Running title: Molecular and genomic epidemiology of *P. knowlesi*

Molecular epidemiology and population genomics of *Plasmodium knowlesi*

Paul CS Divis [1], Balbir Singh [1] & David J Conway [1,2]

[1] Malaria Research Centre, Universiti Sarawak Malaysia, Kota Samarahan, Sarawak, Malaysia

[2] Department of Infection Biology, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, United Kingdom

Keywords: population genetics, emerging infection, genotyping, genome sequence, microsatellites, mitochondrial DNA.

Abstract

Molecular epidemiology has been central to uncovering *P. knowlesi* as an important cause of human malaria in Southeast Asia, and to understanding the complex nature of this zoonosis. Species-specific parasite detection and characterization of sequences were vital to show that *P. knowlesi* was distinct from the human parasite species that had been presumed to cause malaria. With established sensitive and specific molecular detection tools, informative molecular surveys of malaria infections subsequently indicated the distribution of *P. knowlesi* infections in humans, wild primate reservoir host species, and different mosquito vector species. The importance of studying *P. knowlesi* genetic polymorphism was indicated initially by analysis of a few nuclear loci as well as the mitochondrial genome, and subsequently by multi-locus microsatellite analyses and whole-genome sequencing. Different human infections have distinct *P. knowlesi* genotypes, reflecting the diverse local parasite reservoirs in macaques, although individual human infections are less genetically complex than those of wild macaques which experience more frequent superinfection. Multi-locus analyses revealed deep population subdivisions, structured both geographically and in relation to different primate reservoir host species. Simplified genotypic discrimination assays enabled efficient large-scale surveillance of the sympatric *P. knowlesi* subpopulations within Malaysian Borneo. The whole-genome sequence analyses have also identified loci under recent positive natural selection in the *P. knowlesi* genome, with evidence that different loci are affected in different populations. These provide a foundation to understand recent adaptation of the zoonotic parasite populations, and to track and interpret any future changes as they emerge.

1. Molecular detection in discovery of *Plasmodium knowlesi* as a significant zoonosis

Utilization of molecular tools to study the epidemiology of malaria two decades ago in Malaysian Borneo first led to the discovery of *Plasmodium knowlesi* as a significant cause of human malaria (Singh et al., 2004). At that time, the main cause of malaria in Sarawak state was *P. vivax*, followed by *P. falciparum* and infections diagnosed by microscopy as '*P. malariae'*, although the epidemiology of these '*P. malariae'* infections appeared unusual. While *P. vivax* infections were widely distributed across all the administrative divisions of the state and affected both adults and children, almost half of the '*P. malariae'* cases were reported in the Kapit Division and were mostly in adults. Elsewhere, *P. malariae* infections have usually had relatively low parasitemia (Garnham, 1966), whereas many of the supposed '*P. malariae'* infections in Kapit required hospitalization and had parasitaemia above 5,000 parasites per µL blood. Initial analysis by nested PCR assays based on the small subunit ribosomal (SSUr) RNA genes (Singh et al., 1999) of DNA from 5 blood samples of patients at Kapit Hospital infected with '*P. malariae'* revealed that they contained *Plasmodium* DNA but were negative for *P. malariae* and the other 3 human malaria parasite species.

This initial observation suggested that these patients were either infected with a newly emergent *Plasmodium* species or with a variant of *P. malariae* that had been described elsewhere in Asia (Liu et al., 1998). The first phylogenetic analysis of the partial sequence of the SSU rRNA gene of one of these '*P. malariae*' isolates indicated it was genetically identical with the macaque malaria parasite species *P. knowlesi* and clearly very different from actual *P. malariae*. Subsequent analysis of larger portions of the SSU rRNA genes derived from eight of the supposed '*P. malariae*' patients produced similar results, with parasites having at a high level of sequence identity with the H strain of *P. knowlesi* that had been isolated from a long-tailed macaque in Peninsular Malaysia in 1965 (Chin et al., 1965, Coatney et al., 1971). PCR primers for *P. knowlesi* were developed and together with primers for the human malarias were used to examine blood samples from 208 malaria patients admitted at Kapit Hospital between 2000-2002, 141 (68%) of whom were diagnosed by microscopy as having '*P. malariae'* infections. The nested PCR assays revealed that none of the patients was actually infected with *P. malariae*, and that 120 (58%) had either single *P. knowlesi*

infections or *P. knowlesi* infections mixed with one of the other human malaria parasite species. The reason for the misdiagnosis by microscopy was that asexual forms of *P. knowlesi* look morphologically similar to those of *P. malariae* on stained slides (Lee et al., 2009), indicating that it is essential to utilize molecular detection methods for correct identification of *P. knowlesi*.

2. Molecular surveys of the distribution of *P. knowlesi* infections in humans

2.1. *P. knowlesi* in humans

The findings of a large focus of human infections with *P. knowlesi* in the Kapit Division of Sarawak, led to subsequent reports of human cases of knowlesi malaria at other locations, all of which required the use of molecular methods for identification. The first of these was of a man who had acquired his infection in 2000 in southern Thailand (Jongwutiwes et al., 2004), followed by an extensive study showing the widespread distribution of knowlesi malaria in the states of Sarawak and Sabah in Malaysian Borneo and in Pahang State in Peninsular Malaysia (Cox-Singh et al., 2008) (Table 1). By the end of 2009, knowlesi malaria had been described in Myanmar (Zhu et al., 2006), Singapore (Ng et al., 2008), the Philippines (Luchavez et al., 2008) and Vietnam (Van den Eede et al., 2009), and reports over the following years indicated that the range extended to Kalimantan (Figtree et al., 2010) and Sumatra (Lubis et al., 2017) in Indonesia , Brunei (UK_Health_Protection_Agency, 2011), Laos (Iwagami et al., 2018), and the Andaman and Nicobar Islands of India (Tyagi et al., 2013).

Many of the initial reports from individual countries were of a single or a relatively small number of knowlesi malaria cases, based mainly on malaria patients at hospitals (Table 1). However, as more extensive studies using molecular detection methods were undertaken at communities as well as hospitals, it became clear that human infections with *P. knowlesi* were more prevalent and widespread than previously thought in Thailand (Putaporntip et al., 2009, Jongwutiwes et al., 2011), Myanmar (Jiang et al., 2010), Malaysia (Vythilingam et al., 2008, Cooper et al., 2020, Joveen-Neoh et al., 2011, Naing et al., 2011, Siner et al., 2017, Yusof et al., 2014, Zhu et al., 2006, Cox-Singh et al., 2008), Laos (Pongvongsa et al., 2018),

Vietnam (Marchand et al., 2011, Pongvongsa et al., 2018) and Indonesia (Herdiana et al., 2018). Furthermore, asymptomatic infections of *P. knowlesi* were described in Vietnam (Pongvongsa et al., 2018, Van den Eede et al., 2009), Malaysian Borneo (Siner et al., 2017, Fornace et al., 2016b) and Peninsular Malaysia (Jiram et al., 2016), indicating the importance of community-based sampling to more fully understand the epidemiology of knowlesi malaria.

2.2. *P. knowlesi* in primate reservoir hosts

Long-tailed macaques (*Macaca fascicularis*) were first implicated as natural hosts for *P. knowlesi* by early studies at the Calcutta School of Tropical Medicine in India in the 1930's on macaques imported from Singapore (Knowles and Das Gupta, 1932). Over the following decades, *P. knowlesi* was detected by slide examination of blood from long-tailed macaques sampled elsewhere in Southeast Asia, as well as from pig-tailed macaques (*M. nemestrina*), and occasionally from leaf monkeys (Coatney et al., 1971). More recently, using molecular detection methods, the presence of *P. knowlesi* has been confirmed in wild long-tailed macaques in Peninsular Malaysia (Akter et al., 2015, Vythilingam et al., 2008), Malaysian Borneo (Lee et al., 2011), Thailand (Putaporntip et al., 2010), Singapore (Jeslyn et al., 2011, Li et al., 2021), the Philippines (Gamalo et al., 2019) and Laos (Zhang et al., 2016) (Table 2). Molecular methods have also detected *P. knowlesi* in pig-tailed macaques in Thailand (Putaporntip et al., 2010) and Malaysian Borneo (Lee et al., 2011), in dusky leaf monkeys (*Trachypithecus obscurus*) in Thailand (Putaporntip et al., 2010), and in a stump-tailed macaque (*Macaca arctoides*) in Thailand (Fungfuang et al., 2020) (Table 2).

2.3. *P. knowlesi* in mosquito vectors

Anopheles hackeri was the first vector described for *P. knowlesi* in 1962 and was discovered by allowing the mosquito to feed on a rhesus macaque (*M. mulatta*) followed by observations of the blood stages. More recently, molecular detection assays for malaria parasites have improved the means to correctly identify sporozoites within mosquitoes, thereby making it simpler to identify vectors of particular malaria parasite species. Molecular approaches were first employed in entomological studies to identify vectors of

knowlesi malaria in the Kapit Division of Sarawak, and *An. latens* was incriminated as the vector (Tan et al., 2008, Vythilingam et al., 2006). Similar approaches were utilised in other parts of Malaysian Borneo (Ang et al., 2020, Ang et al., 2021, Brown et al., 2020, Hawkes et al., 2019, Wong et al., 2015), Peninsular Malaysia (Jiram et al., 2012, Vythilingam et al., 2008), and Vietnam (Jiram et al., 2012, Nakazawa et al., 2009, Vythilingam et al., 2008), showing that the main vector species for knowlesi malaria varied at different locations (Table 3). Furthermore, in addition to members of the Leucosphyrus Group which were initially thought to be the only vectors capable of transmitting *P. knowlesi* in nature, members of the Umbrosus Group were also found to be capable of transmitting *P. knowlesi* in the Betong district of Sarawak, Malaysian Borneo (Ang et al., 2021).

3. Early utility of a few genetic loci for analysis of *P. knowlesi* polymorphism

3.1. Initial informative studies involving sequencing of individual genes

A common purpose of sequencing an individual parasite gene is to confirm the identification of the parasite species within a sample from a vertebrate host or mosquito. Although such analysis is limited in scope, it is preferable to merely counting a sample as positive by detecting a PCR product using species-specific primers. A main reason is that natural nucleotide polymorphisms normally commonly occur among parasite samples of any species, and identification of nucleotide differences among different samples can help confirm that PCR-positivity is not a result of laboratory PCR-contamination. Indeed, such natural sequence polymorphisms within-species often exist in loci that are widely used for discriminating different species, such as the small subunit ribosomal RNA (SSUrRNA) gene, or single-copy essential protein-coding genes.

Extending this, analysis of a second or third gene for species confirmation gives added confidence that contamination or misidentification has not occurred. This was done in the initial characterization of *P. knowlesi* as a major zoonosis in Malaysian Borneo (Singh et al., 2004), which is illustrated here by showing sequence variation in the circumsporozoite (*csp*) gene in the first clinical isolates of *P. knowlesi* that were analysed when the existence of the zoonosis was not already known (Figure 1). This was done at the time for further

confirmation, as the sequences of the SSUrRNA gene in the samples had also been analysed in parallel, showing nucleotide polymorphism that clearly indicated that the samples were from a natural *P. knowlesi* population that was unexpected in humans (Singh et al., 2004).

Following this first molecular description of the zoonosis, characterization of *P. knowlesi* clinical infections in Thailand included analysis of SSUrRNA gene sequences, which showed polymorphism as would be expected from a natural zoonotic infection (Putaporntip et al., 2009, Jongwutiwes et al., 2004). Although only a few local samples from macaques positive for *P. knowlesi* were available at the time, they showed sequences similar to those in the human infections, consistent with the local human infections being zoonotic (Putaporntip et al., 2009). Subsequent analysis in Thailand included a few more samples of *P. knowlesi* from macaques along with human cases, and analysed sequences of the merozoite surface protein 1 (*msp1*) gene which gave consistent results, showing that the diversity of sequences among different patients was similar to the diversity of sequences from local macaques (Jongwutiwes et al., 2011).

There are ongoing examples of local identification of zoonoses, where description of natural sequence variation is a part of the initial description, and this is now being done in analyses of *P. cynomolgi* in humans alongside *P. knowlesi*. For example, samples of zoonotic *P. knowlesi* and *P. cynomolgi* infections in the south of Thailand show polymorphism in the mitochondrial cytochrome b (*cytb*) gene within each parasite species, as expected for local zoonoses (Figure 2) (Putaporntip et al., 2021). It is known that both of these species are transmitted to humans from wild macaques, although the numbers of cases of *P. knowlesi* are much higher than those of *P. cynomolgi*.

Analysis of *P. knowlesi* in multiple samples of humans and macaques from the same area was first focused on the Kapit District of Sarawak in Malaysian Borneo, where most zoonotic cases were first described. This showed that the macaques had more *P. knowlesi* genotypes per infection, as illustrated for *csp* gene sequence data (Figure 3), most individual human infections having an unmixed parasite allele sequence and most macaque infections containing multiple alleles, consistent with the expectation that transmission is more common among macaques which causes superinfection of different parasite genotypes.

However, at the local population level the sequence variation overall was similar among different samples from humans and macaques, with no evidence of only a restricted set of genotypes being seen in either host species (Figure 3).

A study of sequence variation in the merozoite surface protein 1 gene (*msp1*) in a small number of *P. knowlesi* isolates from humans and macaques of Thailand suggested higher haplotype diversity in *P. knowlesi* isolated from humans than those of macaques (Putaporntip et al., 2013). However, the numbers of samples were very small, and as *msp1* is under strong natural selection in most parasite species its use as a single marker was not ideal. This emphasized the need to study different genetic loci, ideally including those in which variation is largely selectively neutral.

It is clear that sequence analysis of one or a few genes is useful for confirmation of parasite species, and to give an initial assessment of whether individual samples are genetically different from each other. Beyond this, application of sequencing and genotyping for molecular epidemiological studies of malaria benefit from analysis of multiple genetic loci. The information from the sequence of each gene is very limited, and distribution of polymorphism seen among samples for any one gene is not reliably representative of what would be seen in other genes. As frequent recombination normally occurs in malaria parasite populations, the patterns of polymorphism seen at different loci in the genome are not linked, so analysis of multiple loci is required to understand how individual samples or geographical sub-populations are related.

3.2. Mitochondrial genome sequencing and haplotype relationships

The two extrachromosomal genomes of the parasite, in the mitochondria and apicoplast organelles, are exceptional in this respect. As they are not subject to recombination, analysis of sequences of these organellar genomes allow a phylogenetic perspective on relatedness between individuals or population samples. The greatest information on the mitochondrial and apicoplast lineages is derived from analysis of the entire genomes of 6 kb and 35 kb respectively. For *P. knowlesi*, such analysis was first performed for the mitochondrial genome by comparing samples from humans and wild macaques in the area

of Malaysian Borneo where most cases were initially described (Lee et al., 2011). This confirmed that there was greater within-infection sequence diversity in most macaque infections compared to each human infection, but that the sequence variants were shared among samples from humans and macaques suggesting that there is no distinct genetic subpopulation infecting humans (Figure 4).

Further analysis of the *P. knowlesi* mitochondrial DNA SNPs and haplotype tree structure in the same study suggested that the parasite derived from an ancestral parasite population more than approximately 100,000 years ago, before human settlement in Southeast Asia. As is the case today, the local macaque species were likely the major natural hosts throughout this time. However, the shape of the mitochondrial DNA phylogeny also suggested that *P. knowlesi* may have had significant population expansion approximately 30,000-40,000 years ago, at a time when Borneo was part of mainland Southeast Asia and when the human population was growing in the region (Lee et al., 2011).

Another question arising from early gene sequencing studies is whether *P. knowlesi* had significant geographical population substructure, and initial analyses of sequence variation in individual genes indicated divergence between parasites from Borneo and mainland Southeast Asia. Several separate studies focused on genes encoding the Duffy binding protein (DBP)-α (Fong et al., 2015), one of its paralogues DBP-γ (Fong et al., 2016), a normocyte-binding protein homologue NBPXa (Ahmed et al., 2016), mitochondrial cytochrome oxidase I (Yusof et al., 2016), as well as the SSUrRNA gene (Yusof et al., 2016), each giving concordant overall results. Each of these indicated that, as would be expected, there is significant genetic subdivision between geographical populations of the parasite that have been separated by the South China Sea since the end of the last ice age approximately 13,000 years ago. However, to gain sufficient information for a more complete understanding of population genetic structure requires simultaneous analysis of a larger number of loci. Different approaches may be taken to genotype multiple loci, with most powerful and complete information being derived by whole genome sequencing as explained below.

4. Multi-locus microsatellite analyses of *P. knowlesi* uncovers population structure

As noted above, to study population genetic structure, the use of multiple genetic markers that are not under selection is recommended. Therefore a *P. knowlesi* genotyping toolkit was developed, based on 10 microsatellite loci distributed across the genome (Divis et al., 2015), chosen from a wide range of potential loci identified from examining the original *P. knowlesi* reference genome sequence (Pain et al., 2008). These assays involve targeted amplification of polymorphic simple sequence repeat loci, using a two-step hemi-nested PCR with dye-labelled internal primer for each locus, enabling some multiplexing of loci in the process of electrophoretic analysis of allele sequence lengths.

This toolkit was initially applied to analyse *P. knowlesi* infections in humans and wild macaques from Kapit division of Sarawak state, Malaysian Borneo, and uncovered a quite unexpected population genetic substructure. Of 167 human *P. knowlesi* infections tested, a statistical Bayesian model-based analysis indicated approximately two thirds were of a genetic subpopulation (termed Cluster 1) which is also associated with long-tailed macaque hosts, while approximately one third were of another genetic subpopulation (termed Cluster 2) otherwise associated with pig-tailed macaque hosts (Divis et al., 2015) (Figure 5). The estimated genetic divergence between the two sympatric *P. knowlesi* subpopulations in Malaysian Borneo was substantial, as determined by the distribution of microsatellite allele frequencies (average fixation index F_{ST} ~ 0.22).

Further surveillance using multilocus microsatellite analysis was then conducted on more human and macaque infections across Malaysia. Analysis of 583 *P. knowlesi* infections from nine localities across Malaysian Borneo from the first (Divis et al., 2015) and subsequent (Divis et al., 2017) studies showed Cluster 1 and Cluster 2 subpopulations to be widely distributed in Malaysian Borneo (Figure 6). Cluster 1 parasites were predominant in frequency at most of the sites, although Cluster 2 was more common at two of the sites (in Kanowit and Miri Districts of Sarawak). Although Cluster 2 subpopulation is associated with pig-tailed macaque reservoir hosts, it is unknown if differences in abundance of this macaque species determines the variation in the distribution, or if it is due to different mosquito vectors being responsible, as more studies are needed on the transmission.

Genetic differentiation between *P. knowlesi* infections in Malaysian Borneo and Peninsular Malaysia was expected, given the geographical separation of macaques in these different landmass areas for approximately 13,000 years since the last glacial period (Ziegler et al., 2007, Fa and Lindberg, 1996). Allopatric divergence in the parasite would be inevitably caused by the ocean barrier between mainland Southeast Asia and Borneo which restricts the movement of the macaque hosts (Liedigk et al., 2015). With the inclusion of more *P. knowlesi* clinical infections from Peninsular Malaysia, microsatellite analysis confirmed a genetically separate subpopulation (Cluster 3) divergent from those in Malaysian Borneo (Figure 6) (Divis et al., 2017), which had been first indicated by genome sequencing of older laboratory-maintained isolates (Figure 5) as described below (Assefa et al., 2015).

5. Whole-genome sequence analysis of *P. knowlesi* subpopulation divergence

Divergence between the genetically-distinct *P. knowlesi* subpopulation Clusters 1 and 2 in Malaysian Borneo was supported by the analysis of whole genome sequences from a subset of the clinical isolates that had been analysed by microsatellite analyses (Assefa et al., 2015) (Figure 5), and also seen in a separate study of genome sequences from a few other clinical isolates in Malaysian Borneo (Pinheiro et al., 2015). The population substructure defined by genome sequences gave concordant classification with the microsatellite analysis, and also revealed the first evidence for a third genetic subpopulation (Cluster 3) which was initially represented by older laboratory isolates that had originated from Peninsular Malaysia.

Genome sequence analysis of 103 clinical isolates sampled across Malaysia shows the three subpopulation clusters (Figure 7) (Hocking et al., 2020, Assefa et al., 2015, Divis et al., 2018). A genome-wide scan shows similar level of nucleotide diversity within both Cluster 1 and Cluster 3 subpopulations, but a lower diversity within the Cluster 2 subpopulation (Figure 7). The reduced genetic diversity of the Cluster 2 subpopulation suggests an initial bottleneck during formation of this subpopulation in the pig-tailed macaque natural hosts (Divis et al., 2018). The timing of genetic divergence between the different *P. knowlesi* subpopulation

clusters remains to be determined, but it is likely to be much more recent than when the different macaque reservoir host species diverged.

Pairwise comparisons of genome-wide scans among the three *P. knowlesi* subpopulations shows substantial divergence in all of the 14 different parasite chromosomes (Figure 8). There are some variations in the levels of divergence genome-wide which are usefully examined with each of the pairwise comparisons. For example, comparison between the sympatric Cluster 1 and Cluster 2 from Malaysian Borneo shows particularly high level of divergence in chromosomes 7, 12 and 13, which exceeds the level of divergence in the comparison between the allopatric Cluster 1 and Cluster 3 (the latter being from Peninsular Malaysia) (Hocking et al., 2020).

Genome-wide scan of fixation indices (F_{ST}) between Cluster 1 and Cluster 2 subpopulations shows heterogeneity across the genome due to the mosaic structure of diversity within Cluster 2 (Hocking et al., 2020, Divis et al., 2018). This enabled the identification of genomic islands of differentiation, and definition of high divergence regions (HDR) and low divergence regions (LDR) (Figure 9). Considering the standard deviations of the mean genome-wide F_{ST} values between Cluster 1 and Cluster 2 subpopulations, contiguous windows of HDRs were mainly covering chromosomes 7, 12 and 13, while none were found in chromosome 3, 5 and 10. The Cluster 2 subpopulation showed reduced mean nucleotide diversity in the HDRs compared to Cluster 1, strongly suggesting bottleneck event in the formation of the Cluster 2 subpopulation (Divis et al., 2018).

An initial observation on the population genetic structure of *P. knowlesi* Cluster 3 in Peninsular Malaysia indicates an unexpected feature that needs more attention. A minority of the clinical cases analysed contained highly-related *P. knowlesi* genome sequences, which contrasts to the general pattern in which all other *P. knowlesi* infections have unrelated sequences reflecting acquisition from a highly diverse reservoir population (Hocking et al., 2020). Although uncommon, clinical cases with this unusual parasite type (provisionally termed Cluster 3C) were identified in different hospitals in 3 different states in Peninsular Malaysia, so further sampling and analysis is needed to indicate if this is an emerging new sub-population or if it reflects an unknown and distinct zoonotic cycle (Hocking et al., 2020).

6. Loci under positive natural selection in the *P. knowlesi* genome

The genome-wide scan also showed a strong skew towards low allele frequency variants in all three *P. knowlesi* subpopulations based on the Tajima's *D* statistics, indicating long-term population size expansion (Divis et al., 2018) (Figure 9) . However, the mean genome-wide *D* values for Cluster 2 were lower than for the sympatric Cluster 1. There are also apparent differences when individual genes expected to be under balancing selection are considered. A few well-studied genes that are likely targets of immunity were indicated to be under balancing selection in a subpopulation-specific manner, including the *csp* and *trap* genes in Cluster 1, the *ama1* gene in Cluster 2, and a 6-cysteine protein gene and *msp7*-like gene in Cluster 3 (Hocking et al., 2020, Assefa et al., 2015, Divis et al., 2018). From these examples, it appears that there may be variation in the strength or targets of balancing selection in the divergent subpopulations of *P. knowlesi*.

Moreover, scans for evidence of recent positive directional selection, indicated by extended haplotype homozygosity on particular chromosomal loci, have implicated different loci in different subpopulations. Particularly, the initial study of Cluster 1 in Malaysian Borneo indicated recent selection on several *P. knowlesi* genomic loci, the strongest signature being identified on chromosome 8 (Assefa et al., 2015). This signature of positive selection on chromosome 8 was confirmed by a secondary analysis of the sequence data performed after mapping to an alternatively-generated reference genome sequence (Diez Benavente et al., 2017). In contrast, analysis of the Cluster 3 subpopulation in Peninsular Malaysia did not indicate any strong signature of selection on chromosome 8, but indicated that there are loci on four other chromosomes that have been under recent positive directional selection (Hocking et al., 2020)(Figure 10). This confirms that adaptation and evolution of the different *P. knowlesi* subpopulations is proceeding independently, and they can be considered as distinct zoonoses, from each of which new and distinct parasite phenotypes may emerge. Understanding this is important for the molecular epidemiology of zoonotic *P. knowlesi* malaria, including tracking of the parasite populations as they adapt to environmental changes.

7. Assays for efficient surveillance of different *P. knowlesi* subpopulations

Given the occurrence of two genetically divergent *P. knowlesi* subpopulations in Malaysian Borneo (Clusters 1 and 2) associated with different macaque reservoir hosts (Divis et al., 2015, Assefa et al., 2015), which was revealed by analysing multiple loci as noted above, a simpler genotyping method has been developed to discriminate these in large scale surveys. Allele-specific PCR primers that were diagnostic for the alternative *P. knowlesi* subpopulation Clusters 1 and 2 were designed based on fixed SNP differences identified from the analysis of whole-genome sequences of clinical isolates. Each PCR assay showed a high level of sensitivity and specificity in detecting the respective subpopulation (Divis et al., 2020, Assefa et al., 2015), indicating a potentially useful method to identify the source of *P. knowlesi* infections in humans associated with different macaque host species in Borneo**.**

This simple PCR method of discriminating the sympatric *P. knowlesi* Clusters 1 and 2 is much less costly and time-consuming compared to multi-locus microsatellite analysis or genome sequencing. A first application of this method was performed, in analysis of samples from 1492 infections that had been previously collected in Malaysian Borneo over a 20-year period. This confirmed that overall approximately 70% of human infections were of the Cluster 1 type and approximately 30% of the Cluster 2 type, with few cases containing both (Figure 11). The relative proportions of these vary across Malaysian Borneo, with Cluster 1 being the most common in many areas, but a few areas having Cluster 2 as the predominant subpopulation. It is most plausible that this spatial variation in the relative proportions of these types in human infections is due to differences in the local relative abundance or habitat preference of the two macaque reservoir host species, although this has yet to be confirmed through systematic sampling in different parts of Malaysian Borneo (Figure 11).

Temporal analysis on 1204 *P. knowlesi* infections in Kapit division in Malaysian Borneo show stable relative frequencies over 20 years, with approximately two thirds of infections being of the Cluster 1 type and one third of the Cluster 2 type (Figure 12). This indicates a steady qualitative pattern of transmission of the two divergent parasite subpopulations from macaque hosts to humans, with the relative frequency also remaining similar throughout different times of year. However, the overall numbers of cases and Cluster 1 infections in particular were higher in the most recent year analysed (Figure 12), and temporally varying numbers of overall cases may be related to environmental changes associated with

deforestation, likely to alter the macaque and mosquito vector behavior and distribution, as was observed in Sabah state of Malaysian Borneo in recent years (Fornace et al., 2016a, Cooper et al., 2020).

8. Adaptation and the future of *P. knowlesi* emerging from local zoonoses

The genome sequence analyses have given clear illustration that the different *P. knowlesi* populations are responding to selection, as expected for natural parasite populations, and future adaptation will depend on environmental conditions. The potential speed of *P. knowlesi* adaptation has been well illustrated by laboratory studies, particularly those indicating adaptation of blood stage parasites to culture growth in different types of erythrocytes, a process depending on use of alternative merozoite ligands for invasion (Lim et al., 2013, Moon et al., 2016, Moon et al., 2013). Indeed, the parasite strain from which the reference *P. knowlesi* genome sequence was derived had been originally isolated decades ago (Chin et al., 1965), then maintained by asexual transfer in laboratory macaques for many years, and the *AP2-G2* gene that is normally involved in malaria parasite sexual stage development for mosquito transmission now has a premature stop codon in this line (Assefa et al., 2015).

Loss-of-function genomic changes selected under artificial laboratory conditions have been seen in other malaria parasite species (Claessens et al., 2017), but specific gain-of-function is generally harder to predict and detect, as there are many possible gene functional alterations and phenotype modifications that may only be potentially discerned after highly focused efforts. Laboratory systems for analysis of *P. knowlesi* in culture, including genetic manipulation (Kocken et al., 2002, Mohring et al., 2019) and mosquito infection (Armistead et al., 2018), are potentially amenable for testing whether naturally occurring changes have effects on some candidate measurable phenotypes. However, testing for other phenotypes would require experimental infection studies in laboratory macaques (Galinski et al., 2018), and there are technical and ethical limits to the availability of the most relevant macaque host species, as well as very few laboratory mosquito colonies of natural vector species.

In natural populations, ecological changes due to environmental modifications may alter mosquito feeding behavior and change the habitats for mosquito breeding, resulting in changes in vector composition (Moyes et al., 2016). Local *P. knowlesi* populations are likely to adapt in response to such changes, as illustrated by data indicating that *P. falciparum* has locally-adapted to different mosquito vectors (Molina-Cruz et al., 2020, Ukegbu et al., 2020). Recent molecular entomological surveys in Sarawak show additional *Anopheles* species being incriminated as *P. knowlesi* vectors (Ang et al., 2020, Ang et al., 2021), so we recommend future studies of parasite genomic variation in samples from naturally-infected vectors of different species from throughout the natural range of *P. knowlesi*. Laboratory methods of selective whole genome amplification are able to increase the *P. knowlesi* genome sequence coverage obtained from samples with low amounts of parasite genomic DNA (Diez Benavente et al., 2019), which will be useful for sequencing parasites from individual mosquitoes or low-level infections in some blood samples.

It should also be noted that *P. knowlesi* population genomic analyses have focused on the core genome, as this is most easily analysed by mapping short-read sequences to a reference genome. In contrast, large multigene families such as *SICAVAR* and *KIR* have not been analysed in population studies, as these need different methods for sequence assembly and analysis (Lapp et al., 2018), and understanding their more extreme polymorphism and distinct evolutionary genetic processes may require specialized approaches.

It is important to know whether *P. knowlesi* is adapting to be transmitted more efficiently from humans to mosquitoes, and whether this will cause more cases to be acquired through a human-mosquito-human route. If so, *P. knowlesi* could eventually become endemic, as happened pre-historically for other human malaria parasite species, all of which were originally acquired from non-human primate reservoir hosts. To fully investigate these questions requires multiple disciplines, alongside molecular and genomic epidemiology, but current understanding of the parasite population structure and adaptation does indicate that there is no intrinsic barrier to such potential changes. To date, the underlying population genomics of *P. knowlesi* appears to be dictated by the common reservoir hosts and the geographical structure, and against this background new patterns of emerging

population structure or adaptation might be detected in future. This should encourage more prospective sampling and genome sequence analysis from throughout the range of *P. knowlesi*, at different sites and in different hosts and vectors.

Figure Legends

Figure 1. Natural sequence polymorphisms in *P. knowlesi* were shown in the initial description of zoonotic cases in Malaysian Borneo. This figure shows data for the circumsporozoite protein (*csp*) gene in the first eight clinical isolates analysed. The top panel indicates that the clinical isolates clustered closely with the two previously available *P. knowlesi* sequences from parasites maintained in laboratory macaques, confirming the species identification, and that the sequences of the individual samples were not identical. The bottom panel shows an alignment of the polymorphic nucleotide sites for these samples together with the two controls. A parallel sequence analysis of the same samples was also performed for the SSUrRNA gene, showing similar evidence of natural polymorphism while also clearly revealing the species identification. (Figure incorporating reproduction with permission) (Singh et al., 2004).

Figure 2. Maximum-likelihood phylogenetic tree inferred from the cytochrome oxidase I (*cox1*) gene in the parasite mitochondrial genome. Isolates of *Plasmodium cynomolgi* and *P. knowlesi* from Thailand are compared with selected sequence data previously obtained from these and other species (GenBank accession numbers in parentheses). Colors and symbols indicate different provinces in Thailand from which the individual human cases were sampled. Bootstrap confidence values exceeding 50% are shown on the branches. Scale bar indicates nucleotide substitutions per site. (Figure reproduced under Creative Commons licence) (Putaporntip et al., 2021).

Figure 3. *P. knowlesi* infections of humans are as genetically diverse from each other as infections of local reservoir hosts, but are less mixed genotypically. The panels show data from analyses of *P. knowlesi* circumsporozoite protein gene (*csp*) gene sequences from infections of macaques and humans in the Kapit division of Sarawak in Malaysian Borneo. (A) Histogram showing proportion of human and macaque individuals with different numbers of *P. knowlesi csp* alleles detected per infection. (B) Diversity of *P. knowlesi csp* alleles as indicated by a Neighbor-Joining tree diagram based on a distance matrix of pairwise sequence differences in the non-repeat region of the gene. Figures on the branches are bootstrap confidence percentages above 70%, and scale bar indicates proportion of

nucleotide differences per site. Note that this is not a phylogeny as intragenic recombination occurs among the different alleles of a nuclear gene such as *csp*, so there is no simple within-species phylogeny for nuclear genes. (Figure reproduced under Creative Commons licence) (Lee et al., 2011).

Figure 4. Analysis of sequence variation in the *P. knowlesi* mitochondrial genome (mtDNA) confirms that most human infections are genotypically unmixed, but that they are as diverse from each other as infections in local reservoir hosts. (A) Histogram showing proportion of human and macaque individuals containing different numbers of *P. knowlesi* mtDNA haplotypes per infection. (B) Schematic diagram showing relationship among 37 mtDNA haplotypes of *P. knowlesi*. Numbers in larger circles represent number of haplotypes and unnumbered circles represent a single haplotype. Each line connecting the circles represents a mutational step and black dots represent hypothetical missing intermediates. In contrast to the nuclear genome, the mitochondrial genome haplotypes are not affected by recombination, so they can represent diverging parasite lineages within the species. (Figure adapted under Creative Commons licence) (Lee et al., 2011).

Figure 5. Two divergent *P. knowlesi* subpopulations in Malaysian Borneo associated with different macaque host species. A) Multi-locus microsatellite analysis on *P. knowlesi* isolated from patients and wild macaques in Kapit division of Sarawak reveals an admixture of two divergent subpopulations in humans associated with long-tailed macaque and pig-tailed macaque hosts. B) Unrooted Neighbor-Joining tree confirms the divergence of two sympatric subpopulations in Malaysian Borneo, with an additional cluster in peninsular Malaysia. (Figure incorporating reproduction under Creative Commons licence) (Divis et al., 2015) and (Assefa et al., 2015).

Figure 6. Population genetic structure of *P. knowlesi* across Malaysia, as inferred by multi-locus microsatellite analysis of 751 infections in humans ('hm') and macaques ('lt' indicates long-tailed *M. fascicularis*, and 'pt' pig-tailed *M. nemestrina*). Three divergent subpopulations of *P. knowlesi* are revealed using different tests, illustrated by (A) STRUCTURE k-means cluster assignment analysis and (B) discriminant analysis of principal components (DAPC). While Cluster 1 and Cluster 2 subpopulations occur in Malaysian Borneo with each associated with different macaque hosts, the Cluster 3 subpopulation

occurs in peninsular Malaysia. (Figure incorporating reproduction under Creative Commons licence) (Divis et al., 2017).

Figure 7. Genome-wide sequence analysis of *P. knowlesi* clinical isolates sampled across Malaysia. A) Population genomic structure of *P. knowlesi* infections as shown by a Neighbour-Joining dendrogram of the pairwise genetic distance based on single nucleotide polymorphisms. B) Genome-wide scans on non-overlapping windows of 50 kb show similar pattern of nucleotide diversity for Cluster 1 and Cluster 3 subpopulations across 14 chromosomes which are overall higher than the Cluster 2 subpopulation. (Figure incorporating reproduction under Creative Commons licence) (Hocking et al., 2020).

Figure 8. Genome-wide scan of divergence between *P. knowlesi* subpopulations. The fixation index (F_{ST}) scan of divergence was performed to compare different pairs of subpopulation clusters: A) Cluster 1 versus Cluster 3, B) Cluster 2 versus Cluster 3, and C) Cluster 1 versus Cluster 2. The F_{ST} scores, marked as solid dots, are mean values within windows of 500 consecutive SNPs with overlapping by 250 SNPs. (Figure incorporating reproduction under Creative Commons licence) (Hocking et al., 2020).

Figure 9. Contiguous high and low divergence regions (HDR and LDR) throughout the *P. knowlesi* genome in comparison of Cluster 1 and Cluster 2 subpopulations in Malaysian Borneo. The highest divergence between Cluster 1 and Cluster 2 is observed in chromosomes 7, 12 and 13 where HDRs cover most of the respective chromosome lengths. Both subpopulations show skew towards low-frequency variants, based on scans of Tajima's *D* values in 10-kb windows genome-wide, with Cluster 1 subpopulation exhibiting less genome-wide variation of the frequency spectrum compared to Cluster 2. (Figure incorporating reproduction under Creative Commons licence) (Divis et al., 2018).

Figure 10. Evidence of *P. knowlesi* genomic regions under recent positive directional selection in Peninsular Malaysia (Cluster 3 subpopulation) using the standardised integrated haplotype score |iHS| index. Examination of the ranges of extended haplotype homozygosity for individual SNPs identified four distinct genomic windows with high |iHS| values. Two of these regions (on chromosomes 1 and 9) spanned across members of *SICAvar*

and *kir* multigene families, while the other two (on chromosomes 7 and 12) did not include *SICAvar* or *kir* genes. These genomic regions are different from those implicated to be under selection in the Cluster 1 subpopulation of *P. knowlesi* in Malaysian Borneo in a separate study (Assefa et al., 2015), indicating that signatures of recent selection are population-specific as expected. (Figure incorporating reproduction under Creative Commons licence) (Hocking et al., 2020).

Figure 11. Proportions of human *P. knowlesi* infections with Cluster 1 and Cluster 2 subpopulations across Malaysian Borneo. Cluster 1 subpopulations predominate in many locations compared to Cluster 2 subpopulations, which is consistent with long-tailed macaques being generally more common than pig-tailed macaques, although there may be other reasons for the variation apart from different reservoir host abundance. The genotyping of different divergent subpopulations *P. knowlesi* can be performed rapidly using simple allele-specific PCR assays. (Figure incorporating reproduction under Creative Commons licence) (Divis et al., 2020).

Figure 12. Temporal analysis of frequencies of the two divergent *P. knowlesi* subpopulations in human cases in Kapit division of Sarawak. (A) The relatively high frequency of the Cluster 1 subpopulation compared to Cluster 2 remained steady since 2000, with low numbers of mixed-genotype infections being detected. (B) Analysis of the later years in more detail showed increasing numbers of Cluster 1 in the most recent year, indicating that ongoing surveillance will be needed to identify if there are changes to the previously described distribution. (Figure incorporating reproduction under Creative Commons licence) (Divis et al., 2020)

References

AHMED, M. A., FONG, M. Y., LAU, Y. L. & YUSOF, R. 2016. Clustering and genetic differentiation of the normocyte binding protein (nbpxa) of *Plasmodium knowlesi* clinical isolates from Peninsular Malaysia and Malaysia Borneo. *Malar J,* 15**,** 241.

AKTER, R., VYTHILINGAM, I., KHAW, L. T., QVIST, R., LIM, Y. A., SITAM, F. T., VENUGOPALAN, B. & SEKARAN, S. D. 2015. Simian malaria in wild macaques: first report from Hulu Selangor district, Selangor, Malaysia. *Malar J,* 14**,** 386.

ANG, J. X. D., KADIR, K. A., MOHAMAD, D. S. A., MATUSOP, A., DIVIS, P. C. S., YAMAN, K. & SINGH, B. 2020. New vectors in northern Sarawak, Malaysian Borneo, for the zoonotic malaria parasite, Plasmodium knowlesi. *Parasit Vectors,* 13**,** 472.

ANG, J. X. D., YAMAN, K., KADIR, K. A., MATUSOP, A. & SINGH, B. 2021. New vectors that are early feeders for Plasmodium knowlesi and other simian malaria parasites in Sarawak, Malaysian Borneo. *Sci Rep,* 11**,** 7739.

ARMISTEAD, J. S., MORAES BARROS, R. R., GIBSON, T. J., KITE, W. A., MERSHON, J. P., LAMBERT, L. E., ORR-GONZALEZ, S. E., SA, J. M., ADAMS, J. H. & WELLEMS, T. E. 2018. Infection of mosquitoes from *in vitro* cultivated *Plasmodium knowlesi* H strain. *Int J Parasitol,* 48**,** 601-610.

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., DURAISINGH, M. T., CONWAY, D. J., PAIN, A. & SINGH, B. 2015. Population genomic structure and adaptation in the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc Natl Acad Sci U S A,* 112**,** 13027-32.

BROWN, R., CHUA, T. H., FORNACE, K., DRAKELEY, C., VYTHILINGAM, I. & FERGUSON, H. M. 2020. Human exposure to zoonotic malaria vectors in village, farm and forest habitats in Sabah, Malaysian Borneo. *PLoS Negl Trop Dis,* 14**,** e0008617.

CHIN, W., CONTACOS, P. G., COATNEY, G. R. & KIMBALL, H. R. 1965. A naturally acquired quotidian-type malaria in man transferable to monkeys. *Science,* 149**,** 865.

CLAESSENS, A., AFFARA, M., ASSEFA, S. A., KWIATKOWSKI, D. P. & CONWAY, D. J. 2017. Culture adaptation of malaria parasites selects for convergent loss-of-function mutants. *Sci Rep,* 7**,** 41303.

COATNEY, G. R., COLLINS, W. E., MCWILSON, W. & CONTACOS, P. G. 1971. *The Primate Malarias*, US Department of Health, Education and Welfare.

COOPER, D. J., RAJAHRAM, G. S., WILLIAM, T., JELIP, J., MOHAMMAD, R., BENEDICT, J., ALAZA, D. A., MALACOVA, E., YEO, T. W., GRIGG, M. J., ANSTEY, N. M. & BARBER, B. E. 2020. *Plasmodium knowlesi* Malaria in Sabah, Malaysia, 2015-2017: Ongoing increase in incidence despite near-elimination of the human-only *Plasmodium* species. *Clin Infect Dis,* 70**,** 361-367.

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**,** 165-71.

DIEZ BENAVENTE, E., FLOREZ DE SESSIONS, P., MOON, R. W., HOLDER, A. A., BLACKMAN, M. J., ROPER, C., DRAKELEY, C. J., PAIN, A., SUTHERLAND, C. J., HIBBERD, M. L., CAMPINO, S. & CLARK, T. G. 2017. Analysis of nuclear and organellar genomes of *Plasmodium knowlesi* in humans reveals ancient

population structure and recent recombination among host-specific subpopulations. *PLoS Genet,* 13**,** e1007008.

DIEZ BENAVENTE, E., GOMES, A. R., DE SILVA, J. R., GRIGG, M., WALKER, H., BARBER, B. E., WILLIAM, T., YEO, T. W., DE SESSIONS, P. F., RAMAPRASAD, A., IBRAHIM, A., CHARLESTON, J., HIBBERD, M. L., PAIN, A., MOON, R. W., AUBURN, S., LING, L. Y., ANSTEY, N. M., CLARK, T. G. & CAMPINO, S. 2019. Whole genome sequencing of amplified *Plasmodium knowlesi* DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. *Sci Rep,* 9**,** 9873.

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**,** 616-624.

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**,** e1004888.

DIVIS, P. C. S., DUFFY, C. W., KADIR, K. A., SINGH, B. & CONWAY, D. J. 2018. Genome-wide mosaicism in divergence between zoonotic malaria parasite subpopulations with separate sympatric transmission cycles. *Mol Ecol,* 2**,** 860-870.

DIVIS, P. C. S., HU, T. H., KADIR, K. A., MOHAMMAD, D. S. A., HII, K. C., DANESHVAR, C., CONWAY, D. J. & SINGH, B. 2020. Efficient surveillance of *Plasmodium knowlesi* genetic subpopulations, Malaysian Borneo, 2000-2018. *Emerg Infect Dis,* 26**,** 1392-1398.

FA, J. E. & LINDBERG, D. G. 1996. *Evolution and Ecology of Macaque Societies*, Cambridge University Press.

FIGTREE, M., LEE, R., BAIN, L., KENNEDY, T., MACKERTICH, S., URBAN, M., CHENG, Q. & HUDSON, B. J. 2010. *Plasmodium knowlesi* in human, Indonesian Borneo. *Emerg Infect Dis,* 16**,** 672-4.

FONG, M. Y., RASHDI, S. A., YUSOF, R. & LAU, Y. L. 2015. Distinct genetic difference between the Duffy binding protein (PkDBPalphaII) of *Plasmodium knowlesi* clinical isolates from North Borneo and Peninsular Malaysia. *Malar J,* 14**,** 91.

FONG, M. Y., RASHDI, S. A., YUSOF, R. & LAU, Y. L. 2016. Genetic diversity, Natural Selection and haplotype grouping of *Plasmodium knowlesi* gamma protein Region II (PkgammaRII): Comparison with the Duffy Binding Protein (PkDBPalphaRII). *PLoS One,* 11**,** e0155627.

FORNACE, K. M., ABIDIN, T. R., ALEXANDER, N., BROCK, P., GRIGG, M. J., MURPHY, A., WILLIAM, T., MENON, J., DRAKELEY, C. J. & COX, J. 2016a. Association between Landscape Factors and Spatial Patterns of *Plasmodium knowlesi* Infections in Sabah, Malaysia. *Emerg Infect Dis,* 22**,** 201-8.

FORNACE, K. M., NUIN, N. A., BETSON, M., GRIGG, M. J., WILLIAM, T., ANSTEY, N. M., YEO, T. W., COX, J., YING, L. T. & DRAKELEY, C. J. 2016b. Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia. *J Infect Dis,* 213**,** 784-7.

FUNGFUANG, W., UDOM, C., TONGTHAINAN, D., KADIR, K. A. & SINGH, B. 2020. Malaria parasites in macaques in Thailand: stump-tailed macaques (*Macaca arctoides*) are new natural hosts for *Plasmodium knowlesi*, *Plasmodium inui*, *Plasmodium coatneyi* and *Plasmodium fieldi*. *Malar J,* 19**,** 350.

GALINSKI, M. R., LAPP, S. A., PETERSON, M. S., AY, F., JOYNER, C. J., KG, L. E. R., FONSECA, L. L., VOIT, E. O. & MAHPIC, C. 2018. *Plasmodium knowlesi*: a superb in vivo

nonhuman primate model of antigenic variation in malaria. *Parasitology,* 145**,** 85-100.

GAMALO, L. E., DIMALIBOT, J., KADIR, K. A., SINGH, B. & PALLER, V. G. 2019. *Plasmodium knowlesi* and other malaria parasites in long-tailed macaques from the Philippines. *Malar J,* 18**,** 147.

GARNHAM, P. C. C. 1966. *Malaria Parasites and other Haemosporidia,* Oxford, Blackwell Scientific Publications.

HAWKES, F. M., MANIN, B. O., COOPER, A., DAIM, S., R, H., JELIP, J., HUSIN, T. & CHUA, T. H. 2019. Vector compositions change across forested to deforested ecotones in emerging areas of zoonotic malaria transmission in Malaysia. *Sci Rep,* 9**,** 13312.

HERDIANA, H., IRNAWATI, I., COUTRIER, F. N., MUNTHE, A., MARDIATI, M., YUNIARTI, T., SARIWATI, E., SUMIWI, M. E., NOVIYANTI, R., PRONYK, P. & HAWLEY, W. A. 2018. Two clusters of *Plasmodium knowlesi* cases in a malaria elimination area, Sabang Municipality, Aceh, Indonesia. *Malar J,* 17**,** 186.

HOCKING, S. E., DIVIS, P. C. S., KADIR, K. A., SINGH, B. & CONWAY, D. J. 2020. Population genomic structure and recent evolution of *Plasmodium knowlesi*, Peninsular Malaysia. *Emerg Infect Dis,* 26**,** 1749-1758.

IWAGAMI, M., NAKATSU, M., KHATTIGNAVONG, P., SOUNDALA, P., LORPHACHAN, L., KEOMALAPHET, S., XANGSAYALATH, P., KAWAI, S., HONGVANTHONG, B., BREY, P. T. & KANO, S. 2018. First case of human infection with *Plasmodium knowlesi* in Laos. *PLoS Negl Trop Dis,* 12**,** e0006244.

JESLYN, W. P., HUAT, T. C., VERNON, L., IRENE, L. M., 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**,** 131-5.

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**,** 1476-8.

JIRAM, A. I., HISAM, S., REUBEN, H., HUSIN, S. Z., ROSLAN, A. & ISMAIL, W. R. W. 2016. Submicroscopic evidence of the simian malaria parasite, *Plasmodium knowlesi*, in an Orang Asli community. *Southeast Journal of Tropical Medicine and Public Health,* 47**,** 591-599.

JIRAM, A. I., VYTHILINGAM, I., NOORAZIAN, Y. M., YUSOF, Y. M., AZAHARI, A. H. & FONG, M. Y. 2012. Entomologic investigation of *Plasmodium knowlesi* vectors in Kuala Lipis, Pahang, Malaysia. *Malar J,* 11**,** 213.

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**,** 1799-806.

JONGWUTIWES, S., PUTAPORNTIP, C., IWASAKI, T., SATA, T. & KANBARA, H. 2004. Naturally acquired *Plasmodium knowlesi* malaria in human, Thailand. *Emerg Infect Dis,* 10**,** 2211-3.

JOVEEN-NEOH, W. F., CHONG, K. L., WONG, C. M. & LAU, T. Y. 2011. Incidence of malaria in the interior division of Sabah, Malaysian Borneo, based on nested PCR. *J Parasitol Res,* 2011**,** 104284.

KHIM, N., SIV, S., KIM, S., MUELLER, T., FLEISCHMANN, E., SINGH, B., DIVIS, P. C., STEENKESTE, N., DUVAL, L., BOUCHIER, C., DUONG, S., ARIEY, F. & MENARD, D. 2011. *Plasmodium knowlesi* Infection in Humans, Cambodia, 2007-2010. *Emerg Infect Dis,* 17**,** 1900-2.

KNOWLES, R. & DAS GUPTA, B. M. 1932. A study of monkey malaria and its experimental transmission to man. *Indian Medical Gazette,* 67**,** 301-320.

KOCKEN, C. H., OZWARA, H., VAN DER WEL, A., BEETSMA, A. L., MWENDA, J. M. & THOMAS, A. W. 2002. *Plasmodium knowlesi* provides a rapid in vitro and in vivo transfection system that enables double-crossover gene knockout studies. *Infect Immun,* 70**,** 655-60.

LAPP, S. A., GERALDO, J. A., CHIEN, J. T., AY, F., PAKALA, S. B., BATUGEDARA, G., HUMPHREY, J., MA, H. C., DE, B. J., LE ROCH, K. G., GALINSKI, M. R. & KISSINGER, J. C. 2018. PacBio assembly of a *Plasmodium knowlesi* genome sequence with Hi-C correction and manual annotation of the SICAvar gene family. *Parasitology,* 145**,** 71-84.

LEE, K. S., COX-SINGH, J. & SINGH, B. 2009. Morphological features and differential counts of *Plasmodium knowlesi* parasites in naturally acquired human infections. *Malar J,* 8**,** 73.

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**,** e1002015.

LI, M. I., MAILEPESSOV, D., VYTHILINGAM, I., LEE, V., LAM, P., NG, L. C. & TAN, C. H. 2021. Prevalence of simian malaria parasites in macaques of Singapore. *PLoS Negl Trop Dis,* 15**,** e0009110.

LIEDIGK, R., KOLLECK, J., BOKER, K. O., MEIJAARD, E., MD-ZAIN, B. M., ABDUL-LATIFF, M. A., AMPENG, A., LAKIM, M., ABDUL-PATAH, P., TOSI, A. J., BRAMEIER, M., ZINNER, D. & ROOS, C. 2015. Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). *BMC Genomics,* 16**,** 222.

LIM, C., HANSEN, E., DESIMONE, T. M., MORENO, Y., JUNKER, K., BEI, A., BRUGNARA, C., BUCKEE, C. O. & DURAISINGH, M. T. 2013. Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat Commun,* 4**,** 1638.

LIU, Q., ZHU, S., MIZUNO, S., KIMURA, M., LIU, P., ISOMURA, S., WANG, X. & KAWAMOTO, F. 1998. Sequence variation in the small-subunit rRNA gene of *Plasmodium malariae* and prevalence of isolates with the variant sequence in Sichuan, China. *Journal of Clinical Microbiology,* 36**,** 3378-3381.

LUBIS, I. N. D., WIJAYA, H., LUBIS, M., LUBIS, C. P., DIVIS, P. C. S., BESHIR, K. B. & SUTHERLAND, C. J. 2017. Contribution of *Plasmodium knowlesi* to multispecies human malaria infections in North Sumatera, Indonesia. *J Infect Dis,* 215**,** 1148-1155.

LUCHAVEZ, J., ESPINO, F., CURAMENG, P., ESPINA, R., BELL, D., CHIODINI, P., NOLDER, D., SUTHERLAND, C., LEE, K. S. & SINGH, B. 2008. Human Infections with *Plasmodium knowlesi*, the Philippines. *Emerg Infect Dis,* 14**,** 811-3.

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**,** 1232-9.

MOHRING, F., HART, M. N., RAWLINSON, T. A., HENRICI, R., CHARLESTON, J. A., DIEZ BENAVENTE, E., PATEL, A., HALL, J., ALMOND, N., CAMPINO, S., CLARK, T. G., SUTHERLAND, C. J., BAKER, D. A., DRAPER, S. J. & MOON, R. W. 2019. Rapid and iterative genome editing in the malaria parasite *Plasmodium knowlesi* provides new tools for *P. vivax* research. *Elife,* 8.

MOLINA-CRUZ, A., CANEPA, G. E., ALVES, E. S. T. L., WILLIAMS, A. E., NAGYAL, S., YENKOIDIOK-DOUTI, L., NAGATA, B. M., CALVO, E., ANDERSEN, J., BOULANGER, M. J. & BARILLAS-MURY, C. 2020. *Plasmodium falciparum* evades immunity of

anopheline mosquitoes by interacting with a Pfs47 midgut receptor. *Proc Natl Acad Sci U S A,* 117**,** 2597-2605.

MOON, R. W., HALL, J., RANGKUTI, F., HO, Y. S., ALMOND, N., MITCHELL, G. H., PAIN, A., HOLDER, A. A. & BLACKMAN, M. J. 2013. Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A,* 110**,** 531-6.

MOON, R. W., SHARAF, H., HASTINGS, C. H., HO, Y. S., NAIR, M. B., RCHIAD, Z., KNUEPFER, E., RAMAPRASAD, A., MOHRING, F., AMIR, A., YUSUF, N. A., HALL, J., ALMOND, N., LAU, Y. L., PAIN, A., BLACKMAN, M. J. & HOLDER, A. A. 2016. Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc Natl Acad Sci U S A*.

MOYES, C. L., SHEARER, F. M., HUANG, Z., WIEBE, A., GIBSON, H. S., NIJMAN, V., MOHD-AZLAN, J., BRODIE, J. F., MALAIVIJITNOND, S., LINKIE, M., SAMEJIMA, H., O'BRIEN, T. G., TRAINOR, C. R., HAMADA, Y., GIORDANO, A. J., KINNAIRD, M. F., ELYAZAR, I. R., SINKA, M. E., VYTHILINGAM, I., BANGS, M. J., PIGOTT, D. M., WEISS, D. J., GOLDING, N. & HAY, S. I. 2016. Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas. *Parasit Vectors,* 9**,** 242.

NAING, D. K. S., ANDERIOS, F. & KIN, Z. 2011. Geographic and ethnic distributions of *Plasmodium knowlesi* infection in Sabah, Malaysia. *International Journal of Collaborative Research on Internal Medicine and Public Health,* 3**,** 9.

NAKAZAWA, S., MARCHAND, R. P., QUANG, N. T., CULLETON, R., MANH, N. D. & MAENO, Y. 2009. *Anopheles dirus* co-infection with human and monkey malaria parasites in Vietnam. *Int J Parasitol,* 39**,** 1533-7.

NG, O. T., OOI, E. E., LEE, C. C., LEE, P. J., NG, L. C., PEI, S. W., TU, T. M., LOH, J. P. & LEO, Y. S. 2008. Naturally acquired human *Plasmodium knowlesi* infection, Singapore. *Emerg Infect Dis,* 14**,** 814-6.

PAIN, A., BOHME, U., BERRY, A. E., MUNGALL, K., FINN, R. D., JACKSON, A. P., MOURIER, T., MISTRY, J., PASINI, E. M., ASLETT, M. A., BALASUBRAMMANIAM, S., BORGWARDT, K., BROOKS, K., CARRET, C., CARVER, T. J., CHEREVACH, I., CHILLINGWORTH, T., CLARK, T. G., GALINSKI, M. R., HALL, N., HARPER, D., HARRIS, D., HAUSER, H., IVENS, A., JANSSEN, C. S., KEANE, T., LARKE, N., LAPP, S., MARTI, M., MOULE, S., MEYER, I. M., ORMOND, D., PETERS, N., SANDERS, M., SANDERS, S., SARGEANT, T. J., SIMMONDS, M., SMITH, F., SQUARES, R., THURSTON, S., TIVEY, A. R., WALKER, D., WHITE, B., ZUIDERWIJK, E., CHURCHER, C., QUAIL, M. A., COWMAN, A. F., TURNER, C. M., RAJANDREAM, M. A., KOCKEN, C. H., THOMAS, A. W., NEWBOLD, C. I., BARRELL, B. G. & BERRIMAN, M. 2008. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature,* 455**,** 799-803.

PINHEIRO, M. M., AHMED, M. A., MILLAR, S. B., SANDERSON, T., OTTO, T. D., LU, W. C., KRISHNA, S., RAYNER, J. C. & COX-SINGH, J. 2015. *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. *PLoS One,* 10**,** e0121303.

PONGVONGSA, T., CULLETON, R., HA, H., THANH, L., PHONGMANY, P., MARCHAND, R. P., KAWAI, S., MOJI, K., NAKAZAWA, S. & MAENO, Y. 2018. Human infection with *Plasmodium knowlesi* on the Laos-Vietnam border. *Trop Med Health,* 46**,** 33.

PUTAPORNTIP, C., HONGSRIMUANG, T., SEETHAMCHAI, S., KOBASA, T., LIMKITTIKUL, K., CUI, L. & JONGWUTIWES, S. 2009. Differential prevalence of *Plasmodium*

infections and cryptic *Plasmodium knowlesi* malaria in humans in Thailand. *J Infect Dis,* 199**,** 1143-50.
PUTAPORNTIP, C., JONGWUTIWES, S., THONGAREE, S., SEETHAMCHAI, S., GRYNBERG, P. & HUGHES, A. L. 2010. Ecology of malaria parasites infecting Southeast Asian macaques: evidence from cytochrome b sequences. *Mol Ecol,* 19**,** 3466-76.
PUTAPORNTIP, C., KUAMSAB, N., PATTANAWONG, U., YANMANEE, S., SEETHAMCHAI, S. & JONGWUTIWES, S. 2021. *Plasmodium cynomolgi* co-infections among symptomatic malaria patients, Thailand. *Emerg Infect Dis,* 27**,** 590-593.
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**,** 213-9.
SINER, A., LIEW, S. T., KADIR, K. A., MOHAMAD, D. S. A., THOMAS, F. K., ZULKARNAEN, M. & SINGH, B. 2017. Absence of *Plasmodium inui* and *Plasmodium cynomolgi*, but detection of *Plasmodium knowlesi* and *Plasmodium vivax* infections in asymptomatic humans in the Betong division of Sarawak, Malaysian Borneo. *Malar J,* 16**,** 417.
SINGH, B., BOBOGARE, A., COX-SINGH, J., SNOUNOU, G., ABDULLAH, M. S. & RAHMAN, H. A. 1999. A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiological studies. *American Journal of Tropical Medicine and Hygiene,* 60**,** 687-692.
SINGH, B., KIM SUNG, L., 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**,** 1017-24.
TAN, C. H., VYTHILINGAM, I., MATUSOP, A., CHAN, S. T. & SINGH, B. 2008. Bionomics of *Anopheles latens* in Kapit, Sarawak, Malaysian Borneo in relation to the transmission of zoonotic simian malaria parasite *Plasmodium knowlesi*. *Malar J,* 7**,** 52.
TYAGI, R. K., DAS, M. K., SINGH, S. S. & SHARMA, Y. D. 2013. Discordance in drug resistance-associated mutation patterns in marker genes of *Plasmodium falciparum* and *Plasmodium knowlesi* during coinfections. *Journal of Antimicrobial Chemotherapy,* 68**,** 1081-1088.
UK_HEALTH_PROTECTION_AGENCY 2011. Imported malaria cases and deaths, United Kingdom: 1992-2011. London: Health Protection Agency.
UKEGBU, C. V., GIORGALLI, M., TAPANELLI, S., RONA, L. D. P., JAYE, A., WYER, C., ANGRISANO, F., BLAGBOROUGH, A. M., CHRISTOPHIDES, G. K. & VLACHOU, D. 2020. PIMMS43 is required for malaria parasite immune evasion and sporogonic development in the mosquito vector. *Proc Natl Acad Sci U S A*.
VAN DEN EEDE, P., VAN, H. N., VAN OVERMEIR, C., VYTHILINGAM, I., DUC, T. N., HUNG LE, X., MANH, H. N., ANNE, J., D'ALESSANDRO, U. & ERHART, A. 2009. Human *Plasmodium knowlesi* infections in young children in central Vietnam. *Malar J,* 8**,** 249.
VYTHILINGAM, I., NOORAZIAN, Y. M., HUAT, T. C., JIRAM, A. I., YUSRI, Y. M., AZAHARI, A. H., NORPARINA, I., NOORRAIN, A. & LOKMANHAKIM, S. 2008. *Plasmodium knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. *Parasit Vectors,* 1**,** 26.
VYTHILINGAM, I., TAN, C. H., ASMAD, M., CHAN, S. T., LEE, K. S. & SINGH, B. 2006. Natural transmission of *Plasmodium knowlesi* to humans by *Anopheles latens* in Sarawak, Malaysia. *Trans R Soc Trop Med Hyg,* 100**,** 1087-8.

WONG, M. L., CHUA, T. H., LEONG, C. S., KHAW, L. T., FORNACE, K., WAN-SULAIMAN, W. Y., WILLIAM, T., DRAKELEY, C., FERGUSON, H. M. & VYTHILINGAM, I. 2015. Seasonal and spatial dynamics of the primary vector of *Plasmodium knowlesi* within a major transmission focus in Sabah, Malaysia. *Plos Neglected Tropical Diseases,* 9.

YUSOF, R., AHMED, M. A., JELIP, J., NGIAN, H. U., MUSTAKIM, S., HUSSIN, H. M., FONG, M. Y., MAHMUD, R., SITAM, F. A., JAPNING, J. R., SNOUNOU, G., ESCALANTE, A. A. & LAU, Y. L. 2016. Phylogeographic evidence for 2 genetically distinct zoonotic *Plasmodium knowlesi* parasites, Malaysia. *Emerg Infect Dis,* 22**,** 1371-80.

YUSOF, R., LAU, Y. L., MAHMUD, R., FONG, M. Y., JELIP, J., NGIAN, H. U., MUSTAKIM, S., HUSSIN, H. M., MARZUKI, N. & MOHD ALI, M. 2014. High proportion of knowlesi malaria in recent malaria cases in Malaysia. *Malar J,* 13**,** 168.

ZHANG, X., KADIR, K. A., QUINTANILLA-ZARINAN, L. F., VILLANO, J., HOUGHTON, P., DU, H., SINGH, B. & SMITH, D. G. 2016. Distribution and prevalence of malaria parasites among long-tailed macaques (Macaca fascicularis) in regional populations across Southeast Asia. *Malar J,* 15**,** 450.

ZHU, H. M., LI, J. & ZHENG, H. 2006. [Human natural infection of Plasmodium knowlesi]. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi,* 24**,** 70-1.

ZIEGLER, T., ABEGG, C., MEIJAARD, E., PERWITASARI-FARAJALLAH, D., WALTER, L., HODGES, J. K. & ROOS, C. 2007. Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the M. silenus group. *Mol Phylogenet Evol,* 42**,** 807-16.

Table 1. Initial reports of human *P. knowlesi* malaria cases at various locations in Asia using molecular methods of detection

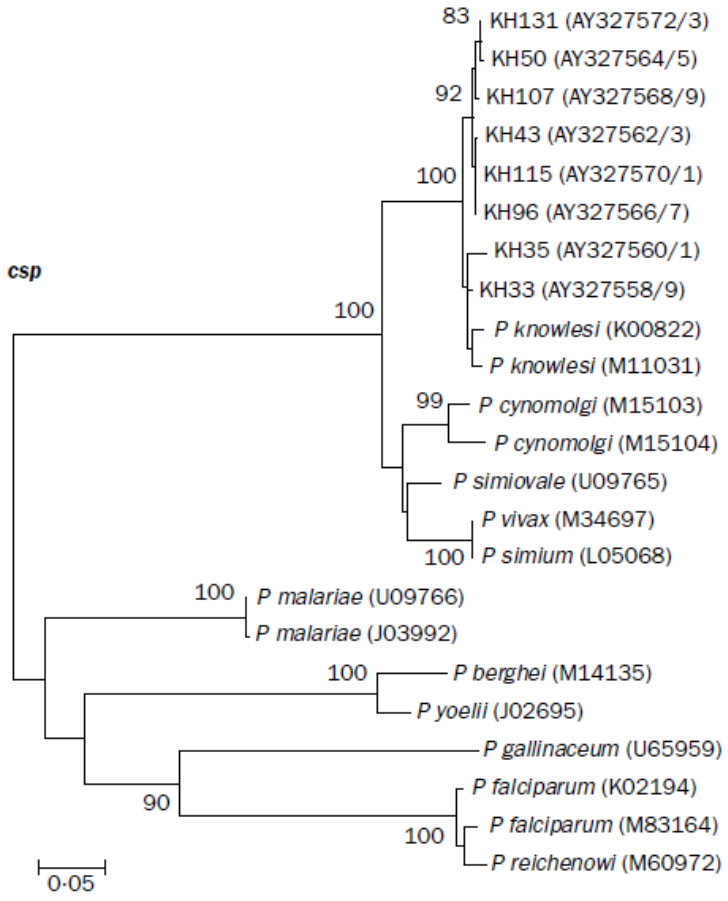
Location	**Detection methods used**	**No of cases**	**Dates of infections**	**Reference**
Sarawak, Malaysian Borneo	Nested PCR and sequencing of SSU rRNA and *csp* genes	120	2000-2002	(Singh et al., 2004)
Prachuap Khiri Khan Province, Thailand	Sequencing of SSU rRNA and *cytb* genes	1	2000	(Jongwutiwes et al., 2004)
Myanmar	Nested PCR	1	1998	(Zhu et al., 2006)
Pahang State, Peninsular Malaysia	Nested PCR	5	2004-2005	(Cox-Singh et al., 2008)
Sabah State, Malaysian Borneo	Nested PCR	41	2003-2005	(Cox-Singh et al., 2008)
Palawan Island, Philippines	Nested PCR	5	2006	(Luchavez et al., 2008)
Singapore	Nested PCR and sequencing of SSU rRNA genes	1	2007	(Ng et al., 2008)
South Kalimantan, Indonesian Borneo	Nested PCR and sequencing of SSU rRNA genes	1	2010	(Figtree et al., 2010)
Cambodia	Nested PCR and sequencing of SSU rRNA genes	2	2007-2010	(Khim et al., 2011)
Southern Vietnam, Vietnam	Nested PCR and sequencing of SSU rRNA genes	3	2004	(Van den Eede et al., 2009)
Brunei	Not stated	1	2006	(UK_Health_Protection_Agency, 2011)
Andaman and Nicobar Islands, India	Nested PCR and sequencing of SSU rRNA and *msp1* genes	53	2004-2010	(Tyagi et al., 2013)
North Sumatra, Indonesia	Nested PCR and hemi-nested PCR	377	2015	(Lubis et al., 2017)
Laos	Real-time PCR and sequencing of *cytb* and *msp1* genes	1	2016	(Iwagami et al., 2018)

Table 2. Prevalence of *Plasmodium knowlesi* infections in non-human primate host species using molecular methods of detection

Locality	Non-human primate	Total no. examined	*P.k.* positive	*% P.k.* positive	Dates of samples	Reference
Sarawak, Malaysian Borneo	*Macaca fascicularis*	82	71	86.6	2004 - 2008	(Lee et al., 2011)
Sarawak, Malaysian Borneo	*Macaca nemestrina*	26	13	50	2004 - 2008	(Lee et al., 2011)
Pahang State, Peninsular Malaysia	*M. fascicularis*	75	10	6.9	2007	(Vythilingam et al., 2008)
Hulu Selangor, Peninsular Malaysia	*M. fascicularis*	70	21	30	2014	(Akter et al., 2015)
Western Catchment Area, Singapore	*M. fascicularis*	3	3	100	2007 & 2009	(Jeslyn et al., 2011)
Western Catchment Area, Singapore	*M. fascicularis*	379	142	37.5	2009 - 2017	(Li et al., 2021)
Narathiwat Province, Southern Thailand	*M. fascicularis*	186	1	0.5	2008 - 2009	(Putaporntip et al., 2010)
Narathiwat Province, Southern Thailand	*M. nemestrina*	373	5	1.3	2008 - 2009	(Putaporntip et al., 2010)
Narathiwat Province, Southern Thailand	*Trachypithecus obscurus*	7	1	14.3	2008 - 2009	(Putaporntip et al., 2010)
Prachuap Kiri Khan Province, Central Thailand	*Macaca arctoides*	32	1	3.1	2017 - 2019	(Fungfuang et al., 2020)
Laos	*M. fascicularis*	44	1	2.3	2013	(Zhang et al., 2016)
Palawan Island, Philippines	*M. fascicularis*	95	18	18.9	2017	(Gamalo et al., 2019)

Table 3. Surveys identifying *P. knowlesi* infections in mosquito vector species using molecular methods of detection

Locality	**Vector identified**	**Detection method**	**Period of study**	**Reference**
Kapit district, Sarawak State, Malaysian Borneo	*An. latens*	Nested PCR	2005-2006	(Vythilingam et al., 2006, Tan et al., 2008)
Lawas district, Sarawak State, Malaysian Borneo	*An. latens, An. donaldi*	Nested PCR and sequencing of SSU rRNA gene	2014-2015	(Ang et al., 2020)
Betong district, Sarawak State, Malaysian Borneo	*An. latens, An. introlatus, An. roperi, An. collessi*	Nested PCR and sequencing of SSU rRNA gene	2015-2016	(Ang et al., 2021)
Kuala Lipis district, Pahang State, Peninsular Malaysia	*An. cracens*	Nested PCR	2007-2008	(Vythilingam et al., 2008, Jiram et al., 2012)
Banggi Island and Kudat district, Sabah State, Malaysian Borneo	*An. balabacensis*	Nested PCR	2013-2015	(Wong et al., 2015)
Ranau district, Sabah State, Malaysian Borneo	*An. balabacensis*	Nested PCR	2015-2016	(Hawkes et al., 2019)
Keningau district, Sabah State, Malaysian Borneo	*An. donaldi*	Nested PCR	2015-2016	(Hawkes et al., 2019)
Pitas district, Sabah State, Malaysian Borneo	*An. balabacensis*	Nested PCR	2016	(Brown et al., 2020)
Khanh Vinh district, Khanh Hoa Province, Vietnam	*An. dirus*	Nested PCR, PCR and sequencing of *csp* gene	2008	(Nakazawa et al., 2009)
Khanh Phu, Khanh Hoa Province, Vietnam	*An. dirus*	Nested PCR and PCR of *csp* gene	2009-2010	(Marchand et al., 2011)



csp	55	523	903	1008	1149	1450	1556	1557	1592	1597	1849	8643	9033	9941	9943	9944	9945	9946	9951	9953	9957	9993	10020	10056
PkH	A	C	C	G	C	A	G	G	A	G	C	G	C	T	A	G	C	T	C	A	G	T	G	C	C
PkN	.	.	A	.	G	C	A	A	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.
KH33	G	.	.	.	G	C	A	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	G
KH35	.	T	.	.	G	C	A	.	.	.	.	.	.	.	G	C	A	G	A	T	.	.	.	.	G
KH43	.	T	.	.	G	C	A	A	.	.	.	A	G	.	G	.	.	.	.	.	A	G	A	T	G
KH50	G	.	.	.	G	C	A	A	G	A	.	A	G	.	G	.	.	.	.	.	A	G	A	.	G
KH96	.	.	.	.	G	C	A	A	.	.	.	A	G	.	G	.	.	.	.	.	A	G	A	T	G
KH107	.	.	.	A	G	C	A	A	.	A	.	A	G	.	G	.	.	.	.	.	A	G	A	T	G
KH115	.	.	.	.	G	C	A	A	.	.	.	A	G	.	G	.	.	.	.	.	A	G	A	T	G
KH131	G	.	.	.	G	C	A	A	G	A	.	A	G	.	G	.	.	.	.	.	A	G	A	T	G

Figure 1.

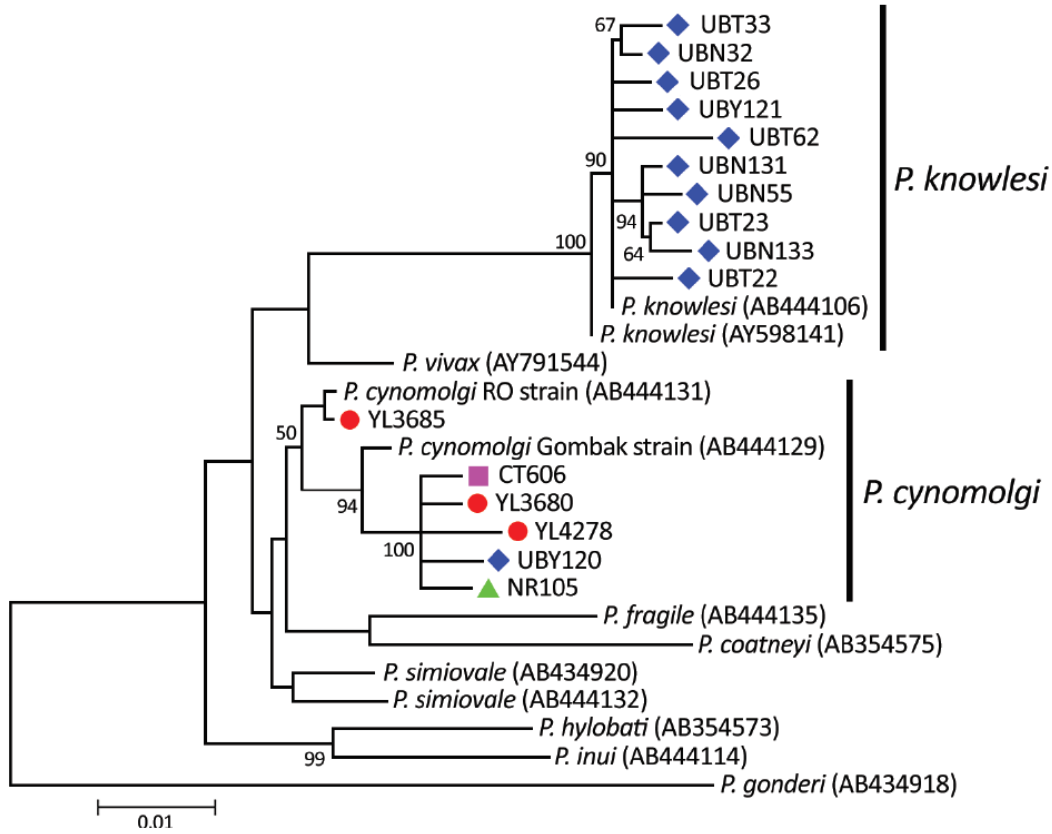

67
UBT33
UBN32
UBT26
UBY121
UBT62
90
UBN131
UBN55
94
UBT23
64
UBN133
UBT22
100
P. knowlesi (AB444106)
P. knowlesi (AY598141)
P. knowlesi
P. vivax (AY791544)
P. cynomolgi RO strain (AB444131)
YL3685
50
P. cynomolgi Gombak strain (AB444129)
CT606
94
YL3680
YL4278
100
UBY120
NR105
P. cynomolgi
P. fragile (AB444135)
P. coatneyi (AB354575)
P. simiovale (AB434920)
P. simiovale (AB444132)
P. hylobati (AB354573)
99
P. inui (AB444114)
P. gonderi (AB434918)
0.01

Figure 2.

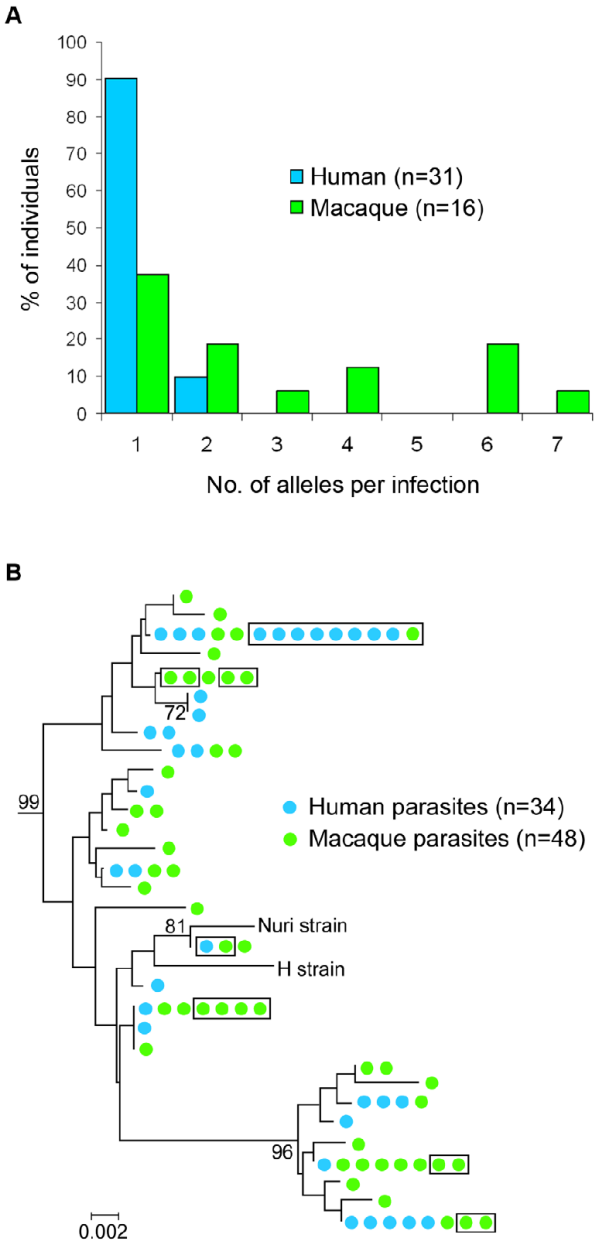

A
100
90
80
70
60
50
40
30
20
10
0
% of individuals
Human (n=31)
Macaque (n=16)
1
2
3
4
5
6
7
No. of alleles per infection
B
72
99
Human parasites (n=34)
Macaque parasites (n=48)
81
Nuri strain
H strain
96
0.002

Figure 3.

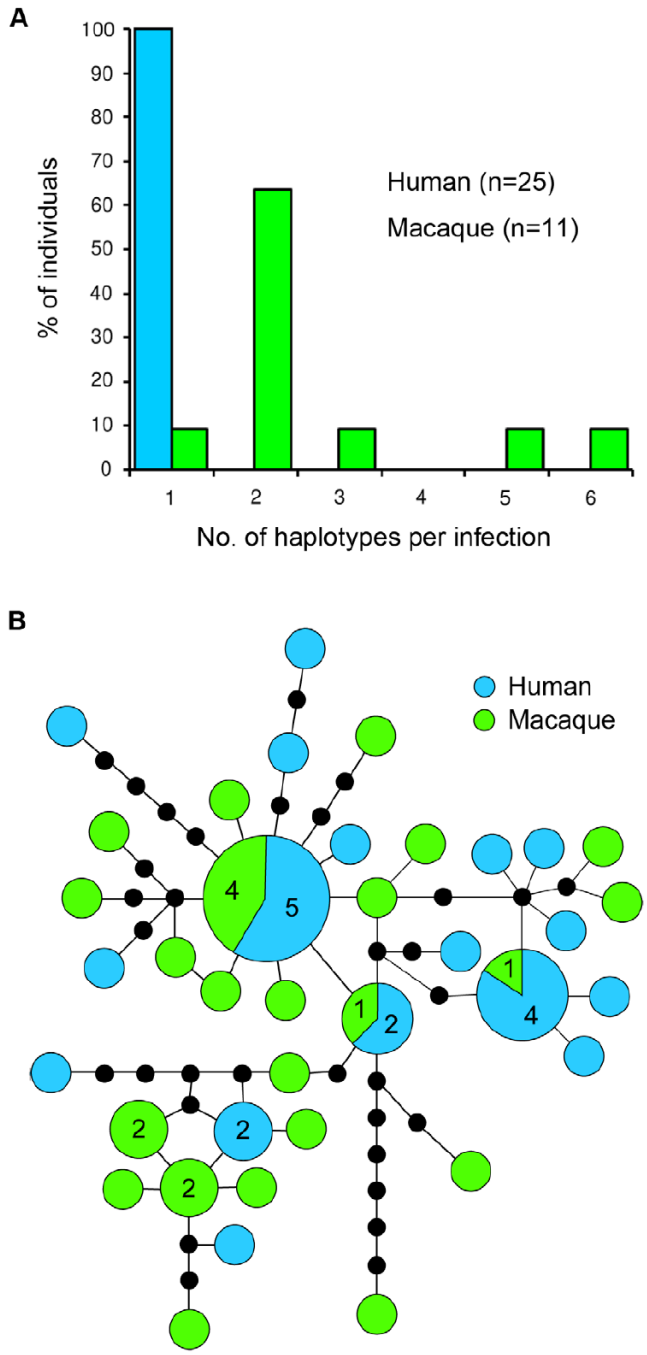

A
100
90
80
70
60
50
40
30
20
10
0
% of individuals
Human (n=25)
Macaque (n=11)
1
2
3
4
5
6
No. of haplotypes per infection
B
Human
Macaque
4
5
1
2
1
4
2
2
2

Figure 4.

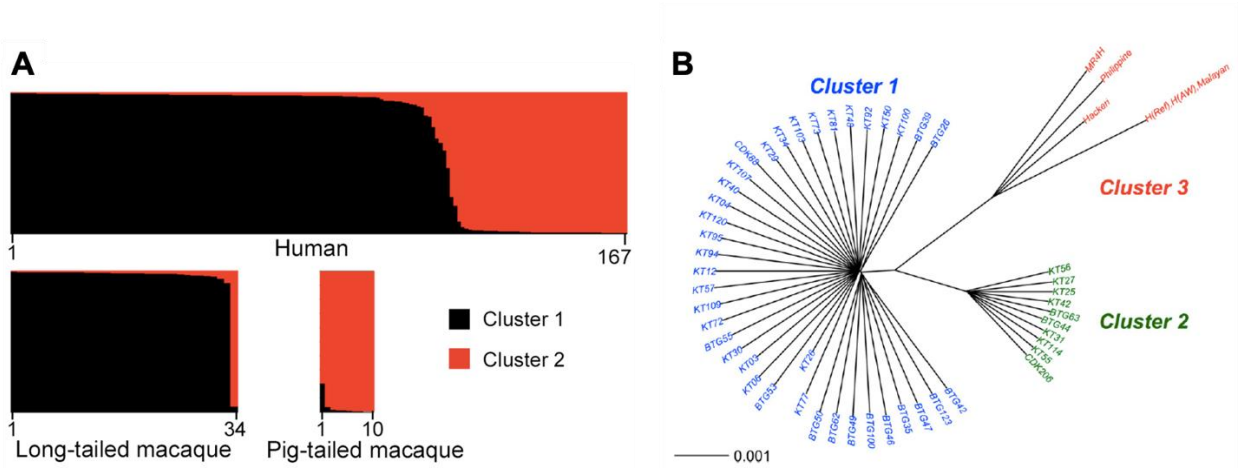

A
1
Human
167
1
34
Long-tailed macaque
1
10
Pig-tailed macaque
Cluster 1
Cluster 2
B
Cluster 1
Cluster 3
Cluster 2
0.001

Figure 5.

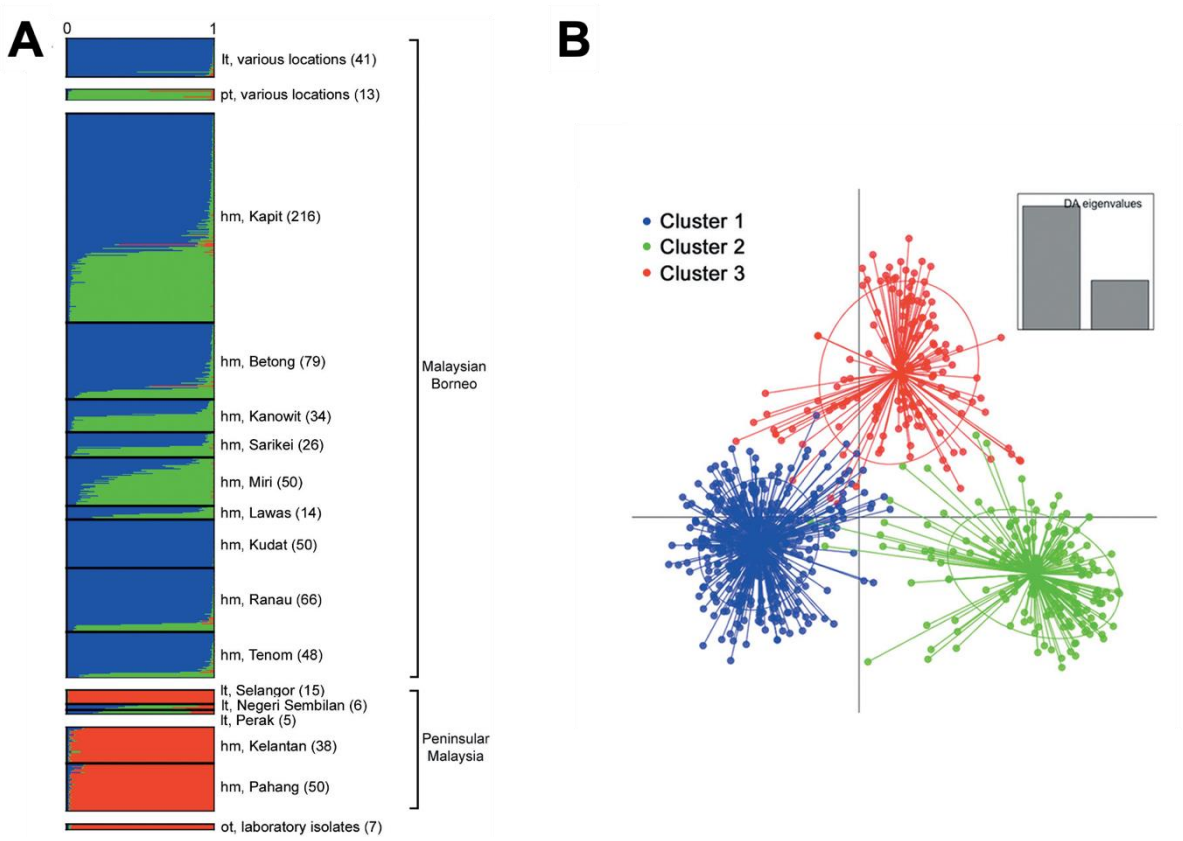

A
0
1
lt, various locations (41)
pt, various locations (13)
hm, Kapit (216)
hm, Betong (79)
hm, Kanowit (34)
hm, Sarikei (26)
hm, Miri (50)
hm, Lawas (14)
hm, Kudat (50)
hm, Ranau (66)
hm, Tenom (48)
Malaysian Borneo
lt, Selangor (15)
lt, Negeri Sembilan (6)
lt, Perak (5)
hm, Kelantan (38)
hm, Pahang (50)
Peninsular Malaysia
ot, laboratory isolates (7)
B
Cluster 1
Cluster 2
Cluster 3
DA eigenvalues

Figure 6.

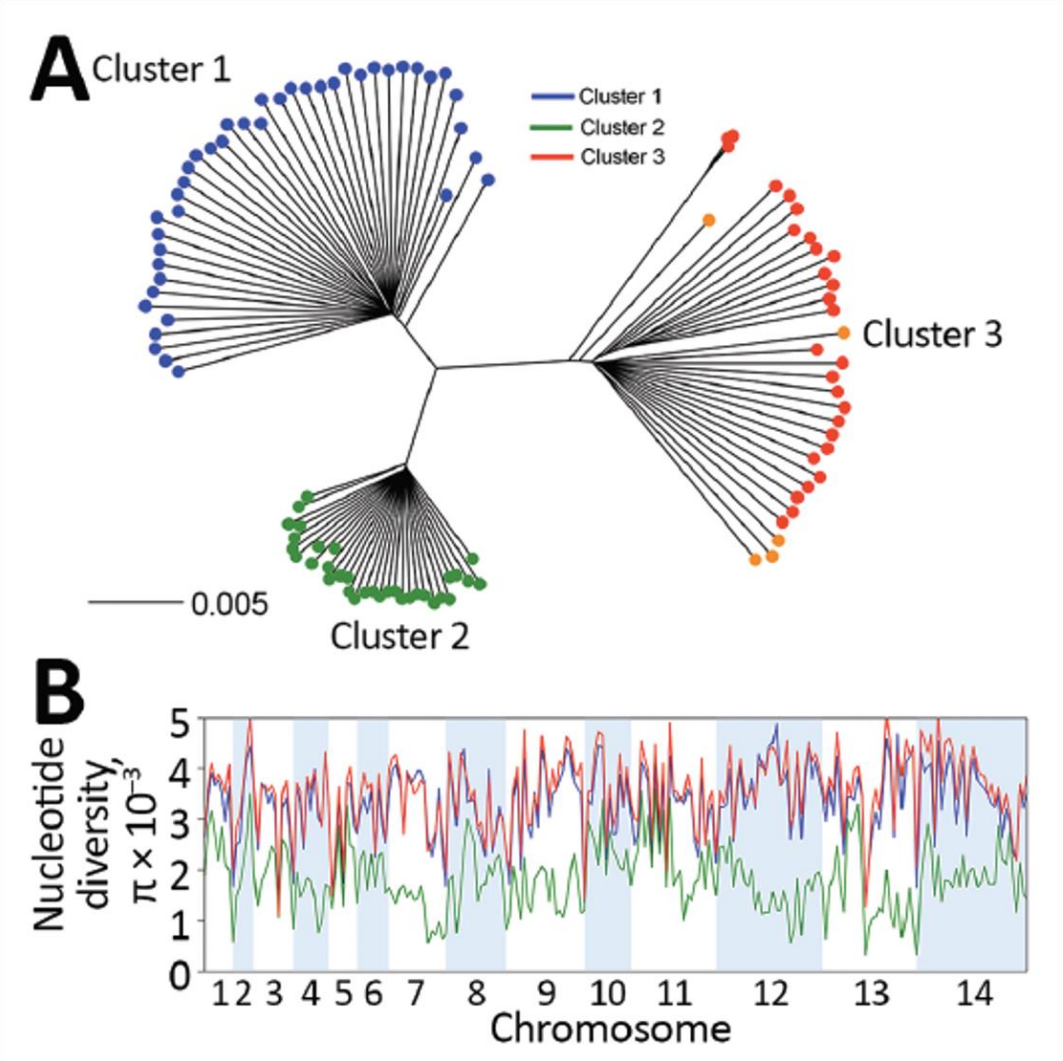

A
Cluster 1
Cluster 1
Cluster 2
Cluster 3
Cluster 3
0.005
Cluster 2
B
Nucleotide diversity, π × 10⁻³
0
1
2
3
4
5
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Chromosome

Figure 7.

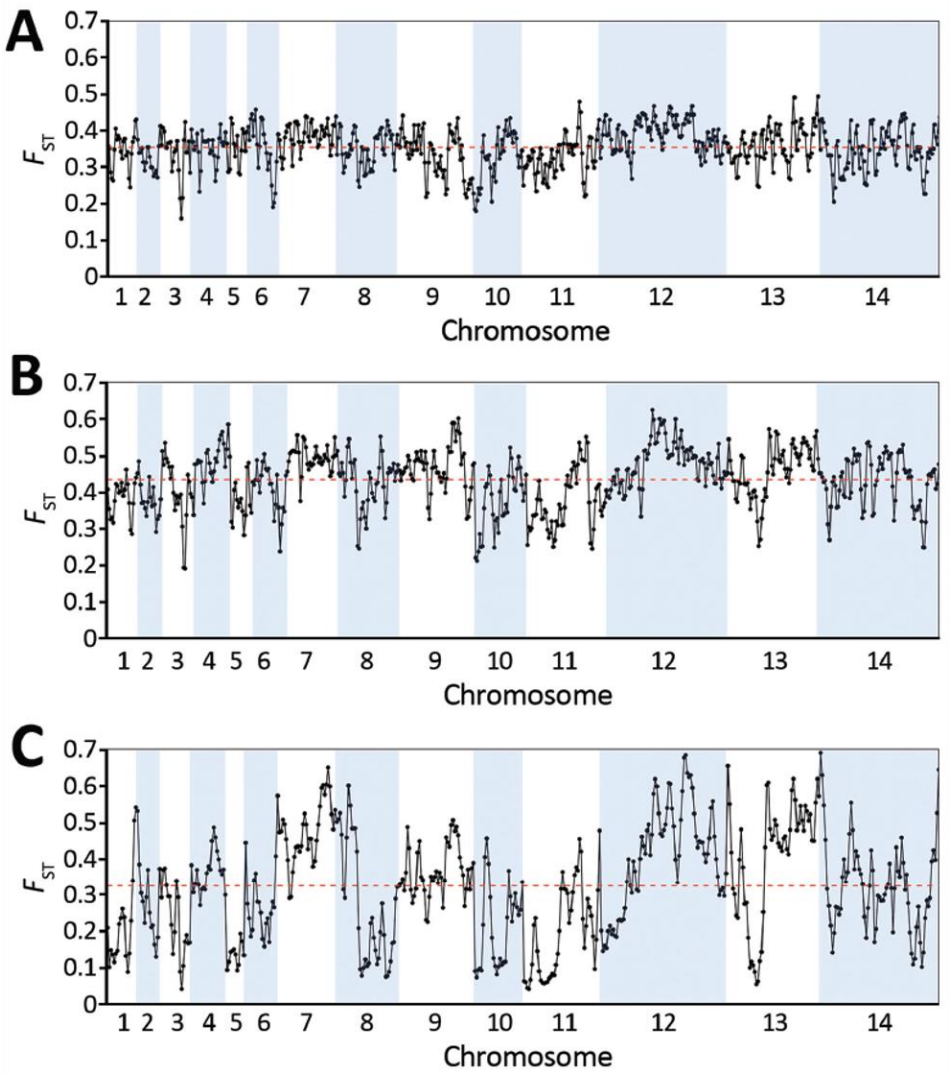

A
B
C
F_{ST}
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Chromosome

Figure 8.

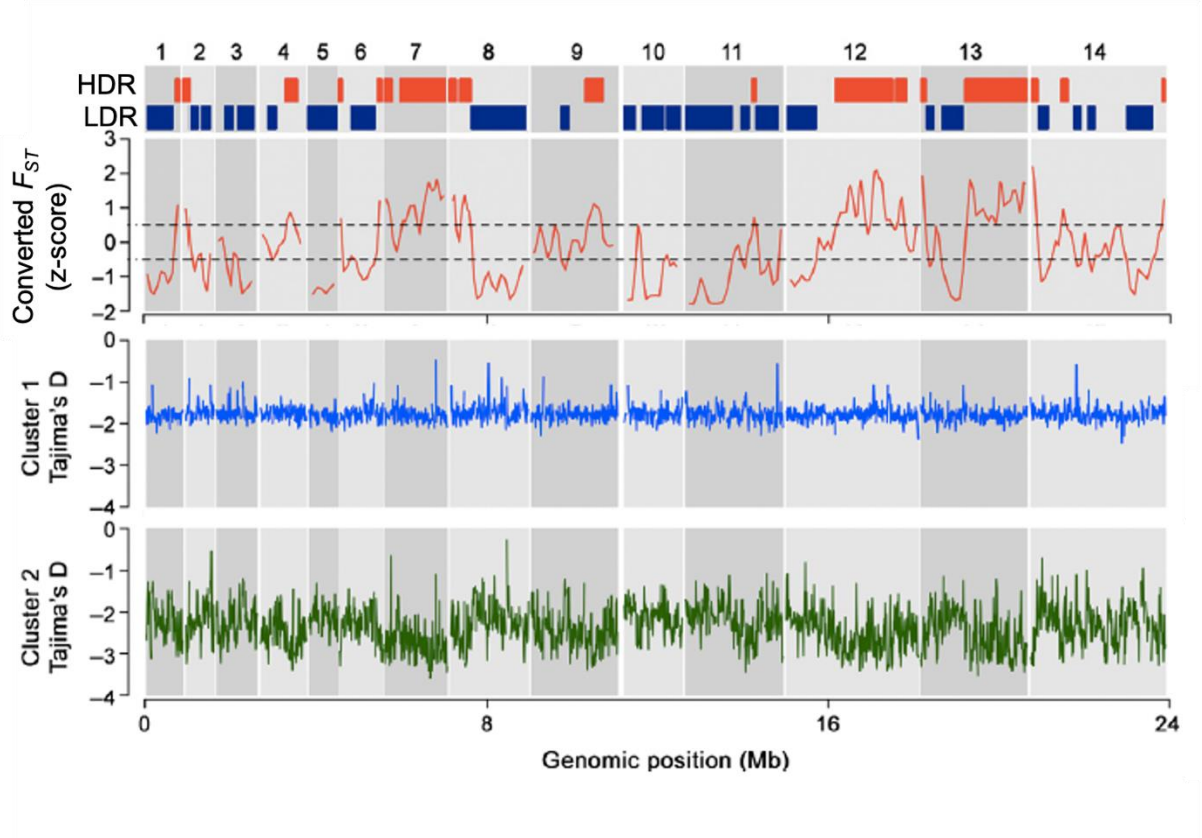

1
2
3
4
5
6
7
8
9
10
11
12
13
14
HDR
LDR
Converted F_{ST} (z-score)
Cluster 1 Tajima's D
Cluster 2 Tajima's D
0
8
16
24
Genomic position (Mb)

Figure 9.

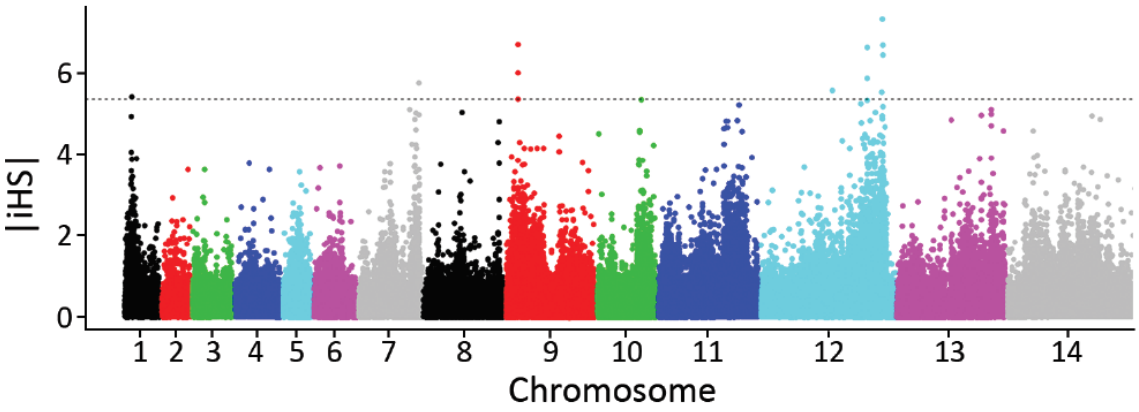


Figure 10.

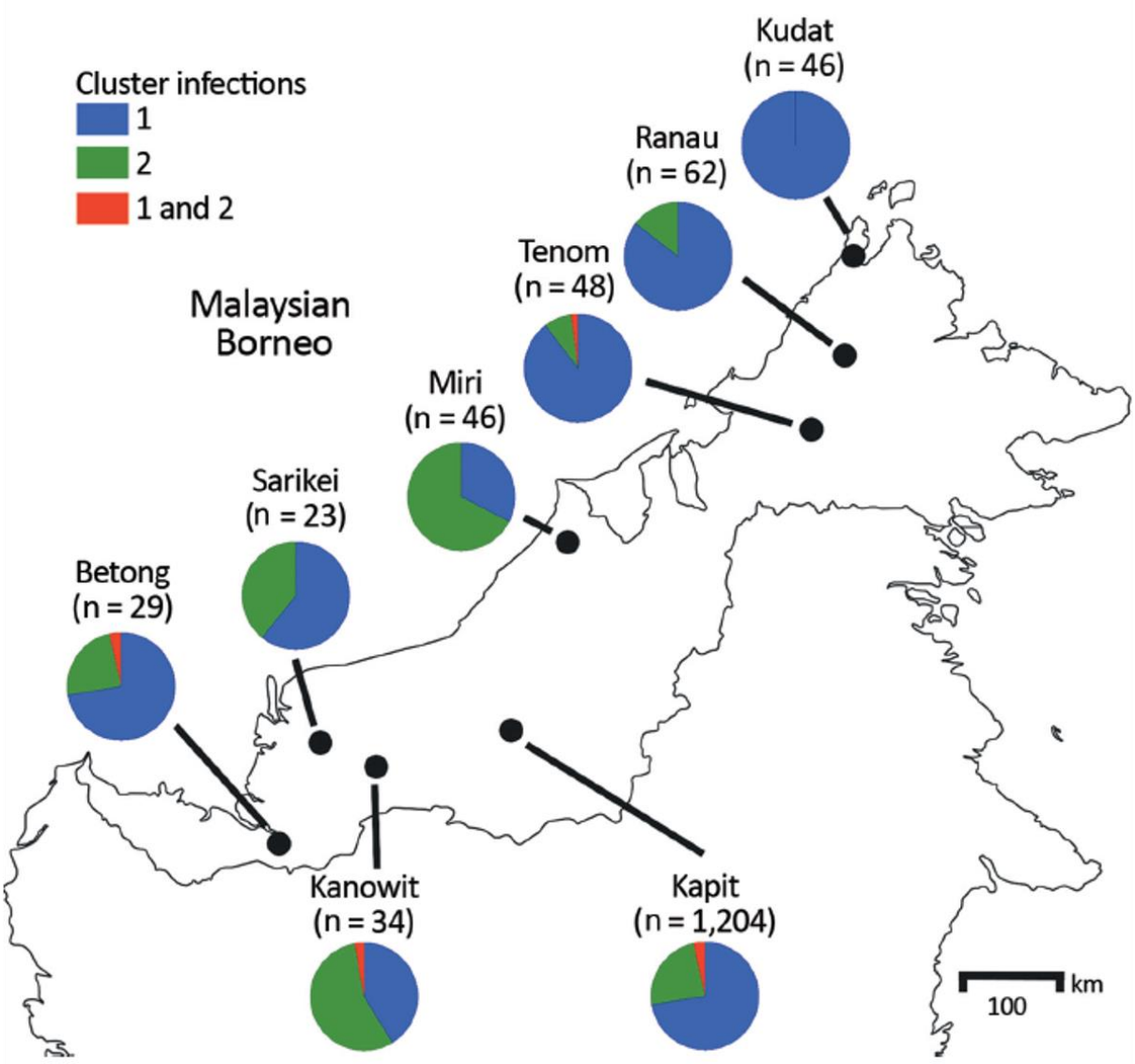

Cluster infections
1
2
1 and 2
Malaysian Borneo
Kudat (n = 46)
Ranau (n = 62)
Tenom (n = 48)
Miri (n = 46)
Sarikei (n = 23)
Betong (n = 29)
Kanowit (n = 34)
Kapit (n = 1,204)
100
km

Figure 11.

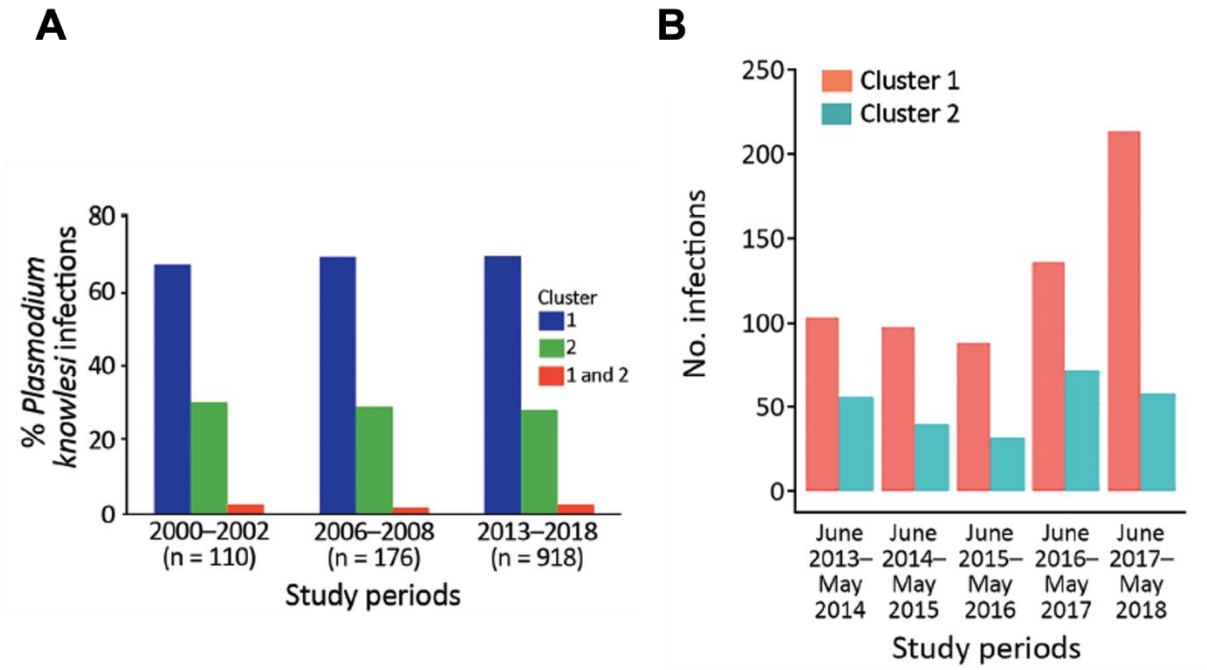

A
% Plasmodium knowlesi infections
80
60
40
20
0
Cluster
1
2
1 and 2
2000–2002
(n = 110)
2006–2008
(n = 176)
2013–2018
(n = 918)
Study periods
B
No. infections
250
200
150
100
50
0
Cluster 1
Cluster 2
June 2013–May 2014
June 2014–May 2015
June 2015–May 2016
June 2016–May 2017
June 2017–May 2018
Study periods

Figure 12.