HHS Public Access

Author manuscript

J Mol Evol. Author manuscript; available in PMC 2019 March 22.

Published in final edited form as:

J Mol Evol. 2018 December ; 86(9): 646–654. doi:10.1007/s00239-018-9875-3.

Delineation of the Genera *Haemoproteus* and *Plasmodium* Using RNA-Seq and Multi-gene Phylogenetics

Jasper Toscani Field[1], **Josh Weinberg**[1], **Staffan Bensch**[2], **Nubia E. Matta**[3], **Gediminas Valkiūnas**[4], and **Ravinder N. M. Sehgal**[1]

[1]Department of Biology, San Francisco State University, 1700 Holloway Ave, San Francisco, CA 94132, USA

[2]Department of Biology, Lund University, Ecology Building, 223 62 Lund, Sweden

[3]Sede Bogotá, Facultad de Ciencias, Departamento de Biología, Grupo de Investigación Caracterización genética e inmunología, Universidad Nacional de Colombia, Carrera 30 No. 45-03, Bogotá 111321, Colombia

[4]Nature Research Centre, Akademijos 2, 08412 Vilnius 21, Lithuania

Abstract

Members of the order Haemosporida are protist parasites that infect mammals, reptiles and birds. This group includes the causal agents of malaria, *Plasmodium* parasites, the genera *Leucocytozoon* and *Fallisia*, as well as the species rich genus *Haemoproteus* with its two subgenera *Haemoproteus* and *Parahaemoproteus*. Some species of *Haemoproteus* cause severe disease in avian hosts, and these parasites display high levels of diversity worldwide. This diversity emphasizes the need for accurate evolutionary information. Most molecular studies of wildlife haemosporidians use a bar coding approach by sequencing a fragment of the mitochondrial cytochrome *b* gene. This method is efficient at differentiating parasite lineages but insufficient for accurate phylogenetic inferences in highly diverse taxa such as haemosporidians. Recent studies have utilized multiple mitochondrial genes (cyt *b*, cox1 and cox3), sometimes combined with a few apicoplast and nuclear genes. These studies have been highly successful with one notable exception: the evolutionary relationships of the genus *Haemoproteus* remain unresolved. Here we describe the transcriptome of *Haemoproteus columbae* and investigate its phylogenetic position recovered from a multi-gene dataset (600 genes). This genomic approach restricts the taxon sampling to 18 species of apicomplexan parasites. We employed Bayesian inference and maximum likelihood methods of phylogenetic analyses and found *H. columbae* and a representative from the subgenus *Parahaemoproteus* to be sister taxa. This result strengthens the hypothesis of genus *Haemoproteus* being monophyletic; however, resolving this question will require sequences of orthologs from, in particular, representatives of *Leucocytozoon* species.

Jasper Toscani Field jtoscanifield@ucmerced.edu.

Accession numbers: *H. columbae* raw reads GenBank ID: SAMN06899305. *H. columbae* transcriptome assembly GenBank ID: GGWD00000000.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00239-018-9875-3) contains supplementary material, which is available to authorized users.

Keywords

Haemoproteus; Phylogenetics; Transcriptome; Avian parasitology

Introduction

Haemosporidian parasites are notable for their diversity and cosmopolitan distribution. These globally distributed parasites are members of the phylum *Apicomplexa* (Garnham 1966; Valkiūnas 2005). Host species include mammals, lizards and birds (Ricklefs and Fallon 2002; Valkiūnas 2005), and several haemosporidian species have been reported in amphibians and fish (Valkiūnas 2005). Genera infecting avian hosts include *Plasmodium, Fallisia, Leucocytozoon* and *Haemoproteus*. The genus *Haemoproteus* currently contains two subgenera: *Haemoproteus* and *Parahaemoproteus*. Members of the subgenus *Haemoproteus* are vectored by hippoboscid flies (Hippoboscidae) and primarily infect Columbiformes such as pigeons and doves but have also been found in marine birds belonging to Pelecaniformes and Charadriiformes (Levin et al. 2012). *Parahaemoproteus* parasites are spread by biting midges (Ceratopogonidae) and infect many bird species, in particular among passerines, raptors and waterfowl. Members of both subgenera produce hemozoin and undergo merogony in cells of fixed endothelial tissues with erythrocyte invasion producing only gametocytes (Valkiūnas 2005; Levin et al. 2012; Palinauskas et al. 2013). The avian haemosporidian sequence database MalAvi (Version 3.2.2, April 4th, 2018) shows 1148 cytochrome *b* lineages of *Haemoproteus*. Of these lineages, 97.38% belong to the subgenus *Parahaemoproteus* and 30 belong to subgenus *Haemoproteus*. Of these 30 lineages, five have been morphologically described as belonging to *Haemoproteus columbae. H. columbae* can cause obstructions in pulmonary, myocardial and hepatic tissue of its hosts as seen in an examination of a deceased Bleeding Heart Dove (*Gallicolumba crinigera)* (Earlé et al. 1993). Marked enlargement of spleen and liver has been reported during heavy *H. columbae* parasitemia, which might exceed 50% of parasitized red blood cells and cause anemia in rock pigeons (Garnham 1966; Valkiūnas 2005). This parasite is globally distributed and has become one of the most well studied *Haemoproteus* species (Valkiūnas 2005; Earlé et al. 1993; Adriano and Cordeiro 2001; Waldenström et al. 2002; Santiago-Alarcón et al. 2010; Waite et al. 2012, 2014).

Focusing on parasite diversity, there are hundreds of morphologically defined species within the genera *Haemoproteus, Leucocytozoon* and *Plasmodium* and they are further diversified with multiple genetically unique lineages (Ricklefs and Fallon 2002; Valkiūnas 2005; Martinsen et al. 2008; Santiago-Alarcón et al. 2010; Borner et al. 2016). The effects of each species on its host can be extremely variable, with some species causing severe anemia and death while others persist in barely detectable chronic infections (Palinauskas et al. 2008; Dimitrov et al. 2015; Valkiūnas and Iezhova 2017). *Plasmodium* parasites have traditionally been the primary focus of research due to their medical importance (Gardner et al. 2002; Bozdech et al. 2003; Otto et al. 2010; Liu et al. 2010; Mbengue et al. 2015). In this aspect, phylogenetics is key in determining how we research these pathogens. Parasites sharing a recent common ancestor offer alternative avenues of researching and understanding species infecting humans (Grech et al. 2006; Lefèvre et al. 2007; Parker et al. 2015; Neher et al.

2016). Lefèvre et al. (2007) discussed the use of a mouse-parasite model system in Grech et al. (2006). Lefèvre et al. (2007) states that while limitations of comparing human-parasite interactions to model systems must be acknowledged, systems as *Plasmodium chabaudi* in mice can be used to understand virulence variation (Lefèvre et al. 2007). Lauron et al. (2014) describes the presence of the *Apical Membrane Antigen 1* and *Rhoptry Neck Protein 2* genes (*ama1* and *ron2*, respectively) in *Plasmodium gallinaceum* transcriptomics and the implications of finding these conserved cell invasion mechanisms in avian haemosporidians (Lauron et al. 2014) With such discoveries in mind, the variation in haemosporidian species diversity coupled with potential speciation of novel pathogenic lineages highlights the necessity of well-resolved phylogenetic trees.

The majority of avian haemosporidian phylogenetic studies are based on single-gene analyses of the cytochrome *b* gene (cyt *b*) (Hellgren et al. 2004, 2007; Valkiūnas 2005; Valkiūnas et al. 2008a, b; Martinez-de la Puente et al. 2011; Carlson et al. 2013; Jasper et al. 2014; Outlaw and Ricklefs 2014; Palinauskas et al. 2015). While valuable, this kind of analysis lacks the depth necessary to answer important questions regarding parasite relationships (Hellgren et al. 2013; Outlaw and Ricklefs 2014; Borner et al. 2016; Bensch et al. 2016). Unresolved phylogenetic relationships can be addressed with multi-gene datasets (Borner et al. 2016). The primary limitation of such methods is that genomic and transcriptomic datasets, required for phylogenomic analyses, are particularly difficult to obtain from haemosporidian parasites (Lauron et al. 2014; Bensch et al. 2016; Videvall et al. 2017; Böehme et al. 2018; Videvall 2018). Recent multi-gene phylogenetic analyses of the avian apicomplexans performed with multiple statistical methods produced well-resolved phylogenies with one exception in the case of haemosporidian parasites, i.e., the placement of subgenus *Haemoproteus* in relation to subgenus *Parahaemoproteus* and genus *Plasmodium* (Martinsen et al. 2008; Bensch et al. 2016; Borner et al. 2016; Pacheco et al. 2018). An early multi-gene study by Martinsen et al. (2008) showed the genus *Haemoproteus* as paraphyletic, with *Plasmodium* forming a sister relationship with subgenus *Parahaemoproteus*. In work by Borner et al. (2016), depending on the evolutionary model, taxa and genes used for analyses, the subgenus *Haemoproteus*, represented by *H. columbae*, formed a sister relationship with either subgenus *Parahaemoproteus* or with the genus *Plasmodium*. Furthermore, the relationships of the *Haemoproteus* subgenera have been open to debate, with one subgenus forming a sister relationship with *Plasmodium* as in Fig. 1a (Martinsen et al. 2008; Santiago-Alarcón et al. 2010; Martinez-de la Puente et al. 2011; Pacheco et al. 2018) or with both subgenera forming a monophyletic clade as in Fig. 1b (Valkiūnas et al. 2010, 2016; Levin et al. 2012; Carlson et al. 2013; Palinauskas et al. 2015; Lutz et al. 2016). While the dataset used in Borner et al. (2016) provided strong resolution for all other clades, it is interesting that *H. columbae*'s placement could not be resolved. Currently available sequencing technology provides an opportunity to address this issue but certain challenges must be acknowledged.

Producing large haemosporidian sequence databases (genomes or transcriptomes) requires knowledge of both parasite genome structure and life cycle (Lauron et al. 2014; Bensch et al. 2016; Böehme et al. 2018). Researchers pursuing genome sequencing must consider that current technology will provide sequences for both host and parasite. Ratios of host-to-parasite DNA generally favor the host by wide margins due to avian host erythrocyte

nucleation (Auburn et al. 2011; Oyola et al. 2012; Bensch et al. 2016; Böehme et al. 2018). The presence of nuclei in avian erythrocytes hinders the parasite's whole-genome sequencing, although recently next generation sequencing and sophisticated parasite DNA isolation protocols have resulted in the sequencing of some avian haemosporidian genomes (Bensch et al. 2016; Böehme et al. 2018) In addition to parasite-host genome ratios, transcriptome assembly requires selecting either mRNA or total organismal RNA for preferential sequencing (Lauron et al. 2014; Videvall et al. 2017). Here we describe the processing of RNA-seq data to produce the first transcriptome of *H. columbae*. We also implement a large gene data set gathered from the transcriptome to attempt to resolve the *Haemoproteus* phylogenetic relationships described in Fig. 1. Included in the data set are sequences from 9 other haemosporidian parasites and 8 more distantly related apicomplexan parasites as detailed in Bensch et al. (2016).

Methods

RNA Collection and Extraction

A Rock Pigeon, *Columba livia domestica*, was captured at the campus of Universidad Nacional de Colombia located at 2560 m above sea level in the city of Bogotá, Colombia. Infection was verified by examining Giemsa stained slides. Whole blood was collected and stored in Trizol® LS reagent (Invitrogen, Grand Island, NY, USA). Samples were imported to San Francisco State University (USDA veterinary permit 114165). Giemsa slides were re-examined to verify infection by a single parasite lineage. RNA was extracted using a Trizol® LS (Invitrogen, Grand Island, NY, USA) extraction protocol in which phase-lock gel tubs were used to separate RNA from DNA and proteins in aqueous phases. Isopropyl ethanol and a high-salt solution were used to precipitate suspended RNA. After re-suspension, the RNA was treated with Ambion® TurboDNAse™ before a size-separation step using Agencourt® RNAClean® XP (Agencourt Bioscience Corporation, Beverly, Massachusetts, USA) beads were applied to remove degraded RNA before re-suspension in DEPC treated water. Additional verification of parasite species was obtained by polymerase chain reaction of the cytochrome *b* gene (Hellgren et al. 2004).

Library Preparation and Sequencing

Library preparation was performed at the University of California, Berkeley Functional Genomics Laboratory. PolyA selection was used for mRNA enrichment via Invitrogen Dynabeads® mRNA Direct™ kit (Life Technologies, Carlsbad, CA, USA). Next, the Ovation® RNA-seq system (NuGEN Technologies, Inc, San Carlos, CA, USA) was used for cDNA synthesis and SPIA amplification. A S220 Focused-Ultrasonicator (Covaris, inc., Woburn, Massachusetts, USA) was used to fragment the cDNA which was then cleaned and concentrated using the MinElute® PCR Purification kit (Qiagen, Valencia, CA, USA). The sequencing library was prepared on an Apollo 324TM™ (Wafergen Biosystems, inc, Fremont, CA, USA) with PrepX ILM 32i DNA Library Kit (Wafergen Biosystems, inc, Fremont, CA, USA) and nine (9) cycles of polymerase chain reaction for library enrichment. Libraries were sequenced on an Illumina Hiseq-4000™ (Illumina, inc, San Diego, CA, USA) with read size selection of 100 base-pairs, paired-end. Raw reads were deposited to the NCBI biosample sequence archive (GenBank Accession No. SAMN06899305).

Raw Sequence Data Processing and Host Separation

The resulting reads from the sequencing described above were collected, and quality was assessed with the FastQC program (Babraham and Bioinformatics 2017). Read quality scores were high enough that very little quality trimming was necessary. Trimming was performed with BBDUK (Joint Genome Institute 2017), a quality filter and adapter removal program, removing sequencing adapters only. The genome of sample host *C. livia* was downloaded from NCBI (Accession: GCA_001887795.1) (National Center for Biotechnology Information 2017). The filtered reads were mapped to the *C. livia* genome using HISAT2 with the "very sensitive" option (Kim et al. 2015). Reads mapping to the *C. livia* genome were not exported, and only unmapped reads were used in the rest of the pipeline. Read output fastq files were relayed to the Trinity (2.1.1) program for de novo contig assembly (Haas et al. 2013). The resulting contigs were clustered using CD-HIT-EST (Li and Godzik 2006) to merge clusters with 97% similarity. The genome of *Haemoproteus tartakovskyi* was download from the MalAvi database (Bensch et al. 2009). The clustered contigs were aligned to the *H. tartakovskyi* genome using BlastX (Altschul et al. 1990) and custom python scripts pulled the aligning sequences with significant matches (e-value: 1e-6). A second round of BlastN (Altschul et al. 1990) and separation was preformed to remove sequences matching to *C. livia* with an identity > = 90%. These successive removal steps ensured that the remaining contigs could be confidently associated with the *H. columbae* parasite. Transcriptome statistics (GC content, contig length, assembled bases and other standard assembly metrics) were calculated using Trinity's built in statistical assessment program and bash scripts (Table 1). This transcriptome shotgun assembly project has been deposited at DDBJ/ENA/GenBank under the accession GGWD00000000. The version described in this paper is the first version, GGWD01000000.

Dataset Alignment and Phylogenetic Analysis

The dataset included 600 protein coding sequences for 17 taxa. The species names and original sources of each taxa dataset are provided in supplementary file (1) taxa sequence datasets were originally collected from the following databases: PlasmoDB (Aurrecoechea et al. 2008), ToxoDB (Gajria et al. 2007), PiroplasmaDB (PiroplasmaDB 2014), CryptoDB (Puiu et al. 2004) and MalAvi (Bensch et al. 2009). The corresponding proteins were found in the transcriptome of *H. columbae* using TblastX (Altschul et al. 1990), and the resulting protein sequences were collected with custom python scripts and added to the Bensch dataset. Sequence alignment was performed with T-Coffee (Notredame et al. 2000) using default parameters. Gene IDs associated with each protein block are listed in supplementary file (2) Aligned protein files were filtered using Gblock (Castresana 2000) to remove gaps and regions of poor alignment. Gblock was run twice, once with default settings allowing no gaps in sequence products. The second Gblock run allowed gaps in alignments using the 50% allowance setting. Gblock alignments were imported to Geneious (version 7.1.0) for final sequence concatenation.

Model selection was preformed using Modelgenerator (version 85, Keane et al. 2006). Maximum likelihood analysis was performed on the concatenated gapped and ungapped datasets in RAxML with 100 bootstrapped phylogenetic trees (version 8.2.10, Stamatakis 2014) using an LG + G + F model. Bayesian analysis was performed with Mrbayes (version

3.2.6, Ronquist and Huelsenbeck 2003) using the LG + G + F evolutionary model. Mrbayes was run for 2 million generations, sampling every 100 generations before a burn in of 25%. Both analysis methods used *Cryptosporidium parvum* and *Cryptosporidium muris* as outgroups as detailed in Bensch et al. (2016). Individual gene trees were produced with RAxML for all 600 genes in the dataset, and a comparison of tree topologies matching the concatenated gapped dataset analysis was performed with the Sumtree program (Sukumaran and Holder 2010).

Results

Sample Parasitemia

The parasitemia was 1.68%, calculated as the number of different parasite stages of *H. columbae* in 10,000 red blood cells.

Read Processing and Transcriptome Assembly

Sequences produced by the Hiseq-4000 totaled 110,331,097 100 nucleotide (nt) paired-end reads (220,662,194 total) equaling 22 Gbp. The BBDUK program removed 84,000 reads as Illumina adapter sequences (0.00038% of total reads). The mapping of reads to the genome of *C. livia* using HISAT2 resulted in 102 million reads remaining unmapped (46% of total reads) that were used for assembly. The assembly resulted in 267,604 contigs. After CD-HIT-EST, 220,867 contigs remained after isoform clustering. The resulting sequences were mapped to the *H. tartakovskyi* genome using BlastX and contigs with significant e-value (< 1e−6) were separated with custom scripts. Only 26,781 contigs passed this filter. At this stage, the GC content was ~ 27% with an average contig length of 716 bp. The separate sequences were then mapped to *C. livia* genome using BlastN and any sequences with 90% identity to *C. livia* were removed. This led to 17,238 contigs passing filter with a GC of 17.78% and an average contig length of 769 bp. For all further purposes, this dataset is referred to as the *H. columbae* transcriptome.

Ortholog Clustering and Phylogenetic Analysis

Ortholog trimming with Gblock produced two distinct datasets. The first dataset consisted of 458 ortholog clusters. No missing data (sequence gaps) were permitted in this dataset. The second dataset contained 600 ortholog clusters but contained gaps in alignments. The discrepancy of 153 orthologs consisted of alignments where at least one species sequence did not align with the region selected for analysis. We kept both datasets for further analyses to check that the data selection did not bias the results. Both alignments are included as supplementary data 1 and 2.

Both maximum likelihood and Bayesian analyses resulted in identical topologies (Fig. 2). Strong support values were found at all nodes, and the recovered topology was identical with regard to the taxa analyzed in Bensch et al. (2016). Trees from all concatenated analyses show *H. columbae* and *H. tartakovskyi* forming a monophyletic clade. Analysis of single-gene tree topologies supported the concatenated dataset with 39% of the single-gene trees supporting the monophyletic relationship of *H. columbae* and *H. tartakovskyi*. ML support values increased slightly between the ungapped and gapped datasets. Bayesian inference was

performed on only the ungapped dataset to prevent biases based on some taxa containing more informative sequence data and fewer alignment gaps. While the number of generations was set to 2 million, analysis was stopped after 1,202,000 generations as the standard deviation of split frequencies had converged beyond 1e−6 after 105,000 generations. *H. columbae* and *H. tartakovskyi* again formed a monophyletic clade with strong posterior probability support.

Discussion

This study provides the first transcriptome data from a *Haemoproteus* parasite. This type of large genomic datasets improves the accuracy of evolutionary reconstructions. Haemosporidian genomic-scale data have also been instrumental in the discovery of many invasion genes in avian parasites as well as resolving the positions of mammalian and avian parasites (Martinez et al. 2013; Lauron et al. 2014; Videvall et al. 2017; Borner et al. 2016; Bensch et al. 2016). Transcriptomic approaches have been effective at characterizing *P. ashfordi* (Videvall et al. 2017), *P. gallinaceum* (Lauron et al. 2014), investigating the effects of parasitism on host gene expression (Videvall et al. 2015) and guiding genome assemblies (Böehme et al. 2018). The transcriptome of *H. columbae* opens new research opportunities for examining invasion gene variation between this and previously published subgenera (Lauron et al. 2014; Videvall et al. 2017). Additionally, sequence-based approaches to protein–protein interactions are feasible, such as ortholog based comparisons and phylogenetic mirror tree methods (Rao et al. 2014). It should be possible to utilize genomic datasets for either *P. ashfordi* or another avian parasite with well-characterized invasion genes, such as *P. gallinaceum* or *P. relictum*, to establish orthologous parasite proteins in *H. columbae* (Böehme et al. 2018). In a preliminary exploration of the transcriptome using Blast (Altschul et al. 1990) sequence matching, we discovered contigs matching portions of the *ama1* and *ron2* genes. Using this knowledge of ortholog data, numerous avian genomes may be used to infer host receptor proteins (Lee et al. 2008; Rao et al. 2014; Videvall et al. 2017).

Our study has added to the work of Bensch et al. (2016) by expanding it as a potential standard dataset for haemosporidian evolutionary analyses. While we must acknowledge that the taxon sampling in this dataset is low, we believe that it represents a firm foundation for future studies when genomes and transcriptomes will be available from other haemosporidian taxa. Here we also address the polytomy found in work by Borner et al. (2016) regarding the evolutionary relationships of *H. columbae* (Borner et al. 2016). The issue of the relationships of genus *Haemoproteus* requires significant literature review. Phylogenetic reconstructions in Martinsen et al. (2008), using a multi-gene analysis, showed *Parahaemoproteus* forming a sister relationship with *Plasmodium*, and the subgenus *Haemoproteus*, represented by *H. columbae*, forming a separate clade. A study by Santiago-Alarcón et al. (2010) describing the relationships of New World columbiform parasites focused on the phylogenetic relationships of subgenus *Haemoproteus* members. The author's results supported the sister relationship of *Parahaemoproteus* and *Plasmodium* while the subgenus *Haemoproteus*, including *H. columbae*, formed a separate clade. The authors performed analyses on two genes, the mitochondrial cyt *b* and the aplicoplast caseinolytic protease C (ClpC) gene. Despite examining the product of the two-gene

analyses, the recovered phylogenetic tree contained multiple polytomies (Santiago-Alarcón et al. 2010). Work by Valkiūnas et al. (2010) again supported a sister relationship of the subgenera *Haemoproteus* and *Parahaemoproteus*: this study focused on parasites primarily found in Columbiformes of the Galapagos Islands with most of the subgenus *Haemoproteus* being represented by *H. multipigmentatus* (Valkiūnas et al. 2010). Investigations by Martinez-de la Puente et al. (2011) found the subgenus *Parahaemoproteus* sister to *Plasmodium* and the subgenus *Haemoproteus* as a more distantly related clade. This study was also based only on cyt *b* analyses and is interesting due to the diversity of *Parahaemoproteus* members included in the study, while *Haemoproteus* was represented by only three lineages of *H. columbae* (Martinez-de la Puente et al. 2011). A recent study by Palinauskas et al. (2015) examined the differentiation of the cryptic parasite species *Plasmodium homocircumflexum* and *Plasmodium circumflexum* and included a *Haemoproteus* lineage for the phylogenetic analysis. This work showed the subgenus *Haemoproteus* and *Parahaemoproteus* as sister clades and *Plasmodium* as more distantly related (Palinauskas et al. 2015). As discussed earlier, work by Borner et al. (2016) displayed the instability of inferred *Haemoproteus* relationships. Depending on the analysis methods used, the topology vacillated to place the subgenus *Haemoproteus* as a sister clade to either *Parahaemoproteus* or *Plasmodium*. The methods included both maximum likelihood analysis and Bayesian inference of nucleotides and proteins (Borner et al. 2016). Work by Lutz et al. (2016) showed a monophyletic relationship for the genus *Haemoproteus*. The authors used mitochondrial cyt *b* sequences along with sequences for a single nuclear and an apicoplast gene. Interestingly, the authors found *Plasmodium* as a paraphyletic clade (Lutz et al. 2016). Recent work by Valkiūnas describing a new malaria parasite *Plasmodium delichoni* included phylogenetic analyses describing a sister relationship of the subgenus *Haemoproteus* and *Parahaemoproteus*, and a more distant relationship with *Plasmodium* (Valkiūnas et al. 2016). Finally, work by Pacheco et al. (2018) used 114 complete mitochondrial genomes of many taxa to infer the relationships of haemosporidians for the genera *Leucocytozoon, Plasmodium, Haemoproteus* and *Hepatocystis*. These phylogenetic reconstructions based on both Bayesian and likelihood methods found a sister relationship between *Plasmodium* and *Parahaemoproteus*. To summarize previous research, four articles supported a *Parahaemoproteus*/subgenus *Haemoproteus* sister relationship (Valkiūnas et al. 2010, 2016; Palinauskas et al. 2015; Lutz et al. 2016), five articles supported a *Plasmodium*/*Parahaemoproteus* sister relationship (Martinsen et al. 2008; Santiago-Alarcón et al. 2010; Martinezde la Puente et al. 2011; Pacheco et al. 2018) and one article obtained mixed results depending on the taxa, evolutionary model and genes analyzed (Borner et al. 2016). Pacheco et al. (2018) bears special mention as the most recent study with the largest dataset in terms of complete mitochondrial genes (Pacheco et al. 2018). With relatively few deep-sequenced haemosporidian taxa available, we can only speculate as to the difference such a selection can have on phylogenetic inference. It should be noted that while much research has been performed supporting both topologies, most articles only used one method of phylogenetic analysis. Additionally, many studies also focused only on cyt *b* analyses, specifically a 478 bp sequence of cyt *b*. Pacheco et al. (2018) stress the necessity of expanding data collection standards beyond sequencing only the cyt *b* gene (Pacheco et al. 2018). Our attempt to clarify the relationships between the subgenus *Haemoproteus* and *Parahaemoproteus* by

using a large number of nuclear genes strengthens the claim of a monophyletic *Haemoproteus* clade.

The research provided here adds to the greater knowledge of haemosporidian genetics but many questions remain. The lack of a *Leucocytozoon* genome or transcriptome prevents us from addressing the parasite relationships of the two subgenera of *Haemoproteus*. Additionally, the low taxon sampling implies that the relationships found here may change as more genomes and transcriptomes from haemosporidians become available. Additionally, coevolutionary studies between parasites and their vectors could add valuable pieces to parasite evolutionary reconstructions, as it seems likely parasites would adapt to closely related vectors (Lauron et al. 2015). Vector involvement in the evolution of haemosporidians remains an understudied area of research. We speculate that a method of sequencing both erythrocyte parasite life cycle stages and the oocyte or sporozoite stages present in the vector would prove advantageous in obtaining complete genetic data.

Conclusions

In closing, our findings are twofold. This study provides the first data about the transcriptome of blood parasites belonging to the family. The transcriptome of *H. columbae* represents a valuable resource for the advancement of haemosporidian genomic studies. Research on parasite antigen polymorphisms from a globally distributed parasite such as *H. columbae* could provide a significant system to study conserved invasion mechanisms. Additionally, our work on evolutionary relationships using large-scale transcriptomic datasets adds to our knowledge of potential haemosporidian evolutionary histories. Our dataset, combined with the original dataset from Bensch et al. (2016), will provide a stable framework to clarify the relationships of apicomplexan parasites in years to come.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by Research Council of Lithuania (Grant MIP-045/2015). This study was also supported by the grant to RNMS, NIH 1SC3GM118210-01A1. The authors would like to thank Elvin J. Lauron, Bradley Bowser and Dr. Frank Cipriano for helping in planning rna-seq and project logistics. We are grateful to Dr. Greg Spicer, Andrew Ontano, Dr. Emily Jane McTavish and Trent Liu for assistance on phylogenetic and scripting advisement.

References

Adriano EA, Cordeiro NS (2001) Prevalence and intensity of *Haemoproteus columbae* in three species of wild doves from Brazil. Mem Inst Oswaldo Cruz 96(2):175–178 [PubMed: 11285493]

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410 [PubMed: 2231712]

Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Manske M, Mangano V, Alcock D, Anastasi E, Maslen G, MacInnis B, Rockett K, Modiano D, Newbold CI, Doumbo OK, Ouédrago JB, Kwiatkowski DP (2011) An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole genome high-throughput sequencing. PLoS ONE 6(7):4–11

Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoekert CJ, Jr, Treatman C, Wang H (2008) PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res 37(Suppl_1):D539–D543 [PubMed: 18957442]

Babraham Bioinformatics (2017) http://www.bioinformatics.babraham.ac.uk/

Bensch S, Hellgren O, Pérez-Tris J (2009) MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. Mol Ecol Resour 9(5):1353–1358 [PubMed: 21564906]

Bensch S, Canbäck B, DeBarry JD, Johansson T, Hellgren O, Kissinger JC, Palinauskas V, Videvall E, Valkiūnas G (2016) The Genome of *Haemoproteus tartakovskyi* and its relationship to human malaria parasites. Genome Biol Evol 8(5):1361–1373 [PubMed: 27190205]

Böehme U, Otto TD, Cotton J, Steinbiss S, Sanders M, Oyola SO, Nicot A, Gandon S, Patra KP, Herd C, Bushell E, Modrzynska KK, Billker O, Vinetz JM, Rivero A, Newbold CI, Berriman M (2018) Complete avian malaria parasite genomes reveal hostspecific parasite evolution in birds and mammals. Genome Res 28:547–560 [PubMed: 29500236]

Borner J, Pick C, Thiede J, Kolawole OM, Kingsley MT, Schulze J, Cottontail VM, Wellinghausen N, Schmidt-Chanasit J, Burmester T (2016) Phylogeny of haemosporidian blood parasites revealed by a multi-gene approach. Mol Phylogenet Evol 94:221–231 [PubMed: 26364971]

Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol 1(1):e5 [PubMed: 12929205]

Carlson JS, Martínez-Gómez JE, Valkiūnas G, Loiseau C, Bell DA, Sehgal RN (2013) Diversity and phylogenetic relationships of hemosporidian parasites in birds of Socorro Island, México, and their role in the re-introduction of the Socorro Dove (Zenaida graysoni). J Parasitol 99(2):270–276 [PubMed: 23043349]

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4):540–552 [PubMed: 10742046]

Dimitrov D, Palinauskas V, Iezhova TA, Bernotienė R, Ilgūnas M, Bukauskaitė D, Zehtindjiev P, Ilieva M, Shapoval AP, Bolshakov CV, Markovets MY, Bensch B, Valkiūnas G (2015) *Plasmodium* spp.: an experimental study on vertebrate host susceptibility to avian malaria. Exp Parasitol 148:1–16 [PubMed: 25450775]

Earlé RA, Bastianello SS, Bennett GF, Krecek RC (1993) Histopathology and morphology of the tissue stages of *Haemoproteus columbae* causing mortality in *Columbiformes*. Avian Pathol 22(1): 67–80 [PubMed: 18670998]

Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ, Jr, Wang H, Brunk BP (2007) ToxoDB: an integrated *Toxoplasma gondii* database resource. Nucl Acids Res 36(suppl_1):D553–D556 [PubMed: 18003657]

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419:498 [PubMed: 12368864]

Garnham PCC (1966) Malaria parasites and other haemosporidia In: Malaria parasites and other haemosporidia. Wiley, Hoboken

Grech K, Watt K, Read AF (2006) Host–parasite interactions for virulence and resistance in a malaria model system. J Evol Biol 19(5):1620–1630 [PubMed: 16910991]

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat Protoc 8(8): 1494 [PubMed: 23845962]

Hellgren O, Waldenström J, Bensch S (2004) A new PCR assay for simultaneous studies of *Leucocytozoon, Plasmodium*, and *Haemoproteus* from avian blood. J Parasitol 90(4):797–802 [PubMed: 15357072]

Hellgren O, Križanauskiene A, Valkiūnas G, Bensch S (2007) Diversity and phylogeny of mitochondrial cytochrome B lineages from six morphospecies of avian Haemoproteus (Haemosporida: Haemoproteidae). J Parasitol 93(4):889–896 [PubMed: 17918371]

Hellgren O, Kutzer M, Bensch S, Valkiūnas G, Palinauskas V (2013) Identification and characterization of the merozoite surface protein 1 (msp1) gene in a host-generalist avian malaria parasite, Plasmodium relictum (lineages SGS1 and GRW4) with the use of blood transcriptome. Malaria J 12(1):381

Jasper MA, Hull JM, Hull AC, Sehgal RNM (2014) Widespread lineage diversity of *Leucocytozoon* blood parasites in distinct populations of western Red-tailed Hawks. J Ornithol 155(3):767–775

Joint Genome Institute (2017) http://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6(1):29 [PubMed: 16563161]

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12(4):357–360 [PubMed: 25751142]

Lauron EJ, Oakgrove KS, Tell LA, Biskar K, Roy SW, Sehgal RN (2014) Transcriptome sequencing and analysis of *Plasmodium gallinaceum* reveals polymorphisms and selection on the apical membrane antigen-1. Malaria Journal 13(1):382 [PubMed: 25261185]

Lauron EJ, Loiseau C, Bowie RC, Spicer GS, Smith TB, Melo M, Sehgal RN (2015) Coevolutionary patterns and diversification of avian malaria parasites in African sunbirds (Family Nectariniidae). Parasitology 142(5):635–647 [PubMed: 25352083]

Lee SA, Chan CH, Tsai CH, Lai JM, Wang FS, Kao CY, Huang CYF (2008) Ortholog-based protein-protein interaction prediction and its application to inter-species interactions. BMC Bioinform 9(12):S11

Lefèvre T, Sanchez M, Ponton F, Hughes D, Thomas F (2007) Virulence and resistance in malaria: who drives the outcome of the infection? Trends Parasitol 23(7):299–302 [PubMed: 17493874]

Levin II, Valkiūnas G, Santiago-Alarcón D, Cruz LL, Iezhova TA, O'Brien SL, Hailer F, Dearborn D, Schreiber EA, Fleischer RC, Ricklefs RE, Parker PG (2012) Hippoboscid-transmitted Haemoproteus parasites (Haemosporida) infect Galapagos Pelecaniform birds: evidence from molecular and morphological studies, with a description of Haemoproteus iwa. Int J Parasitol 41(10):1019–1027

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658–1659 [PubMed: 16731699]

Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, Ndjango JN, Sanz CM, Morgan DB, Locatelli S, Gonder MK, Kranzusch PJ, Walsh PD, Delaporte E, Mpoudi-Ngole E, Georgiev AV, Muller MN, Shaw GM, Peeters M, Sharp PM, Rayner JC, Hahn BH (2010) Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 467(7314):420 [PubMed: 20864995]

Lutz HL, Patterson BD, Peterhans JCK, Stanley WT, Webala PW, Gnoske TP, Hackett SJ, Stanhope MJ (2016) Diverse sampling of East African haemosporidians reveals chiropteran origin of malaria parasites in primates and rodents. Mol Phylogenet Evol 99:7–15 [PubMed: 26975691]

Martinez C, Marzec T, Smith CD, Tell LA, Sehgal RN (2013) Identification and expression of maebl, an erythrocyte-binding gene, in *Plasmodium gallinaceum*. Parasitol Res 112(3):945–954 [PubMed: 23224610]

Martinez-de la Puente J, Martinez J, Aguilar RD, Herrero J, Merino S (2011) On the specificity of avian blood parasites: revealing specific and generalist relationships between haemosporidians and biting midges. Mol Ecol 20(15):3275–3287 [PubMed: 21627703]

Martinsen ES, Perkins SL, Schall JJ (2008) A three-genome phylogeny of malaria parasites (Plasmodium and closely related genera): evolution of life-history traits and host switches. Mol Phylogenet Evol 47(1):261–273 [PubMed: 18248741]

Mbengue A, Bhattacharjee S, Pandharkar T, Liu H, Estiu G, Stahelin RV, Rizk S, Njimoh DL, Ryan Y, Kesinee C, Nguon C, Ghorbal M, Lopez-Rubio J, Pfrender M, Emrich S, Mohandas N, Dondorp AM, Wiest O, Haldar K (2015) A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. Nature 520(7549):683 [PubMed: 25874676]

National Center for Biotechnology Information (NCBI)[Internet] (2017) Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; Accessed Apr 12 2017

Neher RA, Bedford T, Daniels RS, Russell CA, Shraiman BI (2016) Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. Proc Natl Acad Sci USA 113(12):E1701–E1709 [PubMed: 26951657]

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302(1):205–217 [PubMed: 10964570]

Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, Lemieux J, Barrell B, Pain A, Berriman M, Newbold C, Llinas M (2010) New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-SEq. Mol Microbiol 76(1):12–24 [PubMed: 20141604]

Outlaw DC, Ricklefs RE (2014) Species limits in avian malaria parasites (Haemosporida): how to move forward in the molecular era. Parasitology 141(10):1223–1232 [PubMed: 24813385]

Oyola SO, Gu Y, Manske M, Otto TD, Alcock D, Macinnis B, Berriman M, Newbold CI, Kwiatkowski DP, Swerdlow HP, Quail MA (2012) Efficient depletion of Host DNA contamination in malaria clinical sequencing. J Clin Microbiol 51(3):745–751 [PubMed: 23224084]

Pacheco MA, Matta NE, Valkiunas G, Parker PG, Mello B, Stanley CE, Jr, Lenino M, Garcia-Amado MA, Cranfield M, Pond SLK, Escalante AA (2017) Mode and rate of evolution of haemosporidian mitochondrial genomes: timing the radiation of avian parasites. Mol Biol Evol 35(2):383–403

Pacheco MA, Cepeda AS, Bernotienė R, Lotta IA, Matta NE, Valkiūnas G, Escalante AA (2018) Primers targeting mitochondrial genes of avian haemosporidians: PCR detection and differential DNA amplification of parasites belonging to different genera. Int J Parasitol 48(8):657–670 [PubMed: 29625126]

Palinauskas V, Valkiūnas G, Bolshakov CV, Bensch S (2008) Plasmodium relictum (lineage P-SGS1): effects on experimentally infected passerine birds. Exp Parasitol 120(4):372–380 [PubMed: 18809402]

Palinauskas V, Iezhova TA, Križanauskienė A, Markovets MY, Bensch S, Valkiūnas G (2013) Molecular characterization and distribution of Haemoproteus minutus (Haemosporida, Haemoproteidae): a pathogenic avian parasite. Parasitol Int 62(4):358–363 [PubMed: 23557683]

Palinauskas V, Žiegytė R, Ilgūnas M, Iezhova TA, Bernotienė R, Bolshakov C, Valkiūnas G (2015) Description of the first cryptic avian malaria parasite, *Plasmodium homocircumflexum* n. sp., with experimental data on its virulence and development in avian hosts and mosquitoes. Int J Parasitol 45(1):51–62 [PubMed: 25449950]

Parker IM, Saunders M, Bontrager M, Weitz AP, Hendricks R, Magarey R, Suiter K, Gilbert GS (2015) Phylogenetic structure and host abundance drive disease pressure in communities. Nature 520(7548):542–544 [PubMed: 25903634]

Perkins SL (2008) Molecular systematics of the three mitochondrial protein-coding genes of malaria parasites: corroborative and new evidence for the origins of human malaria: full-length research article. DNA Seq 19(6):471–478

Perkins SL (2014) Malaria's many mates: past, present, and future of the systematics of the order Haemosporida. J Parasitol 100(1):11–25 [PubMed: 24059436]

PiroplasmaDB (2014), http://piroplasmadb.org

Puiu D, Enomoto S, Buck GA, Abrahamsen MS, Kissinger JC (2004) CryptoDB: the Cryptosporidium genome resource. Nucleic Acids Res 32:D329–D331 [PubMed: 14681426]

Rao VS, Srinivas K, Sujini GN, Kumar GN (2014) Protein-protein interaction detection: methods and analysis. Int J Proteomics 2014:147648 [PubMed: 24693427]

Ricklefs RE, Fallon SM (2002) Diversification and host switching in avian malaria parasites. Proc R Soc Lond B 269(1494):885–892

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19(12):1572–1574 [PubMed: 12912839]

Santiago-Alarcón D, Outlaw DC, Ricklefs RE, Parker PG (2010) Phylogenetic relationships of haemosporidian parasites in New World *Columbiformes*, with emphasis on the endemic Galapagos dove. Int J Parasitol 40(4):463–470 [PubMed: 19854196]

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313 [PubMed: 24451623]

Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. Bioinformatics 26(12):1569–1571 [PubMed: 20421198]

Valkiūnas G (2005) Avian malaria parasites and other haemosporidia. CRC Press, Boca Raton

Valkiūnas G, Iezhova TA (2017) Exo-erythrocytic development of avian malaria and related haemosporidian parasites. Malaria J 16(1):101

Valkiūnas G, Iezhova TA, Loiseau C, Chasar A, Smith TB, Sehgal RN (2008a) New species of haemosporidian parasites (Haemosporida) from African rainforest birds, with remarks on their classification. Parasitol Res 103(5):1213 [PubMed: 18668264]

Valkiūnas G, Zehtindjiev P, Dimitrov D, Križanauskienė A, Iezhova TA, Bensch S (2008b) Polymerase chain reaction-based identification of Plasmodium (Huffia) elongatum, with remarks on species identity of haemosporidian lineages deposited in GenBank. Parasitol Res 102(6):1185–1193 [PubMed: 18270739]

Valkiūnas G, Santiago-Alarcón D, Levin II, Iezhova TA, Parker PG (2010) A new Haemoproteus species (Haemosporida: Haemoproteidae) from the endemic Galapagos dove *Zenaida galapagoensis*, with remarks on the parasite distribution, vectors, and molecular diagnostics. J Parasitol 96(4):783–792 [PubMed: 20486741]

Valkiūnas G, Ilgūnas M, Bukauskaitė D, Žiegytė R, Bernotienė R, Jusys V, Eigirdas V, Fragner K, Weissenböck H, Iezhova TA (2016) *Plasmodium delichoni* n. sp.: description, molecular characterisation and remarks on the exoerythrocytic merogony, persistence, vectors and transmission. Parasitol Res 115(7):2625–2636 [PubMed: 27000087]

Videvall E (2018) *Plasmodium* parasites of birds have the most AT-rich genes of eukaryotes. Microb Genomics 4:1–9

Videvall E, Cornwallis CK, Palinauskas V, Valkiūnas G, Hellgren O (2015) The avian transcriptome response to malaria infection. Mol Biol Evol 32(5):1255–1267 [PubMed: 25636457]

Videvall E, Cornwallis CK, Ahrén D, Palinauskas V, Valkiūnas G, Hellgren O (2017) The transcriptome of the avian malaria parasite *Plasmodium ashfordi* displays host-specific gene expression. Mol Ecol 26(11):2939–2958 [PubMed: 28267239]

Waite JL, Henry AR, Adler FR, Clayton DH (2012) Sex-specific effects of an avian malaria parasite on an insect vector: support for the resource limitation hypothesis. Ecology 93(11):2448–2455 [PubMed: 23236915]

Waite JL, Henry AR, Owen JP, Clayton DH (2014) An experimental test of the effects of behavioral and immunological defenses against vectors: do they interact to protect birds from blood parasites? Parasites Vectors 7(1):104 [PubMed: 24620737]

Waldenström J, Bensch S, Kiboi S, Hasselquist D, Ottosson U (2002) Cross-species infection of blood parasites between resident and migratory songbirds in Africa. Mol Ecol 11(8):1545–1554 [PubMed: 12144673]

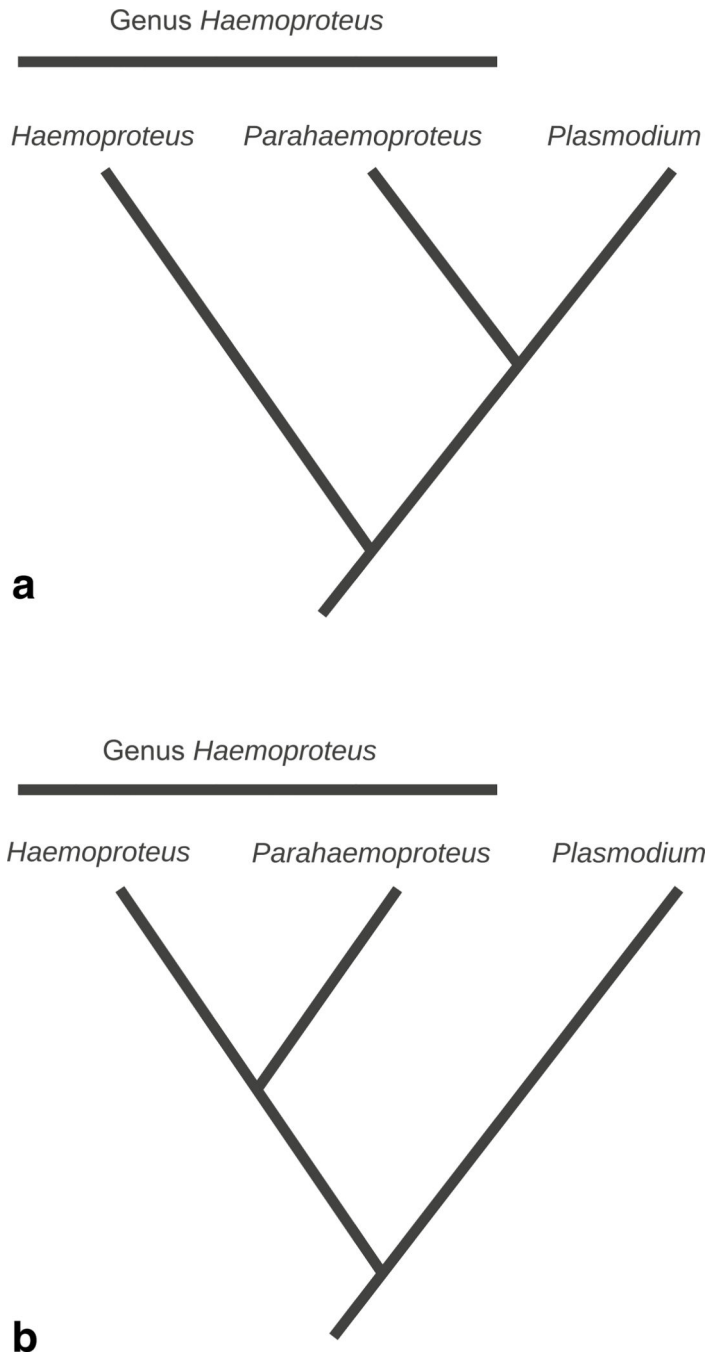

Fig. 1.

Alternate trees describe the prominent hypotheses described by previous research. **a** Describes genus *Haemoproteus* as paraphyletic. **b** Describes genus *Haemoproteus* as monophyletic

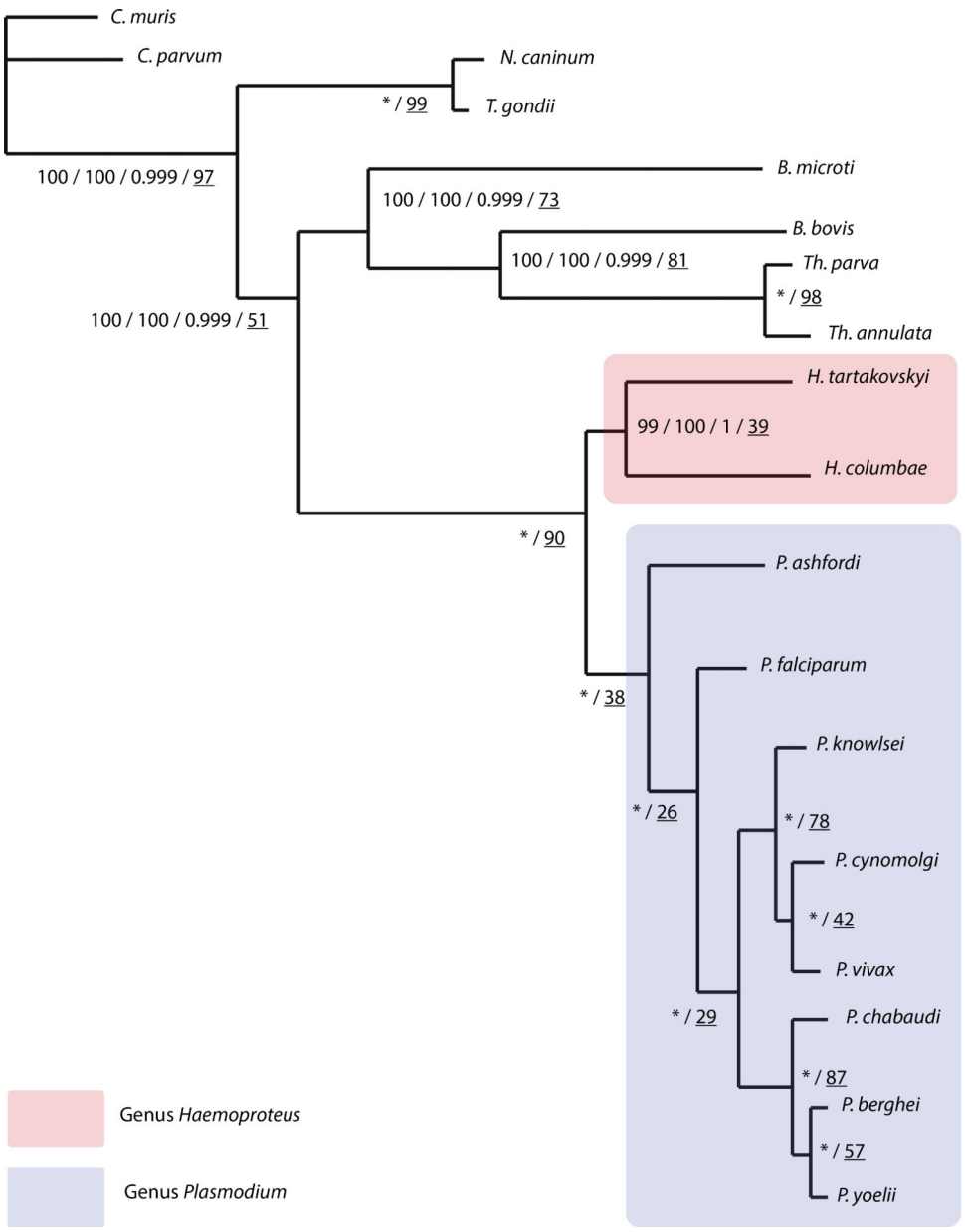

Fig. 2.

Composite phylogenetic tree of apicomplexans focusing on haemosporidian parasites. Tree includes bootstrap support values for maximum likelihood analyzed ungapped and gapped datasets, Bayesian posterior probability and single-gene tree support percentages, respectively (ML-ungapped/ML-gapped/Bayes/single-gene). * indicates 100% bootstrap support and posterior probability for all analysis. Underlined numbers indicate the percentage of single-gene trees supporting the node topology. Bootstrap values produced with RAxML and posterior probability values produced with MrBayes. The scale bar displays branch length in units of evolutionary distance

Table 1

Transcriptome assembly statistics for *H. columbae*

	H. columbae
Total raw reads	220,662,194
Number of reads used for assembling	102,790,708
Assembled contigs before filtering	267,604
Assembled contigs after filtering	17,152
GC content (%)	17.71
Total assembly length (BP)	13,223,089
Number of contigs > 500 bp	9744
Number of contigs > 1000 bp	4258
Longest contig (BP)	9502
Contig N50	1030
Number of bases in gapped dataset	178,347
Number of bases in ungapped dataset	59,028
Bensch et al. (2016) dataset bases	133,965

The number of bases used in both gapped, ungapped and original Bensch et al. (2016) datasets, respectively