

1 **Detection of malaria in insectary-reared *Anopheles gambiae* using near-infrared spectroscopy**

2

3 **Authors:** Marta F. Maia^{1,2,3,4 * §}, Melissa Kapulu^{3,4}, Michelle Muthui³, Martin G. Wagah³, Heather M.
4 Ferguson⁵, Floyd E. Dowell⁶, Francesco Baldini^{5#} and Lisa-Ranford Cartwright^{5#}

5

6 **Affiliations**

7 ¹ Swiss Tropical and Public Health Institute, Socinstrasse 57, 4020-Basel, Switzerland.

8 ² University of Basel, Petersplatz 1, 4001-Basel, Switzerland

9 ^{3§} KEMRI Wellcome Trust Research Programme, P.O. Box 230, 80108 - Kilifi, Kenya.

10 ^{4§} University of Oxford, Centre for Tropical Medicine and Global Health, Nuffield Department of
11 Medicine, Old Road Campus Roosevelt Drive, Oxford OX3 7FZ, UK.

12 ⁵ Institute of Biodiversity, Animal Health and Comparative Medicine, Graham Kerr Building, University of
13 Glasgow, Glasgow G12 8QQ, UK.

14 ⁶ USDA, Agricultural Research Service, Center for Grain and Animal Health Research, 1515 College
15 Avenue, Manhattan, KS, 66502, USA.

16 * Corresponding author

17 # These authors contributed equally to this work.

18 § current affiliation

19

20 **Contact information**

Marta Maia	mmaia@kemri-wellcome.org
Melissa Kapulu	mkapulu@kemri-wellcome.org
Martin Wagah	wgorry@kemri-wellcome.org
Michelle Muthui	mmuthui@kemri-wellcome.org
Heather Ferguson	Heather.Ferguson@glasgow.ac.uk
Floyd Dowell	floyd.dowell@ars.usda.gov
Francesco Baldini	Francesco.Baldini@glasgow.ac.uk
Lisa Ranford-Cartwright	Lisa.Ranford-Cartwright@glasgow.ac.uk

21

22

23 **Abstract**

24 Large-scale surveillance of mosquito populations is crucial to assess the intensity of vector-borne
25 disease transmission and the impact of control interventions. However, there is a lack of accurate, cost-
26 effective and high-throughput tools for mass-screening of vectors. This study demonstrates proof-of-
27 concept that near-infrared spectroscopy (NIRS) is capable of rapidly identifying laboratory strains of
28 human malaria infection in African mosquito vectors. By using partial least square regression models
29 based on malaria-infected and uninfected *Anopheles gambiae* mosquitoes, we showed that NIRS can
30 detect oocyst- and sporozoite-stage *Plasmodium falciparum* infections with 88% and 95% accuracy,
31 respectively. Accurate, low-cost, reagent-free screening of mosquito populations enabled by NIRS
32 could revolutionize surveillance and elimination strategies for the most important human malaria
33 parasite in its primary African vector species. Further research is needed to evaluate how the method
34 performs in the field following adjustments in the training datasets to include data from wild-caught
35 infected and uninfected mosquitoes.

36

37 **Introduction**

38 Malaria is holding back development in endemic countries and remains one of the leading causes of
39 death in children under 5 years-old in sub-Saharan Africa [1-3]. During the past decade, the large-scale
40 roll-out of long-lasting insecticide treated nets and indoor residual spraying across Africa has resulted in
41 a substantial reduction in malaria cases [4]. The WHO's Global Technical Strategy for Malaria 2016-
42 2030 seeks to reduce malaria incidence and related mortality by at least 90% and to eliminate the
43 disease in a minimum of 35 countries [1]. These bold goals will require new interventions that can
44 address residual malaria transmission as well as new tools to better monitor their impact on vector-
45 borne disease transmission. Mosquito surveillance is a cornerstone of the control of malaria and other
46 vector-borne diseases [5]. However, presently, there is no high-throughput, cost-efficient method to
47 identify *Plasmodium* infection and infectiousness in mosquitoes. Molecular methods such as ELISA and
48 PCR are used to determine parasite infection, but these are expensive and laborious [6-8], challenging
49 resource-poor countries with few funds and limited access to reagents and equipment, and thus are
50 unsuitable for large-scale surveillance. A further complication is that typically only 1-2% of mosquitoes
51 may be infected with transmission stage parasites (sporozoites), meaning that very large sample sizes
52 must be tested to accurately quantify site and time-specific estimates of mosquito infection rates as will
53 be required to assess progress towards malaria elimination [9].

54
55 Recent advances indicate several mosquito traits can be accurately identified through analysis of their
56 tissues with near infrared spectroscopy (NIRS) [10-13]. Here, visible and NIR light (wavelength 400-
57 2500 nanometers) is passed through the whole or part of a mosquito specimen and an absorbance
58 spectrum is collected instantly without destroying the sample. Changes in spectral peaks at different
59 wavelengths represent how intensely different molecules absorb light, and thus NIR spectra of
60 mosquitoes are determined by the biochemical composition of their tissues, which are known to differ
61 according to age [14, 15], species [16, 17], microbiome [18], physiological stage [19, 20], and by
62 pathogen infection [20, 21]. Differences in NIR spectra have been used to distinguish young (e.g. <7
63 days old) from older (7+ days old) malaria vectors, to identify morphologically identical *Anopheles*
64 sibling species, and to detect the presence of the endosymbiont *Wolbachia* in *Aedes aegypti*
65 mosquitoes [10-12]. Most recently, NIRS has been used to detect rodent malaria infections in
66 laboratory-reared *Anopheles stephensi* mosquitoes [22] and Zika virus in *Ae. aegypti* mosquitoes [23].

67 The use of NIRS has not previously been investigated on human malaria infected mosquitoes. The
68 presence of the parasite-specific proteins and other biochemical changes induced by malaria infection
69 in the vector may permit these to be distinguished from uninfected mosquitoes using spectral tools such
70 as NIRS [24, 25]

71
72 Parasite infection in the mosquito can be found in two main forms defined by their parasite development
73 stages: midgut oocyst infections occurring around 2-8 days after feeding on infectious blood; and
74 sporozoite infections occurring 9-14 days after infection, characterized by the release of sporozoites
75 from oocysts into the mosquito's haemocoel and salivary glands, enabling the mosquito to infect the
76 next human host. Given the different nature of the two infection stages the NIRS profile of an oocyst-
77 infected mosquito may not be the same as a sporozoite-infected one. For this reason, we aimed to test
78 whether NIRS could successfully identify oocyst and sporozoite infections in *Anopheles* vectors, and
79 estimate if the method's prediction accuracy is dependent on the intensity of infection in the mosquito.

80
81 In this paper, we present the successful application of NIRS to differentiate *Plasmodium falciparum*-
82 infected mosquitoes from uninfected mosquitoes, providing the first evidence of detection of human
83 malaria infections in the *A. gambiae* mosquito vector by this cost-effective, fast and reagent-free
84 method. The development of a tool such as NIRS to measure malaria infection rates in mosquito
85 populations would be of great service to malaria pre-elimination efforts as it would allow the processing
86 of large numbers of mosquitoes increasing the accuracy of the estimates of human exposure to malaria
87 infection across different regions, and advancing malaria vector surveillance in Africa.

88

89 **Results**

90 *Experimental infections*

91 Approximately 750 female *A. gambiae* (Keele line) [26] of different ages (3-6 days old) were offered a
92 blood meal containing NF54 gametocyte cultures in standard membrane-feeding assays (SMFAs) in
93 three independent replicate experiments. Control (uninfected) mosquitoes were generated by feeding
94 approximately 450 mosquitoes the same blood after gametogenesis was completed. Both groups were
95 represented with mosquitoes of similar ages, between 3 and 6 days old (Table 1 and 2 of
96 supplementary information). Mosquitoes were maintained for 7 and 14 days under insectary conditions

97 to allow oocyst (D7) and sporozoite (D14) development, on each day of sampling live mosquitoes were
98 removed, killed and immediately scanned using NIRS.

99

100 Mosquitoes fed on infectious blood were analysed by quantitative polymerase chain reaction (qPCR) for
101 intensity of infection. Additionally, 60 mosquitoes from the control groups (30 from feed 2 and 3
102 respectively) were also analysed by qPCR to confirm the absence of malaria infection. No mosquitoes
103 from these control groups tested positive for infection.

104

105 The minimum number of parasite genomes detectable per mosquito was 10 parasite genomes/ per μ l of
106 DNA extract, calculated from standard curves generated for each qPCR run using a 5-point 10-fold
107 serial dilution of DNA extracted from asexual NF54 cultures synchronized to ring stage. This gives a
108 threshold detection of ~500 parasite genomes per mosquito for the qPCR assay.

109

110 *Near infrared spectra selection*

111 A total of 634 *A. gambiae* (Keele strain) were scanned using NIRS (Table 1). DNA was extracted and
112 analyzed for *P. falciparum* infection by qPCR as described above. Samples with inconclusive qPCR
113 results or poor spectra quality were excluded (n=72). Poor quality or outlier spectra were visually
114 identified by comparing them to all other spectra, and spectra that were prominently flat or prominently
115 noisy were excluded, as described elsewhere [10]. Thus, NIR absorbance spectra and respective
116 infection status data from a final total of 562 mosquitoes were used to estimate the accuracy of NIRS
117 for prediction of malaria infection (Figure 1).

118

119 *Model prediction accuracy*

120 The relationship between spectra and infection was analyzed using partial least square regression
121 (PLS). Training datasets were used to perform multiple leave-one-out cross validations (LOOCV) and
122 develop two calibrations, one for prediction of oocyst infection and another for prediction of sporozoite
123 infection. The calibrations were then validated using test datasets composed of samples with unknown
124 infection status that had not been included in the calibration's training dataset. The number of factors
125 used in the calibration was 12, determined from the prediction residual error sum of squares (PRESS)

126 and regression coefficient plots (see supplementary information). In the PLS model, a value of “1” was
127 assigned to all the actual uninfected samples whereas a value of “2” was assigned to the actual
128 infected mosquitoes (infection as defined by the qPCR results). The PLS calibration derived
129 components used to transform the original spectra of each predicted independent sample into a PLS
130 score; a score value of 1.5 was considered as the threshold for correct or incorrect classification,
131 meaning any mosquito with PLS score below 1.5 was predicted as uninfected and equal or greater than
132 1.5 was predicted as infected. The PLS model showed that NIR spectra from both oocyst and
133 sporozoite infected mosquitoes were distinct from their counterpart uninfected mosquitoes with 91.2%
134 (86.7% - 94.5%) and 92.8% (87.1% - 96.5%) self-prediction accuracy respectively (Figure 2a and 3a).
135 When tested on samples with unknown infections status that had not been included in the training
136 dataset, the calibration maintained high sensitivity and specificity at both detecting oocyst and
137 sporozoite infection, with 87.7% (95%CI: 79.9% -93.3%; Cohen’s kappa=0.75)) and 94.5% (95%CI:
138 87.6% - 98.2%; Cohen’s kappa=0.86) prediction accuracy respectively (Figures 2b and 3b).

139 *Infection load and prediction accuracy*

140 The parasite load in a mosquito is of epidemiological importance as there is evidence of a continual
141 increase in transmission potential with increasing sporozoites numbers [27]. To test if the NIR prediction
142 output scores were affected by parasite load qPCR was done to estimate the relative number of
143 parasite genomes in each infected mosquito (Table 2) and used to evaluate the calibration model’s
144 accuracy. The oocyst-infected mosquitoes in the test data set had a range of infection loads (Median:
145 1925, IQR: [295 to 4883]). Two oocyst-infected mosquitoes were misclassified as uninfected, both of
146 which had relatively low infection loads (357 and 389 parasite genomes/ μ l of DNA extract) (Figure 4a).
147 Generalised linear mixed-effects models were used to investigate the effect of infection load and
148 infection presence on the PLS scores (response variable) of the predicted samples. The age of the
149 mosquitoes on the day of the infectious feed was included as a random effect. It was observed that the
150 presence of oocyst infection influenced the NIRS prediction score (Coefficient: 0.67; 95%CI: 0.41 to
151 0.93; $p < 0.001$) but the infection load did not (Coefficient: -0.000003; 95%CI: -0.0000074 to 0.0000015;
152 p -value: 0.21). The sporozoite-infected mosquitoes in the test dataset had a range of infection loads
153 (Median: 8841, IQR: [2516 to 20112]). Five sporozoite-infected mosquitoes were misclassified as
154 uninfected: two presented with the lowest infection loads of the test dataset (33 and 38 parasite
155 genomes/ μ l of DNA extract); the other three had relatively high infection loads (1156, 6660 and 12591

156 parasite genomes/ μ l of DNA extract) (Figure 4b). The presence of sporozoite significantly affected the
157 PLS scores of the predicted samples (Coefficient: 0.75; 95%CI: 0.51 to 1.00; p-value: <0.001) as did
158 the infection load (Coefficient: 0.0000019; 95%CI: 0.0000013 to 0.0000025; p-value<0.001).

159

160 **DISCUSSION**

161 This is the first study to show that NIRS can be used to accurately detect human malaria in *A. gambiae*
162 mosquitoes. NIRS predicted oocyst infection with 87.7% accuracy (79.9% - 93.3%) and sporozoite
163 infection with 94.5% accuracy (87.6% – 98.2%). The NIRS predictive accuracy for sporozoite infection
164 of >90% in this study concurs with previous work done using the rodent malaria in *Anopheles*
165 *stephensi*, which found that NIRS could detect the presence of sporozoites in infected mosquitoes with
166 77% accuracy [22]. Unlike the previous study, the present calibration model was also capable of
167 identifying oocyst-infected mosquitoes. The PLS calibration of the present study was based on a
168 narrower interval of the electromagnetic spectrum, 500 to 2400 nm, compared to 350 to 2500 nm. This
169 narrower range excludes noise present in the extremities of the spectra due to light source and sensor
170 limitations and therewith improved the prediction accuracy of the calibration model. Furthermore, the
171 previous study used spectra from mosquitoes that had been saturated with chloroform which was used
172 to knock them down. This contamination led to clear chloroform peaks in the NIR spectra which may
173 have added to the noise and reduced prediction accuracy of the calibration. Differences between the
174 vector species and parasite species may also have played a role in the small discrepancy of predictive
175 accuracy between studies. In addition, the experimental approach used in the present study, also
176 permitted to account for the potentially confounding effects of the infected bloodmeal, given that control
177 group had been fed the same blood but with inactivated gametocytes.

178

179 Near infrared light is absorbed differently by diverse biochemical compounds which, in the mosquito,
180 may consistently vary with between species, age and in this case infection status. It is hypothesized
181 that biochemical changes occurring in the mosquito, as a consequence of *P. falciparum* infection, made
182 it possible to distinguish between infected and uninfected mosquitoes using NIRS. Consistent
183 differences between the NIR absorbance spectra of infected and uninfected mosquitoes may be related
184 to the presence of parasite-specific molecules in the infected mosquitoes [28-30]. Also, it is possible
185 that tissue changes may occur in the mosquitos due to their immune response to the parasite which

186 could have an effect on the biochemical composition of the mosquito [28]. Additionally, it is known that
187 *Plasmodium* infection alters metabolic pathways in mosquitoes and leads to higher energy resource
188 storage [31] which may lead to differences in NIRS spectra. More research is needed to better
189 understand the underlying biochemical features that enable NIRS to distinguish between *Plasmodium*-
190 infected and uninfected mosquitoes.

191

192 The prediction accuracy of the NIRS calibration to detect sporozoite infection was influenced not only
193 by the presence of *Plasmodium falciparum* sporozoites but also the parasite load (number of parasite
194 genomes). This was not the case of the calibration to detect oocysts, which was only significantly
195 influenced by the presence of infection in the midgut. It is possible that slight differences in DNA
196 extraction efficiency between samples may have affected the estimate number of parasite genomes in
197 each insect sample and therefore it is imprudent to make conclusions on how strongly infection load
198 may be influencing the PLS output scores. The performance accuracy of NIRS was similar to qPCR
199 (sporozoite detection: Cohens kappa=0.86; oocyst detection: Cohens's kappa=0.75). The strong inter-
200 rate agreement between the two methods, suggests that NIRS may have similar sensitivity and
201 specificity as qPCR at detecting malaria sporozoites in the mosquito host. ELISA is less specific than
202 PCR [32], however due to its low-cost and ease, it is routinely the assay chosen by surveillance
203 programs to measure the proportion of mosquitoes that carry sporozoites and the entomological
204 inoculation rate (EIR). It is possible that EIR estimates could be improved by using a more accurate
205 diagnostic test. However, a direct comparison of NIRS and ELISA was not the objective of this study.
206 Presently NIRS still requires further optimization and validation in the field before being considered as a
207 possible replacement for ELISA in surveillance programs. While the results presented in this paper are
208 promising, NIRS calibrations generated using lab-reared mosquitoes do not necessarily represent the
209 diversity of vectors in the field, providing no guarantee of the robustness of the method when tested on
210 wild-caught mosquitoes. Calibrations must be based on training datasets that capture the diversity of
211 field-mosquitoes reducing confounders that may affect the classification accuracy, including, different
212 mosquito species, age, infection, size, insecticide resistance status, microbiome, and origin. NIRS is a
213 promising technology that may provide an accurate and high-throughput solution to monitoring malaria
214 transmission in the vector as progression towards elimination is made. Such a tool may revolutionize
215 how entomological data is used by control and research programmes given that the same test can

216 report various entomological parameters, including age, species and infection status, therewith
217 compiling vast information of epidemiological importance to understanding how vector populations and
218 malaria transmission are changing. Future research efforts and resources need to be directed at
219 evaluating the best way of generating and optimizing calibrations based on wild-caught mosquitoes for
220 each entomological parameter, and validating these using specimens from different ecological and
221 geographical regions.

222

223 .

224

225 **MATERIALS AND METHODS**

226 *Mosquitoes*

227 Mosquitoes from a colony of *A. gambiae* (Keele line) [26] were reared under standard insectary
228 conditions (26±1°C, 80% humidity, 12 hr light:12 hr dark cycle) at the University of Glasgow, Scotland,
229 UK. Larvae were fed on Tetramin tropical flakes and Tetra Pond Pellets (Tetra Ltd, UK). Pupae were
230 transferred into cages for adult emergence. Adult mosquitoes were fed *ad libitum* on 5% glucose
231 solution containing 0.05% (w/v) 4-aminobenzoic acid (PABA). SMFA was done with 3-6 days old
232 mosquitoes.

233

234 *Parasite culture and standard membrane feeding assays (SMFA)*

235 *P. falciparum* (NF54) parasites were cultured using standard methodology to produce infectious
236 gametocytes [33], using human blood and serum obtained from the Glasgow and West of Scotland
237 Blood Transfusion Service. Standard membrane feeding assays (SMFA) were conducted on three
238 different occasions using gametocytes produced *in vitro*: the first SMFA was done with a high
239 gametocyte density (approx. 1% gametocytes) and the two-subsequent feeds with a lower density (~
240 0.1% gametocytes) to produce more uninfected mosquitoes. For each SMFA, 300 female *A. gambiae*
241 s.s (Keele line) mosquitoes 3-6 days post emergence were distributed in pairs into 6 cups of 50
242 mosquitoes each. In the first SMFA, mosquitoes were 3,4 and 5 days old, in the second SMFA they
243 were 4,5 and 6 days old and in the third SMFA mosquitoes were 3 (2 pairs of cups) and 4 days old.
244 One cup of each pair was offered blood with infectious gametocytes and allowed to feed for 20 minutes.
245 The temperature of the membrane feeders was then reduced to below 30°C for 30 minutes to allow all

246 mature gametocytes to complete gametogenesis [34]. The remaining cups of mosquitoes were then
247 allowed to feed on the same blood, to produce control mosquitoes with zero infection rates, and thus
248 obtain a comparable control sample differing only in the complete absence of parasite infection.

249

250 *Near infrared spectra collection and data analysis*

251 After feeding, the blood-fed mosquitoes in each pot were maintained for 14 days under insectary
252 conditions and examined for oocyst and sporozoite development on day 7 and 14 days post infection
253 respectively. Mosquitoes were killed using chloroform vapour before collecting near infrared
254 absorbance spectra from each individual mosquito without any further processing, using a Labspec 4i
255 NIR spectrometer with an internal 18.6 W light source (ASD Inc, Longmont, CO) and ASD software RS³
256 per established protocols [10], but using a 3.2 mm-diameter bifurcated fibre-optic probe which
257 contained a single 600 micron collection fibre surrounded by six 600 micron illumination fibres. The
258 probe was placed 2.4 mm from a spectralon plate onto which the mosquitoes were placed for scanning.
259 Spectra between 500–2400 nm were analysed through leave-one-out cross validations (LOOCV) using
260 partial least square (PLS) regression in GRAMS Plus/IQ software (Thermo Galactic, Salem, NH). After
261 scanning, each mosquito carcass was stored individually at -80 °C in ATL lysis buffer (QIAGEN) until
262 DNA extraction, to perform qPCR to determine the infection status of the mosquito.

263

264 *DNA extraction and quantitative real-time polymerase chain reaction (qPCR)*

265 DNA was extracted using Qiagen DNeasy Blood & Tissue® DNA extraction kits from mosquito
266 abdomens (for mosquitoes analyzed 7 days post infectious feed) and whole mosquitoes (for
267 mosquitoes killed 14 days post infectious feed) and eluted in 50 µl of water. A 20 µl aliquot of the 50 µl
268 of extracted DNA for each mosquito was transferred to individual wells of DNASTable® 96 well plates
269 (Sigma-Aldrich) and allowed to air dry at room temperature. The plates were shipped to KEMRI
270 Wellcome Trust (Killifi, Kenya) for qPCR analysis. Samples were reconstituted in 20 µl of DNase-free
271 water and *P. falciparum* genome numbers present were quantified by qPCR. Quantification reactions
272 were performed in 15 µL volumes, containing 1.2 µl of 10 mM forward and reverse primers (377F: 5'
273 ACTCCAGAAGAAGAAGAGCAAGC-3'; 377R: 5'-TTCATCAGTAAAAAAGAATCGTCATC- 3'; 7.5 µL
274 of SYBR® Green PCR Master Mix, 1.1 µL of DNase-free water and 4 µL of sample DNA, using an
275 Applied Biosystems 7500 Real-Time PCR System. The cycling profile comprised an initial denaturation

276 of 95 °C for 900 s (holding stage) and then 40 amplification cycles of denaturation 95°C for 30s
277 (seconds), annealing 55°C for 20s and extension 68°C for 30s. At the end of amplification, melt curves
278 were produced with 15s denaturation at 95°C, followed by 60s at 60°C, 30 s at 95°C and 15 s at 60°C.
279 Parasite load was estimated for each sample by comparison with the standard curve drawn from the
280 DNA standards using Applied Biosystems 7500 software v2.0.6. Samples which amplified after 38
281 cycles, or which showed a shift in melt curve or two melt curve peaks were excluded.

282

283 DNA extracted from uninfected mosquitoes (abdomens and cephalothorax) were used as negative
284 controls, in addition to negative controls with no DNA. Standard curves were generated for each qPCR
285 run using a 5-point 10-fold serial dilution of DNA extracted from asexual NF54 cultures synchronized to
286 ring stage, starting with 100,000 parasites/ μ l (100,000 parasites; 10,000 parasites; 1,000 parasites; 100
287 parasites and 10 parasites), run in duplicate.

288

289 *Analysis using PLS leave-one-out cross-validations (LOOCV)*

290 The *P. falciparum* detection model was trained and tested according to previously published
291 methods[10] using partial least square(PLS) regression to develop a calibration based on a training
292 data set, which was then used to predict the infection status of samples contained in a test dataset and
293 therewith validate the prediction accuracy of the calibration.

294

295 Leave-one-out cross validation (LOOCV) was used to determine if NIR spectra of uninfected
296 mosquitoes were distinct from *P. falciparum*-infected mosquitoes, and to give information on the
297 prediction accuracy of the model to distinguish between infected and uninfected mosquitoes. LOOCV
298 is a k -fold cross validation, with k equal to n , the number of spectra in a training dataset. That means
299 that n separate times, the function approximator is trained on all the spectra except for one spectrum
300 and a prediction is made for that spectrum. Multiple LOOCV based on the training dataset were used to
301 develop a calibration file which was then used to test the predictive ability of the model on a spectra
302 collected from a test dataset (Figure 1).

303 The results from the qPCR were used to identify which individual mosquitoes, that had been fed an
304 infectious blood meal, had confirmed oocyst and sporozoite infections. This information was then
305 specified to each spectrum and these were randomly assigned to either the training dataset or the test

306 dataset whilst ensuring the same proportion of different mosquito ages was found in the training and
307 test datasets. All uninfected mosquitoes were from the group that had been fed blood without viable
308 gametocytes. A total of 69 sporozoite-infected and 69 uninfected mosquitoes that had been kept for 14
309 days post SMFA were used to perform multiple LOOCV and generate a calibration file. The same was
310 done using spectra from 121 oocyst-infected mosquitoes and 110 uninfected mosquitoes kept for 7
311 days post SMFA.

312 Two separate LOOCV were run to investigate the prediction accuracy of oocyst-infected vs. uninfected,
313 and sporozoite-infected and uninfected mosquitoes respectively. The models were run on Grams IQ
314 software (Thermo Galactic, Salem, NH) and a total of 12 latent factors were selected by visualizing the
315 prediction residual error sum of squares (PRESS) curve, and choosing the minimum number of factors
316 needed to reduce the prediction error of the model without overfitting it. Actual vs Predicted plots were
317 drawn by plotting the actual constituent values (coded as 1= uninfected and 2= infected) on the x axis,
318 and model predicted values on the y axis. Prediction values were generated according to previously
319 published methods [10], values below 1.5 were considered to be predicted as uninfected and values
320 equal to or above 1.5 predicted as infected

321 The mosquito spectra that had not been included in the training dataset used for developing the
322 calibration were randomly assigned to the test dataset to validate the performance accuracy of the
323 model serving as an independent set of samples. The calibrations generated for detecting *P. falciparum*
324 sporozoite and oocyst infection were validated using 69 sporozoite-infected and 22 uninfected, and, 53
325 oocyst-infected and 56 uninfected, respectively (Figure 1). This was done by generating a calibration
326 with 12 latent factors based on the training dataset which was then loaded into IQPredict software and
327 used to obtain PLS scores of the independent samples based on the predicted probability of infection,
328 with 1= predicted as uninfected, 2=predicted as infected and cut-off value of 1.5.

329

330 *Analysis of prediction accuracy*

331 Sensitivity was calculated to estimate of the model's ability to detect the presence of infection and
332 specificity as the model's ability to detect the absence of infection. Accuracy was calculated as the

333 overall prediction ability of the model (Table 3). Sensitivity, specificity, accuracy and respective exact
334 Clopper-Pearson confidence intervals were calculated using MedCalc for Windows, version 18.0
335 (MedCalc Software, Ostend, Belgium). Cohen's kappa was calculated in STATA/IC Version 13.as a
336 measure of inter-rate agreement between qPCR (reference test) and NIRS. The PLS scores of the
337 predicted independent samples were analyzed using generalized linear mixed-effects model in
338 STATA/IC Version 13.1. The response variable investigated was the PLS score generated from the
339 PLS calibration models. The effects of infection presence and infection intensity (number of parasite
340 genomes) on the PLS prediction value were investigated. Given that the age of a mosquito may affect
341 NIRS spectra and therewith the PLS score, mosquito age was included as a random effect in the
342 model. Regression coefficients for each factor, confidence intervals and p-values were reported. Model
343 selection was done based on the Akaike information criterion (i.e. the lower the AIC value, the better
344 the model).

345

346 **DATA AVAILABILITY**

347 All the data necessary to interpret and replicate the finding on this paper have been made publicly
348 available on the data repository Harvard dataverse (<https://doi.org/10.7910/DVN/YD34OX>). This
349 includes details on the mean number of parasite genomes of each individual sample; NIR spectra of all
350 the specimens (spc files) with specification to whether they had been included in the training dataset or
351 test dataset for oocyst or sporozoite calibration; calibration file (cal file) for oocyst and sporozoite
352 prediction; GRAMS IQ training files (tdfx file) for oocyst and sporozoite prediction; as well as the
353 prediction outputs from IQ Predict for each sample in the test datasets (xls file).

354 **ACKNOWLEDGEMENTS**

355 The authors acknowledge the Swiss National Foundation of Science for the funding provided to MFM
356 through the Marie-Heim Voegtlin fellowship scheme (PMPDP3-164444) and AXA RF fellowship (14-
357 AXA-PDOC-130) and an EMBO LT fellowship (43-2014) for funding to FB. The authors also wish to
358 thank the Elizabeth Peat and Dorothy Armstrong for the production of mosquitoes at the University of
359 Glasgow and Laura Ciuffreda (supported by EU-FP7 MCSA-ITN Lapaso (607350)) for the ring stage

360 synchronized parasite culture. The authors also thank the Initiative to Develop African Leaders Program
361 (IDeAL) for funding Michelle Muthui and Martin Wagah.

362

363 **AUTHORS CONTRIBUTIONS**

364 MFM designed the experiment, cultured the parasites, assisted with the SMFAs, scanned the
365 mosquitoes, analysed the data and drafted the manuscript. MK provided the DNA standards, provided
366 training and commented on the final draft of the manuscript. MM optimized the qPCR method and
367 trained MW. MW performed qPCRs. HF provided mentorship to MFM, was involved in the experimental
368 design and commented on the final manuscript draft. FD provided mentorship to MFM, contributed to
369 the experimental design and data analysis. FB and LRC contributed to the experimental design, setup
370 the parasite culture, led the SMFAs, provided training to MFM in asexual and sexual culture of
371 *Plasmodium falciparum* NF54 as well as contributed to the final manuscript. All authors read and
372 commented on drafts of the manuscript and approved the final version.

373

374 **COMPETING INTERESTS STATEMENT**

375 The authors declare no competing interests. Mention of trade names or commercial products in this
376 publication is solely for the purpose of providing specific information and does not imply
377 recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity
378 provider and employer.

379

380 **REFERENCES**

- 381 1. WHO. Global technical strategy for malaria 2016–2030. Geneva: 2015.
- 382 2. RBM. Lessons Learned from fifteen years of responding to malaria globally: A prototype for sustainable
383 development. New York: Roll Back Malaria Partnership, 2016.
- 384 3. WHO. World Malaria Report 2018. Geneva, Switzerland: 2018.
- 385 4. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on
386 *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526(7572):207-11. Epub 2015/09/17.
387 doi: nature15535 [pii]
388 10.1038/nature15535 [doi]. PubMed PMID: 26375008.
- 389 5. MacDonald G. Epidemiological basis of malaria control. Bull WHO. 1956;15:613-26.

- 390 6. Burkot TR, Zavala F, Gwadz RW, Collins FH, Nussenzweig RS, Roberts DR. Identification of malaria-
391 infected mosquitoes by a two-site enzyme-linked immunosorbent assay. *Am J Trop Med Hyg.* 1984;33(2):227-31.
392 Epub 1984/03/01. PubMed PMID: 6370003.
- 393 7. Stoffels JA, Docters van Leeuwen WM, Post RJ. Detection of Plasmodium sporozoites in mosquitoes by
394 polymerase chain reaction and oligonucleotide rDNA probe, without dissection of the salivary glands. *Med Vet*
395 *Entomol.* 1995;9(4):433-7. Epub 1995/10/01. PubMed PMID: 8541598.
- 396 8. Sandeu MM, Moussiliou A, Moiroux N, Padonou GG, Massougbodji A, Corbel V, et al. Optimized Pan-
397 species and speciation duplex real-time PCR assays for Plasmodium parasites detection in malaria vectors. *PLoS*
398 *One.* 2012;7(12):e52719. Epub 2013/01/04. doi: 10.1371/journal.pone.0052719. PubMed PMID: 23285168;
399 PubMed Central PMCID: PMC3532469.
- 400 9. Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in Plasmodium falciparum
401 transmission: precision, accuracy and costs of metrics. *Adv Parasitol.* 2014;84:151-208. Epub 2014/02/01. doi:
402 10.1016/b978-0-12-800099-1.00003-x. PubMed PMID: 24480314; PubMed Central PMCID: PMC4847140.
- 403 10. Mayagaya VS, Michel K, Benedict MQ, Killeen GF, Wirtz RA, Ferguson HM, et al. Non-destructive
404 determination of age and species of Anopheles gambiae s.l. using near-infrared spectroscopy. *Am J Trop Med*
405 *Hyg.* 2009;81(4):622-30. PubMed PMID: 19815877.
- 406 11. Sikulu M, Killeen GF, Hugo LE, Ryan PA, Dowell KM, Wirtz RA, et al. Near-infrared spectroscopy as a
407 complementary age grading and species identification tool for African malaria vectors. *Parasit Vectors.*
408 2010;3:49. Epub 2010/06/08. doi: 1756-3305-3-49 [pii]
409 10.1186/1756-3305-3-49 [doi]. PubMed PMID: 20525305; PubMed Central PMCID: PMC2902455.
- 410 12. Sikulu-Lord MT, Maia MF, Milali MP, Henry M, Mkandawile G, Kho EA, et al. Rapid and Non-destructive
411 Detection and Identification of Two Strains of Wolbachia in Aedes aegypti by Near-Infrared Spectroscopy. *PLoS*
412 *Negl Trop Dis.* 2016;10(6):e0004759. doi: 10.1371/journal.pntd.0004759. PubMed PMID: 27362709; PubMed
413 Central PMCID: PMC4928868.
- 414 13. Sikulu-Lord MT, Milali MP, Henry M, Wirtz RA, Hugo LE, Dowell FE, et al. Near-Infrared Spectroscopy, a
415 Rapid Method for Predicting the Age of Male and Female Wild-Type and Wolbachia Infected Aedes aegypti. *PLoS*
416 *Negl Trop Dis.* 2016;10(10):e0005040. doi: 10.1371/journal.pntd.0005040. PubMed PMID: 27768689; PubMed
417 Central PMCID: PMC5074478.

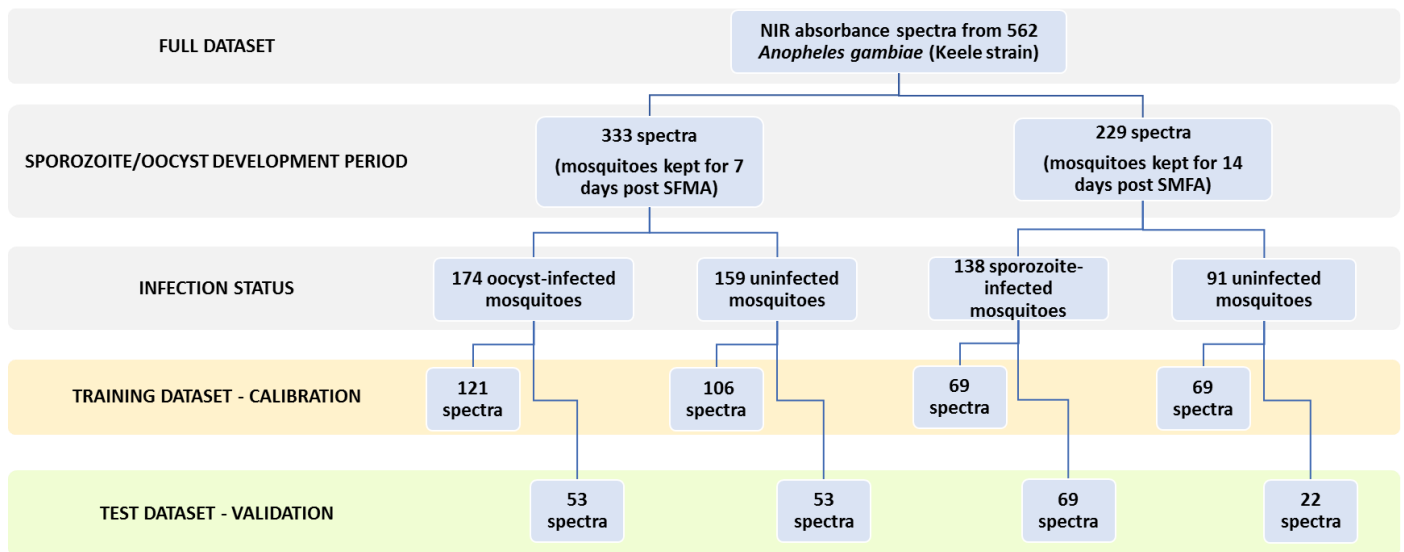
- 418 14. Caputo B, Dani FR, Horne GL, Petrarca V, Turillazzi S, Coluzzi M, et al. Identification and composition of
419 cuticular hydrocarbons of the major Afrotropical malaria vector *Anopheles gambiae* s.s. (Diptera: Culicidae):
420 analysis of sexual dimorphism and age-related changes. *J Mass Spectrom.* 2005;40(12):1595-604. PubMed PMID:
421 16320293.
- 422 15. Sikulu MT, Monkman J, Dave KA, Hastie ML, Dale PE, Kitching RL, et al. Proteomic changes occurring in
423 the malaria mosquitoes *Anopheles gambiae* and *Anopheles stephensi* during aging. *Journal of proteomics.*
424 2015;126:234-44. Epub 2015/06/24. doi: 10.1016/j.jprot.2015.06.008. PubMed PMID: 26100052.
- 425 16. Carlson DA, Service MW. Differentiation between species of the *Anopheles gambiae* Giles complex
426 (Diptera: Culicidae) by analysis of cuticular hydrocarbons. *Ann Trop Med Parasitol.* 1979;73(6):589-92. Epub
427 1979/12/01. PubMed PMID: 539859.
- 428 17. Carlson DA, Service MW. Identification of mosquitoes of *Anopheles gambiae* species complex A and B
429 by analysis of cuticular components. *Science.* 1980;207(4435):1089-91. Epub 1980/03/07. PubMed PMID:
430 7355276.
- 431 18. Baldrige GD, Baldrige AS, Witthuhn BA, Higgins L, Markowski TW, Fallon AM. Proteomic profiling of a
432 robust *Wolbachia* infection in an *Aedes albopictus* mosquito cell line. *Molecular microbiology.* 2014;94(3):537-
433 56. Epub 2014/08/27. doi: 10.1111/mmi.12768. PubMed PMID: 25155417; PubMed Central PMCID:
434 PMC4213348.
- 435 19. Suarez E, Nguyen HP, Ortiz IP, Lee KJ, Kim SB, Krzywinski J, et al. Matrix-assisted laser
436 desorption/ionization-mass spectrometry of cuticular lipid profiles can differentiate sex, age, and mating status
437 of *Anopheles gambiae* mosquitoes. *Analytica chimica acta.* 2011;706(1):157-63. Epub 2011/10/15. doi:
438 10.1016/j.aca.2011.08.033. PubMed PMID: 21995923.
- 439 20. Serrano-Pinto V, Acosta-Perez M, Luviano-Bazan D, Hurtado-Sil G, Batista CV, Martinez-Barnette J, et
440 al. Differential expression of proteins in the midgut of *Anopheles albimanus* infected with *Plasmodium berghei*.
441 *Insect Biochem Mol Biol.* 2010;40(10):752-8. Epub 2010/08/10. doi: 10.1016/j.ibmb.2010.07.011. PubMed PMID:
442 20692341.
- 443 21. Chotiwan N, Andre BG, Sanchez-Vargas I, Islam MN, Grabowski JM, Hopf-Jannasch A, et al. Dynamic
444 remodeling of lipids coincides with dengue virus replication in the midgut of *Aedes aegypti* mosquitoes. *PLoS*
445 *pathogens.* 2018;14(2):e1006853. Epub 2018/02/16. doi: 10.1371/journal.ppat.1006853. PubMed PMID:
446 29447265; PubMed Central PMCID: PMC5814098.

- 447 22. Esperança PM, Blagborough AM, Da DF, Dowell FE, Churcher TS. Detection of *Plasmodium berghei*
448 infected *Anopheles stephensi* using near-infrared spectroscopy. *Parasites & vectors*. 2018;11(1):377. doi:
449 10.1186/s13071-018-2960-z.
- 450 23. Fernandes JN, Dos Santos LMB, Chouin-Carneiro T, Pavan MG, Garcia GA, David MR, et al. Rapid,
451 noninvasive detection of Zika virus in *Aedes aegypti* mosquitoes by near-infrared spectroscopy. *Science*
452 *advances*. 2018;4(5):eaat0496. Epub 2018/05/29. doi: 10.1126/sciadv.aat0496. PubMed PMID: 29806030;
453 PubMed Central PMCID: PMC5966221.
- 454 24. Vignali M, Speake C, Duffy PE. Malaria sporozoite proteome leaves a trail. *Genome Biol*. 2009;10(4):216.
455 Epub 2009/05/14. doi: 10.1186/gb-2009-10-4-216. PubMed PMID: 19435488; PubMed Central PMCID:
456 PMC2688921.
- 457 25. Marie A, Holzmüller P, Tchioffo MT, Rossignol M, Demette E, Seveno M, et al. *Anopheles gambiae*
458 salivary protein expression modulated by wild *Plasmodium falciparum* infection: highlighting of new antigenic
459 peptides as candidates of *An. gambiae* bites. *Parasites & vectors*. 2014;7:599. Epub 2014/12/21. doi:
460 10.1186/s13071-014-0599-y. PubMed PMID: 25526764; PubMed Central PMCID: PMC4287575.
- 461 26. Ranford-Cartwright LC, McGeechan S, Inch D, Smart G, Richterová L, Mwangi JM. Characterisation of
462 Species and Diversity of *Anopheles gambiae* Keele Colony. *PLoS ONE*. 2016;11(12):e0168999
- 463 27. Churcher TS, Sinden RE, Edwards NJ, Poulton ID, Rampling TW, Brock PM, et al. Probability of
464 Transmission of Malaria from Mosquito to Human Is Regulated by Mosquito Parasite Density in Naive and
465 Vaccinated Hosts. *PLoS Pathog*. 2017;13(1):e1006108. Epub 2017/01/13. doi: 10.1371/journal.ppat.1006108.
466 PubMed PMID: 28081253; PubMed Central PMCID: PMC5230737.
- 467 28. Sinden RE. The cell biology of malaria infection of mosquito: advances and opportunities. *Cellular*
468 *microbiology*. 2015;17(4):451-66. Epub 2015/01/06. doi: 10.1111/cmi.12413. PubMed PMID: 25557077;
469 PubMed Central PMCID: PMC4409862.
- 470 29. Robson KJ, Frevert U, Reckmann I, Cowan G, Beier J, Scragg IG, et al. Thrombospondin-related adhesive
471 protein (TRAP) of *Plasmodium falciparum*: expression during sporozoite ontogeny and binding to human
472 hepatocytes. *The EMBO journal*. 1995;14(16):3883-94. Epub 1995/08/15. PubMed PMID: 7664729; PubMed
473 Central PMCID: PMC394467.

- 474 30. Beier JC. Frequent blood-feeding and restrictive sugar-feeding behavior enhance the malaria vector
475 potential of *Anopheles gambiae* s.l. and *An. funestus* (Diptera: Culicidae) in western Kenya. *J Med Entomol.*
476 1996;33(4):613-8. Epub 1996/07/01. PubMed PMID: 8699456.
- 477 31. Zhao YO, Kurscheid S, Zhang Y, Liu L, Zhang L, Loeliger K, et al. Enhanced survival of *Plasmodium*-
478 infected mosquitoes during starvation. *PLoS One.* 2012;7(7):e40556. Epub 2012/07/19. doi:
479 10.1371/journal.pone.0040556. PubMed PMID: 22808193; PubMed Central PMCID: PMC3393683.
- 480 32. Marie A, Boissiere A, Tsapi MT, Poinsignon A, Awono-Ambene PH, Morlais I, et al. Evaluation of a real-
481 time quantitative PCR to measure the wild *Plasmodium falciparum* infectivity rate in salivary glands of *Anopheles*
482 *gambiae*. *Malaria journal.* 2013;12:224. Epub 2013/07/04. doi: 10.1186/1475-2875-12-224. PubMed PMID:
483 23819831; PubMed Central PMCID: PMC3707787.
- 484 33. Carter R, Ranford-Cartwright L, Alano P. The culture and preparation of gametocytes of *Plasmodium*
485 *falciparum* for immunochemical, molecular and mosquito infectivity studies. In: Hyde JE, editor. *Methods in*
486 *Molecular Biology.* 21: Protocols in Molecular Parasitology. Totowa, NJ: Humana Press Inc., ; 1993. p. 67-88.
- 487 34. Sinden RE, Croll NA. Cytology and kinetics of microgametogenesis and fertilization in *Plasmodium yoelii*
488 *nigeriensis*. *Parasitology.* 1975;70(1):53-65. Epub 1975/02/01. PubMed PMID: 1118188.
- 489
490
491
492
493
494
495
496
497
498
499
500
501
502

503 **FIGURES AND LEGENDS**

504



505

506

507 **Figure 1** – Study flow chart showing number of spectra collected, infection status and random

508 assignment of spectra to either training or test dataset.

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523
524
525
526
527
528
529
530
531
532
533

a. NIRS PLS calibration – oocyst infections

b. PLS model validation – oocysts infection



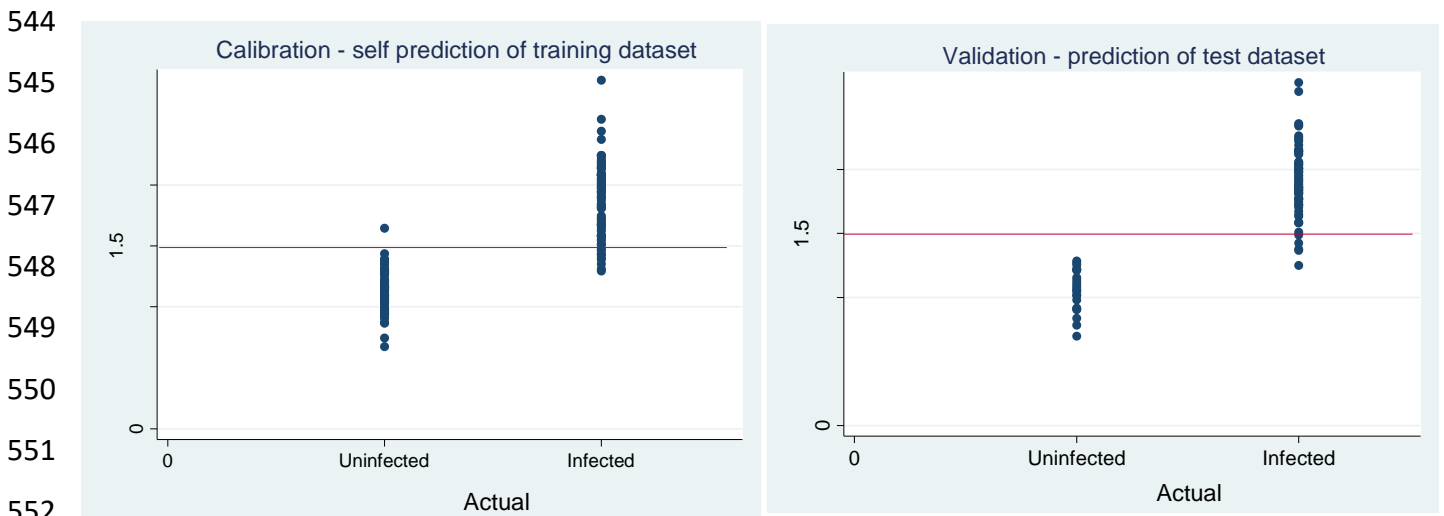
Training dataset				Test dataset			
(self-prediction accuracy)				(samples of unknown infection status)			
		Actual				Actual	
		Infected	Uninfected			Infected	Uninfected
Predicted	Infected	112	11	Predicted	Infected	51	11
	Uninfected	9	95		Uninfected	2	42
		95% CI				95% CI	
Sensitivity	92.6%	86.4% - 96.5%		Sensitivity	96.2%	87.0% - 99.5%	
Specificity	89.6%	82.2% - 94.7%		Specificity	79.3%	65.9% - 89.2%	
Accuracy	91.2%	86.7% - 94.5%		Accuracy	87.6%	79.9% - 93.3%	

534
535
536
537
538
539
540
541
542

Figure 2 – Actual versus predicted plots of oocyst infected mosquitoes investigating NIRS as diagnostic method. Sensitivity, specificity, accuracy and respective 95% confidence intervals of self-prediction of *P. falciparum*-infection in training dataset (left) and prediction of samples of unknown status in test dataset (right) (PLS scores: 1= uninfected, 2= infected and 1.5 as cut-off value).

543 **a. NIRS PLS calibration – oocyst infections**

543 **b. PLS model validation – oocysts infection**



Training dataset (self-prediction accuracy)				Test dataset (samples of unknown infection status)			
		Actual				Actual	
		Infected	Uninfected			Infected	Uninfected
Predicted	Infected	60	1	Predicted	Infected	64	0
	Uninfected	9	68		Uninfected	5	22
		95% CI				95% CI	
Sensitivity	87.0%	76.7% - 93.9%		Sensitivity	92.8%	83.9% - 97.6%	
Specificity	98.6%	92.2% - 100%		Specificity	100%	84.6% - 100%	
Accuracy	92.8%	87.1% - 96.5%		Accuracy	94.5%	87.6 – 98.2%	

554

555 **Figure 3** – Actual versus predicted plots of sporozoite infected mosquitoes investigating NIRS as
 556 diagnostic method. Sensitivity, specificity, accuracy and respective 95% confidence intervals of self-
 557 prediction of *P. falciparum*-infection in training dataset (left) and prediction of samples of unknown
 558 status in test dataset (right) (PLS scores: 1= uninfected, 2= infected and 1.5 as cut-off value).

559

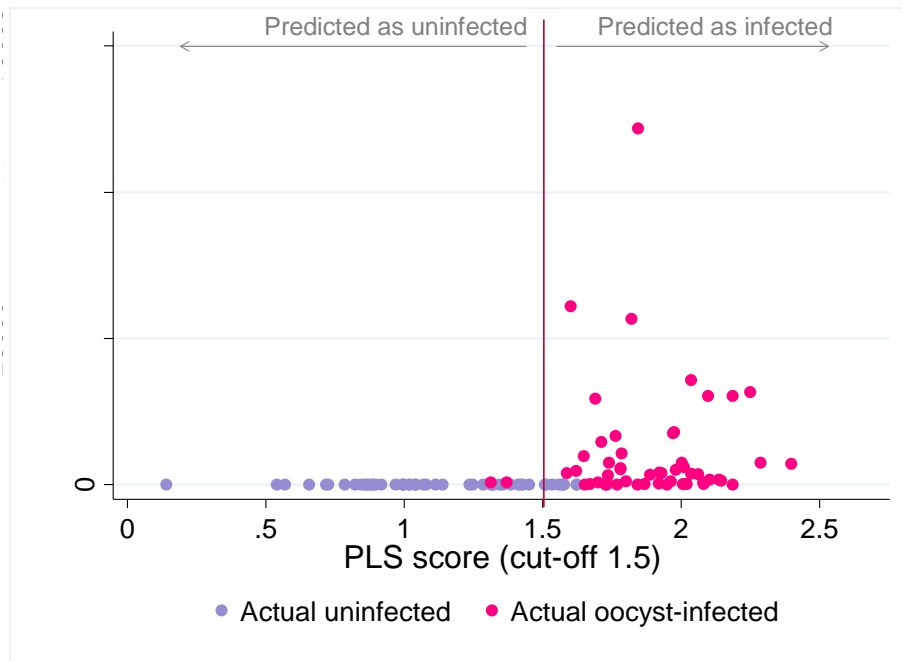
560

561

562

563

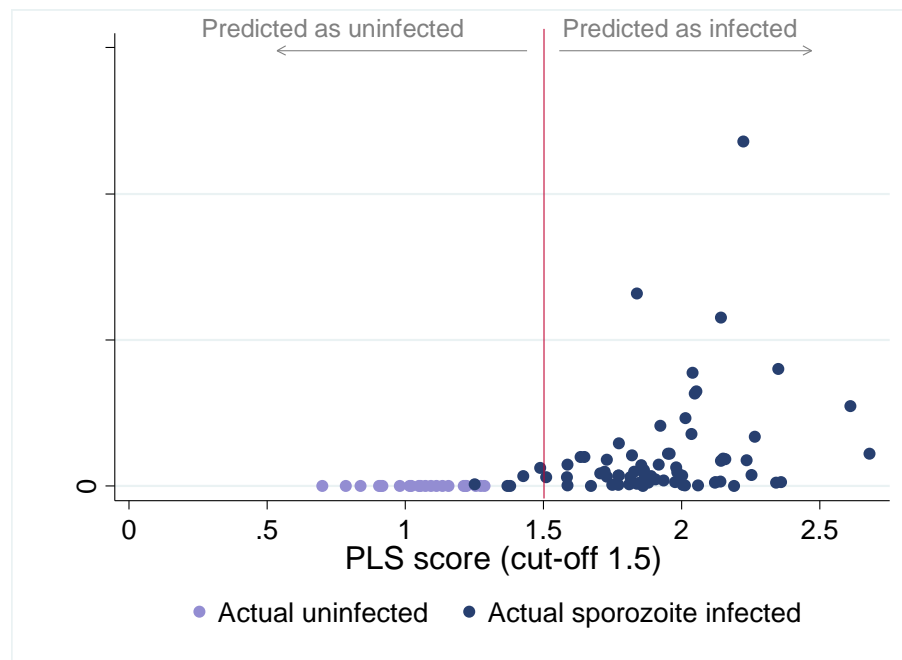
a. Oocyst infections



564

565

b. Sporozoite infections



566

567

568 **Figure 4** – Intensity of *P. falciparum* oocyst (a) and sporozoite (b) infection, quantified as the number of
569 parasite genomes per μl of DNA extract, in *A. gambiae* mosquitoes and prediction value score based
570 on the predicted probability of infection, with 1= predicted as not infected and 2=predicted as infected
571 (cut-off value of 1.5)

572

573 **TABLES AND CAPTIONS**

574

575 **Table 1** – Description of the gametocytemia used for each of the three standard membrane feeding
 576 assays (SMFA), number of days kept post blood feeding, number of mosquitoes processed by
 577 quantitative PCR (qPCR), % prevalence, and the intensity of infection described as the median and
 578 interquartile range (IQR) of the number of parasite genomes present in infected mosquitoes, excluding
 579 mosquitoes with no infection.

SMFA	Estimated gametocytemia	Day post-infectious blood meal	No. mosquitoes tested (n=634)	Positive (N=423)	% prevalence of infection	Intensity of infection: median number of parasite genomes and IQR.
1	1%	7	175	105	60.0%	680 (283-1625)
		14	n.d.	n.d.	n.d.	n.d.
2	0.1%	7	104	73	70.2%	456 (67-2052)
		14	99	47	47%	516 (211 – 6081)
3	0.1%	7	114	85	74.6%	2995 (3210-8881)
		14	142	113	80%	10114 (2540 – 29145)

580

581

582

583

584

585

586

587

588

589

590

591

592

593 **Table 2** - Generalised linear mixed-effects models investigating the effect of infection presence

594 (infected or uninfected) and infection load (number of parasite genomes/ μ l of DNA extract quantified

595 using qPCR) on the PLS score of the predicted samples including mosquito age as a random effect.

	Coefficient	Robust	z	95% Confidence	P value
	s	standard error		intervals	
Oocyst infections					
Infection presence	0.67	0.13	5.11	0.41 to 0.93	<0.001
Infection load	-0.000003	-0.000002	-1.26	-0.0000074 to 0.0000015	0.21
<i>Mosquito age</i> (<i>random effect</i>)	.018	.005	-	0.01 to 0.03	-
Sporozoite infections					
Infection presence	0.75	0.12	6.03	0.51 to 1.00	<0.001
Infection load	0.0000019	0.0000003	5.82	0.0000013 to 0.0000025	<0.001
<i>Mosquito age</i> (<i>random effect</i>)	0.007	0.008	-	0.0032 to 0.087	-

596

597

598 **Table 3**- Sensitivity, specificity and accuracy as measures of the performance of a binary classification

599 test TP – True positives; TN – True negative; FP – False positive; and FN – False negatives

Sensitivity	Specificity	Accuracy
$\frac{TP}{TP + FN}$	$\frac{TN}{TN + FP}$	$\frac{TP + TN}{TP + TN + FP + FN}$

600