Veterinary Parasitology 234 (2017) 13–18

Contents lists available at ScienceDirect

Veterinary Parasitology

journal homepage: www.elsevier.com/locate/vetpar

Research paper

Prediction of lymph node parasite load from clinical data in dogs with leishmaniasis: An application of radial basis artificial neural networks

Rafaela Beatriz Pintor Torrecilha[a], Yuri Tani Utsunomiya[b], Luís Fábio da Silva Batista[c], Anelise Maria Bosco[a], Cáris Maroni Nunes[d], Paulo César Ciarlini[a], Márcia Dalastra Laurenti[c,*]

[a] São Paulo State University (Unesp). School of Veterinary Medicine, Araçatuba. Department of Clinics, Surgery and Animal Reproduction. São Paulo, Brazil
[b] São Paulo State University (Unesp). School of Agricultural and Veterinarian Sciences, Jaboticabal. Department of Preventative Veterinary Medicine and Animal Reproduction. São Paulo, Brazil
[c] USP – Universidade de São Paulo, Departamento de Patologia Veterinária, Faculdade de Medicina Veterinária e Zootecnia, São Paulo, Brazil
[d] São Paulo State University (Unesp). School of Veterinary Medicine, Araçatuba. Department of Support, Production and Animal Health. São Paulo, Brazil

ARTICLE INFO

Article history:
Received 24 September 2016
Received in revised form
16 December 2016
Accepted 18 December 2016

Keywords:
Canis lupus familiaris
qPCR
Machine learning
Leishmania spp.

ABSTRACT

Quantification of *Leishmania infantum* load via real-time quantitative polymerase chain reaction (qPCR) in lymph node aspirates is an accurate tool for diagnostics, surveillance and therapeutics follow-up in dogs with leishmaniasis. However, qPCR requires infrastructure and technical training that is not always available commercially or in public services. Here, we used a machine learning technique, namely Radial Basis Artificial Neural Network, to assess whether parasite load could be learned from clinical data (serological test, biochemical markers and physical signs). By comparing 18 different combinations of input clinical data, we found that parasite load can be accurately predicted using a relatively small reference set of 35 naturally infected dogs and 20 controls. In the best case scenario (use of all clinical data), predictions presented no bias or inflation and an accuracy (i.e., correlation between true and predicted values) of 0.869, corresponding to an average error of ±38.2 parasites per unit of volume. We conclude that reasonable estimates of *L. infantum* load from lymph node aspirates can be obtained from clinical records when qPCR services are not available.

1. Introduction

Visceral leishmaniasis (VL) is a zoonotic disease caused by the protozoan parasite *Leishmania infantum* (synonym *Leishmania chagasi*) (WHO, 2010). Over 90% of the global prevalence of VL is concentrated in India, Bangladesh, Sudan, South Sudan, Ethiopia and Brazil (Alvar et al., 2012). In the Americas, parasites are typically transmitted from host to host by the bites of female *Lutzomyia longipalpis* sand flies. After being inoculated in the skin, promastigote parasites are phagocytized by macrophages or dendritic cells, where they turn into amastigotes, multiply, and disseminate to various tissues (Solano-Gallego et al., 2009, 2011)

The domestic dog (*Canis lupus familiaris*) is a key animal model for VL. Dogs are the most important urban reservoirs of the disease (Dantas-Torres, 2007; Gramiccia and Gradoni, 2005) (Gramiccia and Gradoni, 2005; Dantas-Torres, 2007) and are also highly susceptible to *L. infantum*. They can express a wide range of physical signs such as those of dermititis, alopecia, hyperkeratosis, lymphadenomegaly, ophtalmic alterations, pallor of mucous membranes, splenomegaly, emaciation, fever, epistaxis and onychogryphosis (Baneth et al., 2008). They can also present high levels of circulating anti-*L. infantum* antibodies (Lima et al., 2005) and alterations of biochemical findings, such as uremia and hyperglobulinemia (Freitas et al., 2012). Importantly, the occurrence of such alterations have been previously associated with increased parasite load (PL) in tissues such as lymph nodes, bone marrow, spleen (Manna et al., 2009; Reis et al., 2006) and other peripheral tissues (Torrecilha et al., 2016).

Since PL is associated with clinical alterations, one can hypothesize that patients with similar levels of infection might also have similar clinical profiles, such that knowledge of PL may have a prognostic as well as an epidemiological value (Martínez et al., 2011), which may directly impact therapeutic conduct and decisions in disease control. This association could also be turned into the opposite direction, namely clinical data could be used to predict PL.

* Corresponding author at: USP – Univ de São Paulo. Av Dr Arnaldo, 455–sala 1209, Cerqueira César, São Paulo, 01246–903 Brazil.
E-mail address: mdlauren@usp.br (M.D. Laurenti).

http://dx.doi.org/10.1016/j.vetpar.2016.12.016
0304-4017/

Given PL has typically been determined by limiting dilution (Titus et al., 1985) and real-time quantitative polymerase chain reaction (qPCR) (Francino et al., 2006), which is costly and demands specialized equipment and properly trained personnel, prediction of unobserved PL from clinical data could be an useful tool to generate information about the extent of infection in situations where direct quantification is impracticable.

One important issue in this prediction problem is that the precise nature of the relationships between various clinical parameters and PL is largely unknown. Therefore, devising an accurate statistical model, at least relying on current knowledge, seems intractable without heuristics. In this situation, machine learning techniques come in handy, since they require no explicit model of the relationships between predictand (output) and predictors (inputs), such that they can recognize (non)linear patterns arising from data structure implicitly (Hastie et al., 2009). Radial Basis Artificial Neural Network (RB-ANN) is an example of such technique (Orr, 1996), which has been successfully applied to a wide range of problems, from adaptive flight control (Yilmaz and Özer, 2009) to speech recognition (Venkateswarlu et al., 2011).

In order to test whether clinical data could reliably predict PL, we used a dataset comprised of 35 naturally infected dogs and 20 controls to assess the accuracy of predictions made by a RB-ANN machine trained on an example set of physical signs, serological test and biochemical markers.

2. Material and methods

2.1. Dog samples

A total of 55 owned dogs, including 35 naturally infected dogs and 20 controls, were sampled from 7 distinct endemic areas, distributed among the Brazilian states of Bahia, Minas Gerais, São Paulo and the Federal District. Breeds included English Bulldog (n = 2), Cocker Spaniel (n = 1), Doberman Pinscher (n = 1), Labrador Retriever (n = 20), Golden Retriever (n = 2), German Shepherd (n = 8), Belgian Shepherd (n = 6), White Swiss Shepherd (n = 1), Toy Poodle (n = 1), Rottweiler (n = 11), German Spitz (n = 1) and Dachshund (n = 1). Samples were obtained according to the Ethical Principles in Animal Research by the Brazilian College of Animal Experimentation and approved by the Ethics Committee in the use of animals of the School of Veterinary Medicine and Animal Science from University of São Paulo (Protocol 2391/2011).

2.2. Parasite load data

Lymph node (popliteal or pre-scapular) fine needle aspiration biopsies were collected, maintained in NET buffer (0.15 M NaCl, 50 mM EDTA, 0.1 M Tris- Hcl, ph 7.5) in a final volume of 200 μL and stored at 4 °C. DNA was extracted using a commercial kit (NucleoSpin® Tissue, Machery Nagel, Germany), according to the manufacturer's recommendations. Quantity and quality of DNA were evaluated in a Biospectrometer (Eppendorf, German), and DNA samples were stored at −20 °C until used. For DNA amplification, primers targeting conserved regions of *Leishmania* kDNA were used, namely LEISH-1 (5′-AACTTTTCTGGTCCTCCGGGTAG-3′) and LEISH-2 (5′-ACCCCCAGTTTCCCGCC-3′). Additionally, eukaryotic 18S was used as internal reference of canine DNA (Francino et al., 2006). Amplification was performed in a final volume of 15 μL that consisted of 5 μL of total DNA (10 ng/μL), 7.5 μL of KAPA SYBR® FAST Universal 2X qPCR Master Mix (Kapa Biosystems, USA), 0.5 μL of each primer at a final concentration of 300 nM and 1.5 μL of deionized water. PCR amplification was carried out in a Realplex® thermocycler (Eppendorf, German) using the following conditions: 95 °C for 4 min, followed by 35 cycles of denaturation at 95 °C for 15 s, annealing of primers at 58 °C for 20 s and extension at 72 °C for 8 s. A standard curve (slope = −3.327, intercept = 28.25, efficiency = 1.00, $R^2 = 0.929$) was constructed with dilutions of *L. infantum* DNA (MHOM/BR/72/strain46) from 0.5×10^6 to 0.5×10^{-4} parasites/μL. The amplifications were performed in duplicate and parasite load was obtained via regression of the cycle threshold (Ct) values against the standardized parasite concentrations (Aschar et al., 2016) in Realplex® Software (Eppendorf, German). The results were defined as the number of parasites in volume of lymph node aspirate which yielded 10 ng of DNA. Parasite load were then transformed to a log10(PL + 1) scale for analysis.

2.3. Clinical data

Dog owners provided data on sex, age, vaccination and the use of any form of repellent (e.g., collars or spot-ons). These variables were included as input in all prediction analyses (NULL input set).

All dogs were evaluated by veterinary experts to determine the presence/absence of common physical signs found in leishmaniasis, such as skin alterations (dermatitis, alopecia and hyperkeratosis), ophthalmic alterations (uveitis, blepharitis, ulcer and secretion), onychogryphosis, lymphadenomegaly, pallor of mucous membranes and emaciation. These physical signs were used as either separate inputs (SIGNS1) or encoded as a single indicator variable (SIGNS2). In the first case, indicator variables (0 = absence, 1 = presence) were used for each physical sign individually. In the second case, animals were classified as sick or clinically healthy based on the presence of at least one physical sign.

Serum samples were obtained from whole blood collected via venipuncture for biochemical markers analysis and quantification of anti-*L. infantum* IgG levels. The quantification of biochemical markers was performed in an automated spectrophotometer (BS 200, Shenzhen Mindray Bio-Medical Eletronics Co., Nanshan, China), previously calibrated with serum control levels I and II (Biosystems, Barcelona, Spain). The biochemical markers and their respective quantification methods were: albumin (g/L) by bromocresol green method; urea (mg/dL) by the urease/glutamate dehydrogenase coupled with the UV enzymatic assay; creatinine (mg/dL) by the kinetic alkaline picrate assay; aspartate aminotransferase (AST, IU/L), alanine aminotransferase (ALT, IU/L) by enzymatic UV method, following the International Federation of Clinical Chemistry (IFCC) guidelines; bilirubin (mg/dL) by the diazotined sulfanilic method; uric acid (mg/dL) by the uricase/peroxidase enzymatic method and total plasma protein (g/L) by the biuret method. Globulin levels (g/L) were determined by the difference between total protein and albumin. Biochemical markers were treated as quantitative inputs (BIOCHEM1) or as indicator variables (BIOCHEM2) indicating levels above or below reference values, according to Kaneko et al. (1997).

Indirect enzyme-linked immunosorbent assay (ELISA) using crude *L. infantum* antigen (MHOM/BR/72/strain46) and anti-canine IgG (A40-123AP) conjugated to alkaline phosphatase (Bethyl, USA) was performed according to Laurenti et al., 2014 to evaluate the presence of anti *L. infantum* IgG in serum samples. Eleven dogs from non-endemic area, with negative cytological examination were used to establish the cut line (cut-off) of reaction that refers to the average of the negative controls added to three standard deviations.

2.4. Non-mathematical description of the prediction machine

The objective of this section is to provide a 'pictorial' description of the prediction problem treated in this paper, which may be useful to help grasping the general idea of a RB-ANN. We make clear, however, that this description is shallow and should be interpreted with caution when dissociated from its mathematical formalism. The problem is stated as follows: we have access to clinical data