



OPEN

## Population genetic analysis of *Plasmodium knowlesi* reveals differential selection and exchange events between Borneo and Peninsular sub-populations

Anna Turkiewicz<sup>1</sup>, Emilia Manko<sup>1</sup>, Damiola R. Oresgun<sup>2</sup>, Debbie Nolder<sup>1,3</sup>, Anton Spadar<sup>1</sup>, Colin J. Sutherland<sup>1,3</sup>, Janet Cox-Singh<sup>2</sup>, Robert W. Moon<sup>1</sup>, Yee-Ling Lau<sup>4</sup>, Susana Campino<sup>1✉</sup> & Taane G. Clark<sup>1,5✉</sup>

The zoonotic *Plasmodium knowlesi* parasite is a growing public health concern in Southeast Asia, especially in Malaysia, where elimination of *P. falciparum* and *P. vivax* malaria has been the focus of control efforts. Understanding of the genetic diversity of *P. knowlesi* parasites can provide insights into its evolution, population structure, diagnostics, transmission dynamics, and the emergence of drug resistance. Previous work has revealed that *P. knowlesi* fall into three main sub-populations distinguished by a combination of geographical location and macaque host (*Macaca fascicularis* and *M. nemestrina*). It has been shown that Malaysian Borneo groups display profound heterogeneity with long regions of high or low divergence resulting in mosaic patterns between sub-populations, with some evidence of chromosomal-segment exchanges. However, the genetic structure of non-Borneo sub-populations is less clear. By gathering one of the largest collections of *P. knowlesi* whole-genome sequencing data, we studied structural genomic changes across sub-populations, with the analysis revealing differences in Borneo clusters linked to mosquito-related stages of the parasite cycle, in contrast to differences in host-related stages for the Peninsular group. Our work identifies new genetic exchange events, including introgressions between Malaysian Peninsular and *M. nemestrina*-associated clusters on various chromosomes, including in parasite invasion genes (*DBPβ*, *NBPXα* and *NBPXβ*), and important proteins expressed in the vertebrate parasite stages. Recombination events appear to have occurred between the Peninsular and *M. fascicularis*-associated groups, including in the *DBPβ* and *DBPγ* invasion associated genes. Overall, our work finds that genetic exchange events have occurred among the recognised contemporary groups of *P. knowlesi* parasites during their evolutionary history, leading to apparent mosaicism between these sub-populations. These findings generate new hypotheses relevant to parasite evolutionary biology and *P. knowlesi* epidemiology, which can inform malaria control approaches to containing the impact of zoonotic malaria on human communities.

*Plasmodium knowlesi* is a zoonotic malaria parasite commonly residing in long-tailed (*Macaca fascicularis*) and pig-tailed (*M. nemestrina*) macaques and monkeys (*Presbytis melalophos*). The parasite is also recognized as a significant cause of human malaria, with cases described across most countries in the Southeast Asia and Western Pacific regions<sup>1</sup>. There is especially high burden in Malaysia, where *P. knowlesi* is the remaining malaria species with increasing numbers of cases reported<sup>2,3</sup>. Severe disease develops in 19% of clinical presentations<sup>4</sup>, while

<sup>1</sup>Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. <sup>2</sup>Division of Infection, School of Medicine, University of St Andrews, St Andrews, UK. <sup>3</sup>UK Health Security Agency Malaria Reference Laboratory, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. <sup>4</sup>Universiti Malaya, Kuala Lumpur, Malaysia. <sup>5</sup>Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK. ✉email: susana.campino@lshtm.ac.uk; taane.clark@lshtm.ac.uk

fatalities have been observed in 0.3–1.8% of reported cases in Malaysia<sup>2</sup>. Serological assessments of exposure to *P. knowlesi* in endemic Borneo (Kudat) found evidence of previous infection in 7% of the population<sup>5</sup>.

The vectors accountable for *P. knowlesi* transmission in Malaysia are members of the *Anopheles Leucosphyrus* group, wherein *An. latens* and *An. balbacensis* are primarily reported in Borneo Island, although recent studies have also confirmed the presence of *An. dondali*<sup>6</sup>. Evidence also suggests that *An. collessi* and *An. roperi*, members of the *Anopheles Umbrosus* group (endemic to Southeast Asia and India), infected with *P. knowlesi* are circulating in previously under-studied areas of Sarawak<sup>7</sup>. In Peninsular Malaysia, *An. hackeri* and *An. cracens* are known vectors<sup>8</sup>.

Deforestation and environmental changes associated with the rapid human population growth and development are thought to underlie *P. knowlesi* becoming the predominant cause of human malaria in Malaysia<sup>9</sup>. Malaria elimination strategies across the Southeast Asia region have focused primarily on *P. falciparum* and *P. vivax*, but as shown in South America<sup>10</sup>, neglected *P. vivax* persists long beyond *P. falciparum* decline and elimination. Reducing *P. knowlesi* malaria cases in human communities is hindered by the challenges to conventional control methods, including increased contact between people, mosquito vectors and *P. knowlesi* main primate reservoirs, influencing population dynamics and behaviour of both hosts and vectors<sup>11</sup>.

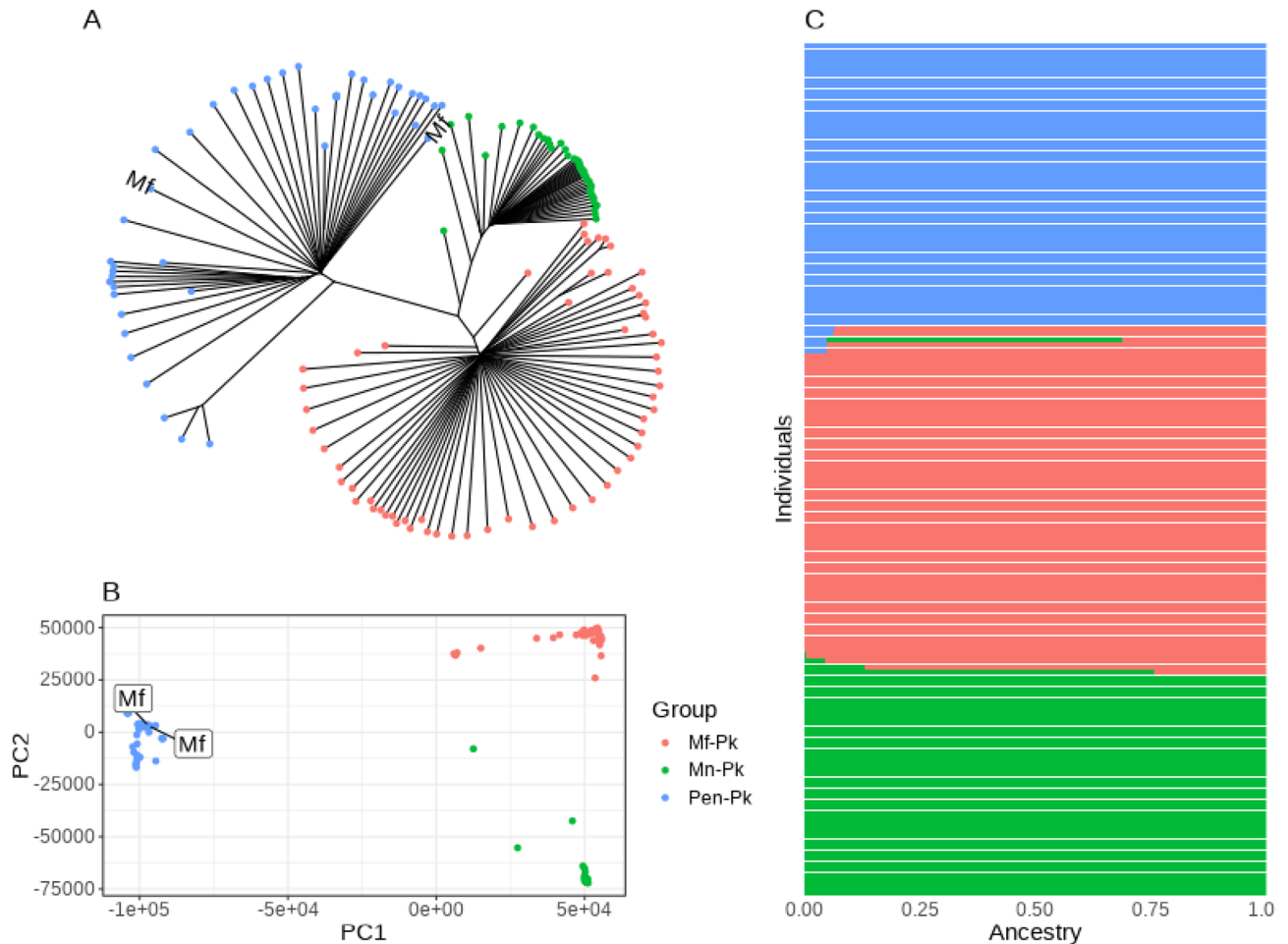
The adaptation of *P. knowlesi* to environmental changes, which in turn affect abundance of both insect and primate hosts, may be observed through analysis of genetic diversity and population structure. Understanding the genomic diversity of *P. knowlesi* using whole-genome sequencing (WGS) data can help reveal transmission patterns and inform disease control. It is now possible to sequence *P. knowlesi* DNA sourced from low parasitaemia infections using a parasite selective whole genome amplification approach<sup>12</sup>. Analysis of WGS data has identified three main clusters determined by location as being either Peninsular Malaysia (Pen-Pk) or Borneo, where in the latter, two groups may be distinguished by association with the respective host *Macaca* species (Mf-Pk - *M. fascicularis*; Mn-Pk - *M. nemestrina*)<sup>13,14</sup>. Genomic dimorphism among the groups have been revealed<sup>15</sup>. Moreover, previous analysis has provided evidence of genetic links between the Borneo populations, and identified chromosomal-segment exchanges spread throughout one of the chromosomes<sup>14</sup>. Utilising the largest collection of *P. knowlesi* WGS data enriched with laboratory samples cultured in *M. fascicularis* blood, we performed a series of population genetic analyses to identify recent evolutionary events and expand the current knowledge of parasite population structure. Evidence of new genetic exchange events between Peninsular (Pen-Pk) and Mn-associated (Mn-Pk) clusters were identified, and novel changes among Borneo groups have been revealed. These discoveries can contribute to an improved understanding of the mosaic genomic architecture observed across all three sub-populations.

## Results

**High levels of genomic variation and three sub-populations.** The analysed data consists of 151 isolates, primarily derived from human infections (n = 149) including field and laboratory samples, but also 2 isolates collected from cultured *M. fascicularis*. Across the dataset, there were 1,883,700 high quality genome-wide SNPs ( $\approx 1$  every 13 bp) identified, including 95 SNPs on the mitochondrial and 640 SNPs on the apicoplast organelle genomes. Within-infection genomic diversity revealed 124 (82.1%) isolates with mono-infections (Mf-Pk, n = 48; Mn-Pk, n = 31; Pen-Pk, n = 45) (S1 Figure), especially prevalent amongst human infections and, as expected, in laboratory strains. A co-infection with *P. vivax* was identified. The principal component analysis (PCA) and neighbour-joining tree reinforced the presence of three clusters, and all newly sequenced isolates (n = 25) fell within one of these groups (Fig. 1A, B). In particular, the macaque and human laboratory blood cultured strains segregated within the Peninsular cluster (Pen-Pk, n = 13). Whereas remaining newly sequenced isolates were identified as belonging to one of the two Borneo groups (Mf-Pk, n = 7; Mn-Pk, n = 5), including the single isolate from Indonesia (PK3), which was linked to the Mf-Pk cluster.

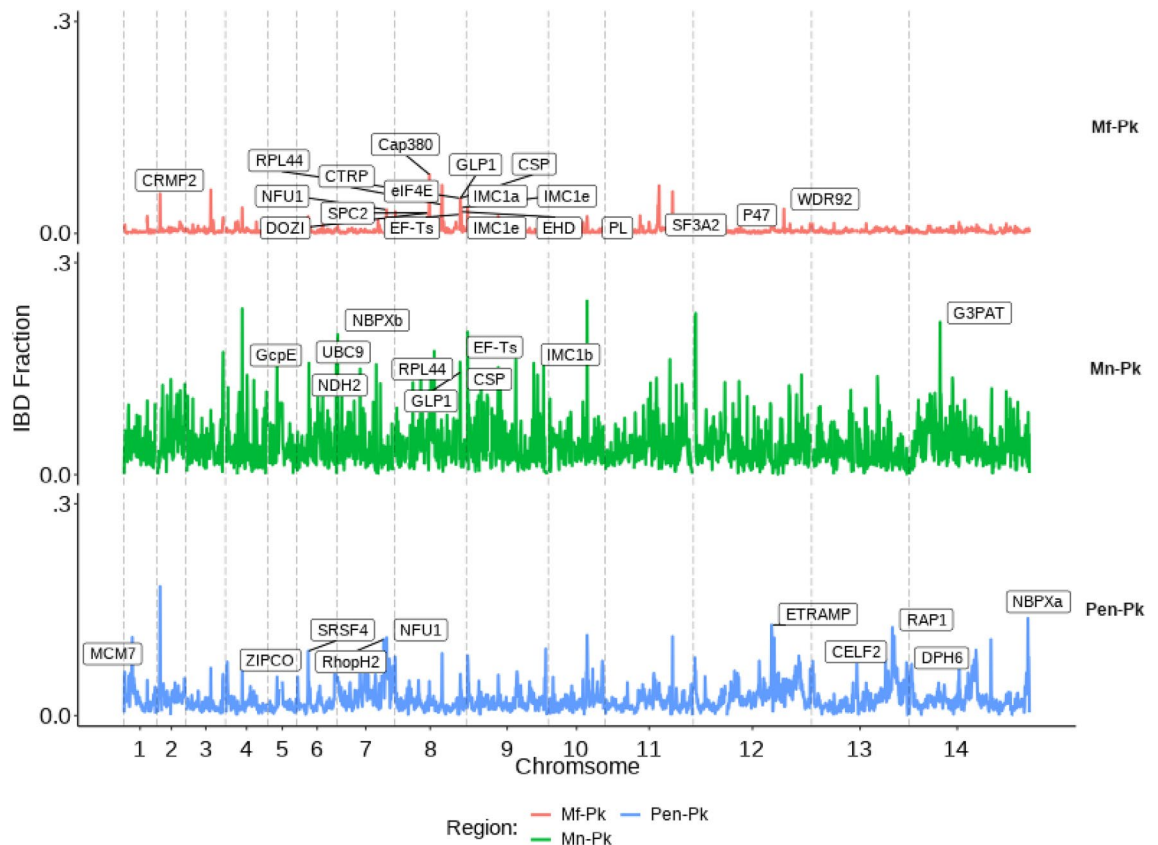
Estimation of individual ancestry using ADMIXTURE software revealed the existence of three ancestral backgrounds (Fig. 1C), consistent with the three PCA-based clusters (Mf-Pk, Mn-Pk, Pen-Pk). Most isolates could be assigned to a single ancestral group, except for three samples isolated from Sarikei, which displayed ancestry from both Borneo clusters. Most isolates were genetically distinct from one another, except for the newly derived human blood-cultured isolates, which were closely related, together with three publicly available clinical isolates (SRR3135172, SRR2222335, SRR2225467). Single representatives from the highly similar groups and all remaining isolates provided a total of 104 genomes available for subsequent population genetics analysis (Mf-Pk, n = 43; Mn-Pk, n = 30; Pen-Pk, n = 31). The overall nucleotide diversity for the Peninsular cluster ( $\pi = 6.60 \times 10^{-3}$ ) and Mf-associated group ( $\pi = 5.95 \times 10^{-3}$ ) were similar, while the Mn-associated samples had relatively lower genetic diversity ( $\pi = 3.48 \times 10^{-3}$ ).

**Identity by descent analysis reveals differences between Borneo and Peninsular Malaysia.** Identity by descent (IBD) analysis was performed for each of the studied groups (Mf-Pk, Mn-Pk, Pen-Pk) to understand the chromosome-level structure and shared ancestry of *P. knowlesi* sub-populations. High IBD values indicate sites of common descent within a group. The fractions of pairwise IBD in the Mn-associated cluster were found to be the greatest (median: 0.087), suggestive of high levels of relatedness among isolates (S2 Figure). In contrast, the Mf-Pk group presented significantly lower pairwise fractions (median: 0.007), while values for Pen-Pk isolates were in-between (median: 0.026) (S2 Figure, S2 Table). Genome-wide IBD fractions, summarised within 10kbp sliding windows, confirmed the differences between Macaque-associated clusters, with the Mn-Pk group being more strongly related compared to the widely divergent Mf-Pk samples (Fig. 2, S2 Table). In Mf-associated groups, several fragments revealed high IBD fraction values, including a region on chromosome 8 (940–950 kbp) containing the Cap380 putative protein (PKNH\_0820800), which is essential for oocyst capsule formation in *P. berghei*<sup>16</sup>. Another region with high IBD was in chromosome 11 (1420–1460 kbp), and



**Figure 1.** Population structure analysis using the whole set of 151 *P. knowlesi* isolates. (A) Neighbour-joining tree, with leaves coloured based on grouping; (B) Principal Component Analysis (PCA) plot, and samples collected from *M. fascicularis* have been marked; (C) ADMIXTURE barplot with predictions for each isolate reflecting the relative abundance of three ancestral sub-populations ( $K=3$ ), with the color scheme the same as (A).

contained the ATP synthase-associated protein, PIMMS43, acyl-CoA synthetase and F-box (PKNH\_1130100 to PKNH\_1131100). In *P. berghei*, the ATP synthase-associated (PKNH\_1130200) protein is essential for parasite transformation from ookinete to further mosquito stages<sup>17</sup>, while PIMMS43 proteins (PKNH\_1130300 and PKNH\_1130900) are required for sporogonic development in the oocyst and production of sporozoites able to infect<sup>18</sup>. The conservation of vector-related genes like PIMMS43 has been recently observed in a rodent malaria parasite<sup>19</sup>. Furthermore, in the cluster, we identified other putative proteins that play an essential role in mosquito-related stages of the parasite cycle including CRMP2, CTRP, IMC1a, and IMC1e<sup>20</sup>. The Mn-associated cluster presents a highly similar structure across the genome, and multiple peaks associated with high IBD values were found, including a region on chromosome 7 (610–620 kbp; PKNH\_0713000 to PKNH\_0713300) containing putative proteins NDH2 and UBC9. The knock-out of NDH2 in *P. berghei* leads to the inability to develop into mature oocysts in the mosquito midgut<sup>21</sup>. While the UBC9 locus is involved in the regulation of the erythrocyte developmental cycle<sup>22</sup>, with its high expression in this stage confirmed in single cell data analysed<sup>23</sup>. A high peak was observed on chromosome 9 (1810–1820 kbp), including putative IMC1b (PKNH\_0939700) found to be an ortholog of a *P. berghei* protein responsible for the mechanical strength and motility of parasite ookinetes<sup>24</sup>. Interestingly, four regions appeared as important in both Borneo clusters (S2 Table). One of the fragments on chromosome 8 (1770–1780 kbp) encompassing five genes (PKNH\_0838400 to PKNH\_0838800), included the circumsporozoite (CSP) protein (PKNH\_0838500), which codes the surface antigen of sporozoites and is expressed in high levels in sporozoites that reached the mosquito salivary gland<sup>25</sup>. The putative 60S ribosomal protein L44 (RPL44 - PKNH\_0838700) was visibly expressed in human blood stages in the analysis of single cell data conducted in *P. knowlesi*<sup>23</sup>. In addition, one of the conserved fragments included the NBPXb gene, which is known to be required for the invasion of host erythrocytes<sup>26</sup> and genetically similar across clusters<sup>12</sup>. Overall, analysis of gene expression patterns for the Borneo group revealed that some of the loci identified in the IBD analysis (69/99) were linked to the blood stages of the *P. knowlesi* life cycle (rings, trophozoite and schizont) (S2 Table). Interestingly, the IBD analysis for the Peninsular cluster (Pen-Pk) demonstrated mostly high peaks containing genes previously linked to vertebrate parasite stages in Plasmodium species, including putative



**Figure 2.** Genome-wide distribution of identity by descent (IBD) fractions across studied populations of *P. knowlesi* isolates, summarised in sliding windows of size 10kbp. Vertical dashed lines represent chromosome boundaries. Genes located in fragments with 1% of highest IBD fraction values have been marked.

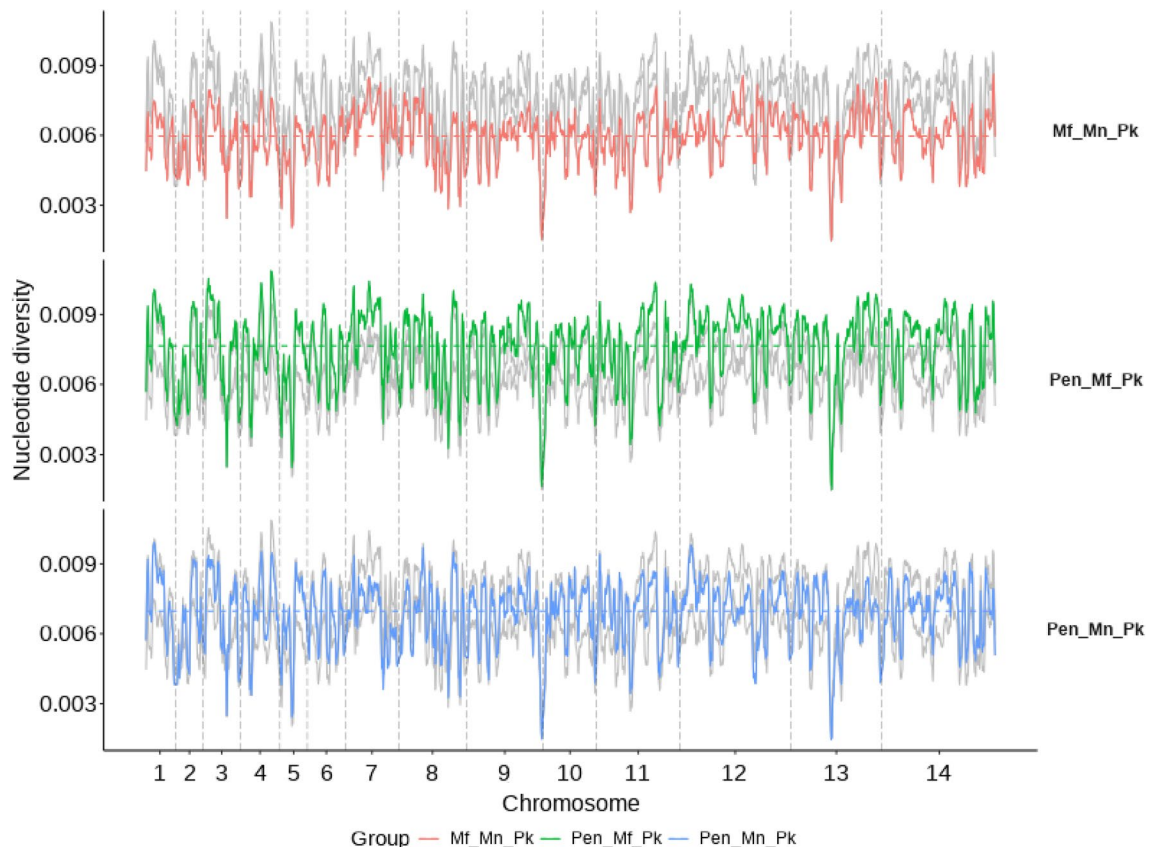
ZIPKO, RhopH2, ETRAMP, RAP1 and NBPXa proteins (PKNH\_0606600, PKNH\_0727900, PKNH\_1246400, PKNH\_1347900 and PKNH\_1472300). In *P. berghei*, the ZIPKO protein is crucial for parasite development inside hepatocytes<sup>27</sup>. Furthermore, in *P. falciparum* RAP1 is required for erythrocyte invasion<sup>28</sup>, while RhopH2 plays a role in the formation of new permeability pathways in infected erythrocytes<sup>29</sup>. The ETRAMP family in *P. falciparum* has been linked to parasite development from ring to trophozoite stages, however expression patterns were not confirmed in the single cell data. Moreover, the NBPXa in *P. knowlesi* is known to be an essential mediator in human erythrocyte invasion<sup>26</sup> and its strong genetic divergence has been reported<sup>12</sup>. In addition, some reported genes displayed signs of high expression in single cell analysis, with exceptionally significant expression found in PKNH\_0623000, PKNH\_0946200 and PKNH\_126300 (elongation factor 1-gamma)<sup>23</sup>. Overall, in general, the high IBD fragments in the Peninsular group contained genes linked with the host-related stages of the parasite cycle.

**Genetic regions underlying high divergence between Peninsular and Borneo clusters.** A genome-wide scan of pairwise nucleotide diversity ( $\pi$ ) was conducted among groups (Fig. 3). Joint nucleotide diversity for the Borneo isolates was the lowest (mean  $\pi$ : 0.0058). The Pen-Pk and Mn-Pk clusters (mean  $\pi$ : 0.0067) had lower values in comparison to the joint nucleotide diversity of Pen-Pk and Mf-Pk (mean: 0.0074). The lowest nucleotide diversity occurred genome-wide in the Borneo clusters (Mf-Pk and Mn-Pk), with identical fragments found on chromosomes 5, 8, 10, 11 and 13 (Fig. 3). Chromosome 8 has been linked with chromosomal-segment exchange events between Borneo populations<sup>14</sup>. The joint nucleotide diversity for the Peninsular and Mn-associated clusters decreased slightly in regions on chromosomes 7, 12 and 13, potentially contributing to the mosaic pattern. Joint nucleotide diversity was greatest when the Mf-Pk and Pen-Pk groups were compared, with no regions displaying significant similarities. Genes presenting low nucleotide diversity across pairwise comparisons and at the same time divergence to the remaining sub-population were found (Table 1). Unsurprisingly, most of these genes ( $n = 50$ , 63.3%) were associated with Borneo clusters and were spread along almost all chromosomes (Table 1). In contrast, the Peninsular and Mn-associated sub-populations revealed a significant number of highly similar genes ( $n = 27$ ), especially on chromosome 7, which carries more than half of the loci identified. The third comparison (Pen-Pk vs. Mf-Pk) indicated less divergence between these two groups, as only 2 genes displayed significant similarity (Table 1).

To further investigate the genes involved in the high genetic divergence between *P. knowlesi* sub-populations, differences in allele frequencies were assessed using the fixation index (Fst). Mean genome-wide Fst values were greatest between Pen-Pk and Mn-associated clusters (mean Fst = 0.322), followed by Pen-Pk versus Mf-Pk (mean

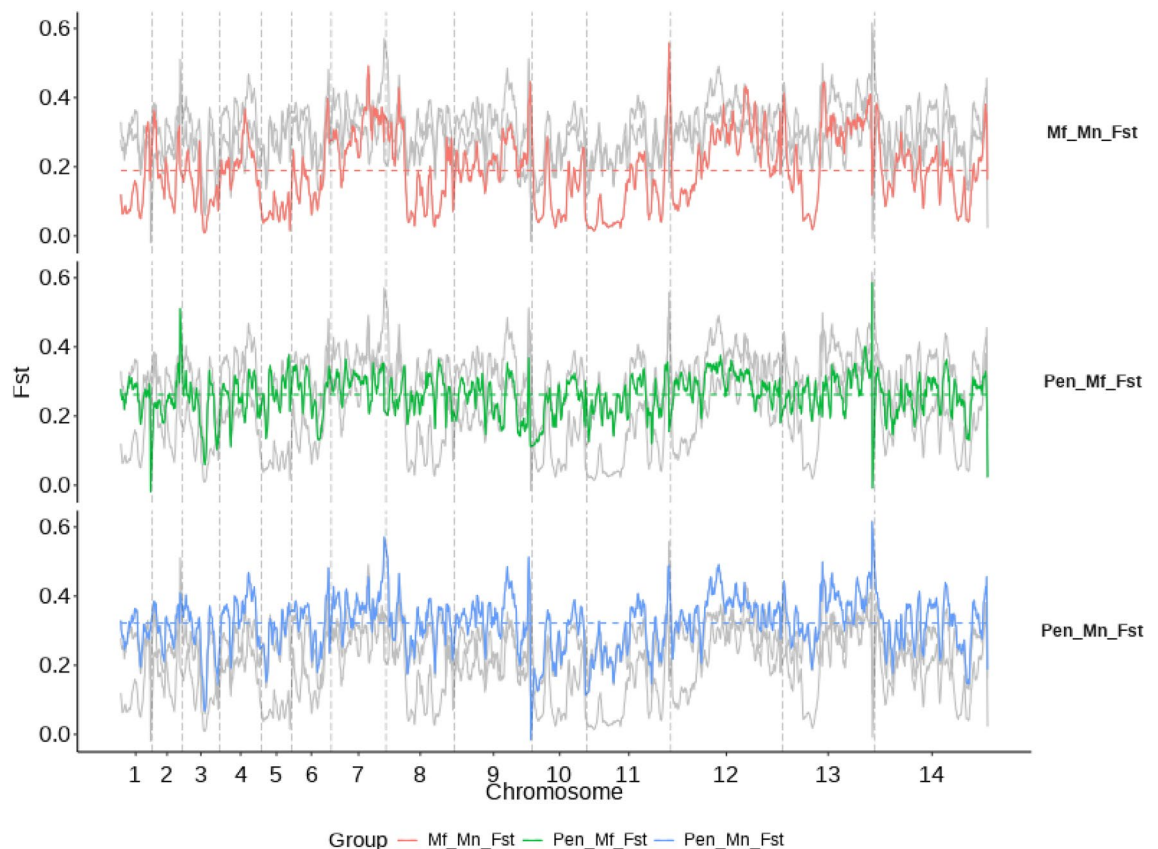
Comparison	High expression - ring [PKNH <sub>1</sub> ]	High expression - Schizont [PKNH <sub>1</sub> ]	High expression - Trophozoite [PKNH <sub>1</sub> ]	Low expression in host-related stages [PKNH <sub>1</sub> ]
Mn-Pk versus Mf-Pk	0604600, 0609000, 0612800, <b>0734800</b> , <b>0836600</b> , 0900700, 0934600 (SRP72), 1264600, 1464500	0101800, 0111000 (DegP), 0515200, <b>0819900 (H2A.Z)</b> , 1347900 (RAP1), 1464500	0515200, 0604600, 0609000, <b>0819900 (H2A.Z)</b> , 0900700, 0926200, 1264600, 1464500	0103400, 0111600, 0404800, 0415700, 0421500, 0501000, <b>0513200 (RRF2)</b> , 0713800 (GLTP), <b>0815200</b> , <b>0820100 (SPC2)</b> , <b>0821300</b> , <b>0824300 (EG5)</b> , <b>0829100</b> , 0904500, 0911900, 0918600 (HSBP), 1026200, 1118800, 1130900, 1132000, 1135000, 1136000, <b>1144800</b> , 1212200 (IF1), 1217900, 1253500, 1263400, 1303000, 1307200, 1314600, 1337700, 1350600, 1444200 (SF3B6), 1446400
Pen-Pk versus Mf-Pk	0100700	0100700	0100700	0602300
Pen-Pk versus Mn-Pk	<b>0734500</b>	<b>0734600</b> , 0812000, 1402300 (GAP)	0714500 (UFD1), <b>0734500</b> , <b>0734600</b> , 1234000	0100500, 0215700, 0713100 (UBC9), 0713500, 0722800, 0727700, 0727800, 0728400, 0728700 (LRR8), 0731600, <b>0733600</b> , <b>0733900 (VPS33)</b> , <b>0734200</b> , 0811800, 1136300, 1234900, 1240100, <b>1249600</b> , 1354000, 1434300

**Table 1.** Genes presenting low nucleotide diversity ( $\pi$ ) between two *P. knowlesi* sub-populations with significant divergence to the third cluster categorised by expression level reported in human-related stages of *P. knowlesi* using single cell experiment data<sup>23</sup>. Bolded genes have been observed in exchange events.



**Figure 3.** Genome-wide scan of nucleotide diversity, summarised in sliding windows of size 100 kbp, with a step size of 10 kbp. Horizontal dashed lines are the mean pairwise nucleotide diversity values for each sub-population, while vertical dashed lines represent chromosome boundaries.

Fst = 0.266) and between Borneo groups (Mn-Pk vs. Mf-Pk; mean Fst = 0.170). The results were consistent with the nucleotide diversity analysis, where regions of low nucleotide diversity between Pen-Pk and Mn-Pk cluster on fragments of chromosome 7, 12 and 13 were highly divergent among the Borneo populations. There are clear mosaic patterns of allele frequency differences, especially visible between the Borneo populations (Fig. 4), and we focus on genes with high numbers of fixed SNPs (Fst = 1) among the comparisons (S3 Table). The high density of point mutations in *P. knowlesi* led to more than half of known genes (n = 2199) with reported fixed



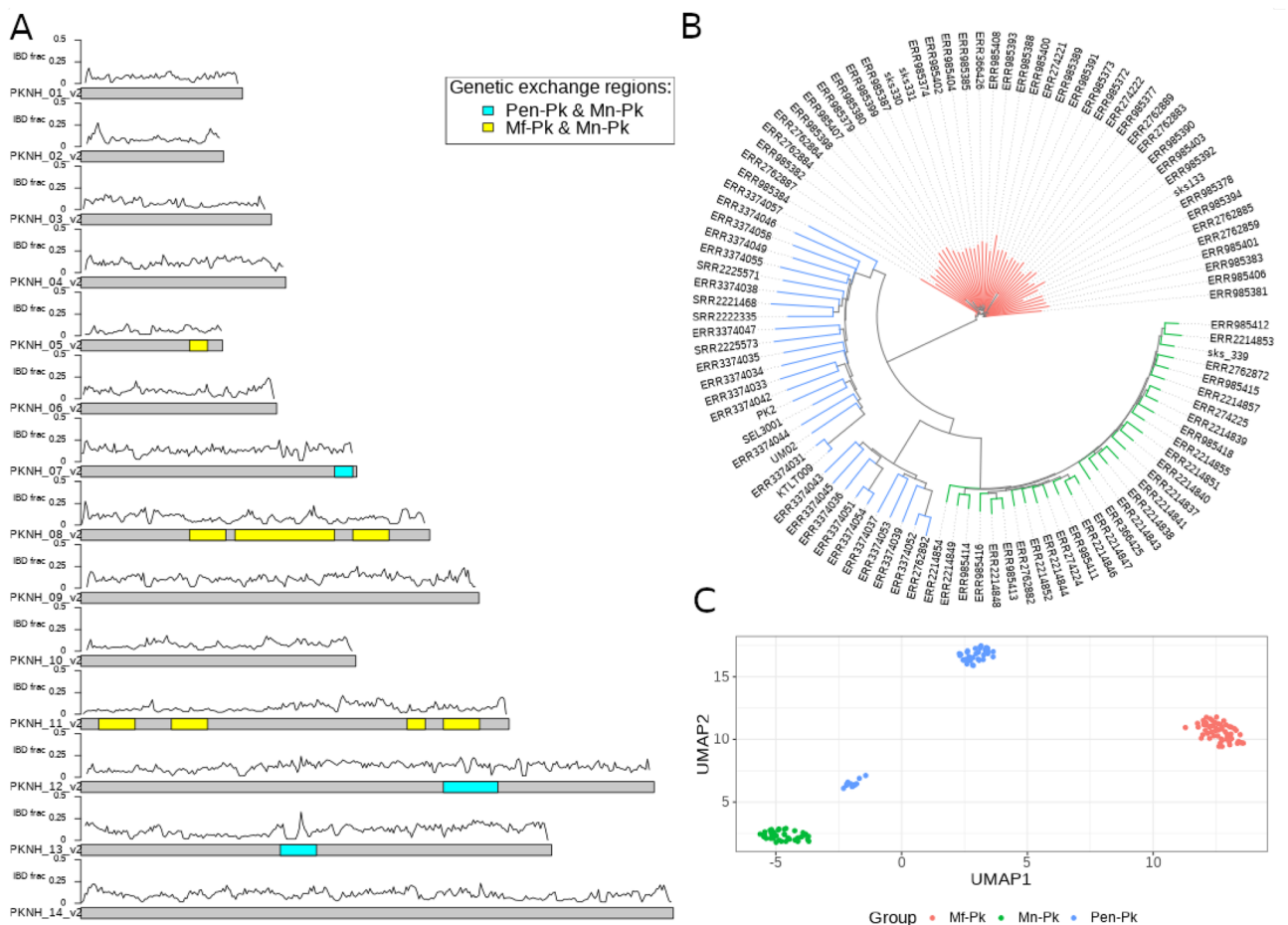
**Figure 4.** Pairwise fixation index (Fst) between sub-populations, summarised in sliding windows of size 100 kbp, with a step size of 10 kbp. Horizontal dashed lines are the mean pairwise Fst values for each sub-population, while vertical dashed lines represent chromosome boundaries.

SNPs. For the Pen-Pk and Mn-Pk comparison, the highest number of genes ( $n = 1,898$ ) with fixed SNPs were found, including ApiAP2 (PKNH\_1232600) (23 SNPs), SETvs (21), CAF1 (21), RRP6 (21), RhopH2 (19), EG5 (18), SEC7 (17), RON2 (16), AP2NAG (16), MRP2 (16), RAP1 (14), SR140 (14), and SLARP (13). These loci are involved across different parasite life cycle stages. The orthologous PfSETvs is involved in controlling expression levels of *var* genes in *P. falciparum*, while RRP6 has a role in the regulation of dynamic chromatin structure<sup>30</sup>. By comparing the Peninsular and Mf-associated groups, 1289 genes were revealed, including EG5 (16 fixed SNPs), two putative AP2 domain transcription factors (PKNH\_1016500 (13); PKNH\_1232600 (14)) known to regulate parasite life cycle transition<sup>31</sup>, and CRMP4 (13) linked to sporozoite egress. More than 80% of genes with a fixed SNP overlapped with the previous comparison (Pen-Pk vs. Mn-Pk), confirming strong differences between Peninsular and Borneo isolates, especially in EG5, ApiAP2 (PKNH\_1016500, PKNH\_1232600), SETvs, and RRP6, with only one fixed SNP between Borneo groups in these loci. The lowest number of genes with fixed SNPs (686) was between the two Borneo clusters, with more than half (53.6%) found in regions on chromosomes 7, 12 and 13 (Fig. 4, S3 Table), including GDV1, LRR8, RON2, IMC1f, CRMP4, and AP2NAG2 genes. Other genes presenting notable quantities of fixed SNPs between Borneo clusters, included NBPXa (57), ApiAP2 (PKNH\_1417900) (48), and the MRP2 protein (47) linked with drug resistance in other Plasmodium species. To further examine the broad function of genes with fixed SNPs, an analysis of Gene Ontology (GO) enrichment (molecular function, biological processes, cellular component) was performed (S4 Table). This analysis revealed the highest number of associated GO sub-groups (62) was linked to the Pen-Pk and Mn-Pk comparison, with the broader molecular function class being the most common (34/62). Similar results were found for Pen-Pk and Mf-Pk, but the Borneo sup-populations differ predominantly with many fixed SNPs in genes linked to broader biological processes (S4 Table).

**Characterisation of genetic exchange events among Borneo isolates.** Using an integrated approach involving Uniform Manifold Approximation and Projection (UMAP) clustering, neighbour-joining tree and nucleotide diversity estimation methods (see Methods), strong signals of introgression (genetic exchange events) were identified among Borneo isolates in chromosomes 5, 8, and 11 (S3 Figure). For the chromosome 5 (600–700 kbp) region, some Mf-associated isolates were found within the Mn-Pk group (S3 Figure A). The related chromosome 5 genes included putative RIO2 (PKNH\_0512800), ACP2 (PKNH\_0513000), ApiAP2 (PKNH\_0513100) and RRF2 (PKNH\_0513200). Analysis conducted on multiple loci throughout chromosome 8 (600–800 kbp, 850–1400 kbp, 1500–1700 kbp) confirmed the clustering of some Mf-Pk isolates within the Mn-associated group (S3 Figure B, C, D). The matching sequence segments were observed in subgroups

of Betong and Sarikei Mf-Pk isolates, consistent with previously observed introgression events of large chromosomal regions between Mf-Pk and Mn-Pk<sup>14</sup>. The chromosome 8 regions included genes linked with sexual development (e.g. HAP2 - PKNH\_0814100, DOZI - PKNH\_0820000) and parasite development in mosquitoes (e.g. oocyst Cap380 - PKNH\_0820800, CTRP - PKNH\_0826900). Host associated genes with evidence of essential asexual development in *P. falciparum*, included HSP60 (PKNH\_0815700), GSK3 (PKNH\_0829800), H2A.Z (PKNH\_0819900), ABCI3 (PKNH\_0821500) and ARK2 (PKNH\_0833500). Numerous genetic exchange events were identified throughout chromosome 11 (100–300 kbp, 500–700 kbp, 1800–1900 kbp, 2000–2200 kbp) (S3 Figure E, F, G, H), mostly involving separation of various Betong and Sarikei Mf-Pk isolates. One of the introgression events (500–700 kbp) contained Mf-Pk samples isolated from Kapit. In addition, singular Mn-Pk isolates were detected among the Mf-Pk sub-population, confirming that genetic exchange events occur in the opposite direction, but with lower frequency and on a smaller scale. Most chromosomal events involved genes linked to vector related parasite stages, including DHHC2 (PKNH\_1140400) and PSOP2 (PKNH\_1103400) proteins, essential for ookinete morphogenesis, as well as SAS6 (PKNH\_1142700) found to be a key protein in male gametogenesis in *P. berghei*. Moreover, close genetic similarity and a lack of separation between Mn-Pk and Mf-Pk clusters demonstrated in those loci accounts for the mosaic structure of the sub-populations. All the selected loci had clear re-branching of the Mn-Pk cluster from the more diverged Mf-associated group.

Using the same analytical approach as above, three regions were identified as highly similar between Pen-Pk and Mn-Pk groups, with evidence of genetic exchange events occurring on chromosome 7 (1400–1500 kbp), 12 (2000–2300 kbp) and 13 (1100–1300 kbp) (Fig. 5, S4 Figure). These regions were identified in the IBD analysis for combined Pen-Pk and Mn-Pk sub-populations, as well as in the nucleotide diversity analysis (Table 1). Almost all isolates involved in exchanges were sourced from the Peninsular region of Kuala Lipis and Sungai Siput, with Taiping and Johor locations contributing to singular events. The fragment on chromosome 7 contains multiple genes, including putative PKAc (PKNH\_0733500), GDV1 (PKNH\_0734100) and ETRAMP (PKNH\_0734700), and are predominantly linked to the host related stages of the Plasmodium cycle. The *P. falciparum* PKAc gene is linked to asexual stage growth, the gametocyte development protein (PfGDV1) is critical for the early sexual differentiation, and ETRAMP gene is fundamental for transformation from rings to trophozoites. In addition, the single cell expression data confirmed high expression of the GDV1 and ETRAMP genes, as well as some



**Figure 5.** Loci and evidence of exchange events between Mn-Pk and Pen-Pk clusters. (A) IBD fraction for the Mn-Pk and Pen-Pk isolates, regions with evidence of exchanges are coloured; (B) Neighbour-joining tree for chromosome 7, 1400–1500 kbp; (C) UMAP plot for chromosome 7, 1400–1500 kbp.

of the neighbouring genes of unknown function (PKNH\_0734500 and PKNH\_0734600)<sup>23</sup>. The largest fragment involving genetic exchange was on chromosome 12 (length 300kbp), and contains more than 50 genes, including the CRMP4 (PKNH\_1245700), ETRAMP (PKNH\_1246400), UTP15 (PKNH\_1248800), ABCB6 (PKNH\_1248900), IMC1f (PKNH\_1249300) and multiple PHIST proteins (PKNH\_1247500, PKNH\_1247600 and PKNH\_1247700). The identified region on chromosome 13 was mostly characterised with *Plasmodium* exported proteins of unknown function. However, two PHIST genes (PKNH\_1324600, PKNH\_1326000) and the inner membrane complex suture component (ISC3 - PKNH\_1326900) were also identified.

**Genetic diversity in invasion genes.** The first evidence that the *P. knowlesi* parasite comprises distinct sub-populations was revealed in the analysis of erythrocyte invasion genes<sup>32</sup>, and later confirmed and extended with the knowledge of inter-group exchange events<sup>12</sup>. We used the neighbour-joining tree and nucleotide diversity approach to characterise the structure of the two reticulocyte binding like (RBL) and three Duffy binding protein like (DBP) genes associated with erythrocyte invasion (S5 Figure). The DBP $\alpha$  gene was the only locus with clear partitioning into three sub-populations coinciding with a whole genome classification (S5 Figure A). For both DBP $\beta$  and DBP $\gamma$  (S5 Figure B, C), the clustering pattern matches the overall genetic structure, but some of the Pen-Pk isolates show evidence of genetic exchange with Mf-Pk and Mn-Pk clusters. A Pen-Pk isolate from Taiping (ERR3374057) aligned more closely with the Mf-associated cluster using the DBP $\beta$  locus. Moreover, in the same locus, a subset of Pen-Pk samples from Singai Siput and Kuala Lipis present with exchange events linking them to the Mn-associated group. These exchanges show that recombination events of the Peninsular group can be linked to both Borneo clusters in the same loci, potentially contributing to the high diversity of the Pen-Pk isolates. The evidence of exchange events among subsets of Pen-Pk isolates with the Mf-associated cluster were also observed in the DBP $\gamma$  locus. This exchange contains four samples, including a historical laboratory strain (SRR2225573, The Philippines), a newly derived clone from *M. fascicularis* blood culture (UM02), and 2 field isolates (ERR3374057, Taiping; ERR3374049, Kuala Lipis), with one linked to the Mf-Pk cluster involving the DBP $\beta$  region. Existence of the recombination events within laboratory maintained isolates, and their significant branch length on the tree, confirmed that such events are non-recent. For the NBPX $\beta$  locus (S5 Figure D, E), subdivision of the Mf-Pk group was observed. However, Mn- and Pen-associated sub-populations were indistinguishable, with a subset of seven Pen-Pk isolates collected from various locations (Kuala Lipis, Taiping and Sungai Siput) being affiliated to the Mn-associated cluster. A similar genetic exchange event was confirmed within the NBPX $\alpha$  locus (S5 Figure D), where a subgroup of ten Pen-Pk isolates (mostly from Kuala Lipis and Sungai Siput) aligned more closely with the Mn-associated sub-population. Once again, the significant length of most of the divergent branches is indicative of the genetic exchange event being non-recent.

**Regions under selection in *P. knowlesi* sub-populations.** A genome-wide scan for genes under recent positive selection was performed for each sub-population using the integrated haplotype score (iHS) approach. Results for the Mf-associated group revealed numerous fragments spreading across all chromosomes (S6 Figure A). Moreover, fragments on chromosome 8 (0.6–0.66 Mbp, 0.9–0.95 Mbp, 1.02–1.05 Mbp and 1.68–1.71 Mbp) and 11 (2.07–2.1 Mbp) were found to be overlapping with the previously shown differentiation of the Betong and Sarikei sub-population<sup>14</sup>. In contrast, the Mn-Pk group shows little evidence of extended haplotypes affirming its low genetic diversity (S6 Figure B). Selective sweeps in the Peninsular cluster were identified on chromosomes 5 and 6 (S6 Figure C), which include inner membrane complex (IMC1m, PKNH\_0613600) and merozoite TRAP-like (MTRAP, PKNH\_0613400) proteins.

Genome-wide scans for positive selection that compared between sub-populations revealed little evidence of selective sweeps for Pen-Pk and Mn-Pk clusters (S6 Figure D, E, F). However, multiple fragments were identified between the Peninsular and Mf-associated groups, including regions on chromosomes 4 (0.82–0.85 Mbp; e.g., LRR5 gene, PKNH\_0419000), 7 (0.58–0.61 Mbp), 11 (2.17–2.20 Mbp; e.g., SIP2 gene, PKNH\_1146400), 12 (2.45–2.48 Mbp; e.g., VPS13 (PKNH\_1264700)), and 13 (0.85–0.88 Mbp; e.g., IF3a (PKNH\_1319200)). Almost all of the fragments enclose putative PIR or SICAvir genes. A comparison of the Borneo associated clusters revealed fragments with selection signals on chromosomes 2 (0.37–0.40 Mbp), 5 (0.30–0.33 Mbp), 7 (1.08–1.11 Mbp) and 14 (1.18–1.21 Mbp). The NPT1 protein (PKNH\_0208700) was identified and is known to be crucial for early stage of sexual development in *P. berghei*.

## Discussion

Despite the widespread occurrence of *P. knowlesi* parasites in Southeast Asia, there are many gaps in the understanding of its evolutionary dynamics, geographical distribution and transmission. Whilst the genetic adaptation of the parasite to the macaque hosts on Borneo Island and geographical differences in comparison to Peninsular Malaysia have been described<sup>12</sup>, the mosaic genome-wide structure occurring between those sub-populations is not well understood. Our study performed genome-wide sequence analyses of the largest collection of *P. knowlesi* isolates yet assembled, and revealed evidence of genetic exchange events between highly divergent Peninsular Malaysia (Pen-Pk) and one of the Borneo groups associated with *M. nemestrina* (Mn-Pk), as well as between Borneo sub-populations. Although, our work extended important whole-genome characterisation of understudied regions of Malaysia and beyond, wider sampling and larger studies are required. Further, inclusion of laboratory isolates cultured in *M. fascicularis* blood provided evidence of non-recent genetic exchange in erythrocyte invasion related genes, whose functional effects can be investigated using *in vitro* models<sup>26</sup>. Nonetheless, inclusion of more isolates derived from wild macaques would be necessary for a more in depth understanding of evolutionary changes of the *Plasmodium* species.

Exploration of *P. knowlesi* population structure across Malaysia revealed three predominant clusters, consistent with previous findings<sup>13,14</sup>. All newly sequenced isolates aligned within those sub-populations, and the



majority had mono infections and single source population ancestry. This observation would have been expected for the laboratory strains, which have undergone *in vitro* culture. However, three samples isolated from Sarikei had inferred mixed ancestry from both Borneo clusters, consistent with multiple genotypes of *P. knowlesi* in human infections and the circulation of underlying sub-populations in that geographical region<sup>15</sup>. A *P. knowlesi* and *P. vivax* co-infection was found, and infections with more than one *Plasmodium* species have been previously observed<sup>33</sup>.

Among all three *P. knowlesi* sub-populations, the Mn-associated group is known to be most genetically conserved, especially in highly divergent regions shown in other sub-populations<sup>34</sup>. It has been suggested that this observation is a result of the Mn-Pk group being an initial bottleneck in the formation of sub-populations<sup>34</sup>. Whereas, the *M. fascicularis* specific (Mf-Pk) and Pen-Pk groups present as genetically divergent throughout the genome<sup>35</sup>. IBD analysis supported these insights, and revealed regions of common ancestry within and between Borneo populations, appearing to be linked with some known vector-related genes like Cap380 and CSP. Multiple regions were found to display high similarity within both Borneo-associated sub-populations, with various fragments on chromosome 8 previously described to have genetic exchange events between the Mf- and Mn-associated genotypes<sup>14</sup>. The identified vector-related genes among the Borneo groups may be linked with the specific ecology of members of the Leucosphyrus Complex *Anopheles* (*An. latens* and *An. balabacensis*) spread across Malaysia, Indonesia, Singapore, Brunei and part of The Philippines<sup>36</sup>. Nonetheless, differences in the habitat zone of *M. fascicularis*, being a canopy roosting macaque while *M. nemestrina* is mainly terrestrial, may expose the animals to distinct sub-populations within the Leucosphyrus Group that are closely related but distinct from mosquitoes in other zones. In contrast, peninsular Malaysia and mainland Southeast Asian countries are associated with *Anopheles* species of the Dirus Complex (e.g. *An. cracens*) as the primary vectors of *P. knowlesi*<sup>36</sup>. Our observed levels of IBD within the Peninsular cluster were found to be remarkably high in loci predominantly related to host-parasite stages. Dissimilarities among sites of common descent between sub-populations could be caused by the geographical separation of the Borneo Island from Peninsular Malaysia.

The mosaic pattern and population differentiation across the Borneo-associated nuclear genome have been highlighted in previous population-based studies, leading to the detection of chromosomal-segment exchange events throughout chromosome 8<sup>14</sup>. Our analysis scanned genome-wide for introgression events across all three sub-populations, extending our knowledge of new exchange events between the Borneo sub-populations located on chromosomes 5 and 11. We identified new regions with evidence of genetic exchanges between previously understudied Mn-Pk and Pen-Pk associated clusters. Introgression events occur on chromosomes 7, 12 and 13, with identified loci possessing patterns of high divergence between the Borneo-associated genotypes, potentially explaining some of the factors driving mosaicism among the sub-populations. These results are consistent with a previous candidate gene analysis<sup>12</sup> and microsatellite analyses that identified traces of Borneo-associated clusters in genomic regions of Peninsular Malaysia<sup>37</sup>. The differences in branch length within and between groups in the neighbour-joining tree analysis suggest that the newly detected exchange events among Mn-Pk and Pen-Pk isolates are potentially caused by historic macaque crossings. Given the likely dates of these past events, the introgression is unlikely to have been facilitated by human host transitions, although further analyses are necessary to properly answer this question. The Pen-Pk genomic regions with exchange events were enriched with host-related genes, whereas Borneo-associated exchange events were primarily connected to mosquito-related stages of the parasite cycle. A potential alternative explanation for the extended regions of clusterisation outside of groups could be incomplete lineage sorting, where such changes have been observed with recent divergence among *P. vivax* species<sup>38</sup>. An analysis of invasion linked loci found that RBL/DBP (DBP  $\alpha$ ,  $\beta$  and  $\gamma$ , NBP X $\alpha$  and X $\beta$ ) genes are highly divergent and mostly coinciding with a whole genome classification. Between Pen-Pk and Mn-Pk associated sub-populations, exchange events were found in all studied invasion genes except DBP $\alpha$ . For the DBP $\beta$  and DBP $\gamma$  genes, individual Pen-Pk isolates were strongly differentiated from their original cluster. Additionally, the DBP $\beta$  locus has singular samples presenting strong affiliation to the Mf-associated cluster, highlighting that multiple introgression events can be found at the same loci. Exchange events revealed on the DBP $\gamma$  gene include a laboratory cultured Philippine strain and a new isolate collected from *M. fascicularis* red blood cell culture. This observation suggests that the genetic exchanges are not recent, especially because of the significant length of the neighbour-joining tree branches and isolation of the laboratory sample in year 1960. The NBPX $\alpha$  and NBPX $\beta$  genes present introgression events where some Pen-Pk isolates cluster with the Mn-Pk associated sub-population. Almost all of the invasion genes show genetic exchange events between Pen-Pk and Mn-Pk associated clusters, highlighting the potential ancestral events between those clusters related to the adaptation of the parasite to different simian hosts. Exchange events in those genes can strongly impact on invasion mechanisms, which can be explored using experimental models<sup>26</sup>.

Our work has provided new insights into *P. knowlesi* evolution, highlighting the genetic exchanges as well as regions of identity among known sub-populations. The *P. knowlesi* data demonstrate significant divergence caused by the geographical separation in the Peninsular cluster and non-human host specification in Borneo groups, whilst, still displaying the ability to recombine when in contact<sup>11</sup>. The fragments of recombination can have an important impact on patterns of diversity among sub-populations and create the observed mosaicism. Our analyses contribute to the investigation of the genetic structure of the parasite by providing evidence of new introgression events. Genetic exchange events found between Pen-Pk and Mn-Pk associated clusters almost entirely overlay with long regions of high divergence shown in Borneo sub-populations, and are non-recent, suggesting a long-term process contributing to the population structure and mosaic patterns of *P. knowlesi* we observe today. In lieu of larger and geographically diverse studies within the wider region, our work has generated new hypotheses as to the historical shape of population movement for both parasite and host, which can be tested in geographically broader sample sets as they become available.

## Methods

**Isolates and sequence data.** A total of 151 *P. knowlesi* isolates with WGS were analysed, including: (i) publicly available data ( $n = 126$ )<sup>12,13,15,34,35</sup>; (ii) isolates sourced from Malaysia ( $n = 15$ ; Peninsular 4, Borneo 11), provided by the University Malaya Medical Centre (spanning July 2008 to December 2014); (iii) isolates from returning travellers to the UK from Malaysia ( $n = 1$ ) and Indonesia ( $n = 1$ ); (iv) laboratory strains ( $n = 8$ ), including two samples isolated from *in vitro* blood cultures of *M. fascicularis* (ERR9751937 and ERR9751954) and six samples from *in vitro* human red blood cell cultures (PkA1-H.1). DNA was extracted from clinical samples using the DNeasy Blood & Tissue Qiagen kit, and underwent parasite selective whole genome amplification, as described previously<sup>12</sup>. Sequencing of the DNA from the newly generated isolates ( $n = 25$ ) was performed on an Illumina HiSeq 4000 platform by The Applied Genome Centre, LSHTM. All WGS data was screened using Centrifuge software to ensure a significant abundance of reads derived from *P. knowlesi* (abundance higher than 0.35), and to confirm host species (*H. sapiens*, *M. fascicularis* or *M. nemestrina*). Ethical approval in written and verbal form was provided by the University of Malaya Medical Centre Medical Ethics Committee (Ref. No: 817.18). Informed consent was obtained for study participation in both Sabah and Kuala Lumpur sites. The UK National Research Ethics Service (Ref: 18/LO/0738) and LSHTM Research Ethics Committee (Ref: 14710) provided approval for the project “Drug susceptibility and genetic diversity of imported malaria parasites from UK travellers”. New data provided in the study can be found on ENA (PRJEB52783). Details of the WGS data, including ENA accession numbers, are provided (S1 Table).

**Bioinformatic analysis.** All raw sequencing data was filtered using trimmomatic software (v0.39; parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36)<sup>39</sup>. Filtered reads were then mapped to the *P. knowlesi* A1-H.1 reference genome<sup>40</sup> using BWA-MEM alignment (v0.7.17) software<sup>41</sup>. SNPs and insertions/deletions (indels) were found using GATK's Base Quality Score Recalibration and HaplotypeCaller (v4.1.4.1), and validation was performed using the ValidateVariants function with default settings<sup>42</sup>. Variants with a Variant Quality Score (VQSLOD) in excess of zero and low levels of missing genotypes (<20%) were retained. Variants occurring in subtelomeric regions or SICAvir genes were excluded. The summary statistics for all analysed samples are provided (S1 Table). The final dataset consists of 151 isolates representing 3 clusters (Cluster 1 - Mf-Pk,  $n = 60$ ; Cluster 2 - Mn-Pk,  $n = 41$ ; Cluster 3 - Pen-Pk,  $n = 50$ ), with 1,883,700 high quality SNPs. Variants were annotated using snpEff (v4.1) software<sup>43</sup>. The multiplicity of infection (MOI) was estimated from biallelic variants using estMOI software<sup>44</sup> and the Fws score, calculated using the mixmoi package (github.com/bahlolab/moimix) for each of the known clusters (Mn-Pk, Mf-Pk and Pen-Pk). Isolates with Fws values  $\geq 0.95$  were taken as infections with a predominant singular genotype, whereas samples below the threshold were assumed to be mixed infections as previously shown<sup>10</sup>.

**Population genetic analysis.** Population structure was explored by performing a PCA and constructing a neighbour-joining tree using the R ape package. PCA was performed on the 151 available isolates using pairwise Manhattan distances based on biallelic SNPs. The ADMIXTURE (v1.3.0) software was applied to estimate individual ancestry<sup>45</sup>, where the number of sub-populations was determined by cross-validation error. The ADMIXTURE analysis was filtered for linkage disequilibrium (LD), where SNPs were identified for pruning using Plink software (with settings: - indep-pairwise 50 10 0.1). Samples with evidence of the multiplicity of infection (Fws < 0.95), low coverage (<five-fold) or identified as highly similar (< 50,000 sites difference) were excluded from any further analyses, to ensure a more robust population genomics analysis. The resulting dataset of 104 high-quality isolates (Mf-Pk,  $n = 43$ ; Mn-Pk,  $n = 30$ ; Pen-Pk,  $n = 31$ ) was used for population genetic analysis. The IBD analysis was performed on biallelic SNPs with a minor allele frequency (MAF) of >5%. This analysis was applied pairwise between each group using the package hmmIBD<sup>46</sup>, and involved sliding windows of size 10kbp. The proportion of pairwise comparisons for isolates presenting evidence of IBD, corrected for fragment length, was plotted by genome location. The top 1% of hits were considered significant.

To measure the amount of polymorphism within each cluster, the average pairwise nucleotide diversity ( $\pi$ ) was estimated genome-wide using the R Pegas package with sliding windows (100 kbp window size, 10 kbp step size). Pairwise nucleotide diversity between groups was calculated for each gene, and the results used to report loci presenting low nucleotide diversity between two clusters, but highly separated from the remaining one. The fixation index (Fst) metric was calculated using VCFtools software on SNPs with a MAF > 5%. A genome-wide scan of the Fst results was performed using sliding windows (100 kbp window size, 10 kbp step size). Genes containing fixed SNPs (Fst = 1) in the pairwise sub-population comparisons were identified, and a gene ontology (GO) term analysis was performed using previously described methods<sup>15</sup>. Gene function across Plasmodium species was established using the PlasmoDB database (plasmodb.org).

All three groups (Mf-Pk, Mn-Pk, Pen-Pk) were screened for recent positive selection using the R rehh package<sup>47</sup>, applied to SNPs with MAF > 5%. Both the within-population integrated haplotype score (iHS) and between-population Rsb score for identification of selection<sup>48</sup> were calculated. Critical regions were identified using 100 kbp sliding windows, which included at least 3 SNPs with a  $p$ -value <  $1 \times 10^{-4}$  for iHS and  $p$ -value <  $1 \times 10^{-5}$  for Rsb<sup>10</sup>. The identification of potential introgression regions was based initially on genome-wide pairwise nucleotide diversity ( $\pi$ ), calculated within sliding windows (window size 100 kbp, step size 50kbp). Regions presenting a low level of diversity among pairwise sub-population comparisons ( $\pi < 0.005$ ) were further investigated using the Fst metric applied within the sliding windows to confirm divergence from the remaining sub-population. To further support these exchange events, the population structure was assessed using the UMAP algorithm applied to 100 kbp windows encompassing the putative region. Fragments displaying differences from whole genome-based clustering, supported by changes in nucleotide diversity and Fst divergence analysis, were further confirmed using a neighbour-joining tree, constructed using the R ape package. The usage of the

UMAP algorithm has previously separated *Plasmodium* species<sup>49</sup>. Single-cell expression data was used to link *P. knowlesi* gene findings to parasite life cycle stages<sup>23</sup>. The mean expression levels were calculated for each gene by stage, and loci revealed in our analysis were checked against the list of the top 20% of (highly) expressed genes<sup>23</sup>.

## Data availability

Previously published WGS data can be found on the European Nucleotide Archive (ENA) using the Run accession codes in S1 Table. The newly generated data can be found in the ENA study accession number PRJEB52783.

Received: 12 October 2022; Accepted: 3 February 2023

Published online: 07 February 2023

## References

- Kantele, A. & Jokiranta, T. S. Review of cases with the emerging fifth human malaria parasite *Plasmodium knowlesi*. *Clin. Infect. Dis.* **52**, 1356–1362. <https://doi.org/10.1093/cid/cir180> (2011).
- Hussin, N. *et al.* Updates on malaria incidence and profile in Malaysia from 2013 to 2017. *Malar. J.* **19**, 1–14. <https://doi.org/10.1186/s12936-020-3135-x> (2020).
- World Health Organisation. *World Malaria Report 2020: 20 Years of Global Progress and Challenges* (World Health Organisation, 2020).
- Kotepui, M., Kotepui, K. U., Milanez, G. D. & Masangkay, F. R. Prevalence of severe *Plasmodium knowlesi* infection and risk factors related to severe complications compared with non-severe *P. knowlesi* and severe *P. falciparum* malaria: a systematic review and meta-analysis. *Infect. Dis. Prev.* **9**, 1–14. <https://doi.org/10.1186/s40249-020-00727-x> (2020).
- Fornace, K. M. *et al.* Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLoS Negl. Trop. Dis.* **12**, 1–16. <https://doi.org/10.1371/journal.pntd.0006432> (2018).
- Ang, J. X. *et al.* New vectors in northern Sarawak, Malaysian Borneo, for the zoonotic malaria parasite *Plasmodium knowlesi*. *Parasit. and Vectors* **13**, 1–13. <https://doi.org/10.1186/s13071-020-04345-2> (2020).
- De Ang, J. X., Yaman, K., Kadir, K. A., Matusop, A. & Singh, B. New vectors that are early feeders for *Plasmodium knowlesi* and other simian malaria parasites in Sarawak, Malaysian Borneo. *Sci. Rep.* **11**, 1–12. <https://doi.org/10.1038/s41598-021-86107-3> (2021).
- Vythilingam, I. *et al.* *Plasmodium knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. *Parasit. Vectors* **1**, 1–10. <https://doi.org/10.1186/1756-3305-1-26> (2008).
- Davidson, G., Chua, T. H., Cook, A., Speldewinde, P. & Weinstein, P. Defining the ecological and evolutionary drivers of *Plasmodium knowlesi* transmission within a multi-scale framework. *Malar. J.* **18**, 1–13. <https://doi.org/10.1186/s12936-019-2693-2> (2019).
- Benavente, E. D. *et al.* Distinctive genetic structure and selection patterns in *Plasmodium vivax* from South Asia and East Africa. *Nat. Commun.* **12**, 1–11. <https://doi.org/10.1038/s41467-021-23422-3> (2021).
- Brock, P. M. *et al.* *Plasmodium knowlesi* transmission: integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology* **143**, 389–400. <https://doi.org/10.1017/S0031182015001821> (2016).
- Benavente, E. D. *et al.* Whole genome sequencing of amplified *Plasmodium knowlesi* DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. *Sci. Rep.* **9**, 1–11. <https://doi.org/10.1038/s41598-019-46398-z> (2019).
- Assefa, S. *et al.* Population genomic structure and adaptation in the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13027–13032. <https://doi.org/10.1073/pnas.1509534112> (2015).
- Diez Benavente, E. *et al.* Analysis of nuclear and organellar genomes of *Plasmodium knowlesi* in humans reveals ancient population structure and recent recombination among host-specific subpopulations. *PLoS Genet.* **13**, 1–16 (2017).
- Pinheiro, M. M. *et al.* *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. *PLoS One* **10**, 1–16. <https://doi.org/10.1371/journal.pone.0121303> (2015).
- Srinivasan, P., Fujioka, H. & Jacobs-Lorena, M. PbCap380, a novel oocyst capsule protein, is essential for malaria parasite survival in the mosquito. *Cell. Microbiol.* **10**, 1304–1312. <https://doi.org/10.1111/j.1462-5822.2008.01127.x> (2008).
- Sturm, A., Mollard, V., Cozijnsen, A., Goodman, C. D. & McFadden, G. I. Mitochondrial ATP synthase is dispensable in blood-stage *Plasmodium berghei* rodent malaria but essential in the mosquito phase. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10216–10223. <https://doi.org/10.1073/pnas.1423959112> (2015).
- Ukegbu, C. V. *et al.* PIMMS43 is required for malaria parasite immune evasion and sporogonic development in the mosquito vector. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7363–7373. <https://doi.org/10.1073/pnas.1919709117> (2020).
- Ramaprasad, A., Klaus, S., Douvropoulou, O., Culleton, R. & Pain, A. *Plasmodium vinckei* genomes provide insights into the pan-genome and evolution of rodent malaria parasites. *BMC Biol.* **19**, 1–22. <https://doi.org/10.1186/s12915-021-00995-5> (2021).
- Tremp, A. Z., Al-Khattaf, F. S. & Dessens, J. T. Distinct temporal recruitment of *Plasmodium alveolins* to the subpellicular network. *Parasitol. Res.* **113**, 4177–4188. <https://doi.org/10.1007/s00436-014-4093-4> (2014).
- Boysen, K. E. & Matuschewski, K. Arrested oocyst maturation in *Plasmodium* parasites lacking type II NADH: ubiquinone dehydrogenase. *J. Biol. Chem.* **286**, 32661–32671. <https://doi.org/10.1074/jbc.M111.269399> (2011).
- Reiter, K. *et al.* Identification of biochemically distinct properties of the small ubiquitin-related modifier (SUMO) conjugation pathway in *Plasmodium falciparum*. *J. Biol. Chem.* **288**, 27724–27736. <https://doi.org/10.1074/jbc.M113.498410> (2013).
- Howick, V. M. *et al.* The malaria cell atlas: single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science* **365**, aaw2619. <https://doi.org/10.1126/science.aaw2619> (2019).
- Tremp, A. Z., Khater, E. I. & Dessens, J. T. IMC1b is a putative membrane skeleton protein involved in cell shape, mechanical strength, motility, and infectivity of malaria ookinetes. *J. Biol. Chem.* **283**, 27604–27611. <https://doi.org/10.1074/jbc.M801302200> (2008).
- Ozaki, L. S., Svec, P., Nussenzweig, R. S., Nussenzweig, V. & Godson, G. N. Structure of the *Plasmodium knowlesi* gene coding for the circumsporozoite protein. *Cell* **34**, 815–822. [https://doi.org/10.1016/0092-8674\(83\)90538-X](https://doi.org/10.1016/0092-8674(83)90538-X) (1983).
- Moon, R. W. *et al.* Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7231–7236. <https://doi.org/10.1073/pnas.1522469113> (2016).
- Sahu, T. *et al.* ZIPCO, a putative metal ion transporter, is crucial for *Plasmodium* liver-stage development. *EMBO Mol. Med.* **6**, 1387–1397 (2014).
- Baldi, D. L. *et al.* RAP1 controls rhoptry targeting of RAP2 in the malaria parasite *Plasmodium falciparum*. *EMBO J.* **19**, 2435–2443. <https://doi.org/10.1093/emboj/19.11.2435> (2000).
- Counihan, N. A. *et al.* *Plasmodium falciparum* parasites deploy RhopH2 into the host erythrocyte to obtain nutrients, grow and replicate. *eLife* **6**, 1–31 (2017).
- Fan, Y. *et al.* Rrp6 regulates heterochromatic gene silencing via ncRNA. *mBio* **11**, 1–14 (2020).
- Jeninga, M. D., Quinn, J. E. & Petter, M. Apiap2 transcription factors in apicomplexan parasites. *Pathogens* **8**, 1–24. <https://doi.org/10.3390/pathogens8020047> (2019).

32. Ahmed, A. M. *et al.* Disease progression in *Plasmodium knowlesi* malaria is linked to variation in invasion gene family members. *PLoS Negl. Trop. Dis.* **8**, 3086. <https://doi.org/10.1371/journal.pntd.0003086> (2014).
33. Lubis, I. N. *et al.* Contribution of *Plasmodium knowlesi* to multispecies human Malaria infections in North Sumatera, Indonesia. *J. Infect. Dis.* **215**, 1148–1155. <https://doi.org/10.1093/infdis/jix091> (2017).
34. Divis, P. C., Duffy, C. W., Kadir, K. A., Singh, B. & Conway, D. J. Genome-wide mosaicism in divergence between zoonotic malaria parasite subpopulations with separate sympatric transmission cycles. *Mol. Ecol.* **27**, 860–870. <https://doi.org/10.1111/mec.14477> (2018).
35. Hocking, S. E., Divis, P. C., Kadir, K. A., Singh, B. & Conway, D. J. Population genomic structure and recent evolution of *Plasmodium knowlesi* Peninsular Malaysia. *Emerg. Infect. Dis.* **26**, 1749–1758. <https://doi.org/10.3201/eid2608.190864> (2020).
36. Moyes, C. L. *et al.* Predicting the geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria in forested and non-forested areas. *Parasit. Vectors* **9**, 1–12. <https://doi.org/10.1186/s13071-016-1527-0> (2016).
37. Divis, P. C. *et al.* Admixture in humans of two divergent *Plasmodium knowlesi* populations associated with different macaque host species. *PLoS Pathog.* **11**, 1–17. <https://doi.org/10.1371/journal.ppat.1004888> (2015).
38. Gilabert, A. *et al.* *Plasmodium vivax*-like genome sequences shed new insights into *Plasmodium vivax* biology and evolution. *PLoS Biol.* **16**, 1–25. <https://doi.org/10.1371/journal.pbio.2006035> (2018).
39. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
40. Benavente, E. D. *et al.* A reference genome and methylome for the *Plasmodium knowlesi* A1–H.1 line. *Int. J. Parasitol.* **48**, 191–196. <https://doi.org/10.1016/j.ijpara.2017.09.008> (2018).
41. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **00**, 1–3 (2013).
42. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* 1st edn. (O'Reilly Media, 2020).
43. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w 1118; iso-2; iso-3. *Landes Biosci.* **6**, 80–92. <https://doi.org/10.1070/qe1980v010n03abeh009978> (2012).
44. Assefa, S. A. *et al.* EstMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics* **30**, 1292–1294. <https://doi.org/10.1093/bioinformatics/btu005> (2014).
45. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-12-246> (2011).
46. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. HmMIBD: software to infer pairwise identity by descent between haploid genotypes. *Malar. J.* **17**, 10–13. <https://doi.org/10.1186/s12936-018-2349-7> (2018).
47. Gautier, M., Klassmann, A. & Vitalis, R. reh2.0: a reimplementation of the R package reh1 to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17**, 78–90. <https://doi.org/10.1111/1755-0998.12634> (2017).
48. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**, 1587–1602. <https://doi.org/10.1371/journal.pbio.0050171> (2007).
49. Real, E. *et al.* A single-cell atlas of *Plasmodium falciparum* transmission through the mosquito. *Nat. Commun.* **12**, 1–13. <https://doi.org/10.1038/s41467-021-23434-z> (2021).

## Acknowledgements

We thank the staff and patients at the University Malaya Medical Centre for providing the original samples for this study. All methods were performed in accordance with the relevant guidelines and regulations. A.T. was funded by a Newton Institutional Links Grant (British Council, no. 261868591). S.C. was funded by BloomsburySET and Medical Research Council UK grants (MR/M01360X/1, MR/R025576/1, MR/R020973/1, and MR/X005895/1). T.G.C. was funded by the Medical Research Council UK (Grant nos. MR/M01360X/1, MR/N010469/1, MR/R025576/1, MR/R020973/1, and MR/X005895/1). The authors declare no conflicts of interest.

## Author contributions

T.G.C. and S.C. conceived and directed the project. D.R.O., D.N., C.J.S., J.C.-S., R.W.M., and Y.-L.L. organised sample collection and processing. S.C. undertook laboratory work including sequencing. A.T. performed bioinformatic analysis under the supervision of S.C. and T.G.C., and together they interpreted the results. A.S. and E.M. provided software. A.T. wrote the first draft of the manuscript. All authors commented on the results and on the manuscript and approved the final submission. A.T. and T.G.C. compiled the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29368-4>.

**Correspondence** and requests for materials should be addressed to S.C. or T.G.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023