

Supplementary Material

A new strategy for canine visceral leishmaniasis diagnosis based on molecular spectroscopy and machine learning

Gustavo Larios¹, Matheus Ribeiro¹, Carla Arruda², Samuel L. Oliveira¹, Thalita Canassa¹, Matthew J. Baker³, Bruno Marangoni¹, Carlos Ramos⁴, Cícero Cena^{1*}

¹ Grupo de Óptica e Fotônica, Instituto de Física, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil.

² Laboratório de Parasitologia Humana, Instituto de Biociências, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil.

³ Pure and Applied Chemistry, University of Strathclyde, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1 RD, UK

⁴ Faculdade de Medicina Veterinária e Zootecnia, UFMS – Universidade Federal de Mato Grosso do Sul, Campo Grande, Brazil.

* cicero.cena@ufms.br

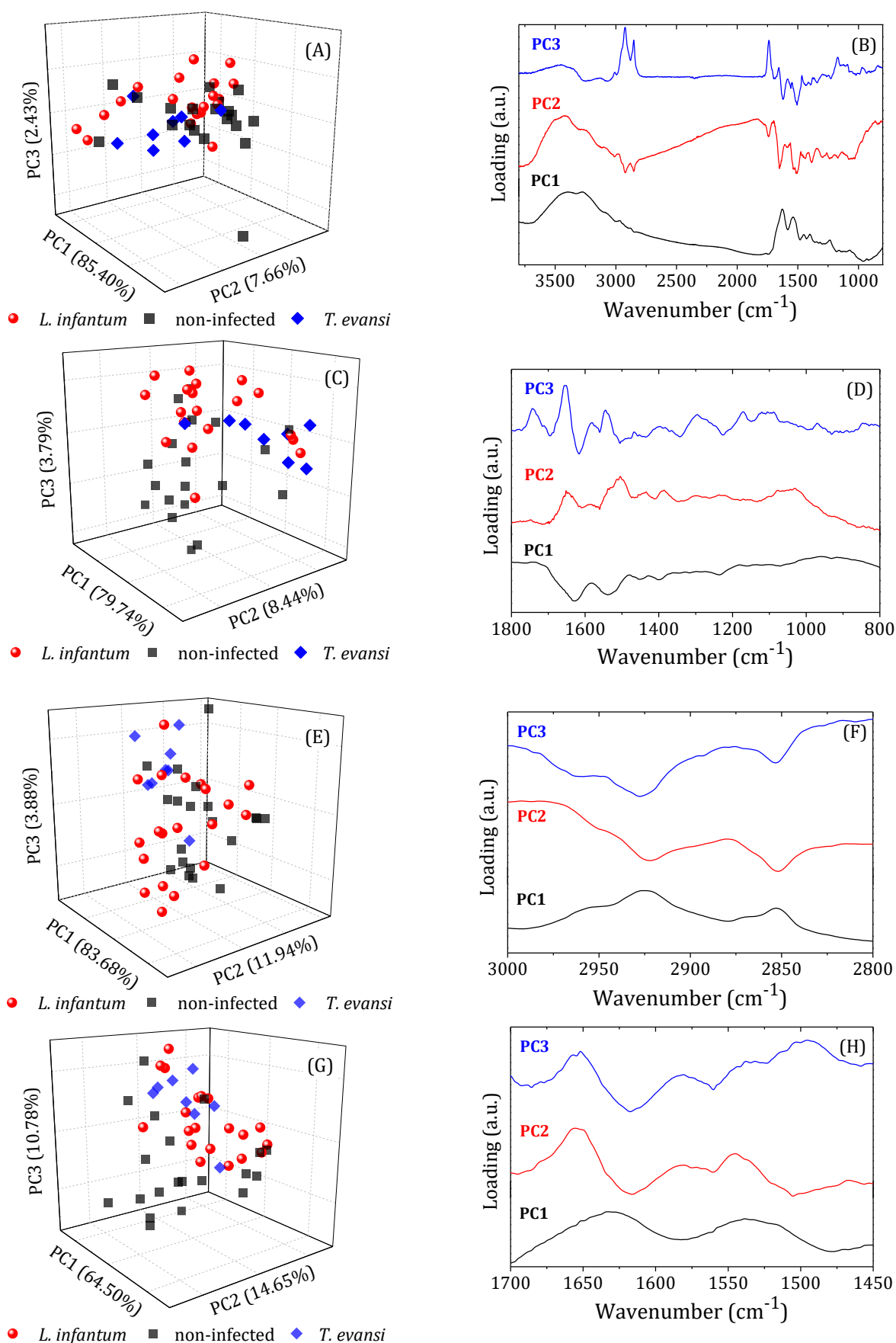


Figure S1: Score plot (left) and loadings (right) from principal component analysis (PCA) for three sample groups; non-infected (black square), *L. infantum* (red circle), and *T. evansi* (blue diamond) infected canine blood serum. (a, b) entire spectral range 4000 until 800 cm^{-1} , involving all molecular groups in the sample. (c, d) 1800 until 800 cm^{-1} range, related to lipids, proteins, carbohydrates, and fatty acids molecules. (e, f) 3000 until 2800 cm^{-1} range, main contributions assigned to lipids. (g, h) 1700 until 1450 cm^{-1} range, mainly related to proteins contributions.

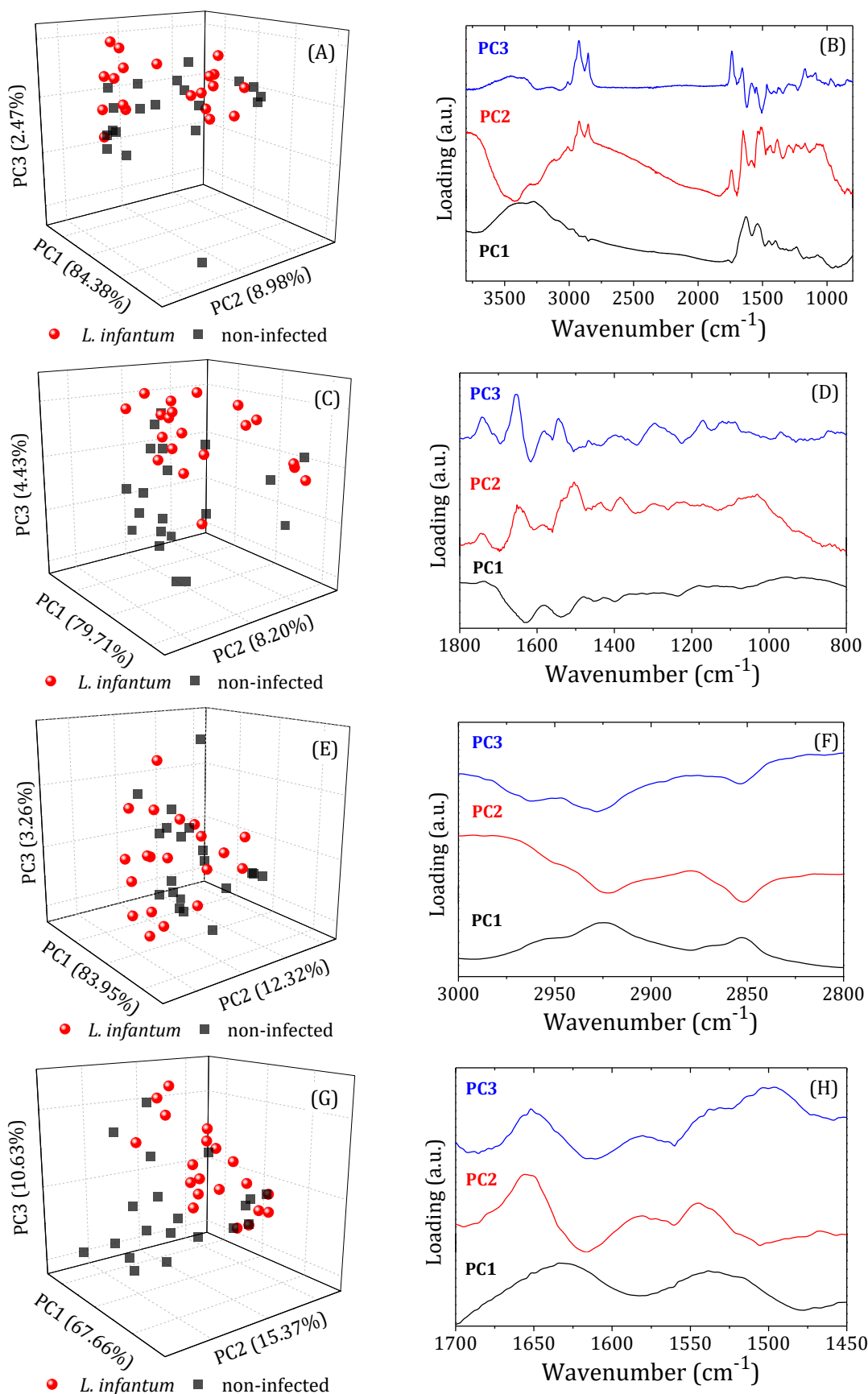


Figure S2: Score plot (left) and loadings (right) from principal component analysis (PCA) for two sample groups; non-infected (black square), and *L. infantum* (red circle) infected canine blood serum. (a, b) entire spectral range 4000 until 800 cm^{-1} , involving all molecular groups in the sample. (c, d) 1800 until 800 cm^{-1} range, related to lipids, proteins, carbohydrates, and fatty acids molecules. (e, f) 3000 until 2800 cm^{-1} range, main contributions assigned to lipids. (g, h) 1700 until 1450 cm^{-1} range, mainly related to proteins contributions.

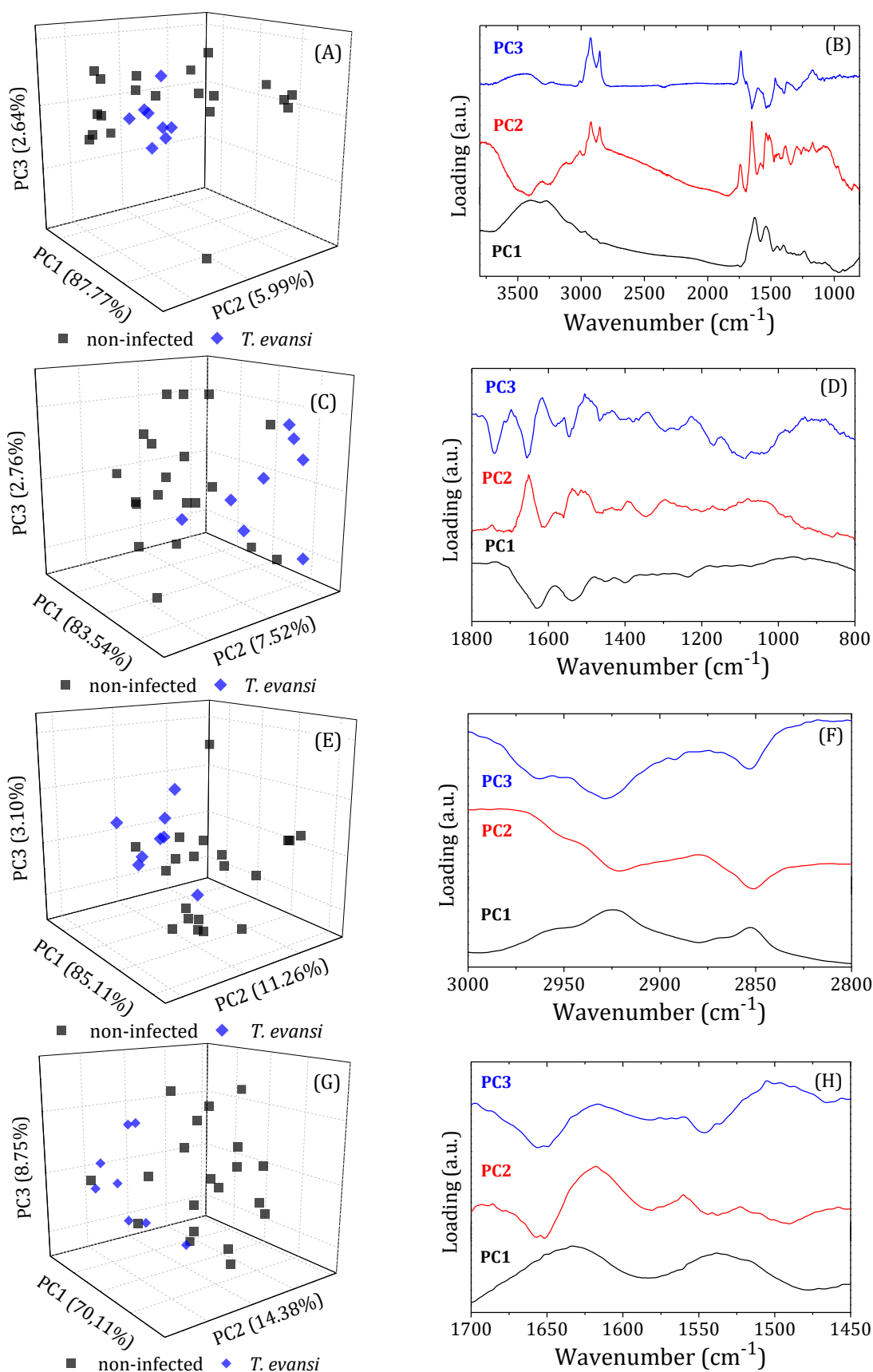


Figure S3: Score plot (left) and loadings (right) from principal component analysis (PCA) for two sample groups; non-infected (black square), and *T. evansi* infected canine blood serum. (a, b) entire spectral range 4000 until 800 cm^{-1} , involving all molecular groups in the sample. (c, d) 1800 until 800 cm^{-1} range, related to lipids, proteins, carbohydrates, and fatty acids molecules. (e, f) 3000 until 2800 cm^{-1} range, main contributions assigned to lipids. (g, h) 1700 until 1450 cm^{-1} range, mainly related to proteins contributions.

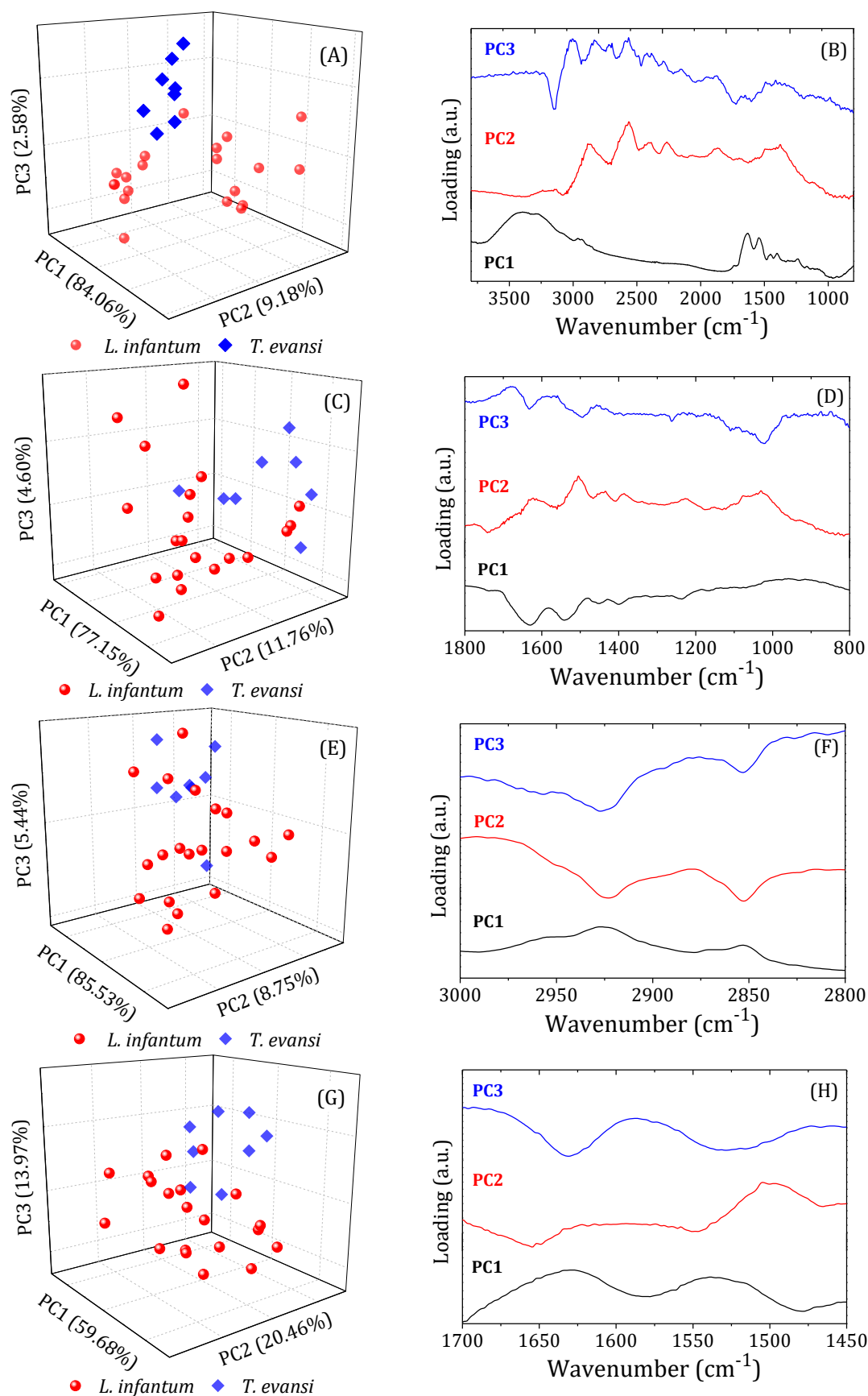


Figure S4: Score plot (left) and loadings (right) from principal component analysis (PCA) for two sample groups; *L. infantum* (red circle), and *T. evansi* (blue diamond) infected, canine blood serum. (a, b) entire spectral range 4000 until 800 cm^{-1} , involving all molecular groups in the sample. (c, d) 1800 until 800 cm^{-1} range, related to lipids, proteins, carbohydrates, and fatty acids molecules. (e, f) 3000 until 2800 cm^{-1} range, main contributions assigned to lipids. (g, h) 1700 until 1450 cm^{-1} range, mainly related to proteins contributions.

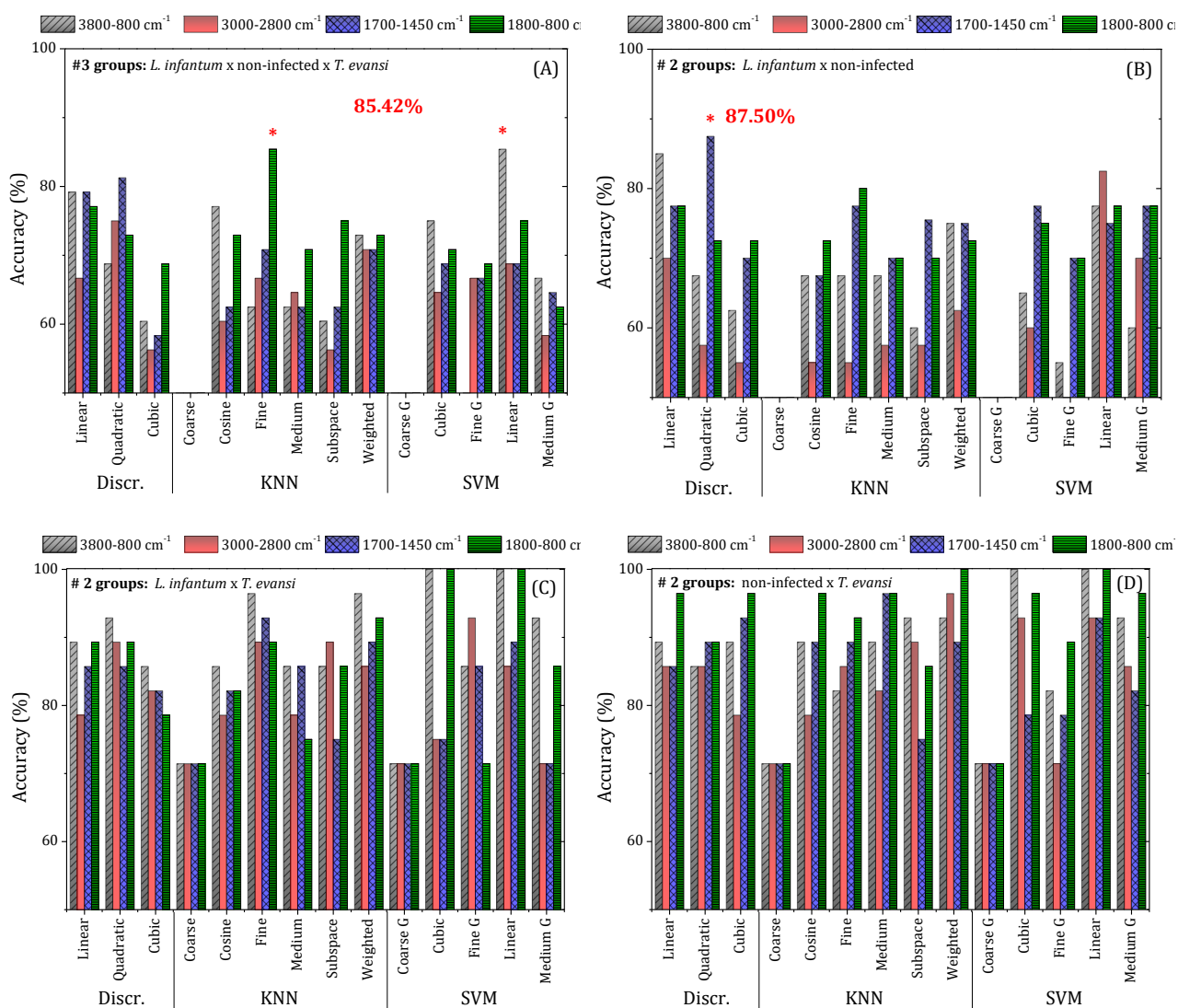


Figure S5: Accuracy of learning machine methods for different spectral ranges: Gray bars (inclined dash pattern) for 3800 until 800 cm⁻¹; Red bars for 3000 until 2800 cm⁻¹; Blue bars (mesh pattern) for 1700 until 1450 cm⁻¹; and Green bars (horizontal dash pattern) for 1800 until 800 cm⁻¹. Groups classification: (A) non-infected versus *L. infantum* versus *T. evansi*. (B) non-infected versus *L. infantum*. (C) *L. infantum* versus *T. evansi*. (D) non-infected versus *T. evansi*, classification. The letter G abbreviated in the label means “gaussian”.

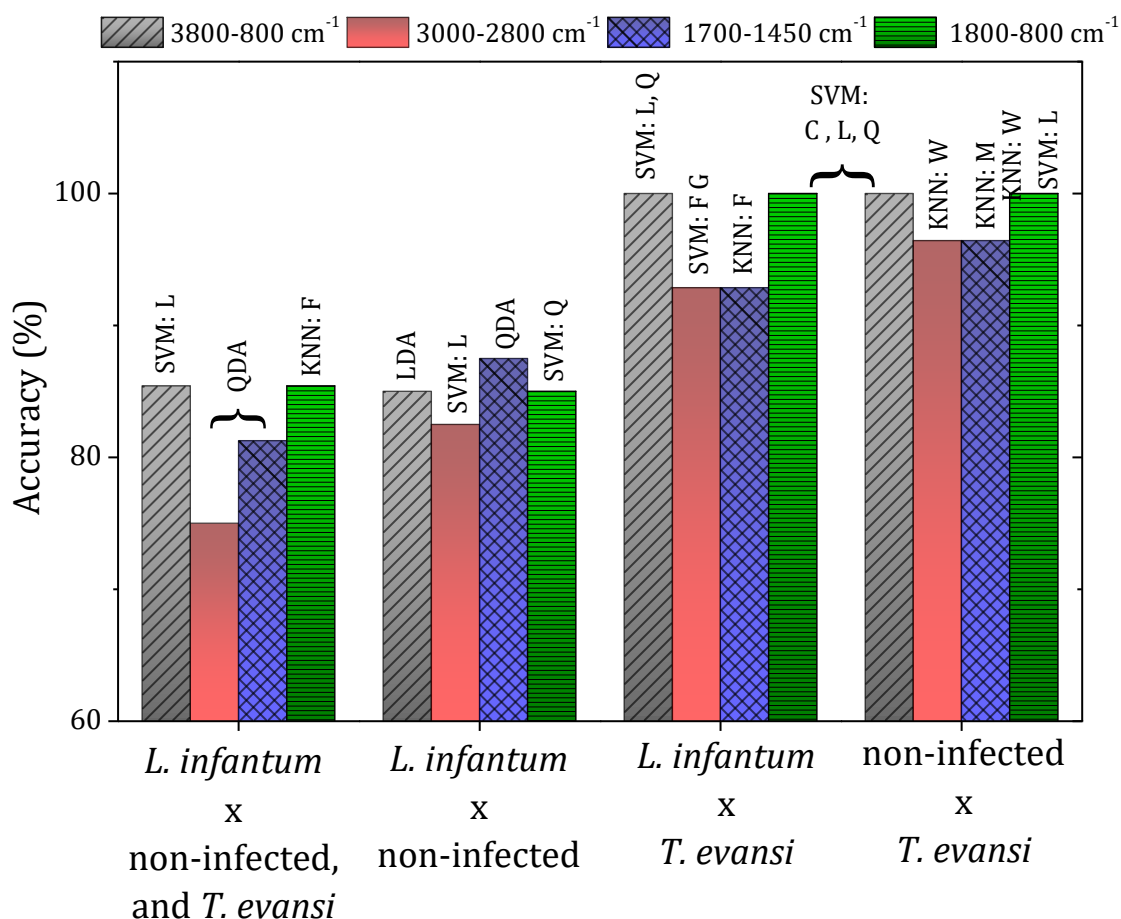


Figure S6: Higher overall accuracy of learning machine methods achieved for different groups classification and spectral ranges: Gray bars (inclined dash pattern) for 3800 until 800 cm^{-1} ; Red bars for 3000 until 2800 cm^{-1} ; Blue bars (mesh pattern) for 1700 until 1450 cm^{-1} ; and Green bars (horizontal dash pattern) for 1800 until 800 cm^{-1} . ML methods: Support vector machine (SVM), with Linear (L), Quadratic (Q), Cubic (C), and Fine Gaussian (F G) functions; k-nearest neighbor (KNN), with Fine (F), Medium (M), and Weighted (W) functions; Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA).