



RESEARCH ARTICLE

Utility of MALDI-TOF MS for determination of species identity and blood meal sources of primary malaria vectors on the Kenyan coast [version 1; peer review: 1 approved, 1 approved with reservations]

Jonathan Karisa ¹⁻³, Kelly Ominde ¹, Mercy Tuwei^{1,2}, Brian Bartilol ¹, Zedekiah Ondieki¹, Harun Musani¹, Caroline Wanjiku ¹, Kioko Mwikali¹, Lawrence Babu¹, Martin Rono ^{1,2}, Mumin Eminov⁴, Charles Mbogo ¹, Philip Bejon ^{1,5}, Joseph Mwangangi¹, Maureen Laroche⁶, Marta Maia ^{1,5}

¹Kenya Medical Research Institute, Wellcome Trust Research Programme, Kilifi, Kenya, 230-80108, Kenya

²The Open University, Milton Keynes, United Kingdom, Walton Hall, Kents Hill, Milton Keynes MK7 6AA, UK

³Pwani University, Kilifi, Kenya, 195-80108, Kenya

⁴Bruker Daltonik GmbH, Bremen, Germany

⁵University of Oxford, Centre for Global Health and Tropical Medicine, Oxford, UK, Oxford, UK

⁶The University of Texas Medical Branch -, Galveston National Laboratory 301 University Blvd, Texas, Galveston TX 77555-1019, USA

V1 First published: 31 Mar 2023, 8:151
<https://doi.org/10.12688/wellcomeopenres.18982.1>
Latest published: 31 Mar 2023, 8:151
<https://doi.org/10.12688/wellcomeopenres.18982.1>

Abstract

Background: Protein analysis using matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS) represents a promising tool for entomological surveillance. In this study we tested the discriminative power of this tool for measuring species and blood meal source of main Afrotropical malaria vectors on the Kenyan coast.

Methods: Mosquito collections were conducted along the coastal region of Kenya. MALDI-TOF MS spectra were obtained from each individual mosquito's cephalothorax as well as the abdomens of blood-engorged mosquitoes. The same mosquitoes were also processed using gold standard tests: polymerase chain reaction (PCR) for species identification and enzyme linked immunosorbent assay (ELISA) for blood meal source identification.

Results: Of the 2,332 mosquitoes subjected to MALDI-TOF MS, 85% (1,971/2,332) were considered for database creation and validation. There was an overall accuracy of 97.5% in the identification of members of the *An. gambiae* (*An. gambiae*, 100%; *An. arabiensis*, 91.9%; *An. merus*, 97.5%; and *An. quadriannulatus*, 90.2%) and *An. funestus* (*An. funestus*, 94.2%; *An. rivulorum*, 99.4%; and *An. lesoni*, 94.1%) complexes. Furthermore, MALDI-TOF MS also provided accurate

Open Peer Review

Approval Status

	1	2
version 1		
31 Mar 2023	view	view

1. **Max Maurin**, Universite Grenoble Alpes, Saint-Martin-d'Hères, France
2. **Petr Halada** , Institute of Microbiology of the Czech Academy of Sciences, Průmyslová, Czech Republic

Any reports and responses or comments on the article can be found at the end of the article.

(94.5% accuracy) identification of blood host sources across all mosquito species.

Conclusions: This study provides further evidence of the discriminative power of MALDI-TOF MS to identify sibling species and blood meal source of Afrotropical malaria vectors, further supporting its utility in entomological surveillance. The low cost per sample (<0.2USD) and high throughput nature of the method represents a cost-effective alternative to molecular methods and could enable programs to increase the number of samples analysed and therefore improve the data generated from surveillance activities.

Keywords

MALDI-TOF MS, entomological surveillance, Anopheles, high-throughput, mass spectrometry, Kenya, Coast

Corresponding author: Jonathan Karisa (kjonathan@kemri-wellcome.org)

Author roles: **Karisa J:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Ominde K:** Investigation, Methodology, Writing – Review & Editing; **Tuweil M:** Investigation, Methodology, Writing – Review & Editing; **Bartilol B:** Investigation, Methodology, Writing – Review & Editing; **Ondieki Z:** Investigation, Methodology, Writing – Review & Editing; **Musani H:** Investigation, Methodology, Writing – Review & Editing; **Wanjiku C:** Investigation, Methodology, Project Administration, Writing – Review & Editing; **Mwikali K:** Data Curation, Formal Analysis, Writing – Review & Editing; **Babu L:** Data Curation, Formal Analysis, Writing – Review & Editing; **Rono M:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Eminov M:** Investigation, Methodology, Writing – Review & Editing; **Mbogo C:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Bejon P:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Mwangangi J:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Laroche M:** Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Maia M:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing

Competing interests: Bruker did not fund or participate in study design or analysis; however, the authors declare that they have obtained technical support for quality control purposes. Bruker provided guidance/assistance in database creation and validation and Mumin Eminov is affiliated with Bruker Daltonik GmbH. No other competing interests were disclosed.

Grant information: This research was supported by the African Academy of Sciences (AAS) under a DELTAS Africa Initiative grant [107769/Z/10/Z, <https://doi.org/10.35802/107769>] as part of the Initiative to Develop African Research Leaders (IDeAL) and from DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the AAS's Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769/Z/10/Z] and the UK government. The views expressed in this publication are those of the authors and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government. This study was also funded by UNITAID (BOHEMIA – Broad One-Health Endectocide-based Malaria Intervention in Africa) and further support was received from The Royal Society FLAIR fellowship grant FLR_R1_190497 awarded to Martin Rono.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Karisa J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Karisa J, Ominde K, Tuweil M *et al.* **Utility of MALDI-TOF MS for determination of species identity and blood meal sources of primary malaria vectors on the Kenyan coast [version 1; peer review: 1 approved, 1 approved with reservations]** Wellcome Open Research 2023, 8:151 <https://doi.org/10.12688/wellcomeopenres.18982.1>

First published: 31 Mar 2023, 8:151 <https://doi.org/10.12688/wellcomeopenres.18982.1>

Introduction

Globally, malaria has significantly declined in the last decades, however countries in sub-Saharan Africa (SSA) continue to endure the high morbidity and highest mortality rates. In 2020, around 241 million cases and 627,000 related deaths were reported globally, with over 90% of the deaths reported in Africa¹. The WHO Global Technical Strategy (GTS) 2016–2030 is not on track having missed the target of reducing malaria case incidence and mortality by at least 40% by 2020². Sustainable vector control strategies capable of addressing current gaps and enabling malaria elimination in SSA will require the development of new malaria vector control tools and approaches and/or improvement of existing ones to malaria vector control, as well as improved vector surveillance systems³. One of the pillars of the global vector control response is the enhancement of vector surveillance. In order to design appropriate and effective vector control strategies, it is imperative to understand the behaviour, biology, and ecology of local vectors. Over 500 *Anopheles* species have been described globally, with approximately 50 species incriminated in malaria transmission as either primary or complementary vectors^{4–7}. In Sub-Saharan-Africa, *Anopheles gambiae* and *An. funestus* complexes dominate^{5,6}; and consist of morphologically indistinguishable members but with distinct biting and resting behaviour. Understanding their composition, distribution, and behaviour including blood feeding patterns and preferred human and animal hosts would be fundamental in designing effective control strategies. In general entomological practice the field, morphologically identical mosquitoes are sorted into complexes using taxonomic keys^{8,9} and sibling species distinguished using molecular methods such as polymerase chain reaction (PCR)^{10–12}. On the other hand, blood meal sources are commonly identified using enzyme linked immunosorbent assays (ELISA)^{13,14}. Although PCR is highly sensitive, it is technically demanding, time-consuming and has a high cost per sample. Contrastingly, ELISA is cheaper, but also equally laborious and impeded by unavailability of antibodies and cross-reactions that sometimes produce spurious results that are difficult to interpret¹⁵.

MALDI-TOF MS is a protein profiling technique with a variety of applications that include microbiology, and entomology. In particular, it has revolutionized clinical microbiology by providing an accurate, rapid, and inexpensive identification of microorganisms¹⁶. Most recently, this tool has been shown to be able to identify mosquito sibling species^{17–20}, blood meal sources^{19,21,22}, as well as pathogen infection^{23–26}. The low-cost per sample, rapidity, and accuracy of MALDI-TOF MS makes it a reliable method for biotyping different mosquito parameters with the potential to exponentially increase the number of specimens analysed by surveillance programs.

This current study evaluated the utility of MALDI-TOF MS in determining the identity of sibling species of the *An. gambiae* and *An. funestus* complexes and associated blood meal sources. To the best of our knowledge, this is the first-time members of the *An. funestus* complex have been characterized using MALDI-TOF MS.

Methods

Study area

Mosquito samples were collected from 2019–2022 in Kilifi, Taita Taveta and Kwale Counties of Kenya. The Coastal region of Kenya encounters bimodal rainfall; long rains happening between April and July and short rains between October and December with mean annual rainfall of 750 to 1,200 mm. The relative humidity ranging between 55 and 65% and mean annual temperature between 20 to 35°C. The altitude ranges between 0 and 400 meters above sea level. Both Kilifi and Kwale counties are inhabited mostly by the Mijikenda; whereas Taita Taveta is mainly inhabited by the Taita and Taveta ethnic groups; communities that rely largely on farming and fishing and that live houses made of mud, or coral rock and roofed using palm leaves (*makuti*)^{27–29}.

In Kilifi County, sampling was conducted in Garithe, Burangi, Marana, Mtondia and Siyu villages^{30,31}. In Kwale, sampling was conducted between June and July 2021 in 12 villages *viz* Fihoni, Jego, Marigiza, Mangwei, Gazi, Madongoni, Kikwezani, Kiwegu, Kidomaya, Mwanamamba, Tsuini and Mukuduru. In Taita Taveta county, sampling was done in Kimundia, Kiwalwa, Mwarusa and Njoro⁶. The villages were selected based on mosquito collection data from previous studies that indicated the distribution of both *An. gambiae* and *An. funestus* complexes^{6,30–32}.

Mosquito collections

In each of the sampling sites, 10 houses were randomly selected from each village. Centers for Disease Control (CDC) light traps were deployed both indoors and outdoors between 1700hrs and 0700hrs. The indoor traps were set up in a room inhabited by at least one person during that night and hung at least 2 m from the ground. The outdoor traps were set at least 5 m from the house and where possible next to a livestock enclosure. In the morning, samples were retrieved from respective traps and transported to the field laboratory in a cooler box for sorting. Additionally, indoor resting mosquitoes were aspirated using Prokopak aspirators in 30 randomly selected houses in the same villages as the CDC-light traps within Kwale county. This was done immediately after retrieval of CDC light traps in the morning and before 0700 hrs. Larval collections from natural breeding sites were also randomly done across the villages using standard dipper method and transported to the laboratory for further processing as previously described²⁷.

Morphological identification and sample preservation

Mosquitoes were morphologically identified to species complex level using taxonomic keys⁸. *Anopheles* mosquitoes were retained and individually dissected into different body parts: legs and wings were used for species identification by PCR; abdomens for blood meal analysis using both MALDI-TOF MS and ELISA (gold standard method); head and thorax for species identification and sporozoite detection by MALDI-TOF MS and ELISA/PCR, respectively. The Anopheline mosquitoes collected in Garithe in 2019 were first placed singly

in 1.5 ml vials and preserved in silica gel for approximately 5 months before being frozen in preparation for analysis. A batch of samples collected in 2021 in Kwale were placed singly in 1.5 ml vials and kept frozen immediately collection and identification. Another batch of samples collected in 2022 in Taita Taveta, Kwale and Kilifi counties were placed *in silica* gel for a few days (≤ 14 days) before freezing them.

MALDI-TOF MS analysis

Protein extraction and plate loading. For species identification, either whole or half of the head and thorax were homogenized in 15 μ l of 70% (v/v) formic acid (FA) (Thermo scientific, Czech Republic) and 15 μ l of 50% (v/v) acetonitrile (ACN) (Thermo scientific, USA) using 106 μ m acid wash glass beads (Sigma-Aldrich, USA) in a Tissue Lyser II (Qiagen, Germany) at 30 Hz for 1 min for three cycles, as previously described²³. For blood meal analysis, abdomens of visibly engorged mosquitoes were crushed in 50 μ l LC-MS grade water (Thermo scientific, USA). A total of 10 μ l of the homogenate was mixed with 10 μ l of 70% FA and 10 μ l of 50% ACN then mixed by vortexing. The samples were then centrifuged at 13,000 rpm for 1 min to separate the debris from the proteins. A total of 1 μ l of the sample was then loaded on a MALDI-target (Bruker Daltonics) plate in quadruplicate and allowed to dry at room temperature. Thereafter, the plate was overlaid with saturated matrix solution (α -cyano-4-hydroxycinnamic acid (Sigma-Aldrich, USA), 50% (v/v) acetonitrile, 2.5% (v/v) trifluoroacetic acid (Thermo scientific, USA) and 47.5% LC-MS grade water (Thermo scientific, USA)) and again allowed to dry. Bacterial Test Standard (BTS) preparation (one spot per plate) was used as positive control and matrix only (four spots per plate) as of the negative control. The plate was then introduced into the Microflex machine (RRID:SCR_019779) (Bruker Daltonics) for spectra acquisition.

Spectra acquisition. Spectra were obtained using the FlexControl software ver. 3.3.0 (Bruker Daltonics). Spectra of mass ranges 2-20 kDa were obtained in a positive linear mode at a frequency of 60 Hz, an acceleration voltage of 20 kV, extraction delay time of 200 ns, and with a maximum laser power energy of 50%. Each spectrum was based on 40 laser shots in six different regions of the sample spot. The spectra were then exported to flexAnalysis software ver. 3.3.0 (Bruker Daltonics) and ClinProTools (RRID:SCR_014341) ver. 4.0 (Bruker Daltonics) (free alternative, Mass-Up) for spectral cleaning and quality control and thereafter for database (DB) creation and blind testing.

Spectra analysis. Spectra quality was visually checked using the FlexAnalysis software ver. 3.3.0 (Bruker Daltonics). Assessing general peak intensity (high intensity), the smoothness of the peaks, the flatness of baseline and its reproducibility compared to other spectra of the same categories. Only spectra of good quality were included for the subsequent analysis. Quality spectra were exported to ClinProTools ver. 4.0 (Bruker Daltonics) for principal component analysis (PCA). To further confirm quality, the spectra were loaded into MALDI Biotyper explorer software ver. 3.3.0 (Bruker Daltonics) (free

alternative, Mass-Up) for data processing, including smoothing, baseline subtraction, normalizing and peak selection. The specificity and reproducibility of the main spectrum profiles (MSPs) of different mosquito species and blood meal sources were checked by cluster analysis using MSP dendrogram and composite correlation index (CCI). MSP clustering was based on mass signals and intensities and with the expectation that mosquitoes of the same sibling species category and their blood meal source cluster on the same branch. On the other hand, CCI was done to assess spectral homogeneity/heterogeneity *i.e.*, variations within and across each sibling species and blood meal source as previously discussed³³. The higher the CCI value, the higher the reproducibility. Composite correlation index CCI value ranged between 0 and 1, with 0 and 1 reflecting no reproducibility and perfect reproducibility, respectively. Composite Correlation index matrix was calculated using MALDI-Biotyper v3.0. software with default settings; mass range 3.0 \pm 12.0 kDa; resolution four; eight intervals; auto-correction off.

Database creation and blind tests (Validation). The reference database containing MSPs was created for species and blood meal identification using spectral fingerprints from the cephalothorax and abdominal sections, respectively, using MALDI-Biotyper software ver. 3.3.0. (Bruker Daltonics). Spectra of good quality from each sibling species and blood meal source randomly selected were loaded in MALDI-Biotyper 3.0 software to create a reference spectra database. This was based on unprejudiced algorithm on intensity, frequency, and peak position of the MSP spectra.

Expectedly, sibling species within the *An. gambiae* and *An. funestus* complexes have high similarity in their protein signatures. Thus, to increase the discriminative power, a minimum of 10 samples per sibling species or blood meal source with high spectral reproducibility were used to create the spectral database^{18,34}. Thereafter, unknown samples were matched against the reference database for which the software assigns a log score value (LSV) ranging from 0-3. Log score value is a biostatistical parameter that provides the level of match between the unknown sample and the reference database³⁵. Log score value ≥ 1.8 was considered as correct identification^{20,36}. For *An. gambiae* sibling species identification by MALDI-TOF MS, ambiguous results (mixture of identification (among the four spots)) (Additional file 1 in *Underlying data*³⁷) can be observed as described earlier³⁸, which we also observed in the current study.

To overcome such challenges, k-nearest neighbor approach was used as previously described by Harju and colleagues³⁴, with slight modifications. Briefly, the MALDI-TOF MS identification ranking list with respective LSV was used to calculate a weighted list score to provide a summary of the list. For each sibling species identified in the ranking list, a weighted LSV was calculated by multiplying the actual LSV to the inverse of their position followed by a summing up the weighted LSV. Given that each sample was spotted in quadruplicate, the mean of the weighted LSV was calculated (Additional file 1 and

Additional file 2 in *Underlying data*³⁷). As a result, the sibling species with the highest mean weighted LSV is regarded as the probable sibling species identity³⁴. This technique has been verified for species identification of closely related organisms and is being considered for use in research by Bruker Daltonics³⁴. When the top hit in at least two of the four spots per sample provided by the MALDI-Biotyper in the ranking list had discordant results with molecular approach, this technique was applied.

Molecular identification and sequencing

Head and thorax of *An. gambiae* and *An. funestus* mosquitoes were subjected to genomic DNA extraction using Chelex protocol as previously described²⁹. A total of 5 µl of the extracts were subjected to a cocktail PCR assay employing primers targeting the intergenic spacer region (IGS) from the 5.8S and 28S coding region, and the internal transcribed region 2 (ITS2) from the 5.8S and 28S coding region flanking the variable ITS2 region for sibling species identification of *An. gambiae* and *An. funestus* complexes following the methods of Scott *et al.*, (1993) and Koekemoer *et al.*, (2002) respectively^{10,11}. *Anopheles gambiae* complex DNA was amplified in a cocktail PCR assay¹¹ in a total reaction volume of 17 µl containing 5 µl 2X green GoTaq master mix (2X Green GoTaq® Reaction Buffer (pH 8.5), 400 µM dATP, 400 µM dGTP, 400 µM dCTP, 400 µM dTTP and 3 mM MgCl₂), 0.5 µl of each of the five primers at a concentration of 10 Mm, 4.5 µl nuclease free water and 5 µl DNA template. The mosquito genomic DNA for *An. funestus* complex were amplified¹⁰ in a total reaction volume of 17.5 µl containing 5 µl 2X green GoTaq master mix, 0.5µl of each of the six primers at a concentration of 10 Mm, 5 µl nuclease free water and 5 µl DNA template. The thermocycler conditions were one cycle of initial denaturation at 95°C for 30 sec followed by 35 cycles of denaturation, annealing and extension at 95°C for 30 min, 55°C for 45 min and 72°C for 30 sec, respectively, with a final elongation at 72°C for 10 min. Amplification products were visualized on 1.5% agarose gel stained with RedSafe™ Nucleic acid staining solution (20,000X) (iNtRON Biotechnology). *An. funestus* s.s., *An. arabiensis* and *An. gambiae* s.s DNA were used as positive controls in their respective assays whereas master mix only was used as negative control.

All samples used for MALDI-TOF MS database creation, those with equivocal results during the test run by MALDI-TOF MS, and a few randomly selected ones were subjected to sequencing using diagnostic primer targeting the ribosomal DNA ITS2 region: ITS2A (Forward primer): TGTGAACTGCA GGACACAT and ITS2B (Reverse primer): TATGCTTAAA TTCAGGGGGT, as previously described¹². Briefly, 20 µl PCR reaction consisting of 10 µl 2X GoTaq Green Master mix (Promega Corporation, USA), 0.5 µl of each of the primers at a concentration of 10 mM was prepared and exposed to the following thermal conditions: 1 cycle of initial denaturation at 95°C followed by 35 cycles of denaturation, annealing and extension at 95°C for 2 min, 52°C for 1 min and 72°C for 30 sec, respectively, with a final elongation at 72°C for 5 min. PCR products were visualized on 1.5% agarose gel stained with

Red safe staining solution. The remaining amplicons were cleaned up using ExoSAP-IT™ Express PCR Product Clean-up (Applied Biosystems; Catalogue Number 75001) as per the manufacturer's guidelines. Lastly, the cleaned-up PCR amplicons were subjected to bi-directional Sanger sequencing as per the manufacturer's guidelines. Briefly, the PCR reaction contained: 2 µl Nuclease free water, 1.5 µl 5X buffer, 1 µl each primer (10 mM) and 2 µl of the cleaned-up PCR amplicons/template. The PCR conditions for this assay were: 96°C for 1 min followed by 45 cycles of 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min.

Using Bio-Edit (RRID:SCR_007361) software (version 7.2.5, 2013), raw ITS2 forward and reverse sequences were checked for quality, insertions, and deletions³⁹. The sequences were edited by removing primer sequences and thereafter aligned. The reverse primer sequence was reverse complimented and using the CAP contig assembly program, a contig from both reverse and forward sequences with a minimum base overlap of 20 bases and match of 85% generated. The resulting contigs were then checked for deletions and insertions and poor-quality sequences (overlapping peaks) excluded from the analysis. The resulting contigs/nucleotides were then compared to reference sequences in the National Center for Biotechnology Information (NCBI) standard nucleic acid databases using the [basic local alignment search tool](#) (BLASTN) (RRID:SCR_001598) using default search parameters.

Blood meal analysis

Direct ELISA assay was used to discriminate blood meal sources by using affinity purified antibody phosphatase labelled [goat anti-bovine IgM](#) 0.1 mg (seraCare, USA, Cat. No: 15-12-03, Lot No.: 111264), KPL affinity purified antibody peroxidase labelled [goat anti-human IgG](#) (H+L) 1.0 mg (seraCare, USA, Cat. No.: 5220-0330; Lot No.: 10266871), KPL Peroxidase-Labelled [rabbit anti-goat IgG](#) (H+L), 0.5 mg (seraCare, USA, Cat. No. 5220-0362 (14-13-06)) and [goat anti-chicken IgG](#) (H+L) 0.5 mg (KPL, USA, Cat. No.: 14-24-06; Lot No: 150194)^{13,14} with slight modifications. Briefly, 1,000 µl 1X phosphate buffered saline (PBS) plain was added to the remaining homogenate crushed in 50 µl LC-MS grade as described above (protein extraction and plate loading section) then vortexed to mix. A total of 50 µl of each homogenate was loaded into a 96 well ELISA microtiter plate (Thermo scientific) and incubated at room temperature for 30 min. The plate contents were aspirated, the plate, washed, 50 µl conjugated mAbs added and plate incubated for 1 hr. The plate contents were then aspirated, plate washed 100 µl ABTS enzyme substrate (2,2'-Azino-bis [3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt) and subjected to a final incubation step for 30 mins before reading the results. For each step, incubation was done at room temperature under subdued light, and washing was done three times using Tween-20 (Sigma-Aldrich, USA). Results were read visually where homogenous greenish blue colour and no colour change was considered a positive and negative result, respectively, as described previously^{14,27}. Serum samples from human, bovine, goat, and chicken blood were used as positive controls and PBS as the negative control.

Data analysis

Data from PCR and MALDI-TOF MS analysis was recorded and cleaned using Microsoft Excel (RRID:SCR_016137) thereafter analysed using R Project for Statistical Computing (RRID:SCR_001905) Version 4.1.1 (2021-08-10) (R Core Team (2021))⁴⁰.

Results

Morphological and molecular identity of wild-collected mosquitoes

A total of 12,038 Anophelines were collected and morphologically identified as belonging to *An. gambiae* and *An. funestus* complexes. From the entire collection, 2,332 were subjected to further morphological and molecular analysis. *Anopheles gambiae* complex was composed of *An. arabiensis* (n=175), *An. merus* (n=250), *An. quadriannulatus* (n=69), unamplified (n=97), whereas *An. funestus* complex was composed of *An. rivulorum* (n=1,104), *An. funestus* (n=296), *An. lesoni* (n=27), *An. parensis* (n=7) and *An. vaneedeni* (n=4) and the rest (n=253) were unamplified (Table 1). For purpose of MALDI-TOF MS database building, we added *An. gambiae* s.s. from an insectary colony (Kilifi strain) (n=50) because we were unable to find this species during our collections. A total of 167 samples consisting of some of the unamplified samples (n=39), those chosen for MALDI-TOF MS database creation (n=55), and a few randomly selected samples (n=73) were selected for Sanger sequencing (Table 1). All mosquito samples used for database creation were incontrovertible. However, with the unamplified samples, results were atypical with two samples previously morphotyped as *An. gambiae* s.l. being identified as *An. pretoriensis* (Additional file 1 in *Underlying data*³⁷). Furthermore, we detected *Anopheles cf. rivulorum* NFL-2015 a new species that is for the first time reported in Kenya and has been implicated in malaria transmission in Eastern Zambia¹² (Additional file 3 in *Underlying data*³⁷).

MALDI-TOF MS analysis

Spectral analysis. Selected spectra from each sibling species in the two complexes, including *An. vaneedeni* and *An. parensis* (Table 1), were used to perform cluster analysis to check for reproducibility and specificity using MALDI Biotyper explorer software ver. 3.3.0. We found three distinct branches: i) *An. rivulorum* and *An. lesoni* ii) *An. vaneedeni*, *An. funestus* and *An. parensis* and iii) for the *An. gambiae* complex where only *An. gambiae* s.s branched separately, while the rest of the members were mixed (Figure 1). The reproducibility was further confirmed using CCI analyses that evaluates the relatedness of MS within and among members of *An. gambiae* and *An. funestus* complexes (Additional file 4 in *Underlying data*³⁷). Respective dark red or blue colour at the intersection square of two groups on the matrix/heat map indicates a close or incongruence relationship. Members of the respective complexes were found to be closely related; but a low correlation of MS spectra was seen between *An. gambiae* and *An. funestus* complexes (Additional file 4 in *Underlying data*³⁷) in agreement with the dendrogram (Figure 1).

Database creation and validation. Of the 2,332 mosquitoes subjected to MALDI-TOF MS, 85% (1,971/2,332) were considered for database creation and validation. The remaining

15% samples were unamplified and were not anophelines and therefore outside the scope of this current study. In the case of *An. parensis* and *An. vaneedeni* we were unable to collect sufficient specimens to allow for database creation and validation (Table 1).

The 1,971 samples consisted of *An. gambiae* (2.5%, n=50), from insectary colony since we did not have specimens from the wild in our collections), *An. arabiensis* (8.9%, n=175), *An. merus* (12.7%, n=250), *An. quadriannulatus* (3.5%, n=69), *An. funestus* s.s. (15.0%, n=296), *An. rivulorum* (56.0%, n=1,104) and *An. lesoni* (1.4%, n=27) (Table 2). Of the 1,971 samples, 5.2% (102/1,971) produced poor quality spectra and flat-lines (suggesting that little or no ionizable proteins were present in the sample), and therefore excluded from this analysis. The remaining samples 4.4% (87/1,971) were used for database creation and 90.4% (1,782/1,971) for validation against the in-house created database. During validation, 44 samples of the 1,782 samples had a low LSV and the identities of another 26 samples were equivocal. Thus, 1,712/1,782 were correctly identified with an accuracy of 96.1% (Table 2). Specifically, there was an 89.1% and 98.3% accuracy in discriminating different members of *An. gambiae* and *An. funestus* complex, respectively (Table 2). However, for *An. funestus*, only three sibling species were included in the database and queried (Table 2). Overall, most of the samples, had an LSV value ≥ 2.00 that is above the cut-off point of 1.80 (Figure 2). In a few scenarios, the LSV was below the 1.8 threshold, probably due to residual blood meals in the head and thorax, which could have interfered with the spectra quality.

For the scenario where the 26 samples of the *An. gambiae* s.l., whose identities were equivocal/ambiguous, we applied a bioinformatics (K-NN) approach to discriminate among the closely related species of *An. gambiae* complex as previously described³⁴. When the top hit in at least two of the four spots per sample provided by the MALDI-Biotyper in the ranking list had discordant results with molecular results, k-nearest neighbor approach was employed (Additional file 1 and 2 in *Underlying data*³⁷). Notably, using this approach, all the samples in the *An. gambiae* complex that had equivocal/discordant results were correctly identified further improving the accuracy in species identification to 95.1% (*An. gambiae*, 100%; *An. arabiensis*, 91.9%; *An. merus*, 97.5%; and *An. quadriannulatus*, 90.2%) (Table 2). Thus, with this approach, the ability to discriminate between *An. funestus* and *An. gambiae* complex rose to 97.5%. To confirm the specificity of MALDI-TOF MS in species discrimination, spectra belonging to *An. pretoriensis*, *An. rufipes*, *Aedes africanus*, *Anopheles cf. rivulorum* NFL-2015, *Culex tritaeniorhynchus* were also queried against the database leading to low log score value (unidentifiable) (Additional file 3 in *Underlying data*³⁷).

Blood meal analysis by ELISA

A total of 223 mosquitoes were subjected for blood meal analysis (Table 3). Of the 206 samples analysed, half of the mosquitoes had fed on bovine (50%), goat (31%), human (12%), and multiple hosts (2.4%) in that order whereas in a small number of samples (4.3%) the blood meals were undetected (Table 3). The human blood index (HBI) was 85.7% for *An. funestus*,

Table 1. Summary table showing the proportion of sibling species of *An. gambiae* and *An. funestus* complexes identified using taxonomic keys and molecular methods and used for MALDI-TOF MS analysis (Parenthesis indicate percentages of each species/category). Abbreviations: MALDI-TOF MS – matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry.

County	Area of collection	Village	Species													Total
			<i>An. funestus</i> complex						<i>An. gambiae</i> complex							
			<i>An. funestus</i> s.s.	<i>An. lesoni</i>	<i>An. parensis</i>	<i>An. rivulorum</i>	<i>An. vaneedeni</i>	Unamplified	<i>An. arabiensis</i>	<i>An. gambiae</i> s.s.	<i>An. merus</i>	<i>An. quadriannulatus</i>	Unamplified	Total		
Kilifi	Burang'i		0	0	0	0	0	0	15	0	1	1	5	22		
	Garithe		0	0	0	0	0	0	69	0	232	0	6	307		
	Insectary		0	0	0	0	0	0	0	50	0	0	0	50		
	Marana		0	0	0	0	0	0	3	0	1	0	3	7		
	Mtondia		0	0	0	0	0	0	15	0	0	0	0	15		
	Sihu		6	0	0	78	0	16	0	0	0	0	0	100		
	Fihoni		135	0	2	42	4	13	1	0	0	0	0	197		
	Gazi		73	0	1	14	0	5	3	0	0	0	0	96		
	Jego		0	2	0	393	0	36	40	0	0	0	70	573		
	Kidomaya		0	0	0	33	0	2	4	0	0	15	0	54		
Kwale	Kikwezani		0	0	1	114	0	38	0	0	1	0	0	154		
	Kiwegu		0	0	0	24	0	5	1	0	0	0	1	31		
	Madongoni		75	2	0	31	0	21	0	0	0	0	1	130		
	Mangwei		0	0	0	13	0	0	1	0	1	2	0	17		
	Marigiza		6	1	2	229	0	6	3	0	4	3	4	258		
	Mukuduru		0	0	0	18	0	6	9	0	10	16	3	62		
	Mwanamamba		1	0	0	1	0	1	3	0	0	0	0	6		
	Tsuini		0	0	1	90	0	18	5	0	0	0	1	115		
	Kimundia		0	2	0	5	0	6	3	0	0	0	3	19		
	Kiwalwa		0	4	0	1	0	48	0	0	0	0	0	53		
Taita Taveta	Mwarusa		0	1	0	3	0	15	0	0	0	0	0	19		
	Njoro		0	15	0	15	0	17	0	0	0	0	0	47		
	Total		296	27	7	1,104	4	253	175	50	250	69	97	2,332		

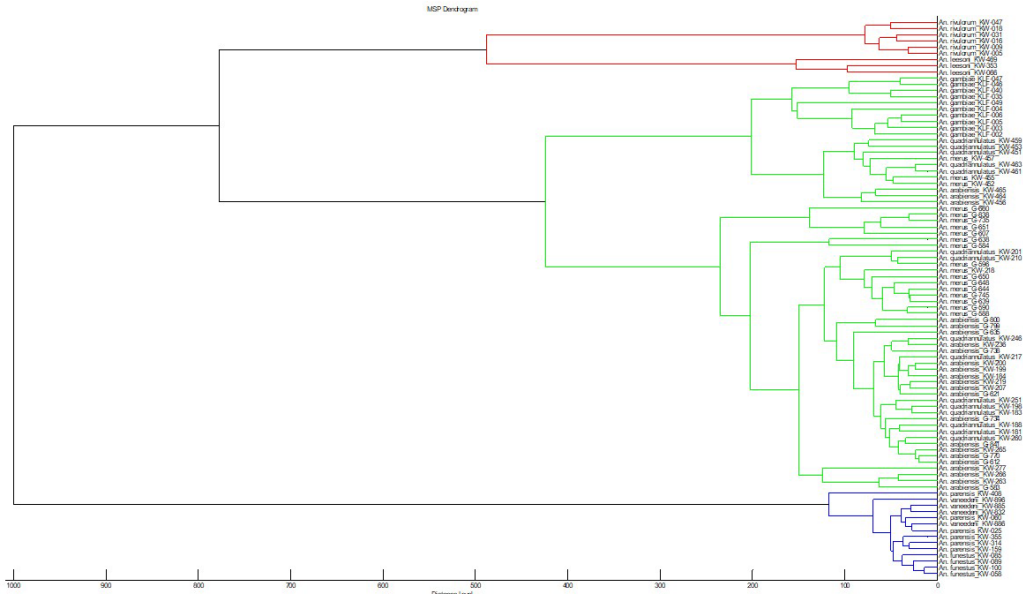


Figure 1. MSP dendrogram of all the MALDI-TOF MS spectra that were used for database creation for species identification as well as *An. vaneedeni* and *An. parensis*. The dendrogram was created using Biotyper v3.0 software and distance units correspond to the relative similarity of MS spectra. Abbreviations: MSP – main spectrum profile; MALDI-TOF MS – matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry.

23.1% for *An. arabiensis*, and 4.3% for *An. merus*. Large proportions of *An. merus* (56/70) and *An. rivulorum* (89/92) obtained their blood meals from goat and bovine, respectively (Table 3). None of the mosquitoes fed on chicken.

MALDI-TOF MS blood meal analysis

All the samples 223 samples analysed by ELISA were subjected to MALDI-TOF MS for blood meal sources identification. Of the 223 samples, 201 (90%) produced high quality and reproducible spectra that were subsequently used for database creation and blind testing (Table 4). The remaining 22 samples generated poor quality spectra or flatlines reflecting little or no proteins and were therefore excluded. Visual inspection of the spectra obtained different hosts using Flex Analysis revealed clearly distinct and highly reproducible spectral profiles within the four biological replicates spotted as well as different samples that had fed on the same host (Figure 3A). Moreover, principal component analysis on the spectra selected for database creation classified them as distinct (Figure 3B).

A total of 20 different spectra were used for database creation for each different blood host. A total 201 samples were queried against the database, yielding 93.8% (167/178) correct identification (Table 4). However, there were some misclassifications in 10 samples: four goat blood meals were identified as mixed meal of bovine and goat; two Bovine-goat blood meals were classified as Bovine; one Bovine-goat mixed blood meal was classified as goat; one human blood meal was classified as bovine; one human and human-goat blood meal could not be identified. The samples (n=3) in which blood meals were inconclusive by ELISA but had quality spectra were also queried against the database, yielding LSVs of the following

hosts: two bovine and one human (above the cut-off (LSV \geq 1.8)) (Figure 4).

Discussion

The study provides more information on the bionomics and diversity of the primary malaria vector and their trophic preferences along the coastal region of Kenya. This information is useful the designing targeted vector control interventions. Different sibling species belonging to *An. funestus* and *An. gambiae* complexes were reported in the region. Furthermore, the study detected *Anopheles cf. rivulorum* NFL-2015 a novel species that is for the first time reported in Kenya and has been implicated in malaria transmission in Eastern Zambia¹² (Additional file 3 in *Underlying data*³⁷). To be able to better our knowledge on the composition, diversity and bionomics of different vectors requires the use of cheaper, high throughput approaches.

The use of MALDI-TOF for entomological surveillance is for the first time being tested in Kenya by determining the identity of sibling species of the *An. gambiae* and *An. funestus* complexes and associated blood meal sources. To the best of our knowledge, this is the first-time members of the *An. funestus* complex have been characterized using MALDI-TOF MS. Moreover, the use of the k-nearest neighbor approach³⁴ to overcome the challenges^{38,42} of MALDI-TOF MS in distinguishing between members of *An. gambiae* complex has been discussed. This study further demonstrated the robustness and ability of MALDI-TOF MS in distinguishing between sibling species as well as blood meal sources of malaria vectors^{17,38,43,44}. The technology was shown to efficiently discriminate members of *An. gambiae* and *An. funestus* complexes, which

Table 2. Summarizes the number of samples of each sibling species that were used for MALDI-TOF MS database creation and validation as well as the accuracy of MALDI-TOF MS to identify each sibling species. Abbreviations: MALDI-TOF MS – matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry; LSV – Log score value; ID – Identification; K-NN – K-Nearest Neighbor; DB – Database.

Species	Sibling species	Total	Flatlines (Poor quality spectra)	DB creation	Validation	Low LSV	Ambiguous (Equivocal) ID	Correct ID	Accuracy (%)	Correct ID (K-NN)	Overall accuracy(%)
An. gambiae	<i>An. gambiae</i> s. s	50	0	10	40	0	0	40/40	100	40/40	100
	<i>An. arabiensis</i>	175	13	26	136	11	11	114/136	83.8	125/136	91.9
	<i>An. merus</i>	250	28	19	203	5	2	196/203	96.5	198/203	97.5
	<i>An. quadriannulatus</i>	69	4	14	51	5	13	33/51	64.7	46/51	90.2
An. funestus	<i>An. funestus</i>	296	12	6	278	16	0	262/278	94.2	262/278	94.2
	<i>An. rivulorum</i>	1104	41	6	1,057	6	0	1,051/1,057	99.4	1,051/1,057	99.4
	<i>An. leesonii</i>	27	4	6	17	1	0	16/17	94.1	16/17	94.1
Total		1,971	102	87	1,782	44	26	1,712/1,782	96.1	1,738/1,782	97.5

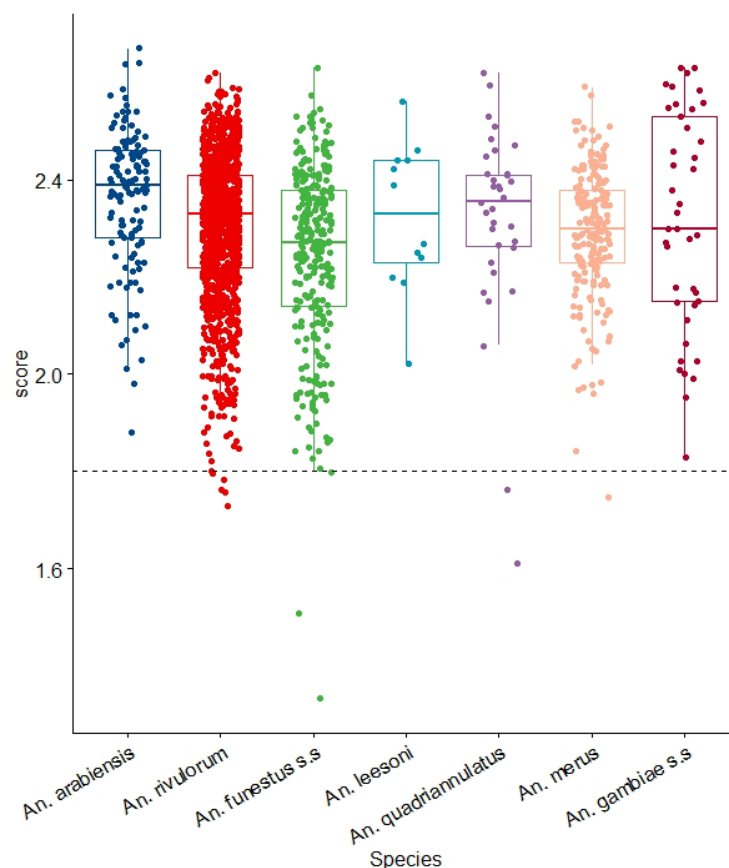


Figure 2. LSVs obtained after MSP reference database query with MS spectra belonging to members of *An. gambiae* and *An. funestus* complexes. Horizontal dashed lines represent the cut-off point for reliable identification (LSV > 1.8). Abbreviations: LSV – Log Score value; MSP – main spectrum profile; MS – mass-spectrometry; A.U. – arbitrary units; An. – Anopheles; s.s. – sensu stricto.

are the main malaria vectors in the region. The standard MALDI-TOF MS approach resulted in 96.1% (1,712/1,782) correctly identified, 1.5% (26/1,782) had ambiguous or equivocal identification and 2.5% (44/1,782) had low log score value. Misclassification could be as a result of closeness of sibling species of *An. gambiae* complex^{38,42}, potential sample degradation or confounding proteomic changes due to mosquito's life history, *i.e.*, physiological age progression: mating, blood feeding, oviposition^{44–49}. In the medical microbiology sector, MALDI-TOF MS has been unable to distinguish between closely related species such as *S. pneumoniae* and other *S. mitis* species group strains^{34,50,51}. To control for species misidentification, k-nearest neighbor approach was developed to calculate a weighted LSV (the sum of their LSV calculated by weighted their inverse position within the ranking list)³⁴. The species with the highest summation of the weighted LSV is considered as the probable species identification. An improvement of the database and the inclusion of the algorithms used to calculate closest matches have also been suggested as means of obtaining a more reliable distinction between closely related organisms³⁴. This K-nearest neighbor technique has been verified for species identification of closely related organisms and is being considered for use in research by Bruker Daltonics in discriminating between closely related organisms. In this current study,

the 26 samples that had ambiguous or equivocal identification were subjected to k-nearest neighbor method of analysis³⁴. Of note, using the weighted LSV, we were able to correctly classify all the samples that had ambiguous or equivocal identification as belonging to the *An. gambiae* complex (Additional file 2 in *Underlying data*³⁷). The application of both standard method for results interpretation and K-nearest neighbor algorithms led to correct identification with an overall accuracy of 1,738/1,782 (97.5%).

Spectra belonging to *An. pretoriensis*, *An. rufipes*, *Aedes africanus*, *Anopheles cf. rivulorum NFL-2015*, *Culex tritaeniorhynchus* were also queried using the MALDI-TOF MS database and were not identifiable as they were not represented in the database, confirming the specificity of the MALDI-TOF MS. Furthermore, some specimens could not be classified and presented with LSV scores below 1.8. The inability of the method to identify these specimens could have been attributed to residual blood meals in the head and thorax^{17,52,53}, as well as protein degradation during shipment, storage, and processing. To improve performance of the database, inclusion spectral profiles of different physiological status and chronological ages would be helpful to improve the performance of MALDI-TOF MS in species discrimination^{17,54}. Despite *An. parensis*

Table 3. Proportion of blood feeding sources among the mosquitoes collected along the coastal region of Kenya (Parenthesis indicate percentages).

	Bovine	Goat	Human	Human-Goat	Bovine-Goat	N/D
Total Tested	N=119	N=64	N=25	N=1	N=5	N=9
Sibling species:						
<i>An. arabiensis</i>	9 (7.6)	6 (9.4)	3 (12.0)	0 (0.0)	2 (40.0)	2 (22.2)
<i>An. funestus s.s.</i>	1 (0.8)	0 (0.0)	18 (72.0)	0 (0.0)	0 (0.0)	2 (22.2)
<i>An. lesoni</i>	0 (0.0)	0 (0.0)	0 (0.00)	1 (100)	0 (0.0)	0 (0.0)
<i>An. merus</i>	5 (4.2)	56 (87.5)	3 (12.0)	0 (0.0)	2 (40.0)	4 (44.4)
<i>An. parensis</i>	2 (1.7)	1 (1.56)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
<i>An. quadriannulatus</i>	2 (1.7)	0 (0.00)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
<i>An. rivulorum</i>	89 (74.8)	0 (0.00)	1 (4.0)	0 (0.0)	1 (20.0)	1 (11.1)
Unamplified	11 (9.24)	1 (1.56)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Village:						
Fihoni	2 (1.7)	1 (1.5)	10 (40.0)	0 (0.0)	0 (0.0)	2 (22.2)
Garithe	6 (5.0)	63 (98.4)	6 (24.0)	0 (0.0)	3 (60.0)	6 (66.7)
Gazi	0 (0.0)	0 (0.0)	4 (16.0)	0 (0.0)	0 (0.0)	0 (0.0)
Jego	42 (35.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)
Kidomaya	6 (5.04)	0 (0.0)	1 (4.0)	0 (0.0)	0 (0.0)	1 (11.1)
Kikwezani	3 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Kiwegu	4 (3.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Madongoni	7 (5.9)	0 (0.0)	3 (12.0)	1 (100)	0 (0.0)	0 (0.0)
Mangwei	0 (0.0)	0 (0.0)	0 (0.00)	0 (0.0)	1 (20.0)	0 (0.0)
Marigiza	36 (30.3)	0 (0.0)	1 (4.0)	0 (0.0)	0 (0.0)	0 (0.0)
Mtondia	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Mukuduru	10 (8.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Mwanamamba	3 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Location:						
Indoor	18 (15.1)	3 (4.7)	19 (76.0)	0 (0.0)	2 (40.0)	4 (44.4)
Outdoor	101 (84.9)	61 (95.3)	6 (24.0)	1 (100)	3 (60.0)	5 (55.6)

and *An. vaneedeni* being reported in the molecular assays, the sample size was not enough for creation of a MALDI-TOF MS reference database and validation.

Our results confirm that the cephalothorax is well-suited for MALDI-TOF-MS database setting for identification of mosquito sibling species^{17,42,43}. The advantage of using the cephalothorax is that the spectra obtained may have the potential,

provided databases are developed, for measuring other entomological endpoints that also reflect proteomic changes in that anatomical compartment *i.e.*, infection status²³ and age of mosquitoes^{55,56}. The application of this technique in distinguishing between infected *vs.* uninfected mosquitoes has been demonstrated using laboratory-reared and artificially infected mosquitoes²³. Currently there are ongoing studies evaluating the applicability and robustness of the technique in determining

Table 4. Summary of blood meal sources of mosquitoes identified by ELISA and MALDI-TOF MS. Abbreviations: MALDI-TOF MS – matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry; ID – Identification; DB – Database; N/D - Not detected; N/A - Not applicable.

Sibling species	Host (ELISA)	Total	Poor quality spectra	DB Creation	Validation	Correct ID (MALDI-TOF)	Accuracy (%)
<i>An. arabiensis</i>	Bovine	9	1	0	8	8	100
	Bovine-Goat	2	1	0	1	0	0
	Goat	6	1	2	3	3	100
	Human	3	0	1	2	2	100
	N/D	2	2	N/A	N/A	N/A	N/A
<i>An. funestus s.s.</i>	Bovine	1	0	0	1	1	100
	Human	18	1	4	12	11	91.7
	N/D	2	0	N/A	N/A	N/A	N/A
<i>An. lesoni</i>	Human-Goat	1	0	0	1	0	0
<i>An. merus</i>	Bovine	5	0	1	4	3	75
	Bovine-Goat	2	0	1	1	0	0
	Goat	56	11	5	40	36	90
	Human	3	0	1	2	2	100
	N/D	4	4	0	N/A	N/A	N/A
<i>An. parensis</i>	Bovine	2	0	0	2	2	100
	Goat	1	0	0	1	1	100
<i>An. quadriannulatus</i>	Bovine	2	0	0	2	2	100
<i>An. rivulorum</i>	Bovine	89	0	5	84	84	100
	Bovine-Goat	1	0	0	1	0	0
	Human	1	0	0	1	0	0
	N/D	1	0	0	1	1	100
Unamplified	Bovine	11	0	0	11	11	100
	Goat	1	1	0	0	0	0
Total		223	22	20	178	167	93.8

Plasmodium falciparum infection status of field-collected mosquitoes as well as age-grading.

In-depth understanding of malaria vector blood meal sources provides information on the risk of vector-borne disease transmission to humans. MALDI TOF MS also provided accurate identification of the blood meal sources irrespective of the sibling species, however, misclassifications were most likely in mixed blood meals.

Conclusions

Accurate and reliable species identification is indispensable as it informs control programs how different vector populations are being affected by interventions. MALDI TOF MS can allow mass screening of mosquitoes, although the approach is high-tech it entails simple lab procedures, permits the processing of hundreds of samples per day, and has a very low-cost per sample (below 0.2 USD). The novel approach could complement or even replace conventional methods for mosquito

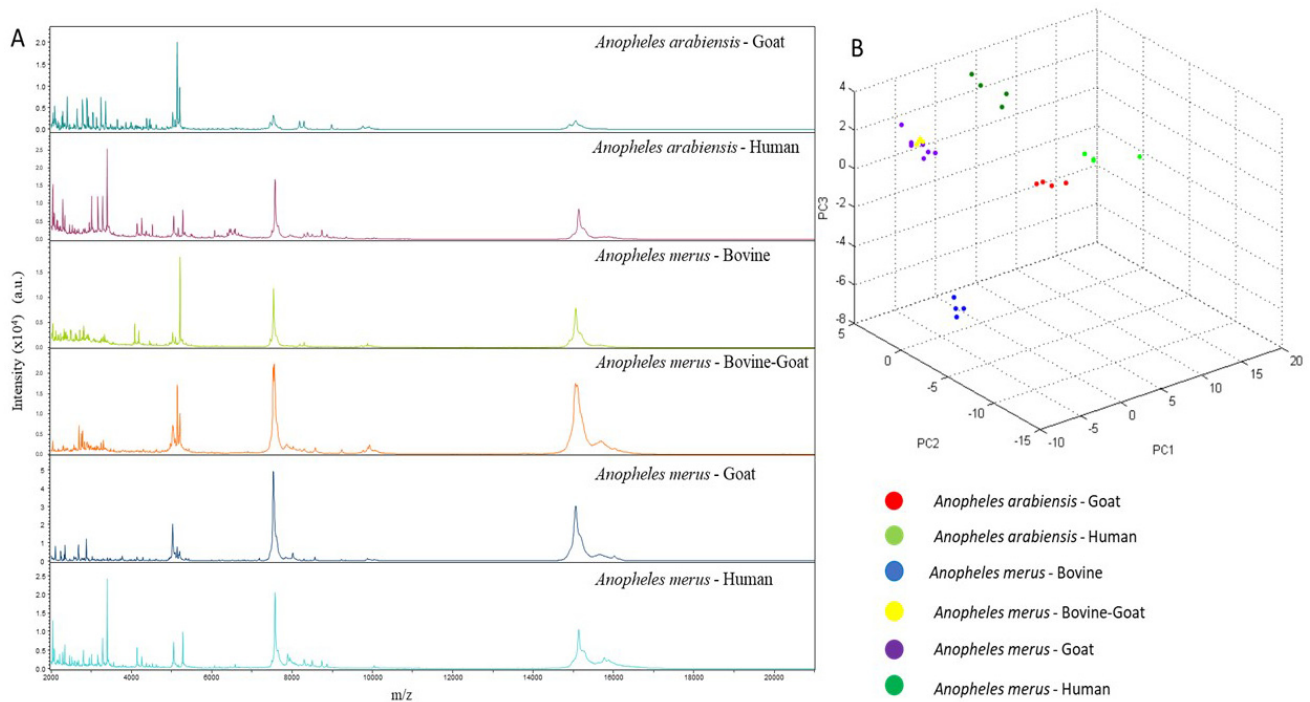


Figure 3. A. Comparison of representative MALDI-TOF MS spectra of different sibling species belonging to respective blood meal sources. B. PCA of MALDI-TOF MS spectra of sibling species of *Anopheles gambiae* and *An. funestus* complexes and their respective blood meal sources. Abbreviations: MALDI-TOF MS – matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry; PCA – Principal component analysis; a.u. – arbitrary units; m/z – mass to charge ratio.

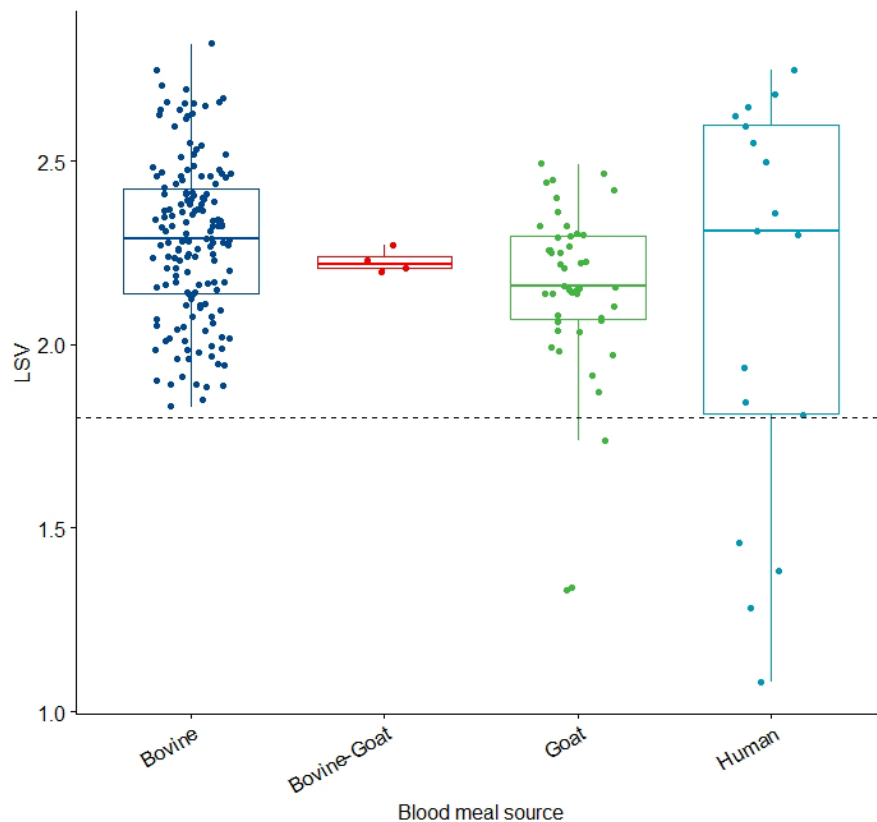


Figure 4. LSVs obtained during validation of the blood meal database created. Horizontal dashed lines represent the threshold value for reliable identification (LSV > 1.8). Abbreviations: LSV – Log Score Value.

species identification and blood meal determination, dramatically reducing costs and allowing surveillance programs to increase the number of samples and associated data resultant from field activities. Further research needs to be done to develop and evaluate databases for prediction of other entomological parameters of interest such as *Plasmodium* infection and age, which if successful could revolutionize entomological surveillance by creating a “silver bullet” assay whereby one test is able to inform various parameters.

Consent for publication

All the authors have reviewed and approved the publication of this paper. This paper has been published with the permission of the Director of the Kenya Medical Research Institute (KEMRI).

Data availability

Underlying data

Harvard Dataverse: Replication Data for: Utility of MALDI-TOF MS for determination of species identity and blood meal sources of primary malaria vectors on the Kenyan coast. <https://doi.org/10.7910/DVN/VYQFNO>³⁷.

This project contains the following underlying data:

- Utility of MALDI_TOF MS_Coast_dataset.tab
- Utility of MALDI_TOF MS_Coast_dataset_codebook.pdf
- Utility of MALDI_TOF MS_Coast_dataset_Readme.txt

- Additional file_1.pptx (Summarizes the calculation of weighted Log score value using k-nearest neighbor approach)

Additional file_2.docx (Show summary of the calculation of weighted LSV of a sub sample of the samples that had ambiguous/equivocal results)

- Additional file_3.docx (Provides a summary of sequencing results for the unamplified samples that were subjected to Sanger sequencing)
- Additional file_4.pptx (Assessment of *Anopheles gambiae* and *An. funestus* complex MS spectra reproducibility using composite correlation index (CCI). All the samples used for database creation in addition to *An. vaneedeni*, *An. parensis* and *An. lesoni* were subjected to analyses using the CCI tool. Levels of MS spectra reproducibility are indicated in red and blue revealing relatedness and incongruence between spectra, respectively.)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Acknowledgements

We are grateful to the Scientific and technical teams at the Centre for Geographic Medicine Research Coast, Kilifi for help in design and implementation of this work. Many thanks to the technical and field staff (Festus Yaa, Gabriel Nzai, Robert Mwakesi and Martha Muturi) who devoted their time and assisted in the field collection of samples and rearing mosquitoes in the insectary.

References

1. WHO: **World malaria report 2021**. 2021. [Reference Source](#)
2. WHO: **World malaria report 2020: 20 years of global progress and challenges**. 2020. [Reference Source](#)
3. WHO: **Global vector control response 2017-2030**. *Global vector control response 2017-2030*. 2017. [Reference Source](#)
4. Bamou R, Rono M, Degefa T, et al.: **Entomological and Anthropological Factors Contributing to Persistent Malaria Transmission in Kenya, Ethiopia, and Cameroon**. *J Infect Dis*. 2021; **223**(12 Suppl 2): S155–S170. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Mwangangi JM, Mbogo CM, Orindi BO, et al.: **Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years**. *Malar J*. 2013; **12**: 13. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Mwangangi JM, Muturi EJ, Muriu SM, et al.: **The role of *Anopheles arabiensis* and *Anopheles coustani* in indoor and outdoor malaria transmission in Taveta District, Kenya**. *Parasit Vectors*. 2013; **6**: 114. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Ogola EO, Fillinger U, Ondiba IM, et al.: **Insights into malaria transmission among *Anopheles funestus* mosquitoes, Kenya**. *Parasit Vectors*. 2018; **11**(1): 577. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Gillies MT, De Meillon B: **The Anophelinae of Africa south of the Sahara (Ethiopian zoogeographical region)**. 1968. [Reference Source](#)
9. Coetzee M: **Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae)**. *Malar J*. 2020; **19**(1): 70. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Koekemoer L, Kamau L, Hunt R, et al.: **A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group**. *Am J Trop Med Hyg*. 2002; **66**(6): 804–811. [PubMed Abstract](#) | [Publisher Full Text](#)
11. Scott JA, Brogdon WG, Collins FH: **Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction**. *Am J Trop Med Hyg*. 1993; **49**(4): 520–529. [PubMed Abstract](#) | [Publisher Full Text](#)
12. Lobo NF, St Laurent B, Sikaala CH, et al.: **Unexpected diversity of *Anopheles* species in Eastern Zambia: implications for evaluating vector behavior and interventions using molecular tools**. *Sci Rep*. 2015; **5**: 17952. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Beier JC, Perkins PV, Wirtz RA, et al.: **Bloodmeal identification by direct enzyme-linked immunosorbent assay (ELISA), tested on *Anopheles* (Diptera: Culicidae) in Kenya**. *J Med Entomol*. 1988; **25**(1): 9–16. [PubMed Abstract](#) | [Publisher Full Text](#)
14. Mwangangi JM, Mbogo CM, Nzovu JG, et al.: **Blood-meal analysis for**

- anopheline mosquitoes sampled along the Kenyan coast. *J Am Mosq Control Assoc.* 2003; **19**(4): 371–375.
[PubMed Abstract](#)
15. Bashar K, Tuno N, Ahmed TU, *et al.*: False positivity of circumsporozoite protein (CSP)-ELISA in zoophilic anophelines in Bangladesh. *Acta Trop.* 2013; **125**(2): 220–225.
[PubMed Abstract](#) | [Publisher Full Text](#)
 16. Seng P, Drancourt M, Gouriet F, *et al.*: Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Infect Dis.* 2009; **49**(4): 543–551.
[PubMed Abstract](#) | [Publisher Full Text](#)
 17. Nabet C, Kone AK, Dia AK, *et al.*: New assessment of *Anopheles* vector species identification using MALDI-TOF MS. *Malar J.* 2021; **20**(1): 33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Rakotonirina A, Pol M, Kainiu M, *et al.*: MALDI-TOF MS: optimization for future uses in entomological surveillance and identification of mosquitoes from New Caledonia. *Parasit Vectors.* 2020; **13**(1): 359.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Tandina F, Niaré S, Laroche M, *et al.*: Using MALDI-TOF MS to identify mosquitoes collected in Mali and their blood meals. *Parasitology.* 2018; **145**(9): 1170–1182.
[PubMed Abstract](#) | [Publisher Full Text](#)
 20. Yssouf A, Parola P, Lindström A, *et al.*: Identification of European mosquito species by MALDI-TOF MS. *Parasitol Res.* 2014; **113**(6): 2375–2378.
[PubMed Abstract](#) | [Publisher Full Text](#)
 21. Niare S, Berenger JM, Dieme C, *et al.*: Identification of blood meal sources in the main African malaria mosquito vector by MALDI-TOF MS. *Malar J.* 2016; **15**: 87.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 22. Tandina F, Niare S, Almeras L, *et al.*: Identification of mixed and successive blood meals of mosquitoes using MALDI-TOF MS protein profiling. *Parasitology.* 2020; **147**(3): 329–339.
[PubMed Abstract](#) | [Publisher Full Text](#)
 23. Laroche M, Almeras L, Pecchi E, *et al.*: MALDI-TOF MS as an innovative tool for detection of *Plasmodium* parasites in *Anopheles* mosquitoes. *Malar J.* 2017; **16**(1): 5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 24. Yssouf A, Almeras L, Terras J, *et al.*: Detection of *Rickettsia* spp in ticks by MALDI-TOF MS. *PLoS Negl Trop Dis.* 2015; **9**(2): e0003473.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 25. Diarra AZ, Almeras L, Laroche M, *et al.*: Molecular and MALDI-TOF identification of ticks and tick-associated bacteria in Mali. *PLoS Negl Trop Dis.* 2017; **11**(7): e0005762.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Tahir D, Almeras L, Varloud M, *et al.*: Assessment of MALDI-TOF mass spectrometry for filariae detection in *Aedes aegypti* mosquitoes. *PLoS Negl Trop Dis.* 2017; **11**(12): e0006093.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 27. Karisa J, Muriu S, Omuoyo D, *et al.*: Urban ecology of arboviral mosquito vectors along the Kenyan coast. *J Med Entomol.* 2021; **58**(1): 428–438.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 28. Munyao V, Karisa J, Munyao CM, *et al.*: Surveillance of Culicine Mosquitoes in Six Villages of Taita-Taveta County, Kenya, With Host Determinations From Blood-Fed Females. *J Med Entomol.* 2020; **57**(6): 1972–1982.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 29. Karisa J, Ominde K, Muriu S, *et al.*: Malaria vector bionomics in Taita-Taveta County, coastal Kenya. *Parasit Vectors.* 2022; **15**(1): 430.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 30. Kipyab PC, Khaemba BM, Mwangangi JM, *et al.*: The bionomics of *Anopheles merus* (Diptera: Culicidae) along the Kenyan coast. *Parasit Vectors.* 2013; **6**: 1–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. Kipyab PC, Khaemba BM, Mwangangi JM, *et al.*: The physicochemical and environmental factors affecting the distribution of *Anopheles merus* along the Kenyan coast. *Parasit Vectors.* 2015; **8**: 221.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 32. Kiuru CW, Oyieke FA, Mukabana WR, *et al.*: Status of insecticide resistance in malaria vectors in Kwale County, Coastal Kenya. *Malar J.* 2018; **17**(1): 3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 33. Boyer PH, Boulanger N, Nebbak A, *et al.*: Assessment of MALDI-TOF MS biotyping for *Borrelia burgdorferi* s.l detection in *Ixodes ricinus*. *PLoS One.* 2017; **12**(9): e0185430.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 34. Harju I, Lange C, Kostrzewa M, *et al.*: Improved differentiation of *Streptococcus pneumoniae* and other *S. mitis* group streptococci by MALDI Biotyper using an improved MALDI Biotyper database content and a novel result interpretation algorithm. *J Clin Microbiol.* 2017; **55**(3): 914–922.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 35. Bruker-Daltonik: MBT Explorer Module User Manual. 2016; 97.
 36. Nebbak A, Willcox AC, Bitam I, *et al.*: Standardization of sample homogenization for mosquito identification using an innovative proteomic tool based on protein profiling. *Proteomics.* 2016; **16**(24): 3148–3160.
[PubMed Abstract](#) | [Publisher Full Text](#)
 37. Karisa J, Ominde K, Tuwei M, *et al.*: Replication Data for: Utility of MALDI-TOF MS for determination of species identity and blood meal sources of primary malaria vectors on the Kenyan coast. V2 edition: Harvard Dataverse. [Dataset]. 2023.
 38. Müller P, Pflüger V, Wittwer M, *et al.*: Identification of cryptic *Anopheles* mosquito species by molecular protein profiling. *PLoS One.* 2013; **8**(2): e57486.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Hall T: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic Acids Symp Ser.* 1999; 95–98.
 40. R-Core-Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020.
 41. Bamou R, Costa MM, Diarra AZ, *et al.*: Enhanced procedures for mosquito identification by MALDI-TOF MS. *Parasit Vectors.* 2022; **15**(1): 240.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 42. Vega-Rúa A, Pagès N, Fontaine A, *et al.*: Improvement of mosquito identification by MALDI-TOF MS biotyping using protein signatures from two body parts. *Parasit Vectors.* 2018; **11**(1): 574.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 43. Briolant S, Costa MM, Nguyen C, *et al.*: Identification of French Guiana anopheline mosquitoes by MALDI-TOF MS profiling using protein signatures from two body parts. *PLoS One.* 2020; **15**(8): e0234098.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Zdobnov EM, Von Mering C, Letunic I, *et al.*: Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science.* 2002; **298**(5591): 149–159.
[PubMed Abstract](#) | [Publisher Full Text](#)
 45. Das S, Radtke A, Choi YJ, *et al.*: Transcriptomic and functional analysis of the *Anopheles gambiae* salivary gland in relation to blood feeding. *BMC Genomics.* 2010; **11**: 566.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 46. Félix RC, Müller P, Ribeiro V, *et al.*: *Plasmodium* infection alters *Anopheles gambiae* detoxification gene expression. *BMC Genomics.* 2010; **11**: 312.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 47. Kumar M, Mohanty AK, Sreenivasamurthy SK, *et al.*: Response to blood meal in the fat body of *Anopheles stephensi* using quantitative proteomics: toward new vector control strategies against malaria. *OMICS.* 2017; **21**(9): 520–530.
[PubMed Abstract](#) | [Publisher Full Text](#)
 48. Hugo LE, Monkman J, Dave KA, *et al.*: Proteomic biomarkers for ageing the mosquito *Aedes aegypti* to determine risk of pathogen transmission. *PLoS One.* 2013; **8**(3): e58656.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 49. Sikulu MT, Monkman J, Dave KA, *et al.*: Proteomic changes occurring in the malaria mosquitoes *Anopheles gambiae* and *Anopheles stephensi* during aging. *J Proteomics.* 2015; **126**: 234–244.
[PubMed Abstract](#) | [Publisher Full Text](#)
 50. Kärpänoja P, Harju I, Rantakokko-Jalava K, *et al.*: Evaluation of two matrix-assisted laser desorption/ionization-time of flight mass spectrometry systems for identification of viridans group streptococci. *Eur J Clin Microbiol Infect Dis.* 2014; **33**(5): 779–788.
[PubMed Abstract](#) | [Publisher Full Text](#)
 51. Ikryannikova L, Filimonova A, Malakhova M, *et al.*: Discrimination between *Streptococcus pneumoniae* and *Streptococcus mitis* based on sorting of their MALDI mass spectra. *Clin Microbiol Infect.* 2013; **19**(11): 1066–1071.
[PubMed Abstract](#) | [Publisher Full Text](#)
 52. Karger A, Kampen H, Bettin B, *et al.*: Species determination and characterization of developmental stages of ticks by whole-animal matrix-assisted laser desorption/ionization mass spectrometry. *Ticks Tick Borne Dis.* 2012; **3**(2): 78–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
 53. Yssouf A, Flaudrops C, Drali R, *et al.*: Matrix-assisted laser desorption/ionization-time of flight mass spectrometry for rapid identification of tick vectors. *J Clin Microbiol.* 2013; **51**(2): 522–528.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 54. Boyer PH, Almeras L, Plantard O, *et al.*: Identification of closely related *Ixodes* species by protein profiling with MALDI-TOF mass spectrometry. *PLoS One.* 2019; **14**(10): e0223735.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 55. Hugo LE, Quick-Miles S, Kay B, *et al.*: Evaluations of mosquito age grading techniques based on morphological changes. *J Med Entomol.* 2014; **45**(3): 353–369.
[PubMed Abstract](#)
 56. Sikulu M, Killeen GF, Hugo LE, *et al.*: Near-infrared spectroscopy as a complementary age grading and species identification tool for African malaria vectors. *Parasit Vectors.* 2010; **3**: 49.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 17 August 2023

<https://doi.org/10.21956/wellcomeopenres.21044.r63323>

© 2023 Halada P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **Petr Halada** 

Institute of Microbiology of the Czech Academy of Sciences, Průmyslová, Vestec, Czech Republic

The work represents a comprehensive study on species and bloodmeal identification of *Anopheles* mosquitoes originating from Kenya. The paper is well-written and the methods properly described. However, there are some drawbacks, especially in the presentation of bloodmeal identification results. I have several specific comments.

Did the authors observe any impact of storage conditions on spectrum quality (silica gel vs. direct freezing, storage time between specimen collection and MALDI-TOF MS)? Could the authors explain a significantly lower ID accuracy for *An. arabiensis* and *An. quadriannulatus* shown in the Table 2.

Page 4: The second part of the paragraph "Spectra acquisition" is partly redundant with the paragraphs "Spectra analysis" and "DTB creation and blind tests". The authors should rephrase those parts.

Page 6: The Mass-Up program is mentioned. Did the authors generate any data using this SW tool? What version of MALDI BioTyper was in fact used, 3.0 or 3.3.0? Some samples gave LSV below 1.8. Did the authors find any signals related to the residual blood in the respective protein profiles?

Did the authors see any differences between results obtained by ELISA and MALDI? In general, protein profiles from abdomen of engorged females are dominating by signals of blood hemoglobins. Could the authors comment why practically no signal of beta-hemoglobin is present in the Figure 3A.

Some imperfections could be found between the data on bloodmeal identification presented in the Tables, Figures and the text. For example, p.6 and Table 3, how was 50% for bovine host calculated? Page 6: "Of the 206 analysed" Why not 223? Table 3: 119 bovine hosts shown for sibling species, but 116 for village.

There are some discrepancies between Table 4 and Figure 3B. The Table 4 shows in total three *An. merus* – Human samples and two *An. merus* – Bovine-Goat samples. However, it seems that there is four *An. merus* – Human samples (dark green) and more than two *An. merus* – Bovine-Goat samples (yellow) in Fig. 3B. Moreover, the number of *An. funestus* s.s. – Human validated samples should be corrected in the Table 4.

Based on Table 4 and Page 8, paragraph “A total of 20 different spectra”, no mixed bloodmeals were correctly identified by MALDI-TOF MS. It is somewhat surprising that four Bovine-Goat hosts are depicted in the Figure 4. Page 8, first sentence: How was the HBI calculated for *An. arabiensis*?

Page 8, paragraph “A total 201 sample were queried” If 20 of 201 samples were used for DB creation, it means that 181 were subjected to MALDI for identification. However, both the following text and Table 4 works with 178 samples only.

In addition, I suggest calculating the overall accuracy for all analysed samples including those with poor quality spectra which might truly reflect the power and efficiency of the employed approach. Is the presented MS-based methodology able to distinguish hosts having hemoglobins with almost identical masses, e.g. human and horse with alpha- and beta-hemoglobins differing by 12 and 10 Da only, respectively.

The Discussion section is mainly devoted to mosquito species identification. The part referring to bloodmeal determination should be extended.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mass spectrometry of biomolecules; MALDI-TOF protein profiling; proteomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have

significant reservations, as outlined above.

Reviewer Report 18 July 2023

<https://doi.org/10.21956/wellcomeopenres.21044.r57657>

© 2023 Maurin M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Max Maurin

Universite Grenoble Alpes, Saint-Martin-d'Hères, Auvergne-Rhône-Alpes, France

This is an interesting study on the identification by MALDI TOF mass spectrometry of malaria vector mosquito species and their blood meal source in Kenya.

The methodology and results of the study are well presented, and the manuscript is of very good quality.

I only have a few minor comments.

Page 6, paragraph "From the entire collection, 2332 were subjected ...". About 20% of collected mosquitoes were identified by both morphological and molecular methods. Were these samples considered representative of the whole mosquito population? Was there any discordance between morphological and molecular identifications?

Page 6, paragraph "The remaining samples, 4.4% (87/1971) were used for database creation ...". Identification accuracy was 89.1% for *An. gambiae* complex and 98.3% for *An. funestus* complex. Although these accuracy scores are high, did the authors try to improve them by creating a database with more samples? Only 4.4% (87 samples) used for database building seems rather low to obtain accurate differentiation of such a high number of mosquito species. For closely related species belonging to the same complex, adding samples might have facilitated their differentiation.

Different parts of mosquitoes were analyzed for molecular identification or MALDI TOF MS and for blood source characterization (legs, wing, head, thorax, and abdomen). The authors could comment on the variations in MS spectra that could be observed depending on the type of mosquito sample analyzed. What would be the most appropriate samples to use in further studies?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Clinical microbiology, bacteriology, zoonoses, arthropod-borne diseases

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
