REVIEW ARTICLE

Species limits in avian malaria parasites (Haemosporida): how to move forward in the molecular era

DIANA C. OUTLAW1* and ROBERT E. RICKLEFS2

1Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39759, USA
2Department of Biology, University of Missouri-St. Louis, One University Boulevard, St. Louis, MO 63121, USA

(Received 4 November 2013; revised 20 December 2013, 3 March and 7 March 2014; accepted 13 March 2014; first published online 9 May 2014)

SUMMARY

Delimiting species of malaria parasites (Haemosporida) has become increasingly problematic as new lineages of parasites are identified solely by molecular information, particularly mitochondrial cytochrome b sequence data. In this review, we highlight some of the issues, both historical and contemporary, that have hindered the development of objective criteria to diagnose, delimit and define species of haemosporidians. Defining species is not the focal interest of most researchers, most of whom merely wish to determine whether lineages identified in their samples match those of other researchers, and if so, where and in which host species. Rather than revisiting all the issues with respect to delimiting and naming species, we instead focus on finding a practical near-term resolution to the ‘species problem’ that utilizes the community’s largest resource: mitochondrial cytochrome b DNA sequences. We recommend a standardized procedure to ‘tag’ these sequences, based on per cent sequence similarity, that will allow researchers to directly assess the novelty, known hosts and geographic distribution of avian malaria parasite lineages.

Key words: Avian malaria, species limits, Haemosporida.

THE SPECIES PROBLEM: WHY NOW?

To understand evolutionary processes that have generated the biodiversity of haemosporidian malaria parasites, we need to circumscribe units of biodiversity – species or, at least, operational taxonomic units (OTUs). Recent discussion of the evolutionary history and diversification of these parasites has been encumbered by disagreement over metrics that can (or should) be used to diagnose and delimit species. Although human malaria parasite species are well-known and well-studied, hundreds of morphologically described species of malaria parasites infect birds (~200), lizards (~150) and non-anthropoid mammals (~20 known), and an increasing number of genetic lineages (i.e. cytochrome b haplotypes) have yet to be attributed to formally described species. Molecular studies over the last 20 years, beginning with Escalante and Ayala (1994), have revealed a previously unexpected level of diversity in these parasites, particularly those of birds, that rivals that of their host species (e.g. Bensch et al. 2000, 2004; Ricklefs and Fallon, 2002; Ricklefs et al. 2004; Beadell et al. 2006). Indeed, the number of papers per year that identify genetic lineages of avian malaria parasites has increased rapidly since 1996 (Scopus ‘avian malaria, cytochrome b’ search by year; Figure S1), and more than 1000 mitochondrial cytochrome b (cyt b) sequences have been recognized among avian malaria parasites (~1500 as of February, 2014; MalAvi database; Bensch et al. 2009). Although disagreement continues about the relevance of sequence differences to species boundaries in protists and bacteria (Adl et al. 2007; Morrison, 2009; Pérez-Ponce de Léon and Nadler, 2010), molecular studies of parasite diversity continue at an accelerating pace. As genomic data become increasingly available, it is timely to discuss the application of molecular tools to defining species limits. Certainly, delimiting species is prerequisite to understanding diversity, distribution and speciation in avian malaria parasites.

Formal descriptions of haemosporidian parasites, with the exception of Plasmodium, are based primarily on limited morphological characters observed in the mature gametocytes preserved on blood smears (e.g. Perkins, 2000; Valkiūnas et al. 2009a, b; Paperna and Yosef, 2010; reviewed in Valkiūnas, 2005). Indeed, over the history of haemosporidian taxonomy, opinions have differed concerning the utility of formally describing a malaria parasite, particularly when its species status is uncertain. Alternatively, one could provide an ambiguous designation,

doi:10.1017/S0031182014000560
e.g. Plasmodium bird.1, which might relegate the
entity to anonymity and minimize its potential
contribution to our understanding of malaria parasite
biodiversity (see Garnham, 1966). It is clear from
the amount of molecular data that has accumulated over
the last 10 years, and from our ambiguity about cir-
cumventing species, or even OTUs, that we remain in
uncharted territory with regard to taxonomically
informative characters for delimiting species (re-
viewed in Perkins et al. 2011). An examination of the
literature shows that several new species of malaria
parasite are described every year and that the criteria
used to define them can differ markedly. In some
cases, species limits are established solely by mol-
ecular data (Rich et al. 2009; Merino et al. 2012; see
also Valkiūnas et al. 2011), while other authors use
only morphology (Zechtindijev et al. 2012), or a com-
bination of both (e.g. Valkiūnas et al. 2009a, b; Falk
et al. 2011).

The Biological Species concept is difficult to eval-
uate in malaria parasites, and, instead, most research-
ers have used variations of the morphological and
phylogenetic species concepts. The morphological
species concept is essentially a similarity species con-
cept (see Garnham, 1966; Valkiūnas, 2005); closely
related to this is the genetic similarity species concept
(Martinsen et al. 2006). The phylogenetic species
concept has been applied by evaluating reciprocal
monophyly between cyt b sequences of named line-
ages (Martinsen et al. 2006). However, surprisingly
few studies of malaria parasites in wildlife have
explicitly addressed agreement across these species
concepts: morphological similarity vs phylogenetic
similarity (reciprocal monophyly) vs genetic simi-
licity. The most thorough is Martinsen et al. (2006),
who evaluated five avian haemosporidian clades
using these three species concepts. In only one case
was there discordance across concepts, i.e. in which a
‘species’ was paraphyletic; some species included
more than one cyt b sequence.

Quality blood smears and sequence data have oc-
casionally provided morphological corroboration of
‘cryptic’ phylogenetic species (e.g. Valkiūnas et al.
2010). One of the most compelling examples involved
Leucocytozoon toddi (Sehgal et al. 2006), in which a
species described on the basis of a single morphotype
was found to represent two genetic lineages present
in different host genera, one in Buteo spp. and one in
Accipiter spp. This marked division caused a re-
evaluation of blood smears from these host groups,
and morphological differences were in fact discovered
between these two lineages (Valkiūnas et al. 2010).
These studies were followed by others showing diver-
sity within other Leucocytozoon morphospecies
(Ishak et al. 2008; Krone et al. 2008). In an extensive
treatment of Plasmodium relictum in birds, Beadell et
al. (2006) found P. relictum to be composed of
multiple, reciprocally monophyletic lineages that
are nonetheless quite closely related to one another.

In a recent study of lizard parasites, Falk et al. (2011)
found four new species of haemosporidian that were
concordant across multi-locus phylogenetic
data, but for which morphological characters were
overlapping and somewhat ambiguous. Finally, in a
very striking recent example, Merino et al. (2012)
describe a new avian parasite species that is geneti-
cally identical in cyt b sequence to another species but
morphologically distinct within one host species
(see ‘Historical philosophies and approaches’ below
for a possible explanation).

AN OVERVIEW OF AVIAN HAEMOSPORIDIAN
DIVERSITY

Malaria parasites—more correctly, haemosporidian
parasites—include six well-known genera of api-
complexans: Plasmodium (parasitic with respect to
mammal vs bird and lizard Plasmodium; Leucocytozoon, Haemoproteus and Parahaemoproteus
(birds); Hepatocystis (mammals) and Polychromophilus (bats), which, combined, infect all amniotes.
Less well-known genera that are not discussed in this
review include Fallisia (Garniidae; Valkiūnas, 2005,
pp. 731) of birds and lizards and the plasmodiids
Although the anthropoid Plasmodium parasites
have received the most attention, owing to their
impact on human health, known malaria parasite
biodiversity is broadest among bird species and popu-
lations. This is primarily due to the long tradition
of taking blood smears and blood samples in field
studies of birds, a practice that is less common among
mammalogists. The earliest work on malaria con-
cerned avian parasites (reviewed in Valkiūnas, 2005,
pp. 9–15), of which the c. 200 described species are
based on morphological characters and/or host taxa.
In contrast, 1517 cytchrome b haplotypes have
been described worldwide, although only 122 are
associated with formally described species (MalAvi;
Bensch et al. 2009). It is important to note here
that avian malaria researchers use different metrics
to define lineages, ranging from a single nucleotide
difference (Bensch et al. 2000, 2004; Hellgren, 2005)
to a combination of genetic differentiation and
host/geographic distribution (Ricklefs et al. 2005;
Svensson-Coelho et al. 2013).

The diversity of avian malaria parasites probably
greatly exceeds the 1500 molecularly defined lineages
known at present. Based on current sampling, avian
malaria parasites are sister to mammalian Plasmodium
and are much more diverse than known mammalian
parasites in terms of vector usage, levels of host
specificity, and geographic range. We have argued
that the most recent common ancestor of extant,
avian malaria parasite genera dates to only 9 MYA
(Ricklefs and Outlaw, 2010)—certainly not much
more than that (Bensch et al. 2013)—and that their
incredible diversity has been achieved primarily
HISTORICAL PHILOSOPHIES AND APPROACHES

The three most fundamental treatises on malaria parasite diversity are Garnham’s, Valkiūnas’s and Telford’s synthetic works (Garnham, 1966; Valkiūnas, 2005; Telford, 2009). In his discussion of Classification and Evolution, Garnham (1966) argued that not naming species, even putative ones, leads to an underestimate of diversity and could actually thwart scientific understanding. He went on to give some historical examples, and stressed that, given the paucity of characters available for description and classification, any and all characters should ‘be seized and used as liberally as possible’, particularly since it is not possible in the majority of cases to view all life-stages of a parasite (the gold standard of species description in malaria parasites). Indeed, the scarcity of characters had led most taxonomists (see below) to use characters associated with the gametocytes (although other stages are quantifiable in Plasmodium only), a point that Garnham treated with some disdain, especially when coupled with ‘host’ as a character. He went on to say that, in gametocytes and even sporozoites (from the dipteran host), morphological variation between parasite species is ‘trifling’. As pessimistic as these statements may seem, Garnham presciently stated that biochemical methods might ultimately replace morphological description (see Garnham, 1966, pp. 64–72).

Valkiūnas (2005) focused exclusively on malaria parasites in birds and took a more conservative approach to the species question than did Garnham (1966). He also argued against the one-host, one-parasite paradigm in favour of detailed descriptions of parasite morphology in the peripheral blood of the vertebrate host with a deliberate de-emphasis of the host, except in some cases in which there is experimental evidence of host preference (e.g. sub-inoculations; p. 232). However, in his taxonomic keys, host taxon is a primary ‘trait’ for species determination in both the subgenus Parahaemoproteus (p. 252) and the genus Leucocytozoon (p. 739). Valkiūnas went on to say that morphometric characters are continuous and overlapping in, for example, the ‘majority of [Leucocytozoon] species’, and indeed that despite these overlapping characters, ‘combinations of these in various species of haemoproteids are rather different and informative’ (pp. 76–78). These statements provide no explicit species criteria, but Valkiūnas also suggested that it would be straightforward to perform the necessary sub-inoculations (albeit only in Plasmodium) to characterize morphological characters of all parasite life-history stages and not to rely solely on peripheral blood stages from naturally infected birds. Indeed, he emphasized that species descriptions from wild birds should be ‘discouraged’ and their validity questioned, and that studies addressing ‘ecology and evolutionary biology based only on generic identification’ (p. 141) are simply not adequate for use in parasitological research and examination. Thus, Valkiūnas emphasizes the importance of training in classical parasitology for species descriptions, and perhaps rightly so.

Telford’s monograph (Telford, 2009) focused entirely on reptilian parasites and, even then, mostly on species descriptions. However, because of the disparity between the number of known reptile haemosporidians from morphology vs genetics, Telford emphasized that ‘until a broad spectrum of known species has been studied…phylogenetic relationships…must remain based on morphological and life history traits’ (Telford, 2009, p. 1). Telford went on to say, as did Valkiūnas (2005), that although the genomic revolution will ultimately meet, one way or another, with described morphospecies, molecular methods will not replace microscopy in the description and naming of species. Indeed, Telford took a somewhat defiant stance, understandable because traditional microscopy methods are often ignored: ‘morphology visible under light microscopy is essential in the definition and description of species, whatever relevance it may possess to genetic relationships among a fauna’ (italics added; Telford, 2009, p. 12). Telford, like Valkiūnas (2005), advocated for microscopy, and provided a detailed list of both direct (i.e. discrete measurements) and indirect (i.e. host-cell deformation) characters that should be used to ‘describe’ species.

It is pertinent here to review the advantages of traditional microscopy methods before focusing on the use of molecular methods. Mixed infections are a formidable problem for molecular methods, because these methods cannot readily distinguish two or more haplotypes of parasite; many researchers simply ignore such samples, but microscopy can be used to distinguish parasites if they have been identified previously. Molecular methods can also be limited because of preferential amplification of some parasite genetic lineages over others in the sample; some of this is due to higher levels of parasitaemia of one parasite vs another, but some is also due to the molecular methods themselves. For example, ‘primers’ that are used to target specific parasite genes may have a greater affinity for some parasite lineages, amplifying those over others, which can lead to under-representation of rare lineages (Zehindjiev et al. 2012); this problem can be addressed with ‘deep sequencing’ (Jarvi et al. 2013) which can uncover all sequence variants of a particular locus in any sample. Another problem is the mistaken assignment of genetic lineages to morphospecies on databases (MalAvi [Bensch et al. 2009] and GenBank; see below). However, one of the most challenging problems involves the biology of infection; insect vectors may

through speciation associated with host-switching (Ricklefs et al. 2004, in review).
infect into a vertebrate host sporozoites that may not be able to reproduce inside that host, while molecular methods will nevertheless amplify DNA from that parasite. Ways around this problem include identification (or not) of the parasite in a blood smear and sub-inoculations. Although the previous issues have been negative with respect to molecular methods, the majority of thorough examinations of concordance between morphology and molecules have indeed been positive: morphology matches molecules, at least at the level of subgenera (e.g. Martinsen et al. 2006; Levin et al. 2012; Zehtindjiev et al. 2012).

**Recent Discussions of the Species Question and Development of Criteria**

In 2011, two reviews addressed several major issues concerning ‘species limits’ and ‘nomenclature’ in malaria parasites: (i) the need for more well-trained parasitologists; (ii) reconciliation of morphological and molecular approaches; (iii) improving how we specify lineages in molecular databases; and (iv) developing criteria to define species (Perkins et al. 2011; Littlewood, 2012). Resolving issues concerning species delimitation is a necessary step towards the goal of determining malaria parasite biodiversity and understanding haemosporidian evolution and distribution. Past disagreement over these issues should not prevent us from coming to a consensus that is transparent and open to modification. Both reviews mentioned above concluded with a call for more cooperation, particularly concerning the species question, among research groups with different but overlapping expertise (i.e. morphometrics and genetics).

Several criteria have been suggested to delimit species in malaria parasites, including concordance among multi-locus sequence data (Falk et al. 2011), absolute values of sequence divergence (Bensch et al. 2004), and fixed nucleotide differences (explored here) which may or may not involve sequence divergence cut-off values (see below). Certainly, monophyly across multiple gene trees would be the gold standard to delimit species, but because amplifying non-mitochondrial parasite DNA has been difficult, this has not been possible in most analyses to this point. Moreover, the markers currently available from the nuclear and apicoplast genomes are not ideally suited to phylogenetic and population genetic analyses because they are either too short and lack informative, synapomorphic characters, or they produce very long terminal branches (i.e. they contain more apomorphic than synapomorphic characters; see Outlaw and Ricklefs, 2010), and in many cases, are not more informative than mitochondrial cyt b (Perkins et al. 2007). Nevertheless, nuclear genes will provide an increasing amount of data that will be useful for delimiting species.

An alternative to seeking congruence among multiple markers has been to use a cutoff level of sequence divergence to establish species boundaries. The current value of this approach rests on the availability of cyt b sequence data and the desirability of leveraging these data for more than simply cataloguing novel lineages. Based on empirical studies and comparisons of well-characterized morphospecies and their counterpart cyt b sequences, researchers have proposed and subsequently ‘tested’ the idea that 5% sequence divergence in cyt b is a reasonable expectation between morphologically differentiated species (Hellgren et al. 2007; reviewed in Valkiūnas et al. 2009a,b). The implication is that in the absence of morphological data, this cutoff could be used to delimit species and estimate the diversity of haemosporidian parasites. However, advocates of this cut-off cautioned that ‘... a molecular criterion of over 5% sequence divergence in cyt b gene for the identification of haemosporidian species should be developed and applied carefully, preferably by linking molecular and microscopical data’ (italics added; Valkiūnas et al. 2009a,b). In a recent treatment of cryptic and novel lineages found in lizard parasites, the results were consistent with the 5% ‘rule’ for cyt b divergence, in that individual cyt b lineages differed by more than 5%. The authors included morphological data (which were not informative), and two mitochondrial genes and one nuclear gene, across which there was concordance, making a strong case using reciprocal monophyly for species delimitation (Falk et al. 2011).

**The Need to Delimit Species, and Criteria for Moving Forward**

Here, we suggest a framework within which we can define species or OTUs in a standardized way so that the malaria parasite research community has a common language. The most straightforward and unambiguous requirement under a phylogenetic species concept is phylogenetic concordance across multiple loci (Perkins, 2000; Martinsen et al. 2006; Ramiro et al. 2012). Concordance across multiple loci (including linkage disequilibrium between non-interbreeding populations; Bensch et al. 2004) is the acid-test for species limits, and the multiple, independent genetic markers required for such analyses should be accessible in the near future. However, because such data are presently unavailable to most researchers, and because most infections have been characterized to date by single mitochondrial gene sequences, it is important to develop a method to circumscribe parasite lineages based on limited genetic information. We to address this issue here and offer a way to move forward.

One goal of haemosporidian research is to develop distributional and evolutionary contexts for individual lineages of parasites. Most malaria parasite lineages from wildlife hosts are given arbitrary
designations, like PTURDUS1 or *Plasmodium* sp. 1, independently by each laboratory. In fact, based on identity (100% GenBank match), 31 avian parasite lineages have at least five names, and five of these have 10 or more names (July, 2013). Without standardized names associated with taxonomic entities designated by accepted criteria, our understanding of evolution and diversity, and, perhaps even more importantly, our ability to communicate effectively about these parasites, will remain wholly inadequate. Garnham (1966) understood the need to name species, even when there were doubts, because ‘apart from the inconvenience [naming species in the absence of total evidence] causes… it tends to inhibit acquisition of knowledge by producing the impression that the life cycle of a parasite bearing the name *Plasmodium* – such as the so-called *Plasmodium* of lizards – must be the same as the cycle of well-known

---

**Fig. 1.** Bayesian maximum clade credibility tree of the avian malaria parasite cytochrome b gene tree based on 479 bp (BEAST 1.7.5; [Drummond and Rambaut, 2007]); posterior probabilities are highest at the terminals and all >0.80 are noted with circles of increasing size at the nodes). All genus-level clades are collapsed except for the focal *Parahaemoproteus* clade. All *Parahaemoproteus* clades that do not contain a sequence from a named lineage are also collapsed. The smallest monophyletic group to include all sequences (purple names) attributed to a named species is highlighted in blue and the specific epithet is adjacent to each clade. Embedded sequences attributed to other species are named in dark green. Two species whose sequences do not form clades are named in light green and in orange. Unnamed lineages are named in black. The full figure is available on request.
members of the genus [i.e. human Plasmodium] (Garnham, 1966, pp. 64). Thus, sequence data for avian malaria parasites are abundant, and increasing rapidly, and now researchers face the additional problem of being provided with ambiguous criteria to determine novel lineages. In many cases, genetic divergence between malaria parasite lineages cannot be linked to morphological variation (because blood smears are lacking; Bensch et al. 2004; Marzal et al. 2011; see also Pérez-Tris et al. 2005), or genetic divergence does not match morphological variation, at least as indicated by species names (Beadell et al. 2006 [P. relictum]). Hence, the utility of genetic lineages, but also of formally named species, is difficult to evaluate in terms of representing independently evolving lineages in the absence of a phylogenetic species concept (Bensch et al. 2004).

One might ask whether it is reasonable to use criteria for species delineation that have been developed for other problematic organisms, such as prokaryotes. Bacteriologists have developed a system, albeit not without controversy and uneven application, that uses 16S rRNA sequence data to define lineages, with 3% divergence often being used to distinguish species (Weisburg et al. 1991; reviewed in Forney et al. 2004). Of course, malaria parasites do not have a complete 16S rRNA sequence in their genome, but researchers have argued for using cyt b sequence divergence to delineate species (Hellgren et al. 2007, but see Levin et al. 2012). Is it biologically meaningful to apply a %-divergence rule to cyt b to delimit avian haemosporidian species? To address the generality of a %-divergence rule, at least up to a point, we reconstructed a Bayesian phylogeny of all available avian malaria parasite lineages (n = 1300 as of July 2013) using the 479 base pairs of cyt b designated by Bensch et al. (2009) to define lineages.

For the genus Parahaemoproteus (chosen because it is the best sampled genus), we identified monophyletic clades of all ‘named’ species for which there were at least two representative sequences in MalAvi (purple taxon names, clades highlighted in blue, in Fig. 1). Overall, there is a great degree of concordance between named species and their constituent clades, which is not surprising given that we used named sequences to define clades and the fact that most named sequences were compared with those for which detailed morphological descriptions are available (e.g. Martinsen et al. 2006; Hellgren et al. 2007). However, it is clear that issues remain with some named sequences. Named species that are embedded (i.e. paraphyletic sequences) in one of these clades are highlighted further in Fig. 1. Constituent sequences are not monophyletic for two named species: two nucleofascialis in light green and six tartakovskyi in orange. In four cases, named species represented by a single sequence are embedded in named species clades: belopolksyi, balmorali, minutus and majoris. Because these are single representative sequences, we cannot assess whether their placement is due to misidentification or whether they represent ‘real’ species that have recently diverged. We calculated the range of pairwise distances in each of the named species (Table 1). Variation in sequence divergence within named species increases with sample size and exhibits a maximum value of 0.041, or about 4%. The average intraspecific pairwise divergence between cyt b sequences is 0.02 (range = 0.001–0.041; Table 1) whereas the average interspecific divergence value for all pairwise comparisons is 0.07 (range = 0.013–0.092), which still includes some comparisons that are clearly due to misnamed lineages (Table S1).

The mean maximum values between the interspecific and intraspecific pairwise divergences differ significantly (P < 0.001). However, it is clear from this analysis that misnamed or misclassified lineages hinder the broad applicability of a %-divergence criterion; any conclusions drawn from this analysis presume that each name represents a biologically meaningful entity, which may not be the case.

In another case, a 3% or 4% rule would classify three genetically divergent, partially allopatric lineages as a single species (Table 2, Fig. 2). Two of the lineages, DR03 and JA02, are found primarily in one host species, the bananaquit (Coereba flaveola), but on different islands in the West Indies (Hispaniola and St. Vincent [DR03, having a disjunct distribution] and Jamaica and the nearby Cayman Islands [JA02]); LA07, although mostly associated with the bananaquit, is widespread geographically, including Mexico and Venezuela where it has been recovered from Icterus spp., but not including Jamaica, Hispaniola, or St. Vincent; it is particularly common on Puerto Rico and Grenada. Fixed differences

### Table 1. Named Parahaemoproteus species in MalAvi, with the number of constituent lineages based on Bayesian phylogenetic analysis of existing cyt b data (Fig. 1) and the range of pairwise sequence divergence values

<table>
<thead>
<tr>
<th>Parahaemoproteus species</th>
<th>Number of host species</th>
<th>Number of constituent lineages (including unnamed lineages)</th>
<th>Sequence divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>balmorali</td>
<td>3</td>
<td>12</td>
<td>0.001–0.039</td>
</tr>
<tr>
<td>belopolksyi</td>
<td>5</td>
<td>30</td>
<td>0.002–0.039</td>
</tr>
<tr>
<td>cyanoitrite</td>
<td>1</td>
<td>9</td>
<td>0.002–0.035</td>
</tr>
<tr>
<td>haemobelo belopolksyi</td>
<td>1</td>
<td>4</td>
<td>0.002–0.006</td>
</tr>
<tr>
<td>lanii</td>
<td>1</td>
<td>8</td>
<td>0.002–0.020</td>
</tr>
<tr>
<td>magnus</td>
<td>2</td>
<td>4</td>
<td>0.002–0.010</td>
</tr>
<tr>
<td>majoris</td>
<td>5</td>
<td>16</td>
<td>0.002–0.010</td>
</tr>
<tr>
<td>micrometeris</td>
<td>3</td>
<td>10</td>
<td>0.002–0.024</td>
</tr>
<tr>
<td>minutus</td>
<td>2</td>
<td>9</td>
<td>0.000–0.010</td>
</tr>
<tr>
<td>motacillae</td>
<td>1</td>
<td>6</td>
<td>0.002–0.010</td>
</tr>
<tr>
<td>parabelopolksyi</td>
<td>2</td>
<td>28</td>
<td>0.002–0.041</td>
</tr>
<tr>
<td>Sanguinis</td>
<td>1</td>
<td>4</td>
<td>0.003–0.010</td>
</tr>
</tbody>
</table>
between these parasite populations in cyt b suggest that they are isolated from one another despite their recent origin (Table 3). Blood smears from these lineages, although available, have not yet been evaluated, but the important point here is that even a few nucleotide differences potentially could be associated with species differences, and this is borne out by our analysis of these same populations using a multi-locus approach (Fig. 2; see also Braga et al. 2011).

Our results suggest that even these very young populations of parasites can be designated as good phylogenetic species, provided we understand the geographic and host-taxonomic context of the constituent populations. Even if one argues that allopatry, as in the case of DR03 and JA02 in Fig. 2, precludes testing reproductive compatibility to determine biological species, the populations are genetically distinct in cyt b (Tables 2 and 3), and the species tree generated from gene trees from markers (four) from all three genomes is highly supported (Fig. 2). None of these lineages is associated with a named species, and each is different from the closest (BLAST) ‘named species’ by at least 5%.

Prospects

Characters used to describe species of malaria parasites, which are relatively few in number, are likely to be conservative because these parasites are highly specialized. This lack of morphological differentiation apparently has led to attaching the same species name to more than one reproductively isolated entity (per Garnham, 1966). This, in turn, has created cases of paraphyly and polyphyly of named species, which we have identified in the avian *Parahaemoproteus* phylogeny (Fig. 1). Conversely, named ‘species’ may harbour considerable species diversity that cannot be identified because the morphological characters are too few. Given the genetic diversity that we see among named haemosporidian lineages, many named species of malaria parasite are like genera in birds, which are typically
defined by morphological characters. Most avian genera contain many species, but many of these were misclassified to genus (i.e. misplaced in the taxonomy) until molecular data began to sort this out, beginning primarily with Sibley and Alquist (1990). Just as ornithologists have done over the past 20 years, malaria parasitologists need to recognize that morphological descriptions are limited for haemosporidians, at least at the level of biological species, and that other characters are required for species delimitation and identification. At the same time, however, few young malaria parasitologists are being trained in classical techniques including simple blood smear preparation. Even with the abundance of genomic data on the horizon (S. Bensch, personal communication), if the field lacks the expertise to tie genetic variation to life-history variation in the parasites, opportunities to understand the broader contexts of parasite evolution and diversification will be missed.

**CONCLUDING REMARKS**

Our analyses show that the genetic variation within named haemosporidian species can be quite variable. This, of course, has been recognized by most malaria taxonomists in recent years and it underlines the case for avoiding the naming of lineages independently of morphological characters. However, in many cases, molecular characters may be required to confirm the monophyly of named species and, in the absence of suitable morphological variation, sequence variation might be the only indication of species distinction. Of course, naming species is not the goal of most researchers, and we outline an approach below that can help to sort out the bewildering sequence diversity of avian malaria parasites.

Developing a uniform standard for naming DNA sequences from haemosporidian parasites is an important goal. The most general and useful scheme at present is the MalAvi database (Bensch et al. 2009). Although MalAvi provides a system to name novel lineages, based on as little as a single nucleotide difference, little headway has been made since the creation of MalAvi in terms of applying its information broadly to the species question. Most malaria parasite researchers are less interested in naming new species than they are in having a protocol that allows them (1) to link their data to a broader database to determine the geographic distribution and host breadth of their lineages (but see Valkiūnas et al. 2009a, b); and (2) to name an existing or novel lineage in such a way that all researchers can easily identify that lineage. The first step would be to align

---

**Table 4. Suggested steps to develop diagnostic standards for species delimitation**

- Adopt a ‘tagging’ criterion for newly generated sequence data (e.g. 99% similarity or 1 fixed difference).
- Develop tagging system based on cyt b.
- Parse existing data into ‘tagged’ blocks of identical/similar sequences.
- Develop input/output pipeline in MalAvi.
- Develop standard procedure to send blood smear slides to experts.
- Facilitate consensus to identify experts who can describe new species and diagnose existing smears for species identification.
- Develop pipeline for processing blood smears.

**Fig. 3. Schematic of a proposed standardized procedure for identifying parasite lineages.**

---
all cyt b sequence data currently on hand, a feature of the MalAvi database (although data that do not overlap MalAvi’s sequence ‘standard’ cannot be incorporated in this scheme), and organize MalAvi lineages into tagged blocks. Beyond that, one may wish to organize sequences into clades, potentially associated with named species, based on an arbitrary level of sequence similarity (e.g. 99%). When new data are generated, a protocol could then include, for example: (1) BLAST to MalAvi and align to all tagged sequences; (2) if the sequence(s) match at a pre-determined level (e.g. 99% sequence similarity, or no fixed differences) with a previously tagged lineage, clade of sequences, or named species, then that tag is applied to the new data and the researcher can easily determine the geographic and host range of the lineage(s); (3) if the lineage is truly novel at some arbitrary level of sequence divergence (e.g. >1%), then it would be given a new tag. Figure 3 shows how the process could work. However, recognizing that malaria ‘species’ undoubtedly contain genetic variation at most diagnostic loci, the application of any such criteria for distinguishing reproductively isolated entities should take into account the geographic and host distribution of the genetic lineages as well (Ricklefs et al. 2005), and such issues could be worked out (see below). Ultimately, the validity of the approach outlined in Fig. 3 would provide a repeatable and useful metric at this stage. Certainly, a useful extension of this thinking would be for the research community to agree to allow a committee of experts (e.g. based on the Malaria Research Coordination Network, www.malar iarcn.org) to extend the applicability of the MalAvi (Bensch et al. 2009) model and further standardize data generation and naming of lineages (Table 4).

Ideally, phylogenetic species should be defined based on multiple loci, i.e. relating gene trees to species trees. This can be done with most mammalian Plasmodium parasites owing to the availability of multiple, independent genetic markers (Martinsen et al. 2008), but it is more difficult for haemosporidian parasites of wild birds and reptiles (Bensch et al. 2004). In these, parasite DNA is mixed with the host DNA, and efforts to isolate parasite DNA are only just beginning to yield additional independent genetic markers (Palinauskas et al. 2013). Certainly, genomic sequencing would solve this issue in terms of identifying markers, but genome sequencing, especially when a minute proportion of all ‘reads’ will be parasite rather than host DNA (Wang et al. 2012), is not cost-effective for most researchers. Currently, major efforts are underway to generate genomic data from malaria parasites of wildlife, and the results are bound to be exciting: whether other loci do, or do not, corroborate current cyt b data are both interesting prospects (see also Palinauskas et al. 2013). We predict that within the next 5 years, sufficient data should be available to define species in a very robust way. However, in the meantime, given the rate at which genetic lineages of malaria parasites are being added to databases, we can usefully adopt a less formal species definition. Moreover, because cyt b gene trees and parasite species trees bear a reasonable correspondence, cyt b is at present the most useful tool that the field has to define species and/or OTUs.

**SUPPLEMENTARY MATERIAL**

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0031182014000560.

**ACKNOWLEDGEMENTS**

We wish to thank Dr Stephen Phillips for inviting us to write this review, and to acknowledge and thank members of our respective lab groups for critiques of previous drafts. Ravinder Sehgal and an anonymous reviewer provided excellent criticism and comments that improved the manuscript.

**REFERENCES**


