Three Divergent Subpopulations of the Malaria Parasite *Plasmodium knowlesi*

Paul C.S. Divis, Lee C. Lin, Jeffrine J. Rovie-Ryan, Khamisah A. Kadir, Fread Anderios, Shamilah Hisam, Reuben S.K. Sharma, Balbir Singh, David J. Conway

Multilocus microsatellite genotyping of *Plasmodium knowlesi* isolates previously indicated 2 divergent parasite subpopulations in humans on the island of Borneo, each associated with a different macaque reservoir host species. Geographic divergence was also apparent, and independent sequence data have indicated particularly deep divergence between parasites from mainland Southeast Asia and Borneo. To resolve the overall population structure, multilocus microsatellite genotyping was conducted on a new sample of 182 *P. knowlesi* infections (obtained from 134 humans and 48 wild macaques) from diverse areas of Malaysia, first analyzed separately and then in combination with previous data. All analyses confirmed 2 divergent clusters of human cases in Malaysian Borneo, associated with long-tailed macaques and pig-tailed macaques, and a third cluster in humans and most macaques in peninsular Malaysia. High levels of pairwise divergence between each of these sympatric and allopatric subpopulations have implications for the epidemiology and control of this zoonotic species.

Plasmodium knowlesi is a zoonotic malaria parasite that has only recently been recognized as a notable cause of malaria (*1*). Although cases have now been seen in most countries in Southeast Asia, the largest numbers have been reported in Malaysia (*1–4*). The extent to which this is a result of varying efforts in diagnosis is unclear, as specific molecular identification is required to discriminate *P. knowlesi* from other malaria parasite species. Moreover, although most reports are of cases presenting with clinical symptoms, asymptomatic infections may also occur (*5*).

Author affiliations: Malaria Research Centre, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia (P.C.S. Divis, K.A. Kadir, B. Singh, D.J. Conway); London School of Hygiene and Tropical Medicine, London, United Kingdom (P.C.S. Divis, D.J. Conway); Universiti Putra Malaysia, Serdang, Malaysia (L.C. Lin, R.S.K. Sharma); Department of Wildlife and National Parks Peninsular Malaysia, Kuala Lumpur, Malaysia (J.J. Rovie-Ryan); Sabah State Public Health Laboratory, Kota Kinabalu, Malaysia (F. Anderios); Institute for Medical Research, Kuala Lumpur (S. Hisam)

DOI: http://dx.doi.org/10.3201/eid2304.161738

The *Plasmodium knowlesi* parasite is transmitted by mosquitoes to humans from monkey reservoir hosts, with different *Anopheles* species of the Leucosphyrus group having been incriminated as potential vectors in different areas (*1*,*6*). Two macaque species, the long-tailed macaque (*Macaca fascicularis*) and the pig-tailed macaque (*M. nemestrina*), are the major reservoirs of infection (*7*,*8*). Human infections in Malaysian Borneo, the portion of Malaysia on the island of Borneo, have divergent genetic subpopulations that are seen in the different macaque species locally, indicating that 2 independent zoonoses may be occurring sympatrically (*9*). Noticeable geographic differentiation of parasites between Malaysian Borneo and peninsular Malaysia was also evident in microsatellite analysis; separate studies have revealed divergence between the 2 regions at unlinked genes encoding the normocyte binding protein (*10–12*) and the Duffy binding protein (*13*,*14*), as well as the 18S rRNA and mitochondrial cytochrome oxidase subunit 1 (*15*). Whole-genome sequencing has confirmed the presence of 2 divergent subpopulations of *P. knowlesi* in Malaysian Borneo and revealed a third divergent cluster of laboratory isolates maintained in laboratories since the 1960s; most of these were recorded to have originated from peninsular Malaysia (*16*).

To resolve the population structure in relation to host species and geography, a new collection of 182 *P. knowlesi* infection samples from humans and wild macaques living in diverse areas of Malaysia was genotyped at 10 microsatellite loci. We first analyzed the new dataset separately and then analyzed a combined dataset incorporating previous multilocus microsatellite data, using several independent and complementary statistical approaches to identify genetic substructure. All analyses revealed that 2 divergent genetic subpopulations of human cases occur sympatrically in Malaysian Borneo, detected separately in long-tailed macaques and pig-tailed macaques in the same region, whereas a third divergent genetic subpopulation occurs in humans and most macaques in peninsular Malaysia. This parasite species has undergone different sympatric and allopatric processes of divergence, which will affect its future adaptation to a changing environmental landscape. Current

differences between the subpopulations need to be recognized in clinical and epidemiologic studies.

Materials and Methods

Study Sites and DNA Samples

We obtained blood samples infected with *P. knowlesi* from human clinical cases at 7 sites and from macaque hosts at 8 sites across Malaysia (Figure 1). We extracted DNA from anticoagulated venous blood samples or dried blood spots, and tested the DNA for the presence of different malaria parasite species by species-specific PCR using methods described previously (*7*). Samples from 134 *P. knowlesi*–positive human cases collected during 2012–2014 that had sufficient DNA for multilocus genotyping originated from Kapit (n = 35), Betong (n = 4), and Lawas (n = 15) in Sarawak state, Malaysian Borneo; from Kudat (n = 20), Ranau (n = 25), and Tenom (n = 22) in Sabah state, Malaysian Borneo; and from Kelantan (n = 13) in peninsular Malaysia.

Samples that were collected during 2007–2014 from 48 *P. knowlesi*–positive macaques had sufficient DNA for multilocus genotyping. Most were from long-tailed macaques, sampled from Selangor (n = 17), Perak (n = 6), and Negeri Sembilan (n = 15) in peninsular Malaysia and from Balingian (n = 1), Limbang (n = 2), Miri (n = 2), and Sarikei (n = 2) in Sarawak; pig-tailed macaque samples were from Limbang (n = 1) and Kapit (n = 2) in Sarawak. We performed the sampling according to the protocols of the Department of Wildlife and National Parks in Malaysia. We included DNA of *P. knowlesi* strain Nuri (kindly provided by Clemens Kocken at the Biomedical Primate Research Centre, the Netherlands) in the genotyping as a control (*17*).

Microsatellite Genotyping of New Samples

We genotyped each of the *P. knowlesi*–positive DNA samples at 10 microsatellite loci (NC03_2, CD05_06, CD08_61, NC0AU: 9_1, NC10_1, CD11_157, NC12_2, NC12_4, CD13_61, CD13_107) using hemi-nested PCR assays specific for *P. knowlesi*, as described previously (*9*). We analyzed fluorescent dye-labeled PCR products by using capillary electrophoresis on the Genetic Analyzer 3730 (Applied Biosystems, Cheshire, UK), with GeneScan 500 LIZ internal size standards, following which we scored alleles and peak heights with GeneMapper version 4.0 software (Applied Biosystems).

The genotypic multiplicity of infection (MOI) was defined as the maximum number of alleles detected at any individual locus. Electrophoretic peak heights above 200 fluorescent units of the expected molecular sizes were scored as alleles, and secondary peaks within an infection sample were scored if they had a height of at least 25% relative to the predominant allele. We determined the multilocus genotype profile of each infection, and allele frequency counts for population samples, by counting the predominant allele at each locus within each infection.

Analysis of Microsatellite Genotypes from Previous Data

We retrieved whole genome sequence data of *P. knowlesi* samples from the European Nucleotide Archive (http://www.ebi.ac.uk/ena), and we obtained the reference genome sequence of strain H from GeneDB (http://www.genedb.org/Homepage/Pknowlesi). Most of the parasite genome short-read Illumina sequences available are from patients sampled in Malaysian Borneo (*12*,*16*), but a few are from older laboratory lines that originated from peninsular Malaysia, as well as 1 supposedly from the Philippines (*16*). Although genome sequences indicate some historical mislabeling or contamination of the laboratory lines, meaning that individual identities are in question, it is clear that most are from peninsular Malaysia (*16*). We aligned the raw short reads to *P. knowlesi* genome strain H by using the

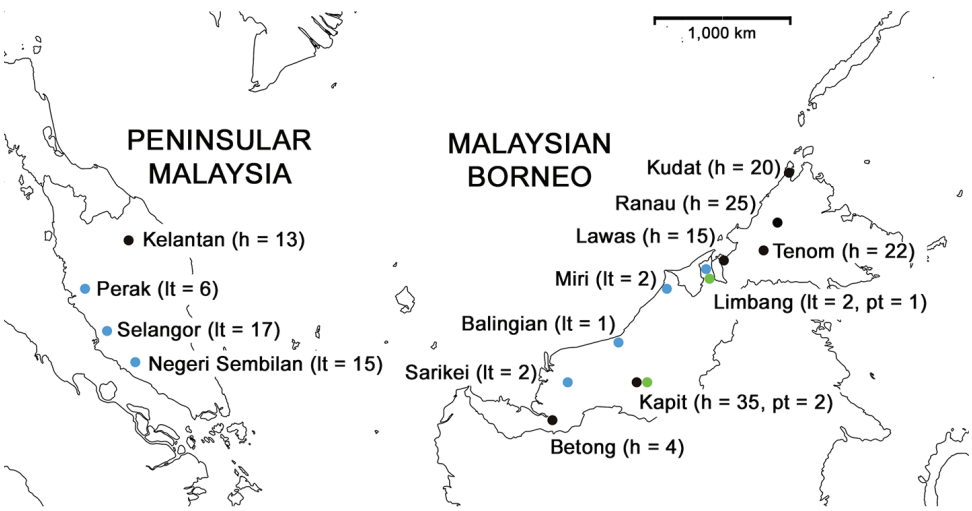


Figure 1. Geographic distribution of DNA samples of *Plasmodium knowlesi* infections derived from 134 humans and 48 macaques across Malaysia. h, human samples; lt, long-tailed macaque samples; pt, pig-tailed macaque samples.

BWA-MEM alignment tool with default parameters (https://arxiv.org/pdf/1303.3997.pdf). We identified lists of indels using the SAMTools and VCFtools software (*18*,*19*) with the following parameters, described elsewhere (*16*): *mpileup –B –Q 23 –d 2000 –C 50 -ugf; varFilter –d 10 –D 2000*. Using ARTEMIS software (*20*), we determined the putative microsatellite allele size by inspecting the indels within the location of the PCR primers used for the second amplification PCR. We assessed the quality of the mapping within the microsatellite allele regions with the minimum depth of short-read coverage at 30-fold.

Analyses of *P. knowlesi* Population Genetic Substructure

We evaluated population genetic structure by Bayesian clustering inference using STRUCTURE version 2.3.4 software (*21*), on samples for which there were no missing data at any locus. First, to allocate the probable ancestral assignment of a genotype into 1 or more *K* clusters, we set the parameters for the admixture model on the basis of correlated allele frequency, without providing the sample source information. However, the sensitivity for population structure analysis can be improved by providing population information, in which an algorithm assumes that the probability of an individual being part of a population varies among locations or sources of origins (*22*). For the second test, we set the parameter to LOCPRIOR. This parameter is informative when population structure signals are weak because of a close relationship between populations. We performed both LOCPRIOR and non-LOCPRIOR parameters in STRUCTURE runs separately with a burn-in period of 50,000 followed by 100,000 Markov chains (MCMC iterations). The simulations were replicated 20 times for *K* values ranging from 1 to 10. The optimal *K* value was calculated based on Evanno's method of ΔK statistics implemented in the STRUCTURE HARVESTER webpage interface (*23*,*24*). For the optimum *K*, we aligned the 20-replicate runs at 10,000 permutations to determine the consensus of cluster scores using CLUMPP version 1.1.2 (*25*).

To evaluate population structure independently, we performed principal coordinate analysis (PCoA) using the GenAlEx package version 6 implemented in Microsoft Excel (*26*). We first generated a genetic distance matrix using the multilocus microsatellite dataset, and we plotted a 2-dimensional PCoA based on the first 2 highest eigenvalues. We calculated the *K*-means clusters using the first and second eigenvectors generated from the PCoA, and subsequently used them to assign each individual infection to the most probable cluster. In addition, we applied the discriminant analysis of principal component from the *adegenet* 2.0.0 packages in R to assess the population structure (*27*). In this procedure, we first transformed genotype data into 40 uncorrelated principal components, and then, using the discriminant function, we partitioned the variances into within-group and among-group components, while optimizing separations between groups.

We calculated pairwise differentiation (F_{ST}) between different subpopulations of *P. knowlesi* by using FSTAT software version 2.9.3.2 (*28*). We estimated the mean allelic diversity across loci, measured as expected heterozygosity (H_E), using FSTAT software. We assessed multilocus linkage disequilibrium with the standardized index of association (I_A^S), calculated by LIAN version 3.7 (*29*), with Monte Carlo simulation of 10,000 data permutations.

Results

Genotypic Diversity within *P. knowlesi* Infections

Of 182 *P. knowlesi* infections genotyped for this study (134 from humans, 45 from long-tailed macaques, and 3 from pig-tailed macaques), 166 (91.2%) yielded complete genotype data for the panel of 10 microsatellite loci, whereas the remainder were each genotyped for at least 7 of the loci (Table; online Technical Appendix 1, http://wwwnc.cdc.gov/EID/article/23/4/16-1738-Techapp1.xlsx).

Among the human cases, single genotype infections were common, and the average number of genotypes per infection (MOI) was less than 2 at all sites sampled. This was expected when these samples were collectively reanalyzed with *P. knowlesi* infections of humans and macaques across Malaysia from previous studies (online Technical Appendix 2 Table 1, http://wwwnc.cdc.gov/EID/article/23/4/16-1738-Techapp2.pdf). We found no notable difference in numbers of genotypes per infection in Malaysian Borneo and peninsular Malaysia (mean MOI values of 1.50 and 1.77, respectively; p = 0.14 by Fisher exact test). In contrast, multiple genotype infections were more common in macaques both in Malaysian Borneo (mean MOI = 2.10, $p = 6.7 \times 10^{-3}$) and peninsular Malaysia (mean MOI = 2.39, $p = 9.8 \times 10^{-4}$) (Table; Figure 2). We counted the predominant allele at each locus per infection for subsequent statistical analyses on population structure.

Analysis of *P. knowlesi* Population Genetic Structure with New Samples

Bayesian clustering analyses using 2 admixture models on the new sample of 166 infections with complete genotype data for the full panel of 10 microsatellite loci identified 3 subpopulation clusters (*K* = 3; Figure 3; online Technical Appendix 2 Figures 1 and 2, panel A), hereafter referred to as clusters 1–3. Human infections in Malaysian Borneo were assigned to clusters 1 and 2, whereas long-tailed macaque infections were all in cluster 1 and pig-tailed macaque infections were mostly in cluster 2 (1 pig-tailed macaque infection was assigned as intermediate between clusters 2 and 3), confirming the existence of 2 major

Table. Summary of *Plasmodium knowlesi* mixed-genotype infections in 134 human and 48 macaque hosts across Malaysia obtained using 10 microsatellite loci*

Host and site	Region	No. samples	No. isolates by no. genotypes detected				% Poly	Average MOI	MS_{10}
			1	2	3	4			
Human									
Kapit	Sarawak	35	27	5	2	1	23	1.34	35
Betong	Sarawak	4	4	0	0	0	0	1.00	3
Lawas	Sarawak	15	7	7	0	1	53	1.67	14
Kudat	Sabah	20	13	6	1	0	35	1.40	20
Ranau	Sabah	25	13	10	2	0	48	1.56	25
Tenom	Sabah	22	11	7	3	1	50	1.73	22
Kelantan	Peninsular Malaysia	13	5	6	2	0	62	1.77	13
Total		134							132
Long-tailed macaque									
Balingian	Sarawak	1	0	1	0	0	100	2.00	1
Limbang	Sarawak	2	0	1	1	0	100	2.50	1
Miri	Sarawak	2	1	1	0	0	50	1.50	1
Sarikei	Sarawak	2	1	0	1	0	50	2.00	2
Selangor	Peninsular Malaysia	17	8	6	2	1	53	1.76	15
Perak	Peninsular Malaysia	6	1	3	2	0	83	2.17	5
Negeri Sembilan	Peninsular Malaysia	15	0	3	6	6	100	3.20	6
Total		45							31
Pig-tailed macaque									
Limbang	Sarawak	1	0	0	1	0	100	3.00	1
Kapit	Sarawak	2	1	0	1	0	50	2.00	2
Total		3							3
All		182							166

*All samples were successfully genotyped at ≥7 loci, and 166 samples had complete genotypes for all 10 microsatellite loci (MS_{10}). MOI, multiplicity of infection; poly, polyclonal infections.

sympatric subpopulations in Malaysian Borneo, as reported previously (*9*,*12*,*16*).

Among the samples from peninsular Malaysia, those from human cases were all assigned to cluster 3, along with most of the infections from wild long-tailed macaques sampled in Kelantan, although long-tailed macaque infections from the other 2 sites had more intermediate cluster assignments, suggesting some ancestral affinity with cluster 2. All the laboratory isolates, originating many years ago mainly from peninsular Malaysia, were clearly assigned to cluster 3, consistent with results of a recent whole genome sequence analysis (*16*).

Analysis of Population Genetic Structure Incorporating New and Previously Acquired Microsatellite Data

To further evaluate the population structure of *P. knowlesi*, we collated the dataset in this study with data from samples analyzed previously (*9*). This yielded a total of 758 *P. knowlesi* infections with the complete panel of 10 microsatellite loci genotyped. This total comprises 166 samples from the present study (Table), 556 previously genotyped samples, 29 samples that had undergone repeat genotyping for all 10 loci completed here (online Technical Appendix 1 Figure 3), and 7 derived from Illumina short-read sequence data.

The admixture STRUCTURE analysis without the LOCPRIOR model identified 2 subpopulation clusters (K = 2; online Technical Appendix 2 Figure 2, panel B, and 4). This was consistent with a previous analysis showing that human cases in the Malaysian Borneo group fell into 2 different genotype clusters, which are also respectively seen in long-tailed and pig-tailed macaque infections, although the current analysis assigned samples from peninsular Malaysia to cluster 2 (previously, they had been grouped into cluster 1). However, incorporation of the LOCPRIOR model showed 3

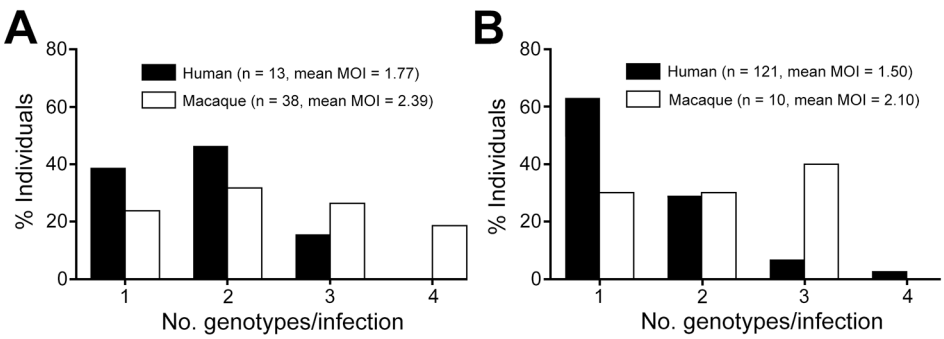


Figure 2. Multiplicity of infection (MOI) for *Plasmodium knowlesi* genotypes in 134 human and 48 macaque hosts across Malaysia. Means of MOI were higher in macaque hosts than in human hosts for both regions, but the values were not statistically significant for A) peninsular Malaysia (p = 0.25 by Fisher exact test) compared with B) Malaysian Borneo (p = 0.01).

subpopulation clusters (K = 3; Figure 4, panel A; online Technical Appendix 2 Figure 4), with most of the isolates from peninsular Malaysia belonging to cluster 3, as also seen with the analysis based solely on the new samples. Overall, this confirms that human *P. knowlesi* infections in Malaysian Borneo are divided into 2 different genetic subpopulations that are associated with different macaque reservoir host species, whereas human infections in peninsular Malaysia belong to a third subpopulation that is also seen in long-tailed macaques at 1 of the sites in peninsular Malaysia.

Robustness and Divergence of Subpopulation Clusters

Using an a priori designation of 3 subpopulation clusters (K = 3), we independently assigned all 758 infections into clusters using PCoA (Figure 4, panel B) and discriminant analysis (Figure 4, panel C), and compared the results with those derived from the STRUCTURE analysis (Figure 4, panel A). These showed highly concordant results (online Technical Appendix 1). PCoA indicated that infections in humans were strongly associated with infections in local macaque reservoir hosts for both Malaysian Borneo and peninsular Malaysia (online Technical Appendix 2 Figure 5). Discriminant analysis also showed clear clustering, with only minimal overlap among the inertia ellipses for the 3 major clusters.

To test the consistency and robustness of cluster assignment for all 758 infections, across the different methods used (Bayesian analysis using STRUCTURE, principal coordinates analysis, and discriminant analysis), we assessed a consensus for each individual (online Technical Appendix 1; online Technical Appendix 2 Table 2). A large majority (86.4%) of infections were assigned into the same cluster by all 3 methods (cluster 1, n = 384; cluster 2, n = 175; cluster 3, n = 96). Most of the remainder (12.9% of the total) had an agreed assignment for 2 of the methods (cluster 1, n = 65; cluster 2, n = 16; cluster 3, n = 17), whereas only 5 (0.7%) showed no agreement across the methods. Omitting the few infections that did not show agreement for 2 or more methods yielded a dataset of 753 *P. knowlesi* infections that grouped into 3 major subpopulation clusters (cluster 1, n = 449; cluster 2, n = 191; cluster 3, n = 113; online Technical Appendix 2 Table 3). We estimated values of allelic diversity (H_E) between 0.51 and 0.83 among different sites at each of the subpopulation clusters (online Technical Appendix 2 Table 4), and we observed similar patterns even without separating the infections by subpopulation cluster assignments (online Technical Appendix 2 Table 5). The index of multilocus linkage disequilibrium yielded various degrees (I_A^S range from -0.007 to 0.305), with loss of significance at majority of the sites among the 3-subpopulation clusters (online Technical Appendix 2 Table 4). However, the degree of significance increased when all infections were not assigned into subpopulation clusters (I_A^S range from -0.002 to 0.242 with p<0.01 at most sites; online Technical Appendix 2 Table 5). Analyses of allele frequencies across all 10 microsatellite loci confirmed strong genetic differentiation among these clusters (F_{ST} = 0.184 between clusters 1 and 2; F_{ST} = 0.152 between clusters 1 and 3;

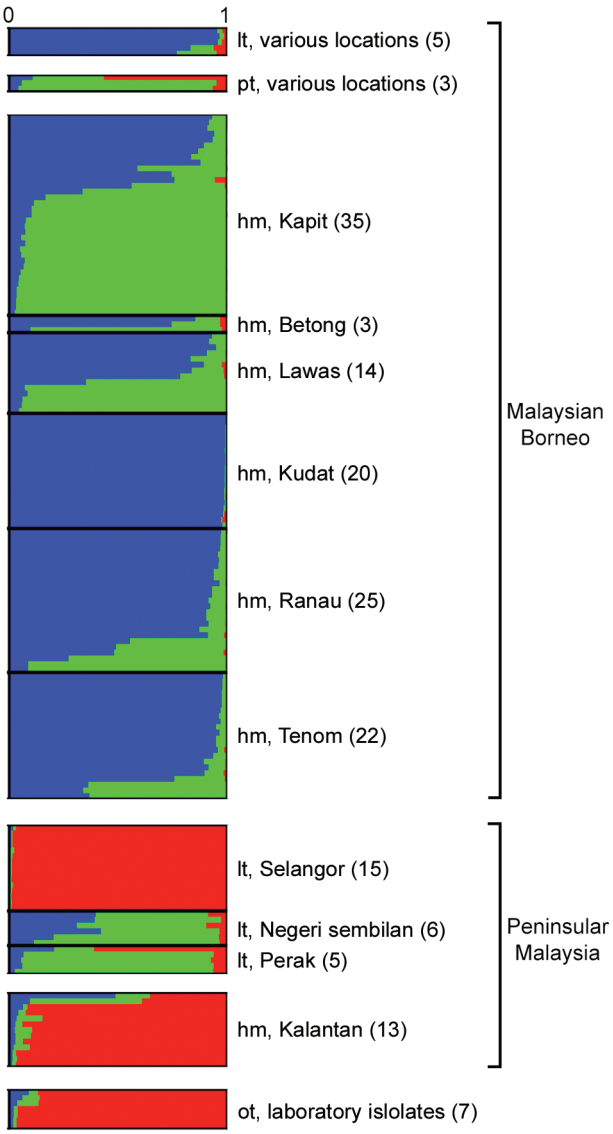


Figure 3. Subpopulation cluster assignments of individual *Plasmodium knowlesi* infections in human and macaque hosts across Malaysia and 7 laboratory isolates. The Bayesian-based STRUCTURE analysis with LOCPRIOR model (*22*) was applied on complete 10-microsatellite loci of 166 *P. knowlesi* infections and 7 laboratory isolates showing 3 subpopulation clusters (K = 3; ΔK = 37.72). Ancestral population clusters are referred to as cluster 1 (blue), cluster 2 (green), and cluster 3 (red). Numbers in parentheses indicate number of isolates. hm, human; lt, long-tailed macaque; pt, pig-tailed macaque; ot, various other sources.

F_{ST} = 0.201 between clusters 2 and 3; p<3.3 × 10^{-4} for each comparison using 3,000 randomized permutations). This indicates deep divergence among the 3 major parasite subpopulations that infect humans, 2 of which are sympatric and predominantly associated with different reservoir hosts (long-tailed and pig-tailed macaques in Malaysian Borneo), and 1 of which is allopatric in a different geographic region (peninsular Malaysia).

Discussion

Three major subpopulations of *P. knowlesi* have been demonstrated in natural human infections in Malaysia. These subpopulations show profound divergence, with pairwise F_{ST} values of ≈0.2, suggesting minimal or no current gene flow between parasites in Malaysian Borneo and peninsular Malaysia, nor between parasites in long-tailed and pig-tailed macaque hosts within Malaysian Borneo.

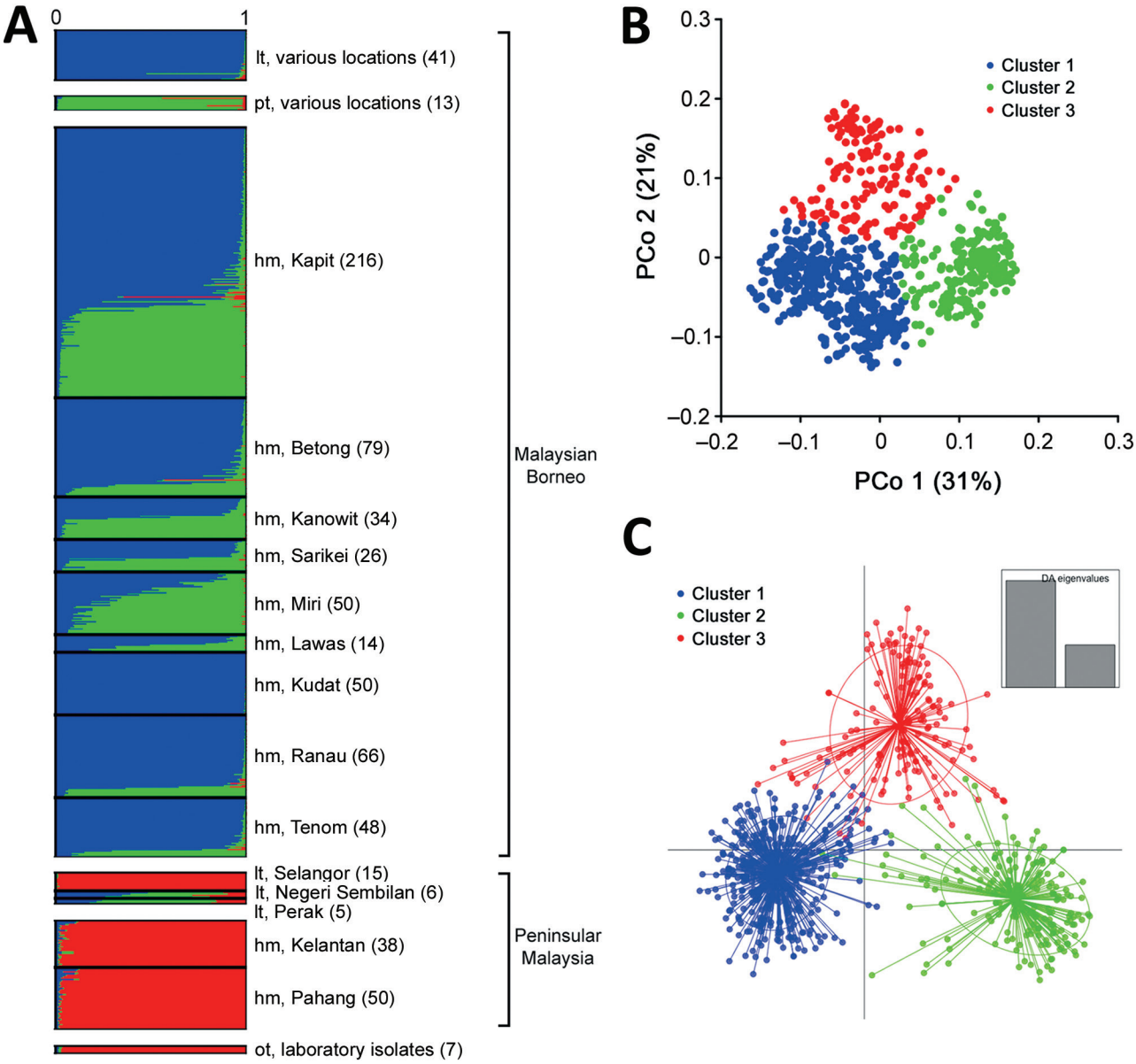


Figure 4. Population genetic structure of combined 751 P. knowlesi infections across Malaysia and 7 laboratory isolates. A) The inference of genetic clusters on complete 10-locus genotype dataset using the STRUCTURE analysis with LOCPRIOR model (22) showed 3 major subpopulation structures (K = 3, ΔK = 98.73), corresponding to those shown in Figure 3. Numbers in parentheses indicate number of isolates. B, C) Using a priori K = 3, individual genotypes were assigned to the most probable subpopulation clusters using independent genetic distance matrix inferred by the principal coordinate analysis (B) and discriminant analysis of principal component (DAPC) (C). In DAPC, clusters depicted as ellipses indicated the variance within the clusters and centered by K-means. hm, human; lt, long-tailed macaque; PCo, principal coordinate; pt, pig-tailed macaque.

The existence of 3 divergent clusters was initially indicated from whole genome sequence-based single nucleotide polymorphism analysis of *P. knowlesi* clinical isolates and laboratory lines (*16*). Whereas 2 of the clusters of genome sequences (clusters 1 and 2) had been seen in clinical infections in Malaysian Borneo, the third (cluster 3) was seen only in old laboratory lines that were originally isolated mostly from peninsular Malaysia. Using microsatellite scoring obtained from genome sequences and combined with genotyping of infections from humans and macaques in the current study, we confirmed that the cluster 3 subpopulation is widespread in peninsular Malaysia. Furthermore, it is divergent from clusters 1 and 2, which account for all infections in Malaysian Borneo and apparently a minority of wild macaque infections in peninsular Malaysia. With smaller numbers of samples, recent studies on sequence diversity in genes encoding the normocyte binding protein (*Pknbpxa*) (*10*) and the Duffy binding protein (*PkDBP*) (*30*), as well as the 18S rRNA gene and the mitochondrial *Cox1* gene, have suggested that parasites in peninsular Malaysia had probably diverged from those in Malaysian Borneo.

It is likely that allopatric divergence occurred as a result of the ocean barrier between Borneo and mainland Southeast Asia, established at the end of the last ice age ≈13,000 years ago, which prevents the movement of wild macaque reservoir hosts (*31*). However, one of the old laboratory lines that was recently sequenced is labeled as having originally been isolated from a long-tailed macaque in "Philippines," and this sequence is clearly assigned to cluster 3 along with the parasites from peninsular Malaysia (*16*), although the islands of the Philippines have never been connected to peninsular Malaysia or any other part of mainland Southeast Asia (*32*). Unless there was a historical mislabeling or previous mixup of parasite material, this finding suggests that wider sampling of *P. knowlesi* in wild macaques will give a more complete understanding of divergence within this zoonotic parasite species (*31*,*33*–*35*). Similarly, the observation that a minority of *P. knowlesi* parasites in long-tailed macaques from peninsular Malaysia are assigned to cluster 2, which has otherwise been seen only in samples from Malaysian Borneo, indicates that additional sampling of macaques from different areas may uncover more features of the parasite population structure.

The sympatric differentiation between cluster 1 and cluster 2 parasites in Malaysian Borneo supports the idea that parasite subpopulations are transmitted independently in long-tailed and pig-tailed macaque populations (*36*,*37*). Although pig-tailed macaques occur mostly in forested areas, long-tailed macaques have a broader habitat range in both forested and nonforested areas (*38*). Because of the absence of parasite samples from pig-tailed macaques in peninsular Malaysia, it is unknown whether there is divergence in *P. knowlesi* between the different macaque host species in this region.

Analysis of genome sequences to derive the frequency distribution of single-nucleotide polymorphism alleles indicates that the cluster 1 subpopulation of *P. knowlesi* has undergone long-term population growth (*16*). It is unknown whether parasites of cluster 2 and cluster 3 subpopulations have a similar demographic history, but genome sequencing of more samples within these subpopulations should be able to address this in the future.

The observation that most infections in all macaque populations are polyclonal, whereas most human cases contain single parasite genotypes, probably reflects a higher intensity of transmission among macaques than from macaques to humans (*9*). It is not yet known whether there are any substantial differences in the clinical course of infections caused by the 3 major subpopulations of *P. knowlesi*; this question should be investigated in a manner that accounts for any confounding variables between different study sites. In any case, recognition of these divergent subpopulations provides a more accurate basis on which to understand and potentially control the transmission of this zoonosis. Furthermore, obtaining whole-genome sequence data from more clinical samples belonging to each of the 3 major types should enable a more thorough investigation of the genomic divergence, and identify loci at which there are signals of recent adaptation that may relate to differences in virulence or transmission.

Acknowledgments

We thank colleagues in Malaria Research Centre, Universiti Malaysia Sarawak, and the London School of Hygiene and Tropical Medicine for laboratory assistance and the Director General of Health in Malaysia for permission to publish this article. We also thank Sarawak Forestry Department, Sarawak Biodiversity Centre, and Department of Wildlife and National Parks Peninsular Malaysia for the permission to collect samples from macaques. We are also grateful to Clemens Kocken for providing *P. knowlesi* DNA control samples.

This study is supported by postgraduate scholarship from the Ministry of Higher Education in Malaysia and grants from UNIMAS (grant nos. 01/(TD03)/1003/2012(01) and F05/SpTDG/1447/2016/4), and from the European Research Council (AdG-2011-294428).

Mr. Divis is a lecturer at the Malaria Research Centre in Universiti Malaysia Sarawak, Malaysia, and a PhD student at the London School of Hygiene and Tropical Medicine, United Kingdom. His research interests include molecular epidemiology, population genetics, and genomics of *Plasmodium knowlesi*.

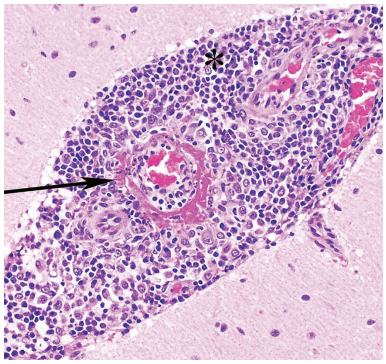
References

1. Singh B, Daneshvar C. Human infections and detection of *Plasmodium knowlesi.* Clin Microbiol Rev. 2013;26:165–84. http://dx.doi.org/10.1128/CMR.00079-12
2. Setiadi W, Sudoyo H, Trimarsanto H, Sihite BA, Saragih RJ, Juliawaty R, et al. A zoonotic human infection with simian malaria, *Plasmodium knowlesi*, in Central Kalimantan, Indonesia. Malar J. 2016;15:218. http://dx.doi.org/10.1186/s12936-016-1272-z
3. Yusof R, Lau YL, Mahmud R, Fong MY, Jelip J, Ngian HU, et al. High proportion of knowlesi malaria in recent malaria cases in Malaysia. Malar J. 2014;13:168. http://dx.doi.org/10.1186/1475-2875-13-168
4. William T, Rahman HA, Jelip J, Ibrahim MY, Menon J, Grigg MJ, et al. Increasing incidence of *Plasmodium knowlesi* malaria following control of *P. falciparum* and *P. vivax* Malaria in Sabah, Malaysia. PLoS Negl Trop Dis. 2013;7:e2026. http://dx.doi.org/10.1371/journal.pntd.0002026
5. Fornace KM, Nuin NA, Betson M, Grigg MJ, William T, Anstey NM, et al. Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia. J Infect Dis. 2016;213:784–7. http://dx.doi.org/10.1093/infdis/jiv475
6. Vythilingam I, Wong ML, Wan-Yussof WS. Current status of *Plasmodium knowlesi* vectors: a public health concern? Parasitology. 2016:1–9. http://dx.doi.org/10.1017/S0031182016000901
7. Lee KS, Divis PC, Zakaria SK, Matusop A, Julin RA, Conway DJ, et al. *Plasmodium knowlesi:* reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog. 2011;7:e1002015. http://dx.doi.org/10.1371/journal.ppat.1002015
8. Vythilingam I, Noorazian YM, Huat TC, Jiram AI, Yusri YM, Azahari AH, et al. *Plasmodium knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. Parasit Vectors. 2008;1:26. http://dx.doi.org/10.1186/1756-3305-1-26
9. Divis PC, Singh B, Anderios F, Hisam S, Matusop A, Kocken CH, et al. Admixture in humans of two divergent *Plasmodium knowlesi* populations associated with different macaque host species. PLoS Pathog. 2015;11:e1004888. http://dx.doi.org/10.1371/journal.ppat.1004888
10. Ahmed MA, Fong MY, Lau YL, Yusof R. Clustering and genetic differentiation of the normocyte binding protein (nbpxa) of *Plasmodium knowlesi* clinical isolates from peninsular Malaysia and Malaysia Borneo. Malar J. 2016;15:241. http://dx.doi.org/10.1186/s12936-016-1294-6
11. Ahmed AM, Pinheiro MM, Divis PC, Siner A, Zainudin R, Wong IT, et al. Disease progression in *Plasmodium knowlesi* malaria is linked to variation in invasion gene family members. PLoS Negl Trop Dis. 2014;8:e3086. http://dx.doi.org/10.1371/journal.pntd.0003086
12. Pinheiro MM, Ahmed MA, Millar SB, Sanderson T, Otto TD, Lu WC, et al. *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. PLoS One. 2015;10:e0121303. http://dx.doi.org/10.1371/journal.pone.0121303
13. Fong MY, Lau YL, Chang PY, Anthony CN. Genetic diversity, haplotypes and allele groups of Duffy binding protein (PkDBPαII) of *Plasmodium knowlesi* clinical isolates from peninsular Malaysia. Parasit Vectors. 2014;7:161. http://dx.doi.org/10.1186/1756-3305-7-161
14. Fong MY, Rashdi SA, Yusof R, Lau YL. Distinct genetic difference between the Duffy binding protein (PkDBPαII) of *Plasmodium knowlesi* clinical isolates from North Borneo and peninsular Malaysia. Malar J. 2015;14:91. http://dx.doi.org/10.1186/s12936-015-0610-x
15. Yusof R, Ahmed MA, Jelip J, Ngian HU, Mustakim S, Hussin HM, et al. Phylogeographic evidence for 2 genetically distinct zoonotic *Plasmodium knowlesi* parasites, Malaysia. Emerg Infect Dis. 2016;22:1371–80. http://dx.doi.org/10.3201/eid2208.151885
16. Assefa S, Lim C, Preston MD, Duffy CW, Nair MB, Adroub SA, et al. Population genomic structure and adaptation in the zoonotic malaria parasite *Plasmodium knowlesi*. Proc Natl Acad Sci U S A. 2015;112:13027–32.
17. Kocken CH, Ozwara H, van der Wel A, Beetsma AL, Mwenda JM, Thomas AW. *Plasmodium knowlesi* provides a rapid in vitro and in vivo transfection system that enables double-crossover gene knockout studies. Infect Immun. 2002;70:655–60. http://dx.doi.org/10.1128/IAI.70.2.655-660.2002
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. http://dx.doi.org/10.1093/bioinformatics/btp352
19. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8. http://dx.doi.org/10.1093/bioinformatics/btr330
20. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16:944–5. http://dx.doi.org/10.1093/bioinformatics/16.10.944
21. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–59.
22. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour. 2009;9:1322–32. http://dx.doi.org/10.1111/j.1755-0998.2009.02591.x
23. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61. http://dx.doi.org/10.1007/s12686-011-9548-7
24. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–20. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x
25. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23:1801–6. http://dx.doi.org/10.1093/bioinformatics/btm233
26. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Res. 2006;6:288–95. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x
27. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11:94. http://dx.doi.org/10.1186/1471-2156-11-94
28. Goudet J. FSTAT (Version 1.2): A computer program to calculate F-statistics. J Hered. 1995;86:485–6.
29. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics. 2000;16:847–9. http://dx.doi.org/10.1093/bioinformatics/16.9.847
30. Putaporntip C, Kuamsab N, Jongwutiwes S. Sequence diversity and positive selection at the Duffy-binding protein genes of *Plasmodium knowlesi* and *P. cynomolgi*: analysis of the complete coding sequences of Thai isolates. Infect Genet Evol. 2016;44:367–75. http://dx.doi.org/10.1016/j.meegid.2016.07.040
31. Liedigk R, Kolleck J, Böker KO, Meijaard E, Md-Zain BM, Abdul-Latiff MA, et al. Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). BMC Genomics. 2015;16:222. http://dx.doi.org/10.1186/s12864-015-1437-0
32. Voris HK. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr. 2000; 27;1153–67. http://dx.doi.org/10.1046/j.1365-2699. 2000.00489.x

33. Esselstyn JA, Widmann P, Heaney LR. The mammals of Palawan Island, Philippines. Proc Biol Soc Wash. 2004; 117:271–302.
34. Meijaard E. Mammals of south-east Asian islands and their Late Pleistocene environments. J Biogeogr. 2003;30:1245–57. http://dx.doi.org/10.1046/j.1365-2699.2003.00890.x
35. Smith DG, Ng J, George D, Trask JS, Houghton P, Singh B, et al. A genetic comparison of two alleged subspecies of Philippine cynomolgus macaques. Am J Phys Anthropol. 2014;155:136–48. http://dx.doi.org/10.1002/ajpa.22564
36. Muehlenbein MP, Pacheco MA, Taylor JE, Prall SP, Ambu L, Nathan S, et al. Accelerated diversification of nonhuman primate malarias in Southeast Asia: adaptive radiation or geographic speciation? Mol Biol Evol. 2015;32:422–39. http://dx.doi.org/10.1093/molbev/msu310
37. Ziegler T, Abegg C, Meijaard E, Perwitasari-Farajallah D, Walter L, Hodges JK, et al. Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the *M. silenus* group. Mol Phylogenet Evol. 2007;42:807–16. http://dx.doi.org/10.1016/j.ympev.2006.11.015
38. Moyes CL, Shearer FM, Huang Z, Wiebe A, Gibson HS, Nijman V, et al. Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas. Parasit Vectors. 2016;9:242. http://dx.doi.org/10.1186/s13071-016-1527-0

Address for correspondence: Paul C.S. Divis, Malaria Research Centre, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia; email: pcsimon@unimas.my

December 2015: **Zoonotic Infections**



- Identifying and Reducing Remaining Stocks of Rinderpest Virus
- Opportunistic Pulmonary *Bordetella hinzii* Infection after Avian Exposure
- Zoonotic Leprosy in the Southeastern United States
- Infection Risk for Persons Exposed to Highly Pathogenic Avian Influenza A H5 Virus–Infected Birds, United States, December 2014–March 2015
- High Prevalence of Intermediate *Leptospira* spp. DNA in Febrile Humans From Urban and Rural Ecuador
- Biological Warfare Plan in the 17th Century—the Siege of Candia, 1648–1669
- Influenza A(H6N1) Virus in Dogs, Taiwan
- Methicillin-Resistant *Staphylococcus aureus* Prevalence among Captive Chimpanzees, Texas, USA, 2012
- Novel *Waddlia* Intracellular Bacterium in Artibeus intermedius Fruit Bats, Mexico
- Tembusu-Related Flavivirus in Ducks, Thailand

- Japanese Macaques (*Macaca fuscata*) as Natural Reservoir of *Bartonella quintana*
- *Onchocerca lupi* Nematode in a Cat, Europe

- Increased Number of Human Cases of Influenza Virus A(H5N1) Infection, Egypt, 2014–15
- Replication Capacity of Avian Influenza A(H9N2) Virus in Pet Birds, Chickens, and Mammals, Bangladesh
- Hendra Virus Infection in Dog, Australia, 2013
- No Evidence of Gouléako and Herbert Virus Infections in Pigs, Côte d'Ivoire and Ghana
- Aquatic Bird Bornavirus 1 in Wild Geese, Denmark
- Vectorborne Transmission of *Leishmania infantum* from Hounds, United States
- Porcine Deltacoronavirus in Mainland China

EMERGING INFECTIOUS DISEASES®

http://wwwnc.cdc.gov/eid/articles/issue/21/12/table-of-contents

Article DOI: http://dx.doi.org/10.3201/eid2304.161738

Three Divergent Subpopulations of the Malaria Parasite *Plasmodium knowlesi*

Technical Appendix 2

Technical Appendix 2 Table 1. Summary of *P. knowlesi* mixed genotype infections in 683 humans and 94 macaques across Malaysia obtained using 10 microsatellite loci*

			No. isolates with the following no. genotypes detected						
Host and Site	Region	N	1	2	3	4	5	%poly	MOI
Human									
Kapit	Sarawak	220	149	62	5	4	0	32	1.38
Betong	Sarawak	81	57	22	2	0	0	30	1.32
Kanowit	Sarawak	34	27	7	0	0	0	21	1.21
Sarikei	Sarawak	26	21	5	0	0	0	19	1.19
Miri	Sarawak	50	43	7	0	0	0	14	1.14
Lawas	Sarawak	15	7	7	0	1	0	53	1.67
Kudat	Sabah	50	32	15	3	0	0	36	1.42
Ranau	Sabah	66	31	31	4	0	0	53	1.59
Tenom	Sabah	48	27	17	3	1	0	44	1.54
Kelantan	Peninsular Malaysia	43	24	17	2	0	0	44	1.49
Pahang	Peninsular Malaysia	50	31	15	3	1	0	38	1.48
Long-tailed macaque									
Kapit	Sarawak	36	4	12	11	7	2	89	2.75
Balingian	Sarawak	1	0	1	0	0	0	100	2.00
Limbang	Sarawak	2	0	1	1	0	0	100	2.50
Miri	Sarawak	2	1	1	0	0	0	50	1.50
Sarikei	Sarawak	2	1	0	1	0	0	50	2.00
Selangor	Peninsular Malaysia	17	8	6	2	1	0	53	1.76
Perak	Peninsular Malaysia	6	1	3	2	0	0	83	2.17
Negeri Sembilan	Peninsular Malaysia	15	0	3	6	6	0	100	3.20
Pig-tailed macaque									
Kapit	Sarawak	12	1	6	3	1	1	92	2.58
Limbang	Sarawak	1	0	0	1	0	0	100	3.00

*All new and old samples (N) were genotyped at ≥7 loci. The percentage of polyclonal infections (%poly) and average genotypic multiplicity of infections (MOI) are shown.

Technical Appendix 2 Table 2. Assignment of combined 753 *P. knowlesi* genotypes into 3 subpopulation clusters determined by a minimum of 2 out of 3 assignment methods*

P. knowlesi population	Subpopulation cluster	DAPC ∧ PCoA ∧ LOC	DAPC ∧ PCoA	DAPC ∧ LOC	PCoA ∧ LOC	Total isolate
LT-various locations in Sarawak	Cluster 1	32	0	4	3	39
	Cluster 2	1	0	0	0	1
	Cluster 3	0	0	0	0	0
PT-various locations in Sarawak	Cluster 1	0	0	0	0	0
	Cluster 2	12	0	0	1	13
	Cluster 3	0	0	0	0	0
Hm-Kapit	Cluster 1	114	0	20	4	138
	Cluster 2	70	1	1	2	74
	Cluster 3	0	0	1	0	1
Hm-Betong	Cluster 1	57	0	10	0	67
	Cluster 2	10	2	0	0	12
	Cluster 3	0	0	0	0	0
Hm-Kanowit	Cluster 1	13	0	3	0	16
	Cluster 2	18	0	0	0	18
	Cluster 3	0	0	0	0	0
Hm-Sarikei	Cluster 1	14	0	0	1	15
	Cluster 2	11	0	0	0	11
	Cluster 3	0	0	0	0	0
Hm-Miri	Cluster 1	16	1	0	0	17
	Cluster 2	28	0	5	0	33
	Cluster 3	0	0	0	0	0
Hm-Lawas	Cluster 1	8	0	0	0	8
	Cluster 2	6	0	0	0	6
	Cluster 3	0	0	0	0	0
Hm-Kudat	Cluster 1	44	0	6	0	50
	Cluster 2	0	0	0	0	0
	Cluster 3	0	0	0	0	0
Hm-Ranau	Cluster 1	49	0	6	2	57
	Cluster 2	8	1	0	0	9
	Cluster 3	0	0	0	0	0
Hm-Tenom	Cluster 1	37	0	5	0	42
	Cluster 2	5	0	1	0	6
	Cluster 3	0	0	0	0	0
LT-Selangor	Cluster 1	0	0	0	0	0
	Cluster 2	0	0	0	0	0
	Cluster 3	14	0	1	0	15
LT-Negeri Sembilan	Cluster 1	0	0	0	0	0
	Cluster 2	2	0	1	1	4
	Cluster 3	0	1	0	0	1
LT-Perak	Cluster 1	0	0	0	0	0
	Cluster 2	4	0	0	0	4
	Cluster 3	1	0	0	0	1
Hm-Kelantan	Cluster 1	0	0	0	0	0
	Cluster 2	0	0	0	0	0
	Cluster 3	33	0	5	0	38
Hm-Pahang	Cluster 1	0	0	0	0	0
	Cluster 2	0	0	0	0	0
	Cluster 3	43	0	5	2	50
Laboratory isolates	Cluster 1	0	0	0	0	0
	Cluster 2	0	0	0	0	0
	Cluster 3	5	0	2	0	7

*Five genotypes from long-tailed macaques in Sarawak (n = 1), humans in Kapit (n = 3), and long-tailed macaques in Negeri Sembilan (n = 1) showed inconsistency in cluster assignment methods are not shown in this table. DAPC, discriminant analysis of principal component; Hm – human; LOC, STRUCTURE analysis with LOCPRIOR model; LT, long-tailed macaque; PCoA, principal coordinate analysis based on genetic distance matrix; PT, pig-tailed macaque.

Technical Appendix 2 Table 3. Summary of subpopulation cluster assignment on combined 758 *P. knowlesi* genotypes according to host and geographic origins

Subpopulation cluster	Malaysian Borneo			Peninsular Malaysia		Laboratory isolate	Total isolate
	LT*	PT	Hm	LT	Hm		
Cluster 1	39	0	410	0	0	0	449
Cluster 2	1	13	169	8	0	0	191
Cluster 3	0	0	1	17	88	7	113
Unassigned	1	0	3	1	0	0	5

*Hm, human; LT, long-tailed macaque; PT, pig-tailed macaque.

Technical Appendix 2 Table 4. Multilocus linkage disequilibrium (standardized index of association) and genetic diversity (expected heterozygosity) of *P. knowlesi* infections for each of the 3 subpopulation clusters separately at different geographic sites*

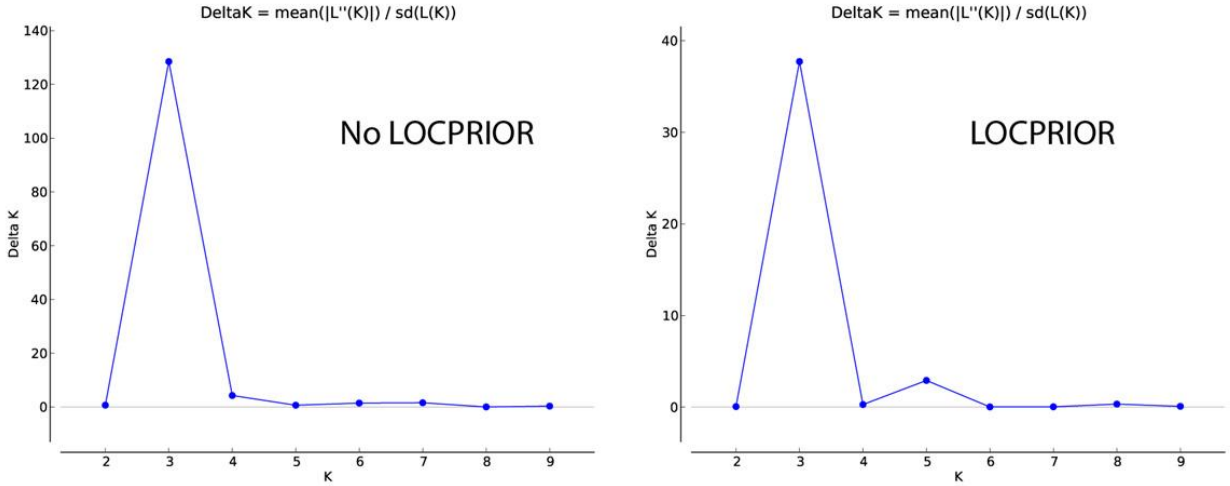
Site	Region	Host	N	I_A^S	p value	H_E (±SE)
Cluster 1						
Kapit	Sarawak	Long-tailed macaque	34	0.026	<0.01	0.70 (0.05)
Kapit	Sarawak	Human	138	0.003	0.19	0.70 (0.06)
Betong	Sarawak	Human	67	0.005	0.25	0.63 (0.08)
Kanowit	Sarawak	Human	16	0.007	0.33	0.67 (0.05)
Sarikei	Sarawak	Human	15	0.003	0.42	0.64 (0.08)
Miri	Sarawak	Human	17	0.038	0.01	0.67 (0.09)
Lawas	Sarawak	Human	8	0.026	0.27	0.71 (0.09)
Kudat	Sabah	Human	50	–0.003	0.65	0.68 (0.07)
Ranau	Sabah	Human	57	0.011	0.02	0.71 (0.06)
Tenom	Sabah	Human	42	0.006	0.22	0.70 (0.06)
Cluster 2						
Kapit	Sarawak	Pig-tailed macaque	12	0.004	0.38	0.58 (0.09)
Kapit	Sarawak	Human	74	0.003	0.27	0.59 (0.10)
Betong	Sarawak	Human	12	0.083	<0.01	0.62 (0.11)
Kanowit	Sarawak	Human	18	–0.009	0.76	0.62 (0.08)
Sarikei	Sarawak	Human	11	0.047	0.04	0.59 (0.10)
Miri	Sarawak	Human	33	0.013	0.10	0.63 (0.08)
Lawas	Sarawak	Human	6	0.076	0.08	0.51 (0.08)
Ranau	Sabah	Human	9	0.099	<0.01	0.64 (0.10)
Tenom	Sabah	Human	6	–0.007	0.66	0.64 (0.08)
Negeri Sembilan	Peninsular Malaysia	Long-tailed macaque	4	0.085	0.26	0.83 (0.06)
Perak	Peninsular Malaysia	Long-tailed macaque	4	0.305	<0.01	0.58 (0.08)
Cluster 3						
Selangor	Peninsular Malaysia	Long-tailed macaque	15	0.044	0.01	0.62 (0.03)
Kelantan	Peninsular Malaysia	Human	38	0.020	0.03	0.71 (0.04)
Pahang	Peninsular Malaysia	Human	50	–0.001	0.56	0.72 (0.04)

All new and old samples (N) were genotyped at complete 10 microsatellite loci. Sites with N<4 were not included in the analyses.
H_E, expected heterozygosity; I_A^S, standardized index of association.

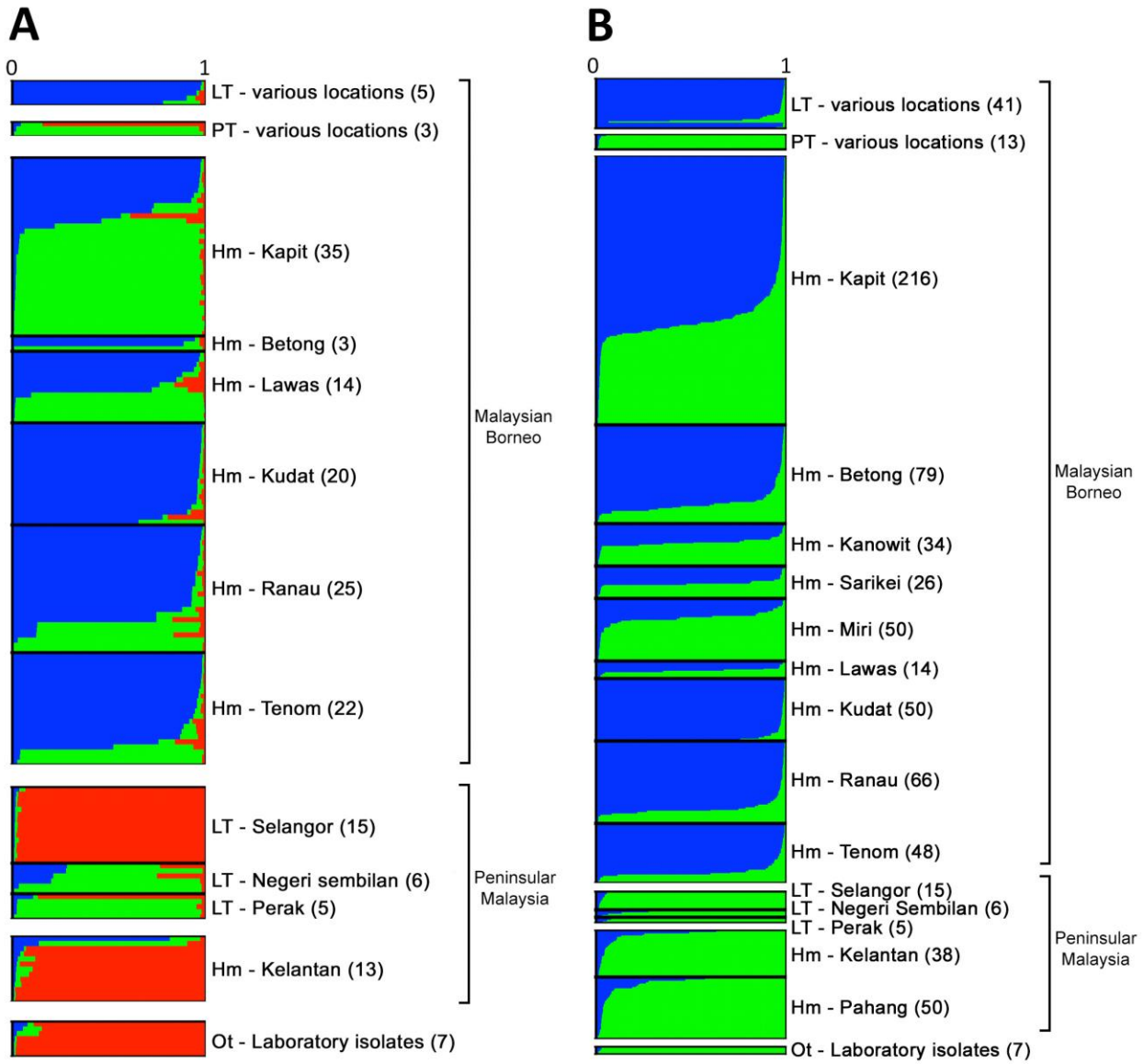
Technical Appendix 2 Table 5. Multilocus linkage disequilibrium (standardized index of association) and genetic diversity (expected heterozygosity) of *P. knowlesi* infections sampled from different host species at each site, without separating by subpopulation cluster assignment*

Site	Region	Host	N	I_A^S	p value	H_E (±SE)
Kapit	Sarawak	Long-tailed macaque	36	0.023	<0.01	0.71 (0.05)
Kapit	Sarawak	Pig-tailed macaque	12	0.004	0.40	0.58 (0.09)
Kapit	Sarawak	Human	216	0.025	<0.01	0.74 (0.04)
Betong	Sarawak	Human	79	0.020	<0.01	0.67 (0.08)
Kanowit	Sarawak	Human	34	0.024	<0.01	0.72 (0.05)
Sarikei	Sarawak	Human	26	0.040	<0.01	0.71 (0.08)
Miri	Sarawak	Human	50	0.031	<0.01	0.60 (0.04)
Lawas	Sarawak	Human	14	0.050	0.01	0.72 (0.05)
Kudat	Sabah	Human	50	–0.003	0.64	0.68 (0.07)
Ranau	Sabah	Human	66	0.026	<0.01	0.75 (0.05)
Tenom	Sabah	Human	48	0.009	0.10	0.73 (0.04)
Selangor	Peninsular Malaysia	Long-tailed macaque	15	0.044	0.01	0.62 (0.08)
Negeri Sembilan	Peninsular Malaysia	Long-tailed macaque	6	0.028	0.24	0.78 (0.06)
Perak	Peninsular Malaysia	Long-tailed macaque	5	0.242	<0.01	0.69 (0.06)
Kelantan	Peninsular Malaysia	Human	38	0.020	0.03	0.71 (0.04)
Pahang	Peninsular Malaysia	Human	50	–0.002	0.56	0.72 (0.04)

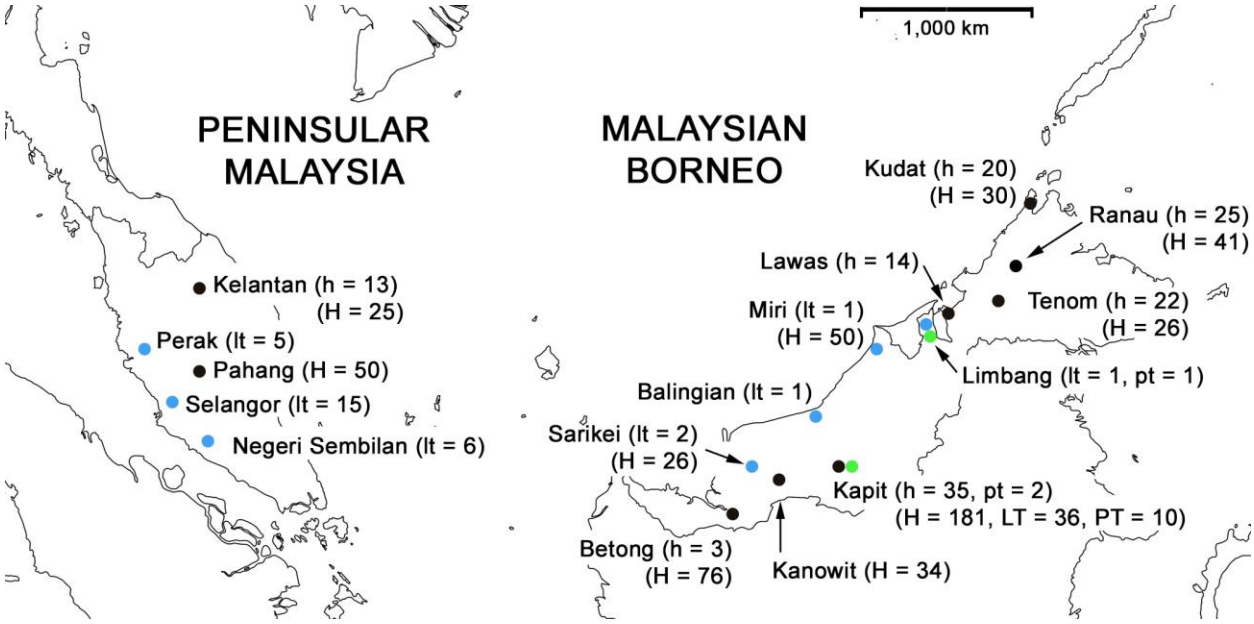
*All new and old samples (N) were genotyped at complete 10-microsatellite loci. Sites with N<4 were not included in the analyses. H_E, expected heterozygosity; I_A^S, standardized index of association.



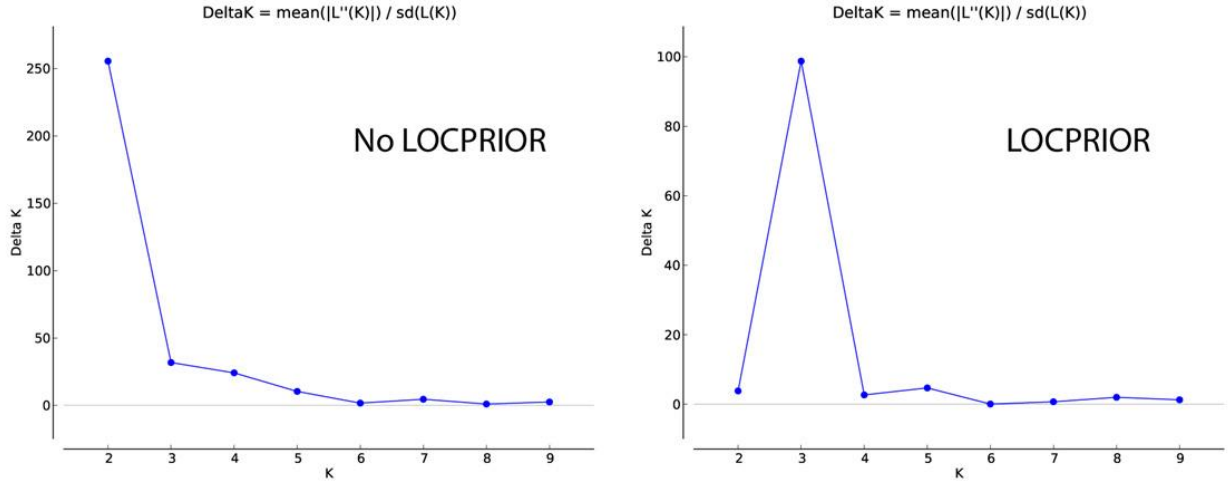
Technical Appendix 2 Figure 1. STRUCTURE analysis on 166 *P. knowlesi* infections across Malaysia and seven laboratory isolates obtained by using 10 microsatellite loci. Using both admixture with and without the LOCPRIOR models, we estimated 3 subpopulation clusters by Evanno's method (K = 3; ΔK = 128.51 for the non-LOCPRIOR model and ΔK = 37.72 for the LOCPRIOR model).



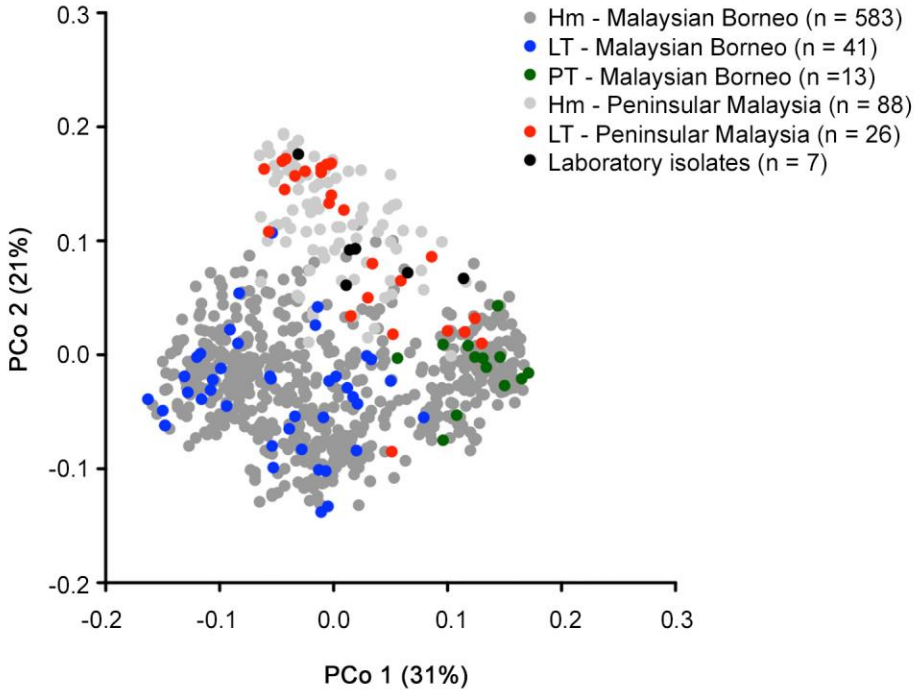
Technical Appendix 2 Figure 2. Cluster assignment of individual *P. knowlesi* genotypes obtained by using the Bayesian-based STRUCTURE algorithm without the LOCPRIOR model. For both analyses, complete 10-locus microsatellite genotyping on (A) 166 *P. knowlesi* infections in the present study and 7 laboratory isolates showed $K = 3$ with $\Delta K = 128.51$, and (B) a combined 758 *P. knowlesi* genotypes showed $K = 2$ with $\Delta K = 255.50$. Ancestral population clusters are referred to as cluster 1 (blue), cluster 2 (green), and cluster 3 (red). Hm, human; LT, long-tailed macaque; PT, pig-tailed macaque; Ot, various sources.



Technical Appendix 2 Figure 3. Total samples of *P. knowlesi* infections derived from 671 humans and 80 macaques across Malaysia with complete 10-microsatellite loci. Samples obtained from humans are marked with black dots and labeled with "h" for new samples and "H" for old samples; samples from long-tailed macaques are marked with blue dots and labeled with "lt" for new samples and "LT" for old samples; and samples from pig-tailed macaques are marked with green dots and labeled with "pt" for new samples and "PT" for old samples.



Technical Appendix 2 Figure 4. STRUCTURE analysis on 758 *P. knowlesi* genotypes obtained using 10 microsatellite loci. Estimated by Evanno's method, analysis of admixture without the LOCPRIOR model showed 2 subpopulation clusters (K = 2, ΔK = 255.50), whereas admixture with LOCPRIOR model showed 3 subpopulation clusters (K = 3, ΔK = 98.73).



Technical Appendix 2 Figure 5. Principal coordinate analysis deduced from genetic distance matrix of 10 microsatellite loci in 751 *P. knowlesi* infections across Malaysia and 7 laboratory isolates. Hm, human; LT, long-tailed macaque; PT, pig-tailed macaque.