Phylogenetic Analysis of Plasmodium 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). Our focus will be on *Plasmodium*, which affects humans and other mammals.

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. Species of *Plasmodium* infect humans, non-human primates, rodents, reptiles, and birds, while insect vectors include sandflies and many genera of mosquitos, the best studied of these being *Anopheles* (CDC).

Plasmodium and its hosts' closely intertwined fitness 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 Law, in that their speciation is synchronous (3). This would result in similar topologies in phylogenetic trees while considering other driving evolutionary factors. We seek to test this hypothesis 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 been faced with many challenges in the past decades. When considering the evolutionary factors on any molecular parasite, it is necessary to parse what genetic loci are affected by the host immune system and membrane dynamics, as these sites can have varied evolutionary rates. The unique biology of the organism can also invalidate a phylogenetic analysis. The genetic locus used for analyzing parsimony in *Plasmodium* was commonly rRNA before it was discovered that the parasite has multiple rRNA loci that are variably expressed throughout its life cycle (5). In addition, a valid outgroup has been debated between the genera *Toxoplasma* and *Leucocytozoon* (4). Here we take these challenges into consideration in making a valid phylogenetic inference.

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**). Outgroups were chosen from amphibian and *Leucocytozoon* species sequences (4). Cytochrome b and rRNA 18S subunit sequences were chosen as the genetic loci for *Plasmodium* and the vertebrate hosts respectively.

Plasmodium sp.	**Vertebrate sp.**	**Region**
P. falciparum	Humans	Tropics
P. vivax	Humans	Brazil
P knowlesi	Old world monkeys (Macaca mulatta)	Malaysia
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)	Dominica
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**

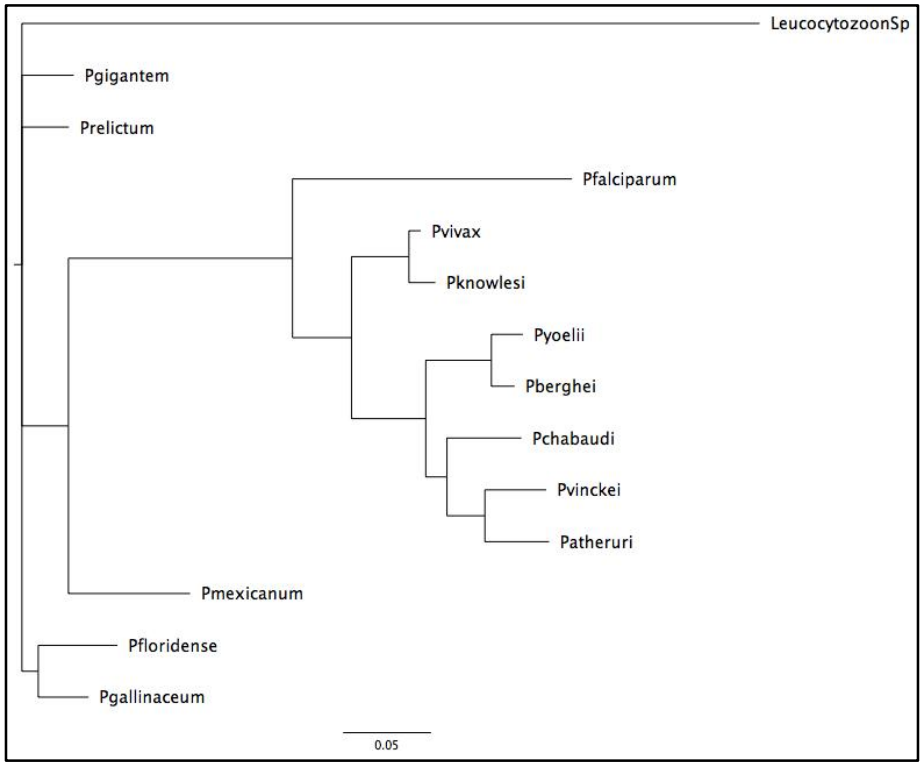
Phylogenetic Analysis

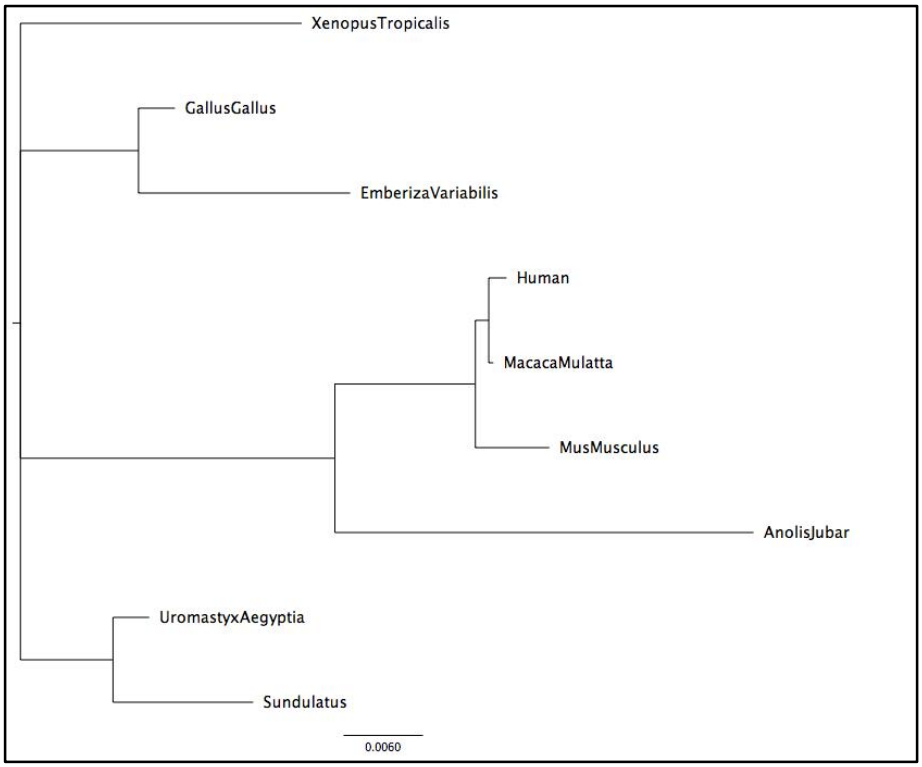
Datasets were aligned with mafft v7.245 (7) using L-INS-I: iterative refinement method (<16) with LOCAL pairwise alignment information. Aligned sequences were then converted to nexus format with PAUP v4.0a (8).

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. Consensus trees were generated for both analyses (See **Figure 1**).

Table 1 *Plasmodium* and vertebrate samples for the analysis are listed. Samples that could not be derived were taken from the species in parenthesis under the intended host. A suitable sequence for *Atherurus africanus* could not be found.

Figure 2
The *Plasmodium* consensus tree (top) and the vertebrate host consensus tree (bottom).





Coevolution Analysis

The Bayesian generated trees were then analyzed with Jane, a tool that uses a heuristic approaches with genetic and fitness algorithms to generate dynamically sorted solutions in coevolution analysis (10). This software was used to create a parasite-host tree (See **Figure 2**)

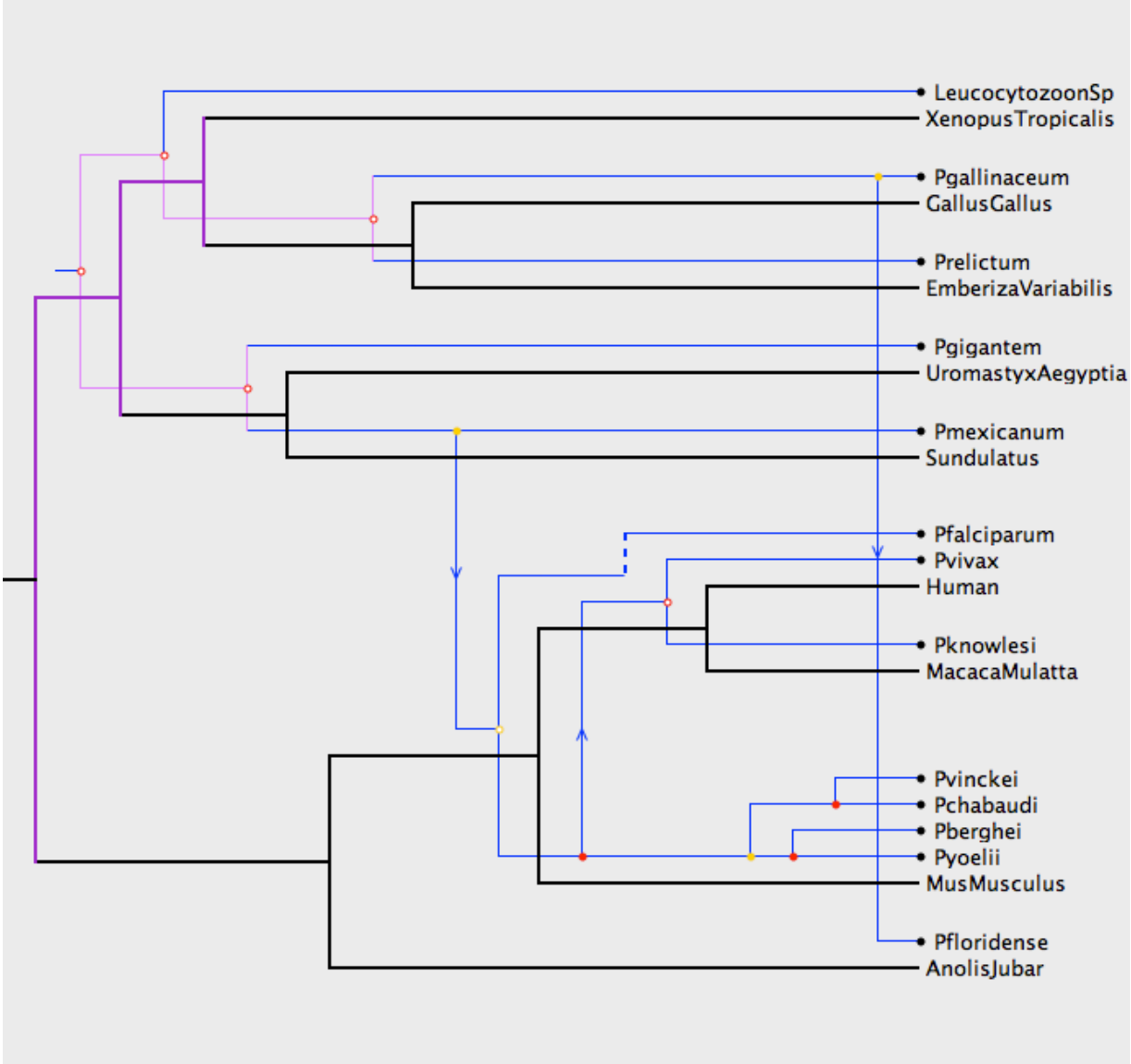


Figure 3

The Methods and Results section need to be organized, and small additions to the Bayesian and Jane analysis will be made. The Discussion section will also be finished