ORIGINAL ARTICLE

WILEY | MOLECULAR ECOLOGY

Genome-wide mosaicism in divergence between zoonotic malaria parasite subpopulations with separate sympatric transmission cycles

Paul C. S. Divis[1,2] | Craig W. Duffy[2] | Khamisah A. Kadir[1] | Balbir Singh[1] | David J. Conway[1,2] (iD)

[1]Faculty of Medicine and Health Sciences, Malaria Research Centre, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

[2]Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, London, UK

Correspondence
Paul C. S. Divis, Faculty of Medicine and Health Sciences, Malaria Research Centre, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia.
Email: pcsimon@unimas.my
David J. Conway, Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, London, UK.
Email: david.conway@lshtm.ac.uk

Abstract

Plasmodium knowlesi is a significant cause of human malaria transmitted as a zoonosis from macaque reservoir hosts in South-East Asia. Microsatellite genotyping has indicated that human infections in Malaysian Borneo are an admixture of two highly divergent sympatric parasite subpopulations that are, respectively, associated with long-tailed macaques (Cluster 1) and pig-tailed macaques (Cluster 2). Whole-genome sequences of clinical isolates subsequently confirmed the separate clusters, although fewer of the less common Cluster 2 type were sequenced. Here, to analyse population structure and genomic divergence in subpopulation samples of comparable depth, genome sequences were generated from 21 new clinical infections identified as Cluster 2 by microsatellite analysis, yielding a cumulative sample size for this subpopulation similar to that for Cluster 1. Profound heterogeneity in the level of intercluster divergence was distributed across the genome, with long contiguous chromosomal blocks having high or low divergence. Different mitochondrial genome clades were associated with the two major subpopulations, but limited exchange of haplotypes from one to the other was evident, as was also the case for the maternally inherited apicoplast genome. These findings indicate deep divergence of the two sympatric *P. knowlesi* subpopulations, with introgression likely to have occurred recently. There is no evidence yet of specific adaptation at any introgressed locus, but the recombinant mosaic types offer enhanced diversity on which selection may operate in a currently changing landscape and human environment. Loci responsible for maintaining genetic isolation of the sympatric subpopulations need to be identified in the chromosomal regions showing fixed differences.

KEYWORDS
adaptation, genomic divergence, host-specificity, introgression

1 | INTRODUCTION

The zoonotic malaria parasite *Plasmodium knowlesi* is a significant cause of human malaria in South-East Asia. Although long known as a malaria parasite of long-tailed and pig-tailed macaques that could potentially infect humans (Coatney, Collin, Warren, & Contacos, 1971), the first large focus of human cases was only detected approximately 15 years ago in Malaysian Borneo (Singh et al., 2004). Since then, infections

have been described from throughout Malaysia (Cox-Singh et al., 2008; William et al., 2013; Yusof et al., 2014) and in almost all countries in South-East Asia (Singh & Daneshvar, 2013). Indeed, *P. knowlesi* is now the most common cause of human malaria in Malaysia (Barber, Rajahram, Grigg, William, & Anstey, 2017), with infections capable of reaching very high parasitaemia and sometimes leading to the death of patients (Cox-Singh et al., 2008; Daneshvar et al., 2009; Rajahram et al., 2016; Singh & Daneshvar, 2013; William et al., 2011).

Multilocus microsatellite genotyping analysis of *P. knowlesi* infections revealed that human infections in Malaysian Borneo comprise two major genetic subpopulations that are, respectively, associated with long-tailed and pig-tailed macaque reservoir hosts (Divis et al., 2015), with significant divergence confirmed by whole-genome sequence analyses of parasites in human infections (Assefa et al., 2015). In most areas of Malaysian Borneo, the number of human clinical infections of the parasite subpopulation type associated with long-tailed macaques (Cluster 1) is higher than those having the type associated with pig-tailed macaques (Cluster 2) (Divis et al., 2017). Further analyses of additional samples have subsequently revealed a third divergent subpopulation of *P. knowlesi* (Cluster 3) on the mainland of South-East Asia which includes Peninsular Malaysia (Divis et al., 2017; Yusof et al., 2016). So far, only *P. knowlesi* parasites of Cluster 3 have been studied in infections of laboratory monkeys (Assefa et al., 2015), and one strain of this type has been adapted to efficiently invade human erythrocytes in culture (Lim et al., 2013; Moon et al., 2013). To develop laboratory studies on the other two major zoonotic populations will require establishment of parasite isolates in controlled monkey infections, or ideally into culture with erythrocytes. Analysis of *P. knowlesi* samples from human clinical infections is relatively straightforward, as most of these are not mixed with other species, whereas most natural *P. knowlesi* infections in macaques occur together with other primate malaria parasite species (Lee et al., 2011).

The first large-scale whole-genome sequence analysis of *P. knowlesi* infections contained clinical samples that were mostly of the Cluster 1 type (N = 38), yielding results indicating that this has undergone long-term population growth, with additional evidence of selection on particular loci (Assefa et al., 2015). There were only 10 Cluster 2 type infections sequenced in the study, which limited investigation of the demographic history of that subpopulation, but these were sufficient to indicate that the level of intercluster divergence varied across the genome, some loci having a concentration of apparently fixed differences and others showing more shared polymorphism (Assefa et al., 2015). A separate simultaneous study reported data from another six infections, confirming the divergence between sympatric subpopulations (Pinheiro et al., 2015), but this did not cumulatively give a much deeper sample. In agreement with the initial study (Assefa et al., 2015), a recent secondary analysis of the previously published data confirmed the existence of genomic regions with shared polymorphisms (Diez Benavente et al., 2017), but did not include any new data.

For a more informed comparison of these important zoonotic parasite subpopulations, a much larger sample of Cluster 2 type *P. knowlesi* genome sequences was obtained in this study. Combining the new data with samples sequenced previously (Assefa et al., 2015; Pinheiro

et al., 2015) yielded a total of 34 Cluster 2 genome sequences that enables a more comprehensive analysis of genomic polymorphism and divergence between the subpopulations. This provides new understanding of the genome-wide variation in divergence of these two sympatric *P. knowlesi* subpopulations, essential for understanding their long-term maintenance and potential for future adaptation.

2 | MATERIALS AND METHODS

2.1 | New *P. knowlesi* DNA samples selected for analysis

Venous blood samples were obtained from patients infected with *P. knowlesi* malaria at Kapit Hospital in Sarawak between March and November 2014, after written informed consent from each patient had been obtained. The collection of blood samples was approved by the Medical Research and Ethics Committee of the Malaysian Ministry of Health and by the Ethics Committee of the London School of Hygiene and Tropical Medicine. Leucocytes were removed by allowing 10 ml of blood to pass through a CF11 cellulose column, to enrich for erythrocytes and thereby increase the proportion of parasite compared to host DNA. Genomic DNA was extracted using QIAamp DNA Mini kits (Qiagen, Germany), and all infections were confirmed to contain only *P. knowlesi* by nested PCR assays testing for all locally known malaria parasite species (Lee et al., 2011). Determination of the genetic subpopulation cluster of each DNA sample was conducted by microsatellite genotyping (Divis et al., 2017), and 21 samples of the Cluster 2 type that had sufficient DNA were selected for whole-genome sequencing. These were mostly single genotype infections as determined by microsatellite typing (Divis et al., 2017).

2.2 | *P. knowlesi* whole-genome sequencing

DNA libraries were constructed using the TruSeq Nano DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Physical shearing of the genomic DNA into fragments having an average size of 550 bp was performed using a M220 Focused-ultrasonicator (Covaris, USA). After denaturation at 95°C for 3 min, amplification of genomic DNA was performed with low number of PCR cycles (eight cycles at 98°C for 20 s, 60°C for 15 s and 72°C for 30 s) followed by a 72°C completion for 5 min. The quality of DNA libraries was assessed using the Agilent High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA USA), while quantitation was performed using the KAPA Library Quantification Kit for Illumina® platform (KAPA Biosystems, Boston, MA, USA). All libraries were then normalized to 4 nM, and up to 12 samples were included on each sequencing run. Paired-end whole-genome sequencing was performed on pooled DNA libraries using MiSeq Chemistry version 3 reagents, on the MiSeq platform (Illumina, San Diego, CA, USA) with a read length of 300 bp. Raw data of short reads generated in FASTQ format were undergone for quality check using the TRIMMOMATIC software (Bolger, Lohse, & Usadel, 2014) with defined parameters (LEADING:3 TRAILING:3, SLIDINGWINDOW:4:10 MINLEN:36).

Trimmed FASTQ reads for individual isolates were then aligned against the version 2.0 of *P. knowlesi* strain H reference genome (www.genedb.org/Homepage/Pknowlesi, genome annotation March 2014, accessed December 2015) using the BURROWS-WHEELER ALIGNER software version 0.7 with the BWA-MEM algorithm and default parameters (Li, 2013). This generated file in the SAM (sequence alignment/map) format, and followed by the conversion into a BAM (binary alignment/map) format using the SAMTOOLS package version 0.1 (Li et al., 2009). Due to the possible effect of PCR amplification bias introduced during the DNA library preparations, read duplications were removed using the *"MarkDuplicates"* command from the Picard toolkit (https://github.com/broadinstitute/picard). The average depth coverage was analysed by the BEDTOOLS version 2 package using the "genomeCoverageBed" command (Quinlan & Hall, 2010).

Re-mapping of short read genome sequences generated from previous studies (Assefa et al., 2015; Pinheiro et al., 2015) against the version 2.0 of *P. knowlesi* strain H reference genome was also performed in the analysis (Table S1). These include 48 isolates from Kapit and Betong in Malaysian Borneo (Sequence Read Archive numbers ERR985372–ERR985419) representing Cluster 1 and Cluster 2 type parasites collected between 2008 and 2013, six isolates from Sarikei in Malaysian Borneo (SRA numbers ERR274221, ERR274222, ERR274224, ERR272225, ERR366425 and ERR366426) and five laboratory isolates ("Nuri" SRA numbers ERR019406, "Hackeri" SRR2221468, "Malayan" SRR2225467, "MR4-H" SRR2225571 and "Philippines" SRR2225573). The reference H strain sequence belongs to Cluster 3 (Assefa et al., 2015), which is approximately equally divergent from Clusters 1 and 2, so no bias is expected in the efficiency of mapping of the sequences to this reference.

2.3 | Single nucleotide polymorphism calling and filtration

The calling of high-quality single nucleotide polymorphisms (SNPs) was performed using several steps, following procedures described previously (Assefa et al., 2015). For each isolate, SNPs were first identified from the BAM file using SAMTOOLS/BCFTOOLS with the following parameters: *mpileup −B −Q 23 −d 2000 −C 50 −ugf; varFilter −d 10 -D 2000*. A high-quality list of potential variant positions (Phred quality, $Q > 30$) was extracted from the resulting variant call format (VCF) file, and a list of unique SNP lists was generated by concatenating all variant positions from all isolates. Using these unique SNP positions, the mapping quality (mq) and base quality (bq) were checked for each isolate to remove positions with an excess of low-quality reads with the requirement of the minimum read depth coverage at 10x. The ratio of read depth values at high-quality (mq = 26; bq = 23) and low-quality (mq = 0; bq = 0) thresholds were calculated for each isolate using customized Perl scripts, and any SNP positions with the ratio below 0.5 were discarded.

Further filtration involved the removal of positions that contained ambiguous sequences (represented as a long stretch of unknown nucleotides "N") in the reference genome. The *SICAVar*, *KIR*, and *pk-fam-a* to *pk-fam-e* multigene families (Pain et al., 2008) and the

subtelomeric regions were also filtered out to avoid ambiguous alignments, which may cause false-positive SNP calls. Subtelomeric regions were here determined by visually inspecting the whole-genome synteny mapping of *P. knowlesi* with the *P. vivax* homolog using the PlasmoDB GBrowse v2.48 (plasmodb.org/cg-bin/gbrowse/plasmodb/), with the boundaries of subtelomeric regions defined as sequences adjacent to the first conserved protein-coding gene (Table S2). After exclusion of subtelomeric regions and the large multigene families, 21.2 Mb (92%) of the 23.0 Mb corresponding to the reference nuclear genome was analysed from each sample.

2.4 | Genomic diversity and population structure

To measure the amount of polymorphism within the parasite population, the average pairwise nucleotide diversity (π) among the sequences from the individual infection samples was calculated. The skewness in allele frequency distributions was estimated by Tajima's D index. Both indices were calculated using the same genome-wide SNP data set in nonoverlapping window sizes of 10 kb and performed using the DIVSTAT software (Soares, Moleirinho, Oliveira, & Amorim, 2015). To illustrate the population substructure, the matrix of pairwise DNA distance among individuals was calculated and the Neighbour-Joining tree was constructed using the APE package version 3.4 in the R environment (Paradis, Claude, & Strimmer, 2004). An independent population structure evaluation was also conducted using principal coordinate analysis (PCoA) with SNPs having no missing data, using the APE package.

To estimate the divergence between the subpopulations, the genome-wide distribution of the fixation index (F_{ST}) between the two-subpopulation clusters was computed with SNPs having minor allele frequencies (MAFs) above 0.1, and above 0.3, using customized R functions. An elevated F_{ST} threshold was set at the 90th percentile of the F_{ST} distributions for all SNPs. Average F_{ST} values were calculated in windows of 500 SNPs with sliding by 250 SNPs. The F_{ST} values for each window were tested for high- or low-differentiated regions against the genome-wide mean F_{ST} value.

Genomic regions with contrasting levels of intercluster divergence were determined empirically by examining the F_{ST} distribution across the genome at two different MAFs (MAF above 0.1 and 0.3). For each MAF analysis, average F_{ST} values were calculated in windows of 200 SNPs (sliding by 100 SNPs), 500 SNPs (sliding by 250 SNPs) and 1,000 SNPs (sliding by 500 SNPs). Mean global F_{ST} values and window F_{ST} values were then converted into standard z-scores in order to standardize the definition of outlier windows for different parameters. Regions of high- or low-F_{ST} windows were observed and compared among the analyses that used different MAF parameters.

Genomic regions were categorized into low divergence regions (LDR; z-scores < −0.5), intermediate divergence regions (IDR), and high divergence regions (HDR; z-scores > 0.5). To determine the contiguous extent of these regions in detail, adjacent outlier windows were merged to form larger adjoining regions. Peak and trough patterns of window z-scores around the thresholds (z-scores < −0.5 and z-scores > 0.5) were taken into consideration in determining the junctions. Each candidate region was demarcated by first and last

SNPs that fell within the merged windows, except for HDRs where SNPs with elevated F_{ST} values were used as start and endpoints.

Patterns of polymorphisms (nucleotide diversity summarized by π and allele frequency spectrum summarized by Tajima's D) in all genomic regions were evaluated using DIVSTAT software. Test runs were performed in nonoverlapping window sizes of 10 kb for each subpopulation. Nonparametric Kruskal–Wallis tests were used to compare among the genomic regions as well as against the genome-wide background.

2.5 | Extra-chromosomal genomes

Population structure and relationships of the sympatric *P. knowlesi* subpopulations were further analysed using the extranuclear DNA, consisting of the nonrecombining genomes of mitochondria and plastid-like apicoplast. The 5.9-kb mitochondrial DNA sequences were obtained from the present whole-genome sequence data and previously published sequences (Assefa et al., 2015; Jongwutiwes et al., 2005; Lee et al., 2011; Pinheiro et al., 2015). Complete mitochondrial sequences were obtained from GenBank database, consisting of 26 haplotypes from human isolates (Accession nos. EU880446–EU880470) and 20 haplotypes from macaque isolates (EU880471–EU880474, EU880477–EU880486, EU880489–EU880493 and EU880499) in Kapit of Malaysian Borneo, and one human isolate from Thailand (AY598141). Three species, *P. coatneyi* (AB354575), *P. cynomolgi* (AB434919) and *P. vivax* (AY791551), that have close evolutionary relationships with *P. knowlesi* were included in the analysis as out-groups. For the apicoplast genome of *P. knowlesi*, 30.6 kb of the DNA sequences that had clear alignment was extracted from the present whole-genome data set as well as from previous data (Assefa et al., 2015; Pinheiro et al., 2015) following mapping and base quality checks as mentioned above.

The derived mitochondrial and apicoplast genome sequences were separately aligned using the CLUSTALX programme version 2 (Larkin et al., 2007), following which nucleotide diversity (π) and haplotype diversity (Hd) was determined using the DNASP version 5 software (Librado & Rozas, 2009). A maximum-likelihood tree was inferred with 1,000 bootstrap replicates and gaps treated as missing data using the PHANGORN packages in R (Schliep, 2011), with the ModelTest algorithm used to determine the best-fit nucleotide substitution model, which was GTR+I+G (General Time Reversible model with a proportion of invariable sites and gamma distribution). For the mitochondrial sequences, major haplotypes were determined with gaps treated as missing data, and the statistical parsimony haplotype network was constructed using the TCS version 1.21 software (Clement, Posada, & Crandall, 2000).

3 | RESULTS

3.1 | Generation of new whole-genome sequences and SNP genotyping

Paired-end Illumina sequencing of 21 new *P. knowlesi* clinical infection samples, selected on the basis of microsatellite genotyping as belonging to Cluster 2 (the type previously associated with pig-tailed macaque as well as human infections), yielded a mean of 6.95 million high-quality reads per sample, which were mapped against the *P. knowlesi* H strain version 2.0 reference genome sequence (Table S3). The mean depth of sequence coverage genome-wide was 52.3-fold (range from 28.7- to 80.3-fold) per sample. In addition, Illumina short read sequence data from another 59 *P. knowlesi* isolates obtained previously (Assefa et al., 2015; Pinheiro et al., 2015) were remapped against the *P. knowlesi* H strain version 2.0 reference genome using the same assembly parameters (Table S1), followed by SNP calling. In the combined data set of 80 infection sequences, a total of 2,109,937 SNPs were identified in the nuclear genome. Following exclusion of those in subtelomeric regions or in the *KIR* or *SICAVAR* multigene families, or that had more than two alleles, 1,669,533 SNPs remained, of which 1,186,073 high-quality SNPs with less than 10% missing calls in all isolates were used for population genomic analyses.

3.2 | Population genetic structure

Consistent with predictions from cluster assignment based on microsatellite genotyping, all 21 of the new *P. knowlesi* clinical infection samples showed genome sequences belonging to the Cluster 2 subpopulation (Figures 1a and S1). Together with previous data, this yielded an overall sample of 34 Cluster 2 isolate sequences, to achieve a similar sample size as previously available for Cluster 1. As is visually apparent from the Neighbour-Joining tree based on the pairwise genetic distances (Figure 1a), the Cluster 2 infections are less genetically diverse (π = 3.43×10^{-3}) than the Cluster 1 infections (π = 5.78×10^{-3}). Furthermore, the Cluster 1 subpopulation demonstrated a homogenous pattern of sequence diversity across the 14 chromosomes (Kruskal–Wallis, p = .23), in contrast with Cluster 2 that showed heterogeneous levels of diversity across the chromosomes (Kruskal–Wallis $p < 10^{-16}$) (Figure S2). In Cluster 2, nucleotide diversity of entire chromosomes ranged from 2.25×10^{-3} (for chromosome 7) to 4.38×10^{-3} (for chromosome 5), but all had a lower diversity than in Cluster 1 (Wilcoxon signed rank $p < 10^{-16}$). In a majority of nonoverlapping 10-kb windows genome-wide, nucleotide diversity (π) indices were lower in Cluster 2 (Figure 1b). Large regions of chromosomes showed contiguous stretches in which diversity was much higher in Cluster 1, and also contiguous stretches in which the diversity was more similar (Figure 1c).

3.3 | Genomic regions of high and low divergence

The genome-wide variation in diversity in Cluster 2 suggested that there might be variation in levels of intercluster divergence. Analysing SNPs with overall minor allele frequencies above 10% (193,068 SNPs), the mean genome-wide fixation index indicated substantial divergence between the two subpopulations (mean $F_{ST} = 0.25$; Figure 2a). The frequency distribution of F_{ST} values was bimodal, one peak having values just above zero and a
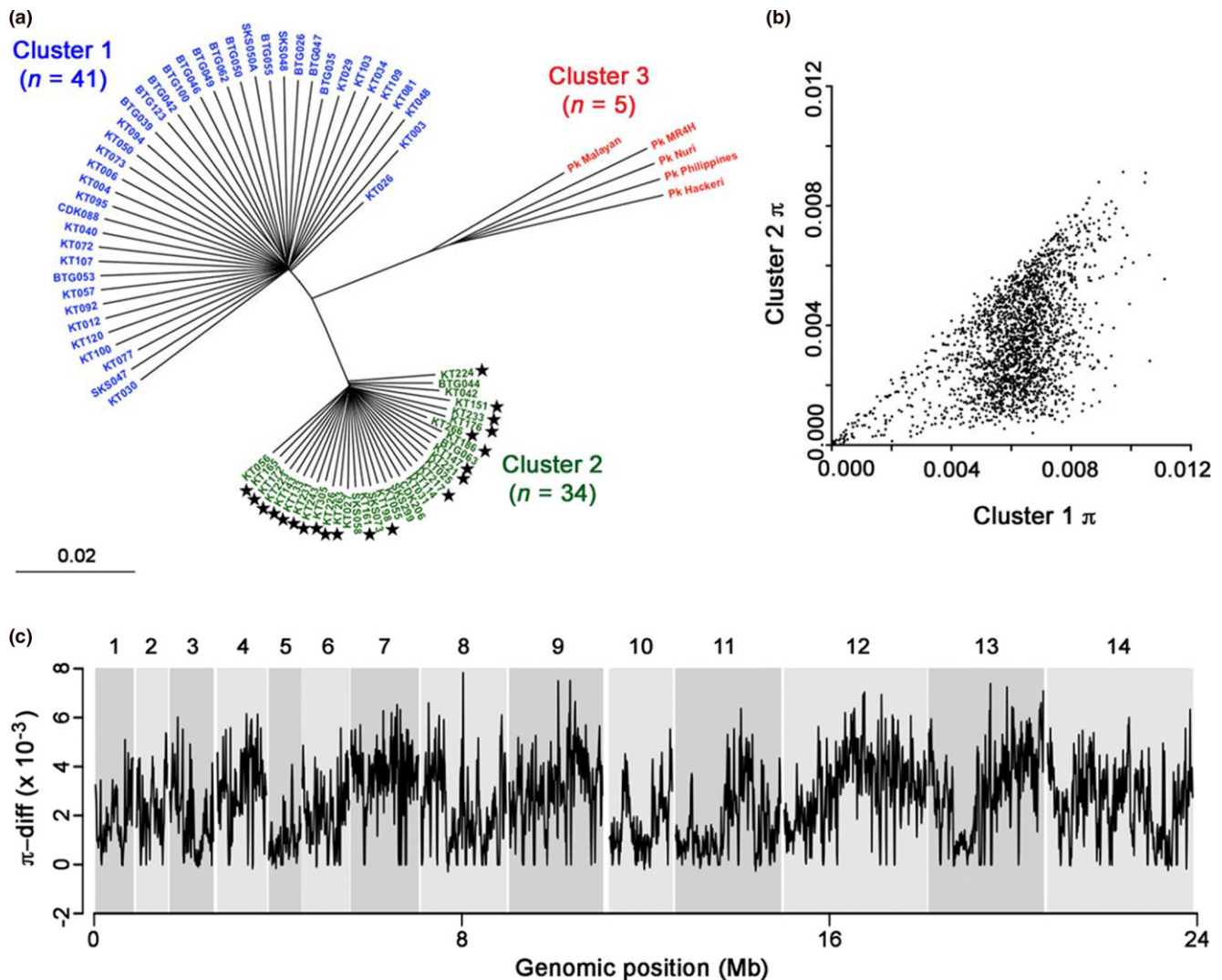
FIGURE 1 Population structure of *Plasmodium knowlesi* indicated by whole-genome sequence data. (a) Neighbour-Joining tree based on a pairwise single nucleotide polymorphism (SNP) difference matrix of 80 *P. knowlesi* isolates. The 21 new genome sequences are indicated with stars, yielding a total sample size for Cluster 2 (*N* = 34) that is similar to that of Cluster 1 (*N* = 41). The scale bar indicates proportions of all SNPs differing between samples. (b) Scatterplot of nucleotide diversity (π) in individual nonoverlapping 10-kb windows of the genome, comparing data for Cluster 1 and Cluster 2 subpopulations. (c) Differences in nucleotide diversity between Cluster 1 and Cluster 2 subpopulations (π-diff) in each of the 10-kb windows of the genome [Colour figure can be viewed at wileyonlinelibrary.com]

second peak having values at or approaching 1.0 (Figure 2b). Very high intercluster F_{ST} values of > 0.8 were seen for 19,116 SNPs, and 7,415 (3.8%) showed complete fixation of alternative alleles ($F_{ST} = 1.0$). A large proportion of low F_{ST} values were removed when analysis focused on SNPs with overall allele frequencies of >0.3 (Figure 2b). Mean F_{ST} values for whole chromosomes ranged from 0.09 (for chromosome 5) to 0.40 for (chromosome 7).

The relative level of population differentiation of all windows of 500 contiguous SNPs across the genome was evaluated by considering standard deviations from the mean genome-wide F_{ST} value (z-score). Genomic regions were identified that contained contiguous windows defining low divergence regions (LDR with z-score < −0.5) and high divergence regions (HDR with z-score > 0.5). This revealed

large genomic blocks of high or low divergence (Figure 2c; Table S4). For example, chromosomes 7, 12 and 13 had HDRs covering most of their respective lengths, whereas chromosomes 3, 5 and 10 showed no HDRs (Figure 2c).

3.4 | Intracluster diversity in genomic regions with contrasting levels of divergence

The relationship of intercluster divergence with the varying nucleotide diversity (π) in Cluster 2 across the genome (Figure 1c) was investigated. Comparing between the two subpopulations, the differences in nucleotide diversity were higher in the HDRs than in the LDRs or in the rest of the genome (Figure 3; Mann–Whitney *U* $p < 10^{-16}$ for both comparisons). Most of the highly differentiated
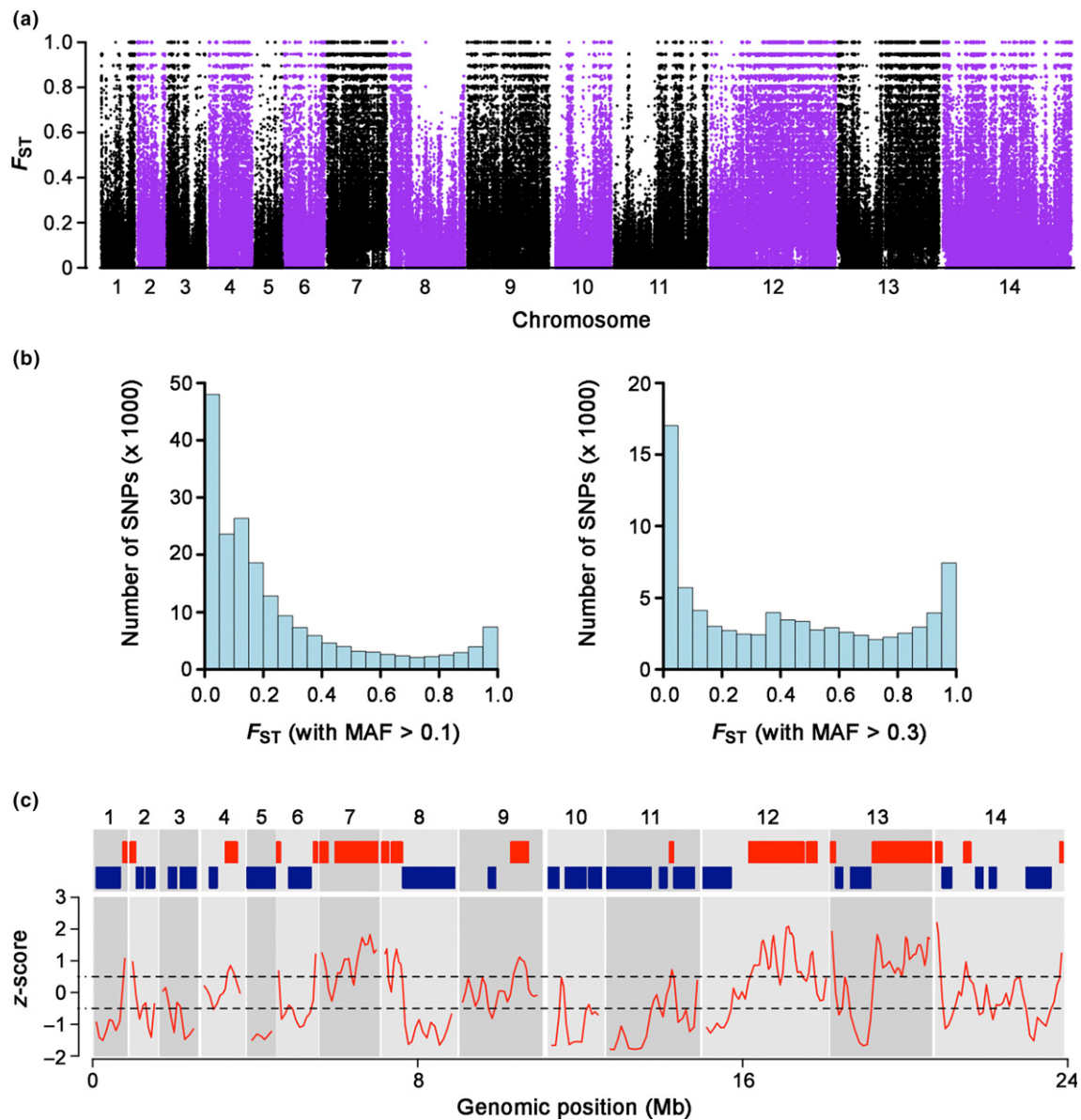
FIGURE 2 Genome-wide plot of divergence between the sympatric *Plasmodium knowlesi* Cluster 1 and Cluster 2 subpopulations in Malaysian Borneo. Each dot shows the F_{ST} value of an individual single nucleotide polymorphism (SNP), of 193,068 SNPs with minor allele frequencies above 0.1. The overall genome-wide mean F_{ST} value is 0.25. (b) Strong bimodal frequency distribution of F_{ST} values for SNPs genome-wide. The left plot shows the distribution of values for 193,068 SNPs with minor allele frequencies (MAF) > 0.1, and the right plot shows the distribution of values for 80,168 SNPs with MAF > 0.3 (the genome-wide average F_{ST} value was 0.42 for SNPs with MAF > 0.3). (c) Contiguous regions of high and low divergence throughout the genome identified by analysis of F_{ST} values of windows of 500 consecutive SNPs converted to standardized z-scores. Thresholds of 0.5 standard deviations above and below the genome-wide average F_{ST} demarcate the regions of high divergence (red blocks) and low divergence (dark blue blocks) [Colour figure can be viewed at wileyonlinelibrary.com]

regions were those in which nucleotide diversity was substantially lower in Cluster 2 (Figure 3).

Reduced nucleotide diversity in HDRs compared to the rest of the genome was specifically seen in Cluster 2 (mean π in HDRs = 2.08×10^{-3}; Mann–Whitney $p < 2.2 \times 10^{-16}$), and not in Cluster 1 (mean π in HDRs = 5.80×10^{-3}; Mann–Whitney $p = 0.25$). Similarly, higher nucleotide diversity in LDRs compared to the rest of the genome was seen specifically within Cluster 2 (Mann–Whitney $p = 2.2 \times 10^{-16}$), and not in Cluster 1 (Mann–Whitney $p = .77$).

Both subpopulations showed strong skew towards low-frequency variants, with mean Tajima's D values of 10-kb windows of the genome for the Cluster 2 subpopulation being even lower than for the Cluster 1 subpopulation (Figure 4a; Cluster 1 mean $D = -1.77$; Cluster 2 mean $D = -2.37$; Wilcoxon Signed Rank $p < 10^{-16}$). Across all 10-kb windows in the genome, there was a weak but highly significant correlation in the distribution of Tajima's D values in the two clusters (Figure 4b; Spearman's $\rho = 0.25$; $p < 10^{-16}$). The allele frequency spectrum as summarized by Tajima's D index was less variable across the 14 chromosomes within the Cluster 1 subpopulation
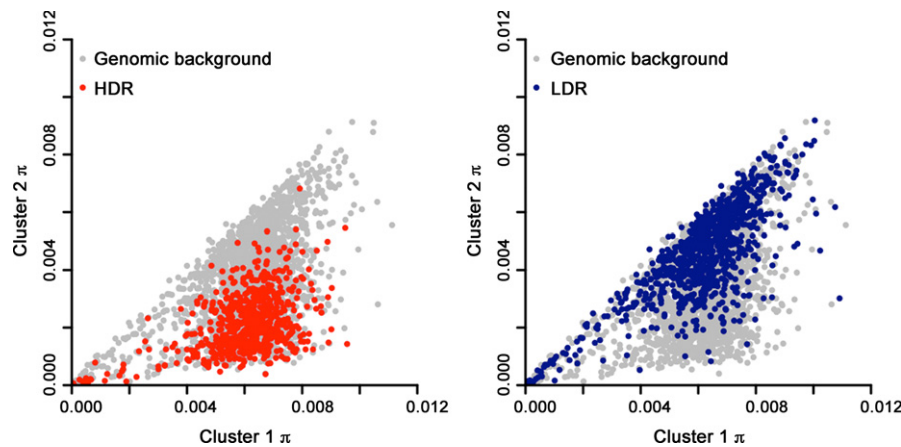
FIGURE 3 Sequence diversity of *Plasmodium knowlesi* Cluster 2 is lowest in regions of the genome that have highest fixation indices in comparison with Cluster 1. Scatterplots show nucleotide diversity in discrete 10-kb windows genome-wide, with red points (left) showing windows in high divergence regions (HDR) and blue points showing windows in low divergence regions (LDR). For HDR, mean $\pi = 5.80 \times 10^{-3}$ for Cluster 1 and 2.08×10^{-3} for Cluster 2. For LDR, mean $\pi = 5.60 \times 10^{-3}$ for Cluster 1 and 4.14×10^{-3} for Cluster 2 [Colour figure can be viewed at wileyonlinelibrary.com]

(Kruskal–Wallis $p = 8.4 \times 10^{-5}$) compared to the Cluster 2 subpopulation (Kruskal–Wallis $p = 1.6 \times 10^{-16}$) (Figure 4c).

The mosaic pattern of genomic diversity in the Cluster 2 subpopulation suggests that a genome-wide scan to identify individual genes with exceptionally high values of Tajima's *D* may not be a robust means of identifying genes under balancing selection within this subpopulation, although the approach may be more straightforwardly applied to the Cluster 1 subpopulation (Assefa et al., 2015). However, the *msp1* merozoite surface protein antigen gene that was previously shown to have a high Tajima's *D* value in Cluster 1 also had a high value in the Cluster 2 subpopulation (*D* = 1.01), suggesting it is likely to be under balancing selection in both. Interestingly, the *ama1* apical membrane antigen gene that did not have a high value in Cluster 1 had an exceptionally high value in Cluster 2 here (*D* = 1.64). The *csp* circumsporozoite protein gene, that had the highest Tajima's *D* value of all genes in Cluster 1, did not have any detected nonrepeat sequence SNPs in Cluster 2. Thus, although an unbiased comparison cannot be straightforwardly performed, these examples indicate that there are some similarities as well as differences in the strength or targets of balancing selection on antigens in the two different parasite subpopulations.

3.5 | Phylogeny and introgression of extra-chromosomal genomes

The analyses of population structure were extended using the maternally inherited extra-chromosomal genomes. Combination of the 5.9-kb mitochondrial sequences generated in this study with previously published sequences yielded a sample size of 129 in total and identification of 77 SNPs. These mitochondrial sequences had a global average nucleotide diversity (π) of 7.9×10^{-4}, with higher values in samples from parasites in Cluster 1 ($\pi = 6.8 \times 10^{-4}$, *n* = 74) than in Cluster 2 ($\pi = 4.9 \times 10^{-4}$, *n* = 46). The genealogical network of mitochondrial genomes contained 56 different haplotypes (Figure 5). The

most common and central core haplotype was detected mainly in parasites of the Cluster 1 subpopulation (25 of 28 isolates). A second common haplotype that was more peripheral in the network was seen mostly in the Cluster 2 subpopulation (15 of 21 isolates), while the third common haplotype was distantly related to this and detected only in Cluster 1 (nine isolates). Most of the closely related haplotypes to each of these were also seen only in the corresponding subpopulation clusters, but there is a group of closely related haplotypes internal in the network seen in parasites of Cluster 1 (13 isolates) which is embedded in part of the network that is otherwise only seen in Cluster 2 parasites (Figure 5). Conversely, a few Cluster 2 isolates have haplotypes that are related to those only seen in Cluster 1. A separate branch of haplotypes was seen in laboratory isolates that had mostly been collected from Peninsular Malaysia. Maximum-likelihood phylogenetic analysis yielded a similar pattern, with haplotype clades being associated but not completely fixed between the Cluster 1 and Cluster 2 subpopulations (Figure S3).

Polymorphism in 30.6 kb of the apicoplast genome could be characterized using the Illumina short read sequence data to identify 520 polymorphic SNPs. With these data, 65 of the 80 isolates were analysed in detail as they had less than 20% missing SNPs, while the remaining 15 samples with more missing SNP data were excluded. The overall nucleotide diversity (π) was 1.79×10^{-3}, and this was higher among the Cluster 1 samples ($\pi = 1.77 \times 10^{-3}$) than Cluster 2 samples ($\pi = 1.12 \times 10^{-3}$). Two major lineages were seen, one of which consisted predominantly of Cluster 1 samples, and the other mainly of Cluster 2 samples (Figure S4), although there were several isolates that had haplotypes of the opposite type to that expected for each cluster.

4 | DISCUSSION

This study analyses the largest ecological sample of sequences representing different subpopulations of a zoonotic eukaryotic parasite
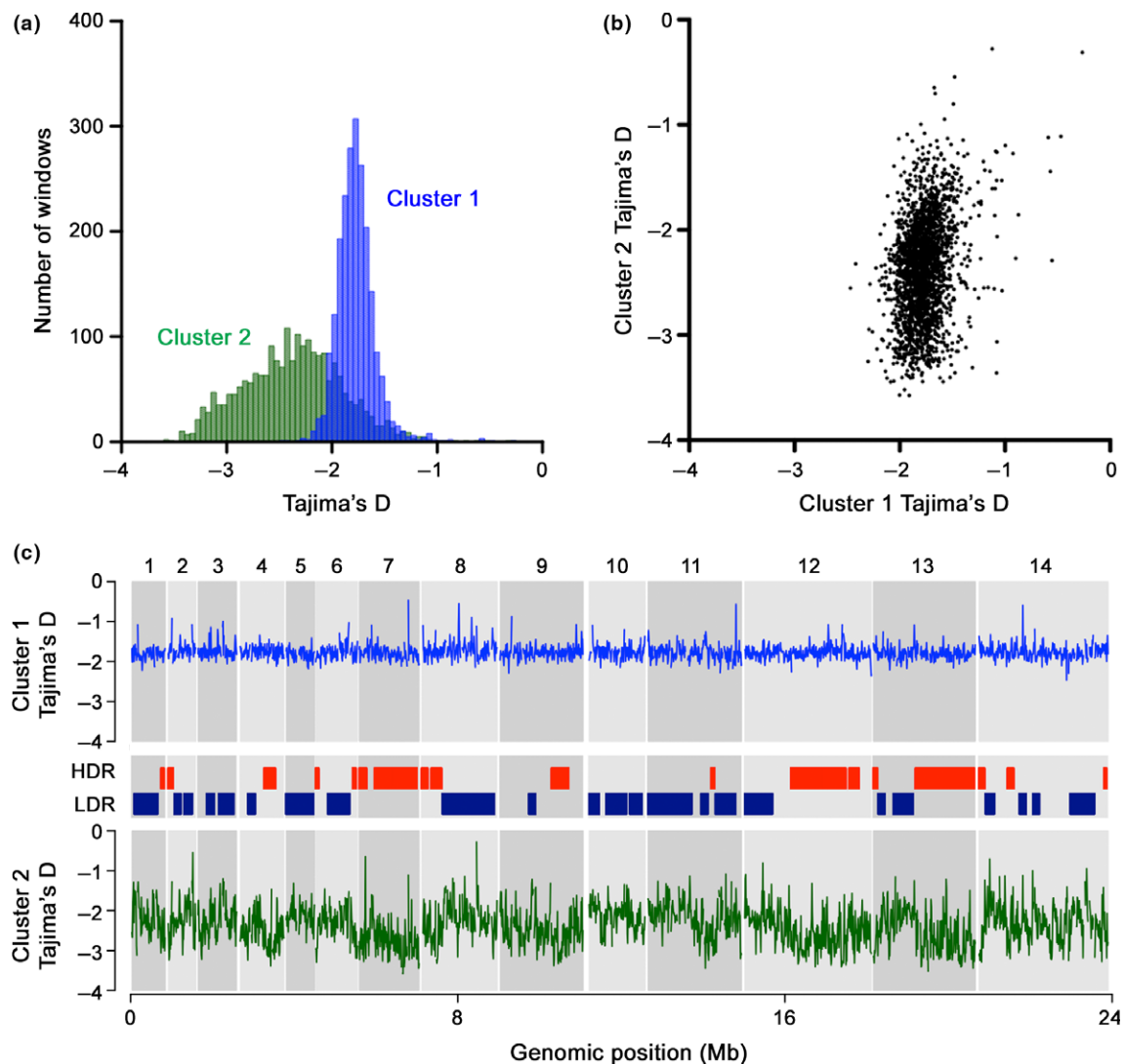
FIGURE 4 Comparison of genome-wide Tajima's *D* distributions between the two major *Plasmodium knowlesi* genetic subpopulations in Malaysian Borneo. (a) Frequency distribution of Tajima's *D* values in nonoverlapping 10-kb windows for Cluster 2 shows more negatively skewed values compared to Cluster 1. (b) Tajima's *D* values for individual 10-kb windows show a weak correlation between the two subpopulations (Spearman's ρ = 0.25), although this is highly significant ($p < 10^{-16}$). (c) Distribution of Tajima's *D* values in nonoverlapping 10-kb windows across all 14 chromosomes presented in alternate dark and light grey blocks. The mean genome-wide value for Cluster 1 is −1.77 and for Cluster 2 is −2.37 [Colour figure can be viewed at wileyonlinelibrary.com]

species. Whole-genome sequencing of new samples from one of the major genetic subpopulations of *P. knowlesi* has clearly revealed the genome-wide patterns of divergence between the sympatric subpopulations, which illuminates aspects of their population history and is essential for understanding their adaptive potential. This provides the most informative overall analysis of population structure of *P. knowlesi* to date, extending the understanding of defined subpopulation clusters that were previously described (Assefa et al., 2015; Divis et al., 2017). These results confirm the distinctness of the two sympatric divergent *P. knowlesi* subpopulations in Malaysian Borneo, supporting the occurrence of independent zoonotic cycles associated with different macaque reservoir host species (Divis et al., 2015; Muehlenbein et al., 2015).

The high differentiation between these two sympatric subpopulations indicates minimal or no ongoing gene flow occurring between

them, and a large number of SNPs showed complete fixation of alternative alleles. However, the pattern of divergence was heterogeneous and bimodally distributed, with large regions of exceptionally high or low divergence interspersed throughout the genome. Reduced genetic diversity of the Cluster 2 subpopulation in highly diverged regions suggests there may have been an initial bottleneck in the formation of this subpopulation. The overall allele frequency spectra were negatively skewed for both subpopulations, signifying long-term population growth, although this was more extreme for the Cluster 2 subpopulation. This gives a more detailed perspective than that previously obtained by analysis of mitochondrial genome sequences, which had already indicated a historical population expansion (Lee et al., 2011). The mitochondrial and apicoplast genomes in *Plasmodium* are inherited together through the female parasite gamete in each transmission cycle (Lim & McFadden, 2010) with
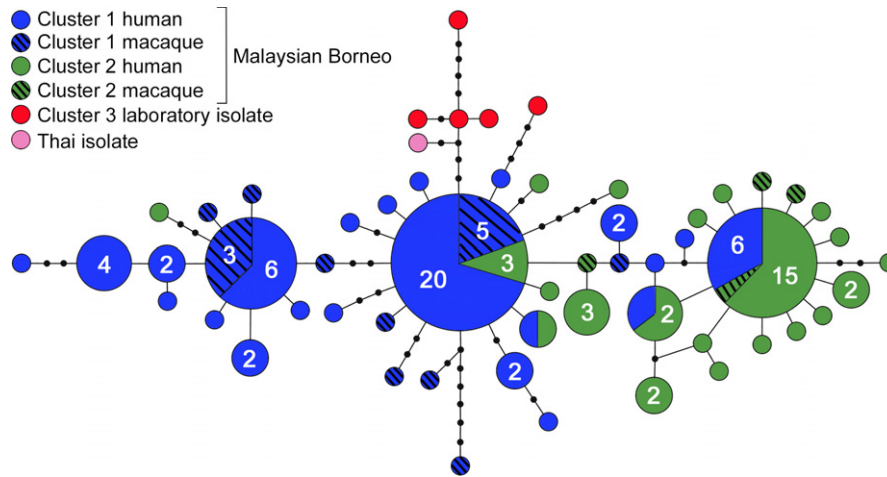
FIGURE 5 Genealogical network based on 129 *Plasmodium knowlesi* mitochondrial DNA genome sequences showing 56 different haplotypes. Sizes of the circles represent relative numbers of samples with each haplotype, with numbers specified where this is more than one. Connecting lines each represent one mutational step, and black dots represent missing intermediate haplotypes [Colour figure can be viewed at wileyonlinelibrary.com]

negligible recombination at the population level, but analyses of these extra-chromosomal genomes here indicates some sharing of different haplotypes between the *P. knowlesi* subpopulations. The mosaic pattern with adjacent large regions of alternating high and low diversity in the genome sequences of the Cluster 2 subpopulation, in contrast to the more consistent high diversity throughout the genome for the Cluster 1 subpopulation, suggests that introgression has probably occurred recently from Cluster 1 into the Cluster 2 population.

Despite the differences at the genomic level, it is not yet known whether these two major sympatric subpopulations exhibit significant phenotypic differences, apart from the previously described association with different macaque reservoir host species (Divis et al., 2015, 2017; Lee et al., 2011). Human *P. knowlesi* infections have been associated with a wide spectrum of disease (Cox-Singh et al., 2010; Daneshvar et al., 2009; Rajahram et al., 2012; William et al., 2011), and there is recent evidence that asymptomatic infections may be more common than previously expected (Fornace et al., 2015; Lubis et al., 2017; Siner et al., 2017), so conducting detailed clinical studies on individuals infected with each parasite subpopulation type is now a priority.

A recent study suggests a link between local deforestation and incidence of *P. knowlesi* infections in an area of Sabah state within Malaysian Borneo (Fornace et al., 2016). Of relevance to this, long-tailed macaques and pig-tailed macaques show different habitat ranges in forested and nonforested areas (Moyes et al., 2016), suggesting that there may be micro foci of infection for each subpopulation cluster, and highlighting the need to examine changes over time. It is clear that future research should include monitoring the proportions of the different *P. knowlesi* subpopulations over time, and potential changes in their genetic composition. Sequencing of *P. knowlesi* genomes from natural macaque infections would be more challenging, given that these are usually

coinfections together with other primate malaria parasite species (Lee et al., 2011), although new methods of sequencing genomes from single parasites could be adapted to address the issue (Trevino et al., 2017). This would ideally be done alongside sampling of infections in local mosquito vector species that could potentially be maintaining the separate zoonotic transmission cycles.

The genome-wide mosaicism, showing bimodal levels of divergence as well as limited discordant occurrence of extra-chromosomal genome lineages, indicate that introgression is likely to have occurred recently between these parasite subpopulations. The recombinant genomes that are now circulating offer a great diversity on which selection may operate, but there is no evidence yet of specific adaptation at introgressed loci. A recent re-analysis of previously published data identified a common shared haplotype in a chromosomal region with low divergence between the subpopulations (Diez Benavente et al., 2017), although an observation that the region had a slightly higher than background proportion of genes predicted to be expressed at a particular developmental stage may not be relevant, as an extended haplotype may result from selection on a single locus rather than on multiple genes.

In contrast, it is likely that at least one of the chromosomal regions showing fixed differences between the clusters contains a locus responsible for maintaining genetic isolation of the sympatric subpopulations, potentially due to transmission in different mosquito vectors, as well as likely adaptation to the different reservoir macaque hosts. Parasites from these sympatric subpopulations have not yet been studied in laboratory infections or adapted to culture, which will be necessary to define phenotypes and enable experimental analyses of differences between them. Despite major technical challenges of such work, efforts should prove worthwhile, as they are likely to reveal parasite phenotypes not present in the old laboratory lines which were sampled from a different part of the parasite species range (Dankwa et al., 2016; Moon

et al., 2016). If there are no parasite subpopulation-specific barriers to infection of mosquito vectors that may be experimentally used, such as *Anopheles cracens* (Amir, Sum, Lau, Vythilingam, & Fong, 2013), it may ultimately be possible to map loci controlling key phenotypes by performing genetic crosses between parental parasites representing the different subpopulations.

DATA ACCESSIBILITY

Paired-end short read genome sequence data for the new parasite infection isolates listed in Table S3 have been deposited in the European Nucleotide Archive, Accession nos. ERS2037781-ERS2037801.

AUTHOR CONTRIBUTIONS

P.C.S.D., B.S. and D.J.C. conceived and designed the study. K.A.K. and B.S. collected and prepared the samples. P.C.S.D. conducted the genome sequencing, bioinformatic S.N.P. calling and nucleotide data deposition. P.C.S.D., C.W.D. and D.J.C. performed data analysis and interpretation. P.C.S.D. and D.J.C. wrote the manuscript, with input from all authors.

ORCID

David J. Conway http://orcid.org/0000-0002-8711-3037

REFERENCES

Amir, A., Sum, J. S., Lau, Y. L., Vythilingam, I., & Fong, M. Y. (2013). Colonization of *Anopheles cracens*: A malaria vector of emerging importance. *Parasites & Vectors*, 6, 81. https://doi.org/10.1186/1756-3305-6-81

Assefa, S., Lim, C., Preston, M. D., Duffy, C. W., Nair, M. B., Adroub, S. A., . . . Singh, B. (2015). Population genomic structure and adaptation in the zoonotic malaria parasite *Plasmodium knowlesi*. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 13027–13032. https://doi.org/10.1073/pnas.1509534112

Barber, B. E., Rajahram, G. S., Grigg, M. J., William, T., & Anstey, N. M. (2017). World malaria report: Time to acknowledge *Plasmodium knowlesi* malaria. *Malaria Journal*, 16, 135. https://doi.org/10.1186/s12936-017-1787-y

Bolger, A. M., Lohse, M., & Usadel, B. (2014). TRIMMOMATIC: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Coatney, G. R., Collin, W. E., Warren, M., & Contacos, P. G. (1971). *The primate malarias*. Washington, DC: U.S. Government Printing Office.

Cox-Singh, J., Davis, T. M., Lee, K. S., Shamsul, S. S., Matusop, A., Ratnam, S., . . . Singh, B. (2008). *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clinical Infectious Diseases*, 46, 165–171. https://doi.org/10.1086/524888

Cox-Singh, J., Hiu, J., Lucas, S. B., Divis, P. C., Zulkarnaen, M., Chandran, P., . . . Krishna, S. (2010). Severe malaria – A case of fatal *Plasmodium knowlesi* infection with post-mortem findings: A case report. *Malaria Journal*, 9, 10. https://doi.org/10.1186/1475-2875-9-10

Daneshvar, C., Davis, T. M., Cox-Singh, J., Rafa'ee, M. Z., Zakaria, S. K., Divis, P. C., & Singh, B. (2009). Clinical and laboratory features of human *Plasmodium knowlesi* infection. *Clinical Infectious Diseases*, 49, 852–860. https://doi.org/10.1086/605439

Dankwa, S., Lim, C., Bei, A. K., Jiang, R. H., Abshire, J. R., Patel, S. D., . . . Duraisingh, M. T. (2016). Ancient human sialic acid variant restricts an emerging zoonotic malaria parasite. *Nature Communications*, 7, 11187. https://doi.org/10.1038/ncomms11187

Diez Benavente, E., de Sessions, P. F., Moon, R. W., Holder, A. A., Blackman, M. J., Roper, C., . . . Clark, T. G. (2017). Analysis of nuclear and organellar genomes of *Plasmodium knowlesi* in humans reveals ancient population structure and recent recombination among host-specific subpopulations. *PLoS Genetics*, 13, e1007008. https://doi.org/10.1371/journal.pgen.1007008

Divis, P. C., Lin, L. C., Rovie-Ryan, J. J., Kadir, K. A., Anderios, F., Hisam, S., . . . Conway, D. J. (2017). Three divergent subpopulations of the malaria parasite *Plasmodium knowlesi*. *Emerging Infectious Diseases*, 23, 616–624. https://doi.org/10.3201/eid2304.161738

Divis, P. C., Singh, B., Anderios, F., Hisam, S., Matusop, A., Kocken, C. H., . . . Conway, D. J. (2015). Admixture in humans of two divergent *Plasmodium knowlesi* populations associated with different macaque host species. *PLoS Pathogens*, 11, e1004888. https://doi.org/10.1371/journal.ppat.1004888

Fornace, K. M., Abidin, T. R., Alexander, N., Brock, P., Grigg, M. J., Murphy, A., . . . Cox, J. (2016). Association between landscape factors and spatial patterns of *Plasmodium knowlesi* infections in Sabah, Malaysia. *Emerging Infectious Diseases*, 22, 201–208. https://doi.org/10.3201/eid2202.150656

Fornace, K. M., Nuin, N. A., Betson, M., Grigg, M. J., William, T., Anstey, N. M., . . . Drakeley, C. J. (2015). Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia. *Journal of Infectious Diseases*, 213, 784–787.

Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Ferreira, M. U., Kanbara, H., & Hughes, A. L. (2005). Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. *Molecular Biology and Evolution*, 22, 1733–1739. https://doi.org/10.1093/molbev/msi168

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). CLUSTAL W and CLUSTAL X version 2.0. *Bioinformatics*, 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lee, K. S., Divis, P. C., Zakaria, S. K., Matusop, A., Julin, R. A., Conway, D. J., . . . Singh, B. (2011). *Plasmodium knowlesi*: Reservoir hosts and

tracking the emergence in humans and macaques. *PLoS Pathogens*, 7, e1002015. https://doi.org/10.1371/journal.ppat.1002015

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXivorg*, arXiv:1303.3997v1302 [q-bio.GN].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMTOOLS. *Bioinformatics*, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Librado, P., & Rozas, J. (2009). DNASP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lim, C., Hansen, E., DeSimone, T. M., Moreno, Y., Junker, K., Bei, A., … Duraisingh, M. T. (2013). Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nature Communications*, 4, 1638. https://doi.org/10.1038/ncomms2612

Lim, L., & McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 365, 749–763. https://doi.org/10.1098/rstb.2009.0273

Lubis, I. N., Wijaya, H., Lubis, M., Lubis, C. P., Divis, P., Beshir, K. B., & Sutherland, C. J. (2017). Contribution of *Plasmodium knowlesi* to multi-species human malaria infections in North Sumatera, Indonesia. *Journal of Infectious Diseases*, 215(7), 1148–1155. https://doi.org/10.1093/infdis/jix091

Moon, R. W., Hall, J., Rangkuti, F., Ho, Y. S., Almond, N., Mitchell, G. H., … Blackman, M. J. (2013). Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 531–536. https://doi.org/10.1073/pnas.1216457110

Moon, R. W., Sharaf, H., Hastings, C. H., Ho, Y. S., Nair, M. B., Rchiad, Z., … Holder, A. A. (2016). Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7231–7236. https://doi.org/10.1073/pnas.1522469113

Moyes, C. L., Shearer, F. M., Huang, Z., Wiebe, A., Gibson, H. S., Nijman, V., … Hy, S. I. (2016). Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas. *Parasites & Vectors*, 9, 242. https://doi.org/10.1186/s13071-016-1527-0

Muehlenbein, M. P., Pacheco, M. A., Taylor, J. E., Prall, S. P., Ambu, L., Nathan, S., … Escalante, A. A. (2015). Accelerated diversification of nonhuman primate malarias in southeast Asia: Adaptive radiation or geographic speciation? *Molecular Biology and Evolution*, 32, 422–439. https://doi.org/10.1093/molbev/msu310

Pain, A., Böhme, U., Berry, A. E., Mungall, K., Finn, R. D., Jackson, A. P., … Berriman, M. (2008). The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*, 455, 799–803. https://doi.org/10.1038/nature07306

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Pinheiro, M. M., Ahmed, M. A., Millar, S. B., Sanderson, T., Otto, T. D., Lu, W. C., … Cox-Singh, J. (2015). *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. *PLoS ONE*, 10, e0121303. https://doi.org/10.1371/journal.pone.0121303

Quinlan, A. R., & Hall, I. M. (2010). BEDTOOLS: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rajahram, G. S., Barber, B. E., William, T., Grigg, M. J., Menon, J., Yeo, T. W., & Anstey, N. M. (2016). Falling *Plasmodium knowlesi* malaria death rate among adults despite rising incidence, Sabah, Malaysia, 2010-2014. *Emerging Infectious Diseases*, 22, 41–48.

Rajahram, G. S., Barber, B. E., William, T., Menon, J., Anstey, N. M., & Yeo, T. W. (2012). Deaths due to *Plasmodium knowlesi* malaria in Sabah, Malaysia: Association with reporting as *Plasmodium malariae* and delayed parenteral artesunate. *Malaria Journal*, 11, 284. https://doi.org/10.1186/1475-2875-11-284

Schliep, K. P. (2011). PHANGORN: Phylogenetic analysis in R. *Bioinformatics*, 27, 592–593. https://doi.org/10.1093/bioinformatics/btq706

Siner, A., Liew, S. T., Kadir, K. A., Mohamad, D. S. A., Thomas, F. K., Zulkarnaen, M., & Singh, B. (2017). Absence of *Plasmodium inui* and *Plasmodium cynomolgi*, but detection of *Plasmodium knowlesi* and *Plasmodium vivax* infections in asymptomatic humans in the Betong division of Sarawak, Malaysian Borneo. *Malaria Journal*, 16, 417. https://doi.org/10.1186/s12936-017-2064-9

Singh, B., & Daneshvar, C. (2013). Human infections and detection of *Plasmodium knowlesi*. *Clinical Microbiology Reviews*, 26, 165–184. https://doi.org/10.1128/CMR.00079-12

Singh, B., Sung, L. K., Matusop, A., Radhakrishnan, A., Shamsul, S. S., Cox-Singh, J., … Conway, D. J. (2004). A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet*, 363, 1017–1024. https://doi.org/10.1016/S0140-6736(04)15836-4

Soares, I., Moleirinho, A., Oliveira, G. N., & Amorim, A. (2015). DIVSTAT: A user-friendly tool for single nucleotide polymorphism analysis of genomic diversity. *PLoS ONE*, 10, e0119851. https://doi.org/10.1371/journal.pone.0119851

Trevino, S. G., Nkhoma, S. C., Nair, S., Daniel, B. J., Moncada, K., Khoswe, S., … Cheeseman, I. H. (2017). High-resolution single-cell sequencing of malaria parasites. *Genome Biology and Evolution*, 9(12), 3373–3383. https://doi.org/10.1093/gbe/evx256

William, T., Menon, J., Rajahram, G., Chan, L., Ma, G., Donaldson, S., … Yeo, T. W. (2011). Severe *Plasmodium knowlesi* malaria in a tertiary care hospital, Sabah, Malaysia. *Emerging Infectious Diseases*, 17, 1248–1255. https://doi.org/10.3201/eid1707.101017

William, T., Rahman, H. A., Jelip, J., Ibrahim, M. Y., Menon, J., Grigg, M. J., … Barber, B. E. (2013). Increasing incidence of *Plasmodium knowlesi* malaria following control of *P. falciparum* and *P. vivax* malaria in Sabah, Malaysia. *PLoS Neglected Tropical Diseases*, 7, e2026. https://doi.org/10.1371/journal.pntd.0002026

Yusof, R., Ahmed, M. A., Jelip, J., Ngian, H. U., Mustakim, S., Hussin, H. M., … Lau, Y. L. (2016). Phylogeographic evidence for 2 genetically distinct zoonotic *Plasmodium knowlesi* parasites, Malaysia. *Emerging Infectious Diseases*, 22, 1371–1380. https://doi.org/10.3201/eid2208.151885

Yusof, R., Lau, Y. L., Mahmud, R., Fong, M. Y., Jelip, J., Ngian, H. U., … Ali, M. M. (2014). High proportion of knowlesi malaria in recent malaria cases in Malaysia. *Malaria Journal*, 13, 168. https://doi.org/10.1186/1475-2875-13-168

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Divis PCS, Duffy CW, Kadir KA, Singh B, Conway DJ. Genome-wide mosaicism in divergence between zoonotic malaria parasite subpopulations with separate sympatric transmission cycles. *Mol Ecol*. 2018;27:860–870. https://doi.org/10.1111/mec.14477