Full Paper

Journal of the Brazilian Chemical Society

https://dx.doi.org/10.21577/0103-5053.20260076

J. Braz. Chem. Soc. **2026**, *37-Spe1*, e-20260076, 1-8

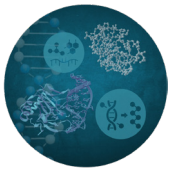
Metabolomics by UPLC-QTOF-MSE in Free-Ranging Neotropical Primates Plasma for Identification of Potential Biomarkers of Toxoplasmosis

Mariana M. Fachi, [a] ***Beatriz Böger,*** [*,a] ***Alexandre de F. Cobre,*** [a] ***Walfrido K. Svoboda,*** [b] ***Paulo Justiniano Ribeiro Junior*** [c] ***and Roberto Pontarolo*** [a]

[a]*Programa de Pós-Graduação em Ciências Farmacêuticas, Universidade Federal do Paraná, 80210-170 Curitiba-PR, Brazil*

[b]*Instituto Latino-Americano de Ciências da Vida e da Natureza, Universidade Federal da Integração Latino-Americana (UNILA), 85870-650 Foz do Iguaçu-PR, Brazil*

[c]*Departamento de Estatística, Universidade Federal do Paraná, 81530-015 Curitiba-PR, Brazil*



Toxoplasmosis, caused by the obligate intracellular protozoan *Toxoplasma gondii*, represents a relevant zoonotic disease affecting both human and animal health. Free-ranging neotropical primates (FRNP) are particularly susceptible to severe clinical outcomes; however, systemic metabolic alterations associated with infection in these species remain poorly characterized. This study aimed to investigate serum metabolic perturbations in FRNP naturally infected with *T. gondii* using an untargeted liquid chromatography-mass spectrometry (LC-MS) approach integrated with multivariate and machine learning (ML) analyses. Serum samples from infected and non-infected FRNP were analyzed by liquid chromatography-high-resolution mass spectrometry (LC-HRMS), yielding 1025 metabolic features following rigorous data preprocessing and quality control procedures. Both unsupervised and supervised ML models were employed to explore group discrimination. Several metabolic features (*m/z* 809.2713, 660.5429, 718.3614, 537.5458, 707.2745, 318.2951) were significantly altered in infected animals. Receiver Operating Characteristic (ROC) analysis indicated moderate discriminatory performance (area under the curve (AUC) = 0.78; sensitivity = 0.88). Collectively, these findings indicate that *T. gondii* infection in FRNP is associated with measurable systemic metabolic perturbations. The identified features should be interpreted as preliminary metabolic signatures, warranting further targeted validation and larger-scale studies to clarify their biological and potential translational relevance.

Keywords: toxoplasmosis, non-human primates, machine learning, liquid chromatography coupled with mass spectrometry

Introduction

Toxoplasma gondii is an obligatory intracellular coccidian, distributed worldwide and capable of infecting cells from different hosts.[1] This parasite presents three infectious stages: tachyzoite (rapid division), bradyzoite (slow division within tissue cysts and an environmental stage), and sporozoite (protected within an oocyst).[1-4]

Although infection in humans is commonly asymptomatic, there is a high susceptibility to the development of severe conditions in free-ranging neotropical primates (FRNP), with high mortality rates.[5] The processes promoting infection in free-ranging neotropical primates (FRNP) are highly complex and involve interactions among physical, biological, and ecological factors, including climatic conditions, host susceptibility, body size and weight, diet, and feeding behavior.[6]

Previous studies have demonstrated that *T. gondii* infection can induce measurable metabolic alterations in different host species, including rodents and humans.[7] Reported alterations include changes in lipid metabolism, amino acid metabolism, and energy-related pathways, suggesting that infection may lead to systemic biochemical perturbations. However, information regarding metabolic alterations associated with toxoplasmosis in free-ranging neotropical primates remains extremely limited. This

*e-mail: beatrizboger@gmail.com
Editor handled this article: Paulo Wender P. Gomes (Guest)
This publication is part of the special issue "Omics Sciences"

knowledge gap highlights the need for studies investigating host metabolic responses in these highly susceptible species.[7]

The diagnosis of toxoplasmosis can be performed by detecting or isolating the parasite, or by conducting serological tests. Direct detection methods include polymerase chain reaction (PCR) and parasite isolation, which identify parasite deoxyribonucleic acid (DNA) or viable organisms but may present limited sensitivity depending on parasite burden and sample type. Serological tests, such as enzyme-linked immunosorbent assay (ELISA) and indirect immunofluorescence assays, detect host antibodies and are widely used for epidemiological studies. However, reported sensitivities and specificities vary substantially depending on the test and host species, and serological methods may not distinguish between active and past infection.[2]

Considering the difficulty in minimizing disease transmission in animals, a simple and routinely available diagnostic procedure is necessary so that clinically ill animals infected with *T. gondii* can be promptly detected and treated appropriately. Thus, metabolomics stands out, an omics technique focusing on detecting metabolites in biological samples through analytical techniques such as liquid chromatography coupled with mass spectrometry, allowing the study of biological information at the biochemical level. Metabolomics has been increasingly applied to investigate infectious diseases, as metabolic profiles may reflect host physiological responses to infection. In recent years, machine learning (ML) approaches have been widely employed to analyze metabolomics data due to their capacity to manage high-dimensional datasets and uncover complex patterns that may not be detected using conventional statistical methods. Previous studies[8-10] have successfully applied ML-based approaches to metabolomics datasets for disease classification and biomarker discovery in infectious and metabolic diseases, highlighting the potential of these techniques for exploratory diagnostic research.

Since metabolomics generates large volumes of complex data, ML is applied for data analysis, pattern recognition, model building, classification, regression, or clustering of highly complex metabolomic data.[11]

Considering that a better understanding of the molecular changes associated with the disease is essential for identifying more advanced diagnostic tests, the aim of this study was to investigate the dynamic changes in the metabolic profile of sera from FRNP during *T. gondii* infection.

Experimental

Sample collection

Blood plasma samples from 147 FRNP were obtained in the Upper Paraná River region, following approval for capture and procedures by IBAMA (license No. 104/04) and by the Animal Experimentation Ethics Committee of the State University of Londrina (registration No 34/05). Further details regarding the collection of samples from FRNP with (test group) and without toxoplasmosis (control group) were previously reported in the study conducted by Svoboda.[12] Samples were classified as infected or non-infected based on serological testing for anti-*Toxoplasma gondii* antibodies. Individuals with detectable antibody levels were classified as infected (test group), whereas individuals with no detectable antibodies were classified as non-infected (control group).

After collection, the samples were transported to the Center for Studies in Biopharmacy at the Federal University of Paraná (CEB-UFPR) and were stored in a freezer at –40 °C until analysis.

Sample preparation

For the analyses, the samples were thawed at room temperature, and then 50 µL aliquots were transferred to Eppendorf tubes and precipitated with 200 µL of acetonitrile (ACN) containing 0.1% formic acid. Protein precipitation using acetonitrile acidified with formic acid is widely used in plasma metabolomics because it promotes efficient protein removal while maintaining a broad range of small metabolites in solution. This extraction approach has been previously described as suitable for untargeted metabolomics workflows.[13]

These samples were agitated for 3 min using a Vortex-Genie 2T model SI T236 and centrifuged (Eppendorf model 5810R) at 14,000 rpm, 4 °C for 10 min. Then, the supernatant was separated and filtered into a vial for subsequent analysis. All sample preparation procedures were performed using sterile disposable tubes and filtered pipette tips to minimize contamination risks. Blank samples were periodically prepared and analyzed in order to monitor potential contamination during sample preparation.

Ultra-performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry with MSE acquisition (UPLC-QTOF-MSE) design and databases

The samples were analyzed in triplicate using an Acquity H-Class ultra-performance liquid chromatograph (UPLC) system from Waters Corporation (Milford, USA),

coupled with a quadrupole-time-of-flight (QTOF) mass spectrometer, Xevo G2-S, Waters (Milford, USA), with an electrospray ionization (ESI) source.

Chromatographic separations were performed using an Acquity UPLC BEH Shield RP18 column (100 × 2.1 mm, 1.7 µm) at 40 °C as the stationary phase. The mobile phase consisted of ultrapure water (Milli-Q system, Millipore) (A) and acetonitrile (B), both containing 0.1% formic acid. Samples were analyzed in gradient elution mode as follows: t_{0-1}: 95% phase A, t_{1-9}: 95-5% phase A, t_{9-11}: 5% A, t_{11-12}: 5-95% A, t_{12-15}: 95% A. The injection volume was 5 µL, and the mobile phase flow rate was 300 µL min^{-1}.

For the spectrometric analysis, high-purity nitrogen was used as the cone and desolvation gas (Peak Scientific Instruments nitrogen generator, Chicago, USA), and argon was used as the collision gas (purity > 99.998% from White Martins Praxair Inc, Curitiba, Brazil). Samples were analyzed in positive ionization mode. Although the acquisition of data in both positive and negative ionization modes is recommended for comprehensive untargeted metabolomics studies, the positive ionization mode was selected in the present study because it provided stable signal intensity and broad metabolite coverage under the experimental conditions employed.

The following conditions were set for the ionization source parameters: +400 V capillary voltage, 40 V cone voltage, 80 V source offset, 150 °C source temperature, 400 °C desolvation temperature, 50 L h^{-1} cone gas flow, 600 L h^{-1} desolvation gas flow, 4 V for low-energy collision energy, and 20-40 V for high-energy collision energy.

During the analytical run, acquired masses were corrected using Lockspray, composed of leucine enkephalin at a concentration of 1000 ng mL^{-1} and with a flow rate of 20 µL min^{-1}, allowing for mass errors of less than 5 ppm. To minimize carryover between injections, the autosampler needle and injection system were washed with strong solvent after each injection according to the instrument cleaning protocol. Blank injections were periodically performed during the analytical sequence to monitor possible carryover effects.

Additionally, calibration checks were performed daily using 0.05 mmol L^{-1} sodium formate as the calibrant, and all data were acquired in centroid mode in MS^E and processed using MassLynx V 4.1 software (Waters Corporation, Milford, USA).

Data processing

In this step, sample preparation was performed to make the raw data suitable for the construction and training of ML-based models. For this purpose, all samples underwent the following preprocessing methods: imputation, transformations, filtering, normalization, scaling, and centering.

Additionally, for data processing, the following parameters were established: (*i*) retention time: from 0.5 to 12 min, removing the time related to column re-equilibration, (*ii*) mass range: from 50 to 900 Da, (*iii*) XIC window (Da): mass accuracy (Da) of acquired data tolerated was 0.02 Da, (*iv*) peak-to-peak baseline noise: baseline noise between peaks was automatically estimated by the software, (*v*) marker intensity threshold (counts): the minimum intensity level for a spectral peak to be considered as a marker was 100, (*vi*) mass window: mass tolerance of 0.02 Da, (*vii*) retention time window: a window of 0.2 min was used for all analyses.

The selected mass range (50-900 Da) was chosen to include the majority of low-molecular-weight metabolites typically detected in plasma metabolomics studies. Although larger molecules such as peptides may occur at higher masses, the selected range provided stable acquisition conditions and adequate signal-to-noise ratios. Future studies may consider extending the mass range to include higher-molecular-weight compounds.

All chromatograms obtained by MassLynx™ software (Waters Corporation) were processed using MarkerLynx™ XS Application Manager software (Waters Corporation). For data analysis in Python, for each sample, the data was organized by correlating the acquired *m/z* values with the maximum detected signal intensity (in percentage).

The scripts used for data preprocessing, machine-learning model development, and statistical analyses were implemented in Python and are publicly available in a GitHub repository website (see Data Availability Statement section). The repository contains the scripts required to reproduce the analyses performed in this study.

Development of a ML model

Initially, an unsupervised ML-based model (Principal Component Analysis - PCA) was conducted to identify the data structure and detect potential outlier samples. The dataset consisted of 147 samples, including infected and non-infected animals. The data were randomly divided into training (70%) and test (30%) subsets, ensuring representation of both groups in each subset. Subsequently, for the prediction of toxoplasmosis diagnosis, various supervised ML-based models were employed: Random Forest Classifier, Extra Trees Classifier, Ada Boost Classifier, Naive Bayes, Dummy Classifier, Gradient Boosting Classifier, Light Gradient Boosting Machine, K Neighbors Classifier (KNN), Logistic Regression,

Support Vector Machine (SVM) with Linear Kernel, Linear Discriminant Analysis (LDA), Extreme Gradient Boosting, Quadratic Discriminant Analysis, Ridge Classifier, and Decision Tree Classifier. For this, 70% of the data was used for the training set (calibration), and the remaining 30% for the test set. Sample selection for the training and test sets was performed randomly.

Furthermore, cross-validation was conducted to select the number of latent variables (LVs) for the ML-based models. LVs correspond to new variables obtained from combinations of the original variables and represent the main sources of variation in the dataset. The use of LVs allows dimensionality reduction and improves model stability. LVs with lower values of cross-validation error, root mean square error of cross-validation (RMSECV), and root mean square error of calibration (RMSEC) were selected. The root mean square error of prediction (RMSEP) was the metric used to assess the predictive ability of the ML-based models; models with lower RMSEP had better performance.

The performance of the model was evaluated considering the metrics of accuracy, sensitivity, and specificity. These metrics were calculated based on the following figures of merit: false positive (FP), false negative (FN), true positive (TP), and true negative (TN). The accuracy of the model was also assessed considering the area under the receiver operating characteristic (ROC) curve (AUC), established based on the training and test samples.

After selecting the best model, a feature importance plot was created to identify the 10 most important molecular mass signals (metabolites) that contributed to the discrimination of toxoplasmosis in FRNPs. It is worth noting that variables with a score higher than 1 were considered statistically significant for group classification.

Metabolite annotation was performed by comparing accurate mass values and isotopic patterns with entries in public metabolomics databases, including LipidMaps and the Human Metabolome Database (HMDB). Molecular formula assignments were based on mass accuracy below 5 ppm and isotopic distribution patterns. Therefore, metabolite identification should be considered putative. Accordingly, metabolite annotations were classified following the Metabolomics Standards Initiative (MSI) reporting guidelines,[14] and should be interpreted as putative (Level 2) identifications.

Results

Figure 1 demonstrates the total ion chromatogram of a sample of FRNP. Baseline correction, peak deconvolution, alignment, normalization, and median centering were applied to process the data, and approximately 1025 masses were found in each sample profile analyzed in positive ion mode. Moreover, the stability and reproducibility of the chromatographic and spectrometric system were evaluated through the analysis of pooled quality control (QC) samples throughout the analytical sequence. QC samples were prepared by pooling equal aliquots from all plasma samples to obtain a representative matrix and were processed following the same extraction protocol as the study samples. The use of pooled QC samples and system suitability assessment followed recommended guidelines for untargeted metabolomics studies.[15]

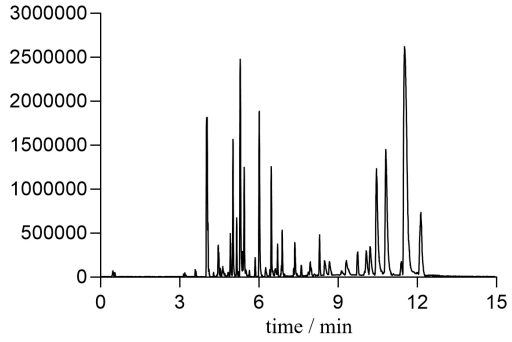


Figure 1. Chromatogram obtained from a sample of non-human primate plasma with toxoplasmosis.

Unsupervised model

Through the PCA model, it was observed that the model was able to discriminate all sample classes (Figure 2). No outlier samples were identified. Furthermore, it was observed that among the samples from the diseased group, there is separation into clusters (data not explored due to study design limitations). The preprocessing methods used were a combination of imputation using median values and autoscaling. Although a tendency toward sub-structuring within the infected group was observed in the PCA analysis, the absence of clinical staging or temporal infection data precluded further interpretation of these patterns.

Supervised model

Table 1 presents the performance of the ML-based models constructed. The Ada Boost Classifier model showed the most promising results for predicting the diagnosis of toxoplasmosis, obtaining the best values for accuracy (0.78), AUC (0.73), sensitivity (0.88), and precision (0.83). Figure 3 shows the ROC curve obtained for this model.

Furthermore, the feature importance analysis derived from the selected model identified the ten most relevant features contributing to classification performance. The

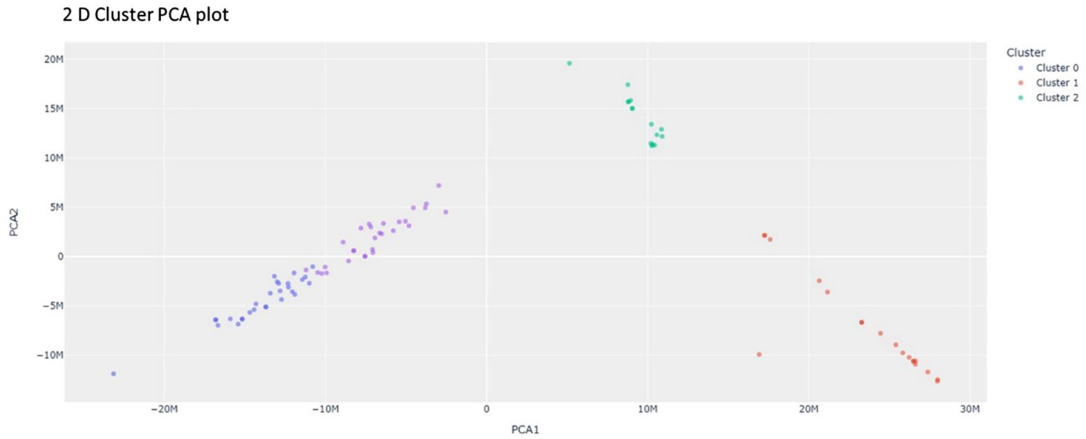


Figure 2. Exploratory analysis of blood samples from animals with toxoplasmosis (represented by red and green colors) and blood samples from animals in the control group (represented by bluish color).

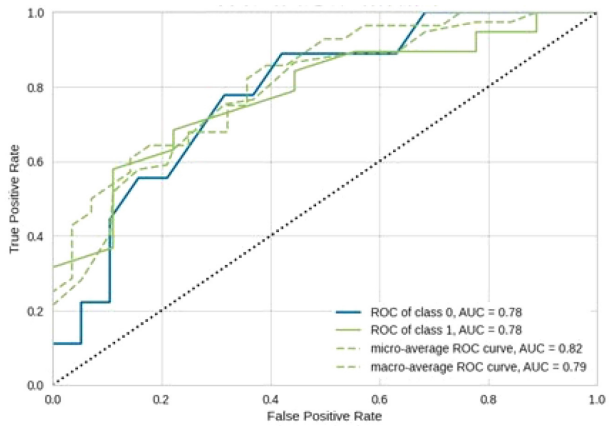


Figure 3. Receiver operating characteristic (ROC) curve for the selected model used to classify samples from animals with toxoplasmosis and the control group.

signals ranked from highest to lowest importance were *m/z* 809.2713 (t_R 7.49 min), *m/z* 660.5429 (t_R 4.51 min), *m/z* 718.3614 (t_R 8.28 min), *m/z* 537.5458 (t_R 3.93 min), *m/z* 707.2745 (t_R 7.54 min), *m/z* 318.2951 (t_R 3.85 min), *m/z* 568.5286 (t_R 6.92 min), *m/z* 774.5705 (t_R 8.19 min), *m/z* 675.6419 (t_R 8.14 min) and *m/z* 353.2914 (t_R 4.57 min).

Discussion

Knowledge about the pathophysiology of oocyst-induced toxoplasmosis is still limited, especially in FRNP. There is a need for new biomarkers that can reliably detect infections and stratify patients infected with *T. gondii* oocysts. Considering the strong need for the identification of new biomarkers that can reliably detect toxoplasmosis

Table 1. Performance of evaluated supervised models

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT / s
Ada Boost Classifier	0.7786	0.7317	0.8800	0.8264	0.8431	0.4436	0.4643	1.1540
Extreme Gradient Boosting	0.7619	0.6833	0.9050	0.7900	0.8382	0.3787	0.4035	1.4850
Random Forest Classifier	0.7524	0.6875	0.9800	0.7486	0.8443	0.3050	0.3272	1.3570
Extra Trees Classifier	0.7524	0.6992	0.9800	0.7486	0.8443	0.3050	0.3272	1.2070
Light Gradient Boosting Machine	0.7476	0.7133	0.9250	0.7595	0.8300	0.3095	0.3248	0.9140
Naive Bayes	0.7357	0.6525	0.9050	0.7686	0.8183	0.3146	0.3346	0.6310
K Neighbors Classifier	0.6667	0.5742	0.7650	0.7700	0.7494	0.2320	0.2571	0.5140
Quadratic Discriminant Analysis	0.6667	0.5000	1.0000	0.6667	0.7994	0.0000	0.0000	0.4580
Dummy Classifier	0.6667	0.5000	1.0000	0.6667	0.7994	0.0000	0.0000	0.4380
Linear Discriminant Analysis	0.6500	0.6092	0.8300	0.6833	0.7427	0.1488	0.1669	0.4730
Logistic Regression	0.6381	0.6750	0.7250	0.7433	0.7251	0.1814	0.1814	1.0500
Gradient Boosting Classifier	0.6381	0.6017	0.7350	0.7233	0.7193	0.1570	0.1755	1.6260
Ridge Classifier	0.6071	0.0000	0.6550	0.6467	0.6449	0.1803	0.1791	0.4690
SVM - Linear Kernel	0.5881	0.0000	0.5500	0.8367	0.6127	0.1562	0.1981	0.5770
Decision Tree Classifier	0.5738	0.5558	0.5950	0.7217	0.6366	0.1036	0.1157	0.4600

Performance metrics included accuracy, area under the receiver operating characteristic curve (AUC-ROC), recall (sensitivity), precision, F1-score, Cohen's kappa, and Matthews correlation coefficient (MCC). TT: training time. Models were trained using 70% of the dataset and evaluated on the remaining 30%.

infections, metabolomics offers a powerful means to investigate parasite-host interactions at the biochemical level and discover new infection biomarkers.

There are already some studies in the literature that have identified certain metabolites related to toxoplasmosis in other species, such as rats, mice[16] and humans.[17] To our knowledge, this was the first study conducted using an untargeted high-resolution liquid chromatography-mass spectrometry method to explore different serum metabolites of FRNP with toxoplasmosis compared to uninfected controls and to identify metabolites that become dysregulated during *T. gondii* infection.

As shown by the unsupervised model, there is a clear separation between the different groups of FRNP. These results indicate that *T. gondii* infection may induce systemic metabolic disturbances in the metabolism of various chemical classes of molecules. Previous metabolomics studies in experimental animal models,[18-20] and in humans[17] have also demonstrated widespread metabolic alterations associated with toxoplasmosis, including perturbations in lipid, amino acid, and energy metabolism. These findings support the biological plausibility of the metabolic differences observed in the present study.

Additionally, in the analyses using supervised algorithms, ten metabolites detected in ESI+ mode were found to be related to the separation of samples from the positive and negative groups for toxoplasmosis. The variability observed in performance metrics across different machine-learning algorithms can be attributed to their distinct mathematical assumptions and learning mechanisms. Linear models (e.g., Logistic Regression, LDA, and linear SVM) assume linear separability and may be more sensitive to high-dimensional feature spaces and multicollinearity. In contrast, tree-based ensemble methods such as AdaBoost, Random Forest, and Gradient Boosting are generally more robust to nonlinear relationships and complex variable interactions, which are common in untargeted metabolomics datasets.[21,22] These differences are consistent with previous reports indicating that ensemble methods often perform better in biological datasets with high feature-to-sample ratios.[21,22]

The molecular formulas for these metabolites were derived from high-resolution mass spectrometry and confirmed by comparison with public databases (METLIN, HMDB) and previous literature. Among the identified metabolites, the feature at *m/z* 318.2951 was tentatively associated with sphingosine ($C_{18}H_{39}NO_3$), a sphingolipid. This metabolite was previously reported in the study conducted by Chen *et al.*,[18] in samples from rats. Sphingolipids play critical roles in mediating cell interactions, modulating the behavior of proteins and cell receptors, and participating in signal transduction.[19] There are already previous studies in the literature confirming that *T. gondii* infection interferes with fatty acid and lipid metabolism by modulating the host's peroxisome proliferator-activated receptor (PPAR) signaling pathway. More specifically, there are reports of positive regulation of metabolites involved in sphingolipid metabolism in both phases of infection (acute and chronic).[18-20]

The mass *m/z* 774.5705 was identified, corresponding to the compound $C_{42}H_{80}NO_9P$. This compound is an oxidized phosphatidylcholine, which is a glycerophospholipid in which a phosphatidylcholine moiety occupies a glycerol substitution site. At least one of the fatty acid chains has undergone oxidation. Like all oxidized lipids, oxidized phosphatidylcholines belong to a group of biomolecules that play a role as signaling molecules.[18,23,24] Elevated levels of phosphatidylcholine are commonly found in patients with toxoplasmosis, playing an important role in the growth of the etiological agent,[24] which may explain the detection of this metabolite.[18,23,24] The remaining eight metabolites also contributed to sample separation, including amino acids, fatty acids, and energy-related metabolites; although not all have been fully characterized, they may play key roles in host response to infection and warrant further functional investigation.

Our study is not without limitations; in the unsupervised model, there was a clear separation of samples from the positive group, possibly related to certain aspects of the disease such as the pathology phase. Previous studies[16,17,20,25] have reported different metabolites in the acute and chronic phases of the disease; therefore, future studies should be conducted considering this aspect.

Additionally, some compounds that allowed the separation of samples in the present study have not been fully elucidated, necessitating new approaches for their identification. The absence of MS/MS spectral confirmation and validation using authentic standards represents a limitation of the present study and should be addressed in future targeted analyses.

Future studies should also explore hyperparameter optimization for machine-learning models, as these can substantially influence predictive performance. Furthermore, although the dataset was randomly divided into independent training (70%) and test (30%) subsets to evaluate model performance, more robust validation strategies could further strengthen the reliability of the results. Future studies should incorporate nested cross-validation, permutation testing, and validation using independent external cohorts to better assess model stability and generalizability. Due to the limited availability of biological samples from free-ranging neotropical primates,

the present study was designed as an exploratory analysis.

The limited sensitivity and specificity of serological methods may also have affected the classification of infected and non-infected individuals, potentially influencing the interpretation of both metabolomics and machine-learning results. In fact, the models achieved moderate predictive performance, with AUC values around 0.78, reflecting the complexity of host-parasite interactions and the heterogeneity of metabolic responses. Moreover, metabolomic analysis was performed only in positive ionization mode within a mass range of 50-900 Da, which may have restricted metabolite detection. Future studies using complementary ionization modes, broader analytical conditions, and additional annotation strategies may provide a more comprehensive metabolic profile.

In particular, GNPS-based tools such as Molecular Networking, Dereplicator, and MolNetEnhancer may improve metabolite identification and chemical characterization. In conclusion, this study provides evidence of systemic metabolic alterations in FRNP infected with *T. gondii* and identifies potential biomarkers for infection stratification. Validation in independent cohorts, refinement of machine-learning models through hyperparameter tuning, and functional characterization of unidentified metabolites are essential next steps to strengthen these findings and improve our understanding of host-parasite interactions in toxoplasmosis.

Conclusions

In conclusion, this study demonstrated that *T. gondii* infection causes significant dynamic changes in the metabolic balance of plasma from *T. gondii*-infected FRNP and the capability of UPLC-QTOF-MSE-based metabolomics to detect potential differential metabolites in the plasma of FRNP. Nonetheless, further studies are needed to fully understand the metabolic alterations in samples infected with *T. gondii*. These findings contribute to the growing body of literature on the application of omics approaches to toxoplasmosis.

Data Availability Statement

The data that support the findings of this study are partially available. The scripts used for data preprocessing, machine learning model development, and statistical analyses are publicly available at: https://github.com/Bboger/-Identification-of-potential-biomarkers-of-toxoplasmosis..git. Due to ethical and legal restrictions related to biological samples from free-ranging neotropical primates, the raw metabolomics datasets are not publicly available but may be obtained from the corresponding author upon reasonable request.

Acknowledgments

The authors would like to thank the UFPR for institutional support and infrastructure. We also acknowledge the Center for Studies in Biopharmacy (CEB-UFPR) for providing laboratory facilities and technical support. Additionally, we thank the Specialization Program in Data Science & Big Data at UFPR for contributing to the development of analytical and computational skills applied in this study.

Author Contributions

All authors of this article participated directly in the planning, execution and writing of this study. M. M. F. was responsible for conceptualization, data curation, formal analysis, investigation, methodology, resources, software, writing (original draft, review and editing); B. B. for conceptualization, methodology, writing (original draft, review and editing); A. F. C. for conceptualization, data curation, formal analysis, investigation, methodology, software and writing review and editing; W. K. S. for conceptualization, project administration, resources, writing (original draft, review and editing); P. J. R. J. for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, visualization and writing review and editing; R. P. for conceptualization, project administration, resources, supervision, validation, visualization, writing (original draft, review and editing).

References

1. Andrade, M. C. R.; Coelho, J. M. C. O.; Amendoeira, M. R. R.; Vicente, R. T.; Cardoso, C. V. P.; Ferreira, P. C. B.; *Ciênc. Rural* **2007**, *37*, 1724. [Crossref]
2. Robert-Gangneux, F.; Dardé, M. L.; *Clin. Microbiol. Rev*. **2012**, *25*, 264. [Crossref]
3. Dubey, J. P.; *Toxoplasmosis of Animals and Humans*; CRC Press: Boca Raton, USA, 2016. [Crossref]
4. Dunay, I. R.; Gajurel, K.; Dhakal, R.; Liesenfeld, O.; Montoya, J. G.; *Clin. Microbiol. Rev.* **2018**, 31. [Crossref]
5. Nishimura, M.; Goyama, T.; Tomikawa, S.; Fereig, R. M.; El-Alfy, E. N.; Nagamune, K.; Kobayashi, Y.; Nishikawa, Y.; *Parasitol. Int.* **2019**, *68*, 79. [Crossref]
6. Niehaus, C.; Spinola, M.; Su, C.; Rojas, N.; Rico-Chavez, O.; Ibarra-Cerdena, C. N.; *Front. Vet. Sci.* **2020**, *7*, 583032. [Crossref]
7. Zhou, C. X.; Zhou, D. H.; Elsheikha, H. M.; Liu, G. X.; Suo, X.; Zhu, X. Q.; *PLoS One* **2015**, *10*, e0139635. [Crossref]
8. Galal, A.; Talal, M.; Moustafa, A.; *Front. Genet.* **2022**, *13*, 1017340. [Crossref]
9. Zhang, X.; Rahnavard, A.; Crandall, K. A.; *Med. Res. Arch.* **2025**, *13*, 6888. [Crossref]

10. Tounta, V.; Liu, Y.; Cheyne, A.; Larrouy-Maumus, G.; *Mol. Omics* **2021**, *17*, 376. [Crossref]
11. Le, A. T.; Wu, M.; Khan, A.; Phillips, N.; Rajpurkar, P.; Garland, M.; Magid, K.; Sibai, M.; Huang, C.; Sahoo, M. K.; Bowen, R.; Cowan, T. M.; Pinsky, B. A.; Hogan, C. A.; *Front. Microbiol.* **2022**, *13*, 1059289. [Crossref]
12. Svoboda, W. K.; *Vigilância de Epizootias em Primatas Não-Humanos (PNH) como Instrumento de Monitoramento de Arboviroses e Outras Viroses de Interesse em Saúde Pública*; PhD Thesis, Universidade Estadual de Londrina: Londrina, Brazil, 2007. [Link] accessed in January 2026
13. Kacerova, T.; Pires, E.; Dixon, A.; Williams, R.; Legge, I.; Hippisley, M.; Yates, A. G.; Smith, D.; Anthony, D. C.; Probert, F.; McCullagh, J. S. O.; *Anal. Chem.* **2025**, *97*, 23014. [Crossref]
14. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M. Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Joachim Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R.; *Metabolomics* **2007**, *3*, 211. [Crossref]
15. Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, A. D.; Lewis, M. R.; Dunn, W. B.; *Metabolomics* **2018**, *14*, 72. [Crossref]
16. Zhou, C. X.; Zhou, D. H.; Elsheikha, H. M.; Zhao, Y.; Suo, X.; Zhu, X. Q.; *Sci. Rep*. **2016**, *6*, 19557. [Crossref]
17. King, E. F. B.; Cobbold, S. A.; Uboldi, A. D.; Tonkin, C. J.; McConville, M. J.; *Methods Mol. Biol.* **2020**, *2071*, 435. [Crossref]
18. Chen, X. Q.; Elsheikha, H. M.; Hu, R. S.; Hu, G. X.; Guo, S. L.; Zhou, C. X.; Zhu, X.-C.; *Front. Cell. Infect. Microbiol.* **2018**, *8*, 189. [Crossref]
19. Merrill Jr., A. H.; *J. Biol. Chem*. **2002**, *277*, 25843. [Crossref]
20. Ma, J.; He, J. J.; Hou, J. L.; Zhou, C.-X.; Elsheikha, H. M.; Zhu, X-Q.; *Front. Microbiol.* **2020**, *11*, 1555. [Crossref]
21. Libbrecht, M. W.; Noble, W. S.; *Nat. Rev. Genet.* **2015**, *16*, 321. [Crossref]
22. Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodcare, R.; *Anal. Chim. Acta* **2015**, *879*, 10. [Crossref]
23. Ricard, J.; Pelloux, H.; Favier, A. L.; Gross, U.; Brambilla, E.; Ambroise-Thomas, P.; *Exp. Parasitol*. **1999**, *91*, 231. [Crossref]
24. Welti, R.; Mui, E.; Sparks, A.; Wernimont, S.; Isaac, G.; Kirisits, M.; Roth, M.; Roberts, C. W.; Botté, C.; Maréchal, E.; McLeod, R.; *Biochemistry* **2007**, *46*, 13882. [Crossref]
25. Zhou, C. X.; Cong, W.; Chen, X. Q.; He, S. Y.; Elsheikha, H. M.; Zhu, X. Q.; *Front. Microbiol.* **2017**, *8*, 2612. [Crossref]

Submitted: December 7, 2025
Final version online: May 5, 2026