Determining species of *Anopheles gambiae* s.l. in Burkina Faso using near-infrared spectroscopy

Bernard Mouonniba SOME[1]*,
Dari Frédéric DA[1]*,
Nicaise Denis Codjo DJEGBE[1,2],
Lawata Inès Géraldine PARE[1, 2],
Roch Koumbobr DABIRE[1]

Abstract

Background: The ability to differentiate accurately *Anopheles gambiae* s.l. species is critical for malaria vector surveillance and control. Traditional methods for species identification, such as morphological analysis and polymerase chain reaction (PCR) are labor-intensive, time-consuming, and reliant on specialized laboratory infrastructure. In recent years, near-infrared spectroscopy (NIRS) has emerged as a promising alternative, offering a rapid, non-destructive, and cost-effective approach for species determination.
Methods: This study explores the use of NIRS to differentiate *An. gambiae*, *An. coluzzii* and *An. arabiensis* that are morphologically sibling mosquito species and the major malaria vectors in Burkina Faso. The methodology involves collecting near-infrared absorbance data from individual mosquito and applying machine-learning algorithms to classify the samples based on their spectral profiles. Thus, laboratory-reared mosquitoes or wild-caught ones from constant and varying ages were included in this study.
Results: Using laboratory-reared *Anopheles gambiae* s.l., NIRS average accuracy in classifying mosquito species of constant age was 93% while the analysis involving mosquitoes of varying ages, the accuracy falls to 59%. In addition to the laboratory data, wild *Anopheles* mosquito data results for constant or varying age achieved an average accuracy of 73% and 83%, respectively.
Conclusion: The study demonstrated that NIRS can determine *Anopheles gambiae complex* species with high accuracy, though it varied depending on experimental conditions such as the mosquitoes age.

Keywords: near-infrared spectroscopy (NIRS); machine learning; *Anopheles gambiae* s.l.

Détermination des espèces de *Anopheles gambiae* s.l. au Burkina

[1] CNRST/Institut de Recherche en Sciences de la Santé, Direction Régionale de l'Ouest, 399 avenue de la liberté, 01 BP 545 Bobo-Dioulasso 01, Burkina Faso.
[2] Université Nazi Boni, Bobo-Dioulasso, Burkina Faso
*** Corresponding author:**
Bernard Mouonniba SOME, bernardsilome40@gmail.com; ORCID ID: https://orcid.org/0009-0009-7346-1207
Dari F. DA, dafrenick@yahoo.fr; ORCID ID: https://orcid.org/0000-0002-9199-9133

DOI : https://doi.org/10.64707/revstss.v48i2.1815

Faso utilisant la technique de la spectroscopie proche infrarouge

Résumé

Introduction : La différenciation avec précision des espèces du complexe *Anopheles gambiae* est essentielle dans la lutte contre les vecteurs du paludisme. Les espèces d'anophèles vectrices du paludisme sont couramment identifiées par les techniques d'identification morphologique et de la biologie moléculaire qui sont très laborieuses, chronophages et ne sont réalisables qu'avec des techniciens qualifiés et dans des laboratoires spécialisés. Ces dernières années, la spectroscopie proche infrarouge (NIRS) est apparue comme une alternative prometteuse, rapide, non destructive et économique pour la détermination des espèces de moustiques.
Méthodes : La méthodologie implique la collecte des absorbances des moustiques dans le proche infrarouge et l'utilisation d'algorithmes d'apprentissage automatique pour déterminer *An. gambiae*, *An. coluzzii* et *An. arabiensis*. Ainsi, des moustiques d'âge constant ou variable, élevés au laboratoire ou collectés dans la nature, ont été inclus dans cette étude.
Résultats : Chez les anophèles d'âge constant de laboratoire, la précision moyenne de la NIRS pour différencier les espèces de *Anopheles gambiae* s.l. était de 93 %. Cette précision a connu une diminution et était de 59 % lorsque l'analyse a porté sur des moustiques de laboratoire d'âges variables. Chez les anophèles sauvages, des précisions moyennes de 73 % et de 83 % ont été obtenues respectivement sur des échantillons de *Anopheles gambiae* s.l. d'âge constant et d'âge variable.
Conclusion : Cette étude a démontré que la précision de la NIRS est assez élevée pour les espèces du complexe *Anopheles gambiae*, bien qu'elle varie en fonction de certaines conditions expérimentales telles que l'âge des moustiques.

Mots clés : Spectroscopie proche infrarouge (NIRS) ; Apprentissage automatique ; *Anopheles gambiae* s.l.

Background

Vector borne diseases, such as malaria, dengue, chikungunya, and Zika, cause significant burden in tropical regions worldwide. As a result, anti-vector measures are among the most critical global public health strategies, being the primary tools to control these diseases. Monitoring the population dynamics of disease-vector, their age distribution and pathogen infection status, is a direct measure of the effectiveness of vector-borne disease control interventions. For a vector to be able to transmit a pathogen in a given area, it must be able to ingest the pathogen, survive to ensure its development and transmit it to another vertebrate (1,2). Consequently, for any vector-borne disease in a specific region, vector species must be accurately recognized prior to

implementing any vector control strategies. In the context of malaria, there are approximately 500 species of *Anopheles* mosquitoes (3), with approximately sixty capable of transmitting the malaria parasite (4,5). These *Anopheles* species involved in *Plasmodium* transmission are typically identified by laborious methods through morphological analysis and molecular biology technique (6). Over the past decade, Near-infrared spectroscopy, a fast, cost-effective, and user-friendly technique, has been tested to identify the *Anopheles* species. The NIRS exploits the relationship between the physical and chemical properties of a material and its ability to absorb light at specific near-infrared wavelengths. This method is particularly effective for studying organic molecules composed of carbon, hydrogen, oxygen and nitrogen elements found in the cuticles of dipteran insects, which have species-specific absorption profiles. These unique compositions allow performing models to identify mosquito species (7). By detecting the molecular vibrations of C-H, O-H and N-H bonds within dipteran cuticle, the technique measures reflectance (R) at different wavelengths and converts it to absorbance (Log1/R) used for analysis. Previous studies conducted in Tanzania have demonstrated that NIRS can reliably distinguish between *Anopheles gambiae* and *Anopheles arabiensis* in both laboratory and field settings with good accuracy (7,8). A subsequent study was conducted four years later on the same species under semi-natural conditions to evaluate the precision of NIRS in distinguishing these two species (9). In addition, recent research conducted in Burkina Faso, involving *Anopheles gambiae* and *Anopheles coluzzii* showed that NIRS is capable to distinguish both *Anopheles* species (10). Since the introduction of NIRS in malaria research, studies have consistently focused on the same species, *Anopheles gambiae* and *Anopheles arabiensis* or *Anopheles gambiae and Anopheles coluzzii*. In this study, we assessed for the first time, the ability of NIRS to differentiate *An. gambiae, An. coluzzii*, and *An. arabiensis*, three key malaria vector species in Burkina Faso.

I. Material and methods

Mosquito rearing and sampling

Three colonies of *Anopheles* mosquitoes (*An. gambiae, An. coluzzii* and *An. arabiensis*) were reared in the insectary (Figure 1) of the "institut de recherche en sciences de la santé" (IRSS) (10,11) and the resulting females were used for NIRS analysis. During experiments, mosquitoes from these three *Anopheles* species were fed on 10% glucose solution *ad libitum*.

Eggs from the three species of *Anopheles* were collected in labelled plastic trays. The resulting pupae were collected over several consecutive days and placed in different emergence cages. From their emergence day, about 20 female *Anopheles* mosquitoes per species were daily sampled, killed with chloroform and subsequently analyzed using NIRS technique. This process was repeated daily until the mosquito population was fully depleted, which occurred around 20 days after emergence. To determine *Anopheles* mosquito species using NIRS, two analyses were performed. The first analysis was carried out on individuals of a same age, specifically 4 days post-emergence. The second analysis consisted in applying the method to a more diverse population of mosquitoes that varied in age, ranging from 1 to 21 days. This approach allowed the assessment of NIRS's accuracy in identifying species across different age groups.

Additionally, wild *Anopheles* mosquitoes with known ages were analyzed using NIRS technique for mosquito species identification. To achieve this, *Anopheles gambiae* s.l. larvae (F0) were collected from breeding sites in localities around Bobo Dioulasso, while gravid or blood-fed females were captured in human dwellings using mouth aspirators (12) between 6:00–9:00 AM. The mosquitoes were transported to the insectary for oviposition, and the first generation (F1) reared under controlled conditions at the same time as the larvae collected in the field (F0). Once adults were four day-olds post-emergence, both F0 and F1 mosquitoes were killed with chloroform, scanned with NIRS and stored in labeled Eppendorf tubes for PCR analysis (6). DNA of each mosquito was extracted with DNAzol, and the conventional PCR (6) was performed to determine *Anopheles* species. The PCR technique was used as the reference method for species identification and study validation.

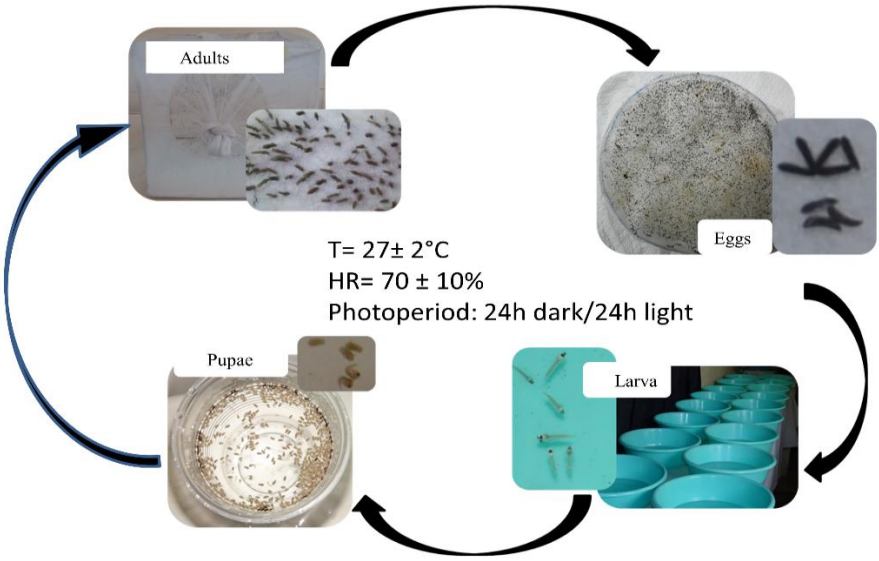

Figure 1: Rearing process of *Anopheles* mosquitoes in the Laboratory

Collecting wild *Anopheles* mosquitoes

Anopheles gambiae s.l. were collected in human dwellings in localities around Bobo Dioulasso, early in the morning using mouth aspirators (12). Once transported to IRSS insectary, all unfed *Anopheles* were sorted, immediately killed with chloroform and then freshly scanned with NIRS. After scanning, the conventional PCR was performed on each cephalothorax to determine *Anopheles* species.

Mosquito scanning

The *Anopheles* females were killed by exposure to chloroform for around 5 minutes. Approximately twenty specimens were placed on a white scanning disc. Each mosquito's cephalothorax was focus under the optical fiber's light beam, which was connected to a spectrometer (Figure 2). The RS3 software (ASD Inc., Boulder, Colorado) was used to collect mosquito spectral. The scanning process was carried out within the electromagnetic spectrum range of 350 to 2500 nm. The machine automatically performed 20 scans and the RS3 spectral acquisition software records the average. The spectral data from the mosquitoes collected in ASD format is converted into numerical text format, which is used for statistical analysis and mathematical modelling.

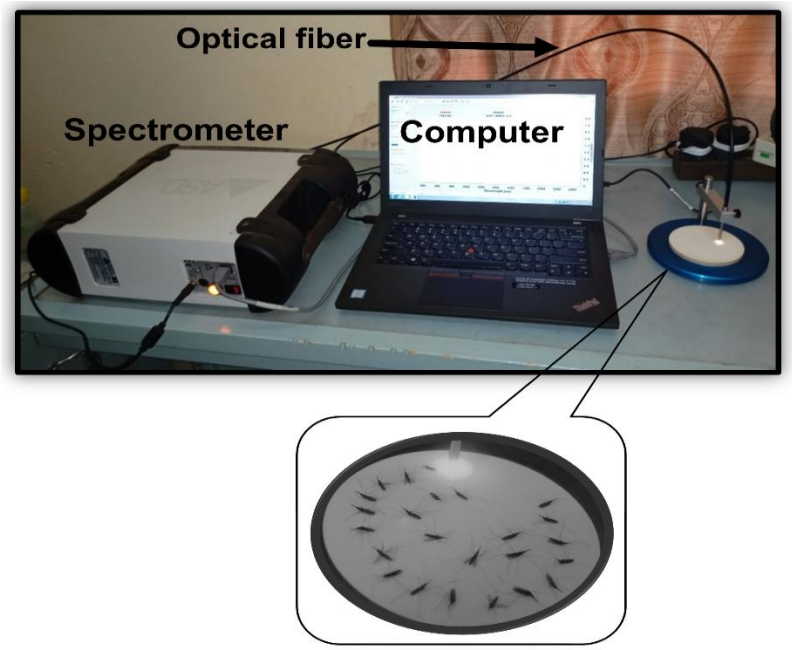

Figure 2: Analysis of *Anopheles* mosquito using NIRS technique

Data analysis

A statistical machine learning approach was used to fit and validate a generalized linear model (GLM). A standard three-stage analysis approach was used to build and assess a predictive model: training, validation and testing. Thus, the dataset was divided into three subsets, each used at a different stage: (i) the training dataset was used to train the model; (ii) the validation dataset was used to validate the model and (iii) the test dataset was used to evaluate the final model.

The proportions of data used in each subset were as follows: 50% for training, 25% for validation and 25% for testing. When species identification involved two *Anopheles* species (*An. gambiae* and *An. coluzzii*), a binomial logistic classification was used, whereas multinomial logistic classification was applied in the case where the test involved three species (*An. gambiae*, *An. coluzzii* and *An. arabiensis*). Two response classes were assigned in the binomial classification: y = 0 for *An. gambiae* and y = 1 for *An. coluzzii,* and three response classes in the multinomial classification: y = 0 for *An. arabiensis,* y = 1 for *An. coluzzii* and y = 2 for *An. gambiae*. The coefficient functions show the most important spectral regions, i.e. the most informative wavelength for prediction, constitutes the key element

of the model. The area under the receiver operating characteristic (ROC) curve (AUC, area under curve) was used to assess the predictive model's performance.

II. Results

NIRS accuracy to determine laboratory-reared *Anopheles gambiae* s.l. species

For mosquitoes of constant age (4 days post-emergence), a sample of 437 individuals (*An. arabiensis*: 133; *An. gambiae*: 144 and *An. coluzzii*: 160) was analyzed. The Figure 3 depicted the spectral profiles of these mosquitoes. The mean AUC across the species was 97% (Figure 4A), and the overall classification accuracy for the three species reached 93% (*An. arabiensis*: 89%; *An. gambiae*: 94% and *An. coluzzii*: 95%; Figure 4B). In a separate analysis involving a population of 1007 mosquitoes of varying ages (*An. arabiensis:* 333; *An. gambiae*: 300 and *An. coluzzii*: 364), the predictive model's performance was evaluated using the AUC of the ROC curve, which was 77% (Figure 4C). The average classification accuracy for these mosquito species was lower: 59% (*An. arabiensis*: 64%; *An. gambiae*: 57% and *An. coluzzii*: 54%; Figure 4D). However, when the model was trained exclusively on *An. gambiae* and *An. coluzzii* of varying ages, the classification accuracy significantly improved, up to 85%.

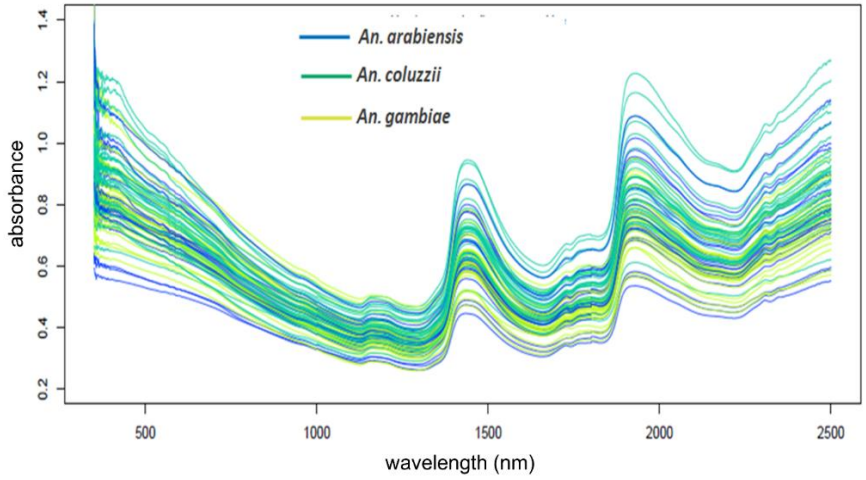

Figure 3: Spectral profiles of laboratory-reared *Anopheles gambiae* s.l. derived from NIRS scanning

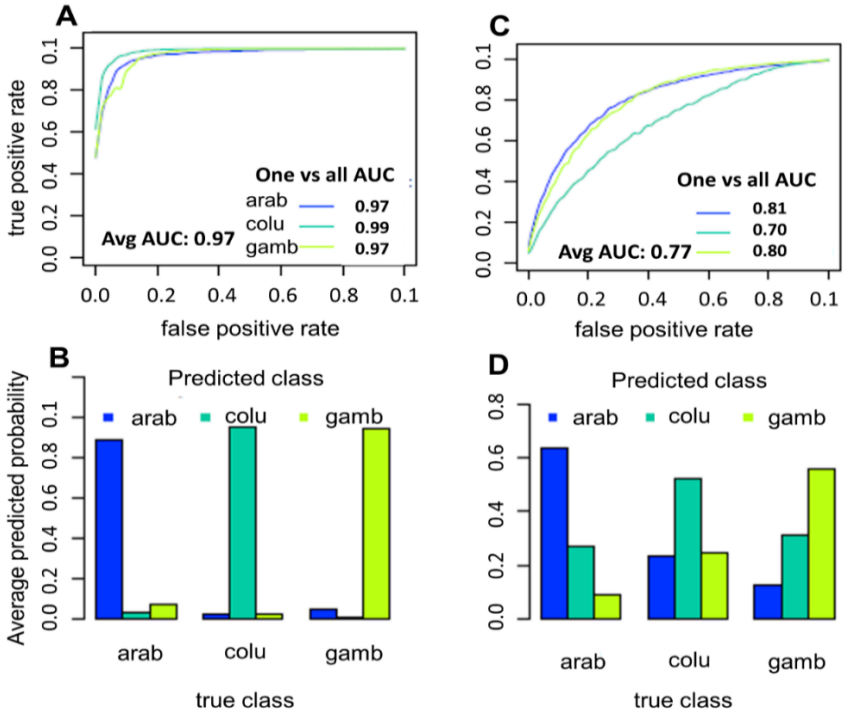

Figure 4: NIRS accuracy in determining laboratory-reared *Anopheles gambiae* s.l. species. (A) ROC curve of*Anopheles* of constant age; (B) NIRS accuracy in classifying *Anopheles* species of constant age; (C) ROC curve of *Anopheles* of varying ages and (D) NIRS accuracy in classifying *Anopheles* species of varying ages. arab=*An. arabiensis*; colu=*An. coluzzii*, gamb=*An. gambiae* and avg AUC=average area under the curve.

NIRS ability to determine wild *Anopheles gambiae* s.l. species of constant age

A sample of 531 wild *Anopheles* mosquitoes (182 F0 and 349 F1) were analysed for all aged four days. The species identified through PCR included *An. arabiensis* (151/538), *An. coluzzii* (148/538), and *An. gambiae* (232/538). Analysis of the spectral data demonstrated that NIRS achieved a classification accuracy of 73% with an average area under the curve (AUC) of 85% (Table I).

Table I: NIRS accuracy in determining wild *Anopheles gambiae* s.l. species of constant age.

	An. arabiensis	*An. coluzzii*	*An. gambiae*	Average
Accuracy	52%	75%	82%	73%
AUC	0.81	0.89	0.87	0.85

NIRS accuracy to determine wild-caught *Anopheles gambiae* s.l. species

The wild *Anopheles* collected (4343 *Anopheles*) included *An. gambiae*, *An. coluzzii* and *An. arabiensis*, which were identified in proportions of 41.56%, 45.66% and 12.78% respectively. Due to the low number of *An. Arabiensis*, this species data was excluded for the analysis. Identification accuracy for *An. gambiae* and *An. coluzzii* was 83% (*An. gambiae*: 84% and *An. coluzzii*: 83%, Figure 5). When the model was trained on laboratory-reared *Anopheles* to predict the species of field-collected mosquitoes, the accuracy decreased to 76% (75% for *An. gambiae* and 77% for *An. coluzzii*).

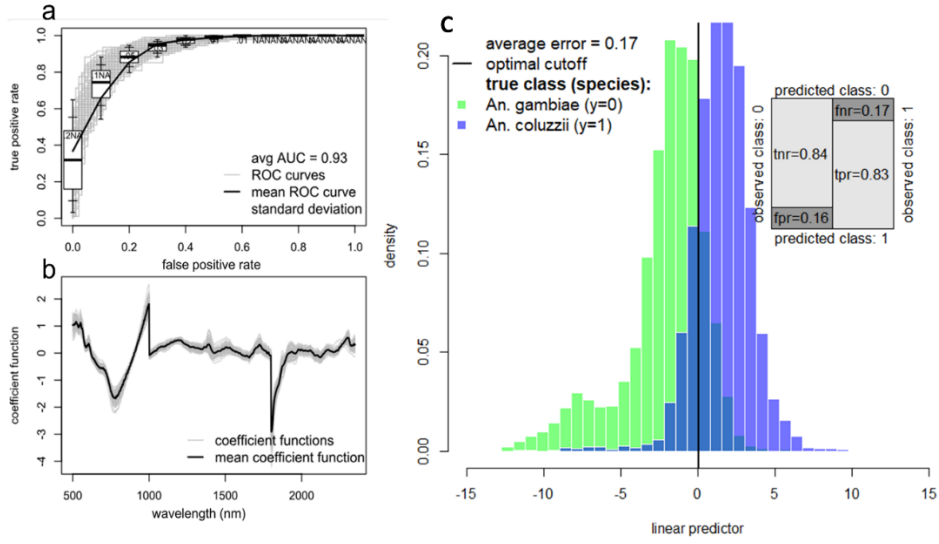

Figure 5: NIRS accuracy to predict wild-caught *Anopheles gambiae* s.l. species. (a) ROC curves; (b) coefficient function; (c) histogram of the linear predictor; the actual classes are colour-coded (blue for *An. gambiae* and green for *An. coluzzii*). The

vertical black line indicates the optimum threshold for classifying mosquitoes as *An. gambiae* or *An. coluzzii*. The shaded area where the two distributions overlap corresponds to misclassified test observations, The confusion matrix (inset) shows the different error rates: false negative rate (fnr), false positive rate (fpr), true negative rate (tnr; *An. gambiae*); true positive rate (tpr; *An. coluzzii*).

III. Discussion

The findings of this study suggest that NIRS can be used to accurately identify the sibling species within the *An. gambiae* complex, not only for laboratory-reared specimens but also for wild-caught *Anopheles* mosquitoes. Notable differences exist in the cuticular hydrocarbons of *Anopheles* species.

For instance, the relative abundance of 13-methylhentriacontanes compared to n-hentriacontanes is higher in *An. arabiensis* (0.79 ± 0.29) than in *An. gambiae* (0.30 ± 0.22) (13). Similarly, the relative abundance of paraffins with 26 and 27 carbon atoms serves as another distinguishing marker for *Anopheles* species differentiation (13).

Under controlled laboratory conditions, the NIRS technique demonstrates higher accuracy in identifying *Anopheles* species of constant age compared to those with varying ages. This indicates that the age is a critical factor influencing NIRS accuracy in *Anopheles* species determination because cuticle component change with mosquito senescence. Considering the impact of age variability, calibrating the machine considering this parameter could enhance prediction accuracy. The high accuracy achieved by NIRS in identifying laboratory-reared *Anopheles* species can be attributed not only to the uniformity in their generational age but also to standardized rearing conditions, including feeding, constant temperature and relative humidity. These controlled factors minimize spectral variations, ensuring reliable species predictions to develop the model using *Anopheles* mosquito of constant age.

Wild *Anopheles* mosquitos' results demonstrated that the NIRS technique achieved high accuracy in identifying species of variable ages compared to those of constant age. This discrepancy with laboratory findings could be attributed to two key factors. First, the constant-age analyses involved three species (*An. arabiensis*, *An. gambiae* and *An. coluzzii*), whereas the variable-age analyses included only two species (*An. gambiae* and *An. coluzzii*) due to the lower number of *An. arabiensis* species which was excluded. Second, the sample size for constant-age analyses was significantly smaller five times fewer

mosquitoes than that of variable-age analyses. It was established that larger sample sizes enhance the NIRS accuracy (14). The NIRS technique could differentiate wild *An. gambiae* and *An. coluzzii* of varying age with an accuracy of 83%, over of accuracy reported in a previous study from Tanzania (7), which include wild *An. gambiae* and *An. arabiensis* species. This study included mosquitoes of various physiological status, such as gravid and blood-fed individuals, without compromising model accuracy which is one of limitations observed in prior research (7). While the exact age of the wild mosquitoes was unknown and the experimental design did not allow for assessing the impact of age on species identification models, an age-based analysis could potentially influence the results. Nonetheless, these findings underscore the robustness of NIRS in identifying wild *Anopheles* species across variable ages under field conditions.

Near-infrared spectroscopy differentiated laboratory-reared *Anopheles* species with higher accuracy compared to their wild counterparts. This discrepancy may be attributed to differences in environmental and nutritional factors affecting the cuticular composition of the mosquitoes. Laboratory-reared *Anopheles* were maintained under controlled conditions, including constant temperature, humidity, and diet, as Tetramine® Baby Fish Food for larvae and a 10% glucose solution for adults. In contrast, wild *Anopheles* likely developed in diverse environmental conditions with varying food sources, such as nectar and blood and different physiological states (parturition, repletion, …). Additionally, the wild mosquito samples may include individuals from multiple generations (F0, F1, ...). Other unexamined factors, such as variations in insecticide resistance level within the same species, could also influence absorption spectra and impact the NIRS accuracy (9). These findings suggest caution when interpreting moderate accuracy levels for wild *Anopheles* species identification.

A decline out-of-sample accuracy, which corresponds to the use of laboratory *Anopheles* to predict wild caught *Anopheles* species, was observed. This suggests structural differences between laboratory-reared *Anopheles* and their wild counterparts.

The findings suggest that NIRS technique could help understanding the distribution and dynamics of mosquito species by allowing to monitor vector populations in real-time. Additionally, this study is the first to include three *Anopheles* species in NIRS-based analyses and aligns with earlier findings highlighting NIRS's usefulness for evaluating relevant

entomological parameters in malaria transmission. These results underline the potential of NIRS as a powerful tool for monitoring *Anopheles* mosquito populations. To enhance comprehensive vector management, future research should explore methods to combine NIRS with other monitoring techniques and expand the database to encompass a greater diversity of mosquito species and environmental conditions.

Limits of the study

NIRS is a technique for determining the chemical composition of materials across several fields. In this study, NIRS demonstrated high accuracy in determining *Anopheles* species, but the accuracy can be improved. Therefore, the algorithms used in *Anopheles* species analyses could be improved to achieve a most effective prediction of the technique. The moderate accuracy in predicting the species of wild *Anopheles* should not be over-interpreted for several environmentally related factors. Indeed, the wild mosquitoes were probably composed of a mixture of individuals from different generations that had evolved in various climatic conditions likely to interact with the effectiveness of the technique. Climatic factors such as temperature and relative humidity can influence NIRS spectra through seasonal rearrangements of hydrocarbons. Thus, various environmental conditions effect on NIRS performance needs to be evaluated.

Acknowledgements

The authors thank inhabitants of Longo and children guardians for their sincere cooperation during mosquitoes sampling and blood donor's recruitment.

Authors' contribution

B.M.S., D.F.D. and R.K.D. conceived the study; B.M.S., N.D.C.D. and L.I.G.P conducted the laboratory and fieldwork; B.M.S. and N.D.C.D. conducted the molecular analysis, B.M.S. was responsible for the data analysis; B.M.S. wrote the first draft of manuscript, all authors approved the final manuscript.

Ethics approval and consent to participate

Ethic committee approval: Reference number: A018-2017/CEIRES.

Competing interest

The authors declare no competing interest.

References

1. Dye C. The Analysis of Parasite Transmission by Bloodsucking Insects. Annu Rev Entomol. 1992 Jan;37(1):1–19.
2. Lord C, Woolhouse M, Heesterbeek J, Mellor P. Vector-borne diseases and the basic reproduction number: a case study of African horse sickness. Medical and veterinary entomology. 1996;10(1):19–28.
3. Harbach RE. The phylogeny and classification of *Anopheles*. In: *Anopheles* mosquitoes-New insights into malaria vectors. IntechOpen; 2013.
4. Fontenille D, Cohuet A, Awono-Ambene P, Kengne P, Antonio-Nkondjio C, Wondji C, et al. Vecteurs de paludisme : du terrain à la génétique moléculaire Recherches en Afrique. Revue d'Épidémiologie et de Santé Publique. 2005 Jun;53(3):283–90.
5. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A global map of dominant malaria vectors. Parasites & vectors. 2012;5(1):1–11.
6. Santolamazza F, Mancini E, Simard F, Qi Y, Tu Z, della Torre A. Insertion polymorphisms of SINE 200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. Malar J. 2008 Dec;7(1):163.
7. Mayagaya VS, Michel K, Benedict MQ, Killeen GF, Wirtz RA, Ferguson HM, et al. Non-destructive Determination of Age and Species of *Anopheles gambiae* s.l. Using Near-infrared Spectroscopy. The American Journal of Tropical Medicine and Hygiene. 2009 Oct 1;81(4):622–30.
8. Sikulu M, Killeen GF, Hugo LE, Ryan PA, Dowell KM, Wirtz RA, et al. Near-infrared spectroscopy as a complementary age grading and species identification tool for African malaria vectors. 2010;7.
9. Sikulu MT. Non-Destructive near Infrared Spectroscopy for Simultaneous Prediction of Age and Species of Two Major African Malaria Vectors: *An. Gambiae* and *An. Arabiensis*. NIR news. 2014 Aug;25(5):4–6.
10. Somé BM, Da DF, McCabe R, Djègbè NDC, Paré LIG, Wermé K, et al. Adapting field-mosquito collection techniques in a

perspective of near-infrared spectroscopy implementation. Parasites & Vectors. 2022 Sep 26;15(1):338.

11. Guissou E, Waite JL, Jones M, Bell AS, Suh E, Yameogo KB, et al. A non-destructive sugar-feeding assay for parasite detection and estimating the extrinsic incubation period of *Plasmodium falciparum* in individual mosquito vectors. Sci Rep. 2021 Dec;11(1):9344.
12. Anthony TG, Trueman HE, Harbach RE, Vogler AP. Polymorphic microsatellite markers identified in individual *Plasmodium falciparum* oocysts from wild-caught *Anopheles* mosquitoes. Parasitology. 2000 Aug;121(2):121–6.
13. Carlson DA, Service MW. Identification of Mosquitoes of *Anopheles gambiae* Species Complex A and B by Analysis of Cuticular Components. Science. 1980 Mar 7;207(4435):1089–91.
14. Lambert B, Sikulu-Lord MT, Mayagaya VS, Devine G, Dowell F, Churcher TS. Monitoring the Age of Mosquito Populations Using Near-Infrared Spectroscopy. Sci Rep. 2018 Dec;8(1):5274.