

RESEARCH ARTICLE

# Admixture in Humans of Two Divergent *Plasmodium knowlesi* Populations Associated with Different Macaque Host Species

Paul C. S. Divis<sup>1,2\*</sup>, Balbir Singh<sup>2\*</sup>, Fread Anderios<sup>3</sup>, Shamilah Hisam<sup>4</sup>, Asmad Matusop<sup>2,5</sup>, Clemens H. Kocken<sup>6</sup>, Samuel A. Assefa<sup>1</sup>, Craig W. Duffy<sup>1</sup>, David J. Conway<sup>1,2\*</sup>

**1** Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Malaria Research Centre, Faculty of Medicine and Health Sciences, University Malaysia Sarawak, Kuching, Sarawak, Malaysia, **3** Sabah State Public Health Laboratory, Kota Kinabalu, Sabah, Malaysia, **4** Institute for Medical Research, Kuala Lumpur, Malaysia, **5** Sarawak State Health Department, Kuching, Sarawak, Malaysia, **6** Biomedical Primate Research Centre, Rijswijk, The Netherlands

\* [paul.divis@lshtm.ac.uk](mailto:paul.divis@lshtm.ac.uk) (PCS); [bskhaira55@gmail.com](mailto:bskhaira55@gmail.com) (BS); [david.conway@lshtm.ac.uk](mailto:david.conway@lshtm.ac.uk) (DJC)



 OPEN ACCESS

**Citation:** Divis PCS, Singh B, Anderios F, Hisam S, Matusop A, Kocken CH, et al. (2015) Admixture in Humans of Two Divergent *Plasmodium knowlesi* Populations Associated with Different Macaque Host Species. PLoS Pathog 11(5): e1004888. doi:10.1371/journal.ppat.1004888

**Editor:** Ananias A. Escalante, Arizona State University, UNITED STATES

**Received:** November 21, 2014

**Accepted:** April 17, 2015

**Published:** May 28, 2015

**Copyright:** © 2015 Divis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The research was supported by grants from UNIMAS to BS (grant number E14054/F05/54PK1/09/2012(01) and 01(TD03)/1003/2013(01)), the UK Medical Research Council to DJC (G1100123), an ERC Advanced Award to DJC (AdG-2011-294428) and a postgraduate scholarship for PCSD from the Ministry of Education in Malaysia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Human malaria parasite species were originally acquired from other primate hosts and subsequently became endemic, then spread throughout large parts of the world. A major zoonosis is now occurring with *Plasmodium knowlesi* from macaques in Southeast Asia, with a recent acceleration in numbers of reported cases particularly in Malaysia. To investigate the parasite population genetics, we developed sensitive and species-specific microsatellite genotyping protocols and applied these to analysis of samples from 10 sites covering a range of >1,600 km within which most cases have occurred. Genotypic analyses of 599 *P. knowlesi* infections (552 in humans and 47 in wild macaques) at 10 highly polymorphic loci provide radical new insights on the emergence. Parasites from sympatric long-tailed macaques (*Macaca fascicularis*) and pig-tailed macaques (*M. nemestrina*) were very highly differentiated (FST = 0.22, and K-means clustering confirmed two host-associated subpopulations). Approximately two thirds of human *P. knowlesi* infections were of the long-tailed macaque type (Cluster 1), and one third were of the pig-tailed-macaque type (Cluster 2), with relative proportions varying across the different sites. Among the samples from humans, there was significant indication of genetic isolation by geographical distance overall and within Cluster 1 alone. Across the different sites, the level of multi-locus linkage disequilibrium correlated with the degree of local admixture of the two different clusters. The widespread occurrence of both types of *P. knowlesi* in humans enhances the potential for parasite adaptation in this zoonotic system.

## Author Summary

Extraordinary phases of pathogen evolution may occur during an emerging zoonosis, potentially involving adaptation to human hosts, with changes in patterns of virulence and

**Competing Interests:** The authors have declared that no competing interests exist

transmission. In a large population genetic survey, we show that the malaria parasite *Plasmodium knowlesi* in humans is an admixture of two highly divergent parasite populations, each associated with different forest-dwelling macaque reservoir host species. Most of the transmission and sexual reproduction occurs separately in each of the two parasite populations. In addition to the reservoir host-associated parasite population structure, there was also significant genetic differentiation that correlated with geographical distance. Although both *P. knowlesi* types co-exist in the same areas, the divergence between them is similar to or greater than that seen between sub-species in other sexually reproducing eukaryotes. This may offer particular opportunities for evolution of virulence and host-specificity, not seen with other malaria parasites, so studies of ongoing adaptation and interventions to reduce transmission are urgent priorities.

## Introduction

The epidemiological emergence of infections can be traced by genotypic analyses, with a high level of resolution when pathogens have a high mutation rate, as illustrated by recently emerged viruses that now have a massive impact on global public health [1,2]. Such analysis is more challenging for eukaryote pathogens with low mutation rate, although it is now clear that the major human malaria parasites *Plasmodium falciparum* and *P. vivax* have been endemic for many thousands of years after having been acquired as zoonotic infections from African apes [3,4]. In contrast, natural human infections by *P. knowlesi* were almost unknown [5] until a large focus of cases in Malaysian Borneo was described a decade ago [6]. Infections have since been reported from throughout southeast Asia, within the geographical range of the long-tailed and pig-tailed macaque reservoir hosts (*Macaca fascicularis* and *M. nemestrina*) and mosquito vectors (of the *Anopheles leucosphyrus* group) [7]. The most highly affected country is Malaysia, where there have been thousands of reported cases and *P. knowlesi* is now the leading cause of malaria in most areas [8,9].

It is vital to determine the causes of this apparent emergence, as *P. knowlesi* can cause severe clinical malaria with a potentially fatal outcome [10–12]. Increasing rates of case detection may reflect better diagnosis, increased transmission by mosquitoes from reservoir host macaques to humans, or parasite adaptation to humans. Molecular tools to discriminate *P. knowlesi* from other malaria parasite species were not widely applied until the zoonosis became known, but analysis of DNA in archived blood samples from Malaysia and Thailand shows that it was already widespread twenty years ago [13,14]. Sequences of parasite mitochondrial genomes and a few nuclear gene loci indicate ongoing zoonotic infection, as human *P. knowlesi* genotypes share most alleles identified in parasites sampled from wild macaques [15–17].

To understand this zoonosis, and to identify whether human-to-human mosquito transmission is occurring, analyses of parasite population genetic structure in humans and macaques should be performed by extensive population sampling and characterisation of multiple putatively neutral loci. This study presents a *P. knowlesi* microsatellite genotyping toolkit and its application to the analysis of a large sample of isolates from human cases at ten different sites, as well as from both species of wild macaque reservoir hosts. Results reveal a profound host-associated sympatric subdivision within this parasite species, as well as geographical differentiation indicating genetic isolation by distance. The existence of two divergent parasite subpopulations, and their admixture in human infections provides unparalleled opportunity for parasite hybridisation and adaptation. Observations of some clinical infections with parasite types that

appear intermediate between the two subpopulations may reflect this process, and are a possible result of human-to-human mosquito transmission.

## Results

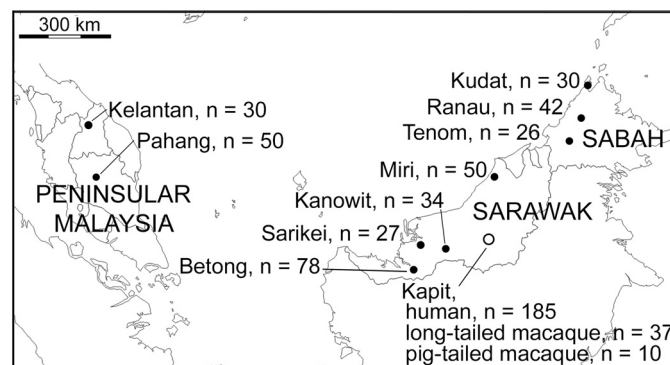
### *P. knowlesi* microsatellites as genetic markers for population studies

Hemi-nested PCR assays were developed for amplification of 19 tri-nucleotide simple sequence repeat loci from throughout the genome of *P. knowlesi* and tested for species-specificity using control DNA from all 10 known parasite species of humans, long-tailed or pig-tailed macaques, as well as human and macaque DNA to identify those suitable for genotyping samples from all hosts (S1 Table). Assays for 11 loci were entirely species-specific for *P. knowlesi*, and 10 of these gave a clear single electrophoretic peak for each allele without any stutter bands (S2 Table). These were used to genotype *P. knowlesi* infections in a total of 599 humans and wild macaques with a high rate of success, 556 (92.8%) scoring clearly for all 10 loci (S3 Table and S1 Dataset). Numbers of alleles at each locus ranged from 7 (for locus NC03\_2) to 21 (for locus CD05\_06) (S1 Dataset).

### Host-dependent genetic structure of *P. knowlesi*

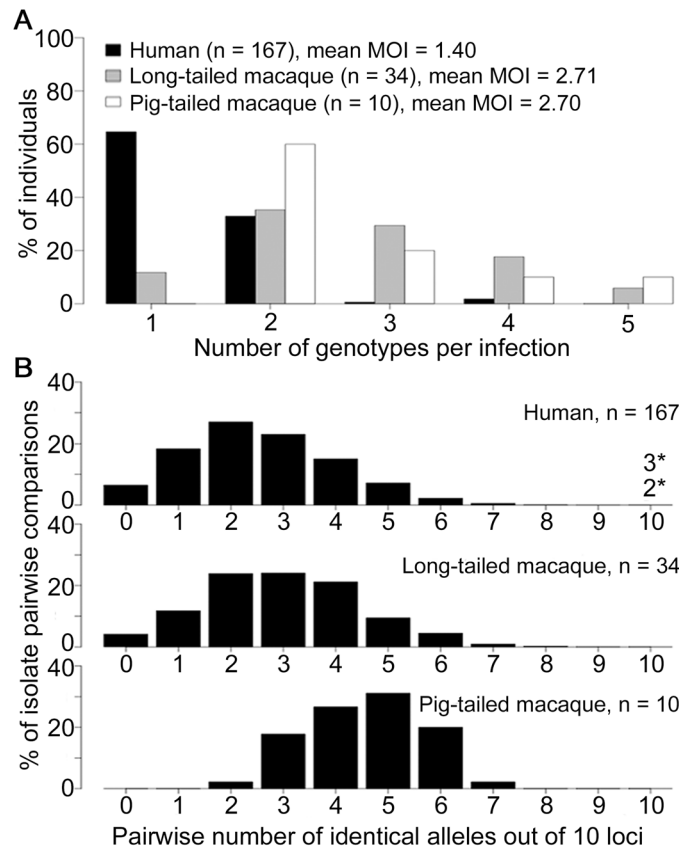
We first compared parasites from different host species sampled from Kapit where high numbers of clinical cases are seen (Fig 1), analysing the infections with complete 10-locus genotype data. Almost all *P. knowlesi* infections in macaques contained multiple genotypes, with no significant difference between long-tailed macaques (88% of 34 were mixed, a mean of 2.71 genotypes per infection) and pig-tailed macaques (100% of 10 were mixed, a mean of 2.70 genotypes per infection;  $P = 0.65$  for comparison between macaque host species), whereas only a minority of human *P. knowlesi* infections had multiple genotypes (35% of 167, a mean of 1.40 genotypes per infection;  $P < 10^{-15}$  for comparison between humans and macaques) (Fig 2A). To allow equally weighted sampling per host, the predominant allele at each locus within each infection was counted for subsequent analysis (S1 Dataset).

Pairwise comparisons of each of the complete 10-locus profiles revealed that all infections in Kapit were genotypically distinct, except for one identical pair and one identical triplet of human infections (Fig 2B, S4 Table). There was a much higher average proportion of shared alleles among pig-tailed macaque infections than among those in long-tailed macaques or



**Fig 1. Map of sampling locations of 599 *P. knowlesi* infections genotyped in this study.** A total of 552 samples were from *P. knowlesi* malaria patients from 10 geographical locations: Peninsular Malaysia (Kelantan and Pahang), Sarawak (Kapit, Betong, Miri, Kanowit and Sarikei), and Sabah (Kudat, Ranau and Tenom). Additionally, 47 samples were from wild macaques (37 long-tailed and 10 pig-tailed macaques) with *P. knowlesi* infections in Kapit.

doi:10.1371/journal.ppat.1004888.g001

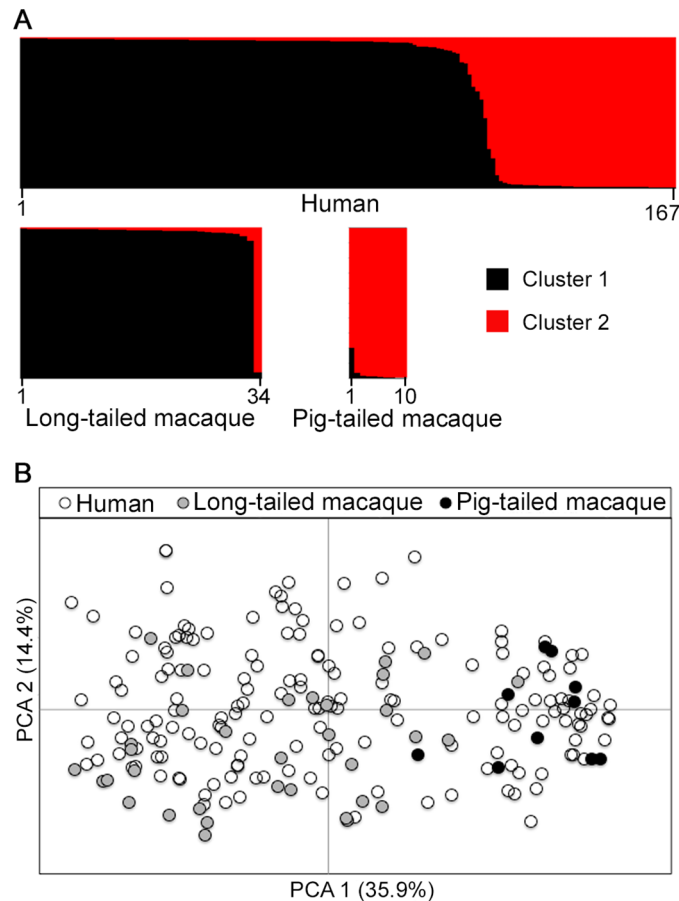


**Fig 2. Multiple genotype *P. knowlesi* infections and diversity among infections in three host species from Kapit.** All data for the multiple genotype infections and distribution of identical alleles between infections derived from complete genotyping of 10 microsatellite loci. (A) Numbers of different *P. knowlesi* genotypes per infection (multiplicity of infection, MOI) showing significant difference between human and macaque infections (Fisher's Exact  $P < 1 \times 10^{-15}$ ), but not between long-tailed and pig-tailed macaques (Fisher's Exact  $P = 0.65$ ). (B) Numbers of identical alleles out of 10 loci in pairwise comparisons of infections, showing a similar diversity among infections in humans and long-tailed macaques, but a higher average identity among infections from pig-tailed macaques. All infections had a different 10-locus genotype, except for 5 of the 167 human infections (there was one pair sharing an identical 10-locus genotype, and a triplet of infections sharing another 10-locus genotype, indicated with asterisks here and shown in [S4 Table](#)).

doi:10.1371/journal.ppat.1004888.g002

humans (medians of 5, 3 and 2 identical alleles out of 10 loci respectively). Analysis of allele frequencies revealed that *P. knowlesi* parasites from pig-tailed macaques are very highly divergent from those in long-tailed macaques ( $F_{ST} = 0.217$ ,  $P < 0.001$ ), whereas those in humans have an intermediate level of relatedness ( $F_{ST} = 0.067$  versus long-tailed macaques,  $F_{ST} = 0.104$  versus pig-tailed macaques;  $P < 0.001$  for both).

A Bayesian model-based STRUCTURE analysis of multi-locus genotype data from all hosts sampled in Kapit clearly indicated the existence of two sub-population clusters of *P. knowlesi* ( $K = 2$ ;  $\Delta K = 936.75$  based on Evanno's estimation of  $K$ -population) (Fig 3A, S1 Fig and S1 Dataset). An individual infection genotype was assigned to be predominantly of a particular cluster if the STRUCTURE analysis score exceeded 0.5 for that cluster. All except one of the long-tailed macaque infections were assigned to the Cluster 1 subpopulation, whereas all pig-tailed macaque infections were assigned to the Cluster 2 subpopulation, while 71% of human infections were assigned to Cluster 1 and 29% to Cluster 2 (Fig 3A, S5 Table). A small minority of those which were primarily assigned to either cluster appeared to have a degree of mixed



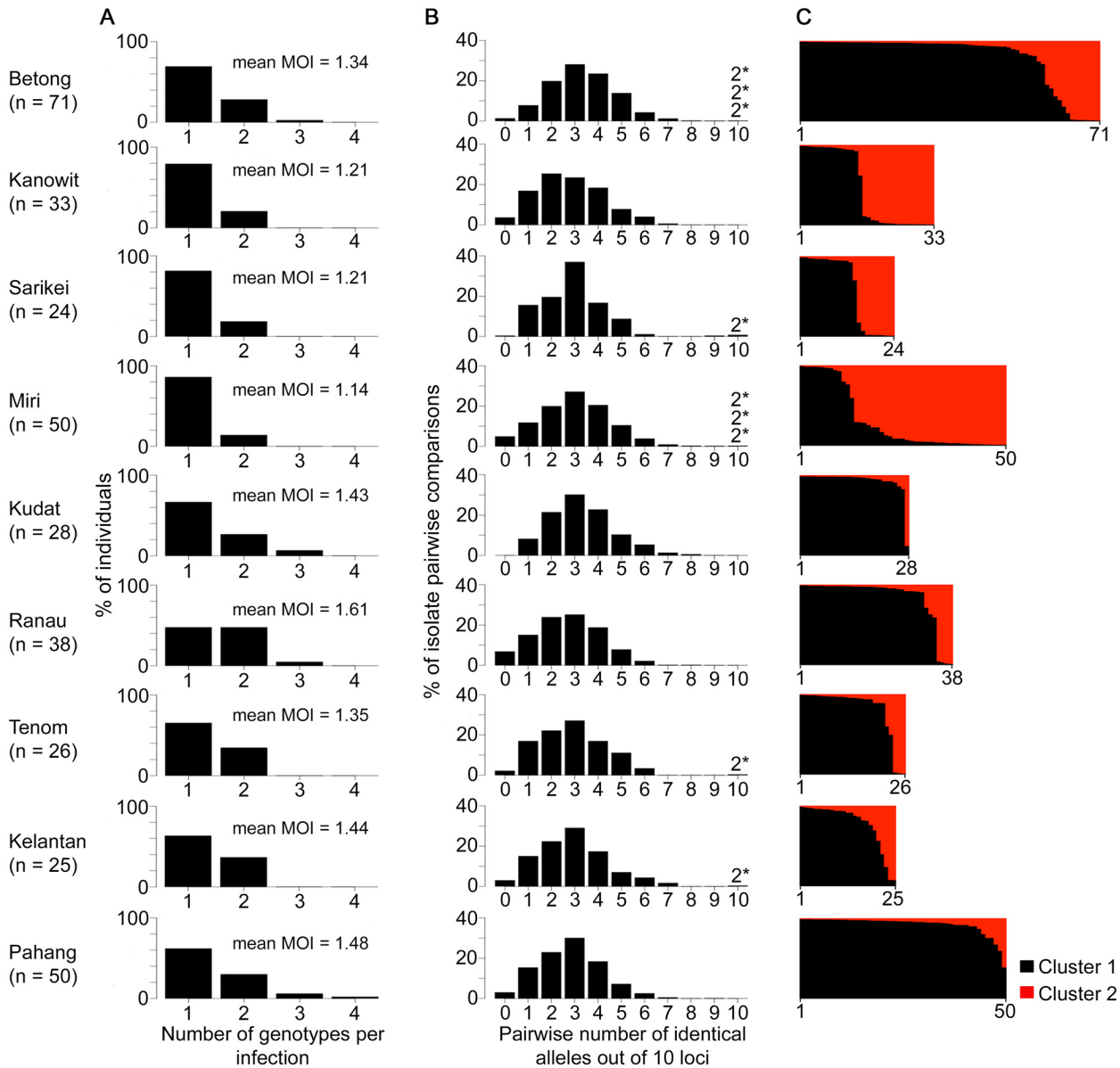
**Fig 3. Population genetic structure of *P. knowlesi* from infections in three host species in Kapit.** Both STRUCTURE and principal component analyses were conducted based on the complete 10-locus genotype dataset (167 from humans, 34 from long-tailed macaques, 10 from pig-tailed macaques). (A) Bayesian model-based STRUCTURE analysis indicates two subpopulation clusters throughout the whole dataset ( $K = 2$ ,  $\Delta K = 936.75$ ), with almost complete partitioning between the two macaque host species. Cluster 1 is shown in black while Cluster 2 is in red. (B) Principal component analysis (PCA) of the genetic divergence among all infections. The percentage of variation captured by each of the first two principal components is shown in brackets. Infections from the different macaque host species are almost completely separated by the first principal component, while human infections are distributed throughout the full range on both axes.

doi:10.1371/journal.ppat.1004888.g003

assignment, with scores nearer 0.5 than either zero or 1.0 for the alternative clusters (S1 Dataset), which is analysed in a separate section below. An independent scan by principal component analysis (PCA) showed an almost complete separation between parasites from long-tailed macaques and pig-tailed macaques along the first principal component, while parasites from humans covered the whole distribution and overlapped with all of the samples from both of the macaque hosts (Fig 3B).

### Geographical population genetic structure of *P. knowlesi*

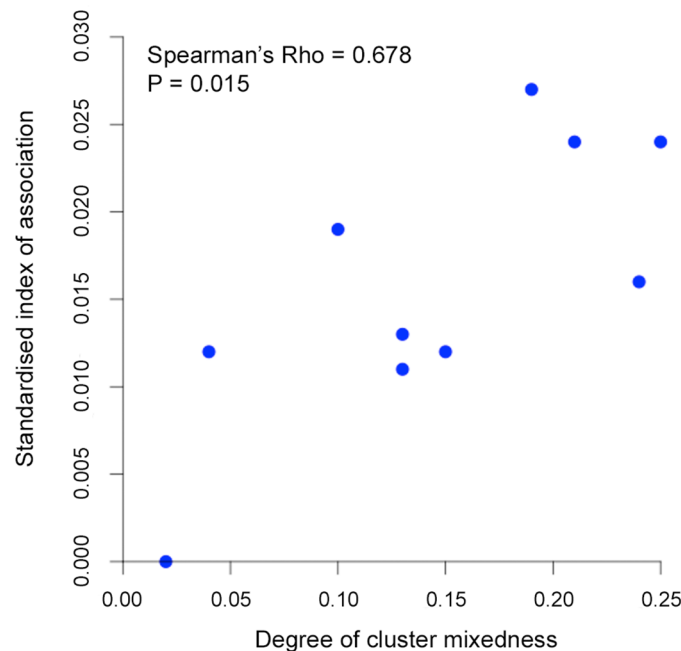
We analysed a further 367 human *P. knowlesi* infections from nine other geographical sites (Fig 1). Most human infections had single *P. knowlesi* genotypes (Fig 4A and S1 Dataset), and there were no differences in the proportions of mixed genotype infections across all sites (Comparison across 10 sites including Kapit: Pearson's  $X^2$ ,  $P = 0.096$ ; 32% of infections having > 1 genotype overall). There were no differences in allelic diversity among the different sites ( $H_E$



**Fig 4. Diversity and genetic structure of *P. knowlesi* in human infections from nine different geographical locations.** All *P. knowlesi* data for analysing multiple genotype infections, distribution of identical alleles between isolates and STRUCTURE analysis were derived from complete genotyping of 10 microsatellite loci. (A) Proportions of infections containing different numbers of genotypes (multiplicity of infection, MOI). Comparison across all 10 geographical locations including Kapit (data in Fig 2) showed no significant differences (Pearson's  $\chi^2$  with 10000 replicates,  $P = 0.096$ ). (B) Distribution of numbers of identical alleles out of 10 loci in pairwise comparisons of infections. In six of the populations, a small number of pairs of identical multi-locus genotypes were seen as indicated with a label here (2\*) and tabulated in detail in S4 Table. (C) Subpopulation clusters inferred by the Bayesian model-based STRUCTURE analysis with Cluster 1 (black) and Cluster 2 (red) corresponding to those identified in Fig 3 ( $K = 2$ ,  $\Delta K = 174.94$ ). Proportions of isolates assigned as Cluster 2 are highest at the sites in Sarawak (top four panels in the figure; locations of all sites are shown in Fig 1).

doi:10.1371/journal.ppat.1004888.g004

estimates between 0.67 and 0.75,  $P > 0.1$  for all pairwise Wilcoxon Signed Rank tests across all 10 loci, S3 Table). Pairwise comparisons among genotypes from different infections showed a similar level of diversity at each site, with a median of 2 or 3 identical alleles out of 10 loci in each site (Fig 4B, S4 Table). Every infection had a different multi-locus genotype, and there were virtually none that shared alleles at more than 7 loci, except for nine pairs of identical haplotypes (three pairs in Betong, three in Miri, and one in each of Sarikei, Tenom and Kelantan)



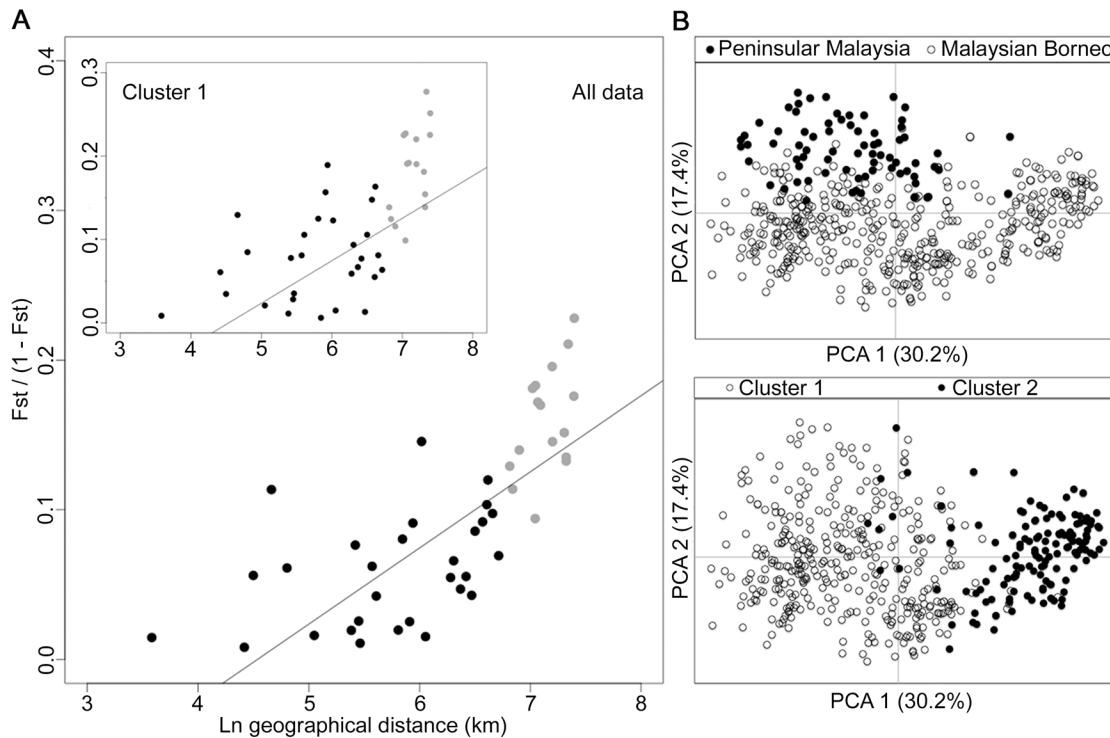
**Fig 5. Correlation between degree of cluster admixture and multi-locus linkage disequilibrium (standardised index of association).** The degree of *P. knowlesi* cluster admixture was estimated as the local cluster mixedness ( $p1 * p2$ ) based on the proportions of infections designated as Cluster 1 ( $p1$ ) and Cluster 2 ( $p2$ ) at each of 10 sampling sites for human infections across Malaysia (Spearman's Rho = 0.678,  $P = 0.015$ ). [S5 Table](#) gives details of each of these values for each population.

doi:10.1371/journal.ppat.1004888.g005

([Fig 4B](#)). Each identical haplotype pair was shared by infections from different individuals sampled at the same site within the same year, except for two of the identical haplotype pairs in Miri, shared by individual infections sampled one and two years apart ([S4 Table](#)).

There were two subpopulation clusters ( $K = 2$ ,  $\Delta K = 174.94$ , [S1 Fig](#)) throughout all of these sites, as had been seen in Kapit, but the relative frequency of the clusters varied geographically ( $P < 0.0001$ , [Fig 4C](#)). The Cluster 1 subpopulation was more frequent overall, but Cluster 2 was also common at each of the sites in Sarawak, particularly in Miri and Kanowit where it was more frequent than Cluster 1 ([S5 Table](#) and [S1 Dataset](#)). Over all human infections, there was a similarly high level of divergence in allele frequencies between the two subpopulation clusters as was seen between parasites from the two different macaque host species ( $F_{ST} = 0.194$ ,  $P < 0.001$ ). As expected, the degree of cluster admixture at each sampling site ( $p1 * p2$ , where  $p1$  and  $p2$  are the local frequencies of Cluster 1 and Cluster 2 respectively) correlated positively with the ( $I^S_A$ ) index of multi-locus linkage disequilibrium (Spearman's Rho = 0.678,  $P = 0.015$ , [Fig 5](#) and [S5 Table](#)).

Analysis of geographical divergence on the basis of  $F_{ST}$  indices derived from population allele frequencies ([S6 Table](#)) identified a pattern strongly consistent with isolation by distance (Mantel test of matrix correlation  $P < 0.0001$ , [Fig 6A](#)). The greatest level of divergence was seen between peninsular Malaysia and Borneo as expected, although isolation by distance was also apparent within Borneo (Mantel test  $P = 0.0016$ ). The overall pattern consistent with isolation by distance remained when only infections with Cluster 1 genotypes were analysed ( $P = 0.0016$ , [Fig 6A](#)). There was a similar trend for the smaller number of samples with Cluster 2 genotypes, although this was not significant ( $P = 0.0922$ , [S2 Fig](#)), indicating that the majority of the geographical differentiation is independent of the Cluster subpopulation structure. A principal



**Fig 6. Isolation-by-distance model and principal component analysis (PCA) of the human *P. knowlesi* isolates.** (A) Relationship between transformed genetic differentiation and natural log of geographical distance (Euclidean distances of population pairs ranged from 36 km to 1631 km) for all pairs of sites across Malaysia (Mantel test of matrix correlation  $P < 0.0001$ ), with a similar relationship when analysing only isolates from the Cluster 1 subpopulation (Mantel test  $P = 0.0016$ ). Black dots denote pairs of populations within Malaysian Borneo and Peninsular Malaysia, while grey dots denote pairs of populations between Malaysian Borneo and Peninsular Malaysia (all data are given in [S6 Table](#)). Evidence of isolation by distance remained when only sites within Borneo were considered (Mantel test  $P = 0.0016$ ). (B) PCA of the whole infection haplotype dataset indicated differentiation of Peninsular Malaysia isolates from Malaysian Borneo isolates by the second principal component axis, whereas isolates defined as Cluster 1 and Cluster 2 by STRUCTURE analysis were almost completely differentiated along the first principal component axis.

doi:10.1371/journal.ppat.1004888.g006

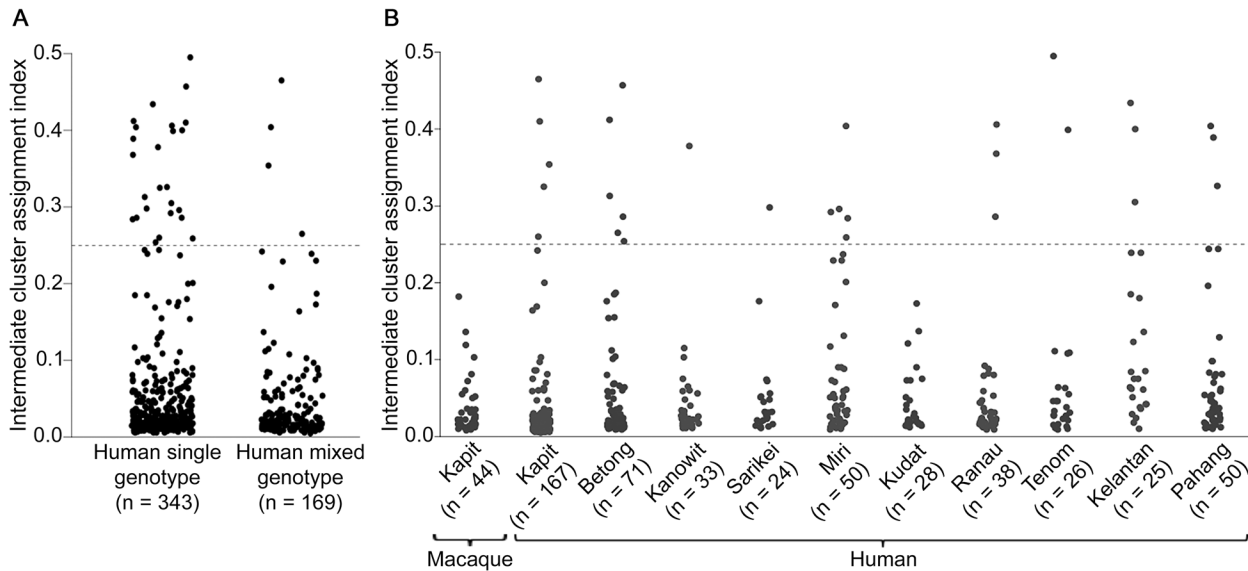
component analysis of all individual infection genotypes showed that most of the overall diversity is among those defined as Cluster 1 by the STRUCTURE analysis (Cluster 2 infections covered only part of the first principal component distribution), and infections from peninsular Malaysia are restricted to part of the second principal component distribution ([Fig 6B](#)).

Combination of the macaque samples together with all of the human samples across the 10 geographical locations confirmed the definition of the two *P. knowlesi* subpopulation clusters, which correspond to those shown above ([S3 Fig](#)). Allele frequency distributions showed that some loci were particularly differentiated between the subpopulation clusters, with  $F_{ST} > 0.3$  for loci NC03\_2 and CD13\_61 ([S4 Fig](#)). The robustness of the two assigned clusters was confirmed even with the exclusion of these most highly differentiated loci in the STRUCTURE analysis ([S5 Fig](#)).

### Evaluation of cluster assignment indices

Most individual infection genotypes had a clear majority of putative ancestry assignment to either Cluster 1 or Cluster 2, but a small minority of infections had a more intermediate profile ([Figs 3A and 4C](#)). Quantitative analysis of the proportional Cluster 1 and Cluster 2 ancestry assignments for each infection genotype based on the STRUCTURE analysis yielded an index of the degree of intermediate cluster assignment for each infection. This has a maximum possible





**Fig 7. Intermediate cluster assignment indices in *P. knowlesi* infections in humans and macaques.** The index for each infection was based on the proportion of shared ancestry between Cluster 1 and Cluster 2 inferred by the STRUCTURE analysis on all samples genotyped at 10 microsatellite loci. An index below 0.25 signifies the individual parasites were predominantly either Cluster 1 or Cluster 2 while those above 0.25 had more intermediate assignment (up to a maximum index value of 0.5). (A) There was no significant difference between single or multiple genotype human infections in the distribution of the indices (Mann-Whitney U test,  $P = 0.20$ ). (B) Intermediate cluster assignment indices were lower in infections from macaques in Kapit than in human infections overall (Mann-Whitney U test,  $P < 0.001$ ), but were not significantly different from human infections in Kapit ( $P = 0.25$ ). Among geographical sites, the indices for human infections were significantly higher in Kelantan than in Kapit, Betong, Kanowit, Kudat or Ranau ( $P < 0.05$  for each pairwise comparison after Bonferroni correction).

doi:10.1371/journal.ppat.1004888.g007

value of 0.5, although most infections had values closer to zero. The intermediate cluster assignment indices showed no difference between single and mixed genotype human infections (Mann-Whitney test  $P = 0.20$ , Fig 7A), whereas both of these independently had higher indices than the macaque infections ( $P < 0.001$  for both comparisons). When analysis was focused on Kapit alone, the distribution of intermediate cluster assignment indices were not significantly different between human and macaque infections ( $P = 0.25$ , Fig 7B). However, there were geographical differences, with human infections from Kelantan having a significantly higher distribution of values compared to five of the sites in Borneo (Mann-Whitney test  $P < 0.05$  for each comparison after Bonferroni correction, Fig 7B). Across the different sites, there was no significant correlation between the local population admixture of both clusters ( $p1 \times p2$ , S5 Table) and the mean or variance of intermediate cluster assignment indices ( $P = 0.33$  and  $P = 0.59$  respectively). Infections which had intermediate cluster assignment (index values  $> 0.25$ ) were not particularly closely related, having a similar degree of allele sharing as seen in the general local populations (S6 Fig, compared with Figs 2B and 4B).

## Discussion

We show that human *P. knowlesi* is an admixture of two divergent parasite populations associated with different forest-dwelling macaque reservoir hosts. In human infections, the long-tailed macaque-associated *P. knowlesi* type (Cluster 1) is most common overall and at most of the geographical sites, while the pig-tailed macaque-associated type (Cluster 2) is also common at sites in Sarawak. The estimate of divergence between these two sympatric parasite subpopulations ( $F_{ST}$  index of  $\sim 0.22$  averaged over 10 microsatellite loci) may be conservative, due to high allelic diversity of the microsatellite loci which restricts the potential upper range of

fixation indices [18,19]. The differentiation varied among the loci, with two of the microsatellite loci being particularly highly differentiated between the clusters ( $F_{ST} \sim 0.35$ ), so the robustness of the two assigned clusters was confirmed by repeat analyses which excluded these. Previous analysis of *P. knowlesi* mitochondrial DNA sequences from a relatively small number of human and long-tailed macaque infections in Kapit did not indicate two divergent lineages [15], although analysis of samples from Sabah suggests that sequences from pig-tailed macaque infections are differentiated from sequences from long-tailed macaque infections [20].

The results confirm that humans have mostly single genotype *P. knowlesi* infections whereas macaques have polyclonal infections, supporting the expectation that there is a higher rate of transmission among macaques [15,21]. The estimated number of genotypes per infection here is a minimum number based on the alleles detected, and it is possible that some infections may have contained additional parasite clones that were not detected, due to having low density in the blood or having similar alleles to the ones detected. The number of *P. knowlesi* genotypes detected per infection in humans is lower than was previously seen in microsatellite analyses of the endemic human malaria parasites *P. falciparum* and *P. vivax* in some of the same areas in Malaysia [22,23], whereas the number of *P. knowlesi* genotypes per infection in macaques is much higher. Levels of multi-locus linkage disequilibrium in *P. knowlesi* here are lower than reported in *P. vivax* or *P. falciparum* in these areas [22,23], indicating that recombination in *P. knowlesi* probably commonly occurs in mosquitoes containing a macaque blood meal with multiple parasite genotypes.

It is unknown how the two sympatric *P. knowlesi* subpopulations are genetically isolated. The observation of a single long-tailed macaque with a *P. knowlesi* Cluster 2 type infection (otherwise only seen in pig-tailed macaques and humans) suggests there is not an absolute barrier in terms of primate host susceptibility, although there are differences in ecology. Additional sampling of both long-tailed and pig-tailed macaques will be important to confirm the host associations of different parasites [20]. Both macaque species are widespread, but long-tailed macaques prefer secondary forest near human settlements where they have access to farms for food, whereas pig-tailed macaques spend more time in ground foraging in primary forests, generally having less frequent contact with humans [24]. There may be differential susceptibility of mosquito species to the respective parasite types, as suggested for subpopulations of another malaria parasite elsewhere [25], or different mosquitoes may feed on the respective macaque host species. Genetic differentiation in *P. knowlesi* was also strongly correlated with geographical distance, overall and for the Cluster 1 parasites. The observation of highest  $F_{ST}$  values between populations from Malaysian Borneo and Peninsular Malaysia was expected, as the South China Sea has separated macaques in these areas since the last glacial period [26], but a test for isolation by distance remained significant when analysing only sites within Borneo.

A small minority of human infections had intermediate cluster assignment indices, which could potentially result from occasional crossbreeding between the two genotypic clusters, although this cannot be concluded from these data alone. Hybridisation between species or subspecies can offer opportunities for adaptation, and has been associated with emergence of novel host-specificity or pathogenicity in other parasitic protozoa [27] and fungi [28]. Switching of host species has occurred repeatedly in malaria parasites of birds [29] and small mammals [30], as well as apes and humans [3,4], but the occurrence of parasite hybridisation and introgression has not been investigated. The potential occurrence of inter-cluster hybridisation in even a minority of human *P. knowlesi* infections, combined with the possibility of human-mosquito-human transmission, may increase the potential for *P. knowlesi* adaptation to the human host or to mosquito species that are more abundant than the currently known forest-associated vectors.

Genome-wide analysis of *P. knowlesi* populations would enable further evaluation of the genetic structure of this zoonotic parasite species, and allow scanning for loci under selection

within each of the two subpopulations. Human clinical isolates containing single species infections would be relatively straightforward to analyse, as *P. knowlesi* sequences would be unmixed with those of other human malaria species. In contrast, as natural macaque infections usually contain a mixture of different malaria parasite species [15], to obtain unambiguous genome sequences it may be necessary to sequence from individual parasites isolated from these hosts [31]. Although experimental studies on *P. knowlesi* are usually conducted *in vivo* in non-human primates [32–34], new approaches to adapt the parasites to *in vitro* growth using human erythrocytes have been successful [35,36]. Analysis of phenotypic differences between the different host-associated types may be investigated using both *in vivo* and *in vitro* experimental systems, while continued epidemiological and clinical surveillance for increasing incidence or disease severity is of the highest priority.

## Materials and Methods

### Ethics statement

Human blood samples were taken after written informed consent had been obtained from patients. This study was approved by the Medical Research and Ethics Committee of the Malaysian Ministry of Health (Reference number: NMRR-12-1086-13607), which operates in accordance to the International Conference of Harmonization Good Clinical Practice Guidelines. Animal sampling was carried out as previously described [15] in strict accordance with the recommendations by the Sarawak Forestry Department for the capture, use and release of wild macaques. A veterinarian took a venous blood sample from each macaque following anesthesia by intramuscular injection of tiletamine and zolazepam, and all efforts were made to minimize suffering by collecting blood at the trap sites and releasing the animals immediately after the blood samples had been obtained. The Sarawak Forestry Department approved the study protocol for capture, collection of blood samples and release of wild macaques (Permit Numbers: NPW.907.4.2–32, NPW.907.4.2–97, NPW.907.4.2–98, 57/2006 and 70/2007). A permit to access and collect macaque blood samples for the purpose of research was also obtained from the Sarawak Biodiversity Centre (Permit Number: SBC-RP-0081-BS).

### *P. knowlesi* samples from humans and macaques

A total of 599 DNA samples from different *P. knowlesi* infections of humans and macaques were analysed from collections performed at 10 different geographical sites (Fig 1). 552 samples were from human *P. knowlesi* malaria patients from all of the sites, eight in Malaysian Borneo (Sarawak and Sabah states) and two in Peninsular Malaysia (Kelantan and Pahang states). For samples from Sarawak, DNA was extracted at the University Malaysia Sarawak (UNIMAS) in Kuching from previously reported blood samples collected between 2000 and 2011 [6,10,37,38] as well as new samples collected in 2012 and 2013, allowing analysis of five sites: Kapit (n = 185), Betong (n = 78), Kanowit (n = 34), Sarikei (n = 27) and Miri (n = 50). Samples from Sabah were collected in 2013, and DNA was extracted by the Sabah Public Health Reference Laboratory, allowing analysis of three sites: Kudat (n = 30), Ranau (n = 42) and Tenom (n = 26). For Peninsular Malaysia, blood samples collected from Kelantan (n = 30) and Pahang (n = 50) underwent DNA extraction at the Institute for Medical Research in Kuala Lumpur. We analysed DNA from blood samples of a total of 47 wild macaques (long-tailed macaque, *Macaca fascicularis* n = 37; pig-tailed macaque, *M. nemestrina* n = 10) previously collected within 30 km radius of Kapit town in Sarawak [15]. The map locations and dates of the individual macaque sampling is shown in S7 Fig. The presence of *P. knowlesi* DNA was confirmed in all samples at UNIMAS by nested PCR assays [6,15].

## Development of *P. knowlesi* microsatellite genotyping markers

A combination of three microsatellite mining tools (iMEX [39], mreps [40], and MSATCOMMANDER [41]) were used to identify simple sequence repeat loci from the *P. knowlesi* reference genome [42]. Loci with perfect tri-nucleotide simple repeat sequences were carefully selected using customised perl-script commands based on narrow criteria to maximise their likely utility for genotyping: i) a minimum of 7 repeat copies in each microsatellite in the reference sequence, ii) located at non-telomeric chromosomal regions as defined by regions syntenic with the *P. vivax* reference genome [43], iii) absence of any homopolymeric tracts adjacent to the microsatellite sequence that could give rise to additional size polymorphism. As a result, 19 trinucleotide repeat loci widely spaced in the genome (S1 Table) were shortlisted and PCR primers were designed using PrimerSelect software (DNASTAR, USA) for hemi-nested PCR assays. The specificity of PCR was tested using DNA controls of all human *Plasmodium* species, common malaria parasites of the Southeast Asian macaques (*P. knowlesi*, *P. coatneyi*, *P. inui*, *P. cynomolgi* and *P. fieldi*), as well as human, long-tailed and pig-tailed macaque DNA. Loci for which primers showed complete specificity of amplification from *P. knowlesi* were tested further for genotyping performance.

## PCR and genotyping protocols

Genotyping of each microsatellite locus was performed using a hemi-nested protocol with a fluorescent dye-labelled inner primer during the second round PCR amplification (primers listed in S1 Table). Both first and second round PCR amplifications were conducted in individual tubes or wells for each locus, in 11  $\mu$ l reaction volume containing 0.2 mM each dNTP (Bio-line, UK), 2 mM MgSO<sub>4</sub>, 1X ThermoPol II reaction buffer (NEB, UK), 0.275 U *Taq* DNA polymerase (NEB, UK), 0.1  $\mu$ M of each forward and reverse primer, and 1  $\mu$ l sample DNA template. The PCR cycling conditions were as follows: initial denaturation at 94°C for 2 min, followed by 28 cycles of 94°C for 30 sec, annealing at 56°C for 30 sec and elongation at 68°C for 30 sec, with a final elongation step at 68°C for 1 min. Final PCR products were pooled into three groups of loci with different product size and dye profiles together with Genescan 500 LIZ molecular size standards (Applied Biosystems, UK) and run on a Genetic Analyzer 3730 capillary electrophoretic system (Applied Biosystems, UK). GENEMAPPER version 4.0 software (Applied Biosystems, UK) was used for scoring of allele electrophoretic size, and quantification of peak heights.

## Genotypic and statistical analyses

Infections containing multiple haploid parasite genotypes were apparent as multiple electrophoretic peaks for a locus corresponding to different alleles. The apparent genotypic multiplicity of infection (MOI) was determined by the locus with the most alleles detected in the infection, considering peaks with height of at least 25% relative to the predominant allele within each isolate. The predominant allele per locus within each infection was counted for subsequent population genetic analyses. Allelic diversity at each locus was measured as the virtual heterozygosity ( $H_E$ ) using FSTAT software version 2.9.3.2 (<http://www2.unil.ch/popgen/softwares/fstat.htm>), and allele frequency distributions were also inspected using GenALEx version 6 [44] within the Microsoft Excel platform. Genetic differentiation between each population was measured by pairwise fixation indices ( $F_{ST}$ ) using FSTAT, with Bonferroni correction on a nominal significance level of 0.05 applied for multiple comparisons across the population pairs. To test for correlation between genetic differentiation and geographical distance, a Mantel test for isolation by distance was performed with Rousset's linearised  $F_{ST}/(1-F_{ST})$  plotted against the natural log of geographic distance using Genepop version 4.2 [45].

The relatedness of haplotypes between individual isolates was assessed by measuring the pairwise proportion of shared alleles, excluding samples with missing data at any locus. A matrix of pairwise similarity among isolates was calculated based on the identical or mismatched alleles from a complete set of loci and the distribution of shared alleles between sample pairs for each population was visualised using a customised perl-script command. To test for non-random allele assortment, multi-locus linkage disequilibrium (LD) was assessed by the standardised index of association ( $I_A^S$ ) using LIAN version 3.6 [46], with significance of the  $I_A^S$  values tested by Monte-Carlo simulation with 10,000 data permutations to generate the null distribution under linkage equilibrium.

To explore evidence of population substructure in the entire population, a Bayesian analysis was performed using the STRUCTURE version 2.3.4 software [47] using samples with no missing data at any locus. Individuals in the population pool were clustered to the most likely population ( $K$ ) by measuring the probability of ancestry using the multi-locus genotype data. The program parameters were set to admixture model with correlated allele frequency, with 50,000 burn-in period and 100,000 Markov chain (MCMC) iterations. To run the simulation,  $K$  value was predefined from 1–10 and the run was performed in 20 replicates for each  $K$ . The most probable  $K$  value was then calculated according to Evanno's method [48] using the webpage interface STRUCTURE Harvester [49]. The assignment of a sample to a subpopulation cluster was based on the inferred cluster scores by STRUCTURE analysis, where samples with inferred cluster scores within a range in relation to the  $K$ -value were assigned together as one subpopulation cluster. The intermediate cluster assignment indices were calculated based on the proportion of shared cluster ancestries per individual isolate inferred by the cluster scores from the STRUCTURE analysis.

We also independently performed a principal component analysis (PCA) using the GenALEx package for the same purpose. Samples with missing data at any locus were excluded, and the genetic distance matrix was generated based on the allelic mismatches between pairs of isolates. A two-dimensional PCA plot was generated considering the first two highest eigenvalues, and genetic clusters were determined based on the eigenvector coordinates along the axes of variation.

## Supporting Information

**S1 Fig. Plots of delta  $K$  ( $\Delta K$ ) based on Evanno's method for the determination of hypothetical ancestral population cluster ( $K$ ) from the STRUCTURE analysis extracted using the STRUCTURE Harvester.** (A) All STRUCTURE analyses of *P. knowlesi* infections strongly indicate the existence of two probable cluster subpopulations when analyses were performed with (A) 44 macaque and 167 human infections from Kapit, Sarawak ( $K = 2$ ,  $\Delta K = 936.75$ ), (B) only 512 human infections from all 10 sampling sites ( $K = 2$ ,  $\Delta K = 174.94$ ) and (C) all 44 macaque and 512 human infections from Malaysia ( $K = 2$ ,  $\Delta K = 136.39$ ). (D) In a similar analysis, reanalysis of 404 macaque and human isolates from the Cluster 1 population did not resolve any further population clusters, as indicated by the very low values of delta  $K$ . (DOCX)

**S2 Fig. Isolation-by-distance model of *P. knowlesi* from human infections of Cluster 1 and Cluster 2 subpopulations.** Positive relationship between transformed genetic differentiation and natural log of geographical distance was observed for Cluster 1 subpopulation, which is represented by open circles (Mantel's test of matrix correlation,  $P = 0.0016$ ). In contrary, no significant relationship was observed for isolates from Cluster 2 subpopulation, which is represented in closed circles (Mantel's test of matrix correlation,  $P = 0.0922$ ). Due to limited number of samples in Cluster 2 subpopulation, isolates from each site of Sabah and Peninsular Malaysia

were grouped together (total  $n = 9$  and  $n = 5$ , respectively) to obtain the maximum number of samples prior to perform the test.

(DOCX)

**S3 Fig. Population genetic structure of *P. knowlesi* infections from all 512 humans, 34 long-tailed macaques and 10 pig-tailed macaques in Malaysia.** (A) Bayesian model-based STRUCTURE analysis corresponds to those mention in Fig 3A where two subpopulation clusters were observed throughout the whole dataset ( $K = 2$ ,  $\Delta K = 136.39$ ). When all of the human and macaque samples were analysed together, little evidence of admixture seen between parasites sampled in the two macaque populations. (B) The principal component analysis (PCA) of all *P. knowlesi* isolates also indicates the infections from different macaque host species were almost completely separated by the first principal component while human infections were widely distributed throughout the full range on both axes.

(DOCX)

**S4 Fig. Allele frequency distributions and genetic differentiations of 10 microsatellite loci between two *P. knowlesi* subpopulation clusters.** Major subpopulation cluster (Cluster 1,  $n = 404$ ) is colour-coded in black while minor subpopulation cluster (Cluster 2,  $n = 152$ ) is in red.

(DOCX)

**S5 Fig. Re-running of Bayesian-approach STRUCTURE analysis on *P. knowlesi* of 512 human and 44 macaque infections using all microsatellite loci except for locus NC03\_2 and locus CD13\_61, which  $F_{ST}$  values  $> 0.3$  between subpopulation clusters obtained from the initial STRUCTURE analysis results.** (A) STRUCTURE analysis run with all loci except for locus NC03\_2 indicated two subpopulation clusters throughout the whole dataset the ( $K = 2$ ,  $\Delta K = 142.50$ ). (B) STRUCTURE analysis run with all loci except for locus CD13\_61 indicated two subpopulation clusters throughout the whole dataset the ( $K = 2$ ,  $\Delta K = 23.17$ ).

(DOCX)

**S6 Fig. Distribution of numbers of identical alleles out of 10 loci in pairwise comparisons of infections with intermediate cluster assignment indices above 0.25 from the overall STRUCTURE analysis.**

(DOCX)

**S7 Fig. Map illustrates the macaque sampling sites from locations within a 30 km radius of Kapit town, Sarawak.** A total of 47 wild macaques were obtained within 30 km radius of Kapit town, Sarawak. Out of these, 37 were long-tailed macaques obtained near Kapit Waterboard from May 2004 to 2006 ( $n = 8$ ), Rumah Braoh from March 2005 to January 2008 ( $n = 18$ ), Sungai Seranau on July 2007 ( $n = 1$ ), Sungai Mujong on August 2007 ( $n = 1$ ), Sungai Sebaby on November 2007 ( $n = 1$ ) Jalan Sungai Sut on March 2008 ( $n = 2$ ), Rumah Belikau on March 2008 ( $n = 4$ ) and Sungai Antaroh on March and April 2008 ( $n = 2$ ). Ten pig-tailed macaques were obtained at Sungai Sut on March 2006 ( $n = 1$ ), Rumah Untang near Sungai Yong from March to April 2008 ( $n = 8$ ) and Sungai Setapang on April 2008 ( $n = 1$ ). Map was accessed and modified from <http://landsatlook.usgs.gov/>.

(DOCX)

**S1 Table. Primers for genotyping of *P. knowlesi* microsatellites and location of loci in the reference genome sequence.**

(DOCX)

**S2 Table. Species-specificity of primers for 19 *P. knowlesi* microsatellite loci.**

(DOCX)

**S3 Table. Numbers of *P. knowlesi* infections genotyped and allelic diversity in each population.**

(DOCX)

**S4 Table. Genotypic data of 10 pairs and one triplet of identical haplotypes detected in six geographical locations.**

(DOCX)

**S5 Table. Degree of local Cluster 1 and Cluster 2 mixedness and test of multi-locus disequilibrium of *P. knowlesi* in human infections sampled at 10 sites across Malaysia.**

(DOCX)

**S6 Table. Pairwise measures of fixation indices ( $F_{ST}$  values above diagonal) and geographical distance (in kilometres below diagonal) across 10 populations of *P. knowlesi* from human infections.**

(DOCX)

**S1 Dataset. Genotypes at *P. knowlesi* microsatellite loci in infections from macaques (n = 47) and humans (n = 552).**

(XLSX)

## Acknowledgments

We are grateful to all patients, as well as clinical staff and field workers who provided and collected the samples for this project. We thank the Director General of Health Malaysia for permission to publish this paper. We also thank our colleagues for assistance, including Dayang Shuaisyah Awang Mohamed and Khamisah Abdul Kadir at the Malaria Research Centre, UNIMAS, for laboratory assistance, Colin Sutherland and colleagues at the UK Malaria Reference Laboratory for providing DNA controls for all human *Plasmodium* species, Alan Thomas and colleagues at BPRC for support in preparation of DNA controls for all macaque *Plasmodium* species, and Ozan Gundogdu and Eloise Thompson for specialised laboratory equipment management.

## Author Contributions

Conceived and designed the experiments: PCSD BS DJC. Performed the experiments: PCSD BS DJC. Analyzed the data: PCSD BS DJC. Contributed reagents/materials/analysis tools: FA SH AM BS CHK SAA CWD. Wrote the paper: PCSD BS DJC.

## References

1. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, et al. (2014) HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346: 56–61. doi: [10.1126/science.1256739](https://doi.org/10.1126/science.1256739) PMID: [25278604](https://pubmed.ncbi.nlm.nih.gov/25278604/)
2. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, et al. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345: 1369–1372. doi: [10.1126/science.1259657](https://doi.org/10.1126/science.1259657) PMID: [25214632](https://pubmed.ncbi.nlm.nih.gov/25214632/)
3. Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, et al. (2010) Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467: 420–425. doi: [10.1038/nature09442](https://doi.org/10.1038/nature09442) PMID: [20864995](https://pubmed.ncbi.nlm.nih.gov/20864995/)
4. Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, et al. (2014) African origin of the malaria parasite *Plasmodium vivax*. *Nat Commun* 5: 3346. doi: [10.1038/ncomms4346](https://doi.org/10.1038/ncomms4346) PMID: [24557500](https://pubmed.ncbi.nlm.nih.gov/24557500/)
5. Coatney GR, Collin WE, Warren M, Contacos PG (1971) *The Primate Malariae*. Washington: U.S. Government Printing Office.
6. Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, et al. (2004) A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* 363: 1017–1024. PMID: [15051281](https://pubmed.ncbi.nlm.nih.gov/15051281/)

7. Singh B, Daneshvar C (2013) Human infections and detection of *Plasmodium knowlesi*. Clin Microbiol Rev 26: 165–184. doi: [10.1128/CMR.00079-12](https://doi.org/10.1128/CMR.00079-12) PMID: [23554413](https://pubmed.ncbi.nlm.nih.gov/23554413/)
8. Yusof R, Lau YL, Mahmud R, Fong MY, Jelip J, et al. (2014) High proportion of knowlesi malaria in recent malaria cases in Malaysia. Malar J 13: 168. doi: [10.1186/1475-2875-13-168](https://doi.org/10.1186/1475-2875-13-168) PMID: [24886266](https://pubmed.ncbi.nlm.nih.gov/24886266/)
9. William T, Jelip J, Menon J, Anderios F, Mohammad R, et al. (2014) Changing epidemiology of malaria in Sabah, Malaysia: increasing incidence of *Plasmodium knowlesi*. Malar J 13: 390. doi: [10.1186/1475-2875-13-390](https://doi.org/10.1186/1475-2875-13-390) PMID: [25272973](https://pubmed.ncbi.nlm.nih.gov/25272973/)
10. Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, et al. (2008) *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. Clin Infect Dis 46: 165–171. doi: [10.1086/524888](https://doi.org/10.1086/524888) PMID: [18171245](https://pubmed.ncbi.nlm.nih.gov/18171245/)
11. Rajahram GS, Barber BE, William T, Menon J, Anstey NM, et al. (2012) Deaths due to *Plasmodium knowlesi* malaria in Sabah, Malaysia: association with reporting as *Plasmodium malariae* and delayed parenteral artesunate. Malar J 11: 284. doi: [10.1186/1475-2875-11-284](https://doi.org/10.1186/1475-2875-11-284) PMID: [22905799](https://pubmed.ncbi.nlm.nih.gov/22905799/)
12. Barber BE, William T, Grigg MJ, Menon J, Auburn S, et al. (2013) A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from *Plasmodium knowlesi* and *Plasmodium vivax* but no mortality with early referral and artesunate therapy. Clin Infect Dis 56: 383–397. doi: [10.1093/cid/cis902](https://doi.org/10.1093/cid/cis902) PMID: [23087389](https://pubmed.ncbi.nlm.nih.gov/23087389/)
13. Lee KS, Cox-Singh J, Brooke G, Matusop A, Singh B (2009) *Plasmodium knowlesi* from archival blood films: further evidence that human infections are widely distributed and not newly emergent in Malaysian Borneo. Int J Parasitol 39: 1125–1128. doi: [10.1016/j.ijpara.2009.03.003](https://doi.org/10.1016/j.ijpara.2009.03.003) PMID: [19358848](https://pubmed.ncbi.nlm.nih.gov/19358848/)
14. Jongwutiwes S, Buppan P, Kosuvin R, Seethamchai S, Pattanawong U, et al. (2011) *Plasmodium knowlesi* malaria in humans and macaques, Thailand. Emerg Infect Dis 17: 1799–1806. doi: [10.3201/eid1710.110349](https://doi.org/10.3201/eid1710.110349) PMID: [22000348](https://pubmed.ncbi.nlm.nih.gov/22000348/)
15. Lee KS, Divis PC, Zakaria SK, Matusop A, Julin RA, et al. (2011) *Plasmodium knowlesi*: reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog 7: e1002015. doi: [10.1371/journal.ppat.1002015](https://doi.org/10.1371/journal.ppat.1002015) PMID: [21490952](https://pubmed.ncbi.nlm.nih.gov/21490952/)
16. Fong MY, Lau YL, Chang PY, Anthony CN (2014) Genetic diversity, haplotypes and allele groups of Duffy binding protein (PkDBP $\alpha$ h1) of *Plasmodium knowlesi* clinical isolates from Peninsular Malaysia. Parasit Vectors 7: 161. doi: [10.1186/1756-3305-7-161](https://doi.org/10.1186/1756-3305-7-161) PMID: [24693997](https://pubmed.ncbi.nlm.nih.gov/24693997/)
17. Putaporntip C, Thongaree S, Jongwutiwes S (2013) Differential sequence diversity at merozoite surface protein-1 locus of *Plasmodium knowlesi* from humans and macaques in Thailand. Infect Genet Evol 18: 213–219. doi: [10.1016/j.meegid.2013.05.019](https://doi.org/10.1016/j.meegid.2013.05.019) PMID: [23727342](https://pubmed.ncbi.nlm.nih.gov/23727342/)
18. Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol 11: 155–165. PMID: [11856418](https://pubmed.ncbi.nlm.nih.gov/11856418/)
19. Meirmans PG, Hedrick PW (2011) Assessing population structure: F(ST) and related measures. Mol Ecol Resour 11: 5–18. doi: [10.1111/j.1755-0998.2010.02927.x](https://doi.org/10.1111/j.1755-0998.2010.02927.x) PMID: [21429096](https://pubmed.ncbi.nlm.nih.gov/21429096/)
20. Muehlenbein MP, Pacheco MA, Taylor JE, Prall SP, Ambu L, et al. (2015) Accelerated diversification of nonhuman primate malaras in southeast Asia: adaptive radiation or geographic speciation? Mol Biol Evol 32: 422–439. doi: [10.1093/molbev/msu310](https://doi.org/10.1093/molbev/msu310) PMID: [25389206](https://pubmed.ncbi.nlm.nih.gov/25389206/)
21. Tan CH, Vythilingam I, Matusop A, Chan ST, 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. doi: [10.1186/1475-2875-7-52](https://doi.org/10.1186/1475-2875-7-52) PMID: [18377652](https://pubmed.ncbi.nlm.nih.gov/18377652/)
22. Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, et al. (2005) Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. J Infect Dis 191: 1558–1564. PMID: [15809916](https://pubmed.ncbi.nlm.nih.gov/15809916/)
23. Abdullah NR, Barber BE, William T, Norahmad NA, Satsu UR, et al. (2013) *Plasmodium vivax* Population Structure and Transmission Dynamics in Sabah Malaysia. PLoS One 8: e82553. doi: [10.1371/journal.pone.0082553](https://doi.org/10.1371/journal.pone.0082553) PMID: [24358203](https://pubmed.ncbi.nlm.nih.gov/24358203/)
24. Fa JE, Lindburg DG (1996) Evolution and ecology of macaque societies: Cambridge University Press. 616 p.
25. Joy DA, Gonzalez-Ceron L, Carlton JM, Gueye A, Fay M, et al. (2008) Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. Mol Biol Evol 25: 1245–1252. doi: [10.1093/molbev/msn073](https://doi.org/10.1093/molbev/msn073) PMID: [18385220](https://pubmed.ncbi.nlm.nih.gov/18385220/)
26. Ziegler T, Abegg C, Meijaard E, Perwitasari-Farajallah D, Walter L, et al. (2007) Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the *M. silenus* group. Mol Phylogenet Evol 42: 807–816. PMID: [17208017](https://pubmed.ncbi.nlm.nih.gov/17208017/)
27. Goodhead I, Capewell P, Bailey JW, Beament T, Chance M, et al. (2013) Whole-genome sequencing of *Trypanosoma brucei* reveals introgression between subspecies that is associated with virulence. MBio 4: e00197–00113. doi: [10.1128/mBio.00197-13](https://doi.org/10.1128/mBio.00197-13) PMID: [23963174](https://pubmed.ncbi.nlm.nih.gov/23963174/)



28. Stukenbrock EH, Christiansen FB, Hansen TT, Duteil JY, Schierup MH (2012) Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proc Natl Acad Sci U S A* 109: 10954–10959. doi: [10.1073/pnas.1201403109](https://doi.org/10.1073/pnas.1201403109) PMID: [22711811](https://pubmed.ncbi.nlm.nih.gov/22711811/)
29. Ricklefs RE, Outlaw DC, Svensson-Coelho M, Medeiros MC, Ellis VA, et al. (2014) Species formation by host shifting in avian malaria parasites. *Proc Natl Acad Sci U S A* 111: 14816–14821. doi: [10.1073/pnas.1416356111](https://doi.org/10.1073/pnas.1416356111) PMID: [25271324](https://pubmed.ncbi.nlm.nih.gov/25271324/)
30. Schaer J, Perkins SL, Decher J, Leendertz FH, Fahr J, et al. (2013) High diversity of West African bat malaria parasites and a tight link with rodent *Plasmodium* taxa. *Proc Natl Acad Sci U S A* 110: 17415–17419. doi: [10.1073/pnas.1311016110](https://doi.org/10.1073/pnas.1311016110) PMID: [24101466](https://pubmed.ncbi.nlm.nih.gov/24101466/)
31. Nair S, Nkhoma SC, Serre D, Zimmerman PA, Gorena K, et al. (2014) Single-cell genomics for dissection of complex malaria infections. *Genome Res* 24: 1028–1038. doi: [10.1101/gr.168286.113](https://doi.org/10.1101/gr.168286.113) PMID: [24812326](https://pubmed.ncbi.nlm.nih.gov/24812326/)
32. Lapp SA, Korir-Morrison C, Jiang J, Bai Y, Corredor V, et al. (2013) Spleen-dependent regulation of antigenic variation in malaria parasites: *Plasmodium knowlesi* SICAvax expression profiles in splenic and asplenic hosts. *PLoS One* 8: e78014. doi: [10.1371/journal.pone.0078014](https://doi.org/10.1371/journal.pone.0078014) PMID: [24205067](https://pubmed.ncbi.nlm.nih.gov/24205067/)
33. Hamid MM, Remarque EJ, El Hassan IM, Hussain AA, Narum DL, et al. (2011) Malaria infection by sporozoite challenge induces high functional antibody titres against blood stage antigens after a DNA prime, poxvirus boost vaccination strategy in Rhesus macaques. *Malar J* 10: 29. doi: [10.1186/1475-2875-10-29](https://doi.org/10.1186/1475-2875-10-29) PMID: [21303498](https://pubmed.ncbi.nlm.nih.gov/21303498/)
34. Murphy JR, Weiss WR, Fryauff D, Dowler M, Savransky T, et al. (2014) Using infective mosquitoes to challenge monkeys with *Plasmodium knowlesi* in malaria vaccine studies. *Malar J* 13: 215. doi: [10.1186/1475-2875-13-215](https://doi.org/10.1186/1475-2875-13-215) PMID: [24893777](https://pubmed.ncbi.nlm.nih.gov/24893777/)
35. Lim C, Hansen E, DeSimone TM, Moreno Y, Junker K, et al. (2013) Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat Commun* 4: 1638. doi: [10.1038/ncomms2612](https://doi.org/10.1038/ncomms2612) PMID: [23535659](https://pubmed.ncbi.nlm.nih.gov/23535659/)
36. Moon RW, Hall J, Rangkuti F, Ho YS, Almond N, et al. (2013) Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A* 110: 531–536. doi: [10.1073/pnas.1216457110](https://doi.org/10.1073/pnas.1216457110) PMID: [23267069](https://pubmed.ncbi.nlm.nih.gov/23267069/)
37. Daneshvar C, Davis TM, Cox-Singh J, Rafe'e MZ, Zakaria SK, et al. (2009) Clinical and laboratory features of human *Plasmodium knowlesi* infection. *Clin Infect Dis* 49: 852–860. doi: [10.1086/605439](https://doi.org/10.1086/605439) PMID: [19635025](https://pubmed.ncbi.nlm.nih.gov/19635025/)
38. Foster D, Cox-Singh J, Mohamad DS, Krishna S, Chin PP, et al. (2014) Evaluation of three rapid diagnostic tests for the detection of human infections with *Plasmodium knowlesi*. *Malar J* 13: 60. doi: [10.1186/1475-2875-13-60](https://doi.org/10.1186/1475-2875-13-60) PMID: [24548805](https://pubmed.ncbi.nlm.nih.gov/24548805/)
39. Mudunuri SB, Nagarajaram HA (2007) IMEx: Imperfect Microsatellite Extractor. *Bioinformatics* 23: 1181–1187. PMID: [17379689](https://pubmed.ncbi.nlm.nih.gov/17379689/)
40. Kolpakov R, Bana G, Kucherov G (2003) mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res* 31: 3672–3678. PMID: [12824391](https://pubmed.ncbi.nlm.nih.gov/12824391/)
41. Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol Ecol Resources* 8: 92–94. doi: [10.1111/j.1471-8286.2007.01884.x](https://doi.org/10.1111/j.1471-8286.2007.01884.x) PMID: [21585724](https://pubmed.ncbi.nlm.nih.gov/21585724/)
42. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, et al. (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* 455: 799–803. doi: [10.1038/nature07306](https://doi.org/10.1038/nature07306) PMID: [18843368](https://pubmed.ncbi.nlm.nih.gov/18843368/)
43. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, et al. (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455: 757–763. doi: [10.1038/nature07327](https://doi.org/10.1038/nature07327) PMID: [18843361](https://pubmed.ncbi.nlm.nih.gov/18843361/)
44. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295.
45. Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resources* 8: 103–106. doi: [10.1111/j.1471-8286.2007.01931.x](https://doi.org/10.1111/j.1471-8286.2007.01931.x) PMID: [21585727](https://pubmed.ncbi.nlm.nih.gov/21585727/)
46. Haubold B, Hudson RR (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* 16: 847–848. PMID: [11108709](https://pubmed.ncbi.nlm.nih.gov/11108709/)
47. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959. PMID: [10835412](https://pubmed.ncbi.nlm.nih.gov/10835412/)
48. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620. PMID: [15969739](https://pubmed.ncbi.nlm.nih.gov/15969739/)
49. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.