Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rapid host species identification from dried bloodspots of *Culex* mosquito blood meals using Mid-Infrared Spectroscopy

Erin Stewart Johnston
erin.johnston0456@gmail.com
University of Glasgow
Mauro Pazmiño-Betancourth
University of Glasgow
Mary Elizabeth Miranda
Field Epidemiology Training Program Alumni Foundation Incorporated
Brandy Alabado
Field Epidemiology Training Program Alumni Foundation Incorporated
Michael Ralph Tanedo
Field Epidemiology Training Program Alumni Foundation Incorporated
Mario González-Jiménez
University of Glasgow
Sarah Cleaveland
University of Glasgow
Troy Adam Laidlow
The University of Sydney
Michael Walsh
University of Sydney
Cameron E. Webb
University of Sydney
Mafalda Viana
University of Glasgow
Victoria J. Brookes
The University of Sydney
Francesco Baldini
University of Glasgow

Research Article

Keywords: Culex, blood meal analysis, mid-infrared spectroscopy, supervised machine learning, mosquito ecology

Posted Date: July 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9969287/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Understanding mosquito blood feeding preferences can help infer vector-borne pathogen transmission dynamics and improve intervention strategies for important human and animal pathogens. The gold-standard tests for identifying mosquito blood meals are expensive, time-consuming methods and decrease in accuracy as blood meals are digested. Mid-infrared spectroscopy with machine-learning (MIRS-ML) can offer a rapid, cost-effective alternative approach to identifying host blood meals but has never been used to analyse blood meals from *Culex* spp., an increasingly important genus of vectors of several high-priority multi-host pathogens such as orthoflaviviruses and filariasis.

Methods

In this study, laboratory-reared *Culex quinquefasciatus* were fed blood from six hosts (human, pig, chicken, cattle, dog and horse). Dried bloodspots were created 6, 12, 24 and 30 hours after digestion in the mosquito by rolling abdominal contents onto filter paper and then scanned using a Bruker ALPHA Fourier-transform infrared spectrometer. The spectra generated were used to train a multilayer perceptron algorithm to predict blood meals from these six host species and then tested on field-collected mosquito blood meals.

Results

For laboratory-reared mosquitoes, the model was 93.6% accurate in predicting host species in comparison to 16.7% random classification, with a small decline in accuracy after 30 hours of digestion. While the accuracy for field-collected bloodspots was lower (67.9%), this likely reflects differences in blood source from wild mosquitoes, such as no anticoagulants and direct feeding from hosts rather than artificial feeders. Model performance was further enhanced by applying a probability threshold, which increased accuracy to 78% in field samples while removing 21% samples.

Conclusions

This study demonstrates the potential application of MIRS-ML with dried bloodspots as a practical, scalable tool for blood meal analysis with scope for improving accuracy through further model development. This method may be of particular use for large-scale surveillance programmes in low-resource settings due to low running and labour costs. Further development of this method using samples from field-collected mosquitoes could enhance classification accuracy and expand the range of identifiable host species.

Background

Vector-borne pathogens reportedly cause 17% of all human infectious diseases cases and over 700,000 human deaths every year (1), and most are transmitted by mosquitoes. For pathogen transmission to

occur, a mosquito needs to feed from an infected host and later feed on a susceptible host, passing on the pathogen. Mosquito feeding patterns and preferences vary between species. For example, *Anopheles arabiensis*, a key malaria vector in West Africa, has a narrow host range preferring cattle but also feeding on humans (2), while *Culex tarsalis*, a key vector of West Nile virus (WNV) in North America, is a generalist, with one study identifying 36 host species from only 69 blood meals (3). Interestingly, intraspecific variation in feeding preference also follows in response to host availability (4). For example, *Cx. tarsalis* has been shown to feed from more mammals than birds later in the summer as ardeid seasonal birds migrated (5). For mosquitoes known to be competent vectors, understanding host-feeding patterns can, therefore, inform pathogen transmission dynamics and control strategies can subsequently be adapted to mosquito feeding behaviour (6). One example of this is zoophylaxis, where animals are used to divert mosquitoes from human hosts, this has worked well for the zoophilic species *An. arabiensis* in Africa, reducing malaria risk when cattle are present (7,8). However, contrasting evidence has been shown in other areas, such as Pakistan where cattle presence increased malaria risk (9,10). Therefore, mosquito feeding patterns should be studied in a variety of geographically and temporally diverse ecological niches because pathogen spread may vary with mosquito behaviour under specific environmental or climatic conditions.

Japanese encephalitis virus (JEV) is another vector-borne pathogen in which transmission is dependent on host species, and an estimated 1.5 billion (± 250 million) people are considered at risk of infection (11). The virus is known to infect a wide range of avian and mammalian species (12), so understanding host feeding behaviour of mosquito vectors is critical to identify the role of different hosts as potential sources of JEV infection in humans. *Culex* spp. mosquitoes are considered the primary vectors of JEV, with ardeid birds thought to act as maintenance hosts, humans, horses and cattle as dead-end hosts, and pigs as an amplifying host (13). However, recent research has questioned these roles (14,15), especially for pigs as regions with little pig farming continue to have JEV outbreaks, implicating other avian or mammalian hosts in transmission of the virus (14–17). In particular, domestic poultry have been shown to be seropositive for JEV, with chickens reaching peak viraemic levels at 2–3 days post-infection and younger chicks with a higher peak viraemia capable of transmitting virus (18–20).

Given the complex transmission cycle of JEV and the potential roles of many different host species, deploying a method that establishes ecological interactions between hosts and vectors could substantially improve the understanding of JEV transmission in endemic countries. In addition, *Culex* mosquitoes can act as vectors for other pathogens of human and veterinary significance with distinct transmission cycles such as WNV, Murray Valley encephalitis virus, Ross River virus, Barmah Forest virus, Usutu virus, Rift Valley fever virus, *Dirofilaria* spp. and avian *Plasmodium* spp. (21). Knowledge of *Culex* spp. blood meal preferences may help infer disease dynamics across this variety of pathogens.

A previous blood meal meta-analysis by Griep et al. (22) of 29,990 blood meals from 70 different *Culex* species globally showed a wide diversity in host preferences, varying regionally from avian to human feeding preferences, with no correlation between *Culex* phylogeny and diet. In addition, despite being thought of as preferentially biting avian hosts, the study showed that only 3.28% of *Culex*

tritaeniorhynchus fed on avian sources, whilst 83.22% of *Culex pipiens* spp. and 34.37% of *Culex quinquefasciatus* (Diptera: Culicidae) did so (22). A better understanding of *Culex* spp. regional blood meal feeding preferences is needed, especially in areas where these species play an important role in outbreaks of human and animal diseases.

Many molecular methods have been developed to identify host blood meal sources, using DNA- or antigen-based methods. Antigen-based tests include latex agglutination (23), precipitin tests (24) and ELISA (25). Latex agglutination and precipitin tests were historically important methods that are rapid, cheap, and easy to interpret but have a lower species sensitivity and generalisability (26). However, ELISA methods are more sensitive and specific but rely on host-specific antibodies being produced, thus limiting their generalisability (26).

DNA-based methods include PCR followed by gene sequencing (27), restriction fragment length polymorphism (RFLP) (28) and qPCR (29). Typically, PCR of mitochondrial DNA genes such as Cytochrome c Oxidase Subunit I (COI) and Cytochrome b (Cyt b) is used as a reference for identifying host blood meal sources (26). RFLP is less expensive than sequencing and can differentiate between host species but is not as generalisable because appropriate enzymes must be matched to expected hosts, thus requiring prior knowledge of available hosts. PCR and gene sequencing is highly sensitive and specific, even down to species level. In addition, when non-specific primers are used, it can reveal any host and therefore does not require prior knowledge of host availability. Whilst DNA remains stable longer than antigens at room temperature, degradation due to mosquito blood meal digestion can inhibit DNA amplification (30,31). Whilst qPCR can resolve this, it is less generalisable because it requires the design of specific primers and probes. Overall, DNA-based methods are much more species specific and sensitive than antigen but require repeated procurement of reagents, are time-consuming and require specialist molecular laboratory capacity and training.

Non-molecular techniques can overcome some of the challenges of molecular methods (33) and provide an alternative approach for identifying host blood meal sources, especially in low-middle income countries (LMIC), where budgets and laboratory capacity are typically more constrained, and vector-borne disease is most prevalent and devastating (32). One such method is mid-infrared spectroscopy paired with machine-learning models (MIRS-ML). In brief, this technique measures biochemical bonds in samples using MIRS. This creates a spectra for each sample, from which ML can create a biochemical fingerprint unique to individual host species. MIRS has been used in a variety of biochemical analyses such as food quality analysis (34), soil analysis (35) and clinical applications such as cancer detection (36). This approach has previously been successfully used for vector-borne disease surveillance by directly measuring dry mosquitoes for age-grading *Anopheles* species and *Aedes albopictus* (37,38) and detection of *Plasmodium falciparum* in *Anopheles funestus* (39) with accuracy above 80%. Additionally, it has been used to determine blood meal sources in *An. arabiensis* and *An. funestus* (31,40). While blood meal identification accuracy was high (> 90%), performance declined with prolonged preservation, highlighting the need for improved long-term solutions. To date, the technique has not been used for *Culex* mosquitoes, a generally understudied genus with highly varied diet and of increasing concern for

animal and public health. It also has not been used for dried blood spots on filter paper (DBS) which provide an alternative long-term storage of samples, mitigating problems with prolonged preservation of dried mosquitoes.

The current study aimed to develop a MIRS-ML approach to identify host species from blood meals in *Culex* mosquito vectors. The study comprised a laboratory phase, which built an ML model using a colony of *Cx. quinquefasciatus* fed on different host species and blood meals preserved on filter papers, and a second field phase, which validated the laboratory-based model on field-caught blood-fed mosquitoes from the Philippines and Australia.

Methods

Study overview

This study comprised two phases: an experimental phase, in which colony-reared *Culex* mosquitoes were fed blood from known species to assess the feasibility of MIRS-ML for host blood meal species identification under controlled conditions; and a field phase, in which blood-fed mosquitoes collected from two regions in Australia and the Philippines were used to validate MIRS-ML in natural settings.

Laboratory MIRS-ML model development

Mosquito Rearing

Culex quinquefasciatus were obtained from a colony housed in Medical Entomology, NSW Health Pathology, Westmead Hospital. The colony was established in 1998 and subsequently augmented when required, from field-caught specimens collected from the Sydney region (New South Wales, Australia). Prior to testing, mosquito larvae were reared at 26 ± 2°C (80 ± 5% humidity) on a 12-hour light cycle and provided commercial fish food. Adult females were held in cages for one week and provided with a 10% sucrose and water mixture but were not given access to any blood source. For experiments, 150 females were aspirated into four separate cages and starved of sugar overnight.

Blood with citrate (3.2%) added as an anti-coagulant was fed to the mosquitoes using a Hemotek® membrane blood feeding system (41). Animal blood was purchased from Cardinal Bioresearch Pty Ltd (42) and human blood was purchased from Australian Red Cross Lifeblood (43). To ensure females had sufficient time to feed, blood was available for at least 45 minutes and mosquitoes were disturbed 2–3 times during the feeding session to ensure as many mosquitoes as possible had access to the feeder. Mosquitoes were then retained under the same laboratory conditions as the larvae described above.

To generate samples for testing the accuracy of MIRS-ML to identify host species from blood meals over time, samples were created at 6, 12, 24 and 30-hours post-feeding for each host species. On each occasion, a cage was removed from the insectary and placed in a freezer for up to 15 minutes to

immobilise the mosquitoes. Unfed females and male mosquitoes were discarded. For blood fed specimens, abdominal contents from fed females were squashed onto Whatman™ 903 Proteinsaver Sample Collection Cards (Merck Life Science UK Limited, Dorst, United Kingdom) using individual pipette tips, to create dried blood spots (DBS). Cards were stored at room temperature until analysis.

While cellulose from the Whatman™ 903 Proteinsaver Sample Collection Cards produces characteristic infrared peaks that overlap with blood spectral features (particularly at 3300, 2900, 1650, and 1050 cm^{-1}), the consistent use of identical card substrates across all samples minimised the influence of the systematic variation from this source, allowing the ML algorithms to effectively distinguish host-specific signals from this consistent background (44,45). Additionally, 50 scans of Whatman™ 903 Proteinsaver Sample Collection Cards were taken to compare spectra absorbance to DBS to ensure any interference could be removed.

Mid-infrared spectrometry and machine learning

DBS were analysed using a Bruker ALPHA II Fourier-transform infrared (FTIR) spectrometer equipped with a Platinum ATR device (Bruker Optics), using Bruker OPUS (version 2.3). Spectra of each DBS were obtained at a resolution of 5 cm^{-1} over the 400–4000 cm^{-1} range, with 32 interferograms co-added for both background and sample measurements prior to Fourier transformation. The anvil was deployed to ensure no space between the DBS and the platinum diamond sampling module. Every 5 samples, background measurements of air on top of the diamond were taken exactly the same way to reduce the interference of background CO_2.

No spectral pre-processing was conducted but spectral quality from DBS was assessed using in-house Python™ programme scripts before ML training (see Software availability section). Briefly, this was achieved by removing spectra that were either flat (no features), had high absorbance around the CO_2 region (3500 to 4000 cm^{-1}), were distorted by the anvil, or were low intensity (absorbance at 400 cm^{-1} below 0.11 absorbance units). Prior to supervised ML model training, two spectral regions were excluded. These regions corresponded to absorbance peaks associated with the Whatman™ 903 ProteinSaver sample collection cards and atmospheric CO_2 to prevent confounding the model with background noise.

The resulting spectra were analysed using supervised ML models developed in the Python™ programming language with the *scikit-learn library* (47). Supervised ML models were trained on a dataset containing all four timepoints grouped together and labelled with host type. To ensure that the final model was tested on unseen spectral data, the dataset was stratified and split with 80% for training of models and 20% for validation.

The training dataset was then stratified further (whilst still preserving the percentage samples for each class), into 80% training and 20% testing spectra to select the best performing models before optimising the model. Seven different models were tested: Logistic Regression (LR), Support Vector Machine (SVM),

K-nearest neighbour (KNN), Gaussian naïve Bayes (NB), Random Forest Classifier (RF), XGBoost and Multilayer Perceptron (MLP)(47). The best performing model returned the highest percentage of correctly predicted class (e.g. host species) on the 20% testing spectra. This model was then optimised using hyperparameter tuning via a grid search with 5-fold stratified cross-validation, with balanced accuracy as a scoring method. Model performance across digestion time for each host species was analysed by fitting a binomial general linearised model (GLM) using the host and timepoint as the independent variables and correct prediction as the dependent variable. Finally, the model was validated using the unseen test spectra (20%) from the laboratory DBS, withheld at the beginning of analysis.

Model Probability Score

To assess confidence in the MIRS-ML model predictions, class probabilities were calculated by using the *predict_proba* function from the *scikit-learn library* (47). Since MLP was the best-performing model, this was only applied to MLP class predictions. In MLP models this function applies the softmax transformation to the MLP's raw output (logits), converting them into a normalised probability distribution across all possible classes (48). Specifically, for each DBS, a probability (0–1) was assigned to each possible host species, with all probabilities summing to one. The host with the highest probability was selected as the predicted class label for the DBS, and the corresponding highest probability served as the model probability score, with values closer to one indicating stronger model certainty for predicted host class.

To assess whether a correct prediction was significantly associated with the ML model's probability score, a binomial GLM was fitted to the predicted probability score and whether the prediction was correct or incorrect was used as dependent variable. A second binomial GLM was used to test if host classes that were included (known) or not included (unknown) in the training set had a different probability score; in this model, the ML model probability score was the independent variable and the known/unknown host class the dependent variable. Another application of model probability score would be applying a threshold to filter out incorrectly labelled DBS when PCR result is not known. To test if this would be useful the threshold at which the highest model accuracy is achieved, and highest number of samples retained was calculated. This threshold was then applied to DBS from field-caught Philippines mosquitoes with known hosts.

Field MIRS-ML model validation Field-collection of mosquitoes

To validate the model in natural settings two spectral datasets from blood-fed mosquitoes were collected from the Central Luzon and Cordillera Regions in the Philippines and from Queensland in Australia. PCR and gene sequencing was used as a reference method for identifying host blood meal sources.

Philippines

Carbon dioxide (CO_2) baited BG-Pro traps (Biogents AG, Regensburg, Germany) were deployed at 21 trap sites in Central Luzon and the Cordillera Administrative Region, Philippines in November 2023 (rainy season), August 2024 (rainy season) and February 2025 (dry season). A total of 21 sites was chosen for mosquito trapping in Region III Central Luzon and Cordillera Administrative Region of the Philippines (Fig. 1). At each site, two BG-Pro traps were placed overnight, one baited with yeast solution (200g of sugar was mixed with 10g of yeast and 500ml of warm water) and the other with bicarbonate soda and citric acid (49) as a source of CO_2. For the latter traps, 500g of citric acid was mixed with 500ml of water and transferred to a saline drip bottle. The IV valve was set to drip at two drips/s into an 800ml plastic bottle filled with 250g of bicarbonate soda.

Traps were set and collected at approximately 4pm and 7am the next day, respectively. Captured mosquitoes were then placed in a freezer for up to 30 minutes to immobilise them. Species identification was conducted using ID guides (50,51). Mosquito blood meals were preserved by creating DBS using the same method described above.

Australia

Mosquitoes were collected at 7 trap sites in the Northern Peninsula Area in Queensland, Australia for 10 days from 2023 to 2024. These collections were March 2023 (rainy season), June 2023 (dry season), September 2023 (dry season) and April 2024 (rainy season). At each site, two CDC-style light traps (Pacific BioLogics, Queensland, Australia) were deployed with one using UV light as a bait and the other using CO_2 gas from a tank released at 300ml/min.

Traps, mosquito handling and DBS production were conducted as described for the Philippines study sites. Mosquito species identification was conducted using a pictorial guide (52).

Analysis of field-caught blood fed mosquitoes

DBS were analysed at the University of Glasgow to identify the source of their blood-meals using MIRS-ML, as previously described. After MIRS sampling, DNA was extracted from field-caught DBS using DNeasy® Blood & Tissue kit (Qiagen, Hilden, Germany) according to manufacturer instructions for animal tissues. DNA was eluted in 50µL water to increase final DNA concentration.

To identify the host species of the mosquitoes' blood meals, a universal vertebrate primer set was used to target the cytochrome c oxidase subunit 1 (COI) (27). COI long primers (COI long-F: 5'-AACCACAAAGACATTGGCAC-3' and COI long-R: 5' AAGAATCAGAATARGTGTTG-3') were used in a 25µl reaction with 12.5µl GoTaq® G2 Hot Start Master Mix (Promega), 0.5µM of each primer, and 2µL of template DNA. The thermal cycle was 95°C for 1 minute, followed by 35 cycles at 95°C for 30 seconds, 50°C for 50 seconds and 72°C for 1 minute and a final step at 74°C for five minutes.

After PCR amplification, Sanger sequencing was outsourced to Eurofins Genomics (Eurofins Genomics Europe Shared Services GmbH). The obtained sequences were then compared to the NCBI nucleotide database using a Basic Local Alignment Search Tool (BLAST) search.

Results

Laboratory MIRS-ML model development

A total of 2,561 spectra were created from scanning DBS from mosquitoes fed on six different host species' blood (chicken, cattle, pig, horse, human, dog) and collected at varying timepoints post-feeding. A total of eight spectra were discarded for having high CO_2 measurements or for anvil distortion. Table S1 shows the number of DBS for each timepoint and host.

The mean spectral pattern of each host and Whatman™ 903 Proteinsaver Sample Collection Cards for all timepoints combined is shown in Fig. 2. The spectra have 6 major absorbance peaks at the following chemical bonds N-H / O-H stretching (3300 cm^{-1}), C-H stretching (2900 cm^{-1}), C = O stretching [Amide I band] (1650 cm^{-1}), N-H bend and C-N stretching [Amide II band] (1540 cm^{-1}), C-O Stretching (1040 cm^{-1}) and Fe-N or Fe-O stretch (520cm^{-1}), highlighted in Fig. 2 (45,53).

Differences between hosts or timepoints in the average spectra could not be identified by eye and there was a great variability in the absorbance bands of individual spectra for each host and timepoint (Figure S1). Scanned Whatman™ 903 Proteinsaver Sample Collection Cards without DBS had a large peak in absorbance in the C-O stretching region at 1140cm^{-1} which overlapped with DBS absorbance, especially at the 30-hour post digestion for horse blood (Figure S1). To prevent spectral interference in the ML, this region (wavenumbers 900 to 1100) and the CO_2 region (wavenumbers 1800–2200) were removed from training.

Of the seven different ML models tested using the training dataset, multilayer perceptron (MLP) was the most accurate (Fig. 3A) at 87% accuracy. Therefore, model refinement was further completed on MLP only.

Overall, MIRS-ML was 93.6% accurate for predicting species in the experimental validation dataset (using the optimised MLP model) (Fig. 3B). Accuracy ranged from 89% (dog) to 97% (pig). When considering timepoint as well in the binomial GLM, human was the most accurately classified host overall, and significantly associated with higher odds of correct classification relative to pigs (OR = 3.5, CI = 1.03–12.2, P = 0.046). Model performance declined at 30 hours post-feeding, with the odds of correct classification reduced relative to earlier timepoints (OR = 0.32, CI = 0.11–0.98, P = 0.046), with this timepoint exhibiting the lowest classification accuracy (Table 1). In addition, comparison of averaged infrared spectra revealed clear differences in absorption between correctly classified and misclassified DBS (Figure S2).

Table 1
Percentage accuracy of MIRS-ML model classification of host blood meal from laboratory-reared *Cx. quinquefasciatus* over time by host.

Time (hours)	Accuracy (%)					
	Chicken	Cattle	Dog	Horse	Human	Pig
6	92.9	100	100	92.6	96.9	100
12	100	95.2	90.9	100	90	92.1
24	92	90.5	100	86.4	100	93.8
30	100	79.3	100	100	96.6	70.6

Field MIRS-ML model validation

In the Philippines a total of 4,902 mosquito specimens were collected across the three sampling periods with *Cx. quinquefasciatus* being the most abundant species. A total of 268 blood fed specimens were collected from this sample and DBS produced. Of these, 78 samples could not be confirmed by PCR because host DNA could not be amplified, and two DBS were discarded due to anvil distortion during scanning. Included in the remaining 188 samples were 2 DBS from *Ae. albopictus*, 8 from *Armigeres subalbatus* mosquitoes and 178 DBS from *Cx. quinquefasciatus* mosquitoes. One mosquito had fed from a host that had not been included in the MLP model training (goat) so was also removed from subsequent analysis (this had been identified as human by MIRS-ML).

187 DBS were amplified and sequenced successfully, and these were subsequently used to assess MIRS-ML model accuracy in predicting host species from blood meals. Most mosquitoes had fed from chickens (n = 166), pigs (n = 9), humans (n = 6) and dogs (n = 4) and the MIRS-ML model was overall 67.9% accurate in identifying host blood meal (Fig. 4). Chicken was the most correctly labelled host source (72%) and pig was the most mislabelled (45%).

In Australia, a total of 24,000 mosquito specimens were collected during the one-year sampling period with *Aedes kochi* being the most abundant species and a total of 28 blood fed specimens collected. 15 DBS came from *Ae. kochi*, nine from *Culex sitiens* subgroup, two from unknown *Anopheles* species, one from *Aedes tremulus* and one from an unknown *Aedes* species. Of these 28 DBS, 23 were successfully amplified for host identification. The majority of hosts identified by sequencing were wallabies (n = 14), rats (n = 3) and bandicoots (n = 3) and only three DBS came from hosts on which the model had been trained on (2 cattle and 1 horse). MIRS-ML correctly identified one cow DBS, and incorrectly labelled the other cow DBS as human and the horse DBS as dog.

Model Probability Score

Laboratory Test Dataset

To assess whether model probability score reflects prediction reliability, the MIRS-ML probability score was evaluated using the 515 DBS spectra from the laboratory test dataset. The binomial GLM indicated that MIRS-ML probability score for correctly classified hosts was significantly higher (0.95) than incorrectly classified hosts (0.63) ($p < 0.001$, $X^2 = 92.98$) (Fig. 5A), suggesting the model assigns lower confidence to incorrect predictions, indicating model probability score could be a meaningful measure of prediction reliability.

Field Datasets

Similar to the laboratory results, binomial GLM analysis using 187 DBS spectra from Philippines indicated that higher MIRS-ML probability score was associated with increased odds of correct classification ($p < 0.001$, $X^2 = 44.36$). The mean probability score for correct predictions was 0.94 and incorrect prediction was 0.78 (Fig. 5B).

We then tested if the MIRS-ML probability score was different between host species that were included (known) or not (unknown) in the training set. Analysis using a binomial GLM from 294 DBS spectra of all field data showed that MIRS-ML model probability score was positively associated with the host being known ($p = 0.004$, $X^2 = 8.48$). Figure 5C shows the mean probability of known hosts for 190 DBS (0.87) and unknown hosts for 104 DBS (0.80), further suggesting that the model's ability to decipher spectra that was not trained on is lower than those included in the training.

Given lower model probability score was significantly associated with incorrect predictions, we evaluated whether applying a probability threshold could improve classification performance. The threshold in which both MLP model accuracy and sample retention was highest was 0.73 (Figure S3). After applying this to DBS from field-caught Philippines mosquitoes with known hosts, overall model accuracy increased by 13% (78%). This was achieved by filtering out 40 samples (12 correctly classified and 28 incorrectly classified). The majority of these were from chicken DBS (n = 35) followed by pig (n = 3) and human (n = 2). When applied to Philippines samples with unknown hosts (n = 78), 26 were filtered out. Predicted hosts for the remaining 52 DBS were; chicken (n = 24), human (n = 16), dog (n = 7), pig (n = 4), and horse (n = 1).

Discussion

This is the first application of MIRS-ML methods to identify mosquito host blood meal sources from DBS and from *Culex* mosquitoes. The three key findings are: a) MIRS-ML was 93.6% accurate in predicting host blood meal source from laboratory-reared *Cx. quinquefasciatus* DBS and 67.9% accurate in distinguishing field-caught mosquito DBS from known hosts; b) MIRS-ML model can predict host at late-stage digestion; and c) increased model probability score was associated with correct label prediction and may be useful as a threshold for filtering incorrectly identified hosts.

The MIRS-ML model used an optimised MLP algorithm as it was the most accurate, this is consistent with MLP being a high performing algorithm for identifying blood meals from malaria vectors (31,40),

and may be due to the algorithm's ability to model more complex patterns and data interactions than simpler models like logistic regression or Gaussian naïve Bayes, especially as this model has multi-class classifications (54). In addition, MLP models are known to be suited to data with many features (55) which is the case with this dataset as after removal of between 900–1100 and 1800–2200, 1473 wavenumbers remain as features for model training. When applied to field-collected DBS with confirmed known hosts MIRS-ML model decreased in accuracy (correct identification of 67.9% of mosquito host blood meals; a decrease of 25.7%) but was still substantially higher than expected random prediction (16.7%). This indicates a strong signal for each host type from the MIRS spectra. Furthermore, the model was able to generalise to highly variable field samples that differed significantly to the training set in the following three ways; mosquitoes had fed from an animal rather than a membrane, the blood did not contain anticoagulant, and the digestion time was unknown. Additionally, the field-caught mosquitoes from the Philippines included 10 DBS from *Aedes* and *Armigeres* mosquitoes and the MIRS-ML model correctly predicted 6 out of 10 of these. This indicates that the MIRS signal from DBS could be trained on one genus of mosquito and used to predict from other genus, making it simpler to create MIRS-ML methods for a variety of mosquito-borne diseases. Similarly, previous research has shown that MIRS-ML trained on laboratory data using *An. arabiensis* had 76% accuracy in identifying blood meals from a field-collected different mosquito species (*An. funestus)* (31), further evidencing that MIRS is robust to changes from the training dataset to testing on field samples.

This study demonstrates the potential for MIRS-ML to overcome several challenges inherent in analysis of variable field samples collected in real-world settings. The use of Whatman™ 903 Proteinsaver Sample Collection Cards proved highly effective for fieldwork, because they allowed for room-temperature storage and eliminated the need for specialised equipment. This enables sample collection in remote or logistically challenging locations. The average storage to analysis timeline was 1–3 months, demonstrating the longevity of this storage method. Whatman™ Proteinsaver Cards have been validated for DNA preservation up to 9 years (56).

Previous research has shown that digestion time reduces the ability to amplify DNA via PCR for blood meal digestion in five different genera of mosquitoes (30). This is further evidenced by our research as despite using a universal primer, 29.6% of DBS could not be analysed using PCR and Sanger sequencing. This could be due to a number of factors such as DNA degradation during the digestion process (57), or multi-host feeding which could interfere with Sanger sequencing (58). When these DBS amplified during PCR, the band was often faint and could not be sequenced. Comparatively, MIRS-ML could predict host species during later digestion despite a small reduction in model accuracy 30-hour post-feeding of approximately 4–6 percentage points. Despite this decrease, multiple host species were correctly classified with 100% accuracy suggesting that MIRS-ML may be more suitable for analysing degraded mosquito blood meals. Repeating this experiment to collect only samples at late digestion stages would be useful to clarify differences in the reduction of predictions accuracies and PCR failure point, as this timepoint had the lowest number of samples.

Host species had a slight but significant impact on MIRS-ML model accuracy, because the model had higher odds of correctly predicting human DBS relative to other hosts. One potential explanation is that human DBS were the most abundant sample type (Table S1) especially at timepoint 30 hours post feeding where overall model accuracy tended to be lower. Despite this, overall MIRS-ML accuracy was still much higher than a random prediction (16.7%), suggesting a strong spectroscopic signal from all hosts.

Increased model prediction scores were positively associated with both correct classifications (Figs. 5A and B) and predictions involving host species included in the training dataset (Fig. 5C). This result suggests that misclassifications are more likely when spectral signals are less distinct or overlapping, possibly due to digestion-related degradation, spectral noise, intrinsic similarity between host blood profiles, or when the model encounters a host species not present in its training set. Additionally, misclassified laboratory spectra had a different average absorbance compared to correctly classified spectra (Figure S2), suggesting some samples may have had reduced quality. Therefore, the model's probability score could be used to improve classification accuracy or exclude field samples from unknown hosts, for example by applying a threshold that retains only high-confidence predictions, which are more likely to reflect clear and biologically distinct host-specific signals. Using the threshold approach, the MLP model increased in overall accuracy to 78.2% for field samples from the Philippines with known hosts, excluding 40 DBS where the model probability was below 0.73. When all Philippines field samples are included, a further 26 DBS are removed, representing 24.9% of all Philippines DBS. This is lower than the number of DBS which could not be amplified by PCR (29.6%). This indicates MIRS-ML may be suitable for identifying blood meals late in the mosquito digestion stage and could be further improved up by increasing training sample size from this stage.

There are some limitations to this study. Whilst the accuracy of MIRS-ML decreased over time, the laboratory experiments only extended until 30 hours post-feeding due to laboratory constraints. Previous laboratory experiments in *Cx. quinquefasciatus* mosquitoes show that digestion enzymes peak around 36 hours post-feeding (59). Additionally, some mosquitoes were reluctant to feed, especially from the chicken blood, which limited the number of chicken DBS at 30-hours post-feeding (Table S1). The apparent aversion to chicken blood in this study was unexpected because *Cx. quinquefasciatus* has been shown to be an opportunistic feeder, sometimes even preferring avian blood (60,61). Using live animals for blood feeding experiments would likely overcome these challenges as well as generating a training set more representative of the real-world and without the need for anticoagulants. This would be beneficial as decreased accuracy of results from field-collected DBS in this study may be from the presence of citrate as an anti-coagulant in the training dataset, with MIRS- learning from wavenumbers correlating to the chemical bonds of citrate. Lastly, only eight spectra were discarded due to low quality, which was most likely caused by samples being incorrectly positioned on the scanner, resulting in distorted scans. This could be resolved by re-scanning the sample before DNA extraction, as a benefit of MIRS is that the sample is not destroyed.

Because model accuracy decreased from laboratory to field-collected mosquitoes, further work to improve the accuracy of field-collected mosquitoes is justified. This could involve using field-collected DBS to train models; however, this may prove challenging as only 5.4% and < 0.1% of our total mosquito catch had blood fed from the Philippines and Australia, respectively. Capturing blood fed mosquitoes is difficult, because mosquitoes tend to rest after a blood meal and do not engage in host seeking behaviour. The use of landing traps, resting collections or gravid traps may improve the number of blood feds caught (62,63). Previous research has shown that transfer learning techniques can improve model predictions (31), including a small number of field samples in a larger laboratory generated dataset improved model prediction of host blood source in *An. funestus* mosquitoes from 76% to 90% (31). Including field DBS in the training dataset may improve model accuracy for field use.

Despite these limitations, this study showed the practical advantage of the MIRS-ML method. It was an extremely quick method when compared to conventional PCR methods; for example, one person scanning 203 DBS took one workday to complete, in comparison to the gene sequencing workflow, which took 15 workdays. The initial cost of a Bruker ALPHA Fourier-transform infrared spectrometer equipped with a Platinum ATR device (46) is approximately GBP12,000, but does not require ongoing costs such as reagents. Additionally, this instrument can be used for a variety of other surveillance purposes, such as identification of malaria in human DBS (45), age-grading of mosquitoes (37,40) plus other vectors (64), and mosquito species identification (37,65,66). Therefore, over time this instrument may be cost-effective.

Under the influence of climate change and globalisation, mosquito-borne pathogens are emerging in new regions or re-emerging as environmental conditions change (67). One such example is JEV, which has recently spread to temperate regions of Australia, resulting in significant health and economic consequences (68). With the spread of these pathogens to new regions where major gaps in the understanding of locally important vectors and hosts occur, effective research and surveillance tools will be critical to assist local authorities with the management of health threats. Ensuring rapid and reliable assessment of host-feeding behaviours and pathogen transmission dynamics is critical and MIRS-ML methods may provide an economical solution.

In conclusion, MIRS-ML method shows great promise in identifying mosquito blood meals from *Culex* mosquitoes, can be used on dried blood spots and would be of practical value for field sample collection. Further refinement and training on larger datasets would improve performance and allow for application of MIRS-ML as an alternative to the gold-standard PCR with integration into vector-borne disease surveillance.

Declarations

Acknowledgements

We are grateful to the Philippines Department of Health and Royal Institute of Tropical Medicine & Hygiene for their continued disease surveillance efforts and support for our field collections, including Dr Jessie Fantone, Ms Cecilia Zuasula and Mr Alex Baday. We thank the FETAPFI administrative team for their continued support in our field collections in the Philippines.

We thank Jack Watson from Medical Entomology, NSW Health Pathology, for technical assistance with mosquito rearing and laboratory testing.

We would like to acknowledge and pay respect to the traditional owners of the land on which our NPA fieldwork was conducted; land owned by Injinoo, Umagico and New Mapoon, and the Islander communities of Seisia and Bamaga, and for the NPA council for their continued support in our field collections.

We are grateful to two anonymous reviewers for their constructive feedback that significantly improved this manuscript.

Funding Sources

This work was funded by the University of Glasgow Integrated Infection Biology Programme from the Wellcome Trust (218518/Z/19/Z), University of Sydney-University of Glasgow Ignition Fund and the Sydney Southeast Asia Centre Collaborative Research Grant. FB was funded by Academy Medical Science Springboard Award [ref: SBF007\100094]. MV was supported by the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 852957).

NSW Health Pathology receives funding from NSW Health for mosquito and arbovirus surveillance.

Conflict of Interest

None declared.

Data Availability

The datasets and code used during the current study are available from the following GitHub page.

https://github.com/erin-jpg/Culex_DBS_MIRS_ML_Scripts

All sequencing data will be uploaded to NCBI GenBank database.

Ethics and Consent to Participate declarations

not applicable

Consent to Publish declaration

not applicable

References

1. WHO WHO. Vector-borne diseases [Internet]. [cited 2025 May 13]. Available from: https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases
2. Diatta M, Spiegel A, Lochouarn L, Fontenille D. Similar feeding preferences of Anopheles gambiae and A. arabiensis in Senegal. Trans R Soc Trop Med Hyg. 1998 May 1;92(3):270–2. doi:10.1016/S0035-9203(98)91005-7
3. Thiemann TC, Lemenager DA, Kluh S, Carroll BD, Lothrop HD, Reisen WK. Spatial Variation in Host Feeding Patterns of Culex tarsalis and the Culex pipiens complex (Diptera: Culicidae) in California. J Med Entomol. 2012 Jul;49(4):903–16. doi:10.1603/me11272 PubMed PMID: 22897051; PubMed Central PMCID: PMC3542768.
4. Lyimo IN, Ferguson HM. Ecological and evolutionary determinants of host species choice in mosquito vectors. Trends in Parasitology. 2009 Apr 1;25(4):189–96. doi:10.1016/j.pt.2009.01.005 PubMed PMID: 19269900.
5. Thiemann TC, Wheeler SS, Barker CM, Reisen WK. Mosquito Host Selection Varies Seasonally with Host Availability and Mosquito Density. PLOS Neglected Tropical Diseases. 2011 Dec 20;5(12):e1452. doi:10.1371/journal.pntd.0001452
6. Ochomo E, Rund SSC, Mthawanji RS, Antonio-Nkondjio C, Machani M, Samake S, et al. Mosquito control by abatement programmes in the United States: perspectives and lessons for countries in sub-Saharan Africa. Malaria Journal. 2024 Jan 4;23(1):8. doi:10.1186/s12936-023-04829-3
7. Asale A, Duchateau L, Devleesschauwer B, Huisman G, Yewhalaw D. Zooprophylaxis as a control strategy for malaria caused by the vector Anopheles arabiensis (Diptera: Culicidae): a systematic review. Infect Dis Poverty. 2017 Oct 25;6:160. doi:10.1186/s40249-017-0366-3 PubMed PMID: 29157310; PubMed Central PMCID: PMC5697156.
8. Iwashita H, Dida GO, Sonye GO, Sunahara T, Futami K, Njenga SM, et al. Push by a net, pull by a cow: can zooprophylaxis enhance the impact of insecticide treated bed nets on malaria control? Parasites Vectors. 2014 Jan 28;7(1):52. doi:10.1186/1756-3305-7-52
9. Bouma M, Rowland M. Failure of passive zooprophylaxis: cattle ownership in Pakistan is associated with a higher prevalence of malaria. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1995 Jul 1;89(4):351–3. doi:10.1016/0035-9203(95)90004-7
10. Franco AO, Gomes MGM, Rowland M, Coleman PG, Davies CR. Controlling Malaria Using Livestock-Based Interventions: A One Health Approach. PLOS ONE. 2014 Jul 22;9(7):e101699.

doi:10.1371/journal.pone.0101699

11. Moore SM. The current burden of Japanese encephalitis and the estimated impacts of vaccination: Combining estimates of the spatial distribution and transmission intensity of a zoonotic pathogen. PLOS Neglected Tropical Diseases. 2021 Oct 13;15(10):e0009385. doi:10.1371/journal.pntd.0009385
12. Erlanger TE, Weiss S, Keiser J, Utzinger J, Wiedenmayer K. Past, Present, and Future of Japanese Encephalitis. Emerg Infect Dis. 2009 Jan;15(1):1–7. doi:10.3201/eid1501.080311 PubMed PMID: 19116041; PubMed Central PMCID: PMC2660690.
13. Tiwari S, Singh R, Tiwari R, Dhole T. Japanese encephalitis: A review of the Indian perspective. The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases. 2012 Nov 8;16. doi:10.1016/j.bjid.2012.10.004
14. Lord JS, Gurley ES, Pulliam JRC. Rethinking Japanese Encephalitis Virus Transmission: A Framework for Implicating Host and Vector Species. PLOS Neglected Tropical Diseases. 2015 Dec 10;9(12):e0004074. doi:10.1371/journal.pntd.0004074
15. Zhang H, Li D, Zheng J, Bao J, Wang Z, Qiu Y, et al. Sheep serve as amplifying hosts of Japanese encephalitis virus, increasing the risk of human infection. Science Advances. 2025 May 16;11(20):eads7441. doi:10.1126/sciadv.ads7441
16. Sunwoo JS, Jung KH, Lee ST, Lee SK, Chu K. Reemergence of Japanese Encephalitis in South Korea, 2010–2015. Emerg Infect Dis. 2016 Oct;22(10):1841–3. doi:10.3201/eid2210.160288 PubMed PMID: 27648677; PubMed Central PMCID: PMC5038436.
17. Yap G, Lim XF, Chan S, How CB, Humaidi M, Yeo G, et al. Serological evidence of continued Japanese encephalitis virus transmission in Singapore nearly three decades after end of pig farming. Parasites & Vectors. 2019 May 17;12(1):244. doi:10.1186/s13071-019-3501-0
18. Ayu Mirah Adi AA, Astawa NM, Asri Damayanti PA, Kardena IM, Krisna Erawan IGM, Suardana IW, et al. Seroepidemiological Evidence for the Presence of Japanese Encephalitis Virus Infection in Ducks, Chickens, and Pigs, Bali-Indonesia. Bali Med J. 2016 Nov 14;5(3):189. doi:10.15562/bmj.v5i3.343
19. Cleton NB, Bosco-Lauth A, Page MJ, Bowen RA. Age-Related Susceptibility to Japanese Encephalitis Virus in Domestic Ducklings and Chicks. Am J Trop Med Hyg. 2014 Feb 5;90(2):242–6. doi:10.4269/ajtmh.13-0161 PubMed PMID: 24394476; PubMed Central PMCID: PMC3919224.
20. Turell MJ, Mores CN, Dohm DJ, Komilov N, Paragas J, Lee JS, et al. Laboratory Transmission of Japanese Encephalitis and West Nile Viruses by Molestus Form of Culex pipiens (Diptera: Culicidae) Collected in Uzbekistan in 2004. Journal of Medical Entomology. 2006 Mar 1;43(2):296–300. doi:10.1093/jmedent/43.2.296
21. Madhav M, Blasdell KR, Trewin B, Paradkar PN, López-Denman AJ. Culex-Transmitted Diseases: Mechanisms, Impact, and Future Control Strategies using Wolbachia. Viruses. 2024 Jul;16(7):1134. doi:10.3390/v16071134

22. Griep JS, Grant E, Pilgrim J, Riabinina O, Baylis M, Wardeh M, et al. Meta-analyses of Culex blood-meals indicates strong regional effect on feeding patterns. PLOS Neglected Tropical Diseases. 2025 Jan 24;19(1):e0012245. doi:10.1371/journal.pntd.0012245
23. Boorman J, Mellor PS, Boreham PFL, Hewett RS. A latex agglutination test for the identification of blood-meals of Culicoides (Diptera: Ceratopogonidae). Bulletin of Entomological Research. 1977 Jun;67(2):305–11. doi:10.1017/S0007485300011123
24. Gomes LA, Duarte R, Lima DC, Diniz BS, Serrão ML, Labarthe N. Comparison between precipitin and ELISA tests in the bloodmeal detection of Aedes aegypti (Linnaeus) and Aedes fluviatilis (Lutz) mosquitoes experimentally fed on feline, canine and human hosts. Mem Inst Oswaldo Cruz. 2001 Jul;96:693–5. doi:https://doi.org/10.1590/S0074-02762001000500020
25. Edrissian GhH, Hafizi A. Application of enzyme-linked immunosorbent assay (ELISA) to identification of *Anopheles* mosquito bloodmeals. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1982 Jan 1;76(1):54–6. doi:10.1016/0035-9203(82)90017-7
26. Borland EM, Kading RC. Modernizing the Toolkit for Arthropod Bloodmeal Identification. Insects. 2021 Jan;12(1):1. doi:10.3390/insects12010037
27. Townzen JS, Brower AVZ, Judd DD. Identification of mosquito bloodmeals using mitochondrial cytochrome oxidase subunit I and cytochrome b gene sequences. Medical and Veterinary Entomology. 2008;22(4):386–93. doi:10.1111/j.1365-2915.2008.00760.x
28. Oshaghi MA, Chavshin AR, Vatandoost H. Analysis of mosquito bloodmeals using RFLP markers. Experimental Parasitology. 2006 Dec 1;114(4):259–64. doi:10.1016/j.exppara.2006.04.001
29. Vilela TS, Senne NA, Paulino PG, Araújo TR, de Lima PP de ABM, Gonçalves NP, et al. A multiplex qPCR followed by high-resolution melting analysis for the detection of blood-feeding sources in Culex sp. mosquitoes. Medical and Veterinary Entomology. 2025;39(2):266–77. doi:10.1111/mve.12775
30. Martínez-de la Puente J, Ruiz S, Soriguer R, Figuerola J. Effect of blood meal digestion and DNA extraction protocol on the success of blood meal source determination in the malaria vector Anopheles atroparvus. Malar J. 2013 Mar 21;12(1):109. doi:10.1186/1475-2875-12-109
31. Mwanga EP, Mchola IS, Makala FE, Mshani IH, Siria DJ, Mwinyi SH, et al. Rapid assessment of the blood-feeding histories of wild-caught malaria mosquitoes using mid-infrared spectroscopy and machine learning. Malar J. 2024 Mar 26;23(1):86. doi:10.1186/s12936-024-04915-0
32. Athni TS, Shocket MS, Couper LI, Nova N, Caldwell IR, Caldwell JM, et al. The influence of vector-borne disease on human history: socio-ecological mechanisms. Ecology Letters. 2021;24(4):829–46. doi:10.1111/ele.13675
33. Pasquini C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc. 2003 Apr;14:198–219. doi:10.1590/S0103-50532003000200006
34. Wilson RH, Tapp HS. Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods. TrAC Trends in Analytical Chemistry. 1999 Feb 1;18(2):85–93. doi:10.1016/S0165-9936(98)00107-1

35. Seybold CA, Ferguson R, Wysocki D, Bailey S, Anderson J, Nester B, et al. Application of Mid-Infrared Spectroscopy in Soil Survey. Soil Science Society of America Journal. 2019;83(6):1746–59. doi:10.2136/sssaj2019.06.0205
36. De Bruyne S, Speeckaert MM, Delanghe JR. Applications of mid-infrared spectroscopy in the clinical laboratory setting. Critical Reviews in Clinical Laboratory Sciences. 2018 Jan 2;55(1):1–20. doi:10.1080/10408363.2017.1414142 PubMed PMID: 29239240.
37. González-Jiménez M, Babayan SA, Khazaeli P, Doyle M, Walton F, Reedy E, et al. Prediction of malaria mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning [Internet]. bioRxiv; 2018 [cited 2025 May 11]. p. 414342. Available from: https://www.biorxiv.org/content/10.1101/414342v1 doi:10.1101/414342
38. Foti M, Micocci M, Betancourth MP, Gomez-Uribarri IC, Serini P, Caputo B, et al. Towards Scalable Age-Grading of Aedes albopictus mosquito using Mid-Infrared Spectroscopy and Machine Learning [Internet]. bioRxiv; 2025 [cited 2025 Nov 7]. p. 2025.06.20.660660. Available from: https://www.biorxiv.org/content/10.1101/2025.06.20.660660v1 doi:10.1101/2025.06.20.660660
39. Mwanga EP, Kweyamba PA, Siria DJ, Mshani IH, Mchola IS, Makala FE, et al. Reagent-free detection of Plasmodium falciparum malaria infections in field-collected mosquitoes using mid-infrared spectroscopy and machine learning. Sci Rep. 2024 May 27;14(1):12100. doi:10.1038/s41598-024-63082-z
40. Mwanga EP, Mapua SA, Siria DJ, Ngowo HS, Nangacha F, Mgando J, et al. Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, Anopheles arabiensis. Malaria Journal. 2019 May 30;18. doi:10.1186/s12936-019-2822-y
41. Hemotek. Starter pack – 6 feeders with 3ml reservoirs – Hemotek membrane feeding systems [Internet]. [cited 2025 Oct 13]. Available from: http://hemotek.co.uk/starter-pack-6-feeders-with-3ml-reservoirs/
42. Cardinal Bioresearch Pty Ltd [Internet]. 2020 Nov 30 [cited 2024 Jun 13]. Available from: https://www.cardinalbioresearch.com.au/contact/, https://www.cardinalbioresearch.com.au/contact/
43. Lifeblood ARC. Australian Red Cross Lifeblood [Internet]. [cited 2025 Oct 13]. Available from: https://www.lifeblood.com.au/health-professionals/products
44. Mwanga EP, Minja EG, Mrimi E, Jiménez MG, Swai JK, Abbasi S, et al. Detection of malaria parasites in dried human blood spots using mid-infrared spectroscopy and logistic regression analysis. Malaria Journal. 2019 Oct 7;18(1):341. doi:10.1186/s12936-019-2982-9
45. Mshani IH, Jackson FM, Mwanga RY, Kweyamba PA, Mwanga EP, Tambwe MM, et al. Screening of malaria infections in human blood samples with varying parasite densities and anaemic conditions using AI-Powered mid-infrared spectroscopy. Malaria Journal. 2024 Jun 17;23(1):188. doi:10.1186/s12936-024-05011-z
46. ALPHA II Compact FT-IR Spectrometer [Internet]. [cited 2024 Dec 3]. Available from: https://www.bruker.com/en/products-and-solutions/infrared-and-raman/ft-ir-routine-

spectrometer/alpha-ii-compact-ft-ir-spectrometer.html

47. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011 Nov 1;12(null):2825–30.
48. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning [Internet]. New York, NY: Springer; 2009 [cited 2026 Apr 17]. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-0-387-84858-7 doi:10.1007/978-0-387-84858-7
49. Hong C, Brookes VJ, Zadoks RN, Webb CE. Citric Acid and Sodium Bicarbonate as an Alternative Carbon Dioxide Source for Mosquito Surveillance. Insects. 2025 Jan;16(1):90. doi:10.3390/insects16010090
50. Rattanarithikul R, Harrison BA, Panthusiri P, Coleman RE. Illustrated keys to the mosquitoes of Thailand I. Background; geographic distribution; lists of genera, subgenera, and species; and a key to the genera. Southeast Asian J Trop Med Public Health. 2005;36 Suppl 1:1–80. doi:https://www.proquest.com/openview/97b406ad13c0e08cb28ee0e4cc04e222/1?pq-origsite=gscholar&cbl=34824 PubMed PMID: 15916030.
51. WHO. Pictorial identification key of important disease vectors in the WHO South-East Asia Region [Internet]. 2025 [cited 2025 Jun 27]. Available from: https://iris.who.int/handle/10665/332202
52. Webb C, Doggett S, Russell R. A Guide to Mosquitoes of Australia. Csiro Publishing; 2016. 217 p.
53. Delrue C, Speeckaert R, Oyaert M, Kerre T, Rottey S, Coopman R, et al. Infrared Spectroscopy: A New Frontier in Hematological Disease Diagnosis. Int J Mol Sci. 2023 Nov 30;24(23):17007. doi:10.3390/ijms242317007 PubMed PMID: 38069330; PubMed Central PMCID: PMC10707114.
54. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON.
55. Rumelhart DE, Hintont GE, Williams RJ. Learning representations by back-propagating errors. 1986.
56. Li J, Ulloa GM, Mayor P, Santolalla Robles ML, Greenwood AD. Nucleic acid degradation after long-term dried blood spot storage. Molecular Ecology Resources. 2024;24(6):e13979. doi:10.1111/1755-0998.13979
57. Oshaghi MA, Chavshin AR, Vatandoost H, Yaaghoobi F, Mohtarami F, Noorjah N. Effects of post-ingestion and physical conditions on PCR amplification of host blood meal DNA in mosquitoes. Experimental Parasitology. 2006 Apr 1;112(4):232–6. doi:10.1016/j.exppara.2005.11.008
58. Tchouassi DP, Kisero RO, Rotich G, Dunlap C, Torto B, Muturi EJ. Next generation sequencing improves the resolution of detecting mixed host blood meal sources in field collected arboviral mosquito vectors. Medical and Veterinary Entomology. 2024;38(4):407–15. doi:10.1111/mve.12725
59. Okuda K, Caroci A de S, Ribolla PEM, de Bianchi AG, Bijovsky AT. Functional morphology of adult female *Culex quinquefasciatus* midgut during blood digestion. Tissue and Cell. 2002 Jun 1;34(3):210–9. doi:10.1016/S0040-8166(02)00032-0
60. Hannon ER, Jackson KC, Biggerstaff BJ, Raman V, Komar N. Bloodmeal Host Selection of Culex quinquefasciatus (Diptera: Culicidae) in Las Vegas, Nevada, United States. Journal of Medical Entomology. 2019 Apr 16;56(3):603–8. doi:10.1093/jme/tjy237

61. Fitzpatrick DM, Hattaway LM, Hsueh AN, Ramos-Niño ME, Cheetham SM. PCR-Based Bloodmeal Analysis of Aedes aegypti and Culex quinquefasciatus (Diptera: Culicidae) in St. George Parish, Grenada. Journal of Medical Entomology. 2019 Jun 27;56(4):1170–5. doi:10.1093/jme/tjz037
62. Melgarejo-Colmenares K, Cardo MV, Vezzani D. Blood feeding habits of mosquitoes: hardly a bite in South America. Parasitol Res. 2022 Jul 1;121(7):1829–52. doi:10.1007/s00436-022-07537-0
63. Silver JB. Mosquito Ecology [Internet]. Dordrecht: Springer Netherlands; 2008 [cited 2025 Oct 16]. Available from: http://link.springer.com/10.1007/978-1-4020-6666-5 doi:10.1007/978-1-4020-6666-5
64. Pazmiño-Betancourth M, Casas Gómez-Uribarri I, Mondragon-Shem K, Babayan SA, Baldini F, Rafuse Haines L. Advancing age grading techniques for Glossina morsitans morsitans, vectors of African trypanosomiasis, through mid-infrared spectroscopy and machine learning. Biol Methods Protoc. 2024 Jan 1;9(1):bpae058. doi:10.1093/biomethods/bpae058
65. González Jiménez M, Babayan SA, Khazaeli P, Doyle M, Walton F, Reedy E, et al. Prediction of mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. Wellcome Open Res. 2019 Sep 16;4:76. doi:10.12688/wellcomeopenres.15201.3 PubMed PMID: 31544155; PubMed Central PMCID: PMC6753605.
66. Pazmiño-Betancourth M, Ochoa-Gutiérrez V, Ferguson HM, González-Jiménez M, Wynne K, Baldini F, et al. Evaluation of diffuse reflectance spectroscopy for predicting age, species, and cuticular resistance of Anopheles gambiae s.l under laboratory conditions. Sci Rep. 2023 Oct 28;13(1):18499. doi:10.1038/s41598-023-45696-x
67. Franklinos LHV, Jones KE, Redding DW, Abubakar I. The effect of global change on mosquito-borne disease. The Lancet Infectious Diseases. 2019 Sep 1;19(9):e302–12. doi:10.1016/S1473-3099(19)30161-6
68. Williams CR, Webb CE, Higgs S, Hurk AF van den. Japanese Encephalitis Virus Emergence in Australia: Public Health Importance and Implications for Future Surveillance. Vector-Borne and Zoonotic Diseases. 2022 Nov 9. Located at: 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA. doi:10.1089/vbz.2022.0037

Figures

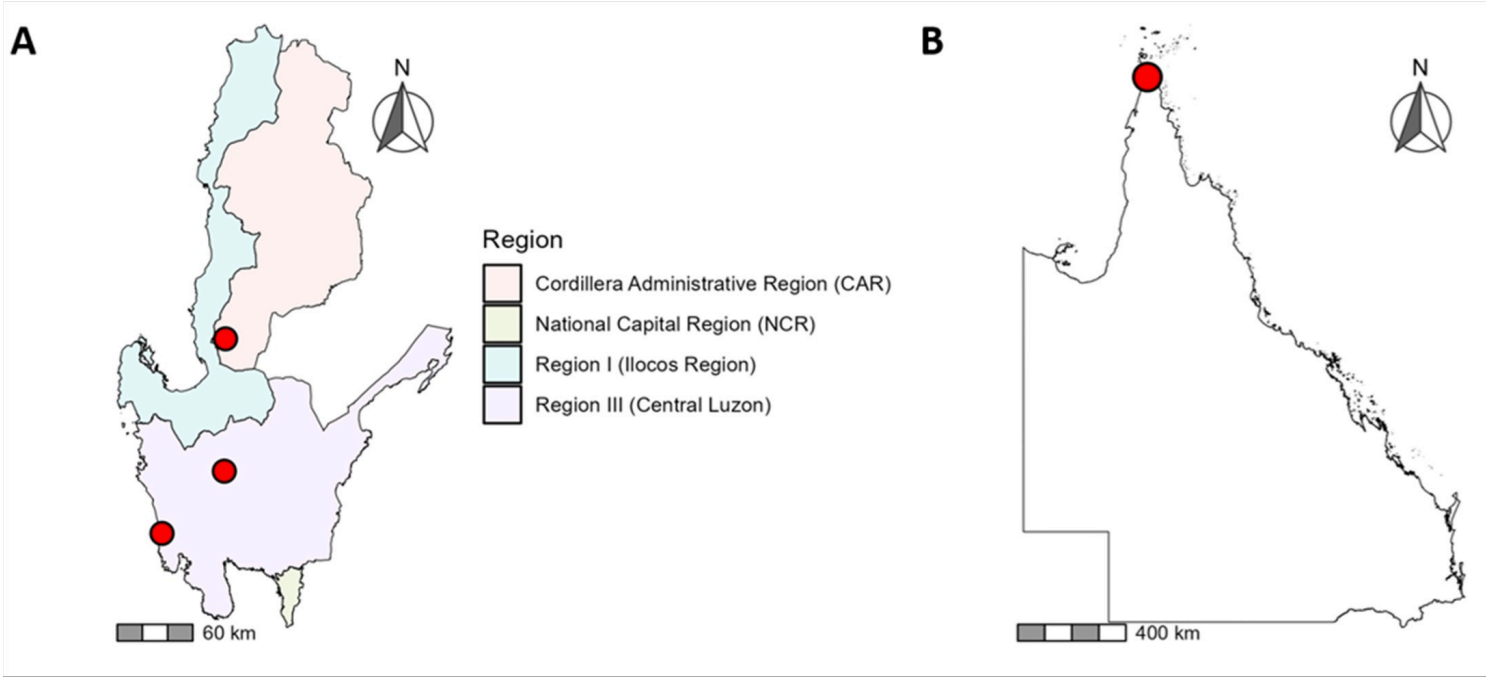


Figure 1

Map of study sites. **A**: Cordillera Administrative Region and Central Luzon regions in the Philippines. **B**: Queensland, Australia. The red dots represent study site locations.

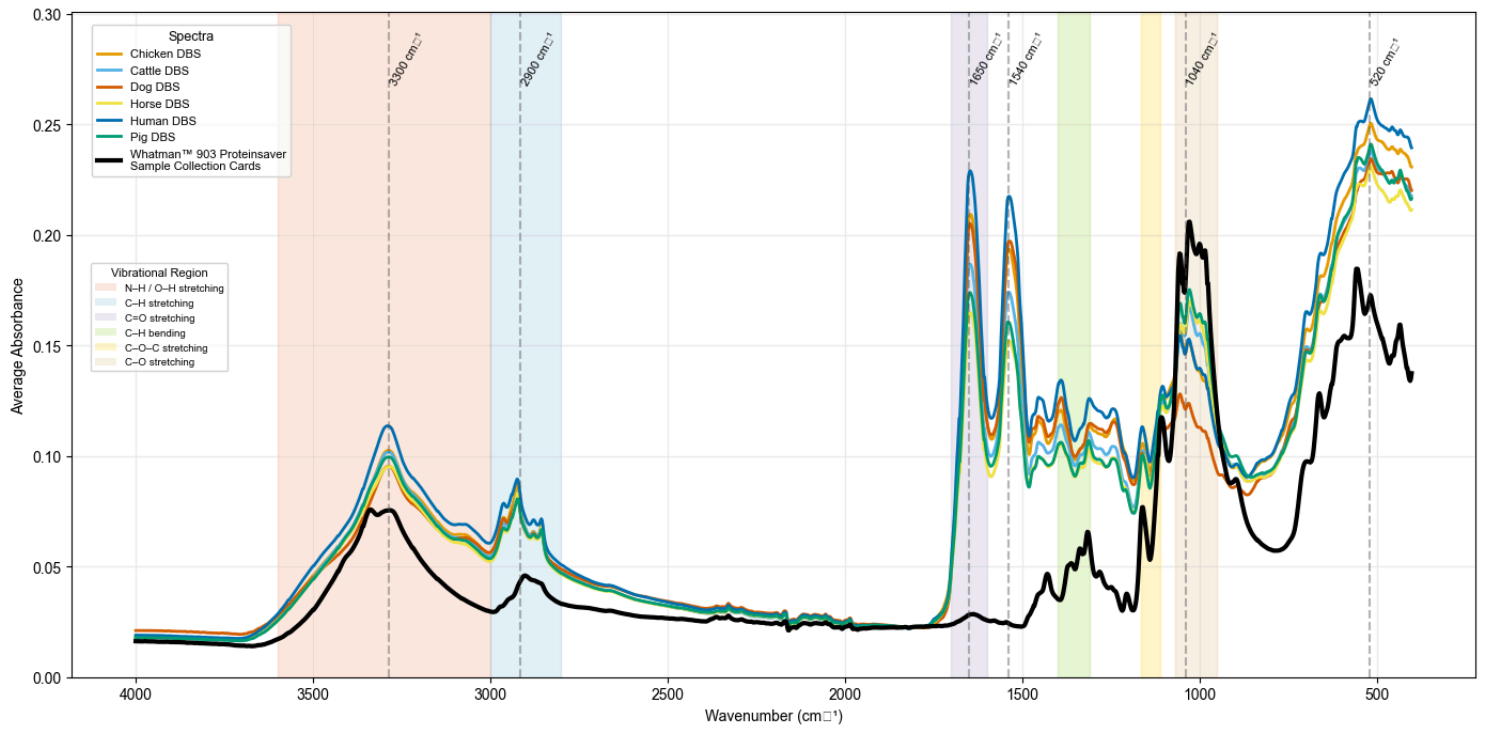


Figure 2

Mean spectra of each host from DBS collected from laboratory-reared *Cx. quinquefasciatus*. Coloured regions show approximate vibrational regions for specific biomolecule groups. Labelled absorbance

peaks correspond to the following chemical bonds N-H/O-H stretching (3300 cm^{-1}), C-H stretching (2900 cm^{-1}), C=O stretching [Amide I band] (1850cm^{-1}), N-H bend and C-N stretching [Amide II band] (1540 cm^{-1}), C-O stretching (1040 cm^{-1}) and Fe-N or Fe-O stretching (520cm^{-1}).

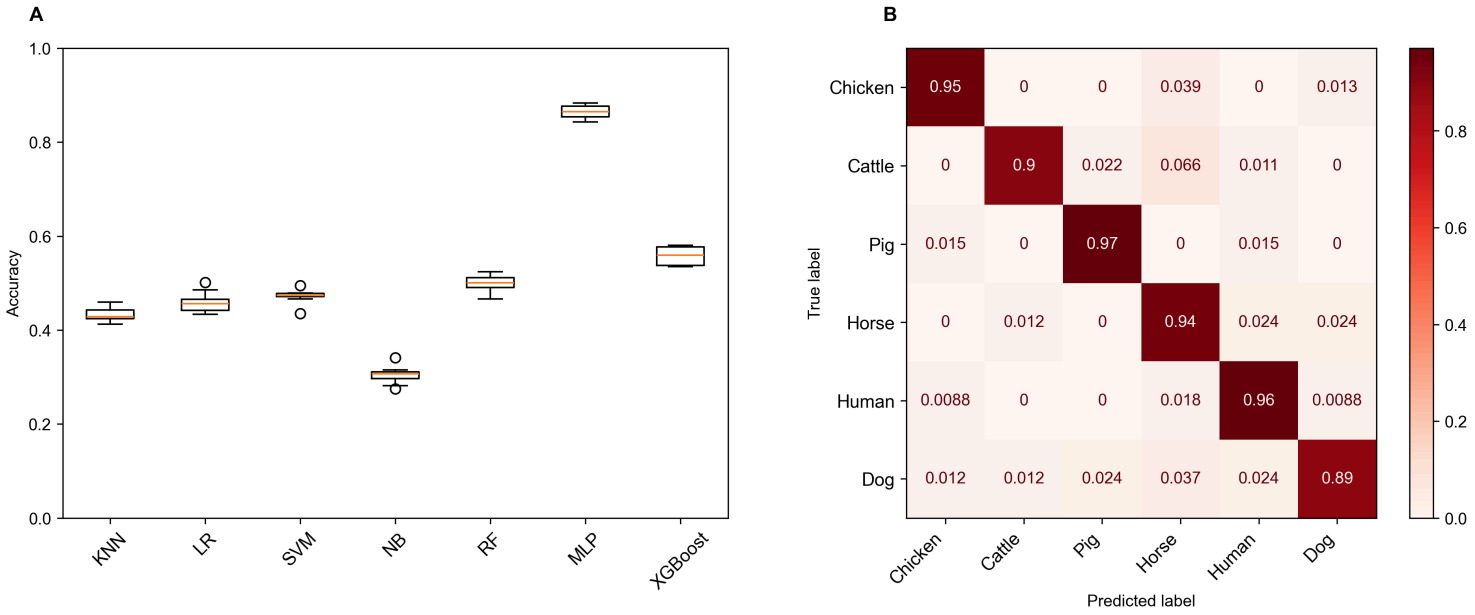


Figure 3

Machine learning model accuracies for classifying blood meal source from DBS collected from laboratory-reared *Cx. quinquefasciatus*. **A:** Comparison of accuracy of the different model algorithms tested: KNN= k-nearest neighbours, LR=logistic regression, SVM= support vector Machine, NB=Gaussian naive bayes, RF=random forest classifier, MLP=multilayer perceptron and XGB= XGBoost**. B**: Normalised confusion matrix for the optimised MLP model, where values represent the percentage of correctly classified samples within each host class, based on 515 DBS from laboratory-reared *Cx quinquefasciatus*.

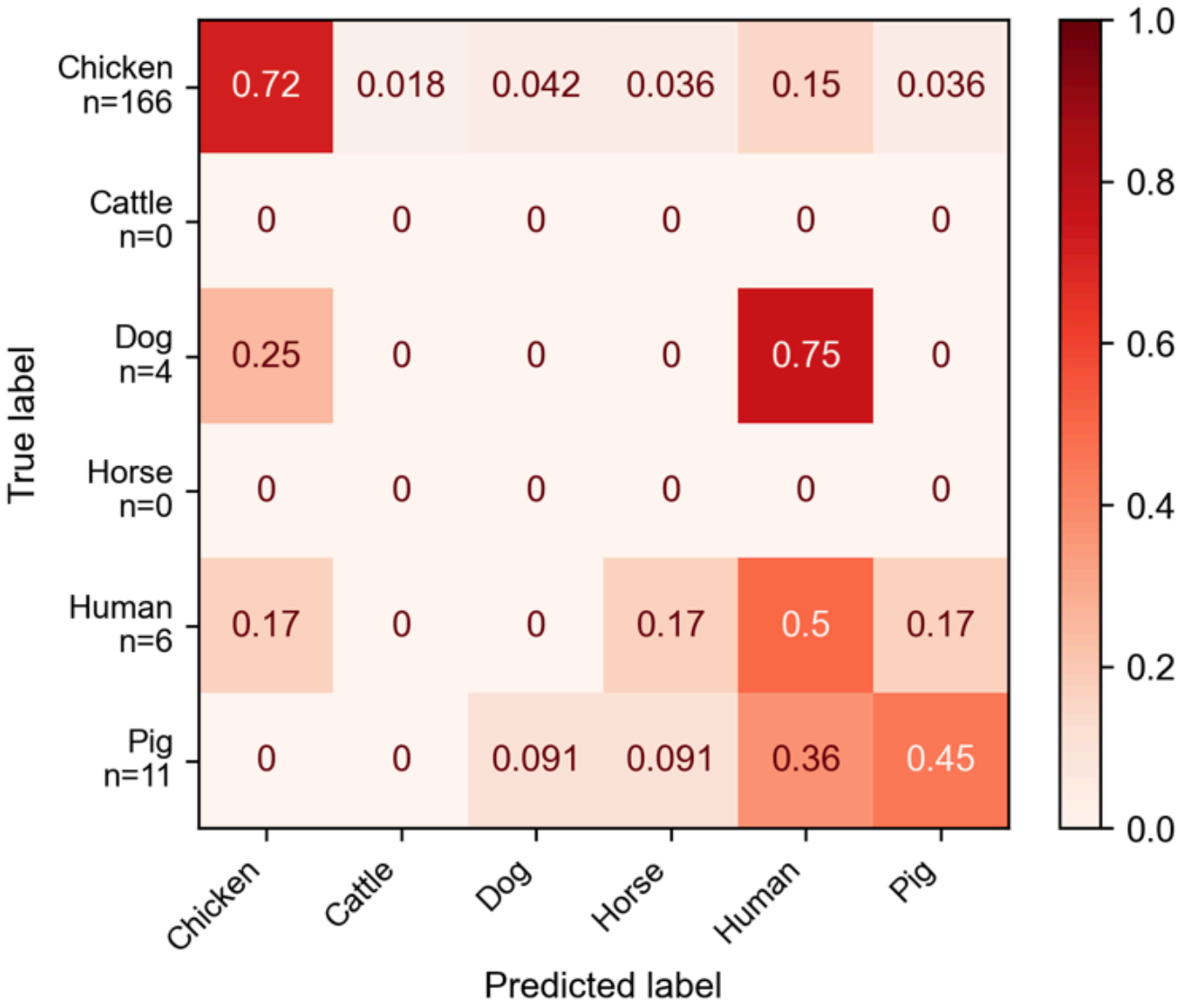


Figure 4

Normalised confusion matrix for the optimised MLP model, where values represent the percentage of correctly classified samples within each known host class, based on 187 DBS from field-caught Philippines mosquitoes.

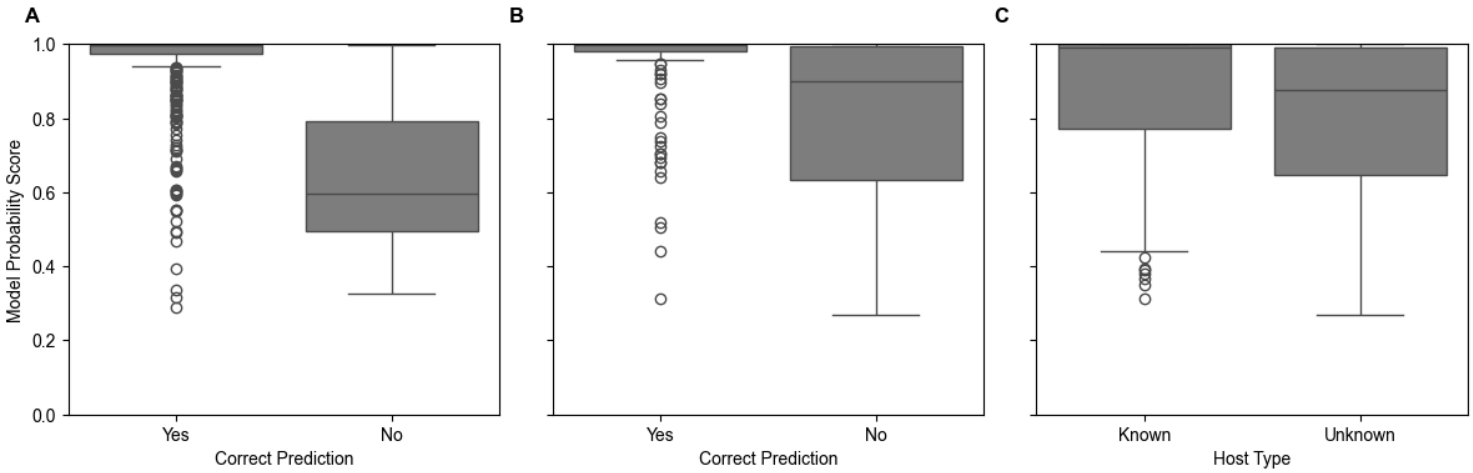


Figure 5

Model probability scores of MIRS-ML model predictions of dried bloodspots (DBS) from mosquito blood meals. **A, B:** MIRS-ML model probability scores of predicted hosts that were predicted correctly (yes) and incorrectly (no) by the optimised multi-layer perceptron model from (A) laboratory-reared *Cx. quinquefasciatus* DBS or (B) from Philippine wild-caught mosquitoes. **C:** MIRS-ML model probability scores of predicted hosts from field-caught DBS from known and unknown labels, indicating host species included or not included in the training set, respectively.

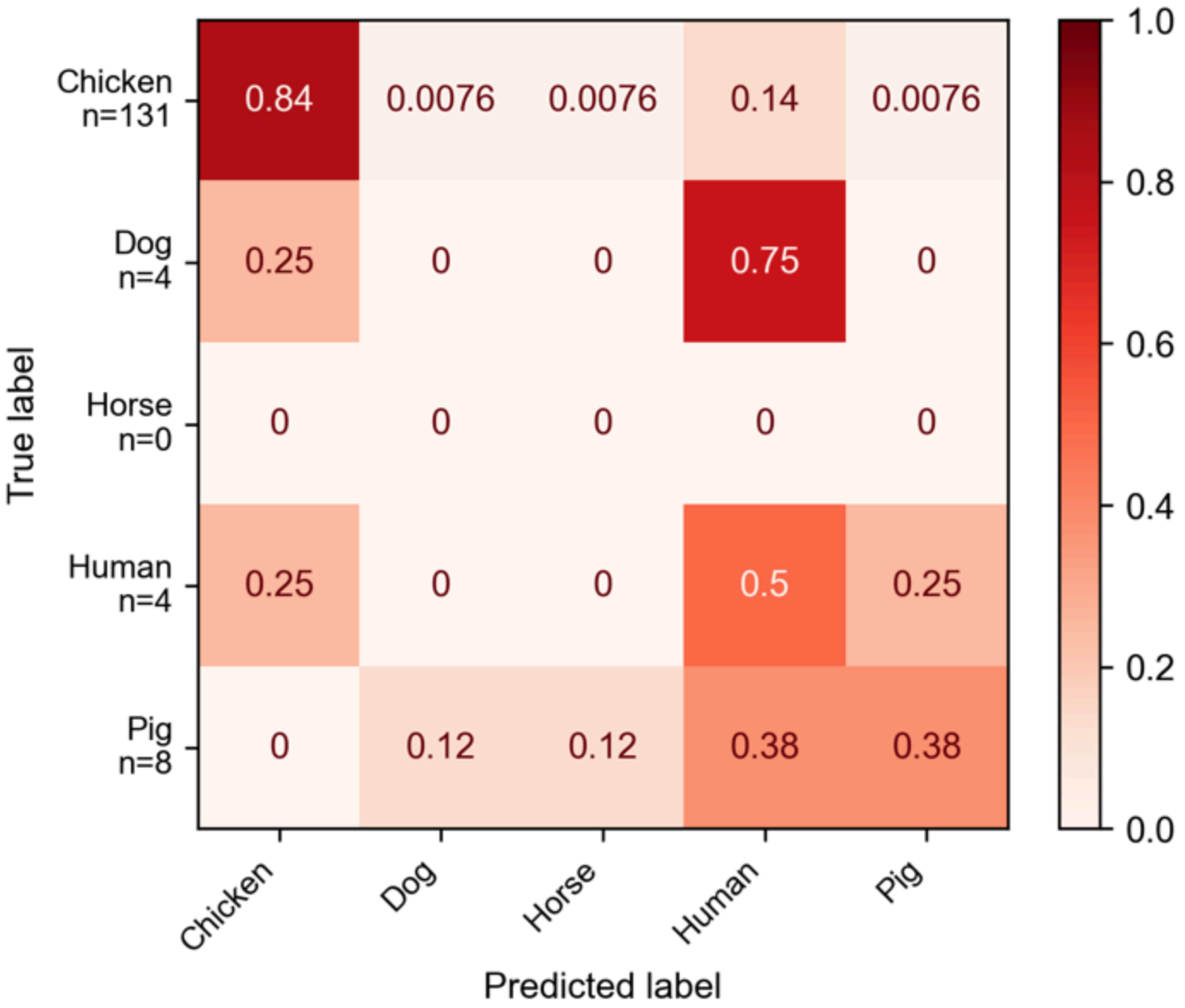


Figure 6

Normalised confusion matrix for the optimised MLP model of samples with a probability greater than 0.73, values represent the percentage of correctly classified samples within each known host class. Based on 147 DBS from field-caught Philippines mosquitoes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CulexMIRSMLSupplementaryMaterial.docx