Origin and diversity of Malaria parasites and other Haemosporida.

M. Andreína Pacheco[1,*], Ananias A. Escalante[1,*]

[1]Biology Department/Institute of Genomics and Evolutionary Medicine (iGEM), Temple University, Philadelphia, Pennsylvania 19122-1801, USA.

***Correspondence:**

M. Andreína Pacheco: maria.pacheco@temple.edu, https://orcid.org/0000-0002-5682-7299

Ananias A. Escalante: ananias.escalante@temple.edu, https://orcid.org/0000-0002-1532-3430

Keywords: Biodiversity, DNA barcode, molecular clock, parasites, phylogenetic, species concept.

Abstract

Symbionts, including parasites, are ubiquitous in all world ecosystems. Understanding the diversity of symbiont species addresses diverse questions, from the origin of infectious diseases to inferring processes shaping regional biotas. Here, we review the current approaches to studying Haemosporida's species diversity and evolutionary history. Despite the solid knowledge of species linked to diseases, such as the agents of human malaria, studies on haemosporidian phylogeny, diversity, ecology, and evolution are still limited. The available data, however, indicate that Haemosporida is an extraordinarily diverse and cosmopolitan clade of symbionts. Furthermore, this clade seems to have originated with their vertebrate hosts, particularly birds, as part of complex community levels processes that we are still characterizing.

Haemosporida: A brief overview

Haemosporidian **species** (see Glossary) are a diverse clade of vector-borne protist symbionts found in almost all terrestrial ecosystems [1,2,3]. All the species belonging to the order Haemosporida (Phylum Apicomplexa) have heteroxenous life cycles (Figure 1) that occur in two types of **hosts**, invertebrates from the order Diptera, and three classes of vertebrates: Reptilia, Aves, and Mammalia [1-17].

Several haemosporidian families and genera have been proposed based on morphology and life history traits (Table 1). There is agreement that the order includes three families: Plasmodiidae Mesnil, 1903; Haemoproteidae Doflein, 1916; and Leucocytozoidae Fallis and Bennett, 1961. A fourth, Garniidae Laison, Landau, and Shaw, 1971, has been proposed but is not widely accepted (see Table 1, [13]). Each family may include several genera and subgenera (Table 1). Plasmodiidae has been the best studied regarding some of its species' biology and diversity, particularly the species belonging to the genus *Plasmodium*. Among those are the agents of humans (e.g., *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, and *Plasmodium malariae* [2,16] and avian malarias such as *Plasmodium relictum* [3]. Indeed, this avian parasite was instrumental in implicating vectors in *Plasmodium* transmission [2].

Plasmodium species in nonhuman primates have been used as research models for understanding parasites that commonly infect humans. Some species are part of zoonotic infections, such as *Plasmodium knowlesi,* which is usually found in macaques in Southeast Asia [2,16]. In addition to those species of interest in human health, around 600 haemosporidian species have been described, mainly in Aves and Reptilia [3,7,18,19]. However, such a number is likely an underestimation.

Beyond being an academic problem, solid symbiont taxonomy allows for addressing fundamental questions, from the emergence of infectious diseases to historical and anthropogenic

processes shaping regional biotas. Putative symbiont species, particularly haemosporidian species, are often studied as a proxy for "disease." Alternatively, a comprehensive view could acknowledge that there is almost no terrestrial ecosystem without these symbionts [4-6,18,19]. Thus, given their abundance and worldwide distribution, haemosporidian species are a component of biodiversity whose function and importance need to be studied. We review our current understanding of these symbionts' diversity, phylogenetic relationships, and origins. We also discuss the challenges of incorporating these symbionts into biodiversity sciences while understanding the origin of evolutionary lineages of parasites that could burden their host populations.

Haemosporidian species discovery and delimitation

Although multiple species concepts exist [20,21], species are ultimately delimited using traits. The value of such **taxa** is that they are hypotheses to be tested using diverse biological data [21]. Haemosporidian species, when delimited, are expected to integrate information on how they disperse, differentiate, and interact with their hosts across geographic, temporal, ecological, and evolutionary contexts. Thus, species allow making inferences about processes within and between organization levels, from organisms to ecosystems [21].

Modern taxonomy aims to reflect evolution requiring that each **taxon** corresponds to a **monophyletic group** [20]. Species are delimited in classic haemosporidian taxonomy using their morphological traits in blood stages as observed on Giemsa-stained blood films with a light microscope [2-4,21]. Thus, the underlying assumption of any taxa defined using morphology is that it reflects a clade [20,21]. However, scaling up morphology to study haemosporidian species diversity and evolution has been challenging.

Morphology reproducibility is affected by the expertise in the fixation and examination of blood films [18,22]. It can change in different host species, as some traits depend on how the host cell changes shape [18,22], and traits in their dipteran hosts are often out of reach because of a lack of information on the vector. Parasitemias below the limit of detection by microscopy or submicroscopic infections are common. Those submicroscopic infections hinder the identification of known species using morphology [21,23-25], description of new ones that require a detailed account of parasite blood stages [6,18,19], and documenting the frequency of species in host populations [25]. Finally, assigning a blood stage's morphology to a taxon is sometimes difficult due to infections with different species (co-infection). Such limitations have been documented in several hosts, including humans [e.g., 25-31]. Thus, morphology has limitations in procuring information about the species (e.g., prevalence and distribution), leading to inadequate data on many taxa, sometimes limited to their original descriptions.

As an alternative, molecular data are widely used for detecting and identifying haemosporidian **lineages**, inferring phylogenies, and studying diversity patterns [21,24,32-37]. It is worth distinguishing species identification (**DNA barcoding** in the strict sense) from discovery/delimitation [21,24,38]. The first involves molecular data linked to described taxa, which is still pending in haemosporidians [18,19,21,39-41]. The second requires a distinct monophyletic group or a character (e.g., a unique insertion); however, the standard remains undefined [24,29,33,34,38]. A minimalist approach for studying diversity uses lineages from a molecular locus, a haemosporidian mitochondrial cytochrome b gene (*cytb*) fragment. Valuable biodiversity studies use such lineages as a proxy for species. This approach has evidenced a previously unknown species diversity as several thousands of *cytb* lineages worldwide have been found [e.g., 21,28,42-46].

It could be argued that a consistently reported lineage is a discrete biological entity that can integrate later other data, including morphological phenotypes and information about their hosts and vectors as they become available. However, since morphological traits are required to describe species [e.g., 47-50], many potential species discovered using only molecular data remain "dark taxa" [51]. The controversy about using non-morphological data in haemosporidian taxonomy is not new [2,3]. Early proponents considered such data crucial in discovering and delimiting species as morphology could not provide consistent answers [1,2]. For example, finding gametocyte stages is critical in classical taxonomy as it may imply a suitable host that sustains transmission. However, gametocytes in a slide may not provide such evidence. E.g., *P. knowlesi* gametocytes are observed in humans, but transmission among humans does not sustain transmission in natural settings. Also, not observing gametocytes by microscopy does not exclude that the host participates in transmission [23,25]. More importantly, the need for evidence that a host is competent for transmission does not supersede that a DNA lineage exists regardless of traits we may ignore, including its morphology and host range.

A consensus is emerging on an integrative approach, but, as expected, consistency between lines of evidence is imperfect [18,19,21,48]. There are **cryptic species** or species with an apparent lack of morphological differences (e.g., *H. jenniae*/*H. iwa* [52] and *Plasmodium ovale* [53]) and limited molecular divergence from distinct morphospecies [34,52,54]. These inconsistencies may create opportunities to understand processes such as **phenotypic plasticity** in morphology; however, a parsimonious explanation in some cases is that morphology lacks the resolution to separate closely related species due to the limited number of traits.

Although molecular data have advantages, there are technical problems. As in the case of morphology, co-infections common in wildlife [35,55,56] may lead to **chimeras** when a PCR

amplicon is sequenced directly. Thus, examining the electropherograms, cloning, and deep sequencing is required. Second, primers can selectively amplify, as evidenced by *Leucocytozoon* species in birds that may not be detected by commonly used primers [48]. Thus, new data must be generated to design primers that detect such taxa. Perhaps more critical is linking other sources of information to molecular lineages, which are perceived as a temporary need rather than species. The taxonomic load has been mitigated by recording detection-based occurrences databases like MalAvi [46] by using specific sets of primers targeting a region of the parasite mitochondrial *cytb* gene [33,35-37,44-46]. However, a standard for using molecular data to delimit species is still needed [21,24,33], as the fragment of *cytb* gene has limited phylogenetic information to address the systematic of Haemosporida [21,24,33-35].

It is important to highlight that molecular criteria to delimit species require understanding the locus or loci genetic diversity, mutation rates, and the putative species population structure [34]. Studies indicate that the *cytb* and other mitochondrial genes, when the data has quality regarding the number of base pairs [35], evolve at a mutation rate that allows species delimitation, at least compared to morphological species [21,30-35]. However, knowledge of the putative species' population structure lags [24]. A source of concern is that hybridization has led to incomplete mitochondrial gene lineage sorting in other eukaryotes. However, it has not been detected in Haemosporida [34] because their life cycles are prone to inbreeding. Notably, the strength of **purifying selection** changes between haemosporidian clades; thus, interpreting differences may not be the same in all groups [34]. Regardless of these considerations, the *cytb* gene has been an extraordinary tool for unveiling the diversity of haemosporidian species [21].

Nowadays, putative haemosporidian species are first detected molecularly rather than morphologically. A case study is *Plasmodium* species discovered in African apes. Such species

with whole nuclear genomes available now were proposed on high-quality but limited mitochondrial data (no *cytb* fragment) lacking morphological descriptions [e.g., 57,58]. Indeed, many old reports may have considered them into a single morphospecies [2,16,29].

It could be argued that having molecular criteria to delimit species is essential in biodiversity sciences. A fruitful discussion can focus on the steps to delimit species, regardless of whether those were proposed using morphological traits or molecular data. Nevertheless, when using molecular data is critical to agree on the data type (base pairs of a locus or loci), the ecological samplings (reproducibility and detection in well-identified hosts across sites), and how to interpret that information to consider a DNA lineage and its variants as part of a species [21,24,33,34]. Such criteria can increase the contribution of biodiversity studies [21,28,33,37] to delimit putative species where researchers can focus on aggregating other data types, including morphology and host range.

M**olecular taxonomic inferences, from species delimitation to phylogenies.**

Single gene approaches.

The 18s rRNA has been the locus of choice for phylogenetic studies and species detection in protists [e.g., 17]. It was used to diagnose [59] and study the origins of *Plasmodium* infecting humans [60,61]. However, non-concerted evolution among paralogs in *Plasmodium* [62] and across Haemosporida [63] has made its use in phylogenetic analyses and species delimitation challenging. Other genes, such as the one encoding the circumsporozoite surface protein (CSP), were used due to the interest in the protein itself [64,65]. Although gene-encoding antigens under positive selection may fail to estimate the species trees, they can detect known species, particularly if the variation of such locus is documented [66].

As an alternative, the *cytb* was used for phylogenetic analysis adding a new locus for detecting species in haemosporidians [67]. Some of the advantages of mitochondrial genes are: (1) The mitochondrial genomes are conserved, allowing the design of primers [e.g., 34,35,42-45], (2) it occurs in multiple copies per symbiont, making it sensitive for detection, (3) it is reproducible and unambiguously digitalized facilitating the discovery of cryptic species [21,31,59,47,48,67], (4) it separates species that have been identified using morphology [21,34], and (5) allows for phylogenetic analyzes even when information about other traits is unavailable. A setback is that although a *cytb* fragment has been incredibly useful in detecting variation [21,33,37], the limited number of base pairs (few informative phylogenetic sites) does not always allow for robust phylogenetic reconstructions or species delimitation [35]. Furthermore, using different *cytb* gene regions hinders lineage comparison across some studies [44,45].

Multigene approaches

Two major approaches have been used. One target loci across mitochondrial (e.g., cytochrome c oxidase subunit 1 gene: *cox1* and *cytb* gene), apicoplast (e.g., *clpc* gene), and nuclear genomes (e.g., *dhfr*, *asl*, *ef2*)[68-78]. The other instead uses the complete (>5500 bp) or partial mitochondrial genome, a single locus with multiple genes [14,15,21,30,34,79,80]. Thus far, both provide consistent results, particularly if no outgroup is used to root the tree (Figure 2). Nevertheless, the same taxa are not sampled in all phylogenies, so studies cannot be thoroughly compared.

The mitochondrial genome is a single locus with the advantage (or disadvantage, depending on its use) that it is not expected to recombine and is not saturated at the time scales under consideration. Mitochondrial genes have comparable codon usages and AT content across taxa

185 [14,34]; this is important because differences in AT content could lead to model misspecification
186 in phylogenetic analyses [76,77]. However, long-range PCR, cloning, or next-generation long-read
187 sequencing may be required to resolve mixed infections. Given their cost, these tools are not
188 widely available in the research community. Using multi-loci data does not solve the mixed
189 infections problems because lineages could be differentially amplified by a particular set of
190 primers targeting a locus. The possibility of chimeras worsens due to the lack of reference genomes
191 when using next-generation sequencing.

192 Although there are multiple phylogenetic studies, their comparison is not easy not only
193 because they usually include different taxa, as stated early, but also because the number of
194 informative sites reported is inaccurate. Adding a sequence with missing sites (e.g., Ns in some
195 regions of the sequences) eliminates such sites from the complete analysis. Thus, the taxa with
196 worse data drive the phylogenetic analysis as the alignment has missing data. If such sequences
197 are not critical, a solution is to exclude them from the study and keep more informative sites with
198 fewer sequences. Alternatively, missing data could be dealt with explicitly if such a level of
199 sophistication is required given the question [81]. Whatever the choices, a good practice is to report
200 the number of sites in the alignment shared by all species, excluding gaps and missing data, not
201 simply the alignment length as reported by the software. Irrespective of these cautionary notes, the
202 information obtained from multigene data, whether a single locus such as the mitochondrial
203 genome or multi-loci, is consistent and informative. Indeed significant advances have been made
204 in our understanding of haemosporidian evolution (Figure 2, Box 1).

205

206 *Haemosporidian phylogeny and taxonomic implications*

Molecular phylogenies have been enriched by an expanded sampling, including novel putative taxa (e.g., *Plasmodium* in lemurs [14,22], *Haemoproteus catharti* [82], *Haemoproteus antigonis* [83], *Haemoproteus pulcher* [50]), and data for genera historically less studied (e.g., *Nycteria*, *Polychromophilus*, *Hemocystidium*, *Plasmodium* in lemurs). Still, several phylogenetic relationships are unsolved (Figure 2). However, there is extraordinary progress in our understanding of Haemosporida evolution. In particular, molecular evidence shows that the classical taxonomy does not comply with modern standards (Box 1). The genus *Plasmodium* is not a monophyletic group [32,34,67,70,72,73], as was proposed by some classical taxonomists [2]. Although there is not enough evidence to establish the phylogenetic relationships of newly proposed *Haemoproteus* species using morphology, such as *H. catharti*, *H. antigonis* and *H. pulcher,* the molecular data indicate that the genus *Haemoproteus* is **polyphyletic**.

The phylogenetic relationships of genera *Hepatocystis*, *Nycteria,* and *Polychromophilus* with others remain unsolved [15,34,73,74,76], requiring better taxa sampling. The relative position of *Hepatocystis* may be affected by including the lemur *Plasmodium* species in some analyses (see Figure 2 for comparison)[14,15,34]. Lemur *Plasmodium* species, thus far, are a diverse understudied monophyletic group whose relationship with other primate malarias needs to be elucidated [15].

Plasmodium species infecting Aves and the available evidence from Reptilia suggest that they are a monophyletic group with several distinct clades (Figure 2)[34,76]. However, lizard Haemosporida from outside the Americas are poorly represented in molecular studies. The clade from *Plasmodium* found in ungulates is separated from other mammals *Plasmodium* species, evidence that may indicate a host switch from mammals to birds [75]. Although the taxonomic sampling is incomplete, it can be speculated that the well-documented symbiotic relationship

between ungulates and birds [84] could have favored such an event. *Haemocystidium,* a genus sometimes considered as *Plasmodium,* others as *Haemoproteus*, is now a distinct clade (see Figure 2). [6,85-88]. Indeed, some *Haemoproteus* have been reclassified as *Haemocystidium* by incorporating molecular data. Molecular phylogenetic analyses also support the division of *Haemoproteus* into two subgenera (Figure 2), *H*. (*Haemoproteus*) and *H*. (*Parahaemoproteus*) [3,70]. These subgenera with similar morphologies in their vertebrate host show differences in their vectors (hippoboscid vs. ceratopogonid flies) and vector stages (sporogonic development) [3].

One consistent result is an association of groups of vectors at the family/subfamily level with specific haemosporidian clades [3,34,70]. *Leucocytozoon* (*Leucocytozoon*) spp. is a monophyletic group where species are transmitted by simuliid flies (Simuliidae). *Haemoproteus* (*Haemoproteus*) spp. are transmitted by hippoboscid flies (Hippoboscidae) and *H*. (*Parahaemoproteus*) spp. by biting midges (Ceratopogonidae) [3]. Interestingly, *Leucocytozoon* (*Akiba*) *caulleryi* that share a common ancestor with *H*. (*Parahaemoproteus*) spp. is also transmitted by biting midges. Mosquitoes (Culicidae) [3] transmit most *Plasmodium* species infecting lizards, birds, and mammals. Some exceptions are *P*. (*Paraplasmodium*) *mexicanum* and *P*. (*Paraplasmodium*) *chiricahua* that are transmitted by sand flies (Phlebotominae) [7], and the two species appear as sister taxa in all phylogenies [e.g., 34,76]. Finally, the limited evidence of vectors for *Plasmodium* in ungulates implicates *Anopheles* mosquitoes [75], like the other *Plasmodium* in mammals. Regardless of these patterns, the link of vector families with haemosporidian clades deserves additional exploration. For example, *Haemoproteus* DNA has been detected in different mosquito genera (*Anopheles*, *Aedes*, *Verrallina*, *Culex*, *Coquillettidia*) [89]. However, it was not proven that those *Haemoproteus* could complete the cycle in such vectors [90].

253 Understanding the diversification of the Haemosporida and their character evolution (e.g.,
254 vectors or malaria pigment) requires a clear root for the Haemosporida phylogeny. Using *Theleria*
255 as an outgroup, it was proposed that the common ancestor of *Haemoproteus* and *Leucocytozoon*
256 diverged from the rest early on in the evolution of the order, a consistent proposal with a root
257 statistically estimated [76,91] (Figure 2b). Piroplasmida (*Babesia* or *Theleria*) and Haemosporida
258 were considered "close" [61], but they seem inappropriate to root the Haemosporida tree due to a
259 risk of **long-branch attraction effect**. The fact that Piroplasmida genomes have 4 chromosomes
260 and Haemosporida species such as *Plasmodium* have 14 gives an idea of the potential divergence.
261 Likewise, the mitochondrial genome of Haemosporida and the known Piroplasmida have different
262 organizations [34]. Statistical estimates of the root [34,91] have not been concordant, perhaps
263 because of the rate heterogeneity and differences in the taxa sampled. Regardless of the root, all
264 evidence nowadays points to a radiation of the Haemosporida with modern birds [15,34,88].
265 What seems a critical finding is that some morphological and life history traits that define
266 different taxa are convergent across clades (Table 1)[67]. However, whether to revise the
267 taxonomy is a controversial matter. Given the practical importance of the current taxonomy in
268 fields such as human and veterinary medicine, these taxonomic problems are unlikely to be solved
269 in the short term [76,77].
270
271 **Molecular clock: Timing the radiation of haemosporidians**
272 Timing the events leading to the radiation of Haemosporida allows testing inferences on how
273 these symbionts were part of processes driving the origins of biotas. Molecular dating initially
274 focused on the origin of human parasites, assuming a constant rate of evolution. The earliest
275 attempt used a universal rate for the 18S SSU rRNA concluding that *Plasmodium*, and by

276 extension, all Haemosporida, was older than their vertebrate hosts [see 92], generating some
277 controversies in the field [93]. More recent studies on the mitochondrial genome assumed a strict
278 clock on the *Plasmodium* mitochondrial genome [79], even when the constant rate assumption was
279 rejected. The origin of the genus *Plasmodium* was estimated to be 22 - 41 Ma (million years ago)
280 using an *ad hoc* method. Other studies also assumed a constant rate [94] on a fragment of *cytb*
281 gene finding younger times (e.g., the global diversity of the Haemosporida across terrestrial
282 vertebrates took place at 16.2 Ma).

283 The first study using nuclear genomic data estimated the origin of *Plasmodium* species to be
284 between 64 and 120 Ma [95], clearly older than those from mitochondrial genes. The method
285 estimated relative times against a reference distance between two species; the original study used
286 *P. vivax* and *P. knowlesi*. Although a constant rate for single-copy protein genes orthologs is
287 assumed, it is not a universal rate for all genes. However, the method assumes a single rate to
288 transform the relative times into absolute times [95].

289 Beyond assuming a constant rate in one way or another, the early studies described above are
290 not comparable in data and methods used to estimate rates and times [77]. Nowadays, there is an
291 agreement that using a constant rate in *ad hoc* methods is not a robust approach to molecular dating
292 [96]. Alternatively, several approaches allow for rate variation using **bayesian methods**. There are
293 reviews on such methods [96], so only a brief description is provided. Recently, molecular dating
294 in Haemosporida has been driven using mitochondrial genomes because there is a broad sampling
295 of taxa (Figure 3). The impact of two sets of assumptions has been explored using bayesian
296 methods. The first concern was how to model the rate variation among lineages. The two standard
297 models differ in assuming autocorrelation or no correlation in the rate variation among the
298 lineages. The second set of assumptions involves the informative **calibration constraints** for

299 reliable timescales. Calibrations are derived from **fossil** data and biogeographic events from the
300 organisms included in the analysis. However, there are limited and controversial haemosporidians
301 reports in the fossil records that are difficult to use for molecular dating [11,12,97], so host data
302 have been used as calibrations [21]. Although not a typical secondary calibration, the calibrations
303 used in Haemosporida still require additional assumptions, such as biogeographic events on the
304 host and their fossils, assuming that such events affect the parasite (Figure 3). Thus, such
305 calibration constraints should be described so others can evaluate them [14,34]. Given the
306 assumptions, molecular time estimates make some scenarios more parsimonious than others. A
307 critical problem seems to be that all available calibrations are in *Plasmodium* from mammals
308 emphasizing the importance of seeking new calibrations [21], particularly within other genera in
309 Aves and Reptilia. Nevertheless, the calibrations that have been used are internally consistent, so
310 removing one calibration does not yield incongruent time estimates when the analyses are
311 compared [15,34,35]. The time priors used thus far provide an incomplete but plausible framework
312 for **molecular clock** studies.

313 A first study used two calibrations within primates, compared different scenarios, and the two
314 Bayesian methods that assume autocorrelation or no correlation (independent) on the rates of
315 mitochondrial genome lineages [14]. A first observation is that the estimates with the
316 autocorrelation model yielded slightly older times than the independent rate models. Some
317 **credibility intervals** overlapped with previous estimates from Hayakawa et al. [79]. This study
318 used the parasite's origin in lemurs to validate the time estimates from separate calibrations [14].
319 The proposed scenarios were investigated in a separate study in rodent malarias [98], producing
320 similar results. Significantly, those estimations were also rescued when new data were included
321 [15].

Expanded studies on mitochondrial genomes, including data from avian species (genera *Leucocytozoon*, *Haemoproteus*, and *Plasmodium*) and reptiles (*Haemocystidium*), still found that different models of evolutionary rate variation across lineages, independent or autocorrelated, affect time estimates, particularly in avian haemosporidian clades [34]. Although the credibility intervals overlapped between the two models, using calibrations that may reduce the variance in the time estimates is critical. Overall, the time estimates that are consistent across the mtDNA genomes studies [14,15,34,88,98] with the best sampling in terms of taxa (Figure 3), indicate that: (1) The african ape parasites that are part of the clade that includes *Plasmodium falciparum*, the agent of the most aggressive form of human malaria, diverged early on from other primate parasites, including all the other species infecting humans (particularly with the origin of Catarrhini primates), and considering the evolutionary histories of their extant hosts, their origins likely took place in Africa; (2) the origin of the lemur malaria parasites sampled so far are consistent with a scenario where they originated when they were introdueced into Madagascar by terrestrial vertebrates [14,15]; (3) several primate malaria parasites, including lemur *Plasmodium* species, appear to be younger than the origin of their hosts as a clade, consistent with our current understanding of frequent host-switches [30,76]; (4) the time estimates for the origin of Haemosporida (age nodes = 77.8 and 79.9 Ma, Ref) overlapped with the calibration constrain proposed for the time to the common ancestor of palaeognathous birds (56.8–86.8 Ma [99], https://fossilcalibrations.org accessed on January 2023) and with molecular estimates for the origin of modern birds [e.g., 100]; and (5) all scenarios explored indicate that the significant haemosporidian parasite genera, as described in the classical taxonomy, could diversified concomitantly with the radiation of their vertebrate host orders after the Cretaceous–Paleogene (K–Pg) boundary (66 Ma) [15,34,88].

Although limited in the taxa sampled, times estimates using the combination of mtDNA plus partial apicoplast genomes (~6kb) showed some differences with the molecular clock estimates using solely mitochondrial data in the clade that includes *P. falciparum*. For example, when considering the mtDNA alone [15,34,88], the time estimates for the origin of the *P. falciparum* clade were 17.3–28.71 Ma, which coincides with the time of origin of the Hominidae that includes *Homo*, *Pan*, *Gorilla*, and *Pongo* (11.6–33.9 Ma [99] https://fossilcalibrations.org last accessed on January 2023). Then, when the apicoplast loci were included, the same clade was younger (7.1–10.9 Ma)[15], coinciding with the extant Homininae (*Homo*, *Pan*, and *Gorilla*; credible interval (CrI): 8.4 – 9.6 Ma; https://www.timetree.org/). An element that may explain the older mtDNA time estimates is that rate differences in the primate *Plasmodium* clades may affect the *P. falciparum* clade estimates [30,34]. Considering these factors, a scenario of a common ancestor for the clade that includes *P. falciparum* that coincides with the origin of Homininae seems plausible and could explain the extant host range of these parasites.

Nevertheless, the best way to move the field forward is to seek additional calibration constraints, ideally within parasite genera such as *Haemoproteus* and *Leucocytozoon,* by discovering clades linked to biogeographic or host-speciation events. Simply adding data from extant species does not solve the problem. Species without calibration may increase the variance in time estimates because of heterogeneity in their rates of evolution. Adding loci is beneficial if they have congruent phylogenetic signals, are not saturated, have similar rate variation across lineages (e.g., no gene under positive natural selection), and do not require different substitution models (e.g., similar GC content).

Knowledge gaps and future directions.

Studies on Haemosporida can benefit from a molecular standard to delimit species. Since the *cytb* gene is widely used, a possibility is using the whole gene or a fragment of the mitochondria that includes it; this allowed the discovery of the *Plasmodium* species in Apes. In addition to phylogenetically informative locus or loci, such a standard should require minimal sampling to detect the proposed species in well-identified hosts across studies. The proposed standard also should address technical problems such as dealing with co-infections. These robust molecular species can facilitate aggregating other information, such as morphology and life history traits, when those become available, particularly addressing the problem of cryptic species.

Regarding the Haemosporida phylogeny, the taxonomic sampling needs to be improved; there is limited or no molecular data in some taxa (reptilian *Plasmodium*, *Saurocytozoon, Fallisia, Garnia, Hepatocystis, Polychromophilus, Nycteria,* and lemur *Plasmodium*). There are only a handful of studies on hosts such as nonpasserine birds and reptiles. Finally, molecular clock studies will benefit from additional calibration points outside of the genus *Plasmodium.*

Concluding Remarks

Putative species are first detected molecularly than using morphology. Although a pragmatic system based on short *cytb* gene sequences has been incredibly informative in assessing biodiversity patterns, it can fail short of delimiting species (see outstanding questions). Thus, a rigorous but realistic molecular standard can move the field forward from a lineage-based approach toward species inference. Such molecularly delimited species should integrate morphological and vector data when available. The use of molecular data should consider sampling at different levels; when a species is delimited, data across individuals and studies are needed. Discovering species in understudied hosts (e.g., bats, reptiles and nonpasserine birds) and haemosporida taxa is also

critical as they may change our understanding of the Haemosporida phylogeny. A molecular Haemosporida phylogeny can update the classical taxonomy to contemporary standards. Although far from perfect, molecular data provides hypotheses about symbionts' rate and mode of evolution. Inferring time incorporates Haemosporida into the discussion on how these symbionts were part of processes driving the origin of biotas. Still, calibrations are needed by searching for clades of Haemosporida, like in *Leucocytozoon* and *Hemoproteus*, with some specificity to clades of hosts with limited geographic distribution or biogeographic events that can provide time points in molecular dating studies.

Acknowledgment and funding information

This work was partly supported by The National Science Foundation (NSF) (grant NSF 2146653 to AAE). The authors thank Dr. Gediminas Valkiūnas for his comments on the early parts of this manuscript and the anonymous reviewers for their positive comments and valuable suggestions. We thank Adithyan Menon for helping with some recompilation data and Ariana Cristina Pacheco Negrin for designing the silhouettes.

References

1. Hewitt, R. (1940) *Bird malaria*, The Johns Hopkins Press, Baltimore
2. Garnham, PCC. (1966) *Malaria parasites and other Haemosporidia*, Oxford: Blackwell Science Ltd
3. Valkiūnas, G. (2005) *Avian malaria parasites and other haemosporidian*, Boca Raton (FL), CRC Press
4. Telford, S.R., Jr. (1983) *Saurocytozoon* Parasites (Haemosporidia: Plasmodiidae) from Southeast Asian Skinks. *J. Parasitol*. 69, 1141-1145
5. Lainson, R. (2012). Atlas of protozoan parasites of the Amazonian fauna of Brazil, Ananindeua: Instituto Evandro Chagas, , Editora IEC, Brazil
6. Telford, S.R., Jr. (1996) Two new species of *Haemocystidium* Castellani & Willey (Apicomplexa: Plasmodiidae) from Pakistani lizards, and the support their meronts provide for the validity of the genus. *Syst. Parasitol*. 34, 197-214

7. Telford, S.R., Jr. (2009) *Hemoparasites of the reptilia*, Boca Raton (FL), CRC Press, Taylor & Francis Group
8. Pineda-Catalan, O. *et al*. (2013) Revision of hemoproteid genera and description and redescription of two species of chelonian hemoproteid parasites. *J. Parasitol*. 99, 1089-1098
9. Lainson, R. *et al*. (1974) Further parasites of the family Garniidae (Coccidiida: Haemosporidiidea) in Brazilian lizards. *Fallisia effusa* gen. nov., sp. nov. and *Fallisia modesta* gen. nov., sp. nov. *Parasitol*. 68, 117-125
10. Gabaldon, G. et al. (1985) *Fallisia* (plasmodioides) *neotropicalis* subgen nov. sp. nov. from Venezuela. *Parasitol*. 90, 217-225
11. Poinar, G. Jr. and Telford, S.R., Jr. (2005) *Paleohaemoproteus burmacis* gen. n., sp. n. (Haemospororida: Plasmodiidae) from an Early Cretaceous biting midge (Diptera: Ceratopogonidae). *Parasitol*. 131, 79-84.
12. Poinar, G. Jr. (2011) *Vetufebrus ovatus* n. gen., n. sp. (Haemospororida: Plasmodiidae) vectored by a streblid bat fly (Diptera: Streblidae) in Dominican amber. *Parasit. Vectors*. 4, 229
13. Telford, S.R., Jr. (1973) Saurian malarial parasites from Guyana: Their effect upon the validity of the family Garniidae and the genus *Garnia*, with descriptions of two new species. *Int. J. Parasitol*. 3, 829-830
14. Pacheco, M.A. et al. (2011) Timing the origin of human malarias: the lemur puzzle. *BMC Evol. Biol*. 11, 299
15. Pacheco, M.A. *et al*. (2022) The evolution of primate malaria parasites: A study on the origin and diversification of *Plasmodium* in lemurs. *Mol. Phylogenet. Evol*. 174, 107551
16. Coatney, G.R. *et al*. (1971) *The primate malarias*, Bethesda, US National Institute of Allergy and Infectious Diseases
17. Cavalier-Smith, T. (2014) Gregarine site-heterogeneous 18S rDNA trees, revision of gregarine higher classification, and the evolutionary diversification of Sporozoa. *Eur. J. Protistol*. 50, 472-495
18. Valkiūnas, G. and Iezhova, T.A. (2018) Keys to the avian malaria parasites. *Malar. J*. 17, 212
19. Valkiūnas, G. and Iezhova, T.A. (2022) Keys to the avian *Haemoproteus* parasites (Haemosporida, Haemoproteidae). *Malar. J*. 21, 269
20. Zachos, F.E (2016) *Species Concepts in Biology: Historical Development, Theoretical Foundations and Practical Relevance*, Switzerland: Springer Nature
21. Santiago-Alarcon, D. and Marzal, A., eds (2020) *Avian Malaria and related parasites in the tropics. Ecology, evolution and systematics*, Springer.
22. Perkins, S.L. (2014) Malaria's many mates: past, present, and future of the systematics of the order Haemosporida. *J. Parasitol.* 100, 11-25
23. Bousema, T. *et al*. (2014) Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat. Rev. Microbiol*. 12, 833-840
24. Oliveira, L. *et al*. (2020) Molecular diversity and coalescent species delimitation of avian haemosporidian parasites in an endemic bird species of South America. *Parasitol. Res*. 119, 4033-4047

25. Pacheco, M.A. *et al*. (2022) Great-tailed Grackles (*Quiscalus mexicanus*) as a tolerant host of avian malaria parasites. *PLoS One*. 17, e0268161
26. Valkiūnas, G. *et al*. (2003) High prevalence of blood parasites in hawfinch *Coccothraustes coccothraustes*. *J. Nat. Hist*. 37, 2647–2652
27. Pérez-Tris, J. and Bensch, S. (2005) Diagnosing genetically diverse avian malarial infections using mixed-sequence analysis and TA-cloning. *Parasitol*. 131, 15-23
28. Silva-Iturriza, A. *et al*. (2012) Prevalence of avian haemosporidian parasites and their host fidelity in the central Philippine islands. *Parasitol. Int*. 61, 650-657
29. Pacheco, M.A. *et al*. (2013) Malarial parasite diversity in chimpanzees: the value of comparative approaches to ascertain the evolution of *Plasmodium falciparum* antigens. *Malar. J.* 12, 328
30. Muehlenbein, M.P. *et al*. (2015) Accelerated diversification of nonhuman primate malarias in Southeast Asia: adaptive radiation or geographic speciation? *Mol. Biol. Evol*. 32, 422-439
31. Clark, N.J. *et al*. (2016) Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. *J. Anim. Ecol*. 85, 1461-1470
32. Schaer, J. et al. (2013). High diversity of West African bat malaria parasites and a tight link with rodent *Plasmodium* taxa. *Proc. Natl. Acad. Sci. USA*. 110, 17415-17419
33. Outlaw, D.C. and Ricklefs, R.E. (2014) Species limits in avian malaria parasites (Haemosporida): how to move forward in the molecular era. *Parasitol.* 141, 1223-1232
34. Pacheco, M.A. *et al.* (2018) Mode and rate of evolution of haemosporidian mitochondrial genomes: Timing the radiation of avian parasites. *Mol. Biol. Evol*. 35, 383-403
35. Pacheco, M.A. *et al*. (2018) Primers targeting mitochondrial genes of avian haemosporidians: PCR detection and differential DNA amplification of parasites belonging to different genera. *Int. J. Parasitol*. 48, 657-670
36. Chumnandee, C. *et al.* (2021) Molecular characterization of *Polychromophilus* parasites of *Scotophilus kuhlii* bats in Thailand. *Parasitol*. 148, 495-499
37. Fecchio, A. *et al*. (2022) Beta diversity, prevalence, and specificity of avian haemosporidian parasites throughout the annual cycle of Chilean Elaenia (*Elaenia chilensis*), a Neotropical austral migrant. *Parasitol.* 149, 1760-1768
38. Collins, R.A. and Cruickshank, R.H. (2013) The seven deadly sins of DNA barcoding. *Mol. Ecol. Resour*. 13, 969-975
39. Falk, B.G. *et al*. (2011) Tree-based delimitation of morphologically ambiguous taxa: a study of the lizard malaria parasites on the Caribbean Island of Hispaniola. *Int. J. Parasitol*. 41, 967-980
40. González, A.D. et al. (2022). Is *Haemoproteus gabaldoni* a valid species? An approach from morphology and molecular tools applied to parasites of Anseriformes. *Acta Trop*. 233, 106540
41. Matta, N.E. *et al.* (2023) Morphometric and molecular characterization of an unpigmented haemosporidian parasite in the Neotropical turnip-tailed gecko (*Thecadactylus rapicauda*). *Parasitol*. 150, 221-229
42. Perkins, S.L. and Schall, J.J. (2002) A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J. Parasitol*. 88, 972-978

43. Ricklefs, R.E. and Fallon, S.M. (2002) Diversification and host switching in avian malaria parasites. *Proc. Biol. Sci*. 269, 885-892
44. Fallon, S.M. *et al*. (2003) Detecting avian malaria: an improved polymerase chain reaction diagnostic. *J. Parasitol*. 89, 1044-1047
45. Hellgren, O. *et al*. (2004) A new PCR assay for simultaneous studies of *Leucocytozoon*, *Plasmodium*, and *Haemoproteus* from avian blood. *J. Parasitol.* 90, 797-802
46. Bensch, S. *et al*. (2009) MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Mol. Ecol. Resour*. 9, 1353-1358
47. Palinauskas, V. *et al*. (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, 51-62
48. Lotta, I.A. *et al*. (2019) Disentangling *Leucocytozoon* parasite diversity in the neotropics: Descriptions of two new species and shortcomings of molecular diagnostics for leucocytozoids. *Int. J. Parasitol. Parasites Wildl*. 9, 159-173
49. Valkiūnas, G. *et al*. (2021). Description of *Haemoproteus asymmetricus* n. sp. (Haemoproteidae), with remarks on predictability of the DNA haplotype networks in haemosporidian parasite taxonomy research. *Acta Trop*. 218, 105905
50. Vanstreels, R.E.T. *et al*. (2022) A new haemosporidian parasite from the Red-legged Seriema *Cariama cristata* (Cariamiformes, Cariamidae). *Int. J. Parasitol. Parasites Wildl*. 18, 12-19
51. Page, R.D. (2016) DNA barcoding and taxonomy: dark taxa and dark texts. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 371, 1702
52. Levin, I.I. *et al*. (2012) Novel *Haemoproteus* species (Haemosporida: Haemoproteidae) from the swallow-tailed gull (Lariidae), with remarks on the host range of hippoboscid-transmitted avian hemoproteids. *J. Parasitol*. 98, 847-854
53. Sutherland, C. J. et al. (2010) Two nonrecombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J. Infect. Dis.* 201,1544–1550
54. Valkiūnas, G. *et al*. (2019) Molecular characterization of six widespread avian haemoproteids, with description of three new *Haemoproteus* species. *Acta Trop*. 197, 105051
55. Bernotienė, R. *et al*. (2016) Avian haemosporidian parasites (Haemosporida): A comparative analysis of different polymerase chain reaction assays in detection of mixed infections. *Exp. Parasitol*. 163, 31-37
56. Ciloglu, A. *et al*. (2022) A novel one-step multiplex PCR protocol to detect avian haemosporidian parasites in the subgenus *Haemoproteus* (Kruse, 1890) used to quantify parasite prevalence in domestic pigeons (*Columba livia*) in Turkey. *Vet. Res. Commun*. doi: 10.1007/s11259-022-09962-z
57. Krief, S. *et al*. (2010) On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. *PLoS Pathog*. 6, e1000765
58. Liu, W. *et al*. (2017) Wild bonobos host geographically restricted malaria parasites including a putative new *Laverania* species. *Nat. Commun*. 8, 1635.

59. Snounou, G. *et al*. (1993). Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of a high prevalence of mixed infections. Mol. Biochem. *Parasitol*. 58, 283-292
60. Waters, A.P. *et al*. (1991). *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proc. Natl. Acad. Sci. USA*. 88, 3140-3144
61. Escalante, A.A. and Ayala, F.J. (1994) Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl. Acad. Sci. USA*. 91, 11373-11377
62. Corredor, V. and Enea, V. (1994) The small ribosomal subunit RNA isoforms in *Plasmodium cynomolgi*. *Genetics*. 136, 857-865
63. Harl, J. *et al*. (2019). The nuclear 18S ribosomal DNAs of avian haemosporidian parasites. *Malar. J.* 18, 305
64. Escalante, A.A. *et al.* (1995) Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. *Mol. Biol. Evol*. 12, 616-626.
65. McCutchan, T.F. *et al*. (1996) Comparison of circumsporozoite proteins from avian and mammalian malarias: biological and phylogenetic implications. *Proc. Natl. Acad. Sci. USA.* 93, 11889-11894
66. Singh, B. *et al*. (2004) A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet.* 363, 1017-1024
67. Escalante, A.A. *et al*. (1998) The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. *Proc. Natl. Acad. Sci. USA*. 95, 8124-8129
68. Bensch, S. *et al*. (2004) Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution* 58, 1617-1621
69. Escalante, A.A *et al*. (2005) A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc. Natl. Acad. Sci. USA*. 102, 1980-1985
70. Martinsen, E.S. *et al*. (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, 261-273
71. Witsenburg, F. *et al*. (2012) The evolutionary host switches of *Polychromophilus*: a multigene phylogeny of the bat malaria genus suggests a second invasion of mammals by a haemosporidian parasite. *Malar. J*. 11, 53
72. Schaer, J. et al. (2015) *Nycteria* parasites of Afrotropical insectivorous bats. *Int. J. Parasitol.* 45, 375-384
73. Borner, J. *et al*. (2016) Phylogeny of haemosporidian blood parasites revealed by a multigene approach. *Mol. Phylogenet. Evol*. 94, 221-231
74. Lutz, H.L. *et al*. (2016) Diverse sampling of East African haemosporidians reveals chiropteran origin of malaria parasites in primates and rodents. *Mol. Phylogenet. Evol*. 99, 7-15
75. Templeton, T.J. *et al*. (2016) Ungulate malaria parasites. *Sci. Rep*. 6, 23230
76. Galen, S.C. et al. (2018) The polyphyly of *Plasmodium*: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *R. Soc. Open Sci.* 5, 171780

77. Escalante, A.A. *et al*. (2022) Why *Plasmodium vivax* and *Plasmodium falciparum* are so different? A tale of two clades and their species diversities. *Malar. J*. 21, 139
78. Tsague, K.J.A. *et al*. (2022) *Hepatocystis* and *Nycteria* (Haemosporida) parasite infections of bats in the Central Region of Cameroon. *Parasitol*. 149, 51-58
79. Hayakawa, T. *et al*. (2008) Big bang in the evolution of extant malaria parasites. *Mol. Biol. Evol*. 25, 2233-2239
80. Perkins, S.L. (2008) Molecular systematics of the three mitochondrial protein-coding genes of malaria parasites: corroborative and new evidence for the origins of human malaria. *Mitochondrial DNA*. 19, 471-478
81. Xi, Z. *et al*. (2016) The impact of missing data on species tree estimation. *Mol. Biol. Evol*. 33, 838-860
82. Greiner, E.C. *et al.* (2011) Hematozoa and a new haemoproteid species from Cathartidae (New World vulture) in South Carolina. *J. Parasitol*. 97, 1137-1139
83. Bertram, M.R. *et al*. (2017) A novel Haemosporida clade at the rank of genus in North American cranes (Aves: Gruiformes). *Mol. Phylogenet. Evol*. 109, 73-79
84. Nunn, C.L. et al. (2011) Mutualism or parasitism? Using a phylogenetic approach to characterize the oxpecker-ungulate relationship. *Evolution*. 65, 1297-1304
85. Pineda-Catalan, O. *et al*. (2013) Revision of hemoproteid genera and description and redescription of two species of chelonian hemoproteid parasites. *J. Parasitol.* 99, 1089-1098
86. Maia, J.P. *et al*. (2016) Reconstruction of the evolutionary history of Haemosporida (Apicomplexa) based on the *cyt b* gene with characterization of *Haemocystidium* in geckos (Squamata: Gekkota) from Oman. *Parasitol. Int*. 65, 5-11
87. González, LP *et al*. (2019) *Haemocystidium* spp., a species complex infecting ancient aquatic turtles of the family Podocnemididae: First report of these parasites in *Podocnemis vogli* from the Orinoquia. *Int. J. Parasitol. Parasites Wildl*. 10, 299-309
88. Pacheco, M.A. *et al*. (2020) A phylogenetic study of *Haemocystidium* parasites and other Haemosporida using complete mitochondrial genome sequences. *Infect. Genet. Evol*. 85, 104576
89. Ishtiaq, F. et al. (2008) Avian haematozoan parasites and their associations with mosquitoes across Southwest Pacific Islands. *Mol. Ecol*. 17, 4545-4555
90. Bukauskaitė, D. et al. (2015) Biting midges (Culicoides, Diptera) transmit *Haemoproteus* parasites of owls: evidence from sporogony and molecular phylogeny. *Parasit. Vectors*. 8, 303
91. Outlaw, D.C. and Ricklefs, R.E. (2011) Rerooting the evolutionary tree of malaria parasites. *Proc. Natl. Acad. Sci. USA*. 108, 13183-13187
92. Ayala, F.J. and Fitch, W.M. (1992) Phylogeny of *Plasmodium falciparum*. *Parasitol. Today*. 8, 74-75
93. Escalante, A.A and Ayala, F.J. (1995) Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proc. Natl. Acad. Sci. USA*. 92, 5793-5797
94. Ricklefs, R.E. and Outlaw, D.C. (2010) A molecular clock for malaria parasites. *Science.* 329: 226-229

95. Silva, J.C. *et al*. (2015) A new method for estimating species age supports the coexistence of malaria parasites and their Mammalian hosts. *Mol. Biol. Evol*. 32, 1354-1364
96. Bromham, L. (2019) Six Impossible Things before Breakfast: Assumptions, models, and belief in molecular dating *Trends Ecol. Evol*. 34, 474-486
97. Poinar, G., Jr. (2005). *Plasmodium dominicana* n. sp. (Plasmodiidae: Haemosporida) from Tertiary Dominican amber. *Syst. Parasitol*. 611, 47–52
98. Ramiro, R.S. et al (2012) Molecular evolution and phylogenetics of rodent malaria parasites. *BMC Evol. Biol*. 12, 219
99. Benton, M.J. et al. (2015) Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron* 18.1.1FC, 1–107
100. Prum, RO *et al*. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. 526, 569-573

Glossary

Bayesian methods: The Bayesian phylogenetic approach combines a tree's prior probability with the data's likelihood to estimate a posterior probability of parameters such as tree topologies and mutation rates. Parameters are not point estimates but rather part of a distribution. The approach can also estimate the time of origin for clades in phylogenies, given calibration constraints as priors.

Calibration constraints: Also called calibration points are evidence for the time of origin of a given clade in a phylogenetic tree used to inform estimates for evolutionary rates. Calibrations are independently provided from a fossil or a biogeographic event. Under a Bayesian framework, calibrations involve a range and a prior distribution (e.g., uniform, exponential).

Chimeras: An artifact in terms of a DNA sequence or morphological entity that erroneously contains information from at least two or more lineages/species present in a given sample.

Credibility interval: In Bayesian statistics, it is an interval within which an unobserved parameter value falls with a particular probability (usually 95%), given the evidence provided by the observed data.

660 **Cryptic species:** Species that cannot be separated morphologically (e.g., the *Leucocytozoon toddi*
661 group).
662 **DNA barcoding:** A method of species identification that uses a genetic marker, usually a short
663 fragment of a locus, to identify a specimen as part of a species or lineage through the comparison
664 with lineages or species documented in a database (e.g., partial *cytb* gene (480bp) sequences using
665 the MalAvi database).
666 **Phenotypic plasticity:** Changes in behavior, morphology, and physiology by an organism in
667 response to environmental stimuli or inputs.
668 **Fossil:** The remains or impression of an organism preserved in petrified form or as a mold or cast
669 in a rock.
670 **Host:** An organism that serves as a habitat for a symbiont where it completes the totality or part
671 of its life cycle.
672 **Long-branch attraction effect:** Arises when the amount of molecular or morphological change
673 accumulated within a lineage is sufficient to cause that, by chance, it appears similar (thus closely
674 related) to another long-branched lineage. Thus, the apparent shared evolutionary history is an
675 artifact of lineages that have undergone a large amount of change rather than an indication that
676 they are related by descent. Such bias is more common when the overall divergence of some taxa
677 results in long branches within a phylogeny.
678 **Lineage:** A continuous line of descent (identical by descent) usually inferred from molecular data.
679 It can be a species or allele (e.g., Haemosporidian *cytb* gene lineages).
680 **Molecular clock:** A model for estimating the time to the common ancestor of two or more
681 lineages. It uses an estimate of the evolutionary rate of nucleic acids or proteins to translate the
682 taxa molecular divergence into absolute time.

683 **Monophyletic group:** It is a group or clade of organisms that includes the most recent common
684 ancestor of the group and its entire descendent species.
685 **Polyphyletic group:** It is a species group that does not include the common ancestor of all group
686 members. As such, it does not reflect the group's evolutionary history.
687 **Purifying selection (or negative selection)** occurs when natural selection removes deleterious
688 alleles (lower reproductive success) that arise through random mutations.
689 **Species:** These are the basic units used in the taxonomic classification of biological diversity.
690 **Taxon (pl. taxa):** Organisms sharing traits that receive a formal taxonomic name. It is expected
691 to be a monophyletic group.
692
693 **Text boxes**
694 **Box 1.**
695 **Haemosporidian phylogenetic hypotheses**

696 Although several phylogenetic relationships are still unsolved, the most complete and recent
697 analyses [15,34,88] have shown:
698 (1) The three genera, *Leucocytozoon*, *Haemoproteus*, and *Plasmodium*, are not monophyletic
699 groups (Figure 2). *Plasmodium* is polyphyletic, part of a clade with the genera *Hepatocystis*,
700 *Nycteria*, and *Polychromophilus*. We cannot infer whether the common ancestor of such a
701 clade can be considered a *Plasmodium*.
702 (2) The positions of genera *Hepatocystis*, *Nycteria,* and *Polychromophilus* are unsolved
703 [15,34,73,74,88]. Differences regarding the relationship of *Hepatocystis* with other taxa may
704 be due to including lemur malarial parasites in some of the analyses [14,15,34]. Lemur
705 parasites seems to be a very diverse clade that shares a common ancestor with other primate
706 parasites from continental Africa. Genus *Nycteria* has been recovered as part of a mammalian

clade comprising *Plasmodium* and *Hepatocystis* or sister to reptilian *Plasmodium* [34,88]. Genus *Polychromophilus* may share a common ancestor with *Plasmodium* species found in ungulates [34,70,75,88] (Figure 2b). Increasing the taxa sampling may help to solve this discrepancy.

(3) *Haemocystidium* species found in reptiles are a monophyletic group [85-88].

(4) Phylogenetic analyses support the two *Haemoproteus* subgenera, *Haemoproteus* and *Parahaemoproteus* [3,70], as they are monophyletic groups (Figure 2). Lineages such as *H. catharti*, *H. antigonis,* and *H. pulcher* have been described as *Haemoproteus*. However, their phylogenetic relationships with the other haemosporidian must be elucidated as *H. catharti* appears out of the *Haemoproteus* clade (Figure 2).

(5) *Leucocytozoon* (*Akiba*) *caulleryi* is not part of the monophyletic group that includes other *Leucocytozoon* species [34]. It shares a common ancestor with *H.* (*Parahaemoproteus*) and is within a well-supported clade that includes *Haemocystidium* and *Plasmodium* species. As a result, the mtDNA phylogeny indicates that the genus *Leucocytozoon* is also polyphyletic. Unfortunately, there is no data for *L. caulleryi* on nuclear genes.

(6) There are two hypotheses regarding the root of the Haemosporida phylogeny. One is estimated by mitochondrial data when no outgroup is included [15]. This estimated root appears as a common ancestor shared by *Leucocytozoon* (*Leucozytozoon*) sp. and all the other genera (Figure 2A). The second uses nuclear data and a piroplasm as an outgroup [76]. It places the root between a clade with most species of *Haemoproteus* as sister taxa to *Leucocytozoon* (*Leucocytozoon*), separated from the other Haemosporida (Figure 2B).

730 **Tables**

731 **Table 1 Characteristics of the Haemosporida families, genera and subgenera.**

Taxa	Invertebrate hosts (vectors)	Vertebrate hosts[e]	Characteristics	Refs.
Family Leucocytozoidae Fallis and Bennett, 1961				
Genus *Leucocytozoon* Berestneff, 1904				
Subgenus *Akiba* Bennett, Garnham and Fallis, 1965	Biting midges (Ceratopogonidae)	Domestic and wild chicken *Gallus gallus* (1 sp.)	Merogony in blood: No With hemozoin pigment: No Gametocytes develop in erythrocytes: In some species, gametocytes develop in erythrocytes, leukocytes, and thrombocytes. In *Akiba*, gametocytes develop only in erythrocytes	[3]
Subgenus *Leucocytozoon* Berestneff, 1904	Black flies (Simuliidae)	Birds (>100 spp.)		
Genus *Saurocytozoon* Lainson and Shaw, 1969[a]	Unknown	Lizards (2-? sp.)	Merogony in blood: No, only in tissues With hemozoin pigment: No Gametocytes develop in erythrocytes: Lymphocytes and reticulocytes.	[4,5]
Family Haemoproteidae Doflein, 1916				
Genus *Haemoproteus* Kruse, 1890				
Subgenus *Haemoproteus* Kruse, 1890	Hippoboscid flies (Hippoboscidae)	Birds (>10 spp.)		[3]
Subgenus *Parahaemoproteus* Bennett, Garnham and Fallis, 1965	Biting midges (Ceratopogonidae)	Birds (>170 spp.)		
Genus *Haemocystidium* Castellani and Willey 1904, emend. Telford, 1996			Merogony in blood: No With hemozoin pigment: Yes Gametocytes develop in erythrocytes: Yes	
Subgenus *Haemocystidium* Castellani and Willey, 1904, emend. Telford, 1996	Unknown	Lizards and snakes (>14-? spp.)		[2,6-8]
Subgenus *Simondia* Garnham, 1966	Tabanid fly (Tabanidae)?	Turtle (>5 spp.)		
Genus *Hepatocystis* Levaditi and Schoen, 1932 emend. Garnham, 1948	Biting midges (Ceratopogonidae: Culicoides)	Bats, monkeys, squirrels, and ungulates (>25-? spp.)	Merogony in blood: No (only liver) With hemozoin pigment: Yes Gametocytes develop in erythrocytes: Yes	[2]

Genus *Nycteria* Garnham and Heisch, 1953	Unknown	Bats (>7-? spp.)	Merogony in blood: No With hemozoin pigment: Yes Gametocytes develop in erythrocytes: Yes	[2]
Genus *Polychromophilus* Dionisi, 1899	Bat flies (Nycteribiidae)	Bats (5-? spp.)	Merogony in blood: No With hemozoin pigment: Yes Gametocytes develop in erythrocytes: Yes	[2]
Family Garniidae Lainson, Landau and Shaw, 1974[b]				
Genus *Fallisia* Lainson, Landau and Shaw, 1974				[9]
Subgenus *Fallisia* Lainson, Landau and Shaw, 1974	Unknown	Lizards (>10-? spp.)	Merogony in blood: Yes With hemozoin pigment: No Gametocytes develop in erythrocytes: *Plasmodioides:* Yes, and thrombocytes, lymphocytes, and/or monocytes *Fallisia*: No *Garnia*: Yes *Progarnia:* No, only in lymphocytes, monocytes, and thrombocytes	
Subgenus *Plasmodioides* Gabaldon, Ulloa and Zerpa, 1985	*Aedeomyia squamipennis* (Culicidae)?	Domestic pigeon, *Columba livia* (1 sp.)		[10]
Genus *Garnia* Lainson, Landau and Shaw, 1971[c]	Unknown	Lizards (>10-? spp.)		[9]
Genus *Progarnia* Lainson, 1995	Unknown	Crocodilians (1 sp.)		[5]
Family Plasmodiidae Mesnil, 1903				
Genus *Plasmodium* Marchiafava and Celli, 1885				
Subgenus *Plasmodium* Marchiafava and Celli, 1885	Mosquitoes (Culicidae: Anophelinae)	Primates (>30-? spp.)		
Subgenus *Laverania* Grassi and Feletti, 1890	Mosquitoes (Culicidae: Anophelinae)	Apes (>8-? spp.)	Merogony in blood: Yes With hemozoin pigment: Yes Gametocytes in erythrocytes: Yes (and for some species reticulocytes)	
Subgenus *Vinckeia* Garnham, 1964[d]	Mosquitoes (Culicidae: Anophelinae)	Non-primate mammals (+Lemurs?) (>15-? spp.)		
Subgenus *Haemamoeba* Grassi and Feletti, 1890	Mosquitoes (Culicidae: Culicinae)	Birds (>12 -? spp.)		

Subgenus *Bennettinia* Valkiūnas, 1997	Mosquitoes (Culicidae: Culicinae)	Birds (1 sp.)	[2,3]
Subgenus *Giovannolaia* Corradetti, Garnham and Laird, 1963	Mosquitoes (Culicidae: Culicinae)	Birds (>16 -? spp.)	
Subgenus *Huffia* Garnham and Laird, 1963	Mosquitoes (Culicidae: Culicinae)	Birds (>4-? spp.)	
Subgenus *Novyella* Corradetti, Garnham and Laird, 1963	Mosquitoes (Culicidae: Culicinae)	Birds (>22-? spp.)	
Subgenus *Carinamoeba* Garnham, 1966	Unknown	Reptiles (>8-? spp.)	
Subgenus *Ophidiella* Garnham, 1966		Snakes (> 5 spp.?)	
Subgenus *Sauramoeba* Garnham, 1966	Mosquitoes (Culicidae: Culicinae)?	Reptiles (>15-? spp.)	
Subgenus *Asiamoeba* Telford, 1988		Reptiles (6 spp.)	[7]
Subgenus *Lacertamoeba* Telford, 1988		Reptiles (>40 spp.)	[7]
Subgenus *Paraplasmodium* Telford, 1988	Sandfly (Phlebotominae)	Reptiles (3 spp.)	[7]

Note: Other suggested genera with limited information.

Family Haemoproteidae Doflein, 1916

Genus *Rayella* (synonym of *Hepatocystis*?) Dasgupta, 1967: host are flying squirrel.
Genus *Dionisia* (synonym of *Polychromophilus*) Landau, Chabaud, Miltgen, Baccam, 1980: host are bats.
Genus *Bioccala* (synonym of *Polychromophilus*?) Landau, Baccam, Ratanaworabhan, Yenbutra, Boulard, Chabaud, 1984: host are bats.
Genus *Biguetiella* (synonym of *Nycteria*?) Landau, Baccam, Ratanaworabhan, Yenbutra, Boulard, Chabaud, 1984: host are *Hipposideros larvatus*.
Genus †*Paleohaemoproteus* Poinar and Telford, 2005 [11]: found in a female biting midge.
Genus *Johnsprentia* Landau, Chavatte and Beveridge, 2012: hosts are flying foxes.
Genus *Sprattiella* Landau, Chavatte, Karadjian, Chabaud, and Beveridge, 2012: hosts are flying foxes.

Family Plasmodiidae Mesnil, 1903

Genus *Mesnilium* Misra, Haldar and Chakravarty, 1972: host are fishes.
Genus *Billbraya* (synonym of *Plasmodium*?) Paperna and Landau, 1990: host are lizards.
Genus *Plasmodium*, Subgenus *Papernaia* (synonym of *Novyella*?) Landau, Chavatte, Peters, Chabaud, 2010: host are birds.
Genus †Vetufebrus (synonym of *Polychromophilus*?) Poinar, 2011 [12]: host could be bats?

[a]Telford [13] considered it a synonym of Plasmodiidae.

[b]Telford [13] considered it a synonym of Plasmodiidae.

[c]Telford [13] considered it a synonym of genus *Plasmodium.*

[d]Subgenus *Vinckeia* [2] was defined as "malaria parasites" found in mammals (including lemurs) below the simian level; however, molecular data suggested that lemur parasites do not conform a monophyletic group with the other species of the subgenus *Vinckeia* [14,15].

[e]Minimum number of described morphospecies.

†Considered as an extinct genus of Haemosporida.

? It is uncertain the total existing species or there is not information available.

732

733 **Figure legends**

734 **Figure 1 General representation of Haemosporida life cycle.** Vector (in blue) and vertebrate
735 (green) parasite stages are shown.

736 **Figure 2 Most recent haemosporidian phylogenetic hypotheses.** (a) Haemosporidian
737 phylogenetic hypotheses based on mitochondrial genomes (≈ 6kb, [15,34]) and (b) nuclear genes
738 [76]. Figures adapted from [15,76].

739 **Figure 3 Timetree of the divergence of the major clade of Haemosporida.** Divergence times
740 were estimated based on the mitochondrial genome using BEAST with the following calibration
741 constraints: (1) The minimum divergence of Macaca/Papio using fossils (6–14.2 Ma), (2) a
742 maximum of 24.44 to a minimum of 34.0 Ma for the human/Macaca split, (3) a range of 20–42
743 Ma for the origin of the lemur parasites, and (4) a range of 16–28.1 Ma proposed for the origin of
744 Bovinae-Antilopinae. Times are shown in My, and the 95% credibility intervals (CrI) for the major
745 clades are gray. Branch/box colors indicate the genus–vector host species relationship. Figures
746 adapted from [15]. Also see [14,15,34] for additional information.

747

Outstanding questions

Can rigorous molecular standards for species delimitation facilitate our understanding of haemosporidian diversity and evolution?

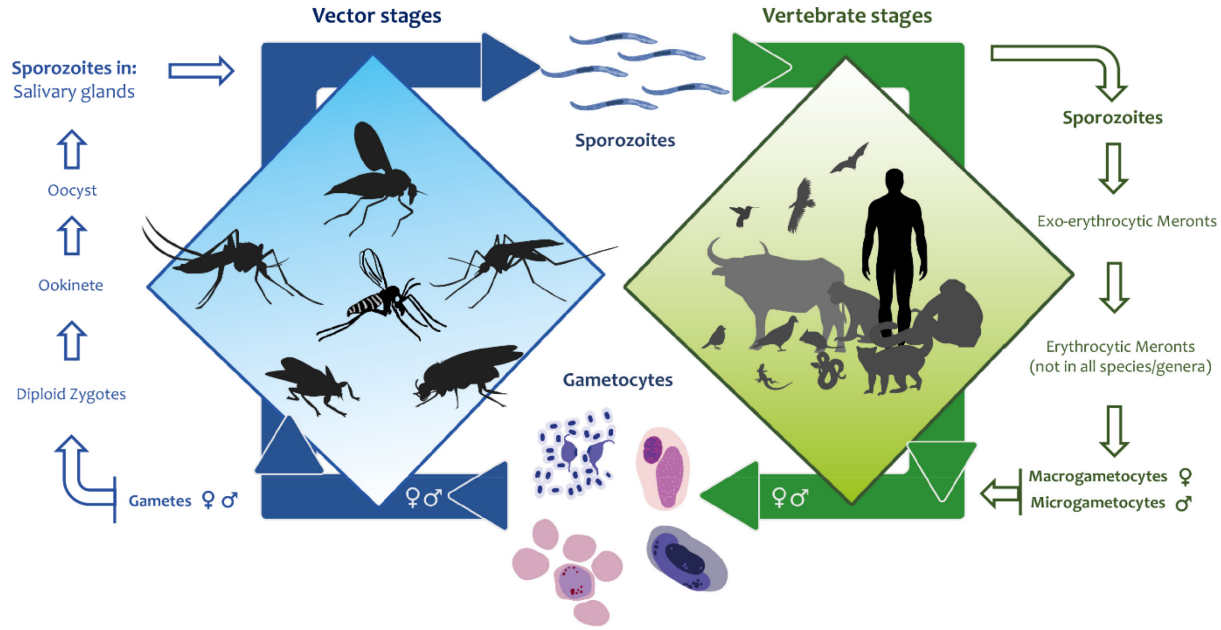
Does the biased taxonomic sampling affect our understanding of haemosporidian diversity?

Are genera without molecular data monophyletic groups?

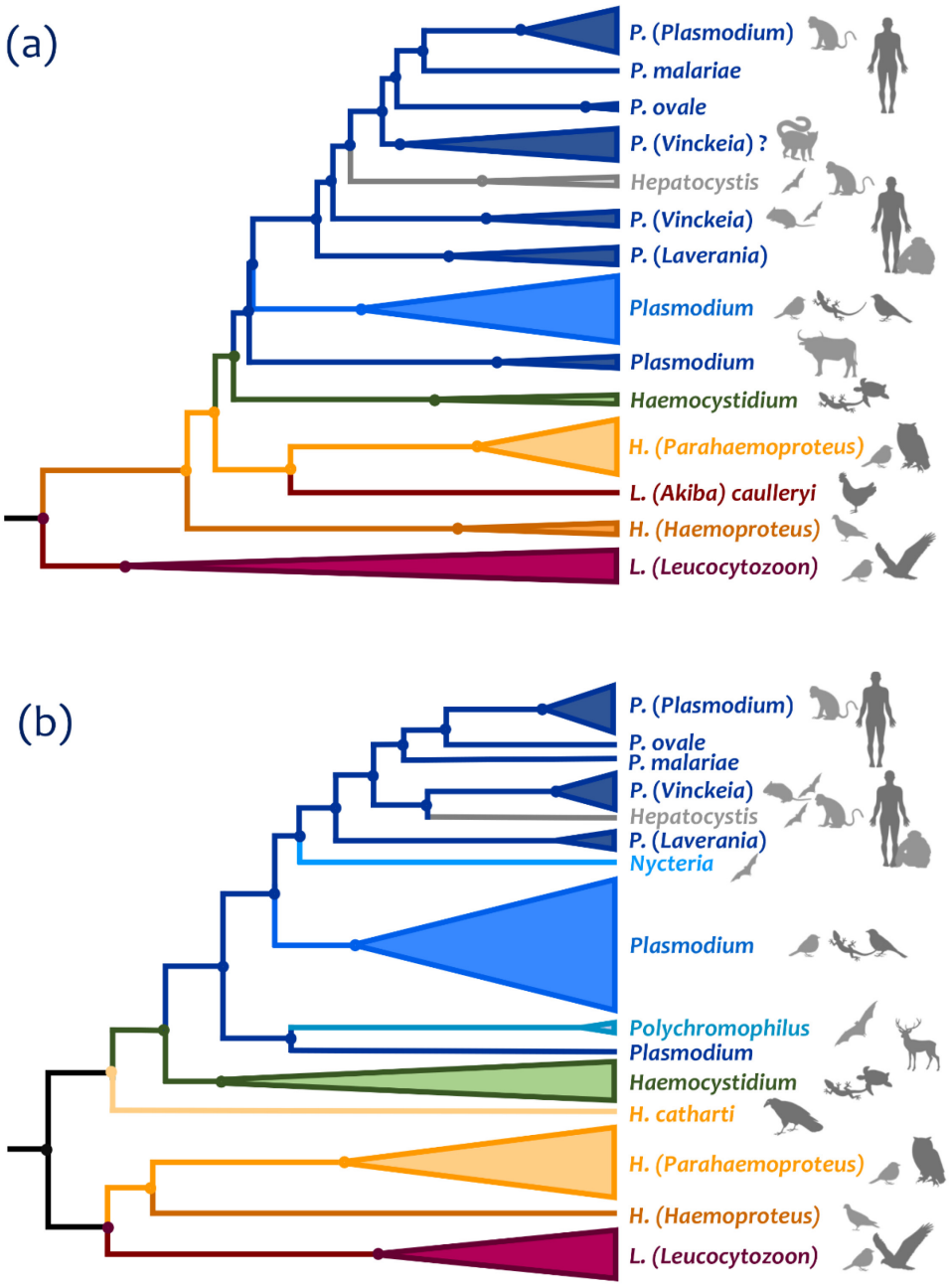
Can the incongruence among phylogenetic studies be explained solely by taxa sampling?

Can calibrations be discovered outside the genus *Plasmodium*?

Figure 1 **General representation of Haemosporida life cycle.** Vector (in blue) and vertebrate (green) parasite stages are shown.



761 Figure 2 **Most recent haemosporidian phylogenetic hypotheses.** (a) Haemosporidian
762 phylogenetic hypotheses based on mitochondrial genomes (≈ 6kb, [15,34]) and (b) nuclear genes
763 [76]. Figures adapted from [15,76].



764

765

Figure 3 **Timetree of the divergence of the major clade of Haemosporida.** Divergence times were estimated based on the mitochondrial genome using BEAST with the following calibration constraints: (1) The minimum divergence of Macaca/Papio using fossils (6–14.2 Ma), (2) a maximum of 24.44 to a minimum of 34.0 Ma for the human/Macaca split, (3) a range of 20–42 Ma for the origin of the lemur parasites, and (4) a range of 16–28.1 Ma proposed for the origin of Bovinae-Antilopinae. Times are shown in My, and the 95% credibility intervals (CrI) for the major clades are gray. Branch/box colors indicate the genus–vector host species relationship. Figures adapted from [15]. Also see [14,15,34] for additional information.