A novel mosquito species identification method based on PCR and capillary electrophoresis

Estelle Chabanol[1], Ottavia Romoli[1], Stanislas Talaga[1], Yanouk Epelboin[1], Katy Heu[1], Ghislaine Prévot[2], and Mathilde Gendrin[1]

[1]Institut Pasteur de la Guyane
[2]Université de la Guyane

March 21, 2023

Abstract

In the Anopheles genus, various mosquito species are able to transmit Plasmodium parasites responsible for malaria, while others are non-vectors. In an effort to better understand the biology of Anopheles species and to quantify transmission risk in an area, the identification of mosquito species collected on the field is an essential but problematic task. Morphological identification requires expertise, well-preserved specimens and high-quality equipment, and it does not allow any subsequent verification when samples are later used in a destructive treatment. Moreover, it involves physical manipulations that are not compatible with experiments requiring fast sampling and processing of specimens, hence species identification is often based on DNA sequencing of reference genes or region such as the Internal Transcribed Spacer 2 (ITS2) region of nuclear ribosomal DNA. Sequencing ITS2 for numerous samples is costly, but the design of species-specific PCR primers is not always possible when local species diversity is high. Here, we introduce a molecular technique of species identification based on precise determination of ITS2 length combined with a simple visual observation, the color of mosquito hindleg tip. DNA extracted from field-collected Anopheles mosquitoes was amplified with universal Anopheles ITS2 primers and analyzed with a capillary electrophoresis device, which precisely determines the size of the fragments. We defined windows of amplicon sizes combined with fifth hind tarsus color, which allow to discriminate the major Anopheles species found in our collections. We validated our parameters via Sanger sequencing of the ITS2 amplicons. This method can be particularly useful in situations with a moderate species diversity, i.e. when the number of local species is too high to define species-specific primers but low enough to avoid individual ITS2 sequencing. This tool will be of interest to evaluate local malaria transmission risk and this approach may further be implemented for other mosquito genera.

A novel mosquito species identification method based on PCR and capillary electrophoresis

Estelle Chabanol[1,2,3,4], Ottavia Romoli[1], Stanislas Talaga[1], Yanouk Epelboin[1], Katy Heu[1], Ghislaine Prévot[2,3], Mathilde Gendrin[1,5,*]

1 Microbiota of Insect Vectors Group, Institut Pasteur de la Guyane, Cayenne, French Guiana

2 Tropical Biome and Immunophysiopathology Laboratory, Université de Guyane, Cayenne, French Guiana

3 Université de Lille, CNRS, Inserm, CHU de Lille, Institut Pasteur de Lille, U1019 – UMR 9017 – CIIL – Center for Infection and Immunity of Lille, F-59000, Lille, France

4 École Doctorale 587, Université de Guyane, Cayenne, French Guiana

5 Institut Pasteur, Université de Paris, Department of Insect Vectors, F-75015, Paris, France

* Correspondence: mathilde.gendrin@pasteur.fr

Abstract: In the *Anopheles* genus, various mosquito species are able to transmit *Plasmodium* parasites responsible for malaria, while others are non-vectors. In an effort to better understand the biology of *Anopheles* species and to quantify transmission risk in an area, the identification of mosquito species collected on the field is an essential but problematic task. Morphological identification requires expertise, well-preserved specimens and high-quality equipment, and it does not allow any subsequent verification when samples are later used in a destructive treatment. Moreover, it involves physical manipulations that are not compatible with experiments requiring fast sampling and processing of specimens, hence species identification is often based on DNA sequencing of reference genes or region such as the Internal Transcribed Spacer 2 (ITS2) region of nuclear ribosomal DNA. Sequencing ITS2 for numerous samples is costly, but the design of species-specific PCR primers is not always possible when local species diversity is high. Here, we introduce a molecular technique of species identification based on precise determination of ITS2 length combined with a simple visual observation, the color of mosquito hindleg tip. DNA extracted from field-collected *Anopheles* mosquitoes was amplified with universal *Anopheles* ITS2 primers and analyzed with a capillary electrophoresis device, which precisely determines the size of the fragments. We defined windows of amplicon sizes combined with fifth hind tarsus color, which allow to discriminate the major*Anopheles* species found in our collections. We validated our parameters via Sanger sequencing of the ITS2 amplicons. This method can be particularly useful in situations with a moderate species diversity, i.e. when the number of local species is too high to define species-specific primers but low enough to avoid individual ITS2 sequencing. This tool will be of interest to evaluate local malaria transmission risk and this approach may further be implemented for other mosquito genera.

Keywords: Species Identification; Capillary Electrophoresis; ITS2; Length Polymorphism; *Anopheles* ; French Guiana

Introduction

Several species of mosquitoes are vectors of viruses or parasites causing serious diseases in humans. In particular, some species of*Anopheles* transmit the *Plasmodium* parasite, which causes malaria in humans **[1]** , and thus strongly impacts human health with more than 247 million cases over 84 countries in 2021**[2]** . Despite global effort to reduce the burden of vector-borne diseases, they remain a sanitary and economic threat in the intertropical area and beyond **[3]** . However, among the 3500 species of mosquitoes, only a minority are vectors of pathogens, therefore the identification of species is an important point in surveillance and studies.

In French Guiana, 245 mosquito species are known **[4,5]**including 22 species of *Anopheles* **[6]** , nine of which belong to the *Nyssorhynchus* subgenus whose elevation to genus level is under consideration **[7–9]** . Among them,*Anopheles darlingi* is known to be the main malaria vector in South America **[10,11]** and *Anopheles aquasalis* , although not being incriminated as a vector in French Guiana, is also considered as one of the principal vectors in other neighboring countries on the continent **[10,12]** . Moreover,*Anopheles medialis* , *Anopheles nuneztovari* and*Anopheles oswaldoi* have been found naturally infected by*Plasmodium* in French Guiana **[13]** , as well as*Anopheles braziliensis* and *Anopheles triannulatus* in Brazil **[14,15]** , but the real extent of their involvement in parasite transmission between humans is currently unknown. Given these differences in their ability to transmit malaria, it is thus important to identify the exact species of *Anopheles* mosquitoes when studying and monitoring field population distributions and dynamics.

Originally, mosquito species identification was only based on visual observation of morphological characteristics with the help of dichotomous taxonomic keys **[16–19]** . This method is effective and relatively accessible when morphological differences are substantial, yet it is arduous and requires advanced skills when variations are subtle. The use of additional equipment and the dissection of internal structures are also sometimes mandatory to distinguish between morphologically close species, as for *Culex*mosquitoes, in which meticulous dissection of male genitalia is required**[20,21]** . Moreover, even with properly trained and experienced professionals the task frequently remains a challenge, misidentifications are common and results for later processed samples cannot be checked afterward in case of doubt **[22]** .

Alternative methods have been developed based on molecular biology techniques. *CO1* (Cytochrome C Oxidase Subunit 1), a mitochondrial gene, and ITS2 (Internal Transcribed Spacer 2), a non-coding nuclear sequence located between *5.8S* and *28S* ribosomal RNA genes, are often used for molecular identification of animals. *CO1* or ITS2 DNA barcoding are two of the most prevalent techniques for species identification today, with databases usually already available. Amplification of DNA regions with species-specific sizes via multiplex PCR (Polymerase Chain Reaction) is often used as a substitute to sequencing to discriminate a limited number of species found locally**[23–26]** , notably *Anopheles coluzzii* and*Anopheles gambiae* found in Africa **[27,28]** and other species from the Gambiae Complex **[29,30]** . When a higher number of species are involved, PCR may be combined with a treatment with restriction enzymes in the RFLP (Restriction Fragment Length Polymorphism) technique **[31,32]** .

Other approaches for mosquito species identification are currently being considered. Mosquito protein profiling using MALDI-TOF technology is promising **[33]** , even though it requires a substantial investment in terms of equipment and the databases are still under development **[34]** . Nowadays, increasingly powerful software solutions make it possible to perform morphometry of mosquito wings**[35]** and the use of artificial intelligence may allow for automatic visual identification of mosquitoes **[36]** . The sound and frequency of wing beats is also used to develop identification tools that are meant to be accessible on a simple smartphone**[37]** . Yet the implementation of these tools needs more development and advanced computational knowledge.

In this study, we have developed a new method to distinguish between nine *Anopheles* species from French Guiana based on precise discrimination of their natural ITS2 sequence size polymorphism using capillary electrophoresis, combined with a simple morphological observation, the color of the hindleg tip (fifth hind tarsus, Ta-III_5).

Materials and Methods

Mosquito Sampling

Mosquitoes were sampled in five different locations in French Guiana: La Césarée (CE) [5° 0' 57.32" N, 52° 31' 41.36" W], Le Galion (GA) [4° 47' 0.08" N, 52° 24' 42.74" W], Cacao (CA) [4° 34' 32.59" N, 52° 28' 4.81" W], Blondin (BL) [3° 52' 37.21" N, 51° 48' 48" W] and Trois Palétuviers (TP) [4° 2' 59.46" N, 51° 40' 9.77" W]**(Figure S1)** . While CE, GA and CA are considered as malaria-free areas, BL and TP are located in a transborder region with higher risks of malaria resurgence. In 2017, a malaria outbreak was observed among inhabitants of TP village **[38]** , thus it remains a region of interest considering malaria control programs.

Field missions took place over a four-year period, from July to November of 2018, 2019, 2020 and 2021. Mosquitoes were collected using Mosquito Magnet$^{?}$ traps (WoodStream^{TM}). Mosquito tissues were preserved in 70 to 100 % ethanol and kept at -80 degC upon arrival at the laboratory until further processing.

Hosted file

`image1.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Figure S1: Maps of the geographical situation of French Guiana in South America **(A)** and mosquito sampling sites in French Guiana **(B)** . La Cesaree (CE), Le Galion (GA), Cacao (CA), Blondin (BL) and Trois Paletuviers (TP) are the five collection locations.

DNA Extraction

DNA was extracted from midguts, carcasses or legs. DNA extraction was performed with ZymoBiomics 96 MagBead DNA Kit (Zymo Research) for midguts and with HighPrep Blood & Tissue DNA Kit (MagBio) for carcasses and legs. Samples were mixed using 0.5 mm glass beads in 2 mL screwcap tubes with a Precellys Evolution (Bertin Technologies) bead beater homogenizer. Automatic DNA extraction was performed with the KingFisher Duo Prime system (Thermo Scientific). DNA was eluted in ZymoBiomics DNAse/RNAse

Free Water from the kit for midguts and MB Elution Buffer supplied in the kit for carcasses and legs. Extracted DNA was kept at -80 degC until further utilization.

ITS2 PCR Amplification

PCR amplification of the ITS2 region was performed with Hot FirePol DNA Polymerase Kit (Solis Biodyne). Total PCR mix volume per sample was 50 µL with 48 µL premix and 2 µL DNA. The premix contained 36.8 µL H2O, 5 µL 10X Buffer B1, 3 µL 25 mM MgCl2, 1 µL 10 mM Deoxynucleotide (dNTP) Solution Mix (New England BioLabs), 1 µL 10 nM Forward ITS2 Primer, 1 µL 10 nM Reverse ITS2 Primer and 0.2 µL (5 U/µL) Hot FirePol DNA Polymerase. The DNA was either undiluted for extracted midguts and legs or diluted at 1/10 in DNAse/RNAse free water for extracted carcasses.*Anopheles* ITS2 primer sequences are 5'-TGTGAACTGCAGGACACAT-3' (Forward) and 5'-TATGCTTAAATTCAGGGGGTAG-3' (Reverse)**[31,39]** . PCR reactions were done on a SimpliAmp Thermal Cycler (Applied Biosystems) using the cycle: 95 °C x 15 min, [95 °C x 30 s, 56 °C x 45 s, 72 °C x 40 s] x 35 cycles, 72 °C x 10 min, 4 °C x [?]. Amplified DNA was either used directly after or stored at 4 degC or -20 degC until use.

Agarose Gel

1 % agarose gel was made with 1 g UltraPure Agarose (Invitrogen), 100 mL 0.5X TBE Buffer diluted from TBE Buffer 10X (Biosolve) and 10 µL Midori Green Advance (Nippon Genetics). 8 µL of each sample was mixed with 2 µL DNA Gel Loading Dye 6X (Thermo Scientific) and 10 µL of 100 bp DNA Ladder Ready to Load size marker (Solis Biodyne) was used as reference. Migration took 45 min at 120 V with Owl EC300XL Compact Power Supply (Thermo Scientific). Gel visualization and analysis were done using Alliance Q9 Advanced imaging system and software (Uvitec Cambridge).

Capillary Electrophoresis

Capillary electrophoresis was carried out in a QIAxcel Advanced System (Qiagen). QIAxcel DNA High Resolution Kit was chosen for higher precision and samples were run using OM800 method, QX Alignment Marker 15 bp – 1 kb and QX DNA Size Marker 50 – 800 bp. Results were treated with QIAxcel ScreenGel Software and manually checked to ensure that detected peaks were real and that noise signals were eliminated. Reports were generated and ITS2 amplicon lengths were recovered for further analysis.

Sequencing and Alignment

PCR products were kept in PCR plates sealed with adhesive film or caps and sent oversea via French post services to be sequenced. Sanger sequencing was handled by Microsynth facility in Lyon, France. Forward and reverse sequencing were realized with the same set of primers than for the PCR reaction. Fragment sequences were aligned using BLAST (NCBI, NIH) against the standard databases (nucleotide collection nr/nt)**[40,41]** and *Anopheles* species identifications were retrieved for forward and reverse sequences individually and integrated in our data and metadata files.

Reference sequences of ITS2 region of 14 *Anopheles* species found in French Guiana are available online on the NIH website (PopSet: 870902931) **[31]** .

Analysis and Graphics

Method development and routine was mainly done in Excel tables and figures were generated with R Studio (2022.02.3+492 version) using the Tidyverse packages.

Results of Table S2 were analyzed by comparing the interval size of each triplicate on day 1 and on day 2 via Wilcoxon rank sum test and Kruskal-Wallis rank sum test on R Studio.

Results

ITS2 region naturally differs in size between *Anopheles* species of French Guiana **(Figure 1A)** , ranging from 400 bp to 680 bp for*Anopheles eiseni* and *Anopheles minor* respectively**[31]** . As ITS2 size generally differs for more than 7 bp among species, we hypothesized this length polymorphism may allow species

identification after PCR amplification and electrophoresis, so that any additional manipulation or sequencing is not required. Some species however share identical or similar ITS2 amplicon size (< 3 bp). For instance, *An. braziliensis* and *An. oswaldoi* both have a 488 bp-long ITS2 sequence, and the species identification would be impossible using exclusively the size information. We decided to collect a simple morphological observation upon mosquito sampling in the field, the color of their three last hind tarsi (Ta-$III_{3,4,5}$), in particular the fifth hind tarsus (Ta-III_5, at the tip of the hindleg) **(Figure 1A)** . This morphological data does not require advanced skills in taxonomy but is enough to discriminate some species with similar ITS2 length. *Anopheles braziliensis* and *An. darlingi* have their three last hind tarsi totally white while the other species like*An. oswaldoi* have a dark basal band on their fifth hind tarsus. Additionally, *Anopheles* outside of the *Nyssorhynchus*subgenus have a mix of white and black on their three last hind tarsi.

While differences in amplicon size are visible after PCR amplification and migration on an agarose gel, the analysis of gel images remains approximative and does not allow precise determination of fragment sizes**(Figure 1B)** . With an agarose gel, fragment sizes are determined using the size marker as a reference, yet gel homogeneity and migration speed are not strictly controlled and slight differences in migration between wells cannot be corrected in the absence of internal controls. Capillary electrophoresis migration is another technique that allows automatization of the process and standardization of the fragment size detection, using internal controls of defined sizes as references**(Figure 1C)** . The precision is up to 3 bp for the most precise migration settings on fragments shorter than 500 bp. Additionally, the electropherogram allows to verify the quality of the fragment amplification **(Figure 1D)** . We decided to set up a method to identify our *Anopheles* samples based on capillary electrophoresis analysis combined with observation of fifth hind tarsus color.

Hosted file

`image2.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Figure 1: The method is based on natural variation of ITS2 sequence size between *Anopheles* species coupled with visual observation of mosquito fifth hind tarsus color **(A)** . Species-specific amplicon sizes are visible on an agarose gel**(B)** but more precisely measured with capillary electrophoresis including internal reference size controls **(C)** . Additionally, capillary electrophoresis allows quality check of fragment amplification and migration **(D)** .

We collected *Anopheles* mosquitoes from five different locations in French Guiana **(Figure S1)** over a four-year period from 2018 to 2021. The color of the fifth hind tarsus was noted for each mosquito sample at the time of capture. Back in the laboratory, we extracted the DNA of each individual, amplified the ITS2 sequence and ran capillary electrophoresis **(Figure 2)** .

Our method development was divided in three phases based on different mosquito collections, with a first phase of establishment of parameters (step 1) and two phases of adjustment of these parameters (steps 2 and 3) **(Figure 2)** . During step 1, we sequenced ITS2 in 167*Anopheles* samples, identified species by BLAST and ran capillary electrophoresis to link mosquito species with observed ITS2 amplicon size. During steps 2 and 3, we collected 163 and 73 mosquitoes respectively, we tested the method identification process, coupling ITS2 size and fifth hind tarsus color to determine the *Anopheles*species, and we verified the result by sequencing again 100 % of our ITS2 amplicons. This enabled us to identify errors, adjust size intervals and add new species to the method. Finally, we applied this method routinely in our laboratory on a total of 372 samples and more than 99 % could be identified without the need for sequencing**(Figure 2)** . Only 2/372 samples needed to be sequenced because of uncertain identification.

Hosted file

`image3.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Figure 2: Organizational chart of the methodology. Sequencing of samples during method development

allowed definition and adjustment of species-specific size intervals. In routine, sequencing is optional and only applied for uncertain species identification.

The ITS2 signal detected after capillary electrophoresis generally consisted of a single and well-distinct peak on the electropherogram**(Figure 1C)** with the exception of *An. nuneztovari*mosquitoes for which 80 % samples (37/46) had a profile consisting of multiple peaks, including a main one around 497 bp. This pattern seemed specific to *An. nuneztovari* . Additionally, four non-*Anopheles* mosquito samples were included within the first sample set (three *Psorophora* and one *Culex*) and gave rise to multiple-peak profiles around 350 bp, which were clearly distinguishable from any *Anopheles* mosquitoes **(Figure 3A)** .

ITS2 size intervals were defined at step 1 simply by taking the minimum and maximum size detected among all the samples of a species as minimum and maximum of the interval for this species. Six *Anopheles*species were detected among the 167 samples of step 1, namely *An. braziliensis* , *An. darlingi* , *Anopheles ininii* , *An. medialis* , *An. nuneztovari* and *An. triannulatus***(Figure 3A)** . Two species were added at step 2, *An. aquasalis* and *Anopheles peryassui* , and one species was added at step 3, *An. oswaldoi* **(Figure 3A)** . Some species intervals were also slightly modified at steps 2 and 3 according to new data sets, with 0-to-6-bp adjustments on each side of the intervals **(Table S1)** . At the end, final amplicon size intervals were defined for nine*Anopheles* species **(Figure 3B)** .

Altogether, the capillary electrophoresis migration is more accurate for fragments smaller than 500 bp due to the specificity of the machine. The shortest intervals were set for *An. medialis* (expected size: 413 bp – median observed size: 417 bp), *An. aquasalis* (expected size: 485 bp – median observed size: 486 bp) and *An. ininii*(expected size: 495 bp – median observed size: 490 bp), they have an amplitude of 7 bp, 7 bp and 8 bp respectively **(Table S1)** . While these short intervals can be explained by the low number of samples for these species, a relatively short interval was also observed for*An. braziliensis* (expected size: 488 bp – median observed size: 488 bp), one of the most abundant species: its total amplitude is 15bp but more than 75 % samples fall in a 6 bp interval (486 - 492 bp). For fragments larger than 500 bp, greater variation between samples of the same species appeared, with a size amplitude of 28 and 32 bp for*An. darlingi* (expected size: 546 bp – median observed size: 564 bp) and *An. triannulatus* (expected size: 564 bp – median observed size: 581 bp) respectively. However, the standard deviations for *An. darlingi* and *An. triannulatus* are only 6 and 7 bp respectively, and more than 75 % fragments still fall into a 15 bp interval for *An. darlingi* (555 - 570 bp) and *An. triannulatus* (573 - 588 bp).

We checked that intervals were actually explained by the precision of the capillary electrophoresis rather than to sequence variation in each species by amplifying and running again *An. darlingi* and*An. triannulatus* DNA samples from both extremes of their respective intervals. We found no difference in size when these samples were processed in the same PCR and run in the same capillary electrophoresis batch **(Table S2)** . An overnight -20 °C storage slightly increased interval amplitude by 140 % between replicated samples when compared to samples processed straight after PCR (**Table S2** – $p = 0.0023$ – $W = 47.5$, Wilcoxon rank sum test), hence we suspect that variations may be due to the storage of some samples at 4 °C or -20 °C for one or several days between PCR and migration.

Some overlaps appeared between the size intervals of several*Anopheles* species **(Figure 3B)** . The intervals of*An. darlingi* and *An. triannulatus* overlapped between 563 and 576 bp, but the leg color distinction made it possible to discriminate these species. Similarly, several species-specific intervals overlapped between 479 and 494 bp, for *An. aquasalis* ,*An. braziliensis* , *An. ininii* , *An. nuneztovari* and*An. oswaldoi* . *Anopheles braziliensis* can be distinguished from the rest by their fifth hind tarsus color and *An. nuneztovari* mosquitoes by their tendency to multiple peak profiles.

Hosted file

`image4.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Figure 3: Individual mosquito ITS2 amplicon lengths from the three development steps **(A)** and definitive

size intervals defined for eight *Anopheles* species **(B)** . Each dot represents a mosquito with its fifth hind tarsus color information recorded at time of capture (white or black dots, sometimes erroneous) or missing (grey dots). Species were determined by sequencing during development steps, allowing to link *Anopheles* species and ITS2 size.

When determination is not possible with our method, samples would still need to be sequenced, but the remaining species (*An. aquasalis* ,*An. ininii* and *An. oswaldoi*) are only found sporadically.*Anopheles braziliensis* , *An. darlingi* , *An. nuneztovari* and *An. triannulatus* represented 90 % of the samples **(Figure 4A)** , including 44 % of *An. darlingi*after all three development steps and 68 % in total with routine samples. In comparison, we collected only 1, 2, 3, 4 and 6 samples of*An. peryassui* , *An. ininii* , *An. oswaldoi* , *An. aquasalis* and *An. medialis* respectively.

Our method allowed us to identify correctly more than 80 % of samples during development phases (steps 2 and 3), and even over 99 % when used in routine **(Figure 4B)** . During steps 2 and 3 of development, 1.2 and 2.7 % of samples were assigned to the wrong species by the method, and the errors were detected after validation by sequencing. The misidentifications were due to wrong fifth hind tarsus color data (2/163*Anopheles* samples) at step 2 and to requirement for adjustment of*An. aquasalis* interval (2/73) at step 3. At these steps, we were not able to determine the species for 12 and 16 % of samples respectively, due to three factors: missing fifth hind tarsus color information; wrong fifth hind tarsus color information that led to contradictory results; interval overlaps. In routine, the remaining uncertainties were very low (0.54 %; 2/372) due to rare missing fifth hind tarsus color and to interval overlaps.

We used sequencing as a classical way of species identification, at step 1 to establish the link between species and ITS2 size, and at steps 2 and 3 to verify and optimize our method. Sequencing turned out to be a source of errors too with 3.0 and 3.7 % of misidentification at steps 1 and 2 respectively **(Figure 4C)** . In addition, 7.8, 9.2 and 1.4 % samples led to different species identification when sequenced from forward and reverse primers at steps 1, 2 and 3 respectively. This may be due to cross contamination occurring during processing, shipping and/or sequencing.

Hosted file

`image5.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Figure 4: Distribution of *Anopheles* species identified during the three development steps and in routine **(A)** . Percentages of method **(B)** and sequencing **(C)** results leading to correct identification, misidentification or uncertainties. Missing or wrong fifth hind tarsus color information are the main causes of indetermination in the method and wrong fifth hind tarsus color information led to wrong result in some cases.

Discussion

In this paper, we describe a method of mosquito species identification based on a traditional endpoint PCR and a precise amplicon size detection, avoiding the need for sequencing. We found that ITS2 sequence size differences between species are detectable by capillary electrophoresis. This enabled us to assign species-specific profiles for nine *Anopheles* species (seven species from the*Nyssorhynchus* subgenus and two from the *Anopheles*subgenus) present in French Guiana. Identification is based on ITS2 amplicon size information combined with a simple morphological observation, the color of the fifth hind tarsus. This additional piece of data is quick and easy to acquire, by eye or under a stereomicroscope, when collecting mosquitoes in the field and does not require any advanced knowledge in taxonomy.

This species identification method has the advantage of being simple and relatively quick to set up, and can be used in routine in a laboratory avoiding the requirement for sequencing. During the three steps of method development, we were not able to sequence our samples on-site and we had to send samples in 96-well plates between French Guiana and France for sequencing. For 7.8, 9.2 and 1.4 % of samples, at steps 1, 2 and 3 of method development respectively, it was not possible to obtain definitive results as forward and reverse sequencing outcomes were different. Moreover, we observed that 3.0 and 3.7 % of sample sequencing

results, at steps 1 and 2 of method development respectively, were erroneous. We were able to detect these errors because of contradictory results between sequencing and our method, whether because of fragment size not matching the putative species or because of conflicting leg color. Without the implementation of the method, these errors would have gone unnoticed. In case of doubt after sequencing, we amplified the ITS2 region again and sent the sample for sequencing to check the initial result. Hence, sequencing can be a source of error, especially when samples need to be shipped, as cross-contaminations may occur during plate processing, shipment and sequencing. In our case, plates were sealed and packaged carefully but we noticed that plate caps that have undergone temperature changes during PCR and storage in the freezer are slightly easier to open. Moreover, we used adhesive PCR plate seals for one shipment of two plates (one of step 1 and one of step 2), which resulted in more misidentifications and uncertain results. During subsequent shipments (steps 2 and 3), we used plate caps that were replaced with new ones just before shipping and observed a relatively low error rate, yet did not investigate this further.

In our method, missing or erroneous leg color is the major source of problem. It rarely led to incorrect identification (1.2 % samples species at step 2), but more often led to uncertainties (10, 12 and 0.27 % at steps 2, 3 and Routine, respectively), which represents the majority of method uncertainties (12, 16 and 0.54 % samples, at steps 2, 3 and routine, respectively). Wrong leg color specifically led to 4.3 and 8.2 % indeterminations at steps 2 and 3 respectively, indicating that efforts to ensure correct annotation of this information at the time of capture can significantly lead to improvement in determination rates. A simple double check by a colleague may be the key, as these data are usually collected in a repetitive way, sometimes at night, in a non-usual environment. Nonetheless, this problem cannot be completely solved, as some mosquito samples lack both hind legs. One could decide to exclude them from the study, yet we were able to identify the species of 54/94 mosquitoes (steps 1, 2 and 3) with missing fifth hind tarsus color information, and of 16/17 in routine. The fifth hind tarsus color element is therefore a piece of data that should not be neglected, but even if it is missing, identification remains possible in many cases.

The second key point for the reliability of our method is, obviously, the proper use of the capillary electrophoresis device and its components. The channels of the capillary electrophoresis cartridge are prone to clogging if they are used incorrectly and the results can be erroneous in these cases. The only recommendation on this point is to follow the supplier's instructions carefully and check results individually, in order to detect any aberrant results. In case of aberrant results, we either excluded the problematic channel or replaced the cartridge.

While the current method has been developed on *Anopheles* species from French Guiana, this approach may be extended to species identification of individual fieldwork specimens in any taxon. Widening the scale besides *Anopheles* (*Nyssorhynchus*), our method clearly discriminated our *Anopheles* samples from four mosquitoes of the *Culex* and *Psorophora* genera during step 1, and we expect development steps to require much less specimens if the species of interest are more distant from each other. More testing would be needed to determine whether all mosquito genera have a significantly different profile from one another and whether, within these genera, species identification is possible. Our tool may then facilitate identification of a broad range of mosquitoes, and be particularly advantageous during epidemics in order to target species that are potential vectors of arboviruses or parasites. It could be compatible with sampling by non-experts combined with photographs during collaborative work or citizen science projects. Finally, when presence/absence information are sufficient, pools of different samples sharing the same morphological trait (in our case, the color of Ta-III_5) could be processed.

To specify the interest of our method with currently available methods, several aspects can be considered: the possibility to go back and check results, the level of local diversity and the cost-effectiveness. Our method allows to easily check results as DNA extracts can be stored for a long time. They can be used for a second similar analysis by PCR and capillary electrophoresis as well as for sequencing of ITS2 or any other sequence, which may notably apply to population genetics studies. The other methods of molecular biology have similar advantages, while morphological observations and audio recordings may not, unless samples have been properly stored.

In terms of diversity, we detected nine different *Anopheles*species, including four dominant ones. With a higher local diversity, too many overlaps between intervals would reduce the efficiency of the method with the current experimental conditions. The latter can be improved, for instance by running capillary electrophoresis straight after PCR, which could allow to deal with a slightly higher diversity. Sequencing would remain the best option in case of high species diversity. With a lower diversity of only two or three species, multiplex PCR with species-specific primers would remain the simplest and cheapest method. Hence, our species identification method is a reliable alternative for locally moderate species diversity.

Considering costs, capillary electrophoresis requires an initial equipment investment, yet the processing cost of each sample is much lower with the proposed method than with sequencing. When considering the initial cost of the device and of capillary electrophoresis reagents on one hand, shipping and sequencing cost of samples on the other hand (and ignoring the cost of labour and our interest in the other applications of the device), our investment on the capillary electrophoresis apparatus would be repaid after 4,000 to 5,000 samples. Beyond that, more than 5 \euro per sample can be saved. Hence, episodic needs may rather be addressed with sequencing, while our approach seems reasonable for a laboratory requiring steady species identification. Alternatively, even with very low consumable cost per sample [**42**] , the MALDI-TOF approach requires a 10-fold higher investment and maintenance budget than capillary electrophoresis. It would thus require a much higher number of samples to be more cost-effective than our method.

In sum, we introduce PCR and capillary electrophoresis combined with simple morphological observation as a convenient method to discriminate species of different field-collected samples. We specifically deployed it to identify nine *Anopheles* species found in French Guiana, mainly from the *Nyssorhynchus* subgenus. After deployment, this method allows to save time and money and to keep control over experimental schedule.

Data Accessibility

Sequencing data will be submitted to a public online database prior to publication.

Competing Interests

The authors declare no conflict of interest.

Author Contributions (CRediT) [Contribution Roles Taxonomy]

Conceptualization: E.C., G.P. and M.G.

Data Curation: E.C.

Formal Analysis: E.C. and M.G.

Funding Acquisition: G.P. and M.G.

Investigation: E.C., O.R., S.T., Y.E. and K.H.

Methodology: E.C. and M.G.

Project Administration: E.C., G.P. and M.G.

Supervision: G.P. and M.G.

Validation: E.C. and M.G.

Visualization: E.C. and M.G.

Writing – Original Draft Preparation: E.C. and M.G.

Writing – Review & Editing: E.C., O.R., S.T., Y.E., G.P. and M.G.

All authors have read and agreed to the published version of the manuscript.

Acknowledgements

We thank Romuald Carinci for training on Mosquito Magnet[?] traps and for his support during some field missions. We thank the owners of Val'Ranch (La Cesaree, Macouria) for allowing us to set traps and capture mosquitoes on their property, the inhabitants of Cacao, Blondin and Trois Paletuviers for their warm welcome during our field missions and the management and staff of Saint-Georges medical center for allowing us to work in their facility.

Photo Credit: Romuald Carinci

Funding

This research was funded by French Government's Agence Nationale de la Recherche for Laboratoires d'Excellence for Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-0062-IBEID) to M.G., Agence Nationale de la Recherche for Jeunes Chercheuses Jeunes Chercheurs for Mosquito Microbiota (ANR-18-CE15-0007-MosMi) to M.G., Agence Nationale de la Recherche for Laboratoires d'Excellence (ANR-10-LABX-2501) to G.P., Ecole Doctorale 587 Universite de Guyane 3-year PhD Studentship to E.C. and a Fondation pour la Recherche Medicale 4thyear PhD Studentship (FDT202204015140) to E.C.

References

1. Phillips, M.A.; Burrows, J.N.; Manyando, C.; van Huijsduijnen, R.H.; Van Voorhis, W.C.; Wells, T.N.C. Malaria. *Nat Rev Dis Primers***2017** , *3* , 17050, doi:10.1038/nrdp.2017.50.

2. World Malaria Report 2022. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

3. Global Vector Control Response 2017–2030. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

4. Talaga, S.; Duchemin, J.-B.; Girod, R.; Dusfour, I. The *Culex*Mosquitoes (Diptera: Culicidae) of French Guiana: A Comprehensive Review With the Description of Three New Species. *Journal of Medical Entomology* **2020** , tjaa205, doi:10.1093/jme/tjaa205.

5. Talaga, S.; Gendrin, M. Three New Species of Culex (Melanoconion) (Diptera: Culicidae) from French Guiana Based on Morphological and Molecular Data. *Zootaxa* **2022** , *5205* , 177–189, doi:10.11646/zootaxa.5205.2.5.

6. Talaga, S.; Dejean, A.; Carinci, R.; Gaborit, P.; Dusfour, I.; Girod, R. Updated Checklist of the Mosquitoes (Diptera: Culicidae) of French Guiana. *J Med Entomol* **2015** , *52* , 770–782, doi:10.1093/jme/tjv109.

7. Foster, P.G.; de Oliveira, T.M.P.; Bergo, E.S.; Conn, J.E.; Sant'Ana, D.C.; Nagaki, S.S.; Nihei, S.; Lamas, C.E.; Gonzalez, C.; Moreira, C.C.; et al. Phylogeny of Anophelinae Using Mitochondrial Protein Coding Genes. *R. Soc. open sci.* **2017** , *4* , 170758, doi:10.1098/rsos.170758.

8. Harbach, R.E. An Anopheles by Any Other Name . . . ? *Journal of Medical Entomology* **2018** , doi:10.1093/jme/tjy108.

9. Marinotti, O. What Is in a Name? Anopheles Darlingi versus Nyssorhynchus Darlingi. *Trends in Parasitology* **2021** ,*37* , 856–858, doi:10.1016/j.pt.2021.06.003.

10. Sinka, M.E.; Bangs, M.J.; Manguin, S.; Rubio-Palis, Y.; Chareonviriyaphap, T.; Coetzee, M.; Mbogo, C.M.; Hemingway, J.; Patil, A.P.; Temperley, W.H.; et al. A Global Map of Dominant Malaria Vectors.*Parasites Vectors* **2012** , *5* , 69, doi:10.1186/1756-3305-5-69.

11. Hiwat, H.; Bretas, G. Ecology of Anopheles Darlingi Root with Respect to Vector Importance: A Review. *Parasites Vectors***2011** , *4* , 177, doi:10.1186/1756-3305-4-177.

12. Pimenta, P.F.; Orfano, A.S.; Bahia, A.C.; Duarte, A.P.; Rios-Velasquez, C.M.; Melo, F.F.; Pessoa, F.A.; Oliveira, G.A.; Campos, K.M.; Villegas, L.M.; et al. An Overview of Malaria Transmission from the

Perspective of Amazon Anopheles Vectors. *Mem. Inst. Oswaldo Cruz* **2015** , *110* , 23–47, doi:10.1590/0074-02760140266.

13. Dusfour, I.; Issaly, J.; Carinci, R.; Gaborit, P.; Girod, R. Incrimination of Anopheles (Anopheles) Intermedius Peryassu, An. (Nyssorhynchus) Nuneztovari Gabaldon, An. (Nys.) Oswaldoi Peryassu as Natural Vectors of Plasmodium Falciparum in French Guiana. *Mem. Inst. Oswaldo Cruz* **2012** , *107* , 429–432, doi:10.1590/S0074-02762012000300021.

14. Duarte, A.M.R.; Pereira, D.M.; de Paula, M.B.; Fernandes, A.; Urbinatti, P.R.; Ribeiro, A.F.; Mello, M.H.S.; Matos, M.O.; Mucci, L.F.; Fernandes, L.N.; et al. Natural Infection in Anopheline Species and Its Implications for Autochthonous Malaria in the Atlantic Forest in Brazil.*Parasites Vectors* **2013** , *6* , 58, doi:10.1186/1756-3305-6-58.

15. Daniel-Ribeiro, C.T.; Lourenco-De-Oliveira, R.; Teva, A.; Deane, L.M.; de Oliveira-Ferreira, J. Natural Malaria Infections in Anophelines in Rondonia State, Brazilian Amazon. *The American Journal of Tropical Medicine and Hygiene* **1990** , *43* , 6–10, doi:10.4269/ajtmh.1990.43.1.TM0430010006.

16. Sallum, M.A.M.; Obando, R.G.; Carrejo, N.; Wilkerson, R.C. Identification Keys to the Anopheles Mosquitoes of South America (Diptera: Culicidae). IV. Adult Females. *Parasites Vectors***2020** , *13* , 584, doi:10.1186/s13071-020-04301-0.

17. Coetzee, M. Key to the Females of Afrotropical Anopheles Mosquitoes (Diptera: Culicidae). *Malar J* **2020** , *19* , 70, doi:10.1186/s12936-020-3144-9.

18. Gunathilaka, N. Illustrated Key to the Adult Female Anopheles (Diptera: Culicidae) Mosquitoes of Sri Lanka. *Appl Entomol Zool***2017** , *52* , 69–77, doi:10.1007/s13355-016-0455-y.

19. Rodrigues de Sa, I.L.; Hutchings, R.S.G.; Hutchings, R.W.; Sallum, M.A.M. Revision of the Educator Group of *Culex* (*Melanoconion*) (Diptera, Culicidae). *Journal of Medical Entomology* **2022** , *59* , 1252–1290, doi:10.1093/jme/tjac051.

20. Sallum, M.A.; Forattini, O.P. Revision of the Spissipes Section of Culex (Melanoconion) (Diptera:Culicidae). *J Am Mosq Control Assoc***1996** , *12* , 517–600.

21. Sa, I.L.R. de; Hutchings, R.S.G.; Hutchings, R.W.; Sallum, M.A.M. Revision of the Atratus Group of Culex (Melanoconion) (Diptera: Culicidae). *Parasites Vectors* **2020** , *13* , 269, doi:10.1186/s13071-020-3982-x.

22. Rahola, N.; Gunay, F.; Ozturk, M.; Alten, B.; Aqeehal, H.A.; Saadawi, W.K.; Shaibi, T.; Kavran, M.; Petrić, D.; Mitrović, J.; et al. Assessment of Expertise in Morphological Identification of Mosquito Species (Diptera, Culicidae) Using Photomicrographs. *Parasite***2022** , *29* , 45, doi:10.1051/parasite/2022045.

23. Bang, W.J.; Won, M.H.; Cho, S.T.; Ryu, J.; Choi, K.S. A Multiplex PCR Assay for Six Aedini Species, Including Aedes Albopictus.*Parasites Vectors* **2021** , *14* , 380, doi:10.1186/s13071-021-04871-7.

24. Bang, W.J.; Kim, H.C.; Ryu, J.; Lee, H.S.; Lee, S.Y.; Kim, M.S.; Chong, S.T.; Klein, T.A.; Choi, K.S. Multiplex PCR Assay for the Identification of Eight Anopheles Species Belonging to the Hyrcanus, Barbirostris and Lindesayi Groups. *Malar J* **2021** ,*20* , 287, doi:10.1186/s12936-021-03808-w.

25. Brosseau, L.; Udom, C.; Sukkanon, C.; Chareonviriyaphap, T.; Bangs, M.J.; Saeung, A.; Manguin, S. A Multiplex PCR Assay for the Identification of Five Species of the Anopheles Barbirostris Complex in Thailand. *Parasites Vectors* **2019** , *12* , 223, doi:10.1186/s13071-019-3494-8.

26. Wilai, P.; Namgay, R.; Made Ali, R.S.; Saingamsook, J.; Saeung, A.; Junkum, A.; Walton, C.; Harbach, R.E.; Somboon, P. A Multiplex PCR Based on Mitochondrial COI Sequences for Identification of Members of the Anopheles Barbirostris Complex (Diptera: Culicidae) in Thailand and Other Countries in the Region. *Insects* **2020** , *11* , 409, doi:10.3390/insects11070409.

27. Niang, A.; Epopa, P.S.; Sawadogo, S.P.; Maïga, H.; Konaté, L.; Faye, O.; Dabiré, R.K.; Tripet, F.; Diabaté, A. Does Extreme Asymmetric Dominance Promote Hybridization between Anopheles Coluzzii and Anopheles Gambiae s.s. in Seasonal Malaria Mosquito Communities of West Africa?*Parasites Vectors* **2015** , *8* , 586, doi:10.1186/s13071-015-1190-x.

28. Santolamazza, F.; Mancini, E.; Simard, F.; Qi, Y.; Tu, Z.; della Torre, A. Insertion Polymorphisms of SINE200 Retrotransposons within Speciation Islands of Anopheles Gambiae Molecular Forms. *Malar J***2008** , *7* , 163, doi:10.1186/1475-2875-7-163.

29. Wilkins, E.E.; Howell, P.I.; Benedict, M.Q. IMP PCR Primers Detect Single Nucleotide Polymorphisms for Anopheles Gambiae Species Identification, Mopti and Savanna RDNA Types, and Resistance to Dieldrin in Anopheles Arabiensis. *Malar J* **2006** , *5* , 125, doi:10.1186/1475-2875-5-125.

30. Scott, J.A.; Brogdon, W.G.; Collins, F.H. Identification of Single Specimens of the Anopheles Gambiae Complex by the Polymerase Chain Reaction. *The American Journal of Tropical Medicine and Hygiene***1993** , *49* , 520–529, doi:10.4269/ajtmh.1993.49.520.

31. Vezenegho, S.B.; Issaly, J.; Carinci, R.; Gaborit, P.; Girod, R.; Dusfour, I.; Briolant, S. Discrimination of 15 Amazonian Anopheline Mosquito Species by Polymerase Chain Reaction—Restriction Fragment Length Polymorphism. *Journal of Medical Entomology* **2022** ,*59* , 1060–1064, doi:10.1093/jme/tjac008.

32. Fanello, C.; Santolamazza, F.; Della Torre, A. Simultaneous Identification of Species and Molecular Forms of the *Anopheles Gambiae* Complex by PCR-RFLP: Identification of An. Gambiae Species and Forms. *Medical and Veterinary Entomology* **2002** ,*16* , 461–464, doi:10.1046/j.1365-2915.2002.00393.x.

33. Sánchez-Juanes, F.; Calvo Sánchez, N.; Belhassen García, M.; Vieira Lista, C.; Román, R.M.; Álamo Sanz, R.; Muro Álvarez, A.; Muñoz Bellido, J.L. Applications of MALDI-TOF Mass Spectrometry to the Identification of Parasites and Arthropod Vectors of Human Diseases.*Microorganisms* **2022** , *10* , 2300, doi:10.3390/microorganisms10112300.

34. Costa, M.M.; Guidez, A.; Briolant, S.; Talaga, S.; Issaly, J.; Naroua, H.; Carinci, R.; Gaborit, P.; Lavergne, A.; Dusfour, I.; et al. Identification of Neotropical Culex Mosquitoes by MALDI-TOF MS Profiling. *TropicalMed* **2023** , *8* , 168, doi:10.3390/tropicalmed8030168.

35. Lorenz, C.; Almeida, F.; Almeida-Lopes, F.; Louise, C.; Pereira, S.N.; Petersen, V.; Vidal, P.O.; Virginio, F.; Suesdek, L. Geometric Morphometrics in Mosquitoes: What Has Been Measured? *Infection, Genetics and Evolution* **2017** , *54* , 205–215, doi:10.1016/j.meegid.2017.06.029.

36. Goodwin, A.; Padmanabhan, S.; Hira, S.; Glancey, M.; Slinowsky, M.; Immidisetti, R.; Scavo, L.; Brey, J.; Sai Sudhakar, B.M.M.; Ford, T.; et al. Mosquito Species Identification Using Convolutional Neural Networks with a Multitiered Ensemble Model for Novel Species Detection. *Sci Rep* **2021** , *11* , 13656, doi:10.1038/s41598-021-92891-9.

37. Mukundarajan, H.; Hol, F.J.H.; Castillo, E.A.; Newby, C.; Prakash, M. Using Mobile Phones as Acoustic Sensors for High-Throughput Mosquito Surveillance. *eLife* **2017** , *6* , e27854, doi:10.7554/eLife.27854.

38. Mosnier, E.; Dusfour, I.; Lacour, G.; Saldanha, R.; Guidez, A.; Gomes, M.S.; Sanna, A.; Epelboin, Y.; Restrepo, J.; Davy, D.; et al. Resurgence Risk for Malaria, and the Characterization of a Recent Outbreak in an Amazonian Border Area between French Guiana and Brazil.*BMC Infect Dis* **2020** , *20* , 373, doi:10.1186/s12879-020-05086-4.

39. Saul, A.; Beebe, N.W. Discrimination of All Members of the Anopheles Punctulatus Complex by Polymerase Chain Reaction-Restriction Fragment Length Polymorphism Analysis. *The American Journal of Tropical Medicine and Hygiene* **1995** , *53* , 478–481, doi:10.4269/ajtmh.1995.53.478.

40. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *Journal of Molecular Biology***1990** , *215* , 403–410, doi:10.1016/S0022-2836(05)80360-2.

41. Sayers, E.W.; Bolton, E.E.; Brister, J.R.; Canese, K.; Chan, J.; Comeau, D.C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **2022** , *50* , D20–D26, doi:10.1093/nar/gkab1112.

42. Croxatto, A.; Prod'hom, G.; Greub, G. Applications of MALDI-TOF Mass Spectrometry in Clinical Diagnostic Microbiology. *FEMS Microbiol Rev* **2012** , *36* , 380–407, doi:10.1111/j.1574-6976.2011.00298.x.

Supplementary Data

Hosted file

`image6.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Table S1: ITS2 size intervals and modifications of the intervals at development steps 1, 2 and 3 compared to the reference sequence lengths for each *Anopheles* species.

Hosted file

`image7.emf` available at https://authorea.com/users/597968/articles/630767-a-novel-mosquito-species-identification-method-based-on-pcr-and-capillary-electrophoresis

Table S2: *Anopheles darlingi* and *An. triannulatus*DNA samples from both extremes of their respective intervals during steps 1 – 3 ("expected length") were amplified in two separate PCR and capillary electrophoresis migration was performed twice for each PCR product on two different days. There is no significant effect of the expected length on the averaged amplicon size between PCRs (Kruskal-Wallis rank sum test; *An. darlingi* – $p = 0.16$, chi-squared $= 5.2$, df $= 3$; *An. triannulatus* – $p = 0.93$, chi-squared $= 0.44$, df $= 3$).