

Metabolomic analysis of urine samples of vivax malaria in-patients for biomarker identification

Arjun Sengupta · Soumita Ghosh · Sulabha Pathak ·
Nithya Gogtay · Urmilla Thatte · Maulik Doshi ·
Shobhona Sharma · Haripalsingh M. Sonawat

Received: 6 October 2014 / Accepted: 17 February 2015 / Published online: 10 March 2015
© Springer Science+Business Media New York 2015

Abstract *Plasmodium vivax* is the most dominating species in the Indian subcontinent among the five *Plasmodium* species causing malarial infection in humans. This species is generally believed to be benign and classically thought to cause uncomplicated clinical infection that rarely leads to death or other complication. However, recent studies have shown incidences of *P. vivax* infection leading to severe complications and even death. To understand the biochemical changes at the pathway level and the modality through which the host recovers from the infection, we monitored the small molecular weight metabolites of urine samples of hospital in-patients undergoing treatment for vivax malaria. Metabolomics, involving ^1H NMR spectra of urine samples subjected to multivariate statistical analyses, was employed for this purpose. Urine metabolic profiles were used to identify putative biomarkers and establishing correlation with parasite count in the peripheral blood. The untreated

malaria patients (day1) exhibited altered excretion of taurine, hippurate, citrate, glycine, 3-methylhistidine and alanine in comparison to the self-control profile on day30. The day1–4 urine metabolite profiles are distinct from day5 and day30 profiles. The latter two profiles are indistinguishable. Temporal trajectories of urinary metabolites indicated an inflection point on day3. The day1 ^1H NMR profiles, distinct from the profiles of viral fever patients, correlated with the parasitemia levels in the patients and could in fact be used to gauge the anemia status of the patients. This study demonstrates the potential of urine metabolic profiling in understanding the pathophysiology of *P. vivax* infection and in non-invasive disease diagnosis.

Keywords Metabolomics · NMR · Malaria · Recovery · *Plasmodium vivax* · Metabolism

1 Introduction

Five species of *Plasmodium* are known to cause malaria in humans—*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, and recently, *P. knowlesi*. Most attempts to understand malarial disease biology are centered over the more severe infections caused by *Plasmodium falciparum*. However, *P. vivax* malaria requires more attention for a number of reasons. First, in the Indian subcontinent *P. vivax* dominates in terms of infections/year (Joshi et al. 2008). Second, believed to be benign in the past, recent reports suggest that it too can result in cerebral complications, severe disease manifestations, and chloroquine resistance (Kochar et al. 2009; Mendis et al. 2001; Tekka et al. 2008). This is possibly because of the continuous evolution of the parasite and its changing dynamic interactions with the environment.

Electronic supplementary material The online version of this article (doi:10.1007/s11306-015-0791-x) contains supplementary material, which is available to authorized users.

A. Sengupta · S. Ghosh · H. M. Sonawat (✉)
Department of Chemical Sciences, Tata Institute of Fundamental
Research, Homi Bhabha Road, Mumbai 400 005, India
e-mail: hms@tifr.res.in

S. Pathak · S. Sharma (✉)
Department of Biological Sciences, Tata Institute of
Fundamental Research, Homi Bhabha Road, Mumbai 400 005,
India
e-mail: sharma@tifr.res.in

N. Gogtay · U. Thatte · M. Doshi
Department of Clinical Pharmacology, Seth GS Medical College
& King Edward Memorial Hospital, Parel, Mumbai 400 012,
India

The misdiagnosis of the infection remains another caveat. Malarial infections are frequently misdiagnosed or over-diagnosed in endemic regions of the world, leading to considerable over diagnosis of the infection and consequent misdiagnosis of other fevers (Nankabirwa et al. 2009). One important factor, contributing to this is the unfamiliarity with the microscopic diagnosis technique and the inherent lower sensitivity of the technique (Hanscheid 2003). Small scale studies report misdiagnosis of $\sim 10\%$, a figure that is likely to be increased during large scale field studies (Hanscheid 2003). Therefore, it is important to identify new diagnostic techniques, in particular those that are non-invasive and comparatively rapid. In this paper we show that the urine metabolic profile of *P. vivax* infected patients is distinct from viral fever cases, which are often misdiagnosed for each other. We also identify malaria specific metabolites in the urine of patients suffering from *P. vivax* infection using the metabolomic approach. This technique, involving NMR in conjunction with multivariate data analyses, has been used for over a decade and is specifically termed metabolomics/metabonomics (Nicholson and Lindon 2008).

Several factors, such as genomic make-up and dietary habits, will contribute towards differences in urine metabolites even in a basal healthy state, and a comparison across different patients will have inherent differences. In order to minimize such variations, the urine of each hospital *P. vivax* in-patient was monitored over a period of five days starting from their admission, using ^1H NMR spectroscopy and multivariate data analysis. Each of the patients was parasite-free and discharged by day5 as judged from peripheral blood smear analysis and their clinical features. A 30th day sample of the patients served as a healthy control sample for each patient. Urine samples from viral fever patients were included to ensure that the biomarkers and metabolic profile identified were malaria-specific and not a feature of the febrile condition of the patient.

We find that malarial infection causes discrete changes in the patient's urine profile and such changes are distinct from those observed in patients with viral fever. Monitoring these changes may allow us to understand the underlying molecular mechanism of malarial disease progression. In addition, we also show that the multivariate analysis of urine profile from the patients may be used as a correlate for disease diagnosis and also for evaluating parameters such as the parasitemia associated with malaria.

2 Materials and methods

2.1 Study design and recruitment of the patients

This study was approved by the local Institutional Ethics Committee and written informed consent was obtained

from each patient prior to his/her enrolment in the study. The study was designed in the following way. *P. vivax* infected patients admitted in KEM hospital in Mumbai were recruited for the study after medical screening for malarial infection by examination of thick and thin blood smear. Informed consent was obtained. The day of admission was considered day1. The first midstream urine samples were collected once daily from the patients for a maximum of five days. During this period the patients underwent standard treatment and care for uncomplicated malaria and were usually discharged on day5, although some patients (20) were discharged earlier. Specifically, the patients were treated with 25 mg/kg of chloroquine base for three days followed by 15 mg/day of primaquine phosphate for 14 days. As a follow-up they were called back after a month and day30 urine samples were collected on that day. Absence of infection was confirmed by examining their peripheral blood smear. The day30 sample served as control and was used for comparative analysis with day1–5 samples. 106 *P. vivax* patients and 20 viral fever patients were enrolled in this study. Complete pathological reports were available for 67 of the 106 patients. The pathological tests included CBC, liver function tests (SGOT, SGPT, bilirubin and alkaline phosphatase) and kidney function (blood urea nitrogen and serum creatinine). The day30 sample could be obtained for only 44 patients, and these patients were found to have recovered completely from malaria and were negative by peripheral blood smear test. Comparisons between the infected and self-control urine samples were possible only for these 44 patients. Samples from a subset of the population (61 patients) were used for the prediction of presence of infection for the model we constructed. 0.02 % sodium-azide was added to all the urine samples before storing at -86°C until NMR analysis.

2.2 Preparation of urine samples and NMR experiments

800 μl of urine sample was added to 400 μl of 0.02 M phosphate buffer (an 81:19 (v/v) mixture of 0.2 M Na_2HPO_4 and 0.2 M NaH_2PO_4 ; pH 7.4, made in deionized water passed through a 0.22 μm filter). This was allowed to equilibrate at ambient temperature for 5 min and then centrifuged at $6000\times g$ for 10 min. The supernatant samples were stored at -86°C until further use for the NMR experiments. For acquiring ^1H NMR spectra, the samples were thawed at room temperature and 600 μl of each sample was transferred to 5 mm NMR tube (Norell Inc). 100 μl of 0.01 % DSS in D_2O was added to it for the purpose of internal referencing and field frequency lock and the samples were used for the NMR experiments. The NMR spectra of the samples were acquired in an AVANCE

700 MHz spectrometer (Bruker Biospin) equipped with a broad band inverse probe and BACS automation system. The one dimensional ^1H NMR spectra were acquired by means of a pulse program which takes the form of first transient of the NOESY NMR spectroscopy. The spectra were acquired with 32 transients, each recorded with 64 k data points on a spectral width of 20 ppm. This resulted in acquisition time of 2.28 s; the inter-scan delay applied was 4 s.

The spectra were processed (TOPSPIN 2.1, Bruker) using an additional line broadening factor of 0.1 Hz, Fourier transformed and phase and baseline corrected using an automated program provided by Dr. Eberherd Humpfer (Bruker Biospin, Germany).

In order to assign the metabolites, several 2-dimensional NMR experiments were carried out. These included the J-resolved, COSY, TOCSY and HMQC NMR techniques. Details of these experiments can be found elsewhere (8). While acquiring the 1-dimensional spectra, the water signal was suppressed using continuous irradiation on the water signal during the relaxation delay and the mixing time. Similar strategy was adopted while acquiring the J-resolved, COSY and HMQC spectra where continuous irradiation on the water signal took place during the relaxation delay. During the TOCSY experiments, the water suppression routine involved excitation sculpting and gradient pulses.

2.3 Statistical analysis of data

2.3.1 Multivariate data analyses

PCA and OPLSDA For the purpose of multivariate data analysis, the ^1H NMR spectra were binned into 0.04 ppm window and each bin was integrated and normalized to the total intensity in order to take care of the dilution effects. This procedure was achieved using Amix 3.8.4 (Bruker Biospin, Switzerland). The region comprising of resonances from urea and water (4.5–5.5 ppm) were excluded from this procedure. In addition, several patients were found to be under medication of paracetamol. The NMR spectra of day1 urine samples of these patients had significant resonances from the paracetamol metabolites. The spectral regions (7.54–7.14 ppm, 3.86–3.82 ppm, 3.66–3.58 ppm, 2.94–2.9 ppm, 1.90–1.86 ppm, 1.78–1.74 ppm) corresponding to these were also excluded from the binning process. The data matrix so generated was imported into SIMCA-P+12.0 for further multivariate data analysis. Prior to the modeling, the data matrix was mean centered and Pareto scaled. Initially the day1–5 urine profiles of the patients were compared with the day 30 urine profiles using PCA to check for the pattern in the data and to find out any outliers. PCA is an unsupervised data

analysis technique that is based on the reduction of dimensionality. Briefly, the highly correlated multivariate data structure, such as the urine NMR profiles, are broken down to fewer (usually 2–4) mutually orthogonal variables called the Principal Components (PCs). Each PC is a linear combination of the original variables. This procedure is able to achieve the visualization of the latent pattern in the complex multivariate data structure by means of the scores variables that are generated in the PC space. The variables responsible for the scores pattern can be identified from the loadings, which are the weightage by which the original variables combine to form the PCs. However, one disadvantage of PCA is that it provides information about the most dominating variations present in the data, which may not be the one that is sought. This can be handled by using supervised data analysis techniques such as OPLS. This technique models the variation of a defined Y variable with that of the data matrix. In case of a class entity of the Y variable, this takes the shape of a discriminant analysis and is known as the OPLSDA. This sort of modeling splits the data matrix into two variations, one that is related to Y and the other that is orthogonal. Therefore, two components are generated. The predictive component and the corresponding scores imply the variation in X that is related to Y and the orthogonal component describes the variation in X that is orthogonal to Y. Like for PCA, the variables that are responsible for the variations are found using the loadings on the predictive component. Two parameters, namely, the R^2Y and the $Q^2(\text{cum})$ can be used to judge the model. The former implies the explained variation in Y while the latter is a cross validation parameter and judges the predictive ability of the model. For a discriminant analysis $Q^2(\text{cum})$ simply indicates the goodness of separation of the two classes. In the present case, the day1–5 urine profiles were compared with the day30 profiles using OPLSDA modeling. In addition, several clinical variables were correlated with the urine profiles using the OPLS models. Further, in order to check the malaria specificity of the features found by OPLSDA, the day1 urine samples were also compared with the urine samples from the viral fever patients using PCA and OPLSDA. After building up the models from the known samples, a set of 40 samples (20 day1 and 20 day5) were also used in order to test the predictive ability of the models.

2.3.2 Univariate data analysis

After extracting the spectral variables responsible for the variation between infection stage (day1–5) urine samples from that of day30 and viral fever samples, the corresponding metabolites were identified using a variety of 2-dimensional NMR techniques. In order to identify their (a) temporal variation across the days and (b) correlation

with the clinical variables, the NMR peak(s) belonging to specific metabolites were integrated and normalized to the total spectral intensity using AMIX 3.8.4. For metabolites showing peak shift, a region was integrated and normalized. These data were further compared using the Mann–Whitney rank-sum test. In order to correlate the data with the clinical variables Pearson correlation was used. Specific metabolites, which showed significant variation in the multivariate modeling of day1–day30 and day1–viral fever samples, were used to generate the Receiver Operator Characteristic (ROC) curve to check the feasibility of candidature as a biomarker.

3 Results

3.1 Recruitment of the patients, clinical assays and ^1H NMR spectroscopy

In total, 126 patients were recruited in the study. Of these, 106 were *P. vivax* infected patients and 20 were patients with viral fever. The details of the patient recruitment are described in the methods section. For 67 *P. vivax* infected patients, various tests were performed at the hospital as per the normal clinical protocol for malaria in-patients. This included Giemsa stained peripheral blood smear count for parasitemia, hematological tests including total and differential count of blood cells and tests for liver and kidney functions (Table 1). 47 male and 20 female patients underwent several diagnostic tests. The age range was 5 to 55 years (median = 30 years). As expected several patients showed abnormalities in their clinical tests. For example, 59 patients were found to be anemic, 29 showed renal dysfunction (blood urea nitrogen (BUN): serum creatinine ratio >20 or <10) while 26 patients had liver dysfunction (alkaline phosphatase >112 U/l). 20 viral fever patients (10 male and 10 female) with age 19–52 years were also recruited. 1-dimensional ^1H NMR spectra of all the samples were acquired for multivariate data analysis. Comparative ^1H NMR spectral profiles from day1 and day30 from one patient are shown in the Supplementary information S1. Obvious discrimination was noted due to the paracetamol metabolites. Therefore, spectral region corresponding to these metabolites were excluded for all the multivariate data analyses as described in the Methods section. Chloroquine/primaquine and their metabolites are largely sequestered in liver, spleen, kidneys and lungs (Grundmann et al. 1971). Their circulating and excretory levels are too low-submicromolar range (McChesney et al. 1967; Augustjins et al. 1992) and hence do not show up in the NMR spectra.

3.2 Treatment related alteration of the ^1H NMR urinary profile of the *P. vivax* patients

The ^1H NMR spectra of urine samples of 44 patients, whose day1 and day30 samples were available were subjected to multivariate analyses. Initially a PCA model of all the spectra was built. The resulting scores of the first six significant PCs are shown in the supplementary information S2. A continuous change of the scores was observed in the PC1 and 2. Next, pair-wise OPLSDA was performed to compare the day30 profiles with the day1–5 profiles. This operation generated 5 OPLSDA models (Fig. 1). The descriptive statistical parameters ($R^2\text{Y}$ and $Q^2(\text{cum})$) are provided in the Supplementary information S3. From Figs. 1 and S3, it is noted that the day5 and day30 urine profiles were indistinguishable. The day1–3 samples were clearly distinct from day30 samples, and this distinction decreased for day4 and 5 samples. The spectral variables/bins contributing towards the class distinction were analyzed in conjunction with the Variable Importance on Projection (VIP) plots of the relevant model. Further, the 2D NMR spectra (J-resolved, COSY, TOCSY and HMQC) recorded on selected samples were analyzed to confirm the identity of the metabolites. Briefly, the multiplicities of the NMR peaks were ascertained from the 1-dimensional and J-resolved NMR spectra. For multiplets, COSY and TOCSY were analyzed to ascertain the metabolites. In case of singlet peaks, the HMQC spectra were analyzed to confirm the identity of the metabolites. The significant spectral variables/bins along with the loadings and VIP values and assigned metabolites are shown in Table 2. Supplementary information S4 shows the assigned peaks from the TOCSY and HMQC spectra. Day1 samples showed markedly enhanced excretion of taurine along with decreased excretion of hippuric acid, citric acid, glycine, 3-methylhistidine and alanine. Day2 samples, in addition, showed elevated level of tyrosine and creatinine and lowered level of guanidinoacetate. Day3 onwards, the excretory levels of creatinine and taurine started to recover while day4 profile was distinct from that of day30 only in the elevated excretion of glycine and 3-methylhistidine.

3.3 Temporal trajectories of the significantly altered metabolites

In order to investigate the longitudinal trajectory of the significantly altered metabolites, NMR spectral peaks from each metabolite obtained from the OPLSDA modeling were integrated and normalized. These were further compared across the days and also compared to the day30 profiles using Mann–Whitney rank sum test. The trajectories so obtained are presented in Supplementary information S5. Certain features were uniquely evident by

Table 1 Clinical and phenotypic details of the *P. vivax* infected patients

Number of patients	67	
Males	47	
Females	20	
Age	29.97 ± 11.9	Range 5–55 years (median 30 years)
Hb (g/dl)	10.85 ± 1.76	59 anemic patients
AlkP (U/l) [liver dysfunction >112]	107.53 ± 45	26 patients with high activity
BUN/SCr [renal dysfunction >20 or <10]	16.6 ± 9	16 patients with values >20 and 13 patients with values <10
%PRBC [parasite count]	0.07 ± 0.04	Range 0.01–0.2 % (median 0.07 %)
RBC (E + 6, ml ⁻¹ ; ref: 4–6 × 10 ⁶ μl ⁻¹)	4.28 ± 0.56	21 patients with RBC values lower than the reference
WBC (E + 3, ml ⁻¹ ; ref: 4.5–10.5 × 10 ³ μl ⁻¹)	6.55 ± 2.46	12 patients with WBC values lower than the reference
Platelet (E + 3, ml ⁻¹ ; ref: 150–450 × 10 ³ μl ⁻¹)	118.6 ± 51.7	53 patients with platelet values lower than the reference

The clinical parameters were available for a subset of the population under study (67 patients). All the data reported are in form of mean and standard deviation

Hb hemoglobin, *AlkP* alkaline phosphate, *BUN/SCr* blood urea nitrogen to serum creatinine ratio, *%PRBC* percent of parasitized RBCs/parasitemia

univariate analysis. Interestingly, although the overall profiles of day5 and day30 were similar, levels of certain metabolites were continued to be significantly different. For example, hippurate and guanidinoacetate levels were lower in the day5 urine samples compared to day30 ($p < 0.05$). In addition, most of the metabolites showed significant alteration as indicated by the OPLSDA modeling. However, univariate analysis did not show changes in creatinine level. This discrepancy may be attributed to the variable peak position of creatinine across the samples.

3.4 Correlation of parasitemia with the ¹H NMR urine profile

All the day1 urine NMR profiles (except those used for the prediction set) were used for the OPLS modeling. In these models, parasitemia (no. of parasites per μl of blood) was used as the response variable in an attempt to check the correlation of this variable with the urine profiles and to explore the possibility of extracting this information non-invasively from the urine profile. Parasitemia showed certain correlation with urine profiles ($R^2Y = 0.31$, $Q^2(\text{cum}) = 0.21$), this model resulted after exclusion of the pediatric samples from the data set and excluding the noise region from the NMR spectra. 56 samples were used as the test set to predict the parasitemia. A linear relationship (Supplementary information S6A) was obtained between the actual and predicted parasite count of the test set population ($r = 0.57$, $p < 0.0001$). Further, the OPLS model loading was investigated to check for the bins correlating maximally to the parasite count. These bins were further analyzed using 2-dimensional NMR spectra available to identify the metabolites. 4 such metabolites,

namely, betaine (3.90 ppm, s), creatine (3.94 ppm, s), trigonelline (4.42 ppm, m) and threonine (4.28 ppm, m) have significant correlation to the parasite count. All of the correlations were found to be positive (data not shown). However, only betaine showed significant elevation during day1 with respect to the viral fever samples. Betaine also showed an increasing trend compared to the day30 samples, but this was not statistically significant. The training set and the test set samples were then pooled and the levels of betaine so obtained were used for modeling. The correlation improved upon pooling (S5C).

3.5 Prediction of unknown urinary samples based on ¹H NMR profiles

In order to test the predictability of the models a training set was made from the day1 urine and day5 urine NMR profiles. The day5 urine samples were chosen for this purpose because of little distinction of day5 global profiles with that of day30 (Fig. 1e). A total of 88 samples were used for making the training set (45 day1 samples and 43 day5 samples). The model resulted in very well clustered segregation of the day1 and day5 training set. The $Q^2(\text{cum})$ of the resulting model was found to be 0.64 (Fig. 2a). Further, this model was tested on a test set comprising of 104 samples (61 day1 and 43 day5 samples; 18 day5 samples could not be included as the respective individuals were discharged before the 5th day on 5th day) from day1 and day5 urine profiles. The model identified most of the day1 and day5 samples correctly. This is evident from the predicted scores of the same model (Fig. 2b). Manually, the correct predictions could be calculated and we found that 14 day1 samples and 2 day5 samples were wrongly

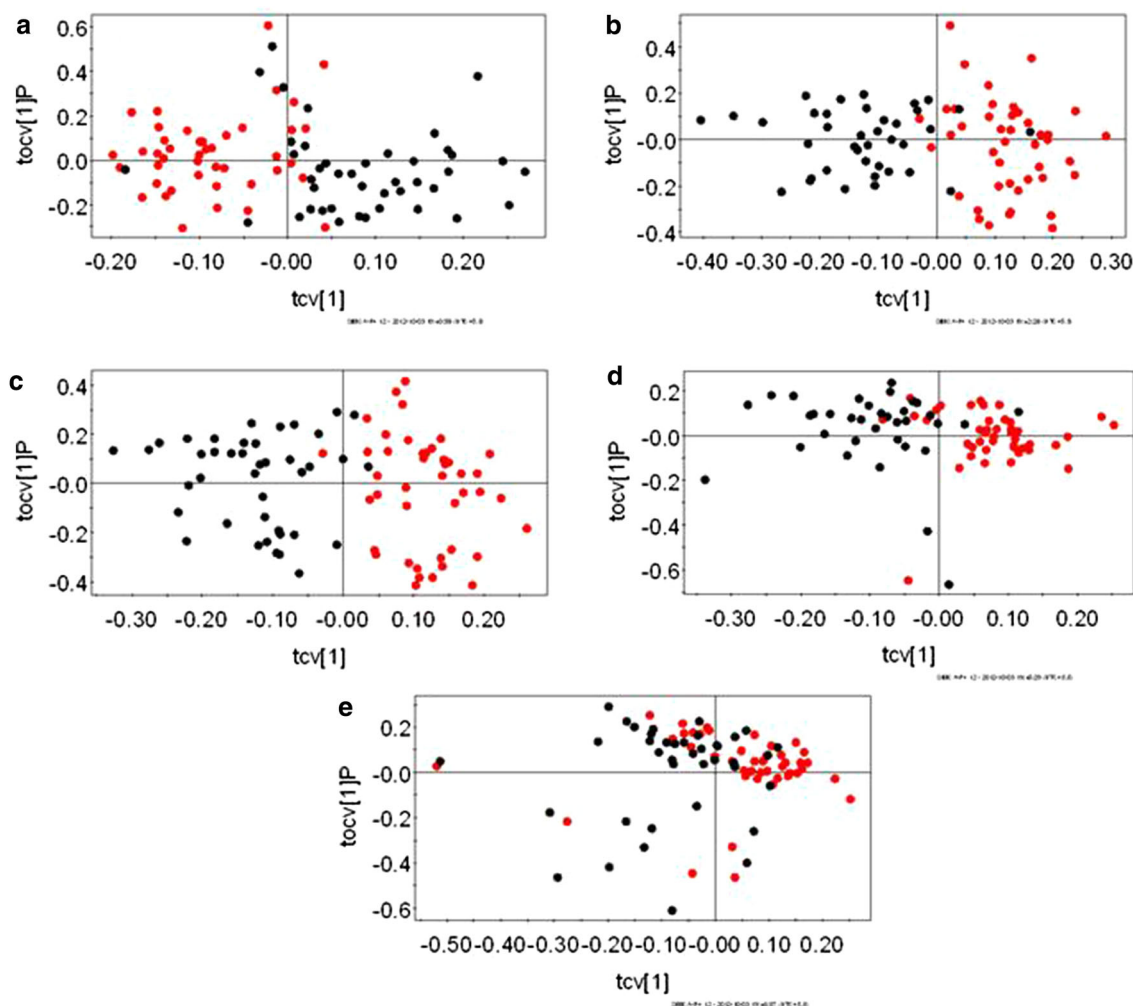


Fig. 1 The cross-validated scores plot from the modeling of urine NMR profiles of *P. vivax* infected patients. Day30 urine NMR profiles of the patients were compared with the day1–5 urine NMR profiles respectively. In each of the figures, the red boxes are the day30 urine NMR profile and the black dots represent day1–5 urine NMR samples

in the figures a–e respectively. In each of the plots, the x-axis represents the predictive component that explains the class specific variation in the spectral variables and the y-axis represents the orthogonal component that explains the variation orthogonal to the class specific variation

predicted. Therefore, out of 104 samples, 16 were misclassified, resulting in a success rate of $\sim 85\%$ of correct identification. The rate of wrong prediction was calculated to be $\sim 15\%$.

3.6 Urinary profiles of *P. vivax* malaria patients are distinct from viral fever patients

Urine samples from patients with viral fever were included in the study as non-malaria febrile controls. ^1H NMR spectra of these samples were compared with the day1 and day30 urine profiles of the *P. vivax* infected patients using OPLSDA analysis. Two OPLSDA models were generated. One compared the urine NMR profiles of viral fever patients with that of day1 *P. vivax* malaria patients and the other with the day30. Both the models resulted in discrete

clustering of day1 and day30 in comparison to the viral fever profiles with the $Q^2(\text{cum})$ of 0.35 (Fig. 3a) and 0.29 (Fig. 3b), respectively. The metabolites that contribute towards the segregation are listed in Table 3. The viral fever patients showed significantly decreased creatinine, taurine, hippurate, citrulline, DMG and taurine and increased suberic acid compared to the day1 profiles of *P. vivax* infected patients. The day30 profiles of the malaria patients, exhibited increased levels of creatinine, hippurate, erythritol, histidine, creatine, DMG, citrulline, glycine and guanidinoacetate (Table 3).

The ^1H NMR resonances of the significantly altered metabolites of day1 and day30 of *P. vivax* infected patients (Table 2) were integrated, normalized with respect to the whole spectrum and compared separately with the corresponding profiles of viral fever patients, using Mann–

Table 2 Spectral variables/bins, loadings, Variable Importance on Projection (VIP) values and the corresponding metabolites assigned using 2-dimensional NMR spectra those were varying significantly across day1–4 and day30

Day compared to day30	Spectral variable elevated in day30	Loadings	VIP	Assigned metabolite	Spectral variable decreased in day30	Loadings	VIP	Assigned metabolite
1	3.94	-0.57	3.45	Hippuric acid	3.42	0.61	4.39	Taurine
	2.54	-0.44	3.31	Citric acid				
	3.54	-0.52	3.30	Glycine				
	7.82	-0.40	2.59	Hippuric acid				
	3.34	-0.38	2.53	3-Methylhistidine				
	7.54	-0.40	2.44	Hippuric acid				
	1.46	-0.37	2.05	Alanine				
2	2.54	0.46	2.92	Citric acid	3.42	-0.45	2.53	Taurine
	3.94	0.55	2.84	Hippuric acid	4.02	-0.34	2.31	Creatinine
	3.54	0.53	2.83	Glycine	6.94	-0.71	2.10	Tyrosine
	7.54	0.46	2.66	Hippuric acid				
	7.82	0.44	2.49	Hippuric acid				
	3.34	0.44	2.26	3-Methylhistidine				
	1.46	0.45	2.21	Alanine				
3	3.78	0.37	2.17	Guanidinoacetate				
	7.54	0.41	2.72	Hippuric acid	6.94	-0.60	1.99	Tyrosine
	3.54	0.50	2.72	Glycine				
	2.54	0.42	2.70	Citric acid				
	7.82	0.38	2.48	Hippuric acid				
	3.78	0.38	2.46	Guanidinoacetate				
	3.34	0.42	2.3	3-Methylhistidine				
4	1.46	0.40	2.03	Alanine				
	3.54	0.37	2.87	Glycine	-	-	-	-
	3.34	0.44	2.49	3-Methylhistidine				

These bins were selected from the OPLSDA models constructed from the urine profiles of day1–4 and day30. Day5 was not compared to day30 because of negligible distinction shown in the relevant OPLSDA model

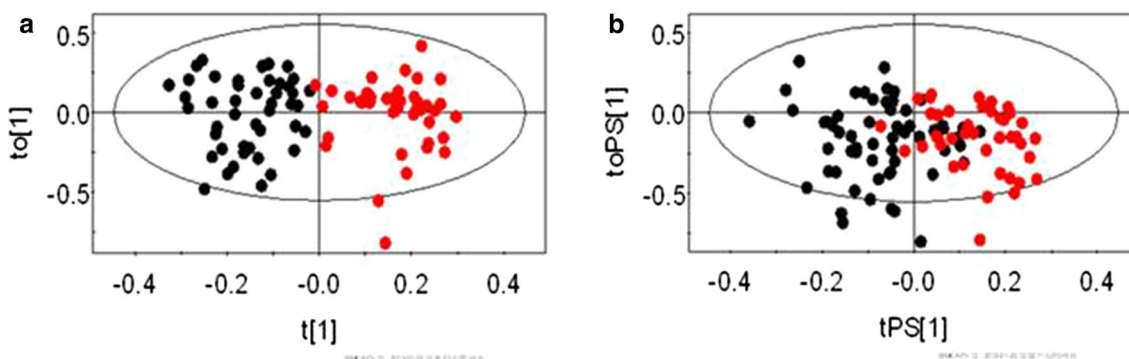


Fig. 2 Representative scores plots from the OPLSDA models of day1 and day5 urine NMR profiles. **a** Scores plot from day1 (black) and day5 (red) test set and **b** scores plot from day1 (black) and day5 (red) prediction set. Model $Q^2(\text{cum}) = 0.72$. In **a**, the x-axis represents the predictive component that explains the class specific variation in the

spectral variables and the y-axis represents the orthogonal component that explains the variation orthogonal to the class specific variation. In **b** the axes represent the same scores predicted on the basis of the test set (Color figure online)

Whitney test (Supplementary information S7). In addition to taurine, citric acid also showed significant alteration between day1 and day30 as well as day1 and viral fever

patients. Therefore, these two metabolites could be considered as putative metabolic markers specific to *P. vivax* infection. In order to check the diagnostic capability,

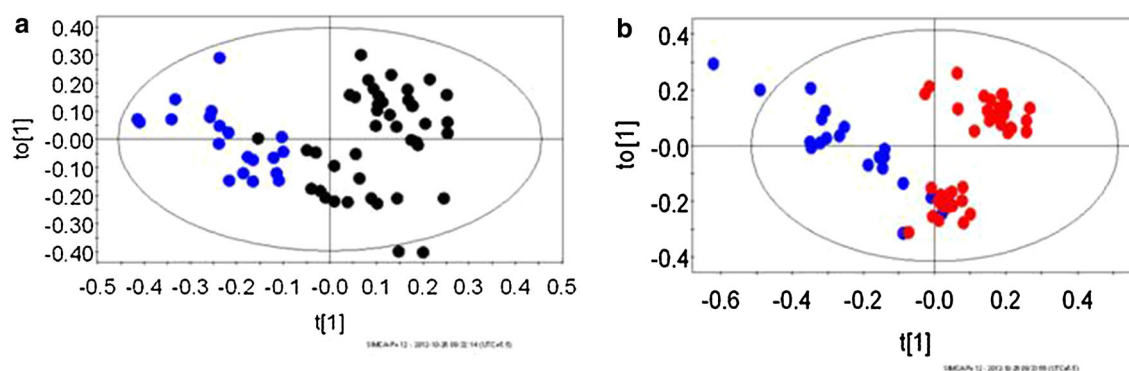


Fig. 3 Representative scores plots from the OPLSDA models of viral fever patients (*blue*) malaria day1 (*black*) and day30 (*red*) urine NMR profiles. **a** Scores plot from malaria day1 (*black*) and viral fever (*blue*) patients' urine sample NMR profiles, Model $Q^2(\text{cum}) = 0.35$ and **b** scores plot from malaria day30 (*red*) and viral fever patients' (*blue*)

urine sample NMR profiles. In each of the plots, the x-axis represents the predictive component that explains the class specific variation in the spectral variables and the y-axis represents the orthogonal component that explains the variation orthogonal to the class specific variation (Color figure online)

Table 3 Spectral variables/bins, loadings, Variable Importance on Projection (VIP) values and the corresponding metabolites assigned using 2-dimensional NMR spectra those were varying significantly across malaria day1 and viral fever patients/malaria day30 and viral fever patients

Malaria day compared to day viral fever profiles	Spectral variable elevated in viral fever	Loadings	VIP	Assigned metabolite	Spectral variable decreased in viral fever	Loadings	VIP	Assigned metabolite				
1	1.26	-0.51	2.02	Suberic/sebacic acid	4.06	0.56	6.46	Creatinine				
					3.90	0.45	2.70	Paracetamol				
					3.42	0.57	2.67	Taurine				
					3.94	0.52	2.38	Hippuric acid				
					3.74	0.43	2.19	Erythritol				
					4.26	0.62	2.15	Unconfirmed				
					3.70	0.41	2.13	Citrulline/DMG				
				30	4.06	0.65	6.57	Creatinine	4.06	0.65	6.57	Creatinine
									3.94	0.75	3.71	Hippurate
									3.74	0.64	2.83	Erythritol
									3.98	0.69	2.73	Histidine
									3.90	0.49	2.48	Creatine
									3.70	0.49	2.42	DMG/citrulline
	3.54	0.58	2.23	Glycine								
	1.78	0.58	2.20	Unconfirmed								
	3.78	0.43	2.02	Guanidinoacetate								

These bins were selected from the OPLSDA models constructed from the urine profiles of day1-viral fever and day30-viral fever OPLSDA models

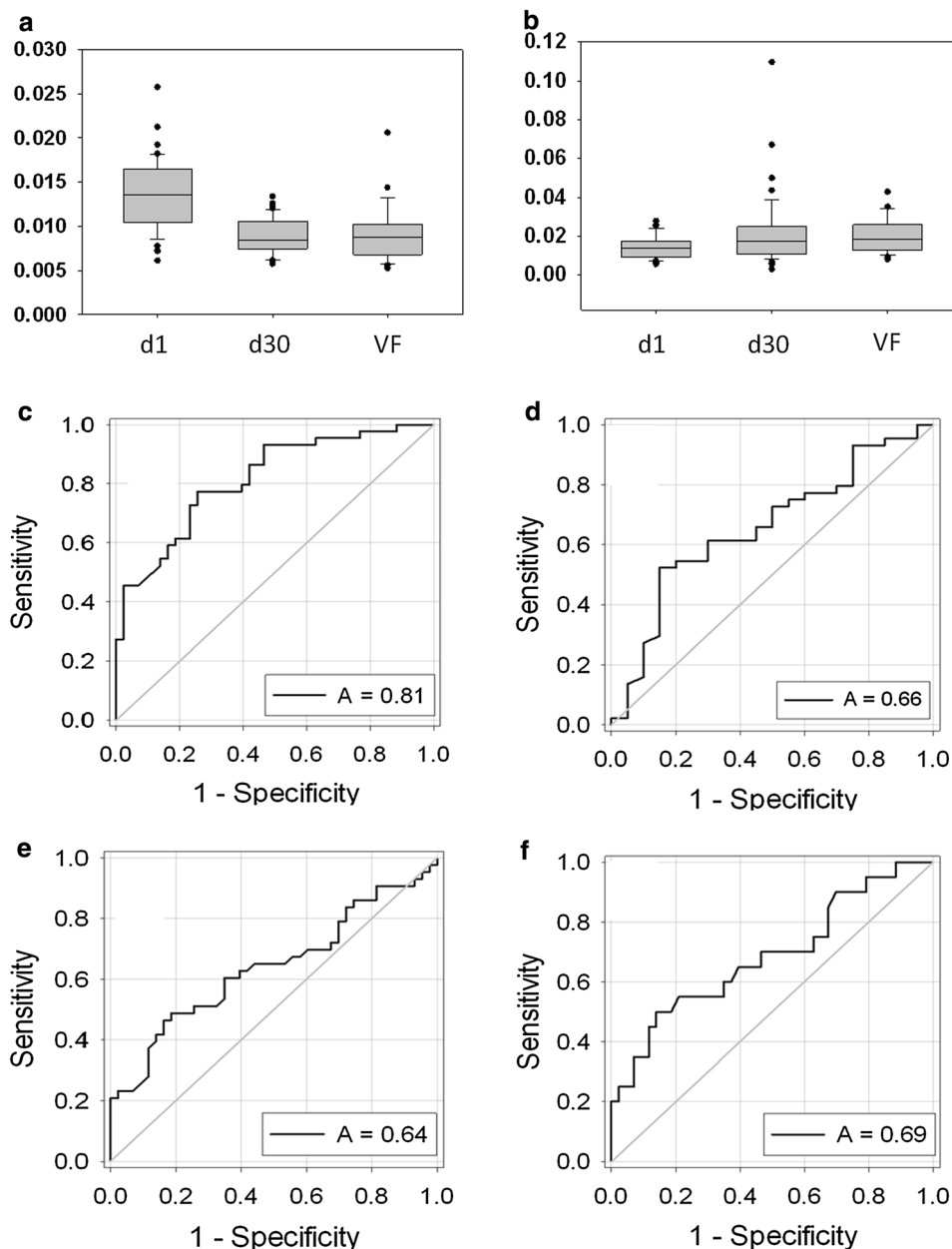
receiver-operator-characteristic curves were generated from these two metabolites' levels of the three sets of profiles (Fig. 4).

3.7 Diagnostic efficiency of excretory level of taurine and citrate

Taurine and citrate were identified as the putative biomarkers for the malarial infection as shown above.

Therefore, the levels of these two metabolites were also used for the prediction. For this purpose, the average levels of taurine and citrate, in terms of normalized peak intensity of the NMR spectra, were obtained from the training set. Then the taurine and citrate levels from the test set samples were also obtained. The levels from the test set were blindly used to assign the identity (day1/day5) of the test set samples. To this end, two different approaches were tried—either of the two metabolites was individually used

Fig. 4 **a, b** the relative levels of excretion of taurine and citrate, respectively, in the urine of day1 (d1) and day30 (d30) of malaria patients and viral fever patients (VF). The normalized spectral intensity (y-axis) of the relevant peaks of the two metabolites is compared here. In both cases $p < 0.05$ when d1–VF are compared. **c–f** The receiver operator characteristics (ROC) curves generated from the intensity values and the class of population. **c, d** The ROC curve generated from urinary taurine levels of d1–d30 and d1–VF population, respectively; **e, f** the ROC curve generated from urinary citrate levels of d1–d30 and d1–VF population, respectively. A = area under the ROC curve in each case



for prediction or they were used together. While using both metabolites for prediction the following criterions were used: (a) If prediction by at least one metabolite is correct, it is taken as correct prediction, (b) If one metabolite predicts wrong and other one could not decide or both predict wrong—it is taken as wrong prediction, and (c) If both metabolites could not decide—it is taken as no decision. Following this procedure, 44.2 % correct and 28.8 % wrong prediction was made using taurine. ~27 % samples could not be judged. Citrate level was able to predict ~52.9 % samples correctly and wrong prediction was made for ~16.4 % of the samples. For ~6.7 % of the samples, the identity could not be judged. While using both

the metabolites together, significant improvement was achieved. ~81 % correct prediction was achieved in this case along with ~16 % of wrong prediction. ~3 % of the sample could not be assigned any identity.

4 Discussion

4.1 Alteration in metabolism upon *P. vivax* infection

Age old diseases such as malaria pose newer global threats that could only be resolved through understanding of the underlying mechanism that modulates the changes in the

host during the infection. Our laboratory has shown that profiling of metabolites in the biofluids and tissues of the infected individuals may be the key in understanding the underlying molecular mechanism of disease progression as well as discovery of biomarkers (Basant et al. 2010; Sengupta et al. 2011a, b; Ghosh et al. 2012). In this report, we longitudinally monitored malaria patients infected by *P. vivax* for a maximum of five days as in-patients in a normal hospital setting. As a follow-up, samples were collected from these patients 25 day after they were discharged from the hospital on completion of treatment (day30). This latter sampling thus provided us with a baseline data for each person. We did NMR spectroscopic profiling of the urine samples obtained from the patients at each day followed by multivariate statistical analysis of those profiles to arrive at our conclusions. In order to confirm the malaria specificity of our results, we compared the urine NMR profiles of malaria patients with urine NMR profiles of viral fever patients.

We have earlier reported a distinct clustering of urine NMR profiles of malaria infected patients from that of uninfected healthy population (Sengupta et al. 2011a, b). Our current analysis suggests significant clustering of the day1 urine NMR profile of individual malaria patient from that of the day30 urine NMR profile (Fig. 1a). In addition, we observed that the distinction improves as the treatment ensues up to the 3rd day and then declines (S3). Day5 urine NMR profiles seem to have the least distinction from that of day30 profiles (S3). In addition, the longitudinal trajectories of hippuric acid, glycine, 3-methylhistidine, alanine and tyrosine also showed similar behavior in terms of inflexion points at day2/3 (Supplementary information S5). Since, urine can be considered as a biofluid that reflects the physiological homeostasis, it is tempting to speculate that physiology started to return to its original state around day3. This observation may be explored further. Specific patterns of the metabolite recovery were observed. For example, day1 urine samples showed significantly decreased excretion of hippuric acid, citric acid, glycine, 3-methylhistidine and alanine along with enhanced excretion of taurine (Table 2). Day2 onwards, excretory level of guanidinoacetate started to fall and that of tyrosine started to rise. Day4 only showed decreased excretion of glycine and 3-methylhistidine. However, investigation into the malaria specificity of these metabolic features revealed that some of them share commonality with the metabolic correlates of viral fever patients. In fact, only hippurate, taurine and citrate showed significant difference between the excretory levels of day1 profiles of *P. vivax* infected patients and viral fever patients. Moreover, the hippurate and taurine levels are also different across the day30 profiles as well as viral fever patients. This suggests that the other metabolic correlates could be the signature of the fever

related changes. Together, they probably indicated changes in gut microfloral co-metabolism, fatty acid β -oxidation and the detoxification process in general. We recognize, however, that some of the responses could be manifestations of drug response. In fact, the responses we observe is a collective of drug, parasite and other effects. It would be difficult to single out the drug related response without an untreated control group which, at this point, is beyond ethical consideration. Together, these observations would lead to a better understanding of the multimodal host response to the *P. vivax* infection.

4.2 Towards non-invasive diagnosis of *P. vivax* infection

It is evident that a simplified non-invasive diagnostic technique for malaria is the key to prevent the misdiagnosis of the infection. This paper reports such a technique by using the ^1H NMR profiling of urine samples obtained from the *P. vivax* infected and recovered patients. Here, we show the possibility to identify the infected and recovered patients by the multivariate profiling of the NMR spectra obtained from the urine of the patients (Fig. 2b). This, to the best of our knowledge, is the first report attempting the non-invasive diagnosis of malarial infection. In addition to this, we could demonstrate a fairly linear trend of the parasitemia of the patients to the urine ^1H NMR scores obtained from the OPLS modeling (S6). This suggests the possibility of a non invasive prediction of parasitemia. Since the parasitemia of the population under study is fairly low (0.07 ± 0.04 %, Table 1), this approach promises the added advantage of detecting/predicting parasite load even at low levels of parasitemia.

In order to address the issue of misdiagnosis of other non-malarial febrile illnesses (NMFI) as malarial infection, it is important to employ appropriate NMFI control population. This was achieved by recruiting viral fever patients and by comparing their urinary metabolic profiles with that of day1 and day30 of the *P. vivax* infected patients. To this end, we could show that the urine profiles of the viral fever patients are clearly distinct from the day1 and day30 urine NMR profiles of malaria patients (Fig. 3). Alteration of certain metabolites in this case would be results of viral fever infection or inherent population difference. These are metabolites which were distinct across the viral fever—day1 samples and viral fever—day30 sets, such as creatinine, erythritol, citrulline and DMG. However, metabolites that are distinct across viral fever day1 and malaria day1–30 sets could be taken as features specific to malarial infection. Taurine and citrate were the metabolites that showed this feature. Of these two, analysis of the ROC curve generated showed that taurine may be a better candidate biomarker for the malarial infection (ROC

curve area 0.81, Fig. 4a). Further, when the levels of taurine and citrate were used together to predict the unknown samples and ~81 % of correct prediction was observed. Compared to this, studies with malaria patients in UK showed 29.4–77.1 % of correct diagnosis depending upon the type of *Plasmodium* spp infection and/or single or mixed infection (Milne et al. 1994). Speculations are that this percentage may decrease with larger sample size (Hanscheld 2003). The problem with blood smear microscopy, specifically with low level transmission, remains to be linked to time taken and personnel expertise. On the other hand, urinary biomarkers may be detected using simple assay protocols. Rapid Diagnostic Tests based on the parasite antigen generally show better sensitivity than blood smear tests (Stauffer et al. 2009; Batwala et al. 2010). The higher costs and lack of quantization remain two major concerns. However, it may be prudent to compare the diagnostic efficacy of the proposed biomarker with already existing techniques with respect to a third technique such as polymerase chain reaction.

Another option is to use the whole NMR profile as the diagnostic parameter. To this end, our data shows very good possibility. We could show ~85 % of correct prediction of infected samples from a cohort of urine profiles belonging to same individuals during infection and after treatment. While comparing this method with the prediction by using taurine alone, we demonstrated that identification by the global profile acts better as the latter resulted in 55 % prediction.

Therefore, we report the first approach to characterize the recovery profile and its utility in the patient management and non-invasive diagnosis of *P. vivax* infection. In fact, this approach can further be extended to other *Plasmodium* infections to validate the specificity of the metabolites we observe to *P. vivax*. Nevertheless, the results presented are very encouraging and raise the possibility of the non-invasive diagnosis of vivax malaria and its severity. This approach may prove useful for rapid diagnosis of infection as well as elucidation of the parasitemia in the infected population.

5 Conclusions

This paper reports the first ever non-invasive study of the recovery profile of *P. vivax* infected patients using urine samples. We delineated the multimodal metabolic response during the malarial infection and recovery. In addition, we could also demonstrate that specific pattern exists in the recovery of the patients that might be useful in optimum management of post-infection drug administration. Moreover, our study also suggests the possibility of non-invasive diagnosis of malarial infection using the metabolic profiling of the urine

from the patients. We could also show that malarial parasitemia and associated disorders such as malarial anemia may also be diagnosed using the urine metabolic profile. Specifically, urinary excretion of metabolites such as taurine and citrate may be of diagnostic interest.

Acknowledgments The study was supported by an in-house research grant of TIFR. We thank the staff of the National Facility for High Field NMR at TIFR for their help in the NMR experiments. AS is grateful to Council of Scientific & Industrial Research, Govt. of India for the SPM fellowship.

Conflict of interest The authors have no conflict of interest to declare.

Compliance with Ethical Requirements The study was approved by the Local Institutional Ethics Committee. Written informed consent was obtained from each patient prior to his/her enrolment in the study.

References

- Augustjns, P., Geusens, P., & Verbeke, N. (1992). Chloroquine levels in blood during chronic treatment of patients with rheumatoid arthritis. *European Journal of Clinical Pharmacology*, *42*, 429–433.
- Basant, A., Rege, M., Sharma, S., & Sonawat, H. M. (2010). Alterations in urine, serum and brain metabolomic profiles exhibit sexual dimorphism during malaria disease progression. *Malaria Journal*, *9*, 110.
- Batwala, V., Magnussen, P., & Nuwaha, F. (2010). Are rapid diagnostic tests more accurate in diagnosis of *Plasmodium falciparum* malaria compared to microscopy at rural health centres? *Malaria Journal*, *9*, 349.
- Ghosh, S., Sengupta, A., Sharma, S., & Sonawat, H. M. (2012). Metabolic fingerprints of serum, brain, and liver are distinct for mice with cerebral and noncerebral malaria: A ¹H NMR spectroscopy-based metabolomic study. *Journal of Proteome Research*, *11*, 4992–5004.
- Grundmann, M., Mikulikova, I., & Vrublovsky, P. (1971). Tissue distribution of subcutaneously administered chloroquine in the rat. *Arzneimittel Forschung*, *21*, 573–574.
- Hanscheld, T. (2003). Current strategies to avoid misdiagnosis of malaria. *Clinical Microbiology & Infection*, *9*, 497–504.
- Joshi, H., Prajapati, S. K., Verma, A., Kanga, S., & Carlton, J. M. (2008). *Plasmodium vivax* in India. *Trends in Parasitology*, *24*, 228–235.
- Kochar, D. K., Das, A., Kochar, S. K., et al. (2009). Severe *Plasmodium vivax* malaria: A report on serial cases from Bikaner in northwestern India. *American Journal of Tropical Medicine and Hygiene*, *80*, 194–198.
- McChesney, E. W., Fasco, M. G., & Banks, W. F. (1967). The metabolism of chloroquine in man during and after repeated oral dosage. *Journal of Pharmacology and Experimental Therapeutics*, *158*, 323–331.
- Mendis, K., Sina, B. J., & Marchesini, P. R. C. (2001). The neglected burden of *Plasmodium vivax* malaria. *American Journal of Tropical Medicine and Hygiene*, *64*(1 suppl), 97–106.
- Milne, L. M., Chiodini, P. L., & Warhurst, D. C. (1994). Accuracy of routine diagnostic tests of malaria in the United Kingdom. *Journal of Clinical Pathology*, *47*, 740–742.

- Nankabirwa, J., Zurovac, D., Njogu, J. N., et al. (2009). Malaria misdiagnosis in Uganda-implications for policy change. *Malaria Journal*, *8*, 66.
- Nicholson, J. K., & Lindon, J. C. (2008). Systems biology: Metabonomics. *Nature*, *455*, 1054–1056.
- Sengupta, A., Basant, A., Ghosh, S., Sharma, S., & Sonawat, H. M. (2011a). Liver metabolic alterations and changes in host intercompartmental metabolic correlation during progression of malaria. *Journal of Parasitology Research*. doi:[10.1155/2011/901854](https://doi.org/10.1155/2011/901854).
- Sengupta, A., Ghosh, S., Basant, A., et al. (2011b). Global host metabolic response to *Plasmodium vivax* infection: A ^1H NMR based urinary metabonomic study. *Malaria Journal*, *10*, 384.
- Stauffer, W. M., Cartwright, C. P., Olson, D. A., et al. (2009). Diagnostic performance of rapid diagnostic tests versus blood smears for malaria in US Clinical Practice. *Clinical Infectious Diseases*, *49*, 908–913.
- Teka, H., Petros, B., Yamuah, L., et al. (2008). Chloroquine-resistant *Plasmodium vivax* malaria in Debre Zeit. *Ethiopia Malaria Journal*, *7*, 220.