Individual and Environmental Properties Determine Rats Role In Host-Pathogen Networks

1 Introduction

Emerging infectious diseases (EID) pose a significant health risk, particularly in developing countries, due to factors such as high disease burden, poor infrastructure, and inadequate water quality [1]. EID events are predominantly caused by zoonotic pathogens, accounting for 60.3% of events between 1940 and 2004, with 71.8% of these originating from wildlife [2]. Rural areas face more pronounced challenges, as exposure to zoonotic pathogens is often linked to agriculture, livestock, wildlife, and outdoor recreation [3]. Rodents, especially the species *Rattus rattus*, commonly known as the black rat, are recognized hosts of at least 60 zoonotic pathogens [4]. In Madagascar, *Rattus rattus* is prevalent in human-inhabited and fragmented environments, serving as a source of various diseases and presenting a significant potential for transmission to humans, primarily through environmental pathways [5].

Disease transmission is fundamentally characterized by heterogeneity at the individual level [6,7]. For example, some hosts are infected by many pathogens, while others are infected by very few. Additionally, the pathogens infecting them can range from being prevalent in the population to rare. However, it remains unclear which host features most significantly impact these infection patterns within *rattus* populations. I hypothesize that host features from different levels of organization contribute to this heterogeneity: (**H1**) *Individual features*, such as co-infection and microbiome composition, can impact an individual's physiology and immune system function [6]. (**H2**) *Class features*, which are traits shared by many individuals, such as sex, mass, and age, often correlate with physiology and behavior [6]. (**H3**) *Environmental features*, for environmentally-transmitted pathogens, such as habitat disturbance, host density, and vegetation, can influence host behavior and pathogen prevalence [8].

2 Research objective

I aim to explore the effect of features from different levels on *Rattus rattus* individuals infection patterns.

3 Methods

I used data from rural Madagascar, consisting of *rattus* individuals (n=540), sampled across three different village areas. In each area, traps were located in grids of different land cover types: semi-intact forest, secondary forest, agriculture (such as rice and vanilla fields), and the village itself. Gut samples were collected from each individual and sequenced to detect parasitic ASVs (ITS-2 gene for nematodes, 18S rRNA for protozoan parasites). Reads below 500 were filtered out. Then, I applied Operational Taxa Units (OTU) clustering on the ASVs by 97% similarity and filtered to pathogenic taxa only (n=40 for protozoa, n=75 for nematodes). OTU clustering is supported by multiple studies as they are more ecologically consistent units for broader ecological pattern analyses [9]. Samples were also sequenced to detect microbial ASVs (by 16s rRNA gene).

I focused on understanding protozoa infection patterns using features from multiple levels. Individual features include microbiome richness and nematode community composition, represented by two principal component (PC) variables created from the nematode data. Class features are traits collected

on each individual, which are often shared, here specifically sex, mass, and age (by dental estimation). For environmental features, I used data from each grid. Vegetation traits (such as number of trees and herb coverage) were reduced to two PC variables. Community data, based on different small-mammals sampling (such as *Mus musculus* and *Microgale mergulus*), was used also. This included both simple indices (*rattus* abundance, abundance of other species, and richness) and two PC variables created from the abundance and richness of each grid's community. The distance from the village was also used as an environmental feature.

Exploring infection patterns of individuals can be effectively accomplished using a host-pathogen network approach [7]. Each *rattus* individual is represented by a node, and the other set of nodes represents different pathogenic protozoa OTUs. A link indicates an infection of the individual by the particular pathogen, thus creating undirected-binary-bipartite network. This approach allows for the use of network analysis tools to capture and compare topological features with ecological significance. I chose to use the role of a node in the network, i.e., its contribution to network structure [7]. Individuals with similar roles have similar infection patterns within the context of the full network. To define each individual's role in the network, I used the SBM (Stochastic Block Model) community detection method. SBM groups nodes by having similar probabilities of connecting to other nodes (belonging to another group). It has been previously used to identify ecological roles [10]. As seen in Figure 1, I found that hosts are separated into four groups, and OTUs into five, with each level having a highly dominant group.

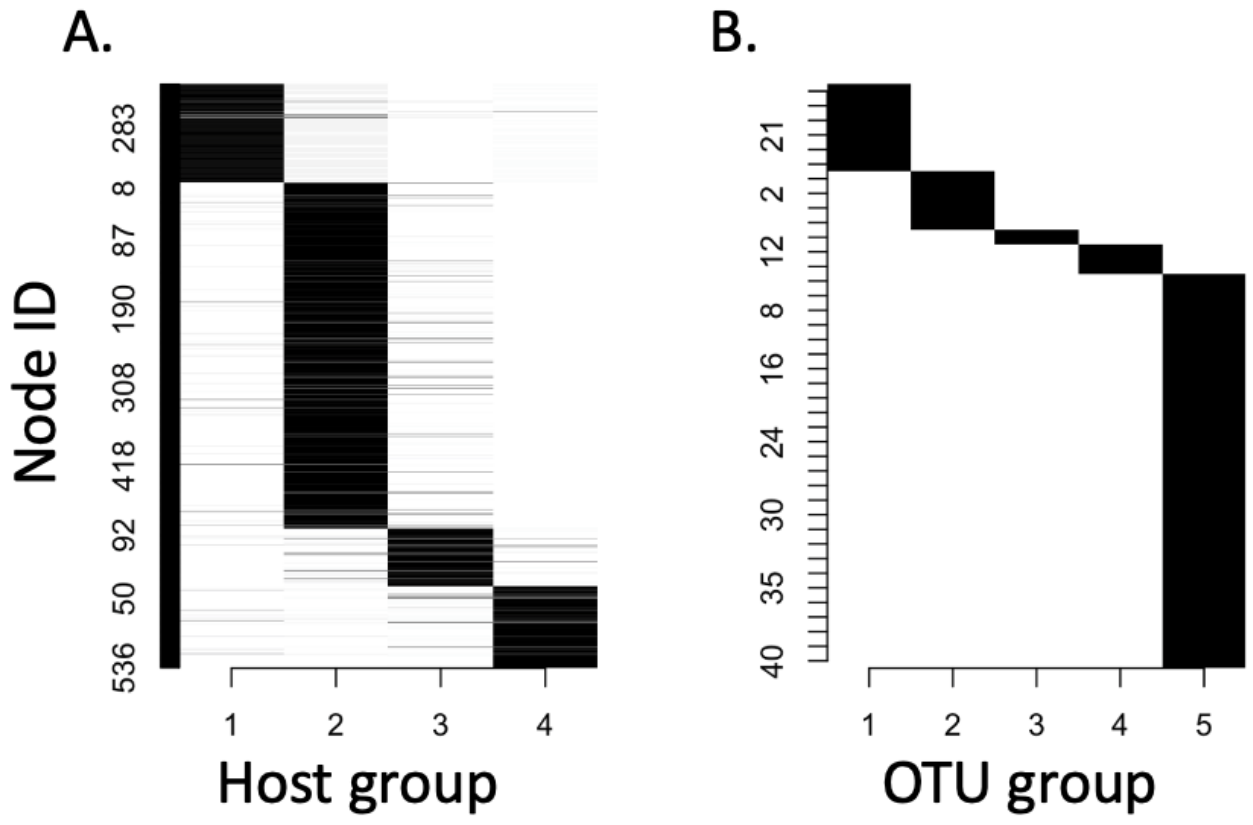

Fig. 1: Group assignments of *rattus* individuals (A) and protozoa OTUs (B). A darker line represents a higher probability of inclusion in the group.

To evaluate the effect of multi-level features on the host SBM-generated group membership (a proxy for ecological role), I used XGBoost, a machine learning algorithm, to create a classification model [11]. As the data is limited in size and contains NAs, I did not use out-of-sample prediction. Instead,

I applied 5-fold cross-validation to reduce the probability of overfitting. I used a confusion matrix to evaluate model performance, specifically the accuracy (i.e., the number of true predictions divided by the total number of predictions). To validate its significance, I ran the model 100 times to predict the SBM-generated memberships, and 100 times with the memberships shuffled randomly (maintaining the proportion of groups as in the true SBM-generated data).

All analysis and data processing were performed using R Statistical Software [12], the RStudio interface [13], the R package `blockmodels` [14] for the SBM analysis, and the R package `XGBoost` [15] for the classification model.

4 Results

As seen in Figure 2, the accuracy of the model is significantly greater than the shuffled predictions ($p < 0.01$). Additionally, the accuracy is significantly greater than the No Information Rate (NIR; $p < 0.01$), which is around 59% and represents the accuracy that would be achieved by always predicting the most frequent class. Salient features, i.e., those with high relative importance to the model, were from multiple levels (Figure 3). For example, mass had a mean importance of 28.5%, microbiome richness 19.7%, and PC1 of grid vegetation 8.9%.

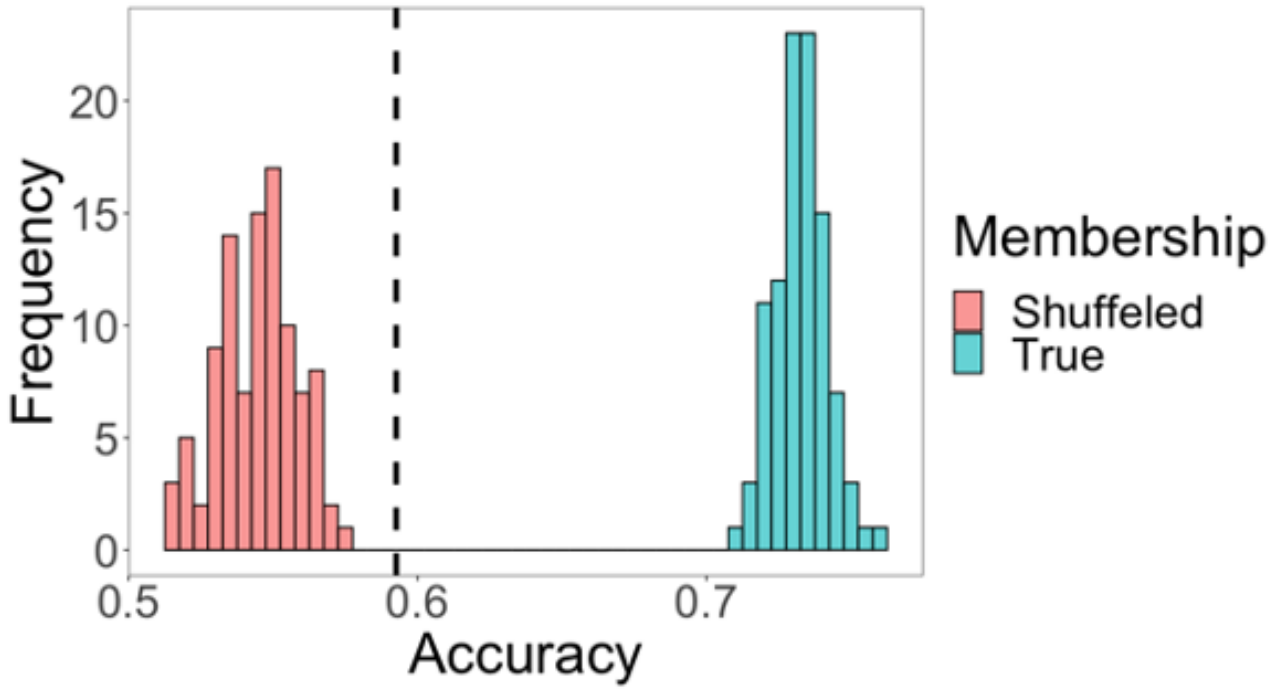

Fig. 2: Accuracy of 100 classification model predictions using the true SBM groups (blue) compared to 100 iterations using shuffled groups (red). The dashed line represents the No Information Rate (NIR) value.

5 Discussion

This study aimed to understand how features from multiple levels (individual, class, and environmental) influence infection patterns of rat individuals by protozoan OTUs. Rat groups, generated by SBM on the host-pathogen network, can be seen as a proxy for similar infection patterns and were significantly predicted using a classification model.

Mass, the most important feature in predicting infection patterns in my model, has been extensively studied in the context of host infection risk and susceptibility. Studies indicate that for certain pathogens, the infection risk is higher in heavier rats [16,17] or lighter rats [18]. It has been hypoth-

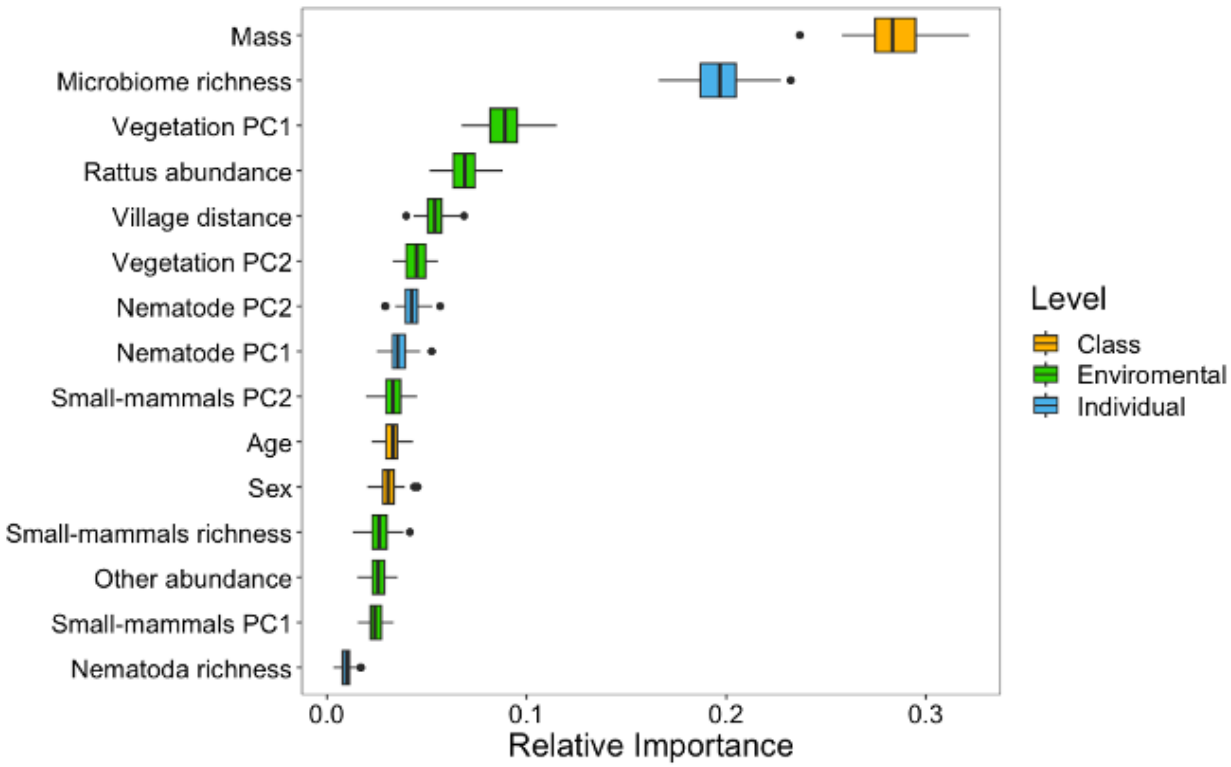

Fig. 3: Boxplots showing the relative importance of features over 100 iterations of running the model, colored by level: yellow for class features, green for environmental features, and blue for individual features.

esized and demonstrated in other cases that larger hosts might harbor more pathogenic organisms [19,20]. However, more infections can negatively impact mass and potentially make the individual more susceptible to co-infection [20]. Although the association of the microbiome community with disease is less known than that of mass, a study has shown that diseases can be associated with reduced microbiome richness [21]. My results support that the importance of the microbiome community on infection patterns should be further researched.

As most of the protozoan OTUs were phylogenetically linked to environmentally transmitted pathogens, usually by the fecal-oral route, it was expected that environmental features would determine individual infection patterns. Vegetation, for example, can influence host diet and pathogen prevalence, as the survival of some pathogens depends on the presence of organic matter [22]. There is also evidence of density-dependent transmission, as *rattus* abundance was an important feature. The cause for this can be suggested that higher density increases the probability of hosts encountering each other's feces (greater fecal contamination per unit area) [23]. Village distance, another important feature, might correlate with factors affecting individual behavior, leading to changing social dynamics, which has been shown to influence fecal-oral transmission [23]. Village distance also represents a habitat fragmentation gradient, which has been found to affect transmission [24]. By altering resource distribution and movement patterns, habitat fragmentation might lead to resource concentration, enhancing environmental contamination in high-use areas.

In conclusion, my results provide evidence for the importance of multilevel features in determining individual roles in the host-pathogen network. As roles in the network were used as a proxy for similar infection patterns, this supports all hypotheses. Additionally, this finding corresponds with many previous studies mentioned before. Further investigation can be done to reveal the effect of each feature on infection patterns. First, I should try to understand the ecological meaning of each SBM-generated group, i.e., what infection pattern it represents. Second, I should analyze how the properties vary among the groups, specifically focusing on the most important features (for example,

if heavier individuals tend to be in a specific group). After this analysis, we can link feature variance to infection patterns and draw more directed conclusions, providing a solid basis for future studies.

References

1. Tosam, M. J., Ambe, J. R. & Chi, P. C. in *Socio-cultural Dimensions of Emerging Infectious Diseases in Africa: An Indigenous Response to Deadly Epidemics* (eds Tangwa, G. B., Abayomi, A., Ujewe, S. J. & Munung, N. S.) 243–253 (Springer International Publishing, Cham, 2019).
2. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451,** 990–993 (2008).
3. Santibañez, S. *et al.* Strengthening Rural States' Capacity to Prepare for and Respond to Emerging Infectious Diseases, 2013-2015. *Southern Medical Journal* **112,** 101–105 (2019).
4. Morand, S. *et al.* Global parasite and Rattus rodent invasions: The consequences for rodent-borne diseases. *Integrative Zoology* **10,** 409–423 (2015).
5. Bublitz, D. C. *et al.* Epidemiology of pathogenic enterobacteria in humans, livestock, and peridomestic rodents in rural Madagascar. *PLoS One* **9,** e101456 (2014).
6. VanderWaal, K. L. & Ezenwa, V. O. Heterogeneity in pathogen transmission: mechanisms and methodology. *Functional Ecology* **30,** 1606–1622 (2016).
7. Runghen, R., Poulin, R., Monlleó-Borrull, C. & Llopis-Belenguer, C. Network analysis: ten years shining light on host-parasite interactions. *Trends in Parasitology* **37,** 445–455 (2021).
8. Paull, S. H. *et al.* From superspreaders to disease hotspots: linking transmission across hosts and space. *Frontiers in Ecology and the Environment* **10,** 75–82 (2012).
9. Schmidt, T. S. B., Matias Rodrigues, J. F. & von Mering, C. Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. *PLoS Computational Biology* **10,** e1003594 (2014).
10. Allesina, S. & Pascual, M. Food web models: a plea for groups. *Ecology Letters* **12,** 652–662 (2009).
11. Tymoteusz, M. *et al.* XGBOOST IN ENVIRONMENTAL ECOLOGY: A POWERFUL TOOL FOR SUSTAINABLE INSIGHTS. *Gospodarka Delo,* 163–170 (2023).
12. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2023).
13. Posit Team. *RStudio: Integrated Development Environment for R* Posit, PBC (Boston, MA, 2023).
14. Leger, J. B. Blockmodels: A R-package for estimating in Latent Block Model and Stochastic Block Model, with various probability functions, with or without covariates. *arXiv preprint arXiv:1602.07587* (2016).
15. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, USA, 2016), 785–794.
16. Costa, F. *et al.* Infections by Leptospira interrogans, Seoul virus, and Bartonella spp. among Norway rats (Rattus norvegicus) from the urban slum environment in Brazil. *Vector Borne and Zoonotic Diseases* **14,** 33–40 (2014).
17. Desvars-Larrive, A. *et al.* Prevalence and risk factors of Leptospira infection in urban brown rats (Rattus norvegicus), Vienna, Austria. *Urban Ecosystems* **23,** 775–784 (2020).
18. Grandón-Ojeda, A. *et al.* Patterns of Gastrointestinal Helminth Infections in Rattus rattus, Rattus norvegicus, and Mus musculus in Chile. *Frontiers in Veterinary Science* **9,** 929208 (2022).

19. Van Der Mescht, L., Le Roux, P. C. & Matthee, S. Remnant fragments within an agricultural matrix enhance conditions for a rodent host and its fleas. *Parasitology* **140,** 368–377 (2013).

20. Sánchez, C. A. *et al.* On the relationship between body condition and parasite infection in wildlife: a review and meta-analysis. *Ecology Letters* **21,** 1869–1884 (2018).

21. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).

22. Gerba, C. P. in *Environmental Microbiology (Second Edition)* (eds Maier, R. M., Pepper, I. L. & Gerba, C. P.) 445–484 (Academic Press, San Diego, 2009).

23. Nunn, C. L., Thrall, P. H., Leendertz, F. H. & Boesch, C. The spread of fecally transmitted parasites in socially-structured populations. *PLoS One* **6,** e21677 (2011).

24. Bordes, F. *et al.* Habitat fragmentation alters the properties of a host-parasite network: rodents and their helminths in South-East Asia. *Journal of Animal Ecology* **84,** 1253–1263 (2015).