



Succinctus

A reference genome and methylome for the *Plasmodium knowlesi* A1-H.1 line



Ernest Diez Benavente^a, Paola Florez de Sessions^b, Robert W. Moon^a, Munira Grainger^c, Anthony A. Holder^c, Michael J. Blackman^{a,c}, Cally Roper^a, Christopher J. Drakeley^a, Arnab Pain^d, Colin J. Sutherland^a, Martin L. Hibberd^{a,b}, Susana Campino^a, Taane G. Clark^{a,e,*}

^a Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

^b Genomics Institute Singapore, Singapore

^c The Francis Crick Institute, 1 Midland Road, London NW1 1AT, United Kingdom

^d King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

^e Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

ARTICLE INFO

Article history:

Received 15 April 2017

Received in revised form 10 August 2017

Accepted 26 September 2017

Available online 16 December 2017

Keywords:

Plasmodium knowlesi

Reference genome

PkA1-H.1

Sequencing

Malaria

Methylation

ABSTRACT

Plasmodium knowlesi, a common parasite of macaques, is recognised as a significant cause of human malaria in Malaysia. The *P. knowlesi* A1H1 line has been adapted to continuous culture in human erythrocytes, successfully providing an in vitro model to study the parasite. We have assembled a reference genome for the PkA1-H.1 line using PacBio long read combined with Illumina short read sequence data. Compared with the H-strain reference, the new reference has improved genome coverage and a novel description of methylation sites. The PkA1-H.1 reference will enhance the capabilities of the in vitro model to improve the understanding of *P. knowlesi* infection in humans.

© 2017 The Authors. Published by Elsevier Ltd on behalf of Australian Society for Parasitology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Plasmodium knowlesi, a common malaria parasite of long-tailed *Macaca fascicularis* and pig-tailed *Macaca nemestrina* macaques in southeastern Asia, is now recognised as a significant cause of human malaria. Clinical outcomes range from high parasitaemia to severe complications including death (Singh and Daneshvar, 2013). Early cases were largely misdiagnosed as *Plasmodium malariae*, a morphologically similar but distantly-related species (Lee et al., 2009). Although sporadic human cases had been described in the 1960s, the public health importance of *P. knowlesi* was first understood with the reporting of a substantial focus of human infections in Malaysian Borneo in 2004 (Singh et al., 2004). Molecular detection has confirmed that human cases of *P. knowlesi* infection are relatively common in eastern Malaysia and occur in other southeastern Asian countries including The Philippines, Indonesia, Vietnam and Myanmar (Kantele and Jokiranta, 2011). The geographical distribution of *P. knowlesi* appears to be constrained by the range of its natural macaque hosts and the Leucosphyrus mosquito group vector (Moyes et al., 2014). There is no evidence of sig-

nificant human-to-human transmission of *P. knowlesi* (Millar and Singh, 2015). Compared with other *Plasmodium* spp., the field of *P. knowlesi* genomics has been understudied, with only one reference genome available. The sequencing of the *P. knowlesi* reference “H-Pk1 (A+) clone” (PKNH, 14 chromosomes, 23.5 Mb, 5188 genes, 37.5% GC content) (Pain et al., 2008) has provided insights into novel genomic features, including highly variable *kir* and *sicavar* protein families, but is incomplete. Genetic diversity among *P. knowlesi* isolates is high compared with other members of the genus. Genome variation in this species exhibits dimorphism (Pinheiro et al., 2015), and there is evidence this may be driven by partitioning between the two distinct macaque hosts, *M. fascicularis* and *M. nemestrina* (Assefa et al., 2015).

Recently, the *P. knowlesi* A1.H1 (PkA1-H.1) line was the first to be successfully adapted to continuous culture in human erythrocytes, providing an in vitro model suitable for genetic modification (Moon et al., 2013). To support in vitro studies with this human-adapted clonal line we have assembled a new reference genome for the PkA1-H.1 line using PacBio (Pacific Biosciences Inc., USA) RS-II long read and Illumina HiSeq short read sequence data. An advantage of PacBio RS-II SMRT cell sequencing is the potential

* Corresponding author at: Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom.

E-mail address: taane.clark@lshtm.ac.uk (T.G. Clark).

to identify modifications on individual nucleotide bases, and thereby provide insights into methylation sites (Morgan et al., 2016). Cytosine and adenine DNA methylation is an epigenetic mark in most eukaryotic cells that regulates numerous processes including gene expression and stress responses. Genome-wide analysis of DNA methylation in *Plasmodium falciparum* has mapped the positions of methylated cytosines. This work has identified a single functional DNA methyltransferase, *PfDNMT* (PF3D7_0727300), which may mediate these genomic modifications (Ponts et al., 2013) and is thus a potential target for anti-malarial drugs. Analyses have revealed that the malaria genome is asymmetrically methylated, in which only one DNA strand is methylated, and this could regulate virulence gene expression and transcription elongation (Ponts et al., 2013). Using PacBio RS-II data, we describe for the first known time, over 40 000 potential modified bases in the *P. knowlesi* genome, including ~5% that were specifically pinpointed as 6-methyladenine modifications, which recently have been shown to have a role in epigenetic regulation of gene expression in other eukaryotic organisms (Greer et al., 2016). Both the PkA1.H-1 reference genome and the associated methylation data are available to support future in vitro and in vivo studies of *P. knowlesi* parasites, and assist the continuing development of anti-malarial drugs and vaccines.

DNA of high quality and molecular weight was purified from phenol chloroform extraction of magnetic activated cell sorted column enriched PkA1-H.1 schizonts (Bates et al., 2010). The DNA (20 µg, A260/280:1.98) was used to prepare 20 kb insert libraries, which were sequenced using nine SMRT cells on the PacBio RS-II device (Pacific Biosciences Inc.) at the Genome Institute of Singapore. The sequencing yielded a total of 365,956 reads with a mean length of 9645 bp. These data were complemented by raw sequences from the Illumina HiSeq2500 platform (500 bp fragment, 150 bp paired end reads), performed at King Abdullah University of Science and Technology, Kingdom of Saudi Arabia, and yielded in excess of 5 million reads. PacBio data were available for the A1-H.1 clone, but were insufficient in themselves to lead to a complete genome i.e. single contig chromosomes (Moon et al., 2016). The de novo assembly of the reads using the HGAP3 software pipeline within the SMRT Portal (Koren et al., 2012) yielded a total of 111 contigs with a theoretical ~145-fold coverage. These contigs were then used as a reference to map with the *bwa-mem* aligner (Li and Durbin, 2009) almost 5 million PkA1-H.1 Illumina 150 bp reads, leading to 7120 single nucleotide polymorphisms (SNPs) and insertion and deletion (indel) corrections. The high quality corrected contigs were ordered with Abacas software (<http://abacas.sourceforge.net>) using the current *Plasmodium knowlesi* H strain (PKNH) reference (version 2.0, www.genedb.org), and manually checked to remove possible errors. This led to 14 complete chromosomes (number of contigs: range 1–5) and mitochondrial (two contigs, copy number seven-fold), and apicoplast (one contig, 1.8 copies) genomes. The PkA1-H.1 genome was then annotated using the Companion webserver (<https://companion.sanger.ac.uk>).

The resulting PkA1-H.1 reference covers 98.2% of the PKNH (v2) reference and 100% of a smaller draft assembly for this line (Moon et al., 2016), and contains only 42 gaps. Compared with the PKNH (v2), PkA1-H.1 has 3993 SNP and 19,936 indel differences. The new genome improved characterisation of *kir* and *sicavar* genes, leading to a total genome length of 24.4 Mb (chromosome size range: 726,979–3,301,832). To assess the performance of the PkA1-H.1 reference, we aligned whole genome sequence data from 60 published human *P. knowlesi* isolates (Assefa et al., 2015; Pinheiro et al., 2015). Raw Illumina sequence data were downloaded for 60 *P. knowlesi* isolates (Assefa et al., 2015; Pinheiro et al., 2015) (Supplementary Table S1). The samples were aligned against the PkA1-H.1 reference using *bwa-mem* (Li and Durbin, 2009) and SNPs

were called using the *Samtools* software suite (Li, 2011), from which a set of high quality SNPs was filtered using previously described methods (Samad et al., 2015; Campino et al., 2016). The alignment process yielded an average coverage of ~143-fold across 99% of the reference genome, and 1,632,024 high quality SNPs (one every 15 bp) were characterised (Fig. 1). These coverage statistics were marginally superior to mapping to PKNH, which yielded average coverage of ~140-fold across 96% of the reference genome (Supplementary Table S1). The improvement includes an increase in coverage in the *kir* and *sicavar* genes, and the closing of several gaps in these genes that exist in the PKNH reference.

The 38-fold sequence coverage obtained during the new sequencing was in excess of the recommended 25-fold for the 6-methyladenine (m6A) and 4-methylcytosine (m4C) modification detection (Morgan et al., 2016). The *RS_Modification_and_Motif_analysis.1* program within the SMRT Analysis Portal was used to perform a methylation analysis (Morgan et al., 2016). Overall, we have identified 41,508 potential modified bases (*P* values <0.01) in the *P. knowlesi* genome (by chromosome: mean 2965; range 1089–5823) (Supplementary Table S2). Of the total, 7231 modifications (~17%) were identified from an independent unanalysed dataset (Moon et al., 2016). Approximately five percent (2218) of the modifications were specifically pinpointed as m6A modifications, a type that has been recently confirmed to play a role in epigenetic regulation of eukaryotic organisms such as *Caenorhabditis elegans* (Greer et al., 2016). Furthermore, 3646 (~9%) modified bases were classified as m4C methylation events. The proportion of the total adenosine and cytosine bases that are modified is 0.11 and 0.25, respectively. These fractions are lower than those reported in a study of *P. falciparum* for 5-methylcytosine (m5C), which estimated that two-thirds of the cytosine bases were methylated (Ponts et al., 2013). Analysis of the distribution of the methylation sites across the genome revealed that 45.3% of the m6A sites are located within gene boundaries. However, the m4C modifications were distributed more evenly, with 50.1% being intragenic (m4C versus m6A, *P* < 0.0004). We calculated the number of modified bases over a 10 kb sliding window, revealing a stable distribution over the different chromosomes (modified bases/kb: mean 1.6; range 1.38–1.81) (Fig. 1). In order to identify genomic regions (islands) with an accumulation of modified bases, the fold change of modified bases per window was compared with the chromosome average. This analysis revealed 85 10 kb regions that present at least an increase of two-fold change over the chromosome average in either of the two PacBio datasets analysed (Table 1, Supplementary Table S3). These regions included four methyltransferase genes, and several loci involved in the ribosome constitution and with DNA/RNA manipulation functionality. Further, the *reticulocyte binding protein NBPA* gene (PKNH_1472300) was also identified, which has been demonstrated to be essential for the infection of human red blood cells in this *P. knowlesi* line (Moon et al., 2016).

The dominant cause of malaria in Malaysia is now *P. knowlesi*. To assist with disease control, a deeper understanding of the biology of this neglected parasite is required, and genomics has the potential to provide useful insights. The use of state-of-the-art technology such as PacBio RS-II SMRT cell sequencing has assisted with the biological understanding of constantly changing parasite populations (Ahmed et al., 2016). Using sequence data from PacBio RS-II and Illumina HiSeq2500 technologies, we have constructed a reference genome for the PkA1-H.1 line, which is the first known *P. knowlesi* line to be successfully adapted to continuous culture in human erythrocytes. The PkA1-H.1 reference is complete to a chromosome level, base-corrected, fully annotated, and spans over 98% of the PKNH “H strain” reference, including the closure of some of its gaps in highly variable *kir* and *sicavar* genes. The completeness of the PkA1-H.1 genome improved alignment of sequences and variant detection from published *P. knowlesi* clinical strains, when

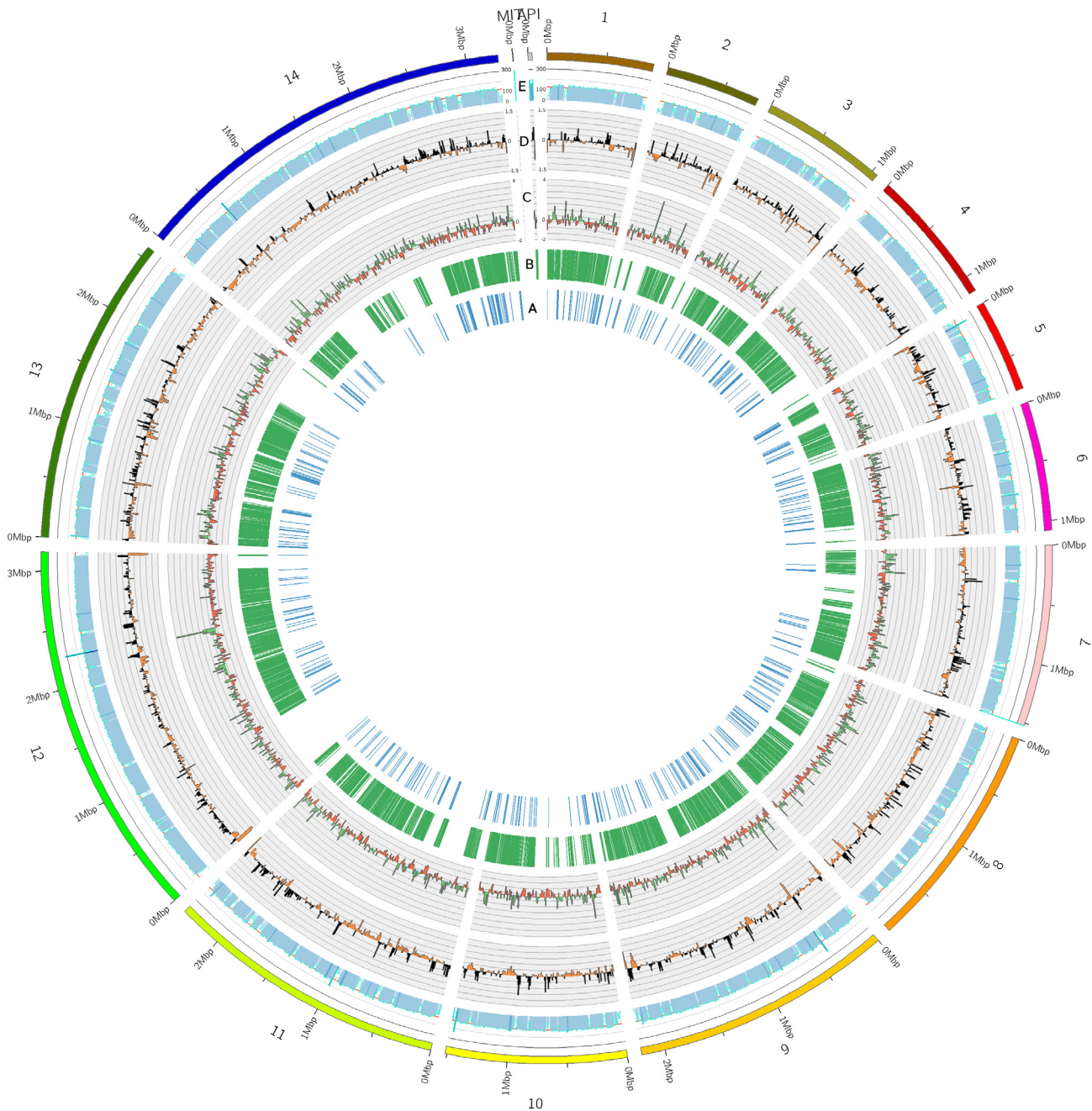


Fig. 1. Analyses using the assembled *Plasmodium knowlesi* A1-H.1 (PKA1-H.1) genome. (A) The 3993 single nucleotide polymorphism differences with PKNH *Plasmodium knowlesi* H strain version 2. (B) The 19,936 insertion and deletion (indel) differences with PKNH version 2. (C) The distribution of the methylated bases per window. (D) Average coverage from mapping 60 publically available samples, using a non-overlapping 10 kb sliding window. (E) The single nucleotide polymorphism density as the fold change of number of single nucleotide polymorphism positions compared with the chromosome average.

compared with using the PKNH genome reference (Supplementary Table S1). The long-read sequencing technology combined with Illumina paired sequences effectively resolved gaps in the genome caused by low complexity regions and large multigene families such as the *sicavar* genes. A more complete genome would assist population genomic studies of genes critical to host–pathogen interactions and virulence. This could include the use of field isolates from across southeastern Asia to investigate host–parasite population structure and to develop molecular barcodes for surveillance (Preston et al., 2014).

The use of the PacBio RS-II technology allowed us to describe for the first time the distribution of methylated bases, particularly m4C and m6A modifications, with single base resolution in the *P. knowlesi* genome. Whilst m4C modifications are usually confined

to prokaryotes, m6A modifications have been previously described as having a role as a transcription regulator in eukaryotes such as *C. elegans* (Greer et al., 2016). The contribution of m6A modifications to epigenetic control of gene expression could be investigated by integration of genomics and methylation with RNAseq transcriptome studies. Concordance of modifications from the same *P. knowlesi* lines identified in PacBio sequence data obtained in an earlier study (Moon et al., 2016) was modest (~20%), but this may be expected as the DNA extracted for sequencing was not derived from synchronised identical parasite cultures. The regions in the genome with the highest density of modified bases included four methyltransferase genes (PKNH_0103500, PKNH_0211900, PKNH_0305100 and PKNH_1416400), which could suggest a role of epigenetic modifications in the regulation of methylation path-

Table 1Regions with putative methylation sites in *Plasmodium knowlesi* line A1-H.1.

Chr.	Window	Fold-change ^a	Gene IDs (<i>PKNH_</i>) ^b
1	90001–100000	2.39, 1.08	0101800, 0101900, 0102000, 0102100
1	220001–230000	2.92, 1.40	0103500, 0103600, 0103700
1	300001–310000	2.16, 1.68	0105200, 0105700
1	410001–420000	2.22, 1.01	0108300, 0108400
1	810001–820000	2.34, 1.25	0117200, 0117400
2	480001–490000	2.42, 1.29	0210600, 0210700, 0210800
2	540001–550000	0.75, 2.00	0211700, 0211800, 0211900, 0212000
3	230001–240000	2.30, 1.26	0304200, 0304300, 0304400
3	260001–270000	2.24, 1.02	0304900, 0305000, 0305100, 0305200
3	310001–320000	2.66, 1.55	0306400, 0306500
3	460001–470000	2.54, 1.66	0310200, 0310300, 0310400
3	750001–760000	2.96, 1.81	0315000, 0315100, 0315200
4	160001–170000	2.39, 0.73	0404100, 0404200
4	170001–180000	2.39, 0.81	0404300, 0404400
4	780001–790000	2.33, 1.07	0418100, 0418200, 0418300, 0418400
4	1010001–1020000	0.69, 2.17	0422400, 0422500, 0422600
5	110001–120000	2.10, 1.49	0502300
5	230001–240000	1.09, 2.01	0505400, 0505500, 0505600
5	570001–580000	2.03, 1.28	0512300, 0512400, 0512500
5	600001–610000	2.32, 1.07	0512700, 0512800
5	710001–720000	1.02, 2.13	0514600, 0514700, 0514800, 0514900, 0515000, 0515100
6	220001–230000	2.66, 0.86	0605300, 0605400, 0605500
6	570001–580000	2.01, 0.98	0612900, 0613000
7	280001–290000	2.98, 1.99	0705200, 0705300, 0705400
7	1180001–1190000	2.36, 1.09	0726600, 0726700, 0726800, 0726900, 0727000
8	240001–250000	2.25, 1.17	0804700, 0804800, 0804900
8	350001–360000	0.81, 2.55	0807600
8	500001–510000	2.06, 1.56	0811300, 0811400, 0811500
8	600001–610000	2.44, 1.29	0813500, 0813600, 0813700
8	1090001–1100000	2.19, 1.00	0823500, 0823600
8	1400001–1410000	2.19, 1.59	0830400, 0830500, 0830600, 0830700
8	1680001–1690000	2.75, 1.70	0837400, 0837500, 0837600
9	270001–280000	2.21, 1.45	0905300, 0905400
9	720001–730000	2.21, 0.83	0915800, 0915900, 0916000
9	830001–840000	2.10, 0.90	0918700, 0918800, 0918900, 0919000
9	1440001–1450000	1.16, 2.04	0932400
9	1660001–1670000	2.10, 0.95	0937000
9	1960001–1970000	2.10, 1.38	0943000, 0943100
9	2010001–2020000	2.71, 1.16	0944300
9	2110001–2120000	2.55, 1.36	0946100
9	2130001–2140000	2.44, 0.49	0946200
10	540001–550000	2.15, 1.06	1011700, 1011800
10	1090001–1100000	2.38, 1.29	1023800, 1023900, 1024000
10	1390001–1400000	2.38, 1.52	1030600, 1030700
10	1410001–1420000	2.26, 1.06	1031200, 1031300, 1031400, 1031500
10	1420001–1430000	2.21, 1.60	1031500, 1031600, 1031700, 1031800
11	300001–310000	2.30, 1.59	1107200, 1107300, 1107400
11	320001–330000	2.24, 1.21	1107700, 1107800, 1107900, 1108000
11	460001–470000	1.19, 2.12	1110700, 1110800
11	500001–510000	2.30, 1.64	1111500, 1111600
11	680001–690000	2.17, 0.94	1114700, 1114800
11	1700001–1710000	2.24, 1.02	1137000, 1137100
11	1710001–1720000	2.04, 1.49	1137200, 1137300, 1137400, 1137500, 1137600
11	2120001–2130000	2.77, 0.73	1144900, 1145000
12	700001–710000	2.23, 1.71	1215700, 1215800, 1215900
12	760001–770000	2.00, 1.06	1217000, 1217200
12	920001–930000	2.45, 1.70	1219600, 1219700, 1219800, 1219900
12	1050001–1060000	2.17, 0.91	1223800
12	1130001–1140000	2.73, 1.06	1225300, 1225400, 1225500, 1225600
12	1730001–1740000	2.39, 1.05	1239200, 1239300, 1239400
12	1750001–1760000	2.06, 0.87	1239600, 1239700
12	1760001–1770000	2.06, 1.49	1239700, 1239800
12	1810001–1820000	2.00, 0.91	1241000
12	2200001–2210000	2.34, 1.02	1248600, 1248700, 1248800
12	2230001–2240000	2.28, 0.87	1249300, 1249400, 1249500
12	2240001–2250000	2.23, 1.19	1249500, 1249600
12	2630001–2640000	2.00, 1.31	1257800, 1257900, 1258000
13	350001–360000	2.36, 1.32	1307500, 1307600, 1307700, 1307800, 1307900
13	400001–410000	2.01, 0.89	1308600, 1308700
13	1100001–1110000	2.54, 0.57	1324500, 1324600
13	1450001–1460000	2.12, 1.06	1331100, 1331200, 1331300
13	1650001–1660000	2.89, 0.72	1336600, 1336900, 1337000
13	1860001–1870000	2.18, 1.57	1341400, 1341500
13	1920001–1930000	2.12, 0.71	1342600

Table 1 (continued)

Chr.	Window	Fold-change ^a	Gene IDs (PKNH_) ^b
13	2250001–2260000	2.60, 1.38	1349600, 1349700
13	2340001–2350000	1.18, 2.10	1351200, 1351300, 1351400
13	2380001–2390000	2.89, 1.43	1352100, 1352200, 1352300
13	2400001–2410000	2.24, 0.99	1352500, 1352600, 1352800
13	2420001–2430000	2.30, 1.28	1353100, 1353200, 1353300
14	740001–750000	2.25, 1.18	1416300, 1416400, 1416500
14	900001–910000	2.19, 1.26	1420200, 1420300, 1420400
14	2000001–2010000	1.68, 0.72	1445800, 1445900, 1446000
14	2830001–2840000	2.19, 1.23	1463300
14	3020001–3030000	2.19, 0.80	1467900, 1468000, 1468100, 1468200
14	3200001–3210000	2.64, 1.04	1472300, 1472400

Chr, chromosome number; PKNH, *Plasmodium knowlesi* H strain.

^a Deviation from chromosomal mean: this study, data analysed from Moon et al. (2016).

^b Bold indicates methylation related annotation.

ways. We also observed a wide range of ribosomal proteins and genes involved in manipulation of the genetic material. Further, the *reticulocyte binding protein NBPXa* gene (PKNH_1472300), which is essential for the infection of human erythrocytes in the A1-H.1 line, was one of the genes showing highest density of modifications in the genome. The modifications associated with the *P. knowlesi* orthologue for the *P. falciparum* mediator *PfDNMT* (PKNH_0211900) were not detected, but are thought to create m5C modifications. It is not possible to detect m5C modifications described in other *Plasmodium* spp., due to limitations in sequencing platform coverage. At least 250× coverage would be required to confidently detect m5C modifications, although some uncharacterised modifications might refer to m5C methylation events (Morgan et al., 2016).

In summary, we have provided a genomic reference and methylation data for the PkA1-H.1 line for researchers to undertake biological and clinical research into *P. knowlesi* and other malaria parasites, potentially assisting with the design of anti-malarial drugs and vaccines, and diagnostic tools.

The reference genome and methylome are available from pathogenseq.lshmt.ac.uk/knowlesi_1/. The underlying raw sequence data is available from the European Nucleotide Archive (accession numbers: PRJEB19298, ERS763679).

Acknowledgements

TGC is funded by the Medical Research Council (MRC), UK (Grant no. MR/K000551/1, MR/M01360X/1, MR/N010469/1). SC and CR are funded by the MRC (Grant no. MR/M01360X/1). RWM is supported by an MRC Career Development Award jointly funded by the MRC and UK Department for International Development. This work was supported in part by the Francis Crick Institute, UK which receives its core funding from Cancer Research, UK (FC001097, FC001043), the MRC (FC001097, FC001043), and the Wellcome Trust, UK (FC001097, FC001043). We gratefully acknowledge the Scientific Computing Group for data management and computer infrastructure at Genome Institute of Singapore for their help. Bioinformatic analysis was performed on the MRC, UK, funded eMedLab computing resource.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ijpara.2017.09.008>.

References

Ahmed, M.A., Fong, M.Y., Lau, Y.L., Yusof, R., 2016. Clustering and genetic differentiation of the normocyte binding protein (nbpXa) of *Plasmodium*

- knowlesi* clinical isolates from Peninsular Malaysia and Malaysia Borneo. *Malar. J.* 15, 241.
- Assefa, S., Lim, C., Preston, M.D., Duffy, C.W., Nair, M.B., Adroub, S.A., Kadir, K.A., Goldberg, J.M., Neafsey, D.E., Divis, P., Clark, T.G., Duraisingh, M.T., Conway, D.J., Pain, A., Singh, B., 2015. Population genomic structure and adaptation in the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc. Nat. Acad. Sci. U.S.A.* 112, 13027–13032.
- Bates, A.H., Mu, J., Jiang, H., Fairhurst, R.M., Su, X.Z., 2010. Use of magnetically purified *Plasmodium falciparum* parasites improves the accuracy of erythrocyte invasion assays. *Exp. Parasitol.* 126, 278–280.
- Campino, S., Benavente, E.D., Assefa, S., Thompson, E., Drought, L.G., Taylor, C.J., Corvett, Z., Carret, C.K., Flueck, C., Ivens, A.C., Kwiatkowski, D.P., Alano, P., Baker, D.A., Clark, T.G., 2016. Genomic variation in two gametocyte non-producing *Plasmodium falciparum* clonal lines. *Malar. J.* 15, 229.
- Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corralles, D., Hsu, C.H., Aravind, L., He, C., Shi, Y., 2016. DNA methylation on N6-adenine in *C. elegans*. *Cell* 161, 868–878.
- Kantele, A., Jokiranta, T.S., 2011. Review of cases with the emerging fifth human malaria parasite, *Plasmodium knowlesi*. *Clin. Infect. Dis.* 52, 1356–1362.
- Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., Adam, M.P., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700.
- Lee, K.S., Cox-Singh, J., Singh, B., 2009. Morphological features and differential counts of *Plasmodium knowlesi* parasites in naturally acquired human infections. *Malar. J.* 8, 73.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25, 1754–1760.
- Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993.
- Millar, S., Singh, B., 2015. Human infections with *Plasmodium knowlesi*—zoonotic malaria. *Clin. Microbiol. Infect.* 21, 640–648.
- Moon, R.W., Sharaf, H., Hastings, C.H., Ho, Y.S., Nair, M.B., Rchiad, Z., Nair, M.B., Rchiad, Z., Knuepfer, E., Ramaprasad, A., Mohring, F., Amir, A., Yusuf, N.A., Hall, J., Almond, N., Lau, Y.L., Pain, A., Blackman, M.J., Holder, A.A., 2016. Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7231–7236.
- Moon, R.W., Hall, J., Rangkuti, F., Ho, Y.S., Almond, N., Mitchell, G.H., Pain, A., Holder, A.A., Blackman, M.J., 2013. Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 531–536.
- Morgan, R.D., Luyten, Y.A., Johnson, S.A., Clough, E.M., Clark, T.A., Roberts, R.J., 2016. Novel m4C modification in type I restriction-modification systems. *Nucleic Acids Res.* 44, 9413–9425.
- Moyes, C.L., Henry, A.J., Golding, N., Huang, Z., Singh, B., Baird, J.K., Newton, P.N., Huffman, M., Duda, K.A., Drakeley, C.J., Elyazar, I.R., Anstey, N.M., Chen, Q., Zimmers, Z., Bhatt, S., Gething, P.W., Hay, S.I., 2014. Defining the geographical range of the *Plasmodium knowlesi* reservoir. *PLoS Negl. Trop. Dis.* 8, e2780.
- Pain, A., Böhme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., Mourier, T., Mistry, J., Pasini, E.M., Pasini, E.M., Aslett, M.A., Balasubramanian, S., Borgwardt, K., Brooks, K., Carret, C., Carver, T.J., Cherevach, I., Chillingworth, T., Clark, T.G., Galinski, M.R., Hall, N., Harper, D., Harris, D., Hauser, H., Ivens, A., Janssen, C.S., Keane, T., Larke, N., Lapp, S., Marti, M., Moule, S., Meyer, I.M., Ormond, D., Peters, N., Sanders, M., Sanders, S., Sargeant, T.J., Simmonds, M., Smith, F., Squares, R., Thurston, S., Tivey, A.R., Walker, D., White, B., Zuiderwijk, E., Churcher, C., Quail, M.A., Cowman, A.F., Turner, C.M., Rajandream, M.A., Kocken, C.H., Thomas, A.W., Newbold, C.I., Barrell, B.G., Berriman, M., 2008. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* 455, 799–803.
- Pinheiro, M.M., Ahmed, M.A., Millar, S.B., Sanderson, T., Otto, T.D., Lu, W.C., Krishna, S., Rayner, J.C., Cox-Singh, J., 2015. *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. *PLoS One* 10, e0121303.

- Ponts, N., Fu, L., Harris, E.Y., Zhang, J., Chung, D.-W.D., Cervantes, M.C., Prudhomme, J., Atanasova-Penichon, V., Zehraoui, E., Bunnik, E.M., Rodrigues, E.M., Lonardi, S., Hicks, G.R., Wang, Y., Le Roch, K.G., 2013. Genome-wide mapping of DNA methylation in the human malaria parasite *Plasmodium falciparum*. *Cell Host Microbe* 14, 696–706.
- Preston, M.D., Campino, S., Assefa, S.A., Echeverry, D.F., Ocholla, H., Amambua-Ngwa, A., Stewart, L.B., Conway, D.J., Borrmann, S., Michon, P., Zongo, I., Ouédraogo, J.B., Djimde, A.A., Doumbo, O.K., Nosten, F., Pain, A., Bousema, T., Drakeley, C.J., Fairhurst, R.M., Sutherland, C.J., Roper, C., Clark, T.G., 2014. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. *Nat. Commun.* 5, 4052.
- Samad, H., Coll, F., Preston, M.D., Ocholla, H., Fairhurst, R.M., Clark, T.G., 2015. Imputation-based population genetics analysis of *Plasmodium falciparum* malaria parasites. *PLoS Genet.* 11, e1005131.
- Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., Thomas, A., Conway, D.J., 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* 363, 1017–1024.
- Singh, B., Daneshvar, C., 2013. Human infections and detection of *Plasmodium knowlesi*. *Clin. Microbiol. Rev.* 26, 165–184.