Phylogenetic Analysis of Plasmodium-Host Coevolution

Timothy J. Nessel, Dennis V. Lavrov

Iowa State University, Department of Department of Ecology, Evolution, and Organismal Biology (EEOB)
https://github.com/tnessel/EEOBFinalProject

Introduction

Malaria is an old and incredibly deadly disease that has plagued humanity for thousands of years, killing more than a million people per year in the modern age (1). The parasites responsible for this disease in a variety of hosts belong to several genera in the phylum Apicomplexa, and include *Haemoproteus, Parahaemoproteus, Hepa- tocystis,* and *Plasmodium* (2). As humans malaria is caused by *Plasmodium,* this genus stimulates the majority of scientific research of this group of parasites. This study will focus on this genus utilizing the wealth of genomic data already produced.

Plasmodium is a versatile and far reaching parasite genus, intrinsic to many arid and tropical biomes around the world. The lifecycle of the parasite includes vertebrate host infection by sporozoites via an insect vector, and later insect infection by gametocytes via the vertebrate. *Plasmodium* infects vertebrate vectors including humans, non-human primates, rodents, reptiles, and birds, and insect vectors including sandflies and many genera of mosquitos, the best studied of these being *Anopheles* (CDC).

Plasmodium and its hosts' closely intertwined survival drives their coevolution, and this relationship is useful when uncovering the evolutionary history of this genus. It has been hypothesized that such parasite-host coevolution follows Fahrenholz's rule: when coevolution of two species occurs, their phylogenies will mirror each other (3). We seek to test this rule in regard to *Plasmodium* and its vertebrate hosts using diverse sampling and modern computational approaches in phylogenetics. Answering this question could lead to a better understanding of the history and future pathogenesis of Apicomplexa.

Phylogenetic analysis of *Plasmodium* has its own unique obstacles. The genetic loci used for analyzing parsimony in *Plasmodium* were commonly rRNA coding genes before it was discovered that the parasite has multiple rRNA genes that are variably expressed throughout its life cycle and subject to varied selection pressures. The cytochrome b loci is used in this analysis to circumvent this obstacle (5). This study was also faced with the challenge of lack of differentiation between *Plasmodium* species and its insect vector. *Plasmodium* species parasitize many of the same insect species leaving few unique host-parasite relationships to draw inferences from, and comparably sparse genomic data has been sampled from many insect vectors. Due to the limits of our analysis, we decided to confine this study to the coevolution of *Plasmodium* and its vertebrate hosts.

Methods

Sample Selection

All samples were accessed through

GenBank. 13 *Plasmodium* samples were chosen to represent the genus across a large range of vertebrate hosts and geographic region. 8 vertebrate hosts were chosen for the coevolution comparison (6). In the event that a valid and consistent sequence could not be found, a sample from the closest relative of the vertebrate host was used (See **Table 1**). Outgroup sequences were chosen from a well-studied amphibian to serve as a closely related species outside of Amniota, in which all malaria vertebrate hosts belong. Outgroup sequences for *Plasmodium* were chosen from a *Leucocytozoon* species due to previously published success (4). Cytochrome b and rRNA 18S subunit sequences were chosen as the genetic loci for *Plasmodium* and the vertebrate hosts respectively, due to consistent rates of evolution.

Plasmodium sp.	**Vertebrate sp.**	**Region**
P. falciparum	Humans	Global Tropics
P. vivax	Humans	Global Tropics
P knowlesi	Old world monkeys (Macaca mulatta)	Asia
P. yoelii	Thamnomys rutilans (Mus Musculus)	C. Africa
P. chabaudi	Thamnomys rutilans (Mus Musculus)	C. Africa
P. berghei	Grammomys surdaster (Mus Musculus)	C. Africa
P. vinckei	Grammomys surdaster (Mus Musculus)	C. Africa
P. atheruri	Atherurus africanus[x]	C. Africa
P. mexicanum	Sceloporus occidentalis (Sceloporus undulates)	C. America
P. floridense	Anolis oculatus (Anolis jubar)	C. America
P. giganteum	Agama agama (Uromastyx aegyptia)	C. Africa
P. gallinaceum	Gallus gallus	Asia
P. relictum	Emberiza hortulana (Emberiza variabilis)	Asia – Europe
Leucocytozoon sp.	Xenopus tropicalis	**OUTGROUP**

Table 1 *Plasmodium* and vertebrate samples for the analysis are listed along with their geographic region. A suitable sequence for *Atherurus africanus* could not be found.

Phylogenetic Analysis

Datasets were aligned with mafft v7.245 (7) using L-INS-I: iterative refinement method (<16) with LOCAL pairwise alignment information. Aligned sequences were then converted to nexus format with PAUP v4.0a (8).

Phylogenetic analysis and tree construction were performed using Bayesian analysis utilizing Markov Chain Monte Carlo. Analysis was performed with MrBayes v3.2 on the *Plasmodium* and host data separately (9). The model used was GTR+ Γ, assuming that the entire cytochrome b and 18S rRNA loci were affected homogenous across their respective sites. Analysis was run with (XYZ) ngen=300000 printfreq=100 samplefreq=100, and all standard deviations of split frequencies were acceptably low. Unpartioned and codon-partioned models were compared by Bayes Factor comparison of marginal likelihoods in stepping stone sampling. Codon-partioned models were chosen for analysis, but the Bayes Factor comparison did not differentiate strong support in the model comparison.

Coevolution Analysis

The Bayesian generated trees were analyzed with Jane, a tool that uses a heuristic approaches with genetic and fitness algorithms to generate dynamically sorted solutions in coevolution analysis (10). Analysis of the parasite and host trees was performed with the following costs for coevolution events: cospeciation = 0, duplication = 1, loss = 1, failure to diverge = 1, and duplication with host switch = 2. P. atheruri was not included in the Jane analysis as its paired host was missing in the dataset. The analysis was run over a simulated 500 generations with a population size of 5,000 per developer recommendations (11). Jane was also used for analysis into region specific parasite-host evolution and figure representation.

Results

Parasite and host consensus trees were generated separately through Bayesian analysis (See **Figure 1**).

The coevolution analysis with Jane predicted cospeciation in 6 host-parasite pairs with 3 duplications with host switching and one loss (See **Figure 2**). Regional analysis of this tree could not be performed due to software malfunction, but a representation of the regional comparison in the tree was generated (See **Figure 3**).

Discussion

The coevolution analysis shows majority cospeciation between *Plasmodium* species and its vectors, though duplication and host switching are present in significant form. It is possible that the Bayesian analysis of lizard and bird infecting *Plasmodium* species was flawed, as the topology of these species seems to lose accuracy compared to that of the mammalian infecting *Plasmodium*. This disruption in the topology would carry through the Jane analysis, and indeed many of the affected species are suspected for dramatic events in coevolution. While this may invalidate this study for use in future inquiry, the analysis can be still be used for a valuable demonstration on Fahrenholz's rule. We show that coevolution is a powerful evolutionary force as seen through the mirroring of topologies. Coevolution analysis can be a valuable tool in sorting through the topologies of rapidly changing parasite genealogies.

More study is needed with regards to the other factors relevant in this coevolution, including the role of the insect host, the impact of geographic region, and variance in species evolutionary rates.

Figures to proceed:

Figure 1
The Plasmodium consensus tree (top) and the vertebrate host consensus tree (bottom).

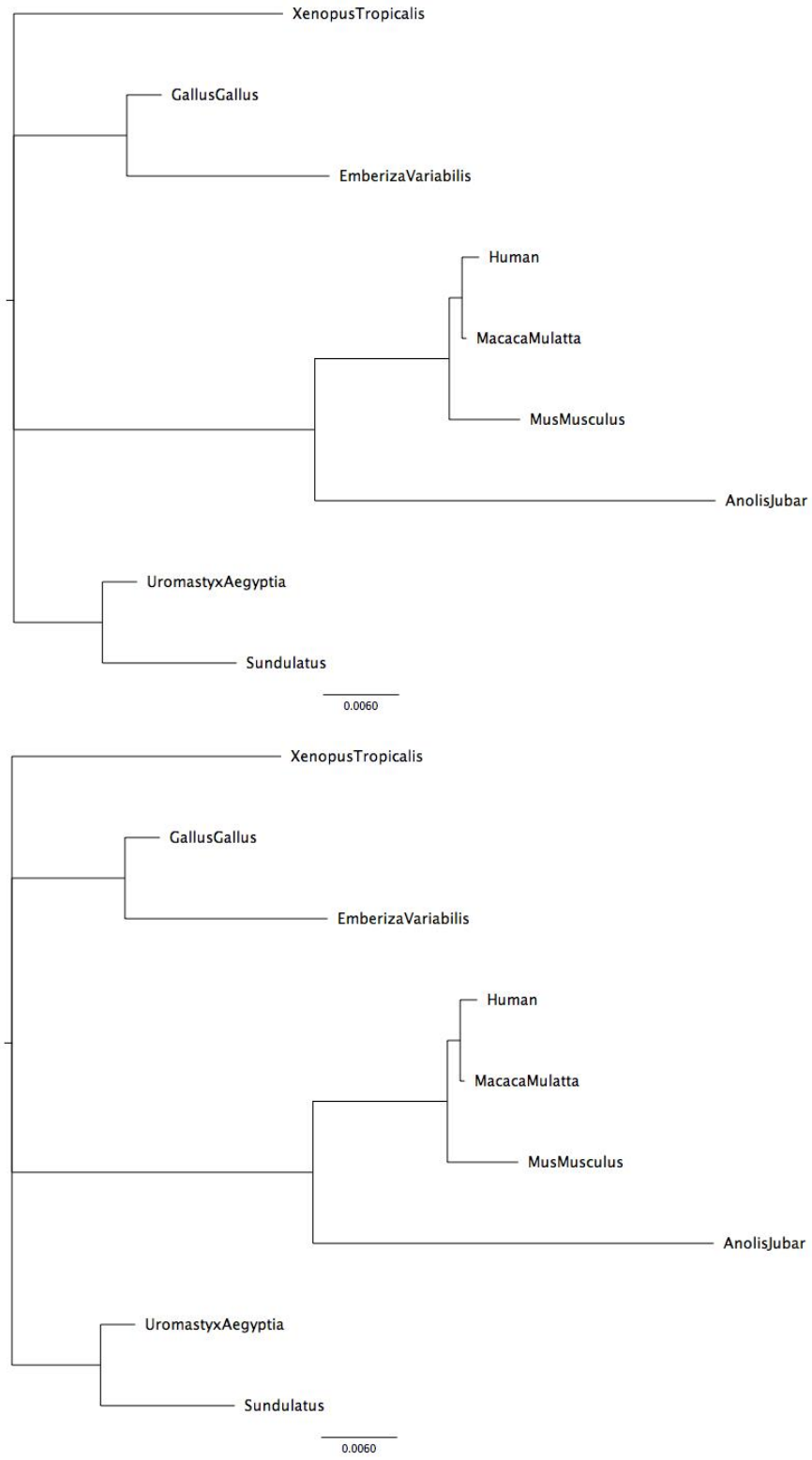


Figure 2
Tree showing *Plasmodium* and host coevolution analysis
Legend shown bottom

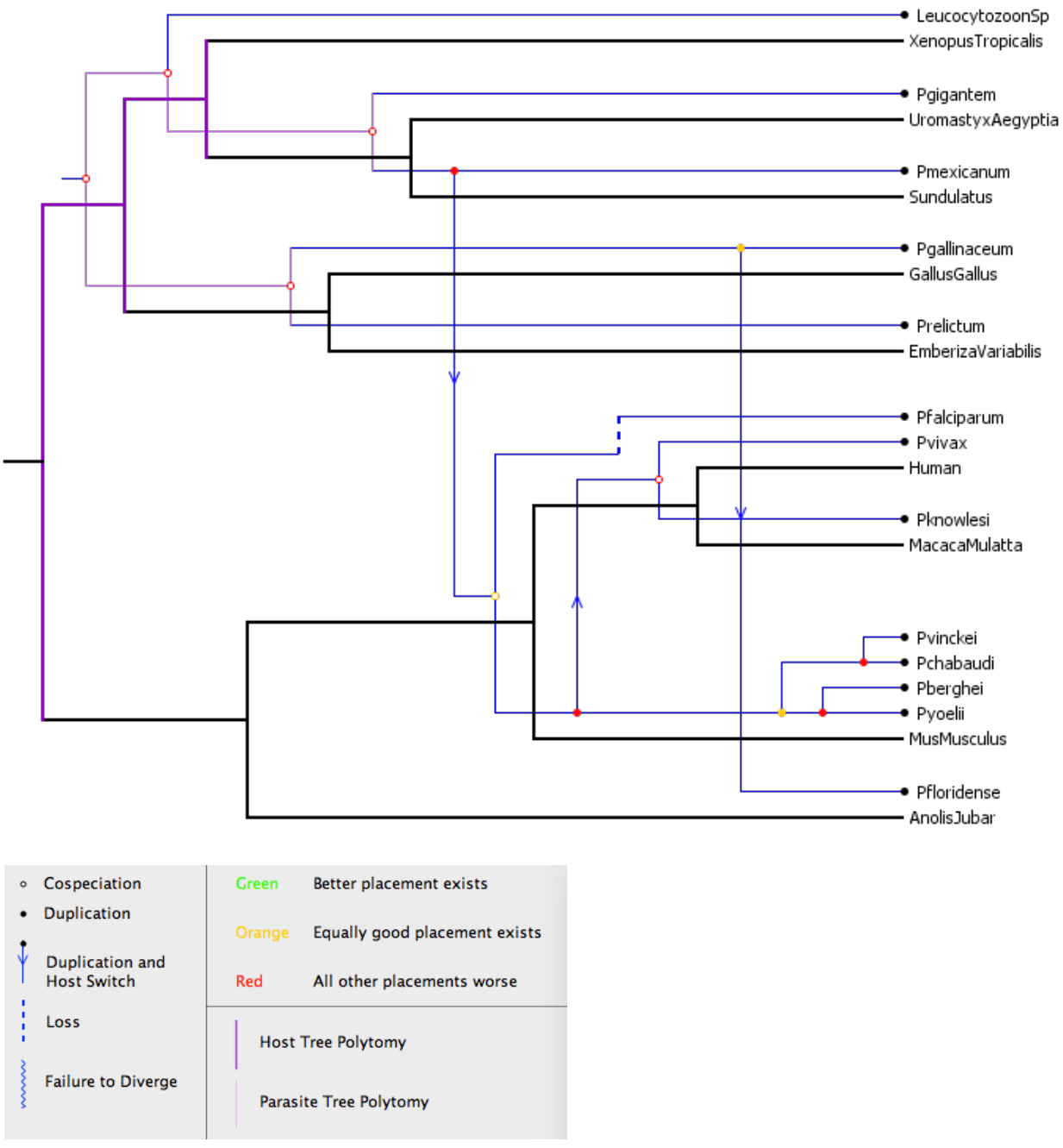
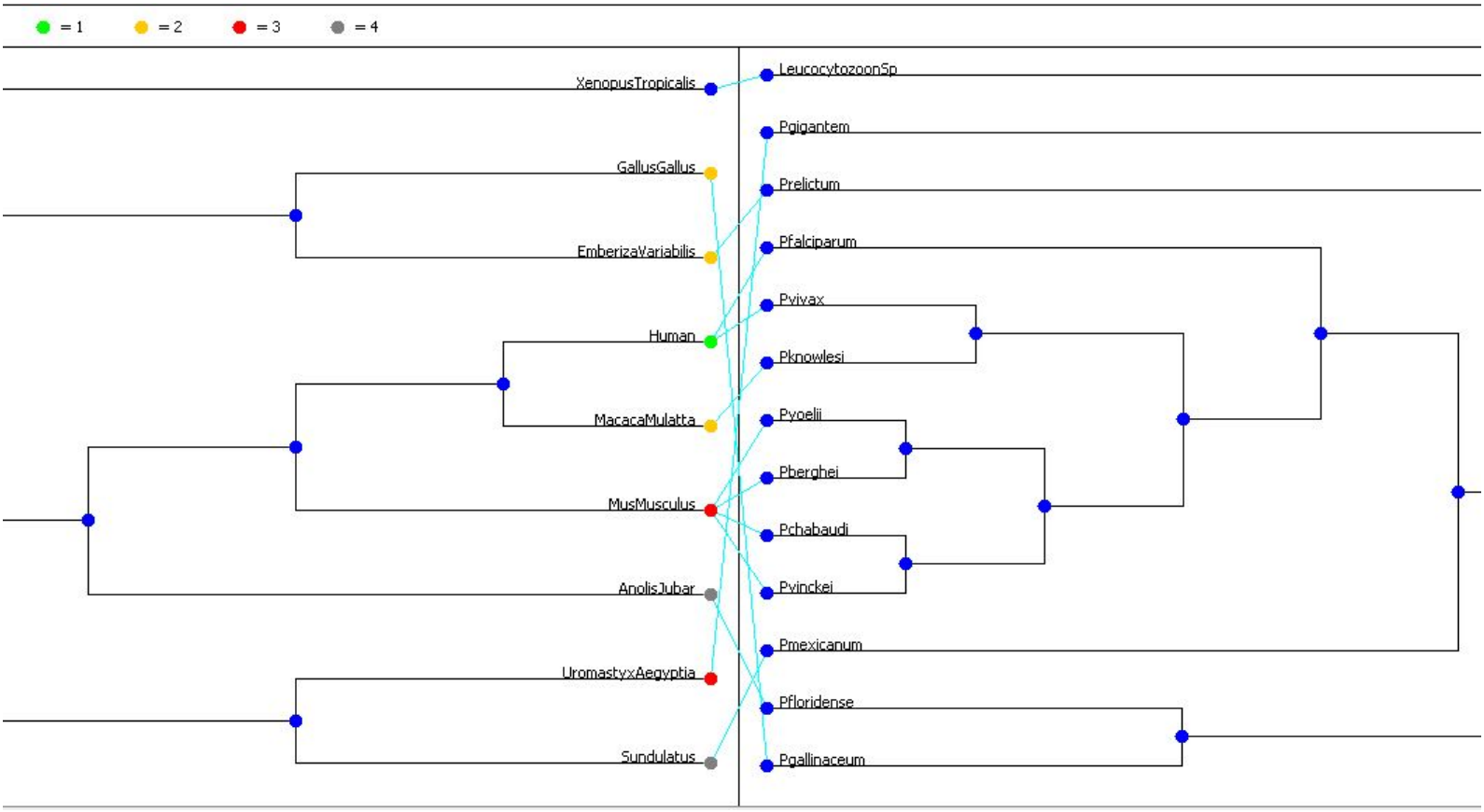


Figure 3
Host and *Plasmodium* trees shown side by side, parasite host relationship shown by cyan connecting lines.
Host nodes colored by region.
Legend:
Green = Global Tropics
Yellow = Asia Tropics
Red = Africa
Grey = Central America



References

1. Carter R, Mendis KN. Evolutionary and Historical Aspects of the Burden of Malaria. *Clinical Microbiology Reviews*. 2002;15(4):564-594. doi:10.1128/CMR.15.4.564-594.2002.
2. Ellen S. Martinsen, Susan L. Perkins, Jos J. Schall, A three-genome phylogeny of malaria parasites.. , *Molecular Phylogenetics and Evolution,Volume 47*, Issue 1, 2008, Pages 261-273, https://doi.org/10.1016/j.ympev.2007.11.012.
3. Hafner MS, Nadler SA. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*. 1988 Mar 17;332(6161):258-9. PubMed PMID: 3347269.
4. Perkins, S.L., Schall, J.J., 2002. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. J. Parasitol. 88, 972–978.
5. Corredor V, Enea V. 1993. Plasmodial ribosomal RNA as phylogenetic probe: a cautionary note. Mol. Biol. Evol. 10:924-926.
6. R. S. Bray, P. C. C. Garnham. Anopheles as Vectors of Animal Malaria Parasites. London School of Hygiene and Tropical Medicine, London, England
7. MAFFT v7.245 (2015/07/22) Copyright (c) 2015 Kazutaka Katoh. MBE 30:772-780 (2013), NAR 30:3059-3066 (2002) http://mafft.cbrc.jp/alignment/software/
8. P A U P v4.0a (build 161) Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony. (*and Other Methods).
9. Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755. Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
10. Ran Libeskind-Hadas. Jane. Harvey Mudd College. https://www.cs.hmc.edu/~hadas/jane
11. Conow, Chris et al. "Jane: A New Tool for the Cophylogeny Reconstruction Problem." *Algorithms for Molecular Biology : AMB* 5 (2010): 16. *PMC*. Web. 20 Apr. 2018.