to prefer distinct habitats; thus, the *plasmodium* parasite is spread by different Anopheles species depending on geographic locations (Nicoletti et al., 2020).

According to Adugna et al., (2020), identification of *Anopheles*species is crucial since each species has a different treatment or preventative strategy. However, identification of *Anopheles*mosquitoes in Limpopo has mainly concentrated on morphological characterization, which is likely to misidentify *Anopheles*species due to high similarity of phenotypic features (Jones et al., 2021), leading to assignment of certain insecticides or treatment to incorrect *Anopheles* species which may be the reason why some insecticides do not kill some Anopheles mosquitoes present in the geographic region of interest. As a results, previous studies found it more convenient to confirm morphological characterisation with molecular characterization through studying genetic diversity of *Anopheles*mosquito species present in the area of interest. Geneticists predominantly believe in molecular characterization since it employs extremely precise molecular procedures such as PCR, Agarose gel electrophoresis, DNA sequencing, and so on, which employ very sensitive components such as genetic markers. Genetic markers are DNA sequences with known positions on chromosomes. They play crucial role in the study of populations, cell identification, as well as species classification (Safdar, 2011).

One of the genetic markers of interest employed in this study's molecular analysis of the *Anopheles* mosquito species is 18S ribosomal DNA (18S rDNA). According to Carranza et al., (1996), all prokaryotic and eukaryotic cells use rDNA as a critical component of their protein production machinery. Due to its high conservation among species and the presence of variable regions, 18S rDNA has been frequently employed for identifying and assessing the genetic diversity of eukaryotes (Kounosu et al., 2019). This sequence conservation is thought to represent functional limitations on the molecules necessary for best translational effectiveness. According to Rackevei et al., (2022), V1 to V9 variable regions of 18S rDNA are commonly used in biodiversity studies. Despite being a member of a multigene family, the 18S rDNA has undergone concerted development to keep all of its copies homogenous(Carranza et al., 1996). This marker is also being used to determine phylogenetic connections between living organisms (Carranza et al., 1996). There are lot of reason for this maker to be used for studying phylogenetic analysis and those reason are detailed in a

review by Woese (1987); Sogin (1991); and Adoutte et al., (1993).

PCR, Agarose gel electrophoresis, Multiple sequence alignment (MSA), and phylogenetic analysis were used to examine the genetic variance of distinct mosquito species. Genetic variation is simply the variance in genomic DNA sequence between individuals within a community/same species (Al-Koofee et al., 2020). PCR was mainly utilized to amplify the areas of interest, and the existence of the target region was determined by agarose gel electrophoresis. Following Agarose gel electrophoresis, samples must be further analyzed with MSA. This method is used to determine the evolutionary links and common patterns among target genes (Sofi et al., 2022). According to a genetic concept, MSA is the alignment of at least three genomic DNA sequences of same length. Computational methods are used to produce and analyze these alignments (Sofi et al., 2022).Identification of a novel protein members after comparing them to related sequences is an additional role played by MSA (Shukla et al., 2022). Since further analysis of research is dependent on the outcome of the MSA, the accuracy of the MSA plays a crucial role. Consequently, creating trustworthy and accurate MSA tools will always be a worthwhile endeavor(Shukla et al., 2022). These techniques, along with targeted genetic markers, are considered to be exceedingly sensitive and very specific, which is why they were chosen to measure mosquito diversity and give trustworthy findings.

Material and Methods

Collection and processing of samples.

A total of one hundred and thirty-four (134) mosquito samples were utilized in this investigation. These specimens were obtained from different regions in Limpopo province including Masisi, Mutale, Tshivaloni, Xikukwani, Makoxa, Thomo, Seloane, Domboni and Musunda as shown in figure 1 and these regions have a unique geographic coordinate (Table 1). Samples were obtained using two different techniques: pits or larvae collecting methods. Stored samples acquired by the Limpopo Malaria Institute were transferred safely and securely from Limpopo to the University of KwaZulu-Natal and stored in the freezer at Genetic Laboratory 3, Dr Moses's laboratory before DNA extraction. In order to prevent DNA mixing in case that mosquitoes were able to feed prior to collection, the abdomen was removed from all samples. Only the legs, head, and thorax were utilized for DNA extraction. The abdomen was removed while the specimen was viewed under the microscope. PCR was used to amplify regions of interest where the successful amplification was confirmed by presence of either one or more bands per tube. The majority of tubes contained a pool of three mosquitoes, however some tubes contained fewer or more than three mosquitoes. The main purpose of this was to simply determine whether the quantity of mosquitoes had any impact on the concentration of DNA after extraction. In this investigation, a total of 42 tubes containing pooled mosquitoes were employed. According to the absorbance at 260/280, which was at least 1.6 for all tubes, the DNA produced was of excellent quality.

Figure 1: South Africa Map showing different geographic areas where mosquito samples were collected.

Table 1 : Different regions of collection and their corresponding geographic coordinate

DNA extraction

Some of the samples that were gathered were undocumented, and those samples were utilized to evaluate the best DNA extraction tool. The first kit used to extract DNA from a pool of mosquitoes was the Promega Wizard® Genomic DNA Purification Kit, but this kit produced very little DNA yield as it employs a micro pestle, which doesn't crush mosquitoes very effectively. However, the beads utilized by the Zymo-Research Quick-DNA Tissue/Insect Miniprep Kit for crushing mosquitos properly crushed the mosquitos, resulting in the best DNA yield suitable for PCR. As a result, the Zymo-Research Quick-DNA Tissue/Insect Miniprep Kit was employed for DNA extraction in this investigation. As previously stated, DNA was obtained from the legs, thorax, and head of mosquitos, with the abdomen left out to avoid DNA mixing if mosquitos were allowed to feed before collection. In this study, 134 mosquito samples were examined; however, mosquitos were pooled, with the majority of tubes holding a pool of three mosquitos, resulting in 42 tubes containing diverse DNA samples. DNA was successfully extracted according to manufacturer's instruction; no modifications

were made. The quantity and purity of the extracted DNA were evaluated using a NanoDrop 2000 UV Visible Spectrophotometer, and the DNA was then kept at -20 C until it was utilized for genetic analysis by polymerase chain reaction (PCR).

Amplification of 18S rDNA gene using PCR .

The polymerase chain reaction was employed to amplify the 18S rDNA region in DNA obtained from pooled mosquitos. Primers that were utilized to amplify 18S rDNA were; Forward primer 5' GAG GGA GCC TGA GAA ACG GCT AC 3' and Reverse primer 5' CCT TCC GTC AAT TCC TTT AAG TTT C 3' (Beebe et al. 1996). The Bio-Rad T100 Thermal Cycler was employed for PCR amplification. A 25µL PCR reaction was set up as follows: 12.5µL of Dream Taq Master Mix, 1µL of each primer (1.9 µM concentration), 1 µL of genomic DNA, and 9.5µL of nuclease-free water. Since the conditions often employed by prior research didn't work well for the primer, it needed to be optimized. As a result, 62.2°C annealing temperature yielded the best results, thus the following PCR conditions were used to amplify 18S rDNA: Denaturation: 94°C for 4 minutes, 1 cycle; annealing: 95°C for 1 minute, 62.2°C for 1 minute, and 72°C for 1 minute, 35 cycles; final extension: 72 C for 7 minutes, 1 cycle; hold: 4 C.

Agarose gel electrophoresis and viewing of amplified PCR products of 18S rDNA.

Availability of target DNA region was checked using agarose gel electrophoresis. A 1.5% agarose gel stained with EtBr was used to run 5 µL of each sample. This type of agarose gel was made by just mixing 100 mL of 1X TAE buffer (Bio Concept, Switzerland) with 1.5g of agarose powder (Cleaver scientific, United Kingdom) in a 1000ml beaker. The powder was dissolved in a buffer by heating the mixture in the microwave for approximately three minutes. The mixture was then allowed to cool before being poured into a gel casting tray with a comb in to produce wells. Then after, the mixture was allowed solidified to form a gel that was used to run 5µL of PCR product in electrophoresis. A hundred bp (100bp) molecular weight marker (New England Biolabs, China) was employed as a reference for comparing DNA fragment sizes of PCR products. One µL of the molecular weight marker was combined with two µL of 6X purple dye. This dye aided in the visualization of a molecular weight marker. Electrophoresis was performed at 100 volts for 60 minutes utilizing Enduro gel XL electrophoresis system (Labnet, United States of America). The ChemiDoc MP imaging machine (Bio-Rad, United States of America) was used to visualize the DNA bands in the gel.

Phylogenetic Analysis.

The 18S ribosomal DNA mosquito samples were sent to Stellenbosch for sequencing. The base calling and trimming were done utilizing FinchTV. The sequences were aligned using the Clustal W multiple alignment function of the BioEdit program, version 7.2.5. MEGA version 6 (Tamura et al., 2013) was used to create phylogenetic trees using the maximum likelihood (ML) technique.

Result

The only mosquito genera investigated in this research were *Culex*and *Anopheles* , although there is one species of mosquito that does not belong to any of the three discovered mosquito genera (*Anopheles* , *Aedes* , and *Culex*); this newly morphologically recognized species is known as *Dimelion* . These different species of mosquitos were collected at different locations within Limpopo province (Figure 1). Among mosquito samples collected,*Anopheles* genera was more prevalent over the other genera,*Culex* . Only six mosquito samples were *Culex* of the 134 species collected, and they were all obtained from Makoxa. 128 Samples belonged to *Anopheles* genera and were collected from different regions in Limpopo (Table 2.). After Polymerase chain reaction amplification, DNA for different samples was run in a 1.5% agarose gel electrophoresis. To amplify samples Tube1- Tube42, which were labelled as T1, T2, etc, 18S r DNA was utilized. As seen in table 3, tubes (T1 -T42) contained a pool of mosquitos of the same species. The marker of interest previously mentioned (18S rDNA) produced varying amplicon sizes (900bp – 910bp) after running them on agarose gel electrophoresis as shown in gel images in figure 2. When repeated for the 3rd time 18S rDNA successfully amplified T26 which didn't amplify at first place. 5ml of PCR mixture was added into each well for each sample and ran at 100 volts for 60 minutes and ChemiDoc imaging system was used to view DNA profiles.

A 100bp ladder stained with purple dye was used as a reference to estimate sizes of PCR products. Prior to sequencing, amplicons were purified to eliminate any residues that may cause interference if not removed, such as non-specific bands.

Table 2. Various mosquito species, their numbers, and the area where they were collected in South Africa, Limpopo province.

Table 3. Pools of mosquito species per Tube (T1-T42) and their corresponding geographic locations and coordinates.

Figure 2. T1-T42 mosquito samples' 18S rDNA amplification profiles. Lane MW indicates a 100bp DNA ladder (New England Biolabs, China). It is also known as molecular weight marker (MWM). It was basically used to estimate PCR product sizes. Lane NC symbolizes the negative control. The obtained PCR products were then run at 100 volts on 1.5% agarose gel with 1X TAE running buffer. To obtain optimal separation of the DNA fragment on the agarose gel, the DNA agarose gel electrophoresis was allowed to run for 60 minutes. In certain mosquito species, the 18S rDNA amplicon is 900 bp in size, whereas in others, it is 910 bp. As a result, some tubes contained two bands (900 bp and 910 bp), showing that the pooled mosquitoes were not all from the same species.

The outcomes of multiple sequence alignment (MSA) in figures 3 do not show any discernible variation in the 18S rDNA region among the mosquito species under investigation. By comparing sequences found in this study with those on Genbank, this study was able to determine whether or not the Limpopo malaria institute's morphological identification of mosquito species was accurate. Even though there are some gaps in the query sequences after alignment, they display little to no variation compared with the reference sequences. However multiple sequence alignment wouldn't give reliable results alone, hence phytogenic diagram of 18S rDNA (Figure 4) was constructed to support MSA results by showing relationship that exist among the species of mosquitos under investigation. Multiple alignment sequences and phylogenetic analysis proved that the "*Dimelion* " species of mosquito, which was once believed to be an entirely novel species identified based only on its morphology, is in fact *An.gambiae* . An additional piece of evidence that *Dimelion* is not a new species of mosquito discovered in Limpopo, but rather *An.gambiae* , is that its sequence revealed a 99.65% identity with *An.gambiae* _OM350318.1 and no *Dimelion* species was observed in hits after blasting the query sequence. *An.gambiae* was not the only species that was mistakenly recognized morphologically; nonetheless, molecular identification disapproved other species' names that were given to them by morphological procedures (Table 3). These mosquito species are believed to have been misclassified based on morphology because, when their 18S rDNA sequences were blasted, they displayed high percentage identities to other mosquito species, but the name assigned to them based on morphological features did not appear in possible hits. However, some sequences (T25_18S rDNA, T26_18S rDNA, and T28_18Sr_DNA) did not match any sequence when blasted on NCBI blast and were too divergent from other sequences to be included in the construction of a phylogenetic tree. As a result, they were not included in the MSA or phylogenetic construction. Five sequences were too short to be aligned with other sequences, leading in an error while attempting to match them with other sequences; hence, they were omitted from MSA. These sequences include T7_18S rDNA (*An.gambiae*), T17_18S rDNA (*An.gambiae*), T33_18S rDNA(*An.listeri*), T36_18S rDNA (*An.pretoriensis*), T42_18S rDNA (unknown).

Looking at the phylogenetic tree for 18S rDNA in figure 4, the existence of an elevated percentage of bootstrap support values, some of which are 100%, indicates that the data in table 3 is accurate. As a result, 18S rDNA phylogenetic analysis verified that the information acquired by comparing query sequences with reference sequences (Table 3) on Genbank through blasting is valid, implying that a large number of mosquitos were mistakenly classified based on their morphological traits. The 18S rDNA phylogenetic tree (figure 4) shows that *tsetse flies* which were used as outgroup are the most recent common ancestor and that two major clades diverged from the 95% node as a common ancestor. A high percentage of bootstrap support values guarantees that each of the involved species is closely linked. Regardless of what morphological traits suggests, species within the same clade and closer to one another are thought to be genetically linked. Despite the six species reported by morphological identification, genetic identification confirmed only seven mosquito

species and one non-mosquito species*(Diaphorina*) among the samples collected. These mosquito species include *An.gambiae* , *An.sundaicus* , *An.melas* ,*An.coluzzii* , *An.merus* , *An.maculipalpis* , and*An.funestus* . The phylogenetic tree (Figure 4) shows that mosquitos of the same species are closely related, suggesting that molecular identification was accurate. However, there are notable exceptions of species that are closely related to various other species due to some reasons that are highlighted in the discussion section.

Figure 3 : DNA bases of Multiple sequence alignment of 18S rDNA targeted in various species of mosquitoes collected from Limpopo. To generate alignment, the BioEdit program was utilized in conjunction with the Clustal W multiple alignment tool.

Table 4 : Molecular characterisation of different mosquito species through database searching and availability of 18S rDNA sequences.

Figure 4: Phylogenetic reconstruction of the various species of mosquitos using the 18S rDNA gene (900bp). Mega X software was used to generate phylogenetic tree, whereby the Maximum likelihood method inferred by kimura 2 parameter model after 100 replicates was utilized. Sequences from the present study start with letter "T", T1-T41 . Bootstrap supports values are shown in percentage (%) , and the scale for the above phylogenetic tree is 0.2.

Discussion

The present investigation portrays genetic characterisation of several mosquito species that have been morphologically defined by the Limpopo Malaria Institute. According to morphological characterization, the species under examination are from the *Anopheles* and *Culex*subfamilies, and there has never been a study like this one in Limpopo. The main aim of this investigation was just basically to confirm or disprove Limpopo Malaria Institute's morphological categorization of mosquito species, as well as to investigate the accuracy of genetic markers in identifying closely related species. The Limpopo Malaria Institute employed the two techniques of collecting mosquitos indicated in 3.2.1 (pits or larvae), and after collection, they utilized standard morphology-based taxonomy approach to identify mosquitos morphologically, and two genera (*Anopheles* and *Culex*) were detected. As previously demonstrated (Table 2), the majority of species identified by morphological features belonged to *Anopheles* (102 species), only 6 were classified as Culex, and 26 were thought to be new species (*Dimelion*) that had not been classified previously, however molecular characterization disproved that.

Gel images show positive results for all samples when the 18S rDNA region were amplified; however, some samples displayed multiple bands per sample, which was thought to be caused by the presence of various species in the pool of mosquitos. However, sequencing only produced one sequence per sample, i.e., one sequence per sample for 18S rDNA marker, with no varying sequences per sample as an indication of various species present in a pool of mosquitoes. Therefore, gel pictures that displayed more than one DNA band per well may be an indication that non-specific bands have been amplified (Bovo et al., 1999). DNA barcoding using 18S rDNA was able to identify species that were mistakenly classified morphologically. The occurrence of high percentage identity (81%-100%) between query and reference sequences on GenBank is evidence of successful molecular characterisation and validation of morphological identification of distinct mosquito species (Tables 3.3). Some sequences of samples that were morphologically described as *An.rufipes* (T25, T26, and T28) matched no sequence on GenBank, suggesting that these are new haplotype of *An.rufipes* found by the current study using 18S rDNA.

The results reported in this study (Table 3) demonstrates that many samples were incorrectly categorized morphologically. This was proven by molecular analysis, which makes use of highly specific molecular techniques (Wilson et al., 2000). Damage to essential distinguishing features, human error, the occurrence of novel or cryptic species, the presence of species showing overlapping or unreported traits, and intraspecific morphological changes are all potential causes of misclassification of mosquito species (Zhang et al., 2022). Despite the fact that the majority of mosquito species were incorrectly classified morphologically, Molecular characterisation using 18S rDNA revealed that certain species were accurately identified. The identification of samples T5, T8, T9, T18, T19, T20, and T21 as *An.gambiae* by 18S rDNA and morphological charac-

terization indicates that the morphological characterization of these species was valid. Individual pools of mosquito species were represented by highly supported clades, with high percentages of bootstrap values up to 100%, validating the morphological identification of diverse mosquito species investigated.

According to morphological characterization of mosquito species, six different species of mosquitos were obtained; however, molecular characterization by 18S rDNA demonstrated that there were actually seven different mosquitos studied plus one non-mosquito species, and among these species of mosquitos, some had been overlooked by morphological characterization and others were mischaracterized (Table 3). These eight species found by molecular characterisation comprise species that were mistakenly categorized by morphology as well as species that were verified to be appropriately classified morphologically by molecular analysis. Seven species of mosquitos found by molecular characterization using 18S rDNA includes *An.gambiae* , *An.sundaicus* ,*An.melas* , *An.culuzzi* , *An.merus* ,*An.maculipalpis* , and *An.funestus* . In addition to these species, 18S rDNA also discovered a non-mosquito species ,*Diaphorina* , which was thought to be *Culex* species, and this serves as a proof that morphological identification alone cannot be trusted when it comes to characterization of various species of mosquitos. Unknown species (T42) were recognized by 18S rDNA as*An. gambiae* . The 18S rDNA based molecular characterization of samples that were not documented , T42(Unknown), discovered that the pool of mosquitos in this sample belong to *An.gambiae* and this was supported by a very high percentage similarity of 18S rDNA query sequence to MG753768.1 (*An.gambiae*) reference (81.32%) and a being relatively small e value (3e-95). The accuracy of 18S rDNA characterization results (Table 3) was supported by a close relationship that exist between mosquitos of the same species and their reference sequences in the phylogenetic construction (Figure 4)

In a phylogenetic tree, species in the same clade and near to one another are genetically connected, i.e., the closer the species are to one another, the closer their genetic relationship is (HAMZAOLU et al., 2017). Other studies made an effort to modify the aforementioned claim in order to conform to the world of science; they claimed that two species are more connected if they share a most recent common ancestor and less related if they share a less recent common ancestor (Lo et al., 2003; Gregory,2008). Furthermore, the presence of a node with a high bootstrap value close to 100% suggests that the species that diverged from that node, also known as a common ancestor, are closely related to one another and that their genetic makeup is identical or contains very little variation. Since previous studies have shown that a very high bootstrap value up to 100% gives researchers confidence in concluding that particular species are siblings. Hence, high bootstraps support values in the 18S rDNA phylogenetic construction and relatively small scale (0.2) imply a close relationship between the species being studied. Phytogenic trees (Figures 4) investigated variation well because it shows that mosquitos of the same species are more closely related, sharing clades and most recent common ancestors.

It is evident from a detailed examination of the 18S rDNA phylogenetic tree that mosquitoes belonging to the same species are genetically related to one another. This tree diagram demonstrates two significant branches that diverged from the 95% node that was determined to represent the most recent common ancestor of all the species analyzed in the current study, and that divergence resulted in the formation of two major clades. *An. gambiae* species are found to be closely linked to one another in the top clade, and their tight relationship is supported by the occurrence of a high bootstrap value of 99% where all of these species diverged from. Regardless of An.maculipalpis being closely related to each other , but their divergence from a common ancestor with *An.gambiae* , supported by 85% bootstrap value, provides evidence that *An.maculipalpis* and *An.gambiae* are genetically linked. Additionally, this tree diagram was able to show that *An. sundaicus* are related to each other, *An. melas* as a reference species, and *An.maculipalpis* , however they are indeed distantly related to *An.gambiae* . *Diaphorina* , a non-mosquito species, was demonstrated to be genetically linked to*An.gambiae* , *An.maculipalpis* , *An. melas* , and*An.sundaicus* by sharing a clade and diverging from a common ancestor. This relationship is supported by a 100% bootstrap support value. Looking at the bottom branch, there are *An. gambiae* and*An.maculipalpis* species that are closer to one another and their reference sequences, providing proof that these species are in fact *An.gambiae* and *An.maculipalpis* . These species, however, are distantly related to the same species in the upper clade, and the occurrence of these species in various geographical locations may be the cause of this variability

(Marcus et al., 2017). There is also the presence of *An.funestus* reference species in this clade, which is distantly linked to T22 *An.funestus* from this study, and this variation might be attributable to a variety of factors, including the one previously stated, mutation, geneflow, or sexual reproduction (Barton, 2010). This reasoning also applies to other species that are distantly linked to their phylogenetic references, such as*Diaphorina* , *An.gambiae* , and others. This information combined offers a proof that 18S rDNA molecular identification was more accurate than morphological identification, as it can seen that the species stated to be linked by molecular characterisation appear to be connected in the phylogenetic tree too (Figure 4).

It is possible that there is still a significant danger of malaria transmission in Limpopo given the presence of so many distinct*Anopheles* species in this province. According to 18S rDNA, this study has discovered a possibly novel haplotype of the *Anopheles*species (T25, T26, and T28 and T32). The main reason behind this fact is that the sequences for these sample show no match when compared to sequences on Genbank. *Anopheles* species are constantly being discovered via the use of molecular methods; for example, new species of*Anopheles nuneztovari* have been discovered in Brazil (Scarpassa et al., 2016). Based on the results, it can be inferred that there is a significant and widespread group of mosquitoes in all of the Limpopo regions that are close to the coast and along the border (Table 1). This is in line with earlier research which showed that the majority of Anopheline mosquito species are found in temperate and tropical climates (Schäfer et al., 2001). All this together causes a rapid spread of malaria in Limpopo and contributes to a rise in malaria cases and malaria-related deaths. According to earlier research, *An. funestus* , *An. arabiensis* , *An.pretoriensis* ,*An.quadriannulatus* , and *An.gambiae* are the only species of Anopheles present and in charge of transmitting malaria in South Africa, specifically in Limpopo (Burke et al., 2019; Dahan-Moss et al., 2020; Braack et al., 2020; Dahan et al., 2020). However, molecular characterisation carried out in the present study discovered numerous more *Anopheles* species that were assumed to be missing in South Africa, including as *An.sundaicus* , *An.melas* ,*An.coluzzii* , *An.merus* , *An.maculipalpis* ,*An.triannulatus* , and *An.darlingi* . These Anopheline mosquito species may have been imported to South Africa by immigrants via luggage or flights from various regions of the world. The rising incidence of arbovirus (malaria) epidemics in Limpopo, and the rapid propagation of such diseases, as well as high volume of the public health consequences (Charrel et al., 2007; Enserink, 2007; Semenza et al., 2014), have prompted multiple calls for South Africa to practice greater vigilance regarding arboviruses, as well as an associated need to comprehend the population status of existing or potential vector mosquitoes (Cornel et al., 2018). Hence prior research investigated the diversity of mosquito species since some scientists observed that mosquitos are treated based on the type or kind of species found in a specific location. Presence of these various species as vectors of malaria which were not known to be present in Limpopo really contributed to increasing cases of malaria in Limpopo due to the fact that they are given wrong treatment. The necessity of understanding the genetic makeup of different mosquito species is demonstrated by the information presented here, and it also made it clear that morphological characterisation of mosquito species alone does not provide trustworthy findings due to its numerous restrictions and lack of specificity.

Conclusion

This study demonstrated that *Anopheles* mosquito species recognition may be accomplished with success using molecular characterization. DNA barcoding results, which utilized the genetic marker 18S rDNA, were equivalent to the morphological identification of some species and also allowed for identification of morphologically misidentified species. It is crucial to highlight that in situations when conventional morphological classification was unsuccessful owing to damaged specimens and unidentifiable characteristics, DNA barcoding successfully identified these species. The newly discovered sequence could potentially be utilized as reference sequences in future mosquito taxonomy investigations. The successful identification of numerous mosquito species in this study will assist with malaria research by determining the real and prospective danger of arbovirus or parasite transmission, as well as providing fundamental knowledge for vector surveillance and enhancing vector control strategies. In the future, molecular identification implementations should include barcoding a greater number of species and incorporating other genetic markers that improve the discriminatory strength of this identification approach. DNA barcoding might also be used in conjunction with next generation sequencing to identify huge numbers of mosquitos at once, considerably reducing the

processing time required for species characterization.

Author Contributions

Mzwandile Thabani Hadebe: Designed and drafted the manuscript - Methodology, Software, Formal analysis, Investigation, Data curation. Moses Okpeku: corrected the drafts and approved the final draft of the manuscript.

Acknowledgments

I acknowledge the support of NRF (National Research Foundation) and SMRC (South African Medical Research Council) for their contributions to this research. I would also like to extend my gratitude to Mr George Makhubela of the Malaria Limpopo Institute for his contribution with sample collection.

Funding Information

This work was supported by NRF—PMDS22061523254

Conflict of interest statement

The authors declare no conflict of interest.

Availability of data

The data supporting findings of this manuscript are available within the article

3.6. References

1. Zheng XL. Unveiling mosquito cryptic species and their reproductive isolation. Insect Mol Biol. 2020 Dec;29(6):499-510. doi: 10.1111/imb.12666. Epub 2020 Aug 18. PMID: 32741005; PMCID: PMC7754467.
2. Adugna F, Wale M, Nibret E. Review of Anopheles Mosquito Species, Abundance, and Distribution in Ethiopia. J Trop Med. 2021 Sep 23;2021:6726622. doi: 10.1155/2021/6726622. PMID: 34603455; PMCID: PMC8486561.
3. Nicoletti M., Chapter Five - Three scenarios in insect-borne diseases, Editor(s): Marcello Nicoletti, Insect-Borne Diseases in the 21st Century, Academic Press, 2020, Pages 99-251, https://doi.org/10.1016/B978-0-12-818706-7.00005-X.
4. Adugna T, Getu E, Yewhalaw D. Species diversity and distribution of Anopheles mosquitoes in Bure district, Northwestern Ethiopia. Heliyon. 2020 Oct 16;6(10):e05063. doi: 10.1016/j.heliyon.2020.e05063. PMID: 33102831; PMCID: PMC7569303.
5. Jones CM, Ciubotariu II, Muleba M, Lupiya J, Mbewe D, Simubali L, Mudenda T, Gebhardt ME, Carpi G, Malcolm AN, Kosinski KJ, Romero-Weaver AL, Stevenson JC, Lee Y and Norris DE (2021) Multiple Novel Clades of Anopheline Mosquitoes Caught Outdoors in Northern Zambia. Front. Trop. Dis 2:780664. doi: 10.3389/fitd.2021.780664.
6. Safdar M,(2011), Genetic Marker : Its Definition, Introduction , Background and Types., Biotech Articles.
7. Kounosu, A., Murase, K., Yoshida, A. et al. Improved 18S and 28S rDNA primer sets for NGS-based parasite detection. Sci Rep 9, 15789 (2019). https://doi.org/10.1038/s41598-019-52422-z
8. Rackevei AS, Borges A, Engstler M, Dandekar T, Wolf M. About the Analysis of 18S rDNA Sequence Data from Trypanosomes in Barcoding and Phylogenetics: Tracing a Continuation Error Occurring in the Literature. Biology. 2022; 11(11):1612. https://doi.org/10.3390/biology11111612.
9. Carranza S, Giribet G, Ribera R, Baguii J, Riutort M, Evidence that Two Types of 18s rDNA Coexist in the Genome of Dugesia (Schmidtea) mediterranea (Platyhelminthes, Turbellaria, Tricladida). ML Biol. Evol. 1996, 13(6):824-832.
10. Woese C. R. 1987. Bacterial evolution. Microbial. Rev. 51: 221-271.
11. Sogin M. L. 1991. Early evolution and the origin of eukaryotes Cut r. Opin. Genet. Dev. 1:457-463.

12. Adoutte A., Philippe H., (1993), The major lines of metazoan evolution: summary of traditional evidence and lessons from ribosomal RNA sequence analysis, Y. PICHON, ed. Comparative molecular neurobiology. Birkhguser Verlag, Base, Switzerland. Pp. I-30.
13. Al-Koofee DAF, Mubarak SMH, "Genetic Polymorphisms," The Recent Topics in Genetic Polymorphisms, May 2020, doi: 10.5772/intechopen.88063.
14. Sofi MY, Shafi A, Masoodi KZ, Chapter 6 - Multiple sequence alignment, Editor(s): Sofi MY, Shafi A, Masoodi KZ, Bioinformatics for Everyone, Academic Press, 2022, Pages 47-53. https://doi.org/10.1016/B978-0-323-91128-3.00011-2.
15. Shukla S, Mishra B, Avashthi H, Muktesh Chandra, Chapter 3 - Biological sequence analysis, Editor(s): Dev Bukhsh Singh, Rajesh Kumar Pathak, Bioinformatics, Academic Press, 2022, Pages 33-47, https://doi.org/10.1016/B978-0-323-89775-4.00003-1.
16. Beebe NW, Foley DH, Cooper RD, Bryan JH, and Saul A. 1996. DNA probes for the Anopheles punctulatus complex. Am J Trop Med Hyg. 54: 395-398.
17. Beard CB, Hamm DM, and Collins FH. 1993. The mitochondrial genome of the mosquito Anopheles gambiae: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol Biol 2: 103-124.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013 Dec;30(12):2725-9. doi: 10.1093/molbev/mst197. Epub 2013 Oct 16. PMID: 24132122; PMCID: PMC3840312.
19. Bovo D, Rugge M, Shiao YH. Origin of spurious multiple bands in the amplification of microsatellite sequences. Mol Pathol. 1999 Feb;52(1):50-1. doi: 10.1136/mp.52.1.50. PMID: 10439841; PMCID: PMC395672.
20. Wilson KM, Schembri MA, Baker PD, Saint CP. Molecular characterization of the toxic cyanobacterium Cylindrospermopsis raciborskii and design of a species-specific PCR. Appl Environ Microbiol. 2000 Jan;66(1):332-8. doi: 10.1128/AEM.66.1.332-338.2000. PMID: 10618244; PMCID: PMC91826.
21. Zhang, C., Luo, C., Yang, R. et al. Morphological and molecular identification reveals a high diversity of Anopheles species in the forest region of the Cambodia–Laos border. Parasites Vectors 15, 94 (2022). https://doi.org/10.1186/s13071-022-05167-0.
22. Wiesemüller B, Rothe H. Interpretation of bootstrap values in phylogenetic analysis. Anthropol Anz. 2006 Jun;64(2):161-5. PMID: 16850767.
23. Weeraratne TC, Surendran SN, Walton C, Karunaratne SHPP. Genetic diversity and population structure of malaria vector mosquitoes Anopheles subpictus, Anopheles peditaeniatus, and Anopheles vagus in five districts of Sri Lanka. Malar J. 2018 Jul 20;17(1):271. doi: 10.1186/s12936-018-2419-x. PMID: 30029664; PMCID: PMC6053832.
24. HAMZAOĞLU, E., Koc, M., & BÜYÜK, İ. (2017). Dianthus ucarii (Caryophyllaceae): a new species from the northwest of Turkey. Turkish Journal of Botany, 41(5), 486-492.
25. Lo N, Watanabe H, Sugimura M. Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. Proc Biol Sci. 2003 Aug 7;270 Suppl 1(Suppl 1):S69-72. doi: 10.1098/rsbl.2003.0016. PMID: 12952640; PMCID: PMC1698037.
26. Gregory, T.R. Understanding Evolutionary Trees. Evo Edu Outreach 1, 121–137 (2008). https://doi.org/10.1007/s12052-008-0035-x
27. Marcus JH, Novembre J. Visualizing the geography of genetic variants. Bioinformatics. 2017 Feb 15;33(4):594-595. doi: 10.1093/bioinformatics/btw643. PMID: 27742697; PMCID: PMC5408806.
28. Frazer KA, Elnitski L, Church DM, Dubchak I, Hardison RC. Cross-species sequence comparisons: a review of methods and available resources. Genome Res. 2003 Jan;13(1):1-12. doi: 10.1101/gr.222003. PMID: 12529301; PMCID: PMC430969.
29. Battey CJ, Ralph PL, Kern AD. Predicting geographic location from genetic variation with deep neural networks. Elife. 2020 Jun 8;9:e54507. doi: 10.7554/eLife.54507. PMID: 32511092; PMCID: PMC7324158.
30. Xia S, Cosme LV, Lutomiah J, Sang R, Ngangue MF, Rahola N, Ayala D, Powell JR. Genetic structure of the mosquito Aedes aegypti in local forest and domestic habitats in Gabon and Kenya. Par-

asit Vectors. 2020 Aug 13;13(1):417. doi: 10.1186/s13071-020-04278-w. PMID: 32791977; PMCID: PMC7427282.
31. Bitomský M, Mládková P, Pakeman RJ, Duchoslav M. Clade composition of a plant community indicates its phylogenetic diversity. Ecol Evol. 2020 Mar 13;10(8):3747-3757. doi: 10.1002/ece3.6170. PMID: 32313633; PMCID: PMC7160181.
32. Scarpassa VM, Cunha-Machado AS, Saraiva JF: Evidence of new species for malaria vector Anopheles nuneztovari sensu lato in the Brazilian Amazon region. Malar J. 2016; 15: 205.
33. Schäfer M, Lundström JO. Comparison of mosquito (Diptera: Culicidae) fauna characteristics of forested wetlands in Sweden. Annals of the Entomological Society of America. 2001 Jul 1;94(4):576-82.
34. Burke, A., Dahan-Moss, Y., Duncan, F. et al. Anopheles parensis contributes to residual malaria transmission in South Africa. Malar J 18, 257 (2019). https://doi.org/10.1186/s12936-019-2889-5.
35. Dahan-Moss, Y., Hendershot, A., Dhoogra, M. et al. Member species of the Anopheles gambiae complex can be misidentified as Anopheles leesoni. Malar J 19, 89 (2020). https://doi.org/10.1186/s12936-020-03168-x.
36. Braack L, Bornman R, Kruger T, Dahan-Moss Y, Gilbert A, Kaiser M, Oliver SV, Cornel AJ, Lee Y, Norris DE, Coetzee M, Brooke B, Jager C. Malaria Vectors and Vector Surveillance in Limpopo Province (South Africa): 1927 to 2018. Int J Environ Res Public Health. 2020 Jun 9;17(11):4125. doi: 10.3390/ijerph17114125. PMID: 32527067; PMCID: PMC7313001.
37. Dahan YL, Jamesboy E, Lobb L, Letinic B, etc., (2020). MALARIA VECTOR SURVEILLANCE REPORT, SOUTH AFRICA, JANUARY -DECEMBER 2019.
38. Charrel RN, de Lamballerie X, Raoult D. Chikungunya outbreaks - the globalization of vectorborne diseases. N Engl J Med. 2007;356:769–71.
39. Enserink M. Infectious diseases. Chikungunya: no longer a third world disease. Science. 2007;318:1860–1.
40. Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, et al. International dispersal of dengue through air travel: importation risk for Europe. PLoS Negl Trop Dis. 2014;8:e3278.
41. Cornel, A.J., Lee, Y., Almeida, A.P.G. et al. Mosquito community composition in South Africa and some neighboring countries. Parasites Vectors 11, 331 (2018). https://doi.org/10.1186/s13071-018-2824-6

Hosted file

Figure 1.South Africa Map showing different geographic areas where mosquito samples were collected..doc available at https://authorea.com/users/764590/articles/743199-18s-rdna-as-a-tool-for-molecular-characterization-of-mosquito-species-from-various-regions-within-limpopo

Hosted file

Figure 2. T1-T42 mosquito samples' 18S rDNA amplification profiles. Lane MW indicates a 100bp DNA ladde available at https://authorea.com/users/764590/articles/743199-18s-rdna-as-a-tool-for-molecular-characterization-of-mosquito-species-from-various-regions-within-limpopo

Hosted file

Figure 3. DNA bases of Multiple sequence alignment of 18S rDNA targeted in various species of mosquitoe available at https://authorea.com/users/764590/articles/743199-18s-rdna-as-a-tool-for-molecular-characterization-of-mosquito-species-from-various-regions-within-limpopo

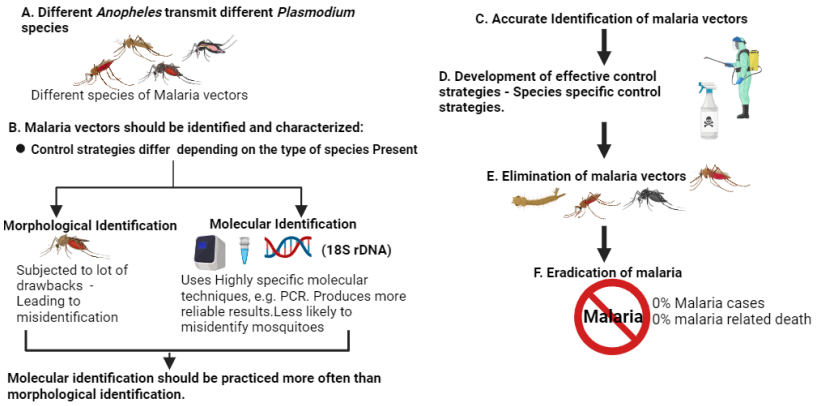
Hosted file

Fihure 4. Phylogenetic reconstruction of the various species of mosquitos using the 18S rDNA gene (900b available at https://authorea.com/users/764590/articles/743199-18s-rdna-as-a-tool-for-molecular-characterization-of-mosquito-species-from-various-regions-within-limpopo

Hosted file

Tables.docx available at https://authorea.com/users/764590/articles/743199-18s-rdna-as-a-

tool-for-molecular-characterization-of-mosquito-species-from-various-regions-within-limpopo



GRAPHICAL ABSTRACT TEXT.

Different *Anopheles* species transmit different malaria parasites, *Plasmodium*, hence they are known as malaria vectors. Malaria vectors should be identified and characterized since treatment/control strategies varies depending on the type of vectors present in the region of interest. However, there are two different ways of identifying malaria vectors namely morphology and molecular identification. Morphological identification is prone to lot of limitations that leads to misidentification of vectors which directly leads to development and assignment of wrong control strategies, which are ineffective. All this information suggests that molecular identification which utilizes highly specific techniques, should be practice more often for precise identification of mosquitoes. Due to its specificity, it is less likely to misidentify malaria vectors, thus it's an ideal basic step for development of specific vector control strategy that would be effective, resulting in elimination of vectors present in a target region. If this approach is implemented, malaria vectors will be managed and the malaria parasites will be eliminated in the Limpopo province by 2030.