Infection, Genetics and Evolution 75 (2019) 103994

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research Paper

Indel-informed Bayesian analysis suggests cryptic population structure between *Plasmodium knowlesi* of humans and long-tailed macaques (*Macaca fascicularis*) in Malaysian Borneo

JustinJ.S. Wilcox*[,1], Abigail Kerschner, Hope Hollocher

Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556-5688, USA

ARTICLE INFO

Keywords:
Malaria
Plasmodium knowlesi
Zoonosis
Macaque
Host specificity
Phylogeny

ABSTRACT

Plasmodium knowlesi is an important causative agent of malaria in humans of Southeast Asia. Macaques are natural hosts for this parasite, but little is conclusively known about its patterns of transmission within and between these hosts. Here, we apply a comprehensive phylogenetic approach to test for patterns of cryptic population genetic structure between *P. knowlesi* isolated from humans and long-tailed macaques from the state of Sarawak in Malaysian Borneo. Our approach differs from previous investigations through our exhaustive use of archival *18S Small Subunit rRNA* (*18S*) gene sequences from *Plasmodium* and *Hepatocystis* species, our inclusion of insertion and deletion information during phylogenetic inference, and our application of Bayesian phylogenetic inference to this problem. We report distinct clades of *P. knowlesi* that predominantly contained sequences from either human or macaque hosts for paralogous A-type and S-type *18S* gene loci. We report significant partitioning of sequence distances between host species across both types of loci, and confirmed that sequences of the same locus type showed significantly biased assortment into different clades depending on their host species. Our results support the zoonotic potential of *Plasmodium knowlesi,* but also suggest that humans may be preferentially infected with certain strains of this parasite. Broadly, such patterns could arise through preferential zoonotic transmission of some parasite lineages or a disposition of parasites to transmit within, rather than between, human and macaque hosts. Available data are insufficient to address these hypotheses. Our results suggest that the epidemiology of *P. knowlesi* may be more complicated than previously assumed, and highlight the need for renewed and more vigorous explorations of transmission patterns in the fifth human malarial parasite.

1. Introduction

Plasmodium knowlesi has recently been acknowledged as an important causative agent of human malaria in Southeast Asia (Singh et al., 2004; Cox-Singh et al., 2008; White, 2008). Although it is generally regarded as a zoonotic agent that humans incidentally acquire through sympatry with macaques, it can persist in human populations and accounts for a substantial portion of human malaria burden in some regions (Cox-Singh and Singh, 2008; Cox-Singh et al., 2008; Barber et al., 2012; William et al., 2013). The ubiquity of *P. knowlesi* in humans of endemic regions provides the potential for transmission between humans and obfuscates the relative roles of zoonotic, enzootic, and (potentially) human-to-human transmission in driving human infections (Singh and Daneshvar, 2013; William et al., 2013). Resolving these relative roles is of paramount concern to the control of *knowlesi*-malaria and the management of human-macaque interactions.

The evidence to date unequivocally suggests that *P. knowlesi* has zoonotic origins in humans and is regularly transmitted between long-tailed macaques (*Macaca fascicularis*) and humans. Several phylogenies have shown co-clustering of human and macaque *P. knowlesi* strains across multiple loci (Singh et al., 2004; Jeslyn et al., 2011; Lee et al., 2011; Jongwutiwes et al., 2011; Putaporntip et al., 2013), and experimental studies provide additional evidence for cross-transmission by showing that *P. knowlesi* from macaques will easily infect humans through direct inoculation (White et al., 2008; Singh and Daneshvar, 2013). *Plasmodium knowlesi* is also transmitted through many of the same vectors (most notably *Anopheles* species of the *Leucosphyrus* subgroup) as dedicated human malarias (Vythilingam et al., 2016) and has been reported to form gametocytes in humans (Singh et al., 2004; Maeno et al., 2017; Davidson et al., 2019), suggesting that human-to-

* Corresponding author at: New York University Abu Dhabi, Abu Dhabi, United Arab Emirates.
E-mail address: jw5478@nyu.edu (J.S. Wilcox).
[1] Present Address: P.O. NYUAD 129188 New York University Abu Dhabi, Abu Dhabi, United Arab Emirates.

https://doi.org/10.1016/j.meegid.2019.103994
Received 10 March 2019; Received in revised form 1 August 2019; Accepted 3 August 2019
Available online 14 August 2019
1567-1348/

human transmission of *P. knowlesi* could be possible as well (Marchand et al., 2011), particularly as mosquito-vectored serial passage of *Plasmodium knowlesi* between humans has been achieved experimentally (Chin et al., 1968).

Natural human-to-human transmission of *P. knowlesi* has not been shown and is difficult to directly assess. Population substructuring between human and macaque hosts may, however, provide indirect evidence for human-to-human transmission. A study in Thailand has, for example, reported clades of *P. knowlesi* that appear to occur more typically in either humans or macaques, consistent with expectations human-to-human transmission (Putaporntip et al., 2013), despite the low incidence of human *knowlesi*-malaria in this country (Jongwutiwes et al., 2011). While studies from Malaysia have, overwhelmingly shown phylogenetic cross-clustering of sequences from parasites of macaques and humans (Singh et al., 2004; Jeslyn et al., 2011; Lee et al., 2011), they may provide some indirect evidence for cryptic partitioning between hosts, by revealing: low haplotype overlap between parasites of human and macaque hosts (Lee et al., 2011); tendencies for overlapping haplotypes to occur at internal nodes of haplotype networks (Lee et al., 2011; Yusof et al., 2016), which are reflective of more ancestral states (Castelloe and Templeton, 1994; Templeton, 1998); and high levels of intraspecific genomic diversity of *P. knowlesi* relative to other human malarial agents with cryptic structuring of this diversity within humans (Assefa et al., 2015; Divis et al., 2015; 2017). While these studies provide a clear role for zoonosis, human-to-human transmission may occur and the primacy of zoonotic transmission as the only source of human *knowlesi*-malaria infections remains an assumption (Assefa et al., 2015). Under these circumstances of potentially incomplete assortment of *P. knowlesi* lineages infecting humans and macaques, more sensitive phylogenetic approaches may detect cryptic divisions between *P. knowlesi* in these hosts.

Here, we utilize a customized, high-throughput pipeline to build a high-resolution indel-informed Bayesian tree of archived *18S* gene sequences from *Plasmodium* spp. and *Hepatocystis* spp., which are reported to be contained within the genus *Plasmodium* (Galen et al., 2018). We seek to improve sensitivity over previous analyses through our application of Bayesian phylogenetic inference (Holder and Lewis, 2003; Yang and Rannala, 2012), inclusive taxon sampling (Pollock et al., 2002; Zwickl and Hillis, 2002; Heath et al., 2008), and inclusion of indels (Belinky et al., 2009; Nagy et al., 2012; Luan et al., 2013). We chose the use of the *18S rRNA* gene given: 1) it is a commonly sequenced gene with a large amount of available sequence data (Quast et al., 2012; Yilmaz et al., 2013); 2) it has been a predominant focus of previous phylogenetic studies on *P. knowlesi* (Singh et al., 2004; Jongwutiwes et al., 2004; Bronner et al., 2009; van Hellemond et al., 2009; Jiang et al., 2010; Yusof et al., 2016); 3) the inclusion of indels may be particularly important to resolving its phylogenetic history (Mu et al., 2011; Nagy et al., 2012). We utilized nearly full-length *18S* gene data, and include SNPs and binary indel data as separate blocks within the same analyses. We analyze our tree using distance-based statistics to directly test for population structure between *P. knowlesi* strains infecting human and long-tailed macaque hosts. We included only sequences from the state of Sarawak in Malaysian Borneo, a hot spot of *P. knowlesi* emergence (William et al., 2013), in this analysis to avoid geographic sampling bias. We evaluate our results in the context of three hypotheses (Fig. 1): 1) that *P. knowlesi* is a strict zoonosis; 2) that *P. knowlesi* displays strict host specificity; 3) that *P. knowlesi* displays patterns of specialization within hosts, but with limited zoonotic transmission.

2. Methods

2.1. Sequence processing

We utilized all *Plasmodium* and *Hepatocystis* spp. sequences in the Silva 128 *18S* gene database (Quast et al., 2012; Yilmaz et al., 2013); this consists of 413 sequences in total. Of these, we could attribute 47 sequences to *P. knowlesi* (see Supplementary Table 1). Sequences were compiled from the Silva 128 database using an automated tcsh-shell script; the *Babesia divergens* sequence AJ439713 was also included with these as an outgroup. This outgroup sequence was chosen as *Babesia* spp. are among the closest relatives of *Plasmodium* spp. with available *18S* sequence information. Sequences were aligned using the Mafft Q-ins-i iterative alignment for structural RNAs (Katoh and Standley, 2013). Redundant sequences were removed using the CD-hit program (Li and Godzik, 2006; Fu et al., 2012), and remaining sequences (N = 320, including the outgroup, of which 28 were attributable to *P. knowlesi*) were realigned using the Q-INS-i iterative alignment. Processed sequences were compiled into Nexus format in two partitions to produce an all-sequence indel-informed tree: one partition contained full sequence data; the other contained only presence-absence data for each base in the alignment. Host and geographic information were recorded for each remaining unique sequence using National Center for Bioinformation Technology (NCBI) Genbank; relevant publications were consulted when necessary. *Plasmodium* spp. possess multiple copies of the *18S* gene that are traditionally classified as A-type that are primarily transcribed during the asexual life stage of the parasite in its vertebrate host, and S-type that are primarily transcribed during the sexual life stage of the parasite during its mosquito host (Waters, 1994; Gardner et al., 2002; Leclerc et al., 2004). As these two types of loci are the result of multiple duplication events, the true homologs of these loci in other *Plasmodium* spp. are unknown, and as A-type and S-type loci are not definitively known to be homologs of one another across *Plasmodium* spp., all *18S* paralog sequences were included in our analysis (Nishimoto et al., 2008; Jiang et al., 2010); ribosome type (i.e. A-type or S-type) was controlled for during down-stream data analysis (see "*2.3 Statistics*"). In order to establish the role of indels in driving our results, we also produced a tree using the full *18S* dataset but excluding the indel partition from the final Nexus file: an all-sequence no-indel tree.

In addition to producing phylogenetic trees based on all *Plasmodium* and *Hepatocystis* spp. sequences available, we also produced trees that contained only *P. knowlesi* sequences to establish the impact of our taxon sampling in driving our conclusions. Two of these trees were type specific and contained only either the (self-reported) A-type (N = 14) or S-type (N = 12) sequences from Malaysia (i.e. A-type-only and S-type-only *P. knowlesi*-specific trees). A third tree contained all *P. knowlesi* sequences remaining after dereplication in the all-sequence analysis, regardless of reported type or host location (N = 28): a general *P. knowlesi*-specific tree. A sequence from *Plasmodium falciparum* (ACBR01000732) was used as an outgroup in all *P. knowlesi*-specific analyses due to its distant position from simian malarial agents (Loy et al., 2017).

2.2. Phylogenetic analysis

Bayesian trees were constructed for all sequence sets (i.e. the all-sequence, and *P. knowlesi*-specific analyses) using MrBayes (version 3.2.6) on Xsede (Ronquist et al., 2012), accessed through the CIPRES Portal. Indel and sequence data were included in analyses for all of these data sets as separate but equally weighted partitions. All terminal positions that contained a gap in any sequence were coded as missing data to ensure that poor coverage would not bias our indel analysis. In addition, an all-sequence tree was constructed without the use of indels to assess the role of indels in driving our results. This tree treated all gaps as missing data. During the construction of all trees, sequence data were analyzed using a Generalized Time Reversible (GTR) model with gamma-distributed rate variation and a proportion of invariable sites; this model was chosen to minimize the assumptions and constraints imposed upon our analysis as this has been shown to be a more conservative approach (Huelsenbeck and Rannala, 2004). The indel partitions, when included, were analyzed using a binary model with rate

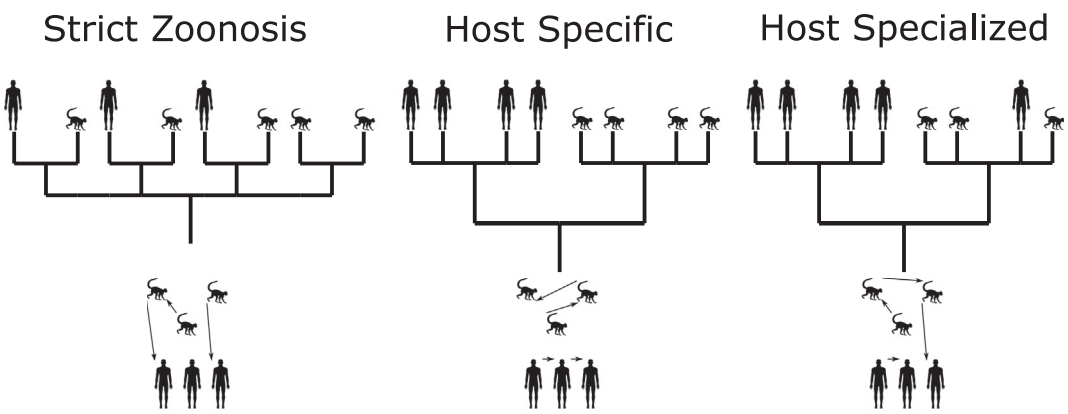

Fig. 1. Hypothesized Modes of Transmission for *Plasmodium knowlesi*. Under strict zoonosis human infections are acquired from macaques and this is reflected by the clustering of human-isolated strains within predominantly macaque-hosted clades. Under host-specific transmission regimes, infections transmit only within hosts, and strains isolated from humans and macaques cluster separately. Under host specialized transmission human and macaque isolated strains tend to cluster apart from one another reflecting transmission within hosts, but some human isolated strains cluster with macaque strains reflecting limited zoonotic transmission.

variation between sites and a proportion of invariable sites. All analyses contained two runs, with four chains per run. The all-sequence analyses had 10,000,000 iterations per run; the *P. knowlesi*-specific analyses utilized 100,000,000 iterations per run. Trees were sampled from each run every 1000 generations. In all cases, consensus trees were built based on majority rule after discarding the first 25% of samples from each run as burn-in.

2.3. Statistics

Statistical tests were performed in R Studio version 1.1.456 (RStudio, 2016), using R version 3.5.3 (R Core Team, 2018). Tests for genetic structuring between sub-populations are often conducted using F_{st}-based tests of allele frequency distributions (Goudet et al., 1996). However, this approach is inappropriate for our dataset because of an almost complete lack of allele frequency reports for *P. knowlesi*. Instead, we analyzed the partitioning of genetic variation in our consensus trees using PerMANOVAs accessed through the "vegan" package (Oksanen et al., 2018). This approach is highly similar to an Analysis of Molecular Variance (AMOVA; see Excoffier et al., 1992), except that it allows for inference of distances between sequences based on a full Bayesian character-based evolutionary model, rather than simple pairwise genetic distances. This use of Bayesian inference has the advantage of allowing us to account for several factors—such as insertion-deletion events, mutation rate variation between sites, and differential mutation rates across lineages (Holder and Lewis, 2003)— that cannot be incorporated into a traditional AMOVA (Excoffier et al., 1992). Branch length information was extracted from our trees using the "APE" package (Paradis et al., 2004). Overall, the structure of the all-sequence tree was assessed in the context of parasite identifications (the species, or genus in the case of *Hepatocystis*, to which a sequence had been ascribed in the database), the broader taxonomic grouping of the host from which a sequence was isolated, and the geographic location from which a sequence was collected, using a multi-way PerMANOVA design with marginal ("type III") sums of squares to control for confounding effects among these multiple variables. PerMANOVAs were also used to assess structuring of *P. knowlesi* sequences between hosts on all trees. Two *18S* rRNA types are reported from *P. knowlesi*: the A-type expressed in asexual life-stages and the S-type expressed in sexual life-stages (Leclerc et al., 2004). To account for this in the all-sequence analysis and the *P. knowlesi*-only sequence analysis, we assessed the genetic structuring of *P. knowlesi* between human and macaque hosts using a PerMANOVA that included rRNA-gene type as a blocking effect within the model and restricted permutations within these types when assessing *p*-value. Sequences were excluded from these analyses when A-type or S-type had not been identified on NCBI Genbank or in the published literature. Several unique sequences were from the same hosts. To control for this in our analysis, isolate identifications were included as a nested effect within host type within all PerMANOVAs. We also assessed differences in the assortment of isolates from human and macaque hosts into different clades using a multinomial z-pooled unconditional exact test (Lydersen et al., 2012), which is similar to the Barnard's Exact Test (Barnard, 1945). We chose this test as it is more powerful than Fisher's Exact test at low sample sizes (Mehta and Senchaudhuri, 2003), while maintaining appropriately low Type I error (Mehrotra et al., 2003). Exact tests were performed with the "Exact" package in R (Calhoun, 2016). These tests were performed separately for A-type and S-type loci, and all isolates were counted only once for each locus type regardless of the number of sequences derived from it. This analysis was only performed for A-type clades on the all-sequence no-indel tree, due to lack of separate S-type clades on this tree. Lower resolution and lack of clearly defined clades precluded exact test analyses of host-type distributions on any of the *P. knowlesi*-only trees. To eliminate potential sampling biases based on geography, we restrict all of our statistical tests on structuring and partitioning of *P. knowlesi* between hosts to sequences and isolates from the state of Sarawak in Malaysian Borneo. All reported *p*-values are two-tailed.

3. Results

3.1. Tree topology

The all-sequence indel-informed dataset produced a highly-structured tree. All branches exhibited posterior probabilities of at least 50%, with a tendency for substantially higher support on more terminal nodes relative to mid-tier divisions. The overall tree was significantly structured by the taxonomic identification of parasites (Psuedo-$F = 3.13$; $p < .0001$; Supplementary Fig. 1), the broader taxonomic classification of their hosts (Psuedo-$F = 3.06$; $p < .0034$; Supplementary Fig. 2), and the geography of collection sites (Psuedo-$F = 4.79$; $p < .0001$; Supplementary Fig. 3). Although sequences were heavily grouped with parasites of the same species (or genus in the case of *Hepatocystis* spp.), polyphyly was observed based on most self-reported *Plasmodium* and *Hepatocystis* species identifications. *Plasmodium gonderi* and *P. juxtanucleare* were the only monophyletic exceptions, although each species only had two sequences represented on the tree. *Plasmodium brasilianum*, *P. cathemerium*, *P. gallinaceum*, *P. lophurae*, and *P. simium* were all only represented by a single sequence and could not be evaluated for monophyly.

Plasmodium knowlesi identification was polyphyletic for the A-type, and these divergent clades were divided into a strictly human-hosted clade (Posterior Probability = 1) and a macaque-host dominated clade

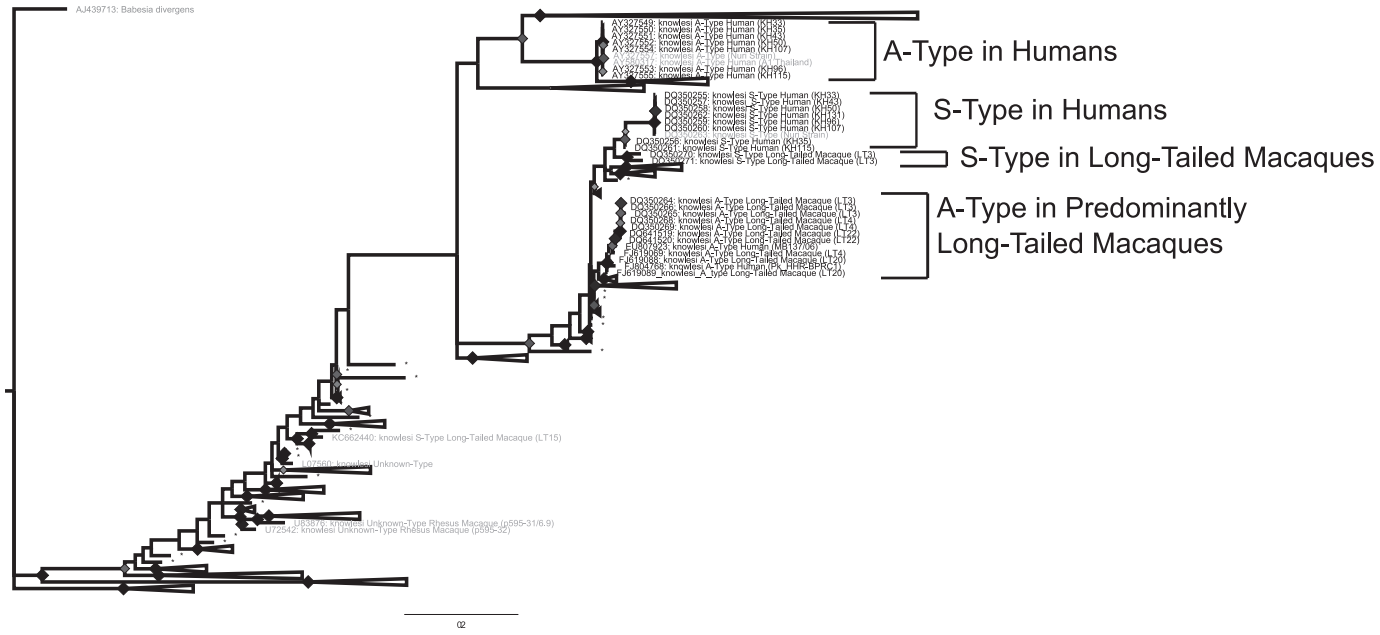

Fig. 2. *Plasmodium knowlesi* Clades on All-Sequence Indel-Informed Bayesian Tree of Archival *18S* Gene Sequences. Only self-designated *P. knowlesi* sequences are shown. Clades that do not contain *P. knowlesi* are collapsed, and lone sequences of non-*knowlesi* species are designated with an asterisk. Individual sequences are labeled with the *P. knowlesi 18S* gene type (A-type or S-type). Sequences without specified *18S* gene types, without specified hosts, that were from outside of Sarawak, lab strains, and outliers were excluded from the analysis and are designated in gray. Remaining *Plasmodium knowlesi* clades are designated and labeled by host types and *18S* gene type. Isolate identifications for each *P. knowlesi* sequence are shown in parenthesis at the end of each sequence label. Node shapes (diamonds) represent posterior probabilities: larger and darker diamonds represent higher posterior probabilities and range in color from lightest gray (posterior probability = .5) to black (posterior probability = 1).

(PP = 1) with unequivocal support. In contrast to the A-type sequences, all S-type sequences were distributed into a single large but weakly-supported clade (PP = 0.5268) with a clear and well-supported division between isolates taken from macaque (PP = 1) and human hosts (PP = 0.91).

3.2. Structuring of genetic variation between humans and macaques in P. knowlesi

We tested for the partitioning of genetic variation between human and macaque hosts in *P. knowlesi* using a PerMANOVA while controlling for isolates by nesting them within hosts and *18S* type (A-type and S-type) included as a blocking effect in the model. We found significant clustering of sequences by human and macaque hosts (pseudo-F = 29.81; p-value = .00025) on the all-sequence indel-informed tree (Fig. 2), and the all-sequence no-indel tree (pseudo-F = 8.70; p = .00542; Supplementary Fig. 4). However, we did not find significant clustering of sequences between human and macaque hosts on the general *P. knowlesi*-specific tree (pseudo-F = 1.37; p = .274) (Supplementary Fig. 5), the A-type-only *P. knowlesi*-specific tree (pseudo-F = 0.68; p = .4625; Supplementary Fig. 6) or the S-type-only (pseudo-F = 1.08; p = .5694; Supplementary Fig. 7) *P. knowlesi*-specific tree. We found significant, non-random (i.e. biased) distributions of isolates from different host types (human or long-tailed macaque) across the two A-type (Barnard's Exact Test: p = .01313) and two S-type (Barnard's Exact Test: p = .01806) clades of the all-sequence indel-informed tree. These results were identical for the A-type sequences in the all-sequence no-indel tree.

4. Discussion

We report phylogenetic patterns of cryptic genetic partitioning between strains of *P. knowlesi* infecting humans and long-tailed macaques across both paralogous A-type and S-type loci of the *18S rRNA* gene. The reasons for this partitioning remain unresolved, but our results provide statistical support for a predisposition of certain *P. knowlesi* strains to occur differentially in human and long-tailed macaque hosts. This could occur as a result of variability in the capacity of different *P. knowlesi* lineages to zoonotically transmit from macaques to humans (i.e. a bias toward zoonotic transmission of particular lineages), or as a result of historic host switching events with independent lineage-specific macaque-to-macaque and human-to-human transmission of *P. knowlesi* in Sarawak (i.e. a bias toward within-host transmission of separate lineages in macaques and humans). These two hypothesized explanations should result in different phylogenetic patterns. We would expect biased-zoonotic transmission of certain strains to result in the nesting of sequences derived from human hosts with macaques within a broader clade of sequences from macaques, reflecting the zoonotic transmission of a subset of related macaque-hosted lineages into humans and the ongoing transmission of these lineages within macaque hosts. In contrast, we would expect historic host-switching events and a bias for within-host transmission of lineages in humans and macaques to result in a deeper division between parasite lineages isolated from different hosts. The very limited sequence data currently available on sympatric *P. knowlesi* infecting macaques and humans in Sarawak prevents us from assessing either of these hypotheses. We do, however, observe A-type and S-type clades composed entirely of sequences isolated from human hosts, and the deep divergence of these from clades that contain macaque sequences. We also note the lack of an A-type clade containing only macaque sequences. Anecdotally, these results are consistent with two alternative transmission pathways leading to *P. knowlesi* infections in humans: a human-to-human transmission pathway and an ongoing zoonotic transmission pathway from long-tailed macaque hosts. This transmission pattern does seem plausible, as the inhabitants of Malaysian Borneo have been reported generally to avoid regions lacking both large areas of jungle and macaques (Md-Zain et al., 2011; Karuppannan et al., 2014; Loft et al., 2016). In addition, the two A-type sequences that were isolated from human hosts and that grouped with macaques were specifically reported from individuals that had spent time in wilderness areas, and in the case of FJ804768, time

with macaques (Bronner et al., 2009; van Hellemond et al., 2009). At present, however, we can conclude only that existing *18S* sequences of *P. knowlesi* infecting humans and long-tailed macaques in Sarawak, Malaysia, are drawn non-randomly from phylogenetically distinct pools.

Our overall tree structure was heavily influenced by taxonomic identification of sequences, but did display polyphyly for some taxa. This result is not unexpected given our use of transcribed paralogous gene loci (Funk and Omland, 2003). It may, however, also be indicative of incorrect taxonomic identifications in the SILVA database (Breitwieser et al., 2017). An examination of the literature from which sequences were derived revealed identifications using: phylogenetic approaches, particularly distance-based approaches, such as neighbor-joining trees (Seethamchai et al., 2008; Li et al., 2013); morphological identifications (Contacos et al., 1972; Schumacher, 1973); hybridizations (Li et al., 1997; Kawamoto et al., 2002); species-specific primers (Win et al., 2004); and archival data on long-term laboratory strains (Kissinger et al., 2002; Wiersch et al., 2005; Carlton et al., 2008). These diverse methods may have given rise to ambiguities in the taxonomic assignment of these sequences. Our results support calls for standardization of sequence-based techniques for *Plasmodium* identification (Kobayashi et al., 2015) and further exploration of *Plasmodium* diversity (Perkins and Schaer, 2016). Standardized methods and broader information on *Plasmodium* diversity may also be important to understanding the relative zoonotic potential of *P. knowlesi* in particular given the high levels of divergence within *P. knowlesi* (Assefa et al., 2015), recent reports of cryptic *Plasmodium* species within macaques (Raja et al., 2018), and the reported problems that this has caused for other phylogenetic studies on *P. knowlesi* (Leclerc et al., 2004).

Our results reveal phylogenetic patterns of host usage that are inconsistent with other studies on *P. knowlesi* (Lee et al., 2011; Jeslyn et al., 2011), including other studies that utilize one or both sets of *18S* loci that we use in our study (Singh et al., 2004; Jongwutiwes et al., 2004; Yusof et al., 2016). Inclusive taxon sampling appears to be the primary factor contributing to our findings, a result that is consistent with reports in other systems (Pollock et al., 2002; Zwickl and Hillis, 2002; Heath et al., 2008). The inclusion of indel information was not decisive to our findings, but did improve resolution and allow for the detection of host-correlated divisions within S-type loci. Our use of Bayesian inference may also be important. Previous phylogenetic studies on population structure of *P. knowlesi* across humans and macaques have relied on Maximum Parsimony (Singh et al., 2004; Lee et al., 2011), Maximum Likelihood (Singh et al., 2004), and Neighbor-Joining (Singh et al., 2004; Jongwutiwes et al., 2004; Bronner et al., 2009; van Hellemond et al., 2009; Jiang et al., 2010; Jeslyn et al., 2011; Lee et al., 2011) approaches. Bayesian approaches can better account for: multiple mutations at the same site, differential mutation rates across lineages, differential substitution rates between bases, and variation in mutation rates across sites relative to the Neighbor-Joining and Maximum Parsimony Methods that have dominated previous analyses (Holder and Lewis, 2003; Yang and Rannala, 2012, and have been shown to be more accurate than maximum-likelihood methods under some circumstances (Beerli, 2005; Hall, 2005).

Our results support the hypothesis of structuring between human and macaque *P. knowlesi* populations. These findings should, however, be interpreted in the context of several important caveats. First, structuring does not imply isolation of populations and our results do not dispute the occurrence of zoonosis as a cause of human *knowlesi*-malaria. Second, we only examined paralogous *18S* rRNA genes. While these have been commonly utilized in previous phylogenetic studies on *P. knowlesi* (e.g.: Singh et al., 2004; Jongwutiwes et al., 2004; Bronner et al., 2009; van Hellemond et al., 2009; Jiang et al., 2010; Yusof et al., 2016), they may not be representative of its general evolutionary history. Third, the *P. knowlesi* sequences in our study were overwhelmingly taken from humans and long-tailed macaques; *P. knowlesi* strains from pig-tailed macaques (*Macaca nemestrina*) are reported to differ from those seen in long-tailed macaques (Divis et al., 2015; 2017). Though long-tailed macaques are believed to be the primary reservoir for human *P. knowlesi*, more information on the *18S* loci of *P. knowlesi* strains found in pig-tailed macaques may provide additional insights into the transmission of this parasite. Our study is also lacking information on *P. knowlesi* occurrence within vectors. The different vectors of *P. knowlesi* do have differential preferences for humans and macaques (Vythilingam et al., 2016), and adaptation to specific mosquito species could therefore drive apparent patterns of host preference in *P. knowlesi*. Finally, the *P. knowlesi* sequences used in our statistical analyses were strictly from the state of Sarawak in Malaysian Borneo. While this relative sympatry minimizes geographic location as a confounding influence on our analysis, it also hinders extrapolation of our findings to other regions. The *Anopheles* vectors of *P. knowlesi* (Vythilingam et al., 2016) and patterns of human-macaque interactions vary widely across geographic and cultural boundaries in Southeast Asia (Fuentes, 2006), and these may give rise to local variation in the zoonotic transmission of *P. knowlesi*. Multiple genetic analyses have shown strong geographic structuring between *P. knowlesi* strains in Peninsular Malaysia and Malaysian Borneo that clearly outweigh any structuring between hosts (Yusof et al., 2016; Divis et al., 2017), suggesting that human-to-human transmission or other lineage-specific biases in the transmission of *P. knowlesi* to humans, if they are occurring, are a predominantly local phenomenon that has arisen more than once in the context of more recent cross host transmission. Ultimately, considerably more information is needed to draw definitive conclusions about within and across host transmission patterns of *P. knowlesi* in any location. Our results do however, raise important questions about the sources of *P. knowlesi* infections in humans in Malaysian Borneo, and demonstrate the need for a data-driven reevaluation of the largely untested assumption that zoonotic transmission acts as the primary, or even exclusive, source of human *knowlesi*-malaria in this hotspot of human *P. knowlesi*-malaria emergence.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2019.103994.

Declaration of Competing Interest

None.

Acknowledgements

We would like to thank Chissa Rivaldi, Benjamin Gombash, Bailee Egan, and Mitsy Wedderburn for contributing feedback during the writing of this manuscript. This research utilized high-performance computing resources provided by the Notre Dame Center for Research and Computing and the New York University Abu Dhabi Dalma high-performance computing cluster. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This research was supported by funds from the College of Science at the University of Notre Dame.

References

Assefa, S., Lim, C., Preston, M.D., Duffy, C.W., Nair, M.B., Adroub, S.A., Kadir, K.A., Goldberg, J.M., Neafsey, D.E., Divis, P., Clark, T.G., 2015. Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi. Proc. Natl. Acad. Sci. 112 (42), 13027–13032.

Barber, B.E., William, T., Dhararaj, P., Anderios, F., Grigg, M.J., Yeo, T.W., Anstey, N.M., 2012. Epidemiology of Plasmodium knowlesi malaria in north-East Sabah, Malaysia: family clusters and wide age distribution. Malar. J. *11* (1), 401.

Barnard, G.A., 1945. A new test for 2× 2 tables. Nature 156, 177.

Beerli, P., 2005. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22 (3), 341–345.

Belinky, F., Cohen, O., Huchon, D., 2009. Large-scale parsimony analysis of metazoan indels in protein-coding genes. Mol. Biol. Evol. 27 (2), 441–451.

Breitwieser, F.P., Lu, J., Salzberg, S.L., 2017. A review of methods and databases for metagenomic classification and assembly. Brief. Bioinform. https://doi.org/10.1093/bib/bbx120.

Bronner, U., Divis, P.C., Färnert, A., Singh, B., 2009. Swedish traveller with Plasmodium knowlesi malaria after visiting Malaysian Borneo. Malar. J. *8* (1), 15.
Calhoun, P., 2016. Exact: unconditional exact test. R package version 1, 7. https://CRAN.R-project.org/package=Exact.
Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., Cheng, Q., 2008. Comparative genomics of the neglected human malaria parasite Plasmodium vivax. Nature *455* (7214), 757.
Castelloe, J., Templeton, A.R., 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. 3 (2), 102–113.
Chin, W., Contacos, P.G., Collins, W.E., Jeter, M.H., Alpert, E., 1968. Experimental mosquito-transmission of Plasmodium knowlesi to man and monkey. Am. J. Trop. Med. Hyg. 17 (3), 355–358.
Contacos, P.G., Collins, W.E., Jeffery, G.M., Krotoski, W.A., Howard, W.A., 1972. Studies on the characterization of Plasmodium vivax strains from Central America. Am. J. Trop. Med. Hyg. *21* (5_Suppl), 707–712.
R Core Team, 2018. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria, . https://www.R-project.org/.
Cox-Singh, J., Singh, B., 2008. Knowlesi malaria: newly emergent and of public health importance? Trends Parasitol. 24 (9), 406–410.
Cox-Singh, J., Davis, T.M., Lee, K.S., Shamsul, S.S., Matusop, A., Ratnam, S., Rahman, H.A., Conway, D.J., Singh, B., 2008. Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. Clin. Infect. Dis. 46 (2), 165–171.
Davidson, G., Chua, T.H., Cook, A., Speldewinde, P., Weinstein, P., 2019. Defining the ecological and evolutionary drivers of Plasmodium knowlesi transmission within a multi-scale framework. Malar. J. *18* (1), 66.
Divis, P.C., Singh, B., Anderios, F., Hisam, S., Matusop, A., Kocken, C.H., Assefa, S.A., Duffy, C.W., Conway, D.J., 2015. Admixture in humans of two divergent Plasmodium knowlesi populations associated with different macaque host species. PLoS Pathog. *11* (5), e1004888.
Divis, P.C., Lin, L.C., Rovie-Ryan, J.J., Kadir, K.A., Anderios, F., Hisam, S., Sharma, R.S., Singh, B., Conway, D.J., 2017. Three divergent subpopulations of the malaria parasite plasmodium knowlesi. Emerg. Infect. Dis. *23* (4), 616.
Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131 (2), 479–491.
Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28 (23), 3150–3152.
Fuentes, A., 2006. Human culture and monkey behavior: assessing the contexts of potential pathogen transmission between macaques and humans. Am. J. Primatol. 68 (9), 880–896.
Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34 (1), 397–423.
Galen, S.C., Borner, J., Martinsen, E.S., Schaer, J., Austin, C.C., West, C.J., Perkins, S.L., 2018. The polyphyly of Plasmodium: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. R. Soc. Open Sci. *5* (5), 171780.
Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419 (6906), 498–511.
Goudet, J., Raymond, M., de Meeüs, T., Rousset, F., 1996. Testing differentiation in diploid populations. Genetics 144 (4), 1933–1940.
Hall, B.G., 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. Mol. Biol. Evol. 22 (4), 1156–1160.
Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46 (3), 239–257.
Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. *4* (4), 275.
Huelsenbeck, J.P., Rannala, B., 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53 (6), 904–913.
Jeslyn, W.P.S., Huat, T.C., Vernon, L., Irene, L.M.Z., Sung, L.K., Jarrod, L.P., Singh, B., Ching, N.L., 2011. Molecular epidemiological investigation of Plasmodium knowlesi in humans and macaques in Singapore. Vector Borne Zoonotic Dis. 11 (2), 131–135.
Jiang, N., Chang, Q., Sun, X., Lu, H., Yin, J., Zhang, Z., Wahlgren, M., Chen, Q., 2010. Co-infections with Plasmodium knowlesi and other malaria parasites, Myanmar. Emerg. Infect. Dis. *16* (9), 1476.
Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Sata, T., Kanbara, H., 2004. Naturally acquired Plasmodium knowlesi malaria in human, Thailand. Emerg. Infect. Dis. 10 (12), 2211.
Jongwutiwes, S., Buppan, P., Kosuvin, R., Seethamchai, S., Pattanawong, U., Sirichaisinthop, J., Putaporntip, C., 2011. Plasmodium knowlesi malaria in humans and macaques, Thailand. Emerg. Infect. Dis. *17* (10), 1799.
Karuppannan, K., Saaban, S., Mustapa, A.R., Zainal Abidin, F.A., Azimat, N.A., Keliang, C.J., 2014. Population status of long-tailed macaque (Macaca fascicularis) in Peninsular Malaysia. Journal of Primatology 3, 2.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.
Kawamoto, F., Win, T.T., Mizuno, S., Lin, K., Kyaw, O., Tantular, I.S., Mason, D.P., Kimura, M., Wongsrichanalai, C., 2002. Unusual Plasmodium malariae-like parasites in Southeast Asia. J. Parasitol. 88 (2), 350–357.
Kissinger, J.C., Souza, P.C., Soares, C.O., Paul, R., Wahl, A.M., Rathore, D., McCutchan, T.F., Krettli, A.U., 2002. Molecular phylogenetic analysis of the avian malarial parasite Plasmodium (Novyella) juxtanucleare. J. Parasitol. 88 (4), 769–773.
Kobayashi, T., Gamboa, D., Ndiaye, D., Cui, L., Sutton, P.L., Vinetz, J.M., 2015. Malaria diagnosis across the international Centers of excellence for malaria research: platforms, performance, and standardization. Am. J. Trop. Med. Hyg. *93* (3_Suppl), 99–109.
Leclerc, M.C., Hugot, J.P., Durand, P., Renaud, F., 2004. Evolutionary relationships between 15 Plasmodium species from new and Old World primates (including humans): a 18S rDNA cladistic analysis. Parasitology 129 (6), 677–684.
Lee, K.S., Divis, P.C., Zakaria, S.K., Matusop, A., Julin, R.A., Conway, D.J., Cox-Singh, J., Singh, B., 2011. Plasmodium knowlesi: reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog. *7* (4), e1002015.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22 (13), 1658–1659.
Li, J., Wirtz, R.A., McCutchan, T.F., 1997. Analysis of malaria parasite RNA from decade-old Giemsa-stained blood smears and dried mosquitoes. Am. J. Trop. Med. Hyg. 57 (6), 727–731.
Li, Y., Wang, G., Sun, D., Meng, F., Lin, S., Hu, X., Wang, S., 2013. A case of Plasmodium ovale wallikeri infection in a Chinese worker returning from West Africa. Korean J. Parasitol. *51* (5), 557.
Loft, M.H., Lee, C.K.C., Tagg, B.R., Loo, J.M.Y., 2016. Modern Malaysian-based students perceive southeast Asian jungle environments as places of high risk and discomfort. Ecopsychology 8 (1), 35–44.
Loy, D.E., Liu, W., Li, Y., Learn, G.H., Plenderleith, L.J., Sundararaman, S.A., Sharp, P.M., Hahn, B.H., 2017. Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. Int. J. Parasitol. 47 (2), 87–97.
Luan, P.T., Ryder, O.A., Davis, H., Zhang, Y.P., Yu, L., 2013. Incorporating indels as phylogenetic characters: impact for interfamilial relationships within Arctoidea (Mammalia: Carnivora). Mol. Phylogenet. Evol. 66 (3), 748–756.
Lydersen, S., Langaas, M., Bakke, Ø., 2012. The exact unconditional z-pooled test for equality of two binomial probabilities: optimal choice of the Berger and boos confidence coefficient. J. Stat. Comput. Simul. 82 (9), 1311–1316.
Maeno, Y., Culleton, R., Quang, N.T., Kawai, S., Marchand, R.P., Nakazawa, S., 2017. Plasmodium knowlesi and human malaria parasites in khan Phu, Vietnam: gametocyte production in humans and frequent co-infection of mosquitoes. Parasitology 144 (4), 527–535.
Marchand, R.P., Culleton, R., Maeno, Y., Quang, N.T., Nakazawa, S., 2011. Co-infections of Plasmodium knowlesi, P. falciparum, and P. vivax among humans and Anopheles dirus mosquitoes, southern Vietnam. Emerg. Infect. Dis. *17* (7), 1232.
Md-Zain, B.M., Tarmizi, M.R., Mohd-Zaki, M., Fuentes, A., Gumert, M., Jones-Engel, L., 2011. Campus monkeys of Universiti Kebangsaan Malaysia: nuisance problems and student's perception. In: Monkeys on the Edge: Ecology and Management of Long-Tailed Macaques and Their Interface with Humans. *Cambridge University Press, Cambridge*, pp. 101–117.
Mehrotra, D.V., Chan, I.S., Berger, R.L., 2003. A cautionary note on exact unconditional inference for a difference between two independent binomial proportions. Biometrics 59 (2), 441–450.
Mehta, C.R., Senchaudhuri, P., 2003. Conditional versus unconditional exact tests for comparing two binomials. Cytel Softw. Corp. 675, 1–5.
Mu, X.J., Lu, Z.J., Kong, Y., Lam, H.Y., Gerstein, M.B., 2011. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 genomes project. Nucleic Acids Res. 39 (16), 7058–7076.
Nagy, L.G., Kocsubé, S., Csanádi, Z., Kovács, G.M., Petkovits, T., Vágvölgyi, C., Papp, T., 2012. Re-mind the gap! Insertion–deletion data reveal neglected phylogenetic potential of the nuclear ribosomal internal transcribed spacer (ITS) of fungi. PLoS ONE *7* (11), e49794.
Nishimoto, Y., Arisue, N., Kawai, S., Escalante, A.A., Horii, T., Tanabe, K., Hashimoto, T., 2008. Evolution and phylogeny of the heterogeneous cytosolic SSU rRNA genes in the genus Plasmodium. Mol. Phylogenet. Evol. 47 (1), 45–53.
Oksanen, J., Guillaume Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2018. Vegan: community ecology package. In: R Package Version 2.5–3, . https://CRAN.R-project.org/package=vegan.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Perkins, S.L., Schaer, J., 2016. A modern menagerie of mammalian malaria. Trends Parasitol. 32 (10), 772–782.
Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. *51* (4), 664.
Putaporntip, C., Thongaree, S., Jongwutiwes, S., 2013. Differential sequence diversity at merozoite surface protein-1 locus of Plasmodium knowlesi from humans and macaques in Thailand. Infect. Genet. Evol. 18, 213–219.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41 (D1), D590–D596.
Raja, T.N., Hu, T.H., Zainudin, R., Lee, K.S., Perkins, S.L., Singh, B., 2018. Malaria parasites of long-tailed macaques in Sarawak, Malaysian Borneo: a novel species and demographic and evolutionary histories. BMC Evol. Biol. *18* (1), 49.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542.
RStudio, 2016. RStudio: Integrated Development Environment for R (Version 0.99.903) [Computer software]. Boston, MA.
Schumacher, R., 1973. Oökinete development of Plasmodium cathemerium in the niidgut of Culex pipiens fatigans (Rangoon). Proiistologica 9 (1), 65–69.
Seethamchai, S., Putaporntip, C., Malaivijitnond, S., Cui, L., Jongwutiwes, S., 2008. Malaria and Hepatocystis species in wild macaques, southern Thailand. Am. J. Trop. Med. Hyg. 78 (4), 646–653.
Singh, B., Daneshvar, C., 2013. Human infections and detection of Plasmodium knowlesi.

Clin. Microbiol. Rev. 26 (2), 165–184.
Singh, B., Sung, L.K., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., Thomas, A., Conway, D.J., 2004. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet 363 (9414), 1017–1024.
Templeton, A.R., 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol. 7 (4), 381–397.
van Hellemond, J.J., Rutten, M., Koelewijn, R., Zeeman, A.M., Verweij, J.J., Wismans, P.J., Kocken, C.H., van Genderen, P.J., 2009. Human Plasmodium knowlesi infection detected by rapid diagnostic tests for malaria. Emerg. Infect. Dis. *15* (9), 1478.
Vythilingam, I., Wong, M.L., Wan-Yussof, W.S., 2016. Current status of Plasmodium knowlesi vectors: a public health concern? Parasitology 1–9.
Waters, A.P., 1994. The ribosomal RNA genes of Plasmodium. Adv. Parasitol. 34, 33–79.
White, N.J., 2008. Plasmodium knowlesi: the fifth human malaria parasite. Clin. Infect. Dis. 46 (2), 172–173.
Wiersch, S.C., Maier, W.A., Kampen, H., 2005. Plasmodium (Haemamoeba) cathemerium gene sequences for phylogenetic analysis of malaria parasites. Parasitol. Res. 96 (2), 90–94.
William, T., Rahman, H.A., Jelip, J., Ibrahim, M.Y., Menon, J., Grigg, M.J., Yeo, T.W., Anstey, N.M., Barber, B.E., 2013. Increasing incidence of Plasmodium knowlesi malaria following control of P. falciparum and P. vivax malaria in Sabah, Malaysia. PLoS Negl. Trop. Dis. *7* (1), e2026.
Win, T.T., Jalloh, A., Tantular, I.S., Tsuboi, T., Ferreira, M.U., Kimura, M., Kawamoto, F., 2004. Molecular analysis of Plasmodium ovale variants. Emerg. Infect. Dis. *10* (7), 1235.
Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. *13* (5), 303.
Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2013. The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42 (D1), D643–D648.
Yusof, R., Ahmed, M.A., Jelip, J., Ngian, H.U., Mustakim, S., Hussin, H.M., Fong, M.Y., Mahmud, R., Sitam, F.A.T., Japning, J.R.R., Snounou, G., 2016. Phylogeographic evidence for 2 genetically distinct zoonotic Plasmodium knowlesi parasites, Malaysia. Emerg. Infect. Dis. *22* (8), 1371.
Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51 (4), 588–598.